An improved procedure for orthogonalising the search vectors in Rosenbrock's and Swann's direct search optimisation methods

By J. R. Palmer*

An improved procedure is presented for generating orthogonal search vectors for use in Rosenbrock's and Swann's optimisation methods. The new procedure shows considerable savings in time and in storage requirements, and deals more satisfactorily with certain cases in which the original method fails.

(Received July 1968)

1. The Rosenbrock and Swann procedure

In the methods described by Rosenbrock (1960) and Swann (1964) for direct search optimisation of a function of n variables, local minima or maxima are sought by conducting univariate searches parallel to each of the n orthogonal unit vectors $\xi_1^0, \xi_2^0, \ldots, \xi_n^0$ in turn, the distances moved in these directions being $d_1, d_2, \ldots, d_n$ respectively. The set of n such searches constitutes one 'stage' of the calculation.

For the next stage, subject to certain restrictions which need not be considered here, a new set of n orthogonal unit vectors $\xi_1^1, \xi_2^1, \ldots, \xi_n^1$ is generated, such that ξ_1^1 lies along the direction of greatest advance for the previous stage, i.e. along the line joining the first and last points of that stage in n-dimensional space. For this purpose, Rosenbrock proposed (and Swann also used) the following calculating sequence:

$$A_k = \sum_{i=k}^{n} d_i \xi_i^0 \tag{1}$$

$$B_k = A_k - \sum_{j=1}^{k-1} (A_k . \xi_j^1)\xi_j^1 \tag{2}$$

$$\xi_k^1 = B_k/|B_k|. \tag{3}$$

Evidently the A_k are obtained by starting with the 'greatest advance' vector A_1, as defined above, and removing from it the successive orthogonal advance vector components $d_i\xi_i^0$. The B_k are derived from the corresponding A_k by removing the components of A_k parallel to all the previously determined ξ_j^1, so that the B_k are mutually orthogonal. Then by dividing each B_k by its modulus, the corresponding unit vector ξ_k^1 is obtained.

2. Failure of the procedure

Swann showed that this procedure breaks down if any of the d_i, for instance d_p (where $1 \leqslant p < n$), is zero. Under these circumstances

$$A_p = \sum_{i=p}^{n} d_i \xi_i^0 = \sum_{i=p+1}^{n} d_i \xi_i^0 = A_{p+1}$$

and $$B_p = A_p - \sum_{j=1}^{p-1} (A_p . \xi_j^1)\xi_j^1$$

and $$B_{p+1} = A_{p+1} - \sum_{j=1}^{p} (A_{p+1} . \xi_j^1)\xi_j^1$$

$$= A_p - \sum_{j=1}^{p-1} (A_p . \xi_j^1)\xi_j^1 - (A_p . \xi_p^1)\xi_p^1$$

whence

$$B_{p+1} = |B_{p+1}|\xi_{p+1}^1 = B_p - (A_p . \xi_p^1)\xi_p^1$$
$$= \{|B_p| - A_p . \xi_p^1\}\xi_p^1. \tag{4}$$

But ξ_{p+1}^1 and ξ_p^1 are orthogonal, from which it follows that

$$B_{p+1} = |B_{p+1}| = |B_p| - A_p . \xi_p^1 = 0$$

so that $\xi_{p+1}^1 = B_{p+1}/|B_{p+1}|$ is undetermined.

In the special case $d_n = 0$ we have

$$A_n = B_n = |B_n| = 0$$

so that $\xi_n^1 = B_n/|B_n|$ is undetermined.

Rosenbrock avoided this difficulty by ensuring that none of the d_i could become zero. In Swann's method, however, one or more of the d_i may become zero: to avoid the trouble described above, the components of the A_k are reordered so as to place those d_i whose values are zero (q in number, say) at the end of the list, and the procedure is then applied only to the first $(n - q)$ components. Swann showed that this still produces a strictly orthogonal set of ξ_k^1 if the d_i concerned are exactly zero, and that if a d_i is taken as zero when its modulus is less than some small quantity (10^{-6}, say), the resulting lack of orthogonality is very small (the scalar products of nominally orthogonal vectors being of the order of 10^{-16}).

3. A new approach to the failing case

It occurred to the present author that it might happen, if B_{k+1} and its modulus were evaluated, that they would each prove to be proportional to d_k, so that in evaluating $\xi_{k+1}^1 = B_{k+1}/|B_{k+1}|$ the quantity d_k would cancel, leaving ξ_{k+1}^1 determinate even if $d_k = 0$, and this was found to be the case, subject to certain reservations.

Thus from (1), (2) and (3) above,

$A_1 = \sum_1 d_i\xi_i^0 = B_1$ (where $\sum_1$ denotes $\sum_{i=1}^{n}$, and correspondingly for other sums)

$$\therefore \quad |B_1| = \sqrt{(\textstyle\sum_1 d_i^2)} = |A_1|$$

$$\therefore \quad \xi_1^1 = A_1/|A_1| = \textstyle\sum_1 d_i\xi_i^0/\sqrt{(\sum_1 d_i^2)}. \tag{5}$$

* *Department of Aircraft Propulsion, The College of Aeronautics, Cranfield, Bedford*

Also, $A_2 = \sum_2 d_i \xi_i^0$

$$\therefore \quad B_2 = \sum_2 d_i\xi_i^0 - (\sum_2 d_i\xi_i^0).\left\{\frac{\sum_1 d_i\xi_i^0}{\sqrt{(\sum_1 d_i^2)}}\right\}\left\{\frac{\sum_1 d_i\xi_i^0}{\sqrt{(\sum_1 d_i^2)}}\right\}$$

$$= \sum_2 d_i\xi_i^0 - \frac{(\sum_2 d_i^2)(\sum_1 d_i\xi_i^0)}{\sum_1 d_i^2}$$

which reduces to

$$B_2 = \frac{d_1^2 \sum_2 d_i\xi_i^0 - d_1\xi_1^0 \sum_2 d_i^2}{\sum_1 d_i^2}$$

$$= \frac{(|A_1|^2 - |A_2|^2)A_2 - (A_1 - A_2)|A_2|^2}{|A_1|^2}$$

$$= \frac{A_2|A_1|^2 - A_1|A_2|^2}{(A_1|^2} \tag{6}$$

$$\therefore \quad |B_2| = \frac{d_1}{\sum_1 d_i^2}\sqrt{[d_1^2 \sum_2 d_i^2 + (\sum_2 d_i^2)^2]}$$

$$= d_1\sqrt{\left(\frac{\sum_2 d_i^2}{\sum_1 d_i^2}\right)} = \frac{|A_2|}{|A_1|}\sqrt{(|A_1|^2 - |A_2|^2)} \tag{7}$$

$$\therefore \quad \xi_2^1 = \frac{d_1 \sum_2 d_i\xi_i^0 - \xi_1^0 \sum_2 d_i^2}{\sqrt{(\sum_1 d_i^2 \sum_2 d_i^2)}}$$

$$= \frac{A_2|A_1|^2 - A_1|A_2|^2}{|A_1||A_2|\sqrt{(|A_1|^2 - |A_2|^2)}} \tag{8}$$

If $d_1 = 0$, these expressions give $B_2 = |B_2| = 0$ and $\xi_2^1 = -\xi_1^0$ i.e. ξ_2^1 remains determinate (unless $\sum_2 d_i^2 = 0$).

Similar results are obtained for B_3, $|B_3|$ and ξ_3^1, from which it appears that in general

$$B_k = \frac{d_{k-1}^2 \sum_k d_i\xi_i^0 - d_{k-1}\xi_{k-1}^0 \sum_k d_i^2}{\sum_{k-1} d_i^2}$$

$$= \frac{A_k|A_{k-1}|^2 - A_{k-1}|A_k|^2}{|A_{k-1}|^2} \tag{9}$$

and $$|B_k| = d_{k-1}\sqrt{\left(\frac{\sum_k d_i^2}{\sum_{k=1} d_i^2}\right)}$$

$$= \frac{|A_k|}{|A_{k-1}|}\sqrt{(|A_{k-1}|^2 - |A_k|^2)} \tag{10}$$

and $$\xi_k^1 = \frac{d_{k-1} \sum_k d_i\xi_i^0 - \xi_{k-1}^0 \sum_k d_i^2}{\sqrt{(\sum_{k-1} d_i^2 \sum_k d_i^2)}}$$

$$= \frac{d_{k-1}A_k - \xi_{k-1}^0|A_k|^2}{|A_{k-1}|\,|A_k|}$$

$$= \frac{A_k|A_{k-1}|^2 - A_{k-1}|A_k|^2}{|A_{k-1}||A_k|\sqrt{(|A_{k-1}|^2 - |A_k|^2)}} \tag{11}$$

for $2 \leqslant k \leqslant n$
and that if $d_{k-1} = 0$, then $\xi_k^1 = -\xi_{k-1}^0$ (unless $\sum_k d_i^2 = 0$).

An inductive proof of the validity of equation (9), and thence of (10) and (11), is given in the Appendix.

It should be noted that, in the particular case $k = n$, (11) gives

$$\xi_n^1 = \frac{d_{n-1}d_n\xi_n^0 - \xi_{n-1}^0 d_n^2}{\sqrt{[(d_{n-1}^2 + d_n^2)d_n^2]}} = \frac{d_{n-1}\xi_n^0 - d_n\xi_{n-1}^0}{\sqrt{[d_{n-1}^2 + d_n^2]}}$$

so that if $d_{n-1} = 0$, $\xi_n^1 = -\xi_{n-1}^0$ (unless $d_n = 0$) (special case of the above) or if $d_n = 0$, $\xi_n^1 = \xi_n^0$ (unless $d_{n-1} = 0$).

Thus it is seen that the ξ_k^1 remain determinate, even if one or more of the d_{k-1} are zero, provided only that $\sum_{i=k}^n d_i^2 \neq 0$. This suggests using equations (5) and (11) to evaluate the ξ_k^1 directly, subject only to a check that $\sum_{i=k}^n d_i^2 \neq 0$, and that the components should *not* be reordered.

It would also appear that this procedure might result in a considerable saving both in arithmetic operations and in working stage requirements, and it will now be demonstrated that this is so.

4. Comparison of the speed and storage requirements of the two procedures

Assuming that the d_k, ξ_k and A_k are already stored in the real arrays *d*[*k*], *xi* [*k*, *i*] and *A*[*k*, *i*] respectively, and that the real array *t*[*k*] has been declared to store $|A_k|^2$ and the real variable *div* to store $|A_{k-1}||A_k|$, the above procedure is described by the following sequence of Algol statements:

```
t[n] := d[n]↑2;
for k := n − 1 step − 1 until 1 do
t[k] := t[k + 1] + d[k]↑2;
for k := n step − 1 until 2 do
  begin div := sqrt (t[k − 1] × t[k]);
    if div ≠ 0·0 then for i := 1 step 1 until n do
      xi[k, i] := (d[k − 1] × A[k, i] − xi[k − 1, i]
                                             × t[k])/div
  end;
div := sqrt (t[1]);
for i:= 1 step 1 until n do xi[1, i] := A[1, i]/div;
```

Since the calculated ξ_k^1 overwrite the previous ξ_k^0, this sequence has the effect of putting $\xi_k^1 = \xi_k^0$ if $\sum_{i=k}^n d_i^2 = 0$, in accordance with Swann's procedure. If none of the d_k is zero, the process requires $(n^2 - 1)$ additions, subtractions or transfers, $(2n^2 - 1)$ multiplications, n^2 divisions and n square root determinations, while the working stage requirement is $(n + 1)$ real variables.

The corresponding sequence for Swann's method requires the previously declared real arrays *B*[*k*, *i*] and *dot*[*j*] for B_k and $A_k.\xi_j^1$ respectively, and the real variable *mod* for $|B_k|$, and reads:

```
for k := 1 step 1 until n do
  begin for j := 1 step 1 until k − 1 do
    begin dot[j] := 0.0;
      for i := 1 step 1 until n do
        dot[j] := dot[j] + A[k, i] × xi[j, i]

    end;
  mod := 0.0;
for i:= 1 step 1 until n do
  begin B[k, i] := A[k, i];
```

```
      for j := 1 step 1 until k − 1 do
      B[k, i]:= B[k, i] − dot[j] × xi[j, i];
      mod := mod + B[k, i] ↑ 2
    end;
  mod := sqrt (mod);
  for i := 1 step 1 until n do xi[k, i] := B[k, i]/mod
    end;
```

This requires $n(n + \frac{1}{2})(n + 1)$ additions, etc., n^3 multiplications, n^2 divisions and n square root determinations, while the working stage requirement is $(n^2 + n + 1)$ real variables. These figures neglect the preliminary reordering process which is necessary, since this is approximately counterbalanced by the reduction in the number of ξ_k^1 which then have to be calculated.

The new procedure is thus seen to have considerable advantages over Swann's (which is itself an improvement on Rosenbrock's in respect of the univariate search) in terms of speed, economy of storage, and ability to deal with the case $d_k = 0$. Specifically, the number of additions, etc., and the working storage requirement are reduced by a factor of the order of n, and the number of multiplications by a factor of the order of $n/2$.

Appendix: Inductive proof of equation (9)

If equations (9) and (10) for B_k and $|B_k|$ respectively, and equation (11) for ξ_k^1 are assumed to be valid for a particular value of $k(>1)$, then using the basic equations (2) and (3) and the explicitly derived equation (5) we have

$$\begin{aligned} B_{k+1} &= A_{k+1} - \sum_{j=1}^{k}(A_{k+1}.\xi_j^1)\xi_j^1 \\ &= A_{k+1} - (A_{k+1}.\xi_1^1)\xi_1^1 - \sum_{j=2}^{k}(A_{k+}\xi_{kj}^{\cdot 1})\xi_j^1 \\ &= A_{k+1} - \left(A_{k+1}.\frac{A_1}{|A_1|}\right)\frac{A_1}{|A_1|} \\ &\quad - \sum_{2=j}^{k}\left(A_{k+1}.\frac{A_j|A_{j-1}|^2 - A_{j-1}|A_j|^2}{|A_{j-1}||A_j|\sqrt{(|A_{j-1}|^2 - |A_j|^2)}}\right) \\ &\quad \left(\frac{(A_j|A_{j-1}|^2 - A_{j-1}|A_j|^2}{|A_{j-1}||A_j|\sqrt{(|A_{j-1}|^2 - |A_j|^2)}}\right) \end{aligned}$$

Now
$$\begin{aligned} A_{k+1}.A_1 &= \sum_{k+1} d_i\xi_i^0 . \sum_{1} d_i\xi_i^0 \\ &= \sum_{k+1} d_i^2 = |A_{k+1}|^2 \text{ (since } k > 1) \end{aligned}$$
and similarly
$$\begin{aligned} A_{k+1}.A_j &= A_{k+1}.A_{j-1} \\ &= |A_{k+1}|^2 \text{ (since } k+1 > j). \end{aligned}$$
Hence
$$\begin{aligned} B_{k+1} &= A_{k+1} - \frac{A_1|A_{k+1}|^2}{|A_1|^2} \\ &\quad - \sum_{j=2}^{k}\left\{\frac{|A_{k+1}|^2(|A_{j-1}|^2 - |A_j|^2)}{|A_{j-1}|^2|A_j|^2}\right\} \\ &\quad \times \left\{\frac{A_j|A_{j-1}|^2 - A_{j-1}|A_j|^2}{|A_{j-1}|^2 - |A_j|^2}\right\} \\ &= A_{k+1} - \frac{A_1|A_{k+1}|^2}{|A_1|^2} \\ &\quad - |A_{k+1}|^2\sum_{j=2}^{k}\frac{A_j|A_{j-1}|^2 - A_{j-1}|A_j|^2}{|A_{j-1}|^2|A_j|^2} \\ &= A_{k+1} - |A_{k+1}|^2\left\{\sum_{j=2}^{k}\left(\frac{A_j}{|A_j|^2} - \frac{|A_{j-1}|}{|A_{j-1}|^2}\right) \right. \\ &\quad \left. + \frac{A_1}{|A_1|^2}\right\}. \end{aligned}$$
But
$$\begin{aligned} \sum_{j=2}^{k}\left(\frac{A_j}{|A_j|^2} - \frac{A_{j-1}}{|A_{j-1}|^2}\right) &= \sum_{j=2}^{k}\frac{A_j}{|A_j|^2} \\ &\quad - \sum_{j=1}^{k-1}\frac{A_j}{|A_j|^2} = \frac{A_k}{|A_k|^2} - \frac{A_1}{|A_1|^2}. \end{aligned}$$
Thus
$$B_{k+1} = A_{k+1} - \frac{A_k|A_{k+1}|^2}{|A_k|^2} = \frac{A_{k+1}|A_k|^2 - A_k|A_{k+1}|^2}{|A_k|^2}. \tag{12}$$

Now (12) is formally the same as (9), with k replaced by $(k + 1)$, so that if (9) is valid for a given value of k, it is also valid for the next higher value of k. But we have already shown in equation (6) that (9) is valid for the case $k = 2$, hence (9) is valid for all k such that $2 \leqslant k \leqslant n$, and consequently (10) and (11) are also valid in this range.

References

ROSENBROCK, H. H. (1960). An Automatic Method for finding the Greatest or Least Value of a Function, *Computer Journal* Vol. 4, pp. 175–184.

SWANN, W. H. (1964). *Report on the Development of a new Direct Search Method of Optimisation*, Imperial Chemical Industries Ltd., Central Instrument Laboratory Research Note 64/3.

Book Review

Semi-Groups of Operators and Approximation, by Paul L. Butzer and Hubert Berens, 1967; 318 pages. (Springer-Verlag, $14.)

This book is concerned with the mathematical aspects of semi-group theory and in particular those aspects which are connected in some way or other with approximation. This theory is of significance in our understanding of the underlying theory of such topics as classical approximation theory, the solutions of partial differential equations and the theory of singular integrals, but is somewhat far removed from the everyday needs of the computing fraternity.

Chapter 1 gives a straightforward presentation of the standard theory of semi-groups of operators. Chapter 2 presents basic approximation theorems for semi-group operators with a study in particular of Dirichlet's problem for the unit disc and Fourier's problem of the ring. Chapter 3 is devoted to the incorporation of approximation theorems for semi-group operators into the theory of intermediate spaces (intermediate between the initial Banach space and the domain of definition of the powers of the infinitesimal generator of the semi-group) and to deep generalisations in the new setting. The last chapter outlines and discusses applications of the previous general theory, including the semi-group of left translations, the singular integrals of Abel-Poisson for periodic functions and of Cauchy-Poisson for functions on the real line, and the singular integral of Gauss-Weierstrass on Euclidean n-space in connection with Sobolev and Besov spaces. There is also a helpful appendix summarising the material in functional analysis that is assumed.

J. K. REID (Falmer)