SEE CGTM 147.

Rotraut Weiss
SLAC Computation Group
Stanford, California

CGTM No. 148
April 1973

A COMPARATIVE STUDY OF MULTIDIMENSIONAL FUNCTION OPTIMIZERS

1. Introduction

A frequent application of digital computers is finding the position of the maximum or minimum of a scalar function of several variables. The cost involved in solving these problems is greatly affected by the algorithm used for the search. There are several commonly used algorithms, each one working best on a particular class of problems.

This study is a comparison of the types of minimizer described below.

To describe the performance of an algorithm, it is necessary to consider the reliability (does the result represent a solution?), the accuracy of the extreme point (how close is the found point X to the true local minimum point A?), the accuracy of the extreme value (how close is the found value F(X) to the true minimum value F(A)?), and the efficiency. The efficiency depends on many parameters. Here, the number of object function evaluations will be used as a measure for the efficiency for the following reason: The type of problems encountered frequently in particle physics applications is that in which the calculation of object function to be optimized requires considerable computation. For these cases, the computation done by the optimizer is overshadowed by the number of evaluations of the object function.

2. Assumption

The attitude taken in this study is that of a black box approach - as little user involvement as possible.

For all functions considered, the true minimum is at $A=(A(I))$, where $-1 \leq A(I) \leq 1$. With the black box philosophy in mind, the origin was chosen as starting point and 1 as stepsize, to ensure that the multidimensional interval was covered.

For MIGRAD, the built-in convergence criterion was used. In STREAM and PRAXIS, an accuracy of 0.01 in the independent variable was chosen; the routine that uses PRAXIS is denoted by PRAXIS2. For those users who want to know how much higher the price for higher accuracy is, the routine PRAXIS8 is included, which is PRAXIS with a prescribed accuracy of 10**(-8).

3. The Class of Minimizers Compared

This comparison concentrates on (FORTRAN) routines that do not require first or higher order derivatives of the function to be optimized. The minimizers compared were PRAXIS (Powell's conjugate axes method),[1] STREAM (Rosenbrock method),[2] and MIGRAD from MINUIT (Davidon variable metric method).[3]

4. The Class of Test Functions

The comparison was done for two types of functions, quadratics and Gaussians, all of 5 variables. If H is an orthogonal matrix, and D a diagonal matrix with D(I)=1 for 1 ≤ 4, then the quadratics were of the form

QUADI(X) = (X-A)-transpose * H-transpose * D * H * (X-A)

D(5,5)=10**i for I=0,1,...,5 ,

and the Gaussians of the form

GAUSI(X) = -EXP(-QUADI(X)), I=0,1,2

Except for QUADO AND GAUSO, the functions change in one direction by far more than in the others. For the quadratics, the starting function value is 10**I, which the minimizer has to bring down to 0. The interesting part about the Gaussians is that it is most unlikely to have picked the starting point between the minimum point and the inflection point; this means the curvature of the function at the starting point is most likely to be of wrong sign.

For each type function, 14 cases were run. In 7 of them, H is the identity matrix; that is, there is no correlation between the parameters. In the other 7 cases, H is randomly (but for all minimizers, and all 8 functions equally) chosen, which gives a random correlation between the parameters.

5. Statistics of Test

The enclosed table shows the statistics of the runs. Since the number of evaluations of the function to be minimized is a measure for the efficiency of the minimizer, the range of these numbers for the 14 runs is tabulated under FUNCTION CALLS, MIN-MAX. ERROR(F) denotes the maximum of F(X)-F(A) for all 14 runs; ERROR(X) the maximum of $|X(I)-A(I)|$ for all I for all 14 runs.

6. Summary and Conclusions

In all cases tested, PRAXIS reliably found the minimum within the specified accuracy (ERROR(X) $\leq$ 1.D-2 for PRAXIS2, $\leq$ 1.D-8 for PRAXIS8) - in many cases, even by far more accurately.

Since $-1 \leq A(I) \leq 1$, a return with ERROR(X) $\geq$ 0.5 can hardly be called a solution. This means that STREAM failed for QUAD4, QUAD5, and that MIGRAD failed for QUAD3 through QUAD5 and the Gaussians.

The average number of FUNCTION CALLS required by PRAXIS is always less than that needed for STREAM; and the difference between the two numbers increases rapidly with the stiffness of the problem. The number of calls to the function for PRAXIS2 is never above 238, while STREAM, in one case, needs 14,904 calls. The FUNCTION CALLS for MIGRAD seem very low; in many cases this is caused by MIGRAD giving up because it is unable to handle the problem.

For the class of functions on which the minimizers in this study were tested, there is no doubt - PRAXIS is, by far, superior.

For further comparison of PRAXIS to other minimizers and different classes of functions, refer to the book by Richard P. Brent.[3] This book contains a detailed description of PRAXIS and a large bibliography on minimizers.

FUNCTION CALLS

Function	Routine	Min	Max	Error(F)	Error(X)
QUAD0	PRA2	59	68	1.D-15	1.D- 8
	PRA8	94	145	1.D-30	1.D-15
	STRM	121	124	0.0	0.0
	MIGD	33	34	1.E- 8	1.E- 2
QUAD1	PRA2	59	111	1.D-23	1.D-12
	PRA8	92	189	1.D-21	1.D-10
	STRM	121	572	1.E-12	1.E- 7
	MIGD	33	111	1.E- 8	1.E- 2
QUAD2	PRA2	59	140	1.D-21	1.D-10
	PRA8	92	207	1.D-17	1.D- 8
	STRM	121	2697	1.E-12	1.E- 6
	MIGD	34	199	1.E- 7	1.E- 4
QUAD3	PRA2	59	152	1.D-19	1.D- 9
	PRA4	92	207	1.D-20	1.D-10
	STRM	121	14904	1.E-10	1.E- 5
	MIGD	44	253	0.7	0.6
QUAD4	PRA2	59	152	1.D-16	1.D- 8
	PRA8	98	221	1.D-22	1.D-11
	STRM	121	6991	1.5	0.8
	MIGD	44	198	1.9	1.0
QUAD5	PRA2	59	165	1.D-13	1.D- 6
	PRA8	98	243	1.D-23	1.D-11
	STRM	121	1084	3.6	1.6
	MIGD	45	287	2.0	1.0
GAUS0	PRA2	81	107	1.D- 9	1.D- 4
	PRA8	144	172	0.0	1.D- 9
	STRM	121	124	0.0	1.E- 3
	MIGD	33	165	0.9	0.6
GAUS1	PRA2	101	130	1.D- 7	1.D- 3
	PRA8	171	222	0.0	1.D- 8
	STRM	122	375	1.E- 6	1.E- 3
	MIGD	33	111	1.0	1.7
GAUS2	PRA2	111	238	1.D- 5	1.D- 2
	PRA8	180	373	0.0	1.D- 9
	STRM	143	1257	1.E- 5	1.E- 2
	MIGD	33	44	0.9	1.3

2/27/73

REFERENCES

1. Richard P. Brent, Algorithms for Minimization without Derivatives, Prentice-Hall, 1973

2. MINF68 - A General Minimizer Routine, Lawrence Radiation Laboratory, Group A Programming Note No. P-190, 1969

3. MINUIT - A program to minimize a function of n variables, compute the covariance matrix, and find the true errors, CERN Computer 6000 Series Program Library, 1969