V. DUMITRU
R. ŞERBAN
Department of Economic Cybernetics
Academy of Economic Studies

Speed Convergence and the Choice of Unidimensional Extremum Point: an Experiment with Rosenbrock's Function

Most of the unconstrained optimization algorithms determine at each iteration the optimum of a function on a certain direction. Obviously, this optimum is in general, not unique nor has the best strategy to follow been demonstrated theoretically. Moreover, we are interested in how the error with which this optimum is determined, influences efficiency in solving the problem. In this study, an experiment with Rosenbrock's test function is carried out in which the optimum is determined approximately by a unidimensional optimization subroutine (STEP) and by establishing with a high accuracy the first extreme point or the best extreme point on the line by solving the equation $\frac{\partial f(x^k + td^k)}{\partial t} = 0$

It has experimetally resulted for the tested function that the most efficient method is that of determining the best extreme point deriving from solving noniteratively (if it can be solved by a formula) the above mentioned equation.

A quite big number of unconstrained nonlinear programming algorithms is known, [1], [4], [5], [6], [8] algorithms which determine an extremum point of a function of n variables $f(x)$, $x \in E^n$.

By knowing at iteration k point x^k, a better point is sought by setting out from x^k on the line d^k so as:

$$x^{k+1} = x^k + td^k \tag{1}$$

where scalar t is the "step" from x^k to x^{k+1} on the d^k direction, determined by min. (max) $f(x^k + td^k)$.)

To determine step t, a unidimensional optimization subroutine is used. Such a subroutine, called "STEP" is given in [2] and is, generally, programming Powell's algorithm [1]. The "STEP" subroutine approximates the first extreme point met on the given line. To compare the efficiency of the various algorithms, by this meaning the capacity of the algorithm of solving the problem with a given accuracy and with minimum computer time, the computer time is supposed to be proportional to the number of estimations function evaluations plus n times the number of valuations of the gradient, thus avoiding the use of the effective time of computing which varies from a computer to another.

The tests made at the Informatics Laboratory of the Cybernetics Department — the Academy of Economic Studies [4], [5] as well as other results mentioned in literature, for example [8], show that the performances of the algorithms particularly depend, besides the initial parameters (for example the initial point) also on the parameters of the unidimensional optimization subroutine, mentioning among the especially the initial exploring step and the exactness of approximating the optimum.

This study presents the conclusions obtained as regards the influence of the precision of unidimensional optimization on the speed convergence of some algorithms based on an experiment with Rosenbrock's function, from two points of view:

a) The more exact determining of the optimum on a line by solving equation

$$f_t'(x^k + td^k) = 0 \tag{2}$$

b) In case equation (2) allows several extreme points, the convergence speed of the algorithms is studied in the event of choosing the first extremum point, as is generally specified in the description of many algorithms or of the best extreme point on the considered line.

Rosenbrock's function

$$f(x_1,\ x_2) = 100\ \ (x_2 - x_1^2)^2 + (1 - x_1)^2 \tag{3}$$

also known as "Rosenbrock's valley", presents a deep valley along parable $x_2 = x_1^2$ and an asymmetrical valley along the negative side of ox_2 axis. The function has a minimum in point $x^* = (1.\ 1)$ of the value $f(x^*) = 0$.

We have choosed as an initial point $x^0 = (-1 \cdot 2.1)$, a point from where the iterations started also in other experiments to which we shall compare the results.

To solve equation (2) for Rosenbrock's function, we programmed a special subroutine, called SSTEPR (Special Step for Rosenbrock) whose parameter ICONT specifies in the case ICONT = 0, the determining of the first extreme on line and in the case INCONT = 1, of the best extreme point.

Equation (2) for function (3) is of the form

$$at^3 + bt^2 + ct + d = 0 \tag{4}$$

where

$$a = 2\,(d_1^k)^4$$
$$b = -3\,(d_1^k)^2(d_2^k - 2d_1^k\,x_1^k)$$
$$c = (d_2^k - 2d_1^k\,x_1^k)^2 + (d_1^k)^2[0.01 - 2\,(x_2^k - (x_1^k)^2)] \qquad (5)$$
$$d = (x_2^k - (x_1^k)^2)\,(d_2^k - 2\,d_1^k\,x_1^k) - 0.01\,d_1^k\,(1 - x_1^k)$$

and x_1^k, x_2^k and d_1^k, d_2^k are the components of point x_k respectively of direction d^k.

The solving of equation (5) is made by the CARDAN subroutine which uses the formulas with the same name.

The following algorithms were tested:

— Cyclic optimization on axes [14]
— Davies — Swann and Campey [13]
— Powell [9]
— The simplex-Cyco mixed algorithm [4]
— Simplex with unidimensional optimization [4]
— Rosenbrock ([1], [4],
— Fletcher-Reeves [7], [5]
— The gradient — Hessian mixed algorithm [5]
— Dawidon-Fletcher-Powell [3], [5]
— The algorithm of parallel tangents [12], [5].

The results of the tests made are given in table 1.

Besides the results obtained by us with SSTEPR, the results previously obtained by V. Dumitru [4], [5] are also given, for comparison, in this table. The bibliographical source from where the results with which we make the comparison were taken is mentioned in the table, in the square brackets in the box for remarks.

We draw the following from the analysis of the results:

a) Either the best or the first extreme point was chosen, in comparison with the previous results, usind subroutine SSTEPRthe number of iterations has decreased and proportionaly to it also the number of valuations of the gradient and of the function.

b) The choosing of the best extreme point has led to a number of iterations much smaller than in the event of choosing the first extreme point.

c) Except for the simplex algorithms with unidimensional optimization (with a usual or weighted centroid) for which the results obtained have a weaker accuracy, in the case of choosing the first extreme on line, the best performances were obtained by the algorithms of Powell (37 valuations), Davidon-Fletcher-Powell (63), Rosenbrock and Davies-Swann-Campey (73), Fletcher-Reeves (75), the simplex mixed -Cyco algorithm (125) and the Algorithm of parallel tangents (145).

In the event of choosing the best extreme point, the performance of the algorithms is approximately the same as in the previous case, the number of function valuations (plus n times the number of gradient estimations, in the case of the algorithms using the gradient), namely: Davies-Swann-Campey's algorithm(11), Rosenbrock (11), Powell (13), the simplex-Cyco mixed algorithm (13), Fletcher-Reeves (38), Davidon-Fletcher-Powell (47), the gradient-hessian mixed algorithm (84) and the algorithm of parallel tangents (119).

THE RESULTS OF THE OPTIMIZATION TESTS FOR ROSENBROCK'S FUNCTION

Min $f(x_1, x_2) = 100\ (x_2 - x_1^2)^2 + (1 - x_1)^2$

Optimal point: $x^* = (1, 1)$; $f(x^*) = 0$

Initial point: $x^0 = (-1 \cdot 2;\ 1)$; $f(x^0) = 24.2$

Method	Subroutine and parameters	Point in which the algorithm ended	No. iterations	No. of. calls for		Remarks
				Function	Gradient	
1	2	3	4	5	6	7
CYCLIC OPTIMIZATION ON AXES	STEP	$x^k = \begin{bmatrix}0.73287\\0.53709\end{bmatrix}$ $f(x^k) = 0.07136$	$k = 200$	3299	—	[4]
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}0.99999\\0.99998\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 70$	211	—	
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}1.00000\\1.00000\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 2$	13	—	
DAVIES SWANN-CAMPEY	STEP	$x^k = \begin{bmatrix}0.99997\\0.99994\end{bmatrix}$ $f(x^k) = 0.8.10^{-9}$	$k = 25$	377	—	[4]
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}0.99999\\0.99999\end{bmatrix}$ $f(x^k) = 0.48.10^{-9}$	$k = 24$	73	—	
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}0.99999\\0.99999\end{bmatrix}$ $f(x^k) = 0.41.10^{-11}$	$k = 2$	11	—	
POWELL	STEP	$x^k = \begin{bmatrix}0.99954\\0.99903\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 13$	286	—	[4]
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}1.00000\\1.00000\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 12$	37	—	
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}1.00000\\1.00000\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 2$	13	—	
	STEP	$x^k = \begin{bmatrix}0.99277\\0.98559\end{bmatrix}$ $f(x^k) = 0.00005$	$k = 35$	550	—	[4]

Method	Subroutine and parameters	Point in which the algorithm ended	No. iterations	No. of calls for		Remarks
				Function	Gradient	
1	2	3	4	5	6	7
SYMPLEX CYCO MIXED ALGORITHM	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}0.99904\\0.99808\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 58$	125	—	
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}1.00000\\1.00000\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 1$	13	—	
SYMPLEX WITH UNIDIMENSIONAL OPTIMIZATION	STEP	$x^k = \begin{bmatrix}1.0285\\1.0542\end{bmatrix}$ $f(x^k) = 0.0021$	$k = 12$	192	—	[4]
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}1.17703\\1.38074\end{bmatrix}$ $f(x^k) = 0.03350$	$k = 21$	122	—	
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}1.16733\\1.35634\end{bmatrix}$ $f(x^k) = 0.03199$	$k = 28$	164	—	
SYMPLEX WITH BALANCED CENTROID AND WITH UNIDIMENSIONAL OPTIMIZATION	STEP	$x^k = \begin{bmatrix}1.0301\\1.0612\end{bmatrix}$ $f(x^k) = 0.00096$	$k = 21$	277	—	[4]
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}1.17895\\1.39099\end{bmatrix}$ $f(x^k) = 0.03213$	$k = 14$	82	—	
ROSENBROCK	STEP	$x^k = \begin{bmatrix}0.99997\\0.99994\end{bmatrix}$ $f(x^k) = 0.8.10^{-9}$	$k = 25$	377	—	[4]
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}0.99999\\0.99999\end{bmatrix}$ $f(x^k) = 0.48.10^{-9}$	$k = 24$	73	—	
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}0.99999\\0.99999\end{bmatrix}$ $f(x\ ) = 0.41.10^{-11}$	$k = 2$	11	—	
FLETCHER-REEVES	STEP	$x^k = \begin{bmatrix}0.99844\\0.99688\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 25$	208	26	
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}0.99750\\0.99497\end{bmatrix}$ $f(x^k) = 0.00001$	$k = 24$	25	25	

Method	Subroutine and parametrs	Point in which the algorithm ended	No. iterations	No. of calls for		Remarks
				Function	Gradient	
1	2	3	4	5	6	7
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}1.00501\\1.01007\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 9$	18	10	
GRADIENT HESSIAN MIXED ALGORITHM	STEP	$x^k = \begin{bmatrix}0.99998\\0.99996\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 39$	393	40	
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}1.00000\\1.00000\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 39$	40	40	
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}0.99999\\0.99999\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 19$	44	20	
DAWIDON-FLETCHER-POWELL	STEP	$x^k = \begin{bmatrix}1.00014\\1.00028\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 23$	131	24	
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}0.99997\\0.99994\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 20$	21	21	
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}0.99999\\0.99998\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 12$	21	13	
PARALLEL TANGENT ALGORITHM	STEP	$x^k = \begin{bmatrix}1.00154\\1.00309\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 17$	345	51	
	SSTEPR (ICONT = 0)	$x^k = \begin{bmatrix}1.00057\\1.000114\end{bmatrix}$ $f(x^k) = 0.00000$	$k = 16$	49	48	
	SSTEPR (ICONT = 1)	$x^k = \begin{bmatrix}1.00899\\1.01789\end{bmatrix}$ $f(x^k) = 0.00008$	$k = 10$	59	30	

REFERENCES

1. ABADIE J. Editor: *Integer and Nonlinear Programming.* North Holland Publishing Company, Amsterdam, 1970.
2. AMUNATEGUI J., DUMITRU V.: *Computer Experiment with an Unconstrained Nonlinear Program.* Economic Computation and Economic Cybernetics Studies and Research, Bucharest, no. 3, 1970.

3. DAVIDON W. C.: *Variable Metric Method for Minimization.* A. E. G. Researh and Development Report. ANL-5990.
4. DUMITRU V.: *Function Optimization without Evaluating Derivatives: Methods and Techniques, Flowcharts, Computer Subroutines and Numerical Results.* Economic Computation and Economic Cybernetics Studies and Research, Bucharest, Part I; no. 2, 1973, Part II: no. 1, 1974.
5. DUMITRU V.: *Gradient Methods of Unconstrained Optimization,* Ec. Comp. and Ec. Cyb. No 4 (1974)
6. FLETCHER R.: *Optimization,* Academic Press London, 1969.
7. FLETCHER R., REEVES C. M.: *Function Minimization by Conjugated Gradients,* Computer J., vol. 7, 1964.
8. KOWALIK J., OSBORNE M. R.: *Methods for Unconstrained Optimization Problems,* Elsevier, N. Y., 1968.
9. POWELL, M. J. D.: *An Efficient Method of Finding the Minimum of a Function of Several Variables without Calculating Derivatives.* The Computer Journal, vol. 7, 1964.
10. POWELL M. J. D:. *Rank One Methods for Unconstrained Optimization.* In [1].
11. ROSENBROCK H. H.: *An Automatic Method for Finding the Greatest of Least Value of a Function,* Computer J., vol. 3, 1960.
12. SHAH B. V., BUEHLER J. J., KEMPTHORNE O.: *Some Algorithms for Minimizing a Function of Several Variables.* J. Soc. Indust. Appl. Math. 12, 1964.
13. SWANN W. H.: *Report on the Development of a New Direct Search Method of Optimization.* ICI Ltd. Central Instrument Laboratory Research Note, 64, 3, 1964.
14. ZADEH N.: *A Note on the Cyclic Coordinate Ascent Method.* Management Science, vol. 16, no. 9, 1970.

VITESSE DE CONVERGENCE ET CHOIX DU POINT D'EXTREMUM UNIDIMENSIONNEL — UN EXPERIMENT A L'AIDE DE LA FONCTION DE ROSENBROCK

La plupart des algorithmes d'optimisation sans restrictions détermine à chaque itération l'optimum de la fonction dans une certaine direction.

Evidemment, cet optimum n'est pas unique en général, et au point de vue théorique, il n'est pas démontré quelle serait la meilleure stratégie à suivre. En outre, il importe, pour l'efficience de la résolution du problème, l'influence de l'erreur dans la détermination de cet optimum. L'auteur présente un expériment à l'aide de la fonction de Rosenbrock où l'optimum est déterminé avec approximation à l'aide d'une sous-routine d'optimisation unidimensinnelle (STEP) déterminant le premier point d'extremum ou le meilleur point d'extremum dans la direction et en donnant la solution de l'équation

$$\frac{\delta f(x^k + td^k)}{\delta t} = 0$$

Pour la fonction vérifiée, il résulte de façon expérimentale, qu'en général, le procédé est celui de déterminer le meilleur point d'extremum obtenu par la résolution de l'équation mentionnée.

СКОРОСТЬ СХОДИМОСТИ И ВЫБОР ОДНОМЕРНОЙ ТОЧКИ ЭКСТРЕМУМА : ЭКСПЕРИМЕНТ С ФУНКЦИЕЙ РОЗЕНБРОКА

Большинство алгоритмов оптимизации без ограничений определяет при каждой итерации оптимум функции по определенному направлению. Несомненно, этот оптимум вообще не является единственным и в теоретическом отношении не представляет собой наилучшую стратегию, которой необходимо придерживаться. Помимо этого, представляет интерес эффективность при решении задачи, погрешность определения этого оптимума.

В настоящей статье проводится эксперимент с функцией Розенброка, при котором оптимум определяется с приближением одной одномерной субрутины (STEP) и путем определения первой точки экстремума или наилучшей точки экстремума в результате решения уравнения:

$$\frac{\delta f(x^k + td^k)}{\delta t} = 0.$$

Для подвергаемой тесту функции экспериментально вытекает, что вообще наиболее эффективным способом является определение наилучшей точки экстремума, вытекающей из решения указанного уравнения.

22.V.1974