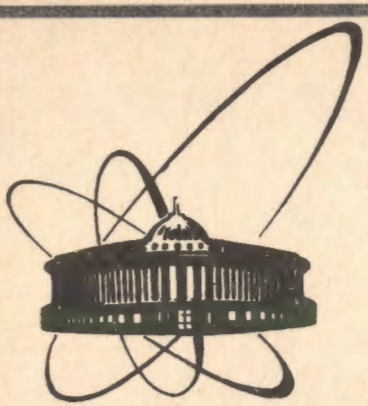
90-552



сообщения
объединенного
института
ядерных
исследований
дубна

E10-90-552

V.I.Ilyushchenko

NEW INTEGRATION METHODS OF GLOBAL OPTIMIZATION BASED ON RIEMANN, LEBESGUE AND PERRON INTEGRAL STRUCTURES

1990

1. Introduction

The existing standard optimization methods of the zero, first and second order perform optimization by computing and comparing the values of the objective function itself, its first and second derivatives, respectively [1,2].

It is a well-known fact that any systematical and/or statistical errors produced in measuring objective functions lead to abnormally great accompanying errors in the values of its first and second derivatives [3,4]. To dump this negative effect, the computer codes developed for processing spectral information with multipeak structures generally use some special filtering techniques [5].

On the other hand, even such an advanced optimization package as MINUIT [6] is totally devoid of any filtering, scaling or smoothing procedures. For example, its basic subroutine DERIVE computes the first derivative, i.e. gradient, and its error by means of the standard numerical central difference formula.

Our computational experience shows that any additional scaling, filtering or smoothing substantially increase both stability and reliability of the standard optimization processes to result in, e.g. a Hessian matrix with substantially improved analytical properties as compared with the Levenberg-Marquardt code. A noticeable improvement is also provided via replacing a simple numerical differentiation code by a more advanced one like DGT3 from program library SSP IBM [4]. The latter code is based on a Lagrange interpolation scheme producing the desirable smoothing effect.

However, a much more drastic gain in optimization stability and reliability is furnished by integration methods [7-10] due to their strong filtering and smoothing action on initial measurement errors of any kind [4].

This paper is devoted to an analysis of the properties of new integration methods of global optimization of arbitrary objective functions by means of Riemann, Lebesgue and Perron integral structures and the integration process itself.

2. The known integration methods of global optimization

Presently, there are known four basic integration methods of global optimization realized in the form of computer algorithms and codes, i.e. linear integral transform [7], psi-transform [8], maximum entropy technique [9] and integration weighing technique [10].

Объединенный институт ядерных исследований БИБЛИОТЕКА

2.1 Linear integral transform [7]

The basic idea of this method is to smooth an initial multimodal objective function F(x) by means of a linear integral transform

$$F(x,a) = \int H(u)*F(x - a*u)du , \qquad (1)$$

where H(u) is a weight function and a=>0 is the averaging parameter. The smoothing action of the linear transform increases proportional to a.

2.2 Psi-transform [8]

This method can be used to solve both linear and nonlinear programming problems. According to this method, an initial objective function F(x) being multivariate and multimodal is transformed into a randomized univariate monotone decreasing function T(a). The global extremum value, $F_{GM}(x)$, can be found from the following condition:

$$T(a) = 0 . \qquad (2)$$

The psi-transform method is based on the partition of F(x) used to construct its Lebesgue integral. Additional functions formed during this transform allow to identify the coordinates of the global extremum, x_{GM}.

However, both above integration methods are limited in scope due to either some analytical properties of F(x) or technical problems encountered in their realization.

2.3 Maximum entropy technique [9]

This technique stems from the integral functional used to compute the so-called Quantity of Information (QI) introduced in 1948 by C.Shannon [11] and N.Wiener [12] and connected with the statistical information measure proposed in 1925 by R.A.Fisher [13] by a simple double differentiation transform.

The value of QI coincides with the negative value of entropy, H, to result in the following integral transform:

$$QI = -H = \int F(x)*\log(F(x))dx . \qquad (3)$$

This expression must be optimized to produce the value of the global minimum for maximum entropy H = MAX ! or minimum information, QI = MIN ! The reduced version of relation (3) contains only a logarithmic integrand.

2.4 Integration weighing technique [10]

This method uses the well-known weighted mean expression:

$$R_{av} = \frac{\sum_{i=1}^{m} (R_i W_i)}{\sum_{i=1}^{m} (W_i)} , \qquad (4)$$

where W_i is the weight of the i-th measurement (R_i). By passing to an integral presentation and assuming $W_i = W_i(R_i)$ one can write down

$$R_{av} = \frac{\int_{-oxo}^{+oxo} R(x)W(x)dx}{\int_{-oxo}^{+oxo} W(x)dx} , \qquad (5)$$

where x is an independent variable.

On the other hand, the conventional Laplace method [14] for the evaluation of an integral like

$$IL = \int_{-oxo}^{+oxo} (G(x)*\exp(\Lambda*F(x)))dx \qquad (6)$$
$$\Lambda \longrightarrow +oxo$$

is based on a substitution of the IL by an IL(Λ) over a small vicinity (+-d) of the point of a maximum, x*, of a function F(x), i.e.

$$IL(\Lambda) = \int_{x^*-d=a}^{x^*+d=b} (G(x)*\exp(\Lambda*F(x)))dx , \qquad (7)$$
$$\Lambda \longrightarrow +oxo$$

The asymptotic estimates given in [14] suggest that, for the points within an interval [a,b], one gets

$$\lim_{\Lambda \longrightarrow +\infty} \frac{\int_{-\infty}^{+\infty} (G(x)*\exp(\Lambda*F(x)))dx}{\int_{-\infty}^{+\infty} \exp(\Lambda*F(x))dx} = G(x^*) , \qquad (8)$$

Thus, by taking F(x) as an objective function and using exp(Λ*F(x)) as a weighing function, i.e. W(x) = exp(Λ*F(x)), one obtains a formula to find out the coordinates of a maximum, R=G(x*), which is positioned within a multidimensional closure C, as follows

$$\lim_{\Lambda \longrightarrow +\infty} \frac{\int\cdots\int_{C} (G(v)*\exp(\Lambda*F(v)))dv}{\int\cdots\int_{C} \exp(\Lambda*F(v))dv} = G(x^*) , \qquad (9)$$

where v is a vector of independent variables.

Numerical computer tests [10] realized with a single precision version of the multiple integration code from LIBCERN for many differentiable and nondifferentiable test functions provided satisfactory final results with Λ =< 200 and a few percent precision.

The integration weighing method is devoid of any inherent limitations when dealing with the nondifferentiable and discontinuous objective functions being out of reach of the standard optimization methods of the zero, first and second order.

3. The Baire classification of functions [15]

A set of all the continuous functions defined within a certain finite interval [a,b] is called the zero Baire class of functions.

If a function F(x) does not enter the zero class, but is representable as

$$F(x) = \lim_{n \longrightarrow \infty} F_n(x) , \qquad (10)$$

where all $F_n(x)$'s are continuous, then F(x) is called a function of the first Baire class.

An analogous definition holds for a function F(x) not entering either the zero or first Baire class, but representable in the form of (10) , where all the functions $F_n(x)$ pertain to the first Baire class – now this function F(x) will be a member of the second Baire class.

One of the main properties of the Baire functions is their measurability. In addition, e.g. a function F(x) specified by a countable set of discontinuity points will belong to the first Baire class.

The second Baire class can be exemplified by the well-known Dirichlet function

$$D(x) = \begin{cases} 1 \text{ for } x = r_k \\ 0 \text{ for } x = r_k \end{cases} , (k=1,\ldots n) , \qquad (11)$$

where r_k are rational points of the segment [0,1].

4. The known integral structures [15-18]

Theory of the integral [16] deals with two groups of integrals specified by different computational structures. The first group is composed of conventional Riemann, Lebesgue and Perron integrals, while the second one includes their Stieltjes modifications. Here we will consider only the representatives of the first group.

4.1 Riemann integral [15,17]

Let us consider an integrand function F(x), where x is an independent variable defined within an interval [a,b]. According to Riemann integral computation structure, this interval must be subdivided into n elementary subintervals (cells) $\Delta x_i = \frac{b-a}{n}$. If the minimum and maximum edges of the function F(x) within each cell Δx_i are equal to $F_i(x)$ and $F^i(x)$, respectively, then the corresponding lower and upper Darboux sums

$$S_d = \sum_{i=1}^{n} F_i(x)\Delta x_i \qquad (12)$$

$$S^d = \sum_{i=1}^{n} F^i(x)\Delta x_i . \qquad (13)$$

The Riemann integral

$$RI = (R)\int F(x)dx \qquad (14)$$

if

$$\lim_{n \to \infty} S_d = \lim_{n \to \infty} S^d = RI . \qquad (15)$$

In the process of global optimization by means of a Riemann integral structure one misses all the local extrema with base widths

$$\Delta x_{LM} \leq \Delta x_i \qquad (16)$$

as is illustrated by Fig.1.

However, the very process of computation of RI provides an additional means to identify the global minimum. This can be accomplished by sorting elementary cell integrals,

$$CI = F_i(x)\Delta x_i \qquad (17)$$

according to their absolute value.

Thus, the external global minimum indicator in the form of RI is supplemented by the internal global minimum indicator, ABS(CI). For example, in the case of maximum entropy technique,

$$RI = -\int P(x)\log(P(x))dx \qquad (18)$$

and the corresponding value of CI is provided by the integration code itself.

4.2 Lebesgue integral [16,18]

As opposed to Riemann integral structure that of Lebesgue starts from a subdivision of F(x) within an interval [$F_*(x), F^*(x)$] (see Fig.2), where F_* and F^* are the minimum and maximum values of F(x).

The cell values of F(x) within each subinterval, $F_i(x)$, are multiplied by corresponding cell subinterval measures, m_i, so that the limit of the sum

$$LI = \lim \sum_{i=1}^{n} F_i m_i \longrightarrow (L)\int F dm , \qquad (19)$$

where LI is Lebesgue integral.

In the process of computing elementary cell Lebesgue measures, m_i, one automatically identifies the global minimum and all the narrow extremum peaks missed in Riemann integration scheme (see Fig.1). This is a simple illustration of the cross-intercept effect.

Unfortunately, in contrast to conventional Riemann scheme, the only known practical Lebesgue integration scheme is that used in the psi-transform technique (see Section 2.2). The latter algorithm is based on an MC random sampling being the slowest method known in the field of interest.

There exists an urgent need for developing more efficient, i.e. direct and fast, computational algorithms to implement the alternative Lebesgue integration structure.

Here again, the very process of integration automatically provides an additional means to identify the global minimum, when going from F_* to F^*. This occurs for the first cell measure

$$\int_a^b dm \neq b - a . \qquad (20)$$

A combination of both integration schemes, those of Riemann and Lebesgue, guarantees a reliable cross-check tool to identify the global extremum position of any integrable function.

4.3 Perron integral [15,16,18]

In passing from smooth continuous convex objective functions, F_0, to nonsmooth discontinuous nonconvex ones, F_2, i.e. from the functions of the zero Baire class to those of the second Baire class, one shifts from the Riemann integral to that of Perron, i.e. to

$$PI = (P)\int F(x)dx . \qquad (21)$$

This integral can be defined by means of the notions of majorating and minorating functions, U and V, respectively, [16]:

$$U \geq PI \geq V \qquad (22)$$

or a superfunction, $\psi(x)$, and a subfunction, $\phi(x)$, to satisfy the following inequality [18]:

$$\bar{D}\ \psi(x) \geq F(x) \geq \underline{D}\ \phi(x) , \qquad (23)$$

where $\bar{D}$ and $\underline{D}$ are the upper and lower derivatives, respectively.

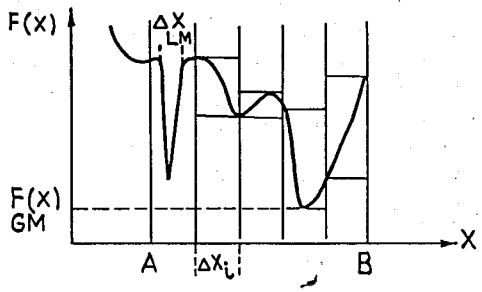


Fig.1.

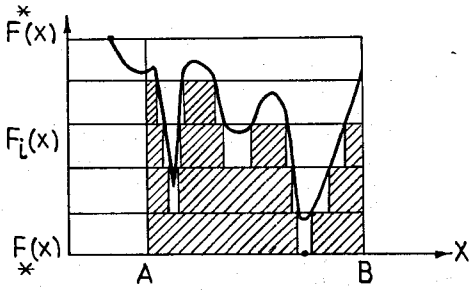


Fig.2.

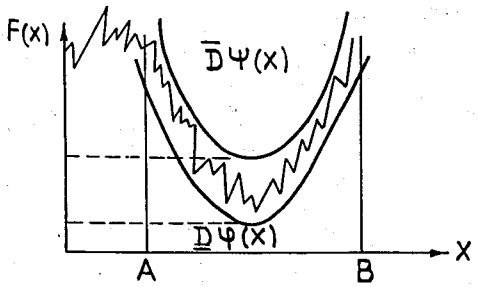


Fig.3.

The method of Variationally Weighted Quadratic Majorants (VWQM) is analyzed in detail in [19], while [20] contains some qualitative results for entropy integral (18).

An elementary illustration of the Perron integration scheme can be seen in Fig.3. Let us consider an integrand F(x) as being discontinuous and piecewise smooth. Then the majorant integral PI* and minorant integral PI* can be used to evaluate the true integral value PI_t as

$$PI_t \sim (PI^* + PI_*)/2 . \qquad (24)$$

The same relation holds for elementary cell integrals.

Here again a powerful effort is needed for implementing basic algorithms into effective computer codes.

5. Conclusion

The development of new integration methods can be viewed as one of the principal modern trends in global optimization research, especially, that of discontinuous objective functions typical of experimental situation.

The integration methods are especially effective in solving the so-called ill-conditioned problems. This efficiency is due to strong error filtering and dumping effect inherent in the very numerical integration process.

The cooperation of Dr. V.V.Tsynya and the staff members of the EC-1055M team is gratefully acknowledged.

References

1. J.R.Dennis and R.B.Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, N.J. (1983).
2. P.E.Gill, W.Murray and M.H.Wright, Practical Optimization Academic Press, London (1981).
3. T.E.Shoup, A Practical Guide to Computer Methods for Engineers, Prentice-Hall, Englewood Cliffs, N.J. (1979).
4. V.I.Ilyushchenko, Numerical differentiation and integration of noisy model data, Preprint JINR, Dubna (1990) (to be published).

5. V.Hnatowicz, V.I.Ilyushchenko and P.Kozma, CPC, 60, 111 (1990).
6. F.James and M.Roos, CPC, 10, 343 (1975).
7. V.Ya.Katkovnik, Linear Estimates and the Stochastic Problems of Optimization, Nauka, Moscow, (1976)(in Russian).
8. F.Archetti, A Sampling Technique for Optimization, Editrice Tecnico Scientifica ,Pisa (1975).
9. B.R.Frieden (ed.), The Computer in Optical Research, Springer, Berlin (1980).
10. V.I.Ilyushchenko, An integration weighing method to evaluate extremum coordinates, Preprint JINR E10-90-410, Dubna, (1990) (submitted to NIM).
11. C.Shannon, Bell Syst. Tech. J., 27, 379 (1948); 27, 623 (1948).
12. N.Wiener, Cybernetics, Wiley, N.Y. (1948)
13. R.A.Fisher, Statistical Methods for Research Workers, Oliver and Boyd, London (1925).
14. N.G.De Bruijn, Asymptotic Methods in Analysis,North Holland, Amsterdam (1958).
15. I.P.Natanson, The Theory of Functions of a Real Variable, GITTL, Moscow (1957).
16. S.Saks, Theory of the Integral, Stechert, N.Y. (1937).
17. E.C.Titchmarsh, The Theory of Functions, Oxford University Press, Oxford (1939).
18. E.Kamke, Das Lebesgue-Stieltjes-Integral, Teubner, Leipzig (1956).
19. V.I.Mudrov and A.A.Ivlev, Newton Majorants in Applied Problems, Radio i Svyaz, Moscow (1987) (in Russian).
20. A.W.Marshall and I.Olkin, Inequalities: Theory of Majorization and its Application, Academic Press, N.Y. (1979).

Received by Publishing Department
on December 10, 1990.