An Algorithm to Locate the Greatest Maxima of Multi-variable Functions

Takao Tsuda* and Mieko Sato*

Abstract

In this paper we propose an algorithm to locate the greatest maximum of a multi-peaked function of many variables.

The main ingredients of this numerical method are the relative peaking of each of the maxima by use of an exponential function and the numerical multidimensional integration.

Detailed discussions are given to some computed examples that evidence the superiority of the scheme proposed to simple random samplings.

1. Introduction

We propose in this paper an algorithm to locate the global (or greatest) maximum of a multi-peaked function defined in the k-dimensional Euclidean space ($k >> 1$).

For functions, each with a single peak, various potential methods have been developed to date[1], and the search for the global maxima often turns out to be successful even in multi-dimensional cases. For multi-variable functions, each having more than one peak, however, global search is mandatory for locating the global maximum. In this case those methods that are useful for singly-peaked functions often result in local maxima before reaching the desired solution, if the starting positions are not appropriately selected. In view of this the only method for multi-peaked multi-variable functions is random sampling. In this paper, therefore, we propose a method more efficient than random sampling. Being more efficient means less frequent function evaluations. Note that most of the computation time is spent for function evaluations.

In the case of a few dimensions, some pathological functions are often used for testing purposes. We do not consider such functions; but rather, those analytic functions that can be approximated by multi-peaked multinomials are treated.

The main characteristics of the algorithm proposed are the relative peaking of each maximum as combined with numerical multiple integration. The scheme of multiple integration is such that the precision is weakly dependent on the number of dimensions, thereby giving results much better than simple-minded random sampling with increasing number of dimensions.

It is difficult to assess the convergence properties of the iteration scheme *a priori* in our case of global search. The fact is, however, that the convergence is satisfactory insofar as the numerical experiments are concerned.

This paper first appeared in Japanese in Joho—Shori (Journal of the Information Processing Society of Japan), Vol. 16, No. 1 (1975), pp. 2~6.

* Department of Electrical Engineering, Hokkaido University

2. Algorithm

2.1 Outline

Let the object function be $F(x_1, x_2, \dots , x_k)$ $[\equiv F(\underset{\sim}{x})]$ $(k >> 1)$. In the rectangular domain $D \equiv [a_1,b_1] \times [a_2,b_2] \times \dots \times [a_k,b_k] \equiv D_1 \times D_2 \times \dots \times D_k$ we look for the point $\bar{\underset{\sim}{x}}$ approximately at which the function assumes the greatest value. Another function such that

$$f(\underset{\sim}{x}) = \exp[\ c(F(\underset{\sim}{x})-F_0)\ ] \tag{1}$$

is defined, where constants c (expansion coefficient) and F_0 (shift constant) are determined in such a way that the exponent of eq.(1) is about zero for $F(\underset{\sim}{x}) = F_{min}$ (observed minimum; $\underset{\sim}{x} \in D$) and does not cause overflow even for $F(\underset{\sim}{x}) = F_{max}$ (observed maximum; $\underset{\sim}{x} \in D$).

We call this operation "shift and expand", which, as seen from fig. 1 , performs the relative peaking of maxima in domain D , i.e. differentiates the highest peak from the other lower peaks by mapping function F to the newly derived function f .

After having carried out such "shift and expand" procedure, we compute, for $i = 1, 2, \dots , k$,

$$x_i^{(1)} = \frac{\int_D x_i \exp[\ c(F(\underset{\sim}{x})-F_0)\ ]\ d\underset{\sim}{x}}{\int_D \exp[\ c(F(\underset{\sim}{x})-F_0)\ ]\ d\underset{\sim}{x}} , \tag{2}$$

which then is an approximation to $\bar{x}_i$, because the exponential functions in eq.(2) is almost $+\infty$ in the neighborhood of $\underset{\sim}{x} = \bar{\underset{\sim}{x}}$. For the multiple integration, use is made of the numerical scheme that has previously been developed[2].

$x_i^{(1)}$ is the first approximation to $\bar{x}_i$, which is indicated by the superscript 1 . Similarly, we denote $D_1 = D_1^{(1)}, \dots , D_k = D_k^{(1)}, D = D^{(1)}$.

Successively the domain of search is contracted. By defining

$$\frac{\int_{D^{(p)}} \phi(\underset{\sim}{x}) \exp[\ c(F(\underset{\sim}{x})-F_0)\ ]\ d\underset{\sim}{x}}{\int_{D^{(p)}} \exp[\ c(F(\underset{\sim}{x})-F_0)\ ]\ d\underset{\sim}{x}} \equiv I^{(p)}[\ \phi(\underset{\sim}{x})\ ] , \tag{3}$$

eq.(2) is then

$$x_i^{(1)} = I^{(1)}[\ x_i\ ] . \tag{4}$$

If one computes

$$[\ \delta_i^{(1)}\ ]^2 \equiv I^{(1)}[\ x_i^2\ ] - (I^{(1)}[\ x_i\ ])^2 , \tag{5}$$

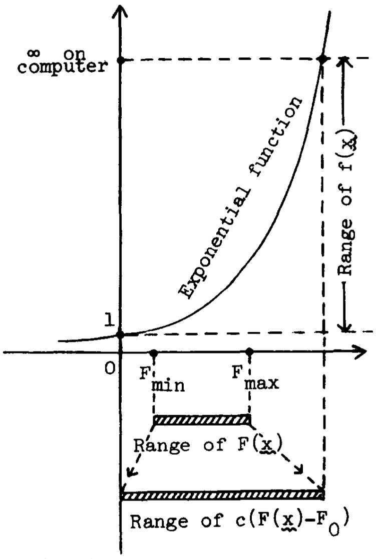

fig. 1
Procedure of SHIFT-AND-EXPAND: The range of $F(\underset{\sim}{x})$ is expanded to the new range of $f(\underset{\sim}{x})$ by exponential function.

then $\delta_i^{(1)}$ gives the width of the highest peak along the length of the x_i-coordinate together with the influence of the adjacent lower peaks on the current estimate of $\bar{x}_i$. $(2k + 1)$ multiple integrals involved in eqs.(4) and (5) are evaluated with common sample points, hence with common function values of F .

The interval of x_i is then reshaped to

$$D_i^{(2)} = [\ x_i^{(1)} - \alpha\delta_i^{(1)} ,\ x_i^{(1)} + \alpha\delta_i^{(1)}\ ] \tag{6}$$

and the computation of (4) and (5) is carried over again. Here α is a positive constant. It should be considered that $D_i^{(1)} \supseteq D_i^{(2)}$ and $D_i^{(2)}$ be not much too small compared with $D_i^{(1)}$. This consideration is necessary to prevent the true value $\bar{x}_i$

from getting out of the interval (6).

After the repetition of the above procedures, it is expected that

$$D^{(1)} \supseteq D^{(2)} \supseteq \ldots \rightarrow 0 , \tag{7}$$

and $\underset{\sim}{x}^{(1)}, \underset{\sim}{x}^{(2)}, \ldots$ will correspondingly give better estimates of $\underset{\sim}{\bar{x}}$. If (7) is not the case, then there are more than one peak comparable in height.

2.2 Details of the algorithm

1) Zero-th order approximation $x_i^{(0)}$ $(i = 1, 2, \ldots , k)$

Do steps 2) and 3) with $D^{(0)} \equiv D$; compute

$$x_i^{(0)} = I^{(0)}[\, x_i \,] . \tag{8}$$

The value of $x_i^{(0)}$ is used when estimating $(\delta_i^{(1)})^2$ in eq.(10) of step 4).
Put $p = 1$ (and $D^{(1)} = D^{(0)}$ for this case only) and do steps 2), 3), 4), 6) and 7).

2) Preliminary computation

The gross range of $F(\underset{\sim}{x})$ is surveyed with random sampling of points in $D^{(p)}$. The number of points sampled is $100 \times k$. Denote the minimum (maximum) value of $F(\underset{\sim}{x})$ thus obtained by F_{min} (F_{max}). For precaution, F_{max} is taken to be $F_{max} + (F_{max} - F_{min})$, otherwise the exponential function of the integrand may easily cause overflow. If, eventually, an overflow has occurred, then step 8) sees to this.

3) Shift and expand

As shown in fig. 1 , the range of F is transformed from $[\, F_{min}, F_{max} \,]$ to $[\, 0, c(F_{max} - F_0) \,]$. $F_0 = F_{min}$ and c is taken so that $\exp[\, c(F_{max} - F_0) \,]$ is 10^{10} below the overflow level.

4) Improvement of the estimates

Compute

$$x_i^{(p)} = I^{(p)}[\, x_i \,] \tag{9}$$

$$(\, \delta_i^{(p)} \,)^2 = I^{(p)}[\, (x_i - x_i^{(p-1)})^2 \,] - (\, I^{(p)}[\, x_i - x_i^{(p-1)} \,] \,)^2 . \qquad (i = 1, 2, \ldots , k) \tag{10}$$

There is good reason to compute (10) instead of

$$(\, \delta_i^{(p)} \,)^2 = I^{(p)}[\, x_i^2 \,] - (\, I^{(p)}[\, x_i \,] \,)^2 , \tag{11}$$

which will incur a cancellation of significant digits with increasing accuracy of the estimates. In eq.(10), the mid-point of the range of integration approximately agrees with the position of the highest peak (i.e. the most recent estimate $x_i^{(p-1)}$). By this consideration the difference between the first-order moment squared and the second-order moment is retained more accurately than otherwise.

5) Convergence criterion

The computation is terminated when d is below d_c (say $d_c = 10^{-5}$), where

$$d = |\, \{\, F(\underset{\sim}{x}^{(p)}) - F(\underset{\sim}{x}^{(p-1)}) \,\} \,/\, F(\underset{\sim}{x}^{(p-1)}) \,| . \tag{12}$$

6) Contraction of the range of integration

If $d_c < d$, then the new range for x_i is computed as

$$D_i^{(p+1)} = [\, x_i^{(p)} - \alpha\delta_i^{(p)},\ x_i^{(p)} + \alpha\delta_i^{(p)} \,] . \tag{13}$$

α is a positive constant (say $\alpha = 2.0$)*.

---

* The choice of $\alpha = 2.0$ is not without reason. In case of the peak being like a normal distribution, the position of this peak is within the interval with probability greater than 95 % .

When $D_i^{(p+1)} \not\subseteq D_i^{(p)}$, the interval is trimmed in such a way that $D_i^{(p+1)} \subseteq D_i^{(p)}$. Conversely, when

$$\delta_i^{(p)} \gg \delta_i^{(p+1)} \tag{14}$$

resulting in some orders of difference between $\delta_i^{(p)}$ and $\delta_i^{(p+1)}$, it is possible that the solution $\bar{x}_i$ is outside the interval $D_i^{(p)}$. In order to keep the interval length (denoted by $|D_i^{(p)}|$ for $D_i^{(p)}$) from rapidly decreasing, the lower bound β is considered for the contraction ratio $|D_i^{(p+1)}| \,/\, |D_i^{(p)}|$ (say $\beta = 1/3 \sim 1/2$). Namely, the wider interval is chosen, from either this case or the case of eq.(13). Also let the relation $D_i^{(p)} \supseteq D_i^{(p+1)}$ hold. The above procedures are followed for $i = 1, 2, \ldots, k$.

7) Repetition

Increase the value of p by one, return to 2) to perform the preliminary computation over the intervals determined in 6). Using the results obtained, do the shift-and-expand of 3) and then 4), 5), 6) and 7).

8) Revision of the expansion coefficient c

While computing (9) and (10), some overflow may arise in the exponential functions. In such cases, the value of c is corrected so that the overflow will not occur. Retaining the current value of p , return to the beginning of 4) and go over the computation again.

9) The lower bound β to the contraction ratio is an arbitrary parameter externally defined. A wise policy is therefore to try recalculations with different values of β ranging over 1/3 to 1/2.

3. Numerical experiments

The following function is considered.

$$f(x_1, x_2, \ldots, x_5) = f_1(x_1) \times f_2(x_2) \times f_3(x_3) \times f_4(x_4) \times f_5(x_5),$$

$$f_1(x_1)=x_1(x_1+13)(x_1-15)\times 0.01, \quad f_2(x_2)=(x_2+15)(x_2+1)(x_2-8)\times 0.01,$$

$$f_3(x_3)=(x_3+9)(x_3-2)(x_3-9)\times 0.01, \quad f_4(x_4)=(x_4+11)(x_4+5)(x_4-9)\times 0.01,$$

$$f_5(x_5)=(x_5+9)(x_5-9)(x_5-10)\times 0.01.$$

Throughout those examples shown below, each numerical integration is done with 2500 sample points (or function evaluations); 500 sample points for the preliminary computation; the value of β equated to 1/3.

[Example 1] Global maximum interior to the intervals; domain of search $-10 \leq x_i \leq 10$ ($i = 1, 2, \ldots, 5$). In this case there are 2^4 local maxima and 2^4 local minima.

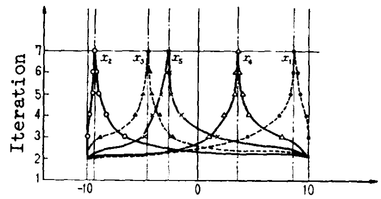

fig. 2
Convergence in Example 1
(—·—: solution)

	x_1	x_2	x_3	x_4	x_5	f_{max}
Solution	8.7564	-9.3582	-4.5720	3.5921	-2.8400	24416.03
Computed result	8.7597	-9.3613	-4.5716	3.5927	-2.8428	24416.01

7 iterations led to convergence. See fig. 2 for interval contractions.

[Example 2] Global maximum at one end of the intervals.

i) Domain of search $-10 \leq x_1 \leq 8$, $-10 \leq x_2 \leq 11$, $-10 \leq x_i \leq 10$ ($i = 3, 4, 5$).

	x_1	x_2	x_3	x_4	x_5	f_{max}
Solution	8.0	11.0	-4.5720	3.5921	-2.8400	27604.21
Computed result	7.9999	-9.3586	-4.5695	3.5923	-2.8406	24139.83

Convergence was attained after 9 iterations. The estimate for x_2 has significantly deviated from the solution, because the object function very slightly differs between $x_2 = 11.0$ and $x_2 = -9.3586$.

ii) Domain of search $-10 \leqq x_1 \leqq 8$, $-10 \leqq x_2 \leqq 12$, $-10 \leqq x_i \leqq 10$ $(i = 3, 4, 5)$.

	x_1	x_2	x_3	x_4	x_5	f_{max}
Solution	8.0	12.0	-4.5720	3.5921	-2.8400	41406.32
Computed result	7.9999	12.000	-4.5719	3.5929	-2.8406	41406.31

Convergence was attained after 11 iterations. See fig. 3 for interval contractions. The global maximum at $x_2 = 12.0$ is significantly greater than other interior local maxima.

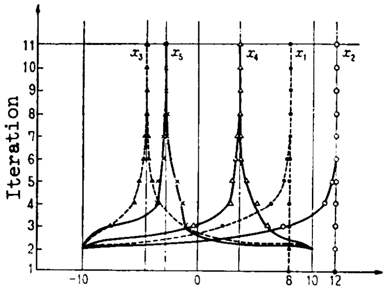

fig. 3
Convergence in ii) of Example 2
(--- : solution)

4. Concluding remarks

The main points are summarized as follows.

(a) Global decision is indispensable for processing the cases of multi-peaked functions. To date the only method to find the approximate solution, or at least the first-order approximation, has been the random-sampling technique. Our attempt in this paper is to propose a method more efficient than random sampling — more efficient in the sense that less frequent function evaluations are required to attain the same objective.

(b) At the moment it is difficult to give theoretical convergence criteria. Numerical experiments, however, indicate that the proposed algorithm works well, especially for a problem such that the global maximum is well inside the domain of search.

(c) Those problems that the global maxima occur on the edge of the rectangular domain of search are often baffling.

(d) Where the greatest maximum is scarcely higher than the next highest peak, the volumes that support the respective peaks are also close to each other in magnitude. Considering this situation, one may derive some sufficient condition for the convergence of the iterative computation. To estimate the magnitude of the above partial volume is as difficult a problem as the solution itself.

Acknowledgment

All the computation was done at the Hokkaido University Computation Center.

References

1) See, for instance, J.Kowalik and M.R.Osborne, Methods for Unconstrained Optimization Problems, American Elsevier Publishing Co., Inc., N.Y. (1968).

2) T.Tsuda, Numerische Math. 20 (1973), 377-391.