Reliable Computing 2 (2) (1996), pp. 187–193

Locating, characterizing and computing the stationary points of a function

MICHAEL N. VRAHATIS and EVANGELIA C. TRIANTAFYLLOU

A method for the localization, characterization and computation of the stationary points of a continuously differentiable real-valued function of n variables is presented. It is based on the combinatorial topology concept of the degree of a mapping associated with an oriented polyhedron. The method consists of two principal steps: (i) localization (and computation if required) of a stationary point in an n-dimensional polyhedron; (ii) characterization of a stationary point as a minimum, maximum or saddle point. The method requires only the signs of gradient values to be correct and it can be successfully applied to problems with imprecise values.

Поиск, классификация и вычисление стационарных точек функции

М. Н. ВРАХАТИС, Е. С. ТРИАНТАФИЛЛУ

Предложен метод поиска, классификации и вычисления стационарных точек непрерывно дифференцируемой вещественной функции n переменных. Метод основан на заимствованном из комбинаторной топологии понятии степени отображения, связанного с ориентированным многогранником. Метод состоит из двух основных шагов: 1) локализация (и, если необходимо, вычисление) стационарной точки в n-мерном многограннике; 2) классификация стационарной точки как точки минимума, максимума или седловой точки. Применение метода требует знания только знаков градиентов, поэтому данный метод может успешно использоваться для решения задач с погрешностями в условиях.

1. Introduction

Several methods for finding the stationary points of a function $f\colon \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$, where $\mathcal{D}$ is open and bounded, have been proposed with many applications in different scientific fields (mathematics, physics, engineering, computer science etc.). Most of these methods require derivative calculations, one dimensional sub–minimization, or/and approximation of the inverse of the Hessian matrix. Even the most efficient methods require precise function and gradient values. In many applications though, such as numerical simulations, precise values are either impossible or time consuming to obtain [6]. These problems can be dealt with by methods that do not require precise function and gradient values [2, 4, 7, 8, 11, 12].

In this contribution a method is presented for the localization, characterization and computation of the stationary points of an n-dimensional real function, which can be applied to problems with imprecise function values, since it requires only the signs of gradient values to be correct. The proposed algorithm implements topological degree theory and especially the concept and properties of the characteristic n-polyhedron by which we avoid all calculations concerning the exact value of the topological degree.

This method consists of two principal parts. In the first part, a stationary point of a continuously differentiable function $f\colon \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$, where $\mathcal{D}$ is open and bounded, is localized within a given n-dimensional polyhedron. This procedure is based on a rootfinding algorithm [9–11]. Specifically, the nonzero value of the topological degree of the gradient of f, ∇f, at the origin $\Theta = (0,\ldots,0)$ relative to the polyhedron P, denoted by $\deg[\nabla f, P, \Theta]$, is examined. In the case where a nonzero value of the topological degree has been obtained a stationary point of the function f is located within the polyhedron. In the second part of the algorithm, a new criterion is proposed for the characterization of the located stationary point as a minimum, maximum or saddle point. This procedure is based on the property of the examined polyhedron to be a characteristic n-polyhedron with a specific orientation on its vertices. This criterion does not require derivatives of ∇f or approximations of them, but only the algebraic sign of ∇f. The located stationary point can be computed if required.

2. The method

To locate and compute a stationary point of a continuous function $f\colon \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$, where $\mathcal{D}$ is open and bounded, we implement a generalized bisection method based on the notion of the *characteristic n-dimensional polyhedron* (CP) [9–11]. To define CP, let B_k^n be the n-digit binary representation of the integer $k-1$, $1 \le k \le 2^n$, counting the leftmost digit first. Then the *n-binary matrix* $\mathcal{M}_n^* = [C_{ij}^*]$, $i = 1,2,\ldots,2^n$, $j = 1,2,\ldots,n$, is the matrix whose entry in the i-th row and j-th column is the j-th digit of B_i^n. By replacing each zero element in the matrix $\mathcal{M}_n^*$ by -1 we get a new $2^n \times n$ matrix $\mathcal{M}_n = [C_{ij}]$, which we call an *n-complete* matrix; i.e. for $n = 2$ we have:

$$\begin{array}{l} B_1^2 = 00 \\ B_2^2 = 01 \\ B_3^2 = 10 \\ B_4^2 = 11 \end{array} \longrightarrow \quad \mathcal{M}_2^* = \begin{bmatrix} B_1^2 \\ B_2^2 \\ B_3^2 \\ B_4^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \longrightarrow \quad \mathcal{M}_2 = \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Suppose now that $\Pi^n = \langle v^1, v^2, \ldots, v^{2^n} \rangle$ is an oriented n-polyhedron in $\mathbb{R}^n$ with 2^n vertices and let $F = (f_1, f_2, \ldots, f_n)\colon \Pi^n \subset \mathbb{R}^n \to \mathbb{R}^n$. Then the *matrix of signs associated with F and Π^n*, denoted $\mathcal{S}(F;\Pi^n)$, is the $2^n \times n$ matrix whose entries in the i-th row are the corresponding coordinates of the vector $\operatorname{sgn} F(v^i) = \big(\operatorname{sgn} f_1(v^i), \operatorname{sgn} f_2(v^i), \ldots, \operatorname{sgn} f_n(v^i)\big)$, where sgn defines the well known sign function. An n-polyhedron Π^n is a CP if $\mathcal{S}(F;\Pi^n) \equiv \mathcal{M}_n$. Under some suitable assumptions on its boundary, a CP always contains at least one solution of the system $F(X) = \Theta$ (CP-criterion), since the absolute value of $\deg[F, \mathrm{CP}, \Theta]$ is equal to one [13]. In order to approximate this solution, a generalized bisection method is used, in combination with the CP-criterion outlined above, which bisects a CP in such a way that the new refined n-polyhedron is also a CP. To do this, we compute the midpoint of a proper 1-simplex (edge) of Π^n and use it to replace that vertex of Π^n for which the vectors of their signs are identical. Finally, the number B of characteristic bisections of the edges of a Π^n required to obtain a new refined CP, $\Pi_\star^n$, whose longest edge length, $\Delta(\Pi_\star^n)$, satisfies $\Delta(\Pi_\star^n) \le \varepsilon$, for some $\varepsilon \in (0,1)$, is given by $B = \left\lceil \log_2\!\big(\Delta(\Pi^n)\,\varepsilon^{-1}\big)\right\rceil$, (for details see [9, 10, 13]).

It is important to notice that the CP-criterion avoids all calculations concerning the topological degree since it requires not its exact value but only its nonzero value.

The procedure outlined above can be implemented for ∇f, in order to determine the stationary points of f. Specifically, the problem of computing a stationary point of a continuously differentiable function $f: \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$, where $\mathcal{D}$ is open and bounded, is equivalent to obtaining the corresponding solution $x^* \in \mathcal{D}$ of the equation $\nabla f(x) = \Theta$. In the case where a nonzero value of $\deg[\nabla f, P, \Theta]$, relative to an n-dimensional polyhedron P, has been obtained, a stationary point of the function f is located within this polyhedron. This procedure makes use only of the algebraic signs of ∇f, while derivatives of ∇f or approximations of them are not required.

Now, using the concept and properties of the CP we can also characterize the located stationary points of a function as minimum, maximum or saddle points. This can be done as soon as a CP is constructed. According to the orientation of its vertices, the included stationary point is characterized and can be accurately computed, if required. If saddle or maxima points are not required then our algorithm does not proceed with their computation. Also, when a stationary point is given the method can easily characterize it. Our experience is that this criterion behaves predictably and reliably. The following theorem clarifies this "characterization" procedure.

Theorem 2.1. *Let $f: \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable in an open neighborhood $\mathcal{D}$ of a point $x^* \in \mathcal{D}$ for which $\nabla f(x^*) = \Theta$ and the Hessian of f at x^*, $\nabla^2 f(x^*)$, is positive definite. Then there exists an oriented* CP, *such that $\mathcal{S}(\nabla f;\ \mathrm{CP}) \equiv \mathcal{M}_n$, which includes the minimizer x^* of f.*

Proof. Clearly, the necessary and sufficient conditions for the point x^* to be a local minimizer of the function f are satisfied by the hypothesis $\nabla f(x^*) = \Theta$ and the assumption that $\nabla^2 f(x^*)$ is positive definite (see for example [1]).

Consider the 2^n vectors $p = (p_1, p_2, \ldots, p_n)$ whose coordinates are nonzero and their signs form all possible combinations of -1 and 1. Then there exists a point $z \in (x^*, x^* + p)$ for which the following relation holds:

$$f(x^* + p) = f(x^*) + \nabla f(z)^\top p.$$

Now, since x^* minimizes f, the inequality $f(x^* + p) > f(x^*)$ holds in every direction p and consequently $\nabla f(z)^\top p$ will be positive. Therefore, the points z can form the vertices of an oriented CP such that $\mathcal{S}(\nabla f;\ \mathrm{CP}) \equiv \mathcal{M}_n$. Thus the theorem is proved.

Based on the results of the above theorem and the properties of a characteristic polyhedron we are able to characterize the located stationary points.

To do this we transform the oriented $\mathrm{CP} = \langle v^1, v^2, \ldots, v^{2^n} \rangle$ so that its n proper 1-simplexes with a common vertex are edges of the polyhedron with vertices formed by the rows of the $2^n \times n$ matrix $\mathcal{R}$, defined as follows. Let $x_j^{\min} = \min\{v_j^1, v_j^2, \ldots, v_j^{2^n}\}$, $x_j^{\max} = \max\{v_j^1, v_j^2, \ldots, v_j^{2^n}\}$ be the minimum and maximum of all the j-th components v_j^i of the CP vertices v^i, respectively. Then we define the matrix $\mathcal{R} = \mathcal{G} + \mathcal{M}_n^* \mathcal{B}$, where $\mathcal{G}$ is the rank-1, $2^n \times n$, matrix with elements in the j-th column having the value $x_j^{\min}$ and $\mathcal{B}$ is the $n \times n$ diagonal matrix with ith element the difference $h_i = x_i^{\max} - x_i^{\min}$. For example, for $n = 2$ we have:

$$\mathcal{R} = \mathcal{G} + \mathcal{M}_2^* \mathcal{B} = \begin{bmatrix} x_1^{\min} & x_2^{\min} \\ x_1^{\min} & x_2^{\min} \\ x_1^{\min} & x_2^{\min} \\ x_1^{\min} & x_2^{\min} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} h_1 & 0 \\ 0 & h_2 \end{bmatrix} = \begin{bmatrix} x_1^{\min} & x_2^{\min} \\ x_1^{\min} & x_2^{\max} \\ x_1^{\max} & x_2^{\min} \\ x_1^{\max} & x_2^{\max} \end{bmatrix}.$$

Next, we construct the matrix $\mathcal{S}(\nabla f;\ \mathcal{R})$ and we distinguish the following cases:

a) If 2^{n-1} rows of $\mathcal{S}(\nabla f;\, \mathcal{R})$ with the same sign in one of their columns are identical with the corresponding rows of $\mathcal{M}_n$, then x^* is considered to be a local minimum.

b) If 2^{n-1} rows of $\mathcal{S}(\nabla f;\, \mathcal{R})$ with the same sign in one of their columns are identical with the corresponding rows of $-\mathcal{M}_n$, then x^* is considered to be a local maximum.

c) Otherwise, if these rows are not identical with the corresponding rows of $\mathcal{M}_n$ or $-\mathcal{M}_n$, x^* is considered to be a saddle point.

3. Numerical applications

The above procedures were implemented using a new portable Fortran program named MINBIS, which has been applied to several test functions. Our experience is that the algorithm behaves predictably and reliably. The results were satisfactory without any redundant function evaluations. Some typical computational results are given below where the reported parameters are: n dimension; $x^0 = (x_1, x_2, \ldots, x_n)$ starting point; $h = (h_1, h_2, \ldots, h_n)$ stepsizes in each coordinate direction used to form the starting polyhedron [9]; δ (= 0.625E−5 for the following examples) positive input parameter (if it is less than the machine precision it is set equal to 0.0625) that is used for the construction of CP [10]; $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$ approximated local stationary point computed within an accuracy of ε (predetermined precision not less than the machine precision, $\varepsilon = 10^{-8}$ for the following examples); NFE the total number of function evaluations for the characterization and computation of a stationary point.

Example 3.1. *Kearfott function*, [5]. The objective function f is given by:

$$f(x) = (x_1^2 + x_2^2 - 2)^2 + (x_1^2 - x_2^2 - 1)^2$$

with four local minima $x_1^* = (-\sqrt{1.5}, -\sqrt{0.5})$, $x_2^* = (-\sqrt{1.5}, \sqrt{0.5})$, $x_3^* = (\sqrt{1.5}, \sqrt{0.5})$, $x_4^* = (\sqrt{1.5}, -\sqrt{0.5})$, one maximum $x_5^* = (0,0)$, and four saddle points $x_6^* = (\sqrt{1.5}, 0)$, $x_7^* = (-\sqrt{1.5}, 0)$, $x_8^* = (0, \sqrt{0.5})$, $x_9^* = (0, -\sqrt{0.5})$. In Table 1 we exhibit indicative results obtained by MINBIS for various instances of the problem.

x^0	h	x_i^*	NFE	Characterization
$(-1.5, -1.5)$	$(1,1)$	x_1^*	67	minimum
$(-1.5, 0.5)$	$(1,1)$	x_2^*	68	minimum
$(0.5, 0.5)$	$(1,1)$	x_3^*	67	minimum
$(0.5, -1.5)$	$(1,1)$	x_4^*	68	minimum
$(-0.5, -0.5)$	$(1,1)$	x_5^*	7	maximum
$(1, -0.5)$	$(1,1)$	x_6^*	71	saddle
$(-1.5, -0.5)$	$(1,1)$	x_7^*	64	saddle
$(-0.5, 0.5)$	$(1,1)$	x_8^*	99	saddle
$(-0.5, -1.5)$	$(1,1)$	x_9^*	142	saddle

Table 1. Kearfott function, $n = 2$

Example 3.2. *Himmelblau function*, [3]. In this case f is given by:

$$f(x) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2.$$

Executing the implemented program, one finds nine stationary points, which are: four local minima $x_1^* = (3,2)$, $x_2^* = (3.584428, -1.848126)$, $x_3^* = (-3.77931, -3.283186)$ and $x_4^* =$

$(-2.805118, 3.131312)$, one maximum $x_5^* = (-0.2708446, -0.9230387)$, and four saddle points $x_6^* = (0.8667755, 2.884255)$, $x_7^* = (-3.073026, -0.8135307)$, $x_8^* = (3.385154, 0.07385179)$ and $x_9^* = (-0.1279613, -1.953715)$. In Table 2 we exhibit some of the corresponding results obtained by MINBIS for various instances of the problem.

x^0	h	x_i^*	NFE	Characterization
$(2,1)$	$(2,2)$	x_1^*	21	minimum
$(1,1)$	$(4,4)$	x_1^*	102	minimum
$(3,-2)$	$(2,2)$	x_2^*	107	minimum
$(-4,-4)$	$(1,1)$	x_3^*	71	minimum
$(-3,3)$	$(2,2)$	x_4^*	84	minimum
$(-0.5,-1)$	$(1,1)$	x_5^*	75	maximum
$(-1,-1)$	$(3,3)$	x_5^*	96	maximum
$(0,0)$	$(3,3)$	x_6^*	48	saddle
$(-5,-2)$	$(3,3)$	x_7^*	69	saddle
$(3.2,-0.2)$	$(0.6,1)$	x_8^*	96	saddle
$(-1,-3)$	$(2,2)$	x_9^*	93	saddle

Table 2. Himmelblau function, $n = 2$

Example 3.3. *Identity function*, [5]. In this case f is given by:

$$f(x) = \frac{1}{2}\sum_{i=1}^{n} x_i^2.$$

This function has the minimum $x^* = (0, 0, \ldots, 0)$. Various starting points and stepsizes were utilized successfully. The algorithm appears to find and characterize the stationary points independently of the dimension n of the problem. In Table 3 we exhibit indicative results obtained by MINBIS for various dimensions n.

n	x^0	h	x_i^*	NFE	Characterization
2	$(-2,-2)$	$(4,4)$	x^*	7	minimum
3	$(-2,-2,-2)$	$(4,4,4)$	x^*	13	minimum
4	$(-2,-2,-2,-2)$	$(4,4,4,4)$	x^*	25	minimum
5	$(-2,-2,\ldots,-2)$	$(4,4,\ldots,4)$	x^*	49	minimum
6	$(-2,-2,\ldots,-2)$	$(4,4,\ldots,4)$	x^*	97	minimum
7	$(-2,-2,\ldots,-2)$	$(4,4,\ldots,4)$	x^*	193	minimum
8	$(-2,-2,\ldots,-2)$	$(4,4,\ldots,4)$	x^*	363	minimum
9	$(-2,-2,\ldots,-2)$	$(4,4,\ldots,4)$	x^*	705	minimum

Table 3. Identity function, $n = 2, 3, \ldots, 9$

Example 3.4. *Extended Kearfott function*, [5]. In this case the components of the gradient are given by:

$$\begin{aligned}\frac{\partial f(x)}{\partial x_i} &= {x_i}^2 - x_{i+1}, \quad i = 1, 2, \ldots, n-1,\\ \frac{\partial f(x)}{\partial x_n} &= {x_n}^2 - x_1.\end{aligned}$$

n	x^0	h	x_i^*	NFE	Characterization
2	$(0.5, 0.5)$	$(1, 1)$	x_1^*	7	minimum
2	$(0.5, 0.5)$	$(1000, 1000)$	x_1^*	41	minimum
3	$(0.5, 0.5, 0.5)$	$(1, 1, 1)$	x_1^*	13	minimum
3	$(0.5, 0.5, 0.5)$	$(1000, 1000, 1000)$	x_1^*	42	minimum
4	$(0.5, 0.5, 0.5, 0.5)$	$(1, 1, 1, 1)$	x_1^*	25	minimum
4	$(0.5, 0.5, 0.5, 0.5)$	$(1000, 1000, 1000, 1000)$	x_1^*	54	minimum
5	$(0.5, 0.5, \ldots, 0.5)$	$(1, 1, \ldots, 1)$	x_1^*	49	minimum
5	$(0.5, 0.5, \ldots, 0.5)$	$(1000, 1000, \ldots, 1000)$	x_1^*	78	minimum
6	$(0.5, 0.5, \ldots, 0.5)$	$(1, 1, \ldots, 1)$	x_1^*	65	minimum
6	$(0.5, 0.5, \ldots, 0.5)$	$(1000, 1000, \ldots, 1000)$	x_1^*	94	minimum

Table 4. Extended Kearfott function, $n = 2, 3, \ldots, 6$

The function has the minima $x_1^* = (1, 1, \ldots, 1)$ and $x_2^* = (0, 0, \ldots, 0)$. Indicative results for various dimensions n are exhibited in Table 4.

The algorithm was tested on several other problems with satisfactory results. It must be noticed that for different starting values or different input parameter δ one gets different number of function evaluations. In all cases, one local stationary point is localized, successfully characterized and computed within the given accuracy.

References

[1] Dennis, Jr., J. E. and Schnabel, R. B. *Numerical methods for unconstrained optimization and nonlinear equations.* Prentice-Hall, 1983.

[2] Grapsa, T. N. and Vrahatis, M. N. *A dimension-reducing method for unconstrained optimization.* J. Comput. Appl. Math., to appear.

[3] Himmelblau, D. M. *Applied nonlinear programming.* McGraw-Hill, NY, 1972.

[4] Kavvadias, D. J. and Vrahatis, M. N. *Locating and computing all the simple roots and extrema of a function.* SIAM J. Sci. Comput., to appear.

[5] Kearfott, R. B. *An efficient degree—computation method for a generalized method of bisection.* Numer. Math. **32** (1979), pp. 109–127.

[6] Kupferschmid, M. and Ecker, J. G. *A note on solution of nonlinear programming problems with imprecise function and gradient values.* Math. Program. Study **31** (1987), pp. 129–138.

[7] Magoulas, G. D., Vrahatis, M. N., Grapsa, T. N., and Androulakis, G. S. *An efficient training method for discrete multilayer neural networks.* Ann. Math. Artif. Intel., to appear.

[8] Magoulas, G. D., Vrahatis, M. N., Grapsa, T. N., and Androulakis, G. S. *Neural network supervised training based on a dimension reducing method.* Ann. Math. Artif. Intel., to appear.

[9] Vrahatis, M. N. *Solving systems of nonlinear equations using the nonzero value of the topological degree.* ACM Trans. Math. Software **14** (1988), pp. 312–329.

[10] Vrahatis, M. N. *CHABIS: A mathematical software package for locating and evaluating roots of systems of non-linear equations.* ACM Trans. Math. Software **14** (1988), pp. 330–336.

[11] Vrahatis, M. N. *An efficient method for locating and computing periodic orbits of nonlinear mappings.* J. Comp. Phys. **119** (1995), pp. 105–119.

[12] Vrahatis, M. N., Androulakis, G. S., and Manoussakis, G. E. *A new unconstrained optimization method for imprecise function and gradient values.* J. Math. Anal. Appl., to appear.

[13] Vrahatis, M. N. and Iordanidis, K. I. *A rapid generalized method of bisection for solving systems of nonlinear equations.* Numer. Math. **49** (1986), pp. 123–138.

Received: November 1, 1995
Revised version: December 8, 1995

Department of Mathematics
University of Patras
Patras, GR–261.10
Greece
E-mail: {vrahatis, ect}@math.upatras.gr