A NEW APPROACH TO GLOBAL OPTIMIZATION USING IDEAS FROM NONLINEAR STABILITY THEORY

Shashikala H., N. K. Sancheti & S. S. Keerthi

Department of Computer Science & Automation
Indian Institute of Science
Bangalore-560 012, India

Abstract - A new method for determing and connecting all the local minima and local maxima of a function on a compact manifold is given. The method is based on properties of stability regions associated with the equilibria of the gradient vector field. Applications of the method include global optimization and nonlinear equation solving.

I. INTRODUCTION

In this paper we give a method for determining the global optimum of a function on a compact manifold. The method determines and connects all the local minima and local maxima of a function on a compact manifold. The aim here is only to describe the basic ideas of the method. Detailed proofs of the results are given in [11]. Let C^2 denote the set of functions from $R^n \rightarrow R$ whose derivatives upto second order are continuous. Given f and $h_i, i = 1, \ldots, m$, to be functions in C^2, define

$$\widetilde{\Omega} = \{x \in R^n : h_i(x) = 0, i = 1, \ldots, m\}. \tag{1.1}$$

Assume that $\widetilde{\Omega}$ is a compact differentiable manifold. Let $x^* \in \widetilde{\Omega}$ be given. We consider the problem of connecting all the local minima and local maxima of f on the set $\Omega \subset \widetilde{\Omega}$ defined by

$$\Omega = \text{the connected component of } \widetilde{\Omega} \text{ containing } x^*. \tag{1.2}$$

Clearly, Ω is a connected, compact differentiable manifold. Further, $\Omega = \widetilde{\Omega}$ if $\widetilde{\Omega}$ is connected. The following definition is standard in optimization literature, and is used in this paper.

Definition 1. Consider

$$X = \{x \in R^n \ : \ \begin{array}{l} h_i(x) = 0, i = 1, \ldots, m, \\ g_j(x) \leq 0, j = 1, \ldots, l\}, \end{array} \tag{1.3}$$

where $h_i, g_j \in C^2$. Let $\bar{x} \in X$ and $J = \{j : g_j(\bar{x}) = 0\}$. $\bar{x}$ is said to be *regular* if the set of gradients, $\{\nabla h_i(\bar{x}), i = 1, \ldots, m\} \cup \{\nabla g_j(\bar{x}) : j \in J\}$ is linearly independent.

Assumption A_1 All $x \in \Omega$ are regular.

The above assumption is generically true. Define the Lagrangian $L : R^n \times R^m \rightarrow R$ of f over Ω as $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$. A point $\bar{x} \in \Omega$ is said to be a *critical point of f over Ω* if $\exists \bar{\lambda} \in R^m$ such that

$$\nabla_x L(\bar{x}, \bar{\lambda}) = \left(\frac{\partial L}{\partial x_1}(\bar{x}, \bar{\lambda}), \cdots, \frac{\partial L}{\partial x_n}(\bar{x}, \bar{\lambda}) \right)^T = 0,$$

where T denotes transpose. By Assumption A_1, the $\bar{\lambda}$ associated with $\bar{x}$ is unique.

Let: Z denote the set of all critical points of f over Ω; and, MIN and MAX represent the set of all local minima and local maxima of f over Ω respectively. Assumption A_1 allows Langrange multiplier rule to be used and so we get $MIN \cup MAX \subset Z$. $SAD = Z \backslash (MIN \cup MAX)$ is the set of all saddle points.

There have been many attempts to solve the problem of global optimization. A good survey of global optimization methods can be found in [1]. Probabilistic methods such as simulated annealing and its variants also solve the global optimization problem [8, 2]. Special homotopy methods for determining all critical points of a polynomial system are well-known [9]. Once all the critical points are known, the global minimum is also known. However these methods have practical value only if the degree of the system is small. Also the extension of the homotopies to general nonlinear systems is unclear. Further, the homotopy curve is in the complex domain and may consist of several components. Deiner [6] uses a homotopy formulation and gives a recursive construction that yields a one dimensional web containing all the critical points. But he needs an algorithm to find all the roots of a system of nonlinear equations. Thus, it is not of much practical value. Canny [4] has given a provably correct algorithm to connect all the critical points of a polynomial over a semi-algebraic set. Apart from its restriction to problems with algebraic description, Canny's method is difficult to implement and has not been proven to be practically useful. Branin [3] tries to get a single trajectory connecting all the critical points by alternately maximizing and minimizing the function. Treccani [12] gives a counterexample where Branin's method fails.

In our approach, we form a directed graph G with all the local minima and local maxima as its vertices. G is a bipartite graph with MIN and MAX as partitions. An edge connecting x_i to x_j has an associated weight vector which encodes a way of connecting x_i to x_j via a continuous path on Ω. We can obtain the global optimum by traversing this graph. In the next section we describe the basic ideas of our method. In section 3 we give some useful applications of our method.

II. OUR METHOD

We begin with some definitions. Let Ω be as defined by (1.2). For $\bar{x} \in \Omega$, denote the tangent space of Ω at $\bar{x}$ by $T_{\bar{x}}(\Omega)$.

Definition 2. Consider x^*, a critical point of f over Ω.

0-7803-0593-0/92 $3.00 1992 IEEE

Let: λ^* denote the Lagrangian multiplier vector associated with x^*; $\nabla^2_x L(x^*, \lambda^*)$ denote the Hessian of L with respect to x at x^*; B be a $n \times (n-m)$ matrix whose columns form an orthogonal basis for $T_{x^*}(\Omega)$; $H = B^T \nabla^2_x L(x^*, \lambda^*) B$; and $\alpha_1, \ldots, \alpha_{n-m}$ be the eigen values of H. We say that x^* is a *hyperbolic critical point* if $\alpha_i \neq 0, \ i = 1, \ldots, (n-m)$.

Assumption A_2 All critical points of f over Ω are hyperbolic.

The above assumption holds generically [10]. By compactness of Ω, there will only be a finite number of hyperbolic critical points of f over Ω [7]. Two differntiable manifolds A and B in Ω are said to be *transversal* if either $T_x(A) + T_x(B) = T_x(\Omega) \ \forall x \in A \cap B$, or $A \cap B = \phi$.

Let $\dot{x} = dx/dt$ and consider the following differential-algebraic systems (in x and λ):

$$\dot{x} = -\nabla_x L(x, \lambda), \quad < \nabla h_i, \dot{x} >= 0, \ i = 1, \ldots, m; \quad (2.1)$$

$$\dot{x} = \nabla_x L(x, \lambda), \quad < \nabla h_i, \dot{x} >= 0, \ i = 1, \ldots, m. \quad (2.2)$$

These systems denote the vector fields in which $\dot{x}$ is the projection of $-\nabla f(x)$ and $\nabla f(x)$ onto $T_x(\Omega)$. Given $x_0 \in \Omega$, let $\Phi^-(t, x_0)$ and $\Phi^+(t, x_0)$ denote the solutions of (2.1) and (2.2) respectively, with $x(0) = x_0$. All solutions of (2.1) and (2.2) which start in Ω will remain in Ω and asymptotically reach one of the critical points of f over Ω [5].

Let $\bar{x}$ be a critical point of f over Ω. Define,

$$W^-(\bar{x}) = \{x \in \Omega : \lim_{t \to \infty} \Phi^-(t, x) = \bar{x}\},$$

and,

$$W^+(\bar{x}) = \{x \in \Omega : \lim_{t \to \infty} \Phi^+(t, x) = \bar{x}\}.$$

If $\bar{x} \in MIN$ then $W^-(\bar{x})$ denotes the region of attraction of $\bar{x}$. Similarly, if $\bar{x} \in MAX$, then $W^+(\bar{x})$ is the region of attraction of $\bar{x}$.

Assumption A_3 $W^+(\bar{x})$ and $W^-(\bar{y})$ intersect transversally for all $\bar{x}, \bar{y} \in Z$.

Assumption A_3 is also generic [5]. Given $x \in R^n$ and $\epsilon > 0$, let $B_\epsilon(x) = \{z : \|z - x\| < \epsilon\}$, where $\|\cdot\|$ denotes the euclidean norm in R^n. We define a local maximum y to be *adjacent* to a local minimum x if $\forall \epsilon > 0 \exists z \in B_\epsilon(x)$ such that $\lim_{t \to \infty} \Phi^+(t, z) = y$. Similarly a local minimum x is said to be *adjacent* to a local maximum y if $\forall \epsilon > 0, \exists z \in B_\epsilon(y)$ such that $\lim_{t \to \infty} \Phi^-(t, z) = x$. With x and y as given above, it is easy to see that y is adjacent to x iff x is adjacent to y. Hence we can simply talk in terms of x and y being adjacent. As stated before, our aim is to connect all the local maxima and local minima of f over Ω. We form a directed graph G as follows. The set of vertices of G is $MIN \cup MAX$. An edge is connected from vertex x to vertex y if x and y are adjacent. The following theorem is proved in [11]. The proof uses the results in [5] strongly.

Theorem 1. Suppose Assumptions A_1, A_2 and A_3 hold. Then G is connected.

Two important remarks should be made here. First, Theorem 1 may not hold if Ω is not compact. Second, if we take Z as the set of vertices and define a corresponding graph $\tilde{G}$ (an extension of G) then Theorem 1 will hold if G is replaced by $\tilde{G}$. In fact, the proof of this modified theorem is much easier. We have chosen G instead of $\tilde{G}$ because saddle points are difficult to reach, i.e., only a set of measure zero starting points in Ω will reach saddles under the flow defined by (2.1) or (2.2).

Consider the construction of G. Let x be a local minimum and $y_1, \ldots, y_l$ be the local maxima adjacent to x. We need an algorithm A^+ which takes x as input and for any given $\epsilon > 0$, outputs y_i and perturbations p_i such that $\|p_i\| \leq \epsilon$ and $\lim_{t \to \infty} \Phi^+(t, x + p_i) = y_i$, for $i = 1, \ldots, l$. In other words A^+ connects a local minimum to all its adjacent local maxima. We also need an algorithm A^- which takes a local maximum as input and outputs all its adjacent local minima with the corresponding perturbations. An exact realization of A^+ and A^- is hard. For practical implementation, A^+ and A^- are replaced by heuristic approximations. Some powerful heuristics for the implementation of A^+ and A^- are described in detail in [11] and briefly sketched further below.

The construction of G is given by the following conceptual algorithm.

Algorithm form_graph(x^*) ($x^* \in \Omega$ is a given point).

1: Set $V_1 = \phi$, $V_2 = \phi$, $E = \phi$, $V = \phi$.
Choose a small $\epsilon > 0$ for use in A^+ and A^-.
2: Integrate (2.1) to find $\bar{x} = \lim_{t \to \infty} \Phi^-(t, x^*)$.
Set: $V_1 = \{\bar{x}\}$, $V = \{\bar{x}\}$
3: While ($V \neq \phi$) do
Begin
Pick $x \in V$. Set $V = V \backslash \{x\}$.
If $x \in MIN$, apply A^+ to find all its adjacent local maxima $y_1, \ldots, y_l$ and corresponding perturbations $p_1, \ldots, p_l$ such that $\lim_{t \to \infty} \Phi^+(t, x + p_i) = y_i$, $i = 1, \ldots, l$, $\|p_i\| < \epsilon$.
Set: $V = V \cup [\{y_1, \ldots, y_l\} \backslash V_2]$;
$V_2 = V_2 \cup \{y_1, \ldots, y_l\}$; and,
$E = E \cup \{(x, y_1, p_1, 1), \ldots, (x, y_l, p_l, 1)\}$.
Else
If $x \in MAX$, apply A^- to find all the adjacent local minima $z_1, \ldots, z_r$ and corresponding perturbations $q_1, \ldots, q_r$ such that $\lim_{t \to \infty} \Phi^-(t, x + q_i) = z_i$, $i = 1, \ldots, r, \|q_i\| < \epsilon$.
Set: $V = V \cup [\{z_1, \ldots, z_r\} \backslash V_1]$;
$V_1 = V_1 \cup \{z_1, \ldots, z_r\}$; and,
$E = E \cup \{(x, z_1, q_1, -1), \ldots, (x, z_r, q_r, -1)\}$.
End.

At the end of the above procedure we get two partitions of vertices V_1 and V_2 and an edge-list, E. With probability one we can say that the $\bar{x}$ determined in step 2 is a local minimum [5]. By Theorem 1, the graph determined by V_1, V_2 and E is the same as G. Thus V_1 is the set of local minima and V_2 is the set of local maxima of f over Ω. Each $e \in E$ is of the form $e = (x, y, p, s)$. $s = 1$ implies that x is a local minimum, y is a local maximum and $\lim_{t \to \infty} \Phi^+(t, x + p) = y$. Similarly $s = -1$ implies that x is a local maximum, y is a local minimum, and $\lim_{t \to \infty} \Phi^-(t, x + p) = y$.

Now we briefly give the ideas behind three heuristics for the implementation of A^-. The ideas for A^+ are similar. Let $x^* \in MAX$. The aim is to determine a set of "small" perturbation vectors with the property that, the solutions of (2.1), starting from the initial conditions generated by applying these perturbations on x^*, will lead to the set of all adjacent local minimizers of x^*. Since a solution of (2.1) is a solution of (2.2) with t reversed, a study of (2.2) near x^* is helpful.

Let $H(x^*)$ and $\alpha_1, \ldots, \alpha_{n-m}$, be as in Definition 2. Generically, $\alpha_1, \ldots, \alpha_{n-m}$ are distinct. For $i = 1, \ldots, n$, let v_i denote the normalized eigen-direction in $T_{x^*}(\Omega)$ corresponding to the eigenvalue α_i. It is a well_known fact in the theory of dynamical systems [10] that a solution of (2.2) terminating at x^* will asymptotically reach x^* along one of the v_i axes. This observation leads us to the *eigen-axis*

heuristic: choose $\{\pm\beta v_i\}$ as a set of perturbation vectors, where β is a small positive number.

Let α_k denote the largest element of $\{\alpha_i\}$ (α_k is the element nearest to zero). It can be shown [10] that only a set of measure zero initial conditions in $W^+(x^*)$ lead to solutions of (2.2) which approach x^* along axes other than v_k. Thus we get the *dominating eigen-axis heuristic*: concentrate the choice of perturbation vectors around the direction, v_k. One way of implementing this heuristic is as follows. Let: C^+ denote the cone in $T_{x^*}(\Omega)$ with apex at x^*, apex angle = θ (a small amount, say 5 degrees) and axis along v_k; C^- denote the symmetrically opposite cone at x^* with axis along $-v_k$; and $NTOL$ be a positive integer, say $NTOL = 3$ (larger the $NTOL$, better the heuristic but more is the computational effort). Choose the perturbation vectors from C^+ and C^- randomly with a uniform distribution until $NTOL$ consecutive perturbations fail to yield a new minimizer.

In the above heuristics, if perturbations are not chosen carefully along the eigen axes, the solutions of (2.1) can go to saddle points. So it may be useful to try perturbations away from the eigen axes. The third heuristic, called as *eigen-orthant heuristic*, is based on this observation. We choose one perturbation vector from the interior of each of the 2^{n-m} orthants of the coordinate system for $T_{x^*}(\Omega)$ formed by $v_1, \ldots, v_{n-m}$. If $(n-m)$ is large then this heuristic is not suitable and it can be replaced by a modification.

In numerical experiments we have found the dominating eigen-axis heuristic to be the most effective. We also note that, saddle points are encountered when the above heuristics are used. However there is no harm in including them in an extended graph of the extrema.

III. APPLICATIONS

Now we describe some examples in constrained global optimization and nonlinear equation solving which have been solved using our method. When applying our method, we will assume that the generic assumptions A_1, A_2 and A_3 always hold.

Constrained global optimization: Consider the global solution of

$$\min \quad f(x) \quad \text{s.t.} \quad x \in X \tag{3.1}$$

where $f \in C^2$ and X is as given by (1.3). Assume: X is compact and connected and, $\forall x \in X$, x is a regular point. Suppose $x^0 \in X$ is given. To solve (3.1) using the ideas of section 2, we introduce new varaibles $z = (z_1, \ldots, z_l)$ and define,

$$\begin{aligned}\Omega = \{(x,z) : x \in R^n, z \in R^l,\\ h_i(x) = 0, i = 1, \ldots, m,\\ g_j(x) + {z_j}^2 = 0, j = 1, \ldots, l\}.\end{aligned}$$

It is easy to verify that Ω is a smooth connected compact manifold and that, $\forall (x,z) \in \Omega$, (x,z) is a regular point. Now we set ${z_j}^0 = \sqrt{-g_j(x^0)}$ and use form_graph(x^0, z^0) to get the set of all the local minima and local maxima of (3.1). This set contains all the global minima and global maxima. Suppose the unconstrained global minimum of a function f is to be found and it is known to exist. This problem can be tackled by choosing a large radius R, setting $X = \{x : \|x\|^2 \leq R^2\}$ and solving (3.1).

Example 1. Consider the minimization of the six hump camel back function defined by

Critical Point	Coordinates	Adjacent from
min1	(-0.0898,0.7126)	-
max1	(-1.23,-0.1623)	min1
max2	(3,1.5)	
max3	(-1.0137,1.5)	
min2	(1.703,0.796)	max1
min3	(0.0898,-0.7126)	
min4	(-1.6071,-0.5686)	
max4	(-3,-1.5)	min2
max5	(3,1.5)	
max6	(1.23,0.1623)	min3
max7	(1.0137,-1.5)	
min5	(1.607,0.5686)	max2
min6	(1.7036,-0.796)	max6
max8	(3,-1.5)	min6

Table 1: Results of Example 1

$$f(x) = 4x_1^2 - 2.1x_1^4 + x_1^6/3 + x_1x_2 - 4x_2^2 + 4x_2^4,$$

subject to

$$-3 \leq x_1 \leq 3, -1.5 \leq x_2 \leq 1.5.$$

This function is known to contain 6 local minima, two of which are also global minima. The set of local minima and maxima obtained by the heuristics described are adequately detailed in Table 1. The last column represents the critical point to which the critical point in the first column was found adjacent. The columns in Tables 2 and 3 (obtained in further examples) should also be interpreted in the same way. The critical point min1 was found by following the negative gradient field from (0.1,0.1). Then form_graph(min1) yielded the results shown in Table 1.

Example 2. Now consider the minimization of the Treccani function defined by

$$f(x) = x_1^4 + 4x_1^3 + 4x_1^2 + x_2^2,$$

subject to

$$-3 \leq x_1 \leq 3, -3 \leq x_2 \leq 3.$$

This function contains two local minima which are also global minima. The set of local minima and maxima obtained by the heuristics described are adequately detailed in Table 2. The critical point min1 was found by following the negative gradient field from (0.1,0.1). Then form_graph(min1) yielded the results shown in Table 2.

Finding all solutions of a system of nonlinear equations: Consider the solution of

$$g(x) = 0 \tag{3.2}$$

where $g : R^n \to R^n$ and each component of g is in C^2. Suppose we want to find all the solutions of (3.2) whose norm is bounded by a fixed number r. In other words, we have to find

$$S = \{x : g(x) = 0, x^T x \leq r^2\}.$$

If $f(x) = \|g(x)\|^2$ and $X = \{x : x^T x \leq r^2\}$ then S is nothing but the set of all global minma of (3.1) with f-value zero. Thus S can be determined by our method.

Critical Point	Coordinates	Adjacent from
min1	(0,0)	-
max1	(3,3)	min1
max2	(3,-3)	
max3	(-1,-3)	
max4	(-1,3)	
min2	(-2,0)	max3
max5	(-3,-3)	min2
max6	(-3,3)	

Table 2: Results of Example 2

Critical Point	Coordinates	Adjacent from
min1	(0.1,0.1,0.1,0.1)	-
max1	(0.0168,-3.971,0.474,0.00157)	min1
max2	(0.474,0.00157,0.0168,-3.971)	
max3	(0.498,0.0028,-0.039,3.968)	
max4	(3.968,0.498,0.0028,-0.039)	
max5	(-3.971,0.474,0.00157,0.0168)	
sad1	(-0.039,3.968,0.498,0.0028)	
min2	(-0.9,-0.9,-0.9,-0.9)	max1
max6	(0.00157,0.0168,-3.971,0.474)	min2

Table 3: Results of Example 3

Example 3. Consider the problem of finding all the roots of the following system of nonlinear equations.

$$\begin{aligned} g_1 &= (x_1 - 0.1)^2 + x_2 - 0.1 \\ g_2 &= (x_2 - 0.1)^2 + x_3 - 0.1 \\ g_3 &= (x_3 - 0.1)^2 + x_4 - 0.1 \\ g_4 &= (x_4 - 0.1)^2 + x_1 - 0.1 \end{aligned}$$

The system has two roots. The starting point used was (0.1,0.2,0.3,0.4) and r=4. The local minima and maxima of $f(x) = \|g(x)\|^2$ obtained by the heuristics are given in Table 3. The critical point min1 was found by following the negative gradient field from (0.1,0.2,0.3,0.4). Then form_graph(min1) yielded the results shown in Table 3. It is easy to verify that both the local minima are the roots of the given set of equations.

References

[1] F. Archetti and F. Schoen, *"A Survey on the Global Optimization Problem: General Theory and Computational Approaches"*, Annals of Operational Research, 1984, pp 87–110.

[2] I. O. Bohavesky, M. E. Johnson and M. L. Stein, *"Generalized Simulated Annealing for Functional Optimization"*, Technometrics, vol. 28, 1986, pp 209–217.

[3] F. H. Branin, *"A widely convergent method for finding multiple solutions of simultaneous nonlinear equations"*, IBM Journal of Research and Development, 1972, pp 504–522.

[4] J. Canny, *The complexity of robot motion planning*, PhD Dissertation, 1987, MIT.

[5] H. Chiang, M. W. Hirsch, F. F. Wu, *"Stability Regions of nonlinear autonomous systems"*, IEEE Transactions On Automatic Control, Vol 33, No 1, 1988.

[6] I. Deiner, *"Trajectory Nets connecting all critical points of a smooth function"*, Mathematical Programming vol. 36, 1986, pp 340–352.

[7] V. Guillemin and A. Pollack, *Differential Topology*, Prentice Hall, Inc. New Jersy, 1974.

[8] S. Kirkpatrick et. al., *"Optimization by Simulated Annealing"*, Science, vol. 220, pp 671–680.

[9] A. P. Morgan, *Solving Polynomial systems using continuation for engineering and scientific problems*, Prentice Hall, Englewood Cliffs,New Jersy, 1987.

[10] J. Palis and Jr. W. de Melo, *Geometric Theory of dynamical systems: An introduction*, NewYork: Springer-Verlag, 1981.

[11] Shashikala H., N. K. Sancheti and S. S. Keerthi, *"An algorithm for connecting all the extrema of a function over a compact manifold: Theory and Applications"*, Technical Report 91-08, Department of CSA , IISc, Bangalore, 1991.

[12] G. Treccani, *"On the convergence of Branin's method: A Counter Example"*, in *Towards Global Optimisation*, edited by L. C. W. Dixon and G. F. Szego – North Holland/American elsevier.