

Globally Descending Method for Unconstrained Global Optimization

*Zhiyou Wu**, *Duan Li* and *Liansheng Zhang*[†]
Chinese University of Hong Kong

April 6, 2008

*Chongqing Normal University

†Shanghai University

Introduction

- Most optimization problems in the world are non-convex in nature. The existence of multiple local optima makes global optimization a great challenge.
- Various approaches in solving general continuous global optimization problems can be divided into two classes, stochastic/heuristic and deterministic.
- In contrast to the stochastic/heuristic methods, deterministic methods are more reliable, however, with a price of expensive computational cost.

- Global solution schemes have been developed for certain problems with special structures, for examples, concave minimization, monotone optimization, and polynomial optimization.
- Branch and bound approach has been the most widely used deterministic approach for a general problem setting. The idea of the branch and bound approach is to divide the feasible region into partitions and discard some non-promising partitions by bounding the objective function in these partitions using estimated lower bounds.

Two-phase scheme

Phase 0: (Initialization) Set an initial point x_0 and $k := 0$.

Phase 1: (Local Search) Perform local search, starting from x_k to find out a local minimizer x_k^* of f .

Phase 2: (Global Search) Construct an auxiliary function such that its minima are of objective values lower than any local minimum of the original problem previously found and select one of them as x_{k+1} . Let $k := k + 1$ and return to Phase 1.

When failing in global search, the incumbent local optimal solution is taken as an approximate global optimal solution to the problem.

Tunneling method

- The concept of the tunneling algorithms was presented by Levy at the Seventh Biennial Conference on Numerical Analysis at Dundee, Scotland, in 1977 and the paper was published in 1985 by Levy and Montalvo, *SIAM Journal on Scientific and Statistical Computing*, Vol. 6, pp. 15-29.
- The idea of the method for finding a better minimizer x_{k+1}^* of f from the current minimizer x_k^* of f is to find roots of $f(x) - f(x_k^*)$, other than x_k^* . A local search is then carried out, starting from one of the identified root, to find out a better minimizer x_{k+1}^* .

Filled Function

- The concept of the filled functions was presented by Ge at the Tenth Biennial Conference on Numerical Analysis at Dundee, Scotland, in 1983, and his paper was published in 1990 in *Mathematical Programming*, Vol. 46, pp. 191-204.
- The filled function method constructs an auxiliary function (a filled function) at the current minimum point x_k^* such that x_k^* becomes a strict maximum of the filled function and the current basin B_k becomes a part of a hill of the filled function.

- A function $p(x, x_1^*)$ is said to be a filled function of $f(x)$ at the local minimizer x_1^* if it satisfies the following:
 1. x_1^* is a maximizer of $p(x, x_1^*)$ and the whole basin B_1^* of $f(x)$ at x_1^* becomes a part of a hill of $p(x, x_1^*)$;
 2. $p(x, x_1^*)$ has no minimizers or saddle points in any basin of $f(x)$ higher than B_1^* ;
 3. if $f(x)$ has a basin B_2^* at x_2^* that is lower than B_1^* , then there is a point $x' \in B_2^*$ that minimizes $p(x, x_1^*)$ on the line through x_1^* and some x'' which is in some neighborhoods of x_2^* .

Problem Formulation

- We consider the unconstrained programming problem:

$$(P) \quad \min_{x \in R^n} f(x).$$

- **Assumption 1** $f(x)$ is continuously differentiable on R^n .
- **Assumption 2** There exist $x_0^0 \in R^n$, $f_0 > 0$ and a box set $\Omega \subset R^n$ such that $x_0^0 \in \Omega$ and $f(x) \geq f(x_0^0) + f_0$ for any $x \in R^n \setminus \text{int}\Omega$, where

$$\Omega = \{x = (x_1, \dots, x_n) \mid c_i \leq x_i \leq d_i, i = 1, \dots, n\}, \quad (1)$$

- Note that $f(x)$ satisfies Assumption 2, if $f(x)$ satisfies the coercive condition, i.e., $f(x) \rightarrow +\infty$ as $\|x\| \rightarrow +\infty$.

- When concerning the global optimal solution(s) of (P) under Assumption 2, the original problem (P) is equivalent to the following problem (P_Ω) ,

$$\begin{aligned} (P_\Omega) \quad & \min && f(x) \\ & \text{s.t.} && x \in \Omega, \end{aligned}$$

i.e., \bar{x}^* is a global minimizer of problem (P) if and only if \bar{x}^* is a global minimizer of problem (P_Ω) when Ω is sufficiently large.

- In the remaining of the paper, we let x^* be the incumbent of an iterative algorithm which satisfies that $f(x^*) \leq f(x_0^0)$, where x_0^0 satisfies Assumption 2.

Globally descending function

- For a given adjustable parameter $r > 0$, we define the following two functions,

$$g_r(t) = \begin{cases} 1, & t \geq 0 \\ -\frac{2}{r^3}t^3 - \frac{3}{r^2}t^2 + 1, & -r < t \leq 0 \\ 0, & t < -r \end{cases}, \quad (2)$$

$$f_r(t) = \begin{cases} t + r & t \leq -r \\ \frac{r-2}{r^3}t^3 + \frac{r-3}{r^2}t^2 + 1, & -r < t \leq 0 \\ 1 & t > 0 \end{cases} \quad (3)$$

- Both $g_r(t)$ and $f_r(t)$ are differentiable on R ,

$$g'_r(t) = \begin{cases} 0, & t \geq 0 \\ -\frac{6}{r^3}t^2 - \frac{6}{r^2}t, & -r < t \leq 0 \\ 0, & t < -r \end{cases} . \quad (4)$$

$$f'_r(t) = \begin{cases} 1 & t \leq -r \\ \frac{3r-6}{r^3}t^2 + \frac{2r-6}{r^2}t, & -r < t \leq 0 \\ 0 & t > 0 \end{cases} . \quad (5)$$

- Without loss of generality, we take point $x_0 \in R^n \setminus \Omega$ as

$$x_0 = (c_1 - 1, \dots, c_n - 1). \quad (6)$$

- Obviously, $\|x - x_0\| \geq 1$ for any $x \in \Omega$ and

$$d = (d_1, \dots, d_n) \quad (7)$$

is the vertex of Ω farthest from x_0 .

- Let

$$\varphi_{r,x^*}(x) = \exp\left(\frac{1}{\|x - x_0\|}\right) g_r\left(f(x) - f(x^*)\right) + f_r\left(f(x) - f(x^*)\right). \quad (8)$$

Note that x^* is not a stationary point of function $\varphi_{r,x^*}(x)$.

- Define

$$D = \frac{1}{\|d - x_0\|^2}. \quad (9)$$

- **Theorem 1** *If Assumption 1 and Assumption 2 hold true, then $\varphi_{r,x^*}(x)$ satisfies the following conditions for any $r > 0$.*

i) *If $x \in \Omega$ satisfies $f(x) \geq f(x^*)$, then x is not a stationary point of $\varphi_{r,x^*}(x)$ and satisfies*

$$\|\nabla^T \varphi_{r,x^*}(x)\| \geq D.$$

ii) *Any local minimizer \bar{x} of $\varphi_{r,x^*}(x)$ on Ω satisfies one of the following two conditions:*

1° $f(\bar{x}) < f(x^*)$,

2° $\bar{x} = d$.

- According to Theorem 1, for any local minimizer or a stationary point \bar{x} of $\varphi_{r,x^*}(x)$ on Ω which is not equal to d , \bar{x} must satisfy $f(\bar{x}) < f(x^*)$. A better lower local minimizer of the original problem (P) can be then obtained by any local searching scheme starting from \bar{x} .

- Let Y be the set of local minima of problem (P_Ω) , and let

$$F = \{f(x) \mid x \in Y\}, \quad (10)$$

$$L(x^*) = \{\bar{x} \in Y \mid f(\bar{x}) < f(x^*)\}, \quad (11)$$

$$\beta_0(x^*) = \min_{x \in L} \left(f(x^*) - f(x) \right). \quad (12)$$

- **Theorem 2** *Suppose that*

- Assumptions 1 and 2 hold and F is a finite set; and*
- $L(x^*) \neq \emptyset$, i.e., x^* is not a global minimizer of problem (P) .*

Then any $\bar{x} \in L(x^)$ will be a local minimizer and a stationary point of $\varphi_{r,x^*}(x)$ on Ω .*

Furthermore, $\varphi_{r,x^}(\bar{x}) < \varphi_{r,x^*}(x^*)$ and $\varphi_{r,x^*}(\bar{x}) < \varphi_{r,x^*}(x)$ for any $x \in \partial\Omega$ when the parameter r is chosen such that*

$$0 < r \leq \frac{\beta_0(x^*)}{2}. \quad (13)$$

- **Theorem 3** *For any x_1 and $x_2 \in \Omega$, if $f(x_1) \geq f(x^*)$, $f(x_2) \geq f(x^*)$, then $\|x_2 - x_0\| > \|x_1 - x_0\|$ if and only if $\varphi_{r,x^*}(x_2) < \varphi_{r,x^*}(x_1)$.*

- we summarize now the properties of function $\varphi_{r,x^*}(x)$.
 1. Any point $x \in \Omega$ (except d) satisfying $f(x) \geq f(x^*)$ is neither a local minimizer nor a stationary point of $\varphi_{r,x^*}(x)$ on Ω (from Theorem 1).
 2. Any local minimizer, \bar{x} , of original problem (P) satisfying $f(\bar{x}) < f(x^*)$ is also a local minimizer and a stationary point of $\varphi_{r,x^*}(x)$ on Ω when parameter r is small enough (from Theorem 2).
 3. In the local search of minimizing $\varphi_{r,x^*}(x)$ on Ω , either we find a point \bar{x} satisfying $f(\bar{x}) < f(x^*)$, or we reach vertex d as the farther the point x from x_0 , the smaller of $\varphi_{r,x^*}(x)$ (from Theorem 3).

- We classify function $\varphi_{r,x^*}(x)$ as a *globally descending function* as it enables us to proceed to a better local solution of the original problem by finding local minimizer or stationary point of such a function.
- **Definition 1** A function $p_{x^*}(x)$ is said to be a *globally descending function of problem (P_Ω) at x^** if $p_{x^*}(x)$ satisfies the following conditions:
 - i) For any stationary point \bar{x} of function $p_{x^*}(x)$ on Ω , $f(\bar{x}) < f(x^*)$ holds;
 - ii) For any local minimizer \bar{x} of $p_{x^*}(x)$ on Ω that is not a vertex of Ω , $f(\bar{x}) < f(x^*)$ holds;
 - iii) If $L(x^*) \neq \emptyset$, i.e., x^* is not a global minimizer of problem (P) , then any $\bar{x} \in L(x^*)$ is a local minimizer and a stationary point of $p_{x^*}(x)$ on Ω .

Furthermore,

$$\begin{aligned} p_{x^*}(\bar{x}) &< p_{x^*}(x^*) \\ p_{x^*}(\bar{x}) &< p_{x^*}(x), \quad \text{for any } x \in \partial\Omega. \end{aligned}$$

iv) *One of the following conditions holds:*

1° *For any x_1 and $x_2 \in \Omega$, if $f(x_1) \geq f(x^*)$ and $f(x_2) \geq f(x^*)$, then $\|x_2 - x_0\| > \|x_1 - x_0\|$ if and only if $p(x_2) < p(x_1)$;*

2° *for any $x_1, x_2 \in \Omega$, if $f(x_1) \geq f(x^*)$ and $f(x_2) \geq f(x^*)$, then $\|x_2 - x_0\| < \|x_1 - x_0\|$ if and only if $p(x_2) < p(x_1)$.*

- *Function $\varphi_{r,x^*}(x)$ satisfies the conditions of i), ii), iii) and 1° of iv). Thus, $\varphi_{r,x^*}(x)$ is a globally descending function at x^* .*

- Let

$$\psi_{r,x^*}(x) = \exp\left(-\frac{1}{\|x - x_0\|}\right) g_r\left(f(x) - f(x^*)\right) + f_r\left(f(x) - f(x^*)\right). \quad (14)$$

- Function $\psi_{r,x^*}(x)$ satisfies the conditions i), ii), iii) and 2° of iv) in Definition 1, thus is another *globally descending function*.

- Function $\psi_{r,x^*}(x)$ also possesses the following properties.

(1) Any $x \in \Omega$ satisfying $f(x) \geq f(x^*)$ satisfies

$$\|\nabla^T \psi_{r,x^*}(x)\| \geq \frac{D}{e}. \quad (15)$$

(2) Any local minimizer \bar{x} of $\psi_{r,x^*}(x)$ on Ω either satisfies $f(\bar{x}) < f(x^*)$ or satisfies $\bar{x} = c$.

- Functions $(\varphi_{r,x^*}(x)$ and $\psi_{r,x^*}(x))$ essentially form a *globally descending function pair* of problem (P_Ω) at x^* . For a *globally descending function pair*, we have some further interesting properties, which are useful in our proposed solution algorithm.
- **Proposition 1** For any direction $d \in R^n$ and any $x \in \Omega$ satisfying $f(x) \geq f(x^*)$, $d^T \nabla \varphi_{r,x^*}(x) > 0$ if and only if $d^T \nabla \psi_{r,x^*}(x) < 0$.
- **Proposition 2** Let u be a unit vector in R^n , i.e., $\|u\| = 1$. Let $d(x) = \frac{x-x_0}{\|x-x_0\|}$. For any $x \in \Omega$ satisfying $f(x) \geq f(x^*)$, we have that $\nabla^T \varphi_{r,x^*}(x)(d(x) + u) = 0$ if and only if $d(x) + u = 0$.

Globally Descending Algorithm (GDA):

Step 0. Select i) a point $x_0 \in R^n \setminus \Omega$ (usually taken as $(c_1 - 1, \dots, c_n - 1)$) and ii) an initial point $x_0^0 \in \Omega$ satisfying Assumption 2. Take a positive integer number K_0 and unit directions u_i , i.e., $\|u_i\| = 1$, $i = 1, \dots, K_0$. Choose a positive number r_0 as the initial value of parameter r . Let $r := r_0$ and $k := 0$.

Step 1. Let x_k^* be the local minimizer of problem (P_Ω) by any local search method starting from the point x_k^0 .

Step 2. Let

$$\begin{aligned}\varphi_{r,x_k^*}(x) &= \exp\left(\frac{1}{\|x - x_0\|}\right)g_r\left(f(x) - f(x_k^*)\right) + f_r\left(f(x) - f(x_k^*)\right), \\ \psi_{r,x_k^*}(x) &= \exp\left(-\frac{1}{\|x - x_0\|}\right)g_r\left(f(x) - f(x_k^*)\right) + f_r\left(f(x) - f(x_k^*)\right).\end{aligned}$$

Consider the following two problems:

$$\min_{x \in \Omega} \varphi_{r, x_k^*}(x) \quad (16)$$

and

$$\min_{x \in \Omega} \psi_{r, x_k^*}(x). \quad (17)$$

Let $d_{0,k} = \frac{x_k^* - x_0}{\|x_k^* - x_0\|}$ and $d_{i,k} = d_{0,k} + u_i$, $i = 1, \dots, K_0$.

Without loss of generality, we assume that, for any $i = 1, \dots, K_0$, $d_{i,k} \neq 0$ (i.e., $d_{0,k} \neq -u_i$), as otherwise we can change this specific u_i . Let $i := 0$.

Step 3. If $i > K_0$, go to *Step 6*, else go to *Step 4* if $d_{i,k}^T \nabla \varphi_{r, x_k^*}(x_k^*) < 0$, or go to *Step 5* otherwise.

Step 4. Starting from point x_k^* , minimize $\varphi_{r, x_k^*}(x)$ along direction $d_{i,k}$ by using any local searching method. If, during

this local search process, $f(y_k^*) < f(x_k^*)$ holds, let $x_{k+1}^0 := y_k^*$, $k := k + 1$, and go to *Step 1*; Otherwise, continue the minimization of $\varphi_{r,x_k^*}(x)$ on Ω . If the local minimizer \bar{x}_{r,x_k^*} is equal to d , let $i := i + 1$ and go to *Step 3*; Otherwise, let $x_{k+1}^0 := \bar{x}_{r,x_k^*}$, $k := k + 1$, and go to *Step 1*.

Step 5. Starting from point x_k^* , minimize $\psi_{r,x_k^*}(x)$ along direction $d_{i,k}$ by any local searching method. If, during the local search process, $f(y_k^*) < f(x_k^*)$ holds, let $x_{k+1}^0 := y_k^*$, $k := k + 1$, and go to *step 1*; Otherwise, continue the minimization of $\psi_{r,x_k^*}(x)$ on Ω . If the local minimizer \bar{x}_{r,x_k^*} is equal to c , let $i := i + 1$ and go to *Step 3*; Otherwise, let $x_{k+1}^0 := \bar{x}_{r,x_k^*}$, $k := k + 1$, and go to *Step 1*.

Step 6. If $r > \mu$, decrease r , for example, by setting $r := \frac{r}{10}$, and go to *Step 2*; Otherwise, stop and x_k^* is an (approximate) global minimizer of problem (P_Ω) .

Quasi Globally Descending Function

- Starting from x_k^* , a local (non-global) minimizer of problem (P) , solving problem (16) or problem (17) may end up with the local minimizer \bar{x}_{r,x_k^*} at d or c , respectively, which is not a better point which we are seeking.
- In such situations, we need to increase the number of search directions, u_i , $i = 1, \dots, K_0$, in order to increase the probability that we find out a better point.
- We develop next a way to overcome this problem by introducing a quasi globally descending function.

- From Assumption 3 that $f(x) \geq f(x_0^0) + f_0$ for any $x \in \partial\Omega$, and the current local minimizer or stationary point x^* satisfies $f(x^*) \leq f(x_0^0)$. Therefore, $f(x) \geq f(x^*) + f_0$ for any $x \in \partial\Omega$.
- Let

$$h_r(t) = \begin{cases} 2, & t \geq r \\ -\frac{4-r}{r^3}t^3 + \frac{6-2r}{r^2}t^2 + t, & 0 < t < r \\ t, & t \leq 0 \end{cases} \quad (18)$$

- $h_r(t)$ is continuously differentiable on \mathbb{R} with

$$h'_r(t) = \begin{cases} 0, & t \geq r \\ -\frac{12-3r}{r^3}t^2 + \frac{12-4r}{r^2}t + 1, & 0 < t < r \\ 1, & t \leq 0 \end{cases} \quad (19)$$

- Choose a given point $x_0 \in R^n \setminus \Omega$ such that $\|x - x_0\| \geq 1$ for any $x \in \Omega$. Let

$$H_{q,r,x^*}(x) = q \left(\exp \left(\frac{1}{\|x - x_0\|} \right) g_r \left(f(x) - f(x^*) \right) + h_r \left(f(x) - f(x^*) \right) \right),$$

where q is used to speed up the the descending rate.

- **Theorem 4** *Suppose that both Assumptions 1 and 2 hold. Then, we have the following for any $r > 0$.*
 - Any $x \in \Omega$ satisfying $f(x) \geq f(x^*) + r$ is not a stationary point of $H_{q,r,x^*}(x)$;*
 - Any x satisfying $\nabla f(x) = 0$ and $0 \leq f(x) - f(x^*) < r$ is not a stationary point of $H_{q,r,x^*}(x)$, in special, $\nabla H_{q,r,x^*}(x^*) \neq 0$.*

• **Theorem 5** *Suppose that F is finite and*

i) Both Assumptions 1 and 2 hold;

ii) x^ is not a global minimizer of problem (P) , i.e., $L(x^*) \neq \emptyset$.*

Then, when $0 < r \leq \frac{\beta_0(x^)}{2}$, any $\bar{x} \in L(x^*)$ is a local minimizer and a stationary point of $H_{q,r,x^*}(x)$ on Ω .*

Furthermore, $H_{q,r,x^}(\bar{x}) < H_{q,r,x^*}(x^*)$ and $H_{q,r,x^*}(\bar{x}) < H_{q,r,x^*}(x)$ for any $x \in \partial\Omega$.*

• **Theorem 6** *Suppose that both Assumptions 1 and 2 hold.*

Let \bar{x} be a local minimizer resulted from minimization of $H_{q,r,x^}(x)$ on Ω starting from x^* . Then $\bar{x} \in \text{int}\Omega$ for any $0 < r \leq f_0$, where f_0 is given in Assumption 2.*

- Although *quasi globally descending function* $H_{q,r,x^*}(x)$ is not a *globally descending function*, it enjoys some nice properties which a *globally descending function* does not.
- Any local minimizer of the *quasi globally descending function* on Ω is in the interior of Ω , i.e., the local search process will not reach the boundary of Ω ;
- Our numerical results reveal much better outcomes of quasi globally descending function $H_{q,r,x^*}(x)$ compared to globally descending functions.
- The local minimizer of $H_{q,r,x^*}(x)$ on Ω can be efficiently obtained by using the optimization subroutine of the optimization Toolbox in Matlab 6.1.

Algorithm *QGDA*:

Step 0. Choose i) a small positive number $\mu > 0$ as the tolerance value of parameter r for terminating the minimization process of problem (P) ; ii) a large positive number $M > 0$ as the tolerance value of q , iii) a point $x_0 \in R^n \setminus \Omega$ such that $\|x - x_0\| \geq 1$ for any $x \in \Omega$ and iv) an initial point $x_0^0 \in \Omega$ such that Assumption 2) is satisfied. In the following examples, we take $\mu = 10^{-10}$ and $M = 10^{10}$. Set q_0 and r_0 as the initial values of parameters q and r , respectively. (In the following examples, we take $q_0 = 100$ and $r_0 = 1$). Let $k := 0$.

Step 1. Let x_k^* be a local minimizer of problem (P_Ω) by implementing a local search procedure starting from the initial point x_k^0 .

Step 2. Let

$$H_{q,r,x_k^*}(x) = q \left[\exp \left(\frac{1}{\|x - x_0\|} \right) g_r \left(f(x) - f(x_k^*) \right) + h_r \left(f(x) - f(x_k^*) \right) \right].$$

Solve the problem:

$$\min_{\Omega} H_{q,r,x_k^*}(x) \quad (20)$$

by a local search method starting from x_k^* . Let \bar{x}_{q,r,x_k^*} be the local minimizer generated from the solution process. If $f(\bar{x}_{q,r,x_k^*}) < f(x_k^*)$, set $x_{k+1}^0 := \bar{x}_{q,r,x_k^*}$, $k := k+1$ and goto *Step 1*; Otherwise ($f(\bar{x}_{q,r,x_k^*}) \geq f(x_k^*)$), goto *Step 3*.

Step 3 If $q < M$, increase q , for example in our numerical calculation, by setting $q := 10q$, and goto *Step 2*; Otherwise goto *Step 4*.

Step 4 If $r > \mu$, set $q = q_0$, decrease r , for example in our numerical calculation, by setting $r := \frac{r}{10}$, and go to *Step 2*; Otherwise, stop and x_k^* is an approximate global minimizer of problem (P) .

We can obtain an approximate global minimizer of problem (P) in finite steps by using the Algorithm *QGDA* and local search methods.

Example 1 Two-dimensional Shubert II Function ($n = 2$)

$$\begin{aligned} \min \quad f_S(x) &= \left(\sum_{i=1}^5 i \cos[(i+1)x_1 + i] \right) \left(\sum_{i=1}^5 i \cos[(i+1)x_2 + i] \right) \\ &\quad + \frac{1}{2} \left[(x_1 + 1.42513)^2 + (x_2 + 0.80032)^2 \right] \quad (21) \\ \text{s.t.} \quad & -10 \leq x_i \leq 10, \quad i = 1, 2. \end{aligned}$$

We take $x_0 = (11, 11)$ and the initial point $x_0^0 = (1, 1)$.

Table 1: Results for Shubert II function by QGDA

k	0		1	
x_k^0	(1, 1)		(-1.4251, -0.8003)	
x_k^*	(1.3119, 1.7980)		(-1.4251, -0.8003)	
$f(x_k^*)$	-0.8464		-186.7309	
q, r		$10^4, 1$		for any $q \leq 10^{10}$ and $r \geq 10^{-5}$
\bar{x}_{q,r,x_k^*}		(-1.4251, -0.8003)		
$f(\bar{x}_{q,r,x_k^*})$		-186.7309		$f(\bar{x}_{q,r,x_k^*}) \geq -186.7309$

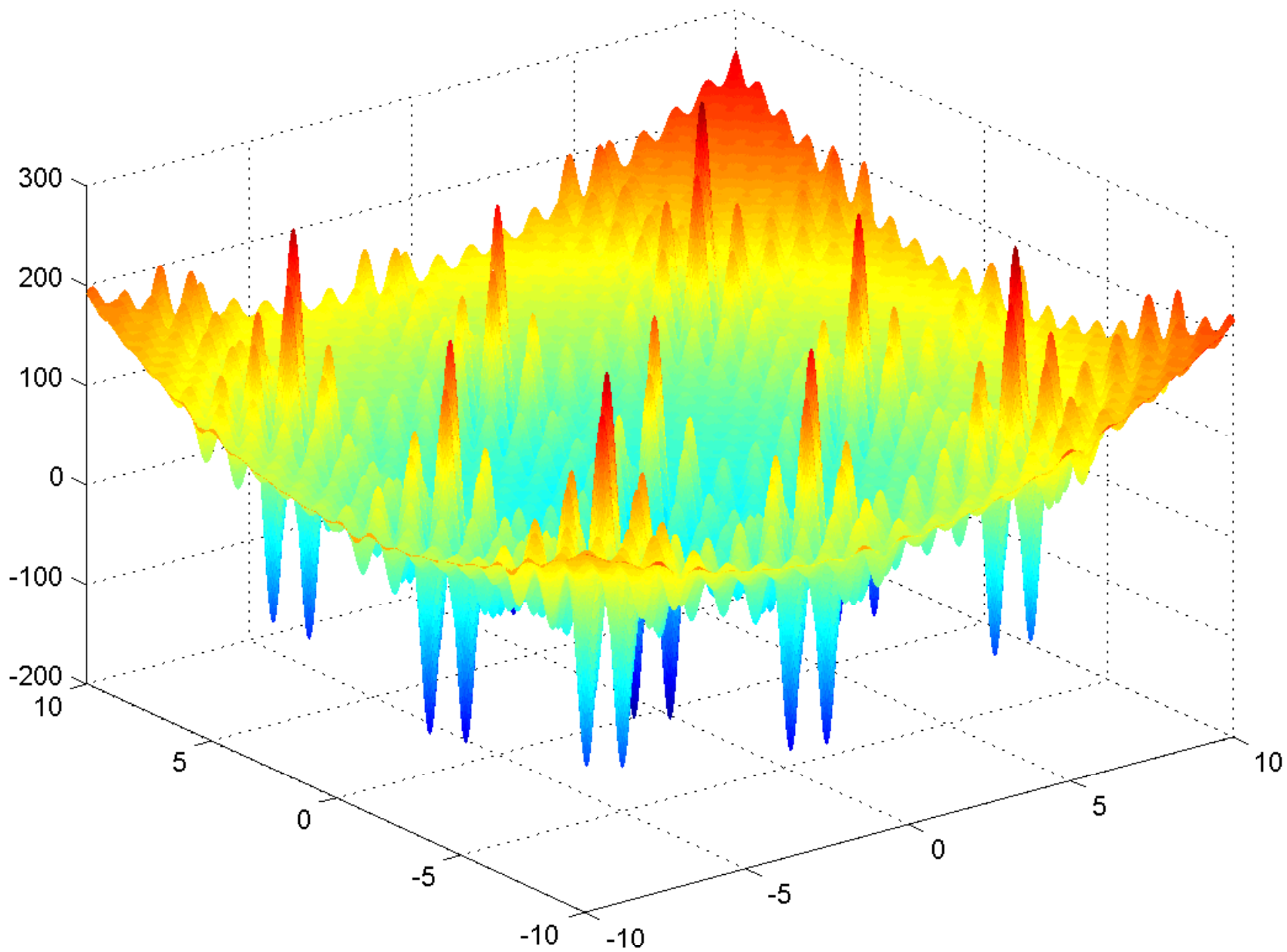


Figure 1: Behavior of **Two-Dimensional Shubert Function**