A finite-step convergent method for unconstrained optimization

Yunqing Huang[a,**], Kai Jiang[a,**]

[a] *School of Mathematics and Computational Science,*
Hunan Key Laboratory for Computation and Simulation in Science and Engineering, Xiangtan University, P.R. China, 411105

Abstract

In our previous work [Adv. Appl. Math. Mech., 2017, 9: 307-323], we proposed a novel optimization algorithm, the hill-climbing method with a stick (HiCS), to address the unconstrained optimization. Numerical results have been demonstrated many satisfactory properties. However, there exist two unsolved issues: convergent analysis and application to high dimensional problems. In this paper, we will give a rigorous theory to guarantee finite-step convergence by introducing a new definition of the suspected extreme point. Meanwhile, an economic sampling strategy based on the regular simplex is developed to treat high dimensional optimization. Finally, the efficiency of the improved HiCS method is demonstrated by several high dimensional examples.

Keywords: Hill-climbing method with a stick, Finite-step convergence, Suspected extreme point, Simplex, High-dimensional unconstrained optimization

1. Introduction

The optimization approaches can be divided into broadly two classes, directional search and model-based [1, 2, 3]. Directional search algorithms first determine the search direction and then the step length along the search direction. Model-based approaches construct and utilize a related simple model to approximate the original problem in a trust region to guide the search process. Inspired by the behavior of the blind for climbing hill, we proposed a new approach, the hill-climbing method with a stick (HiCS), to treat unconstrained optimization problems in our previous work [4]. The main idea of the HiCS, at each search step, is comparing function values on a surface surrounding the current iterator. It requires the comparison of function values, and does not need the search direction or construct a surrogate model in a trust-region. Numerical results have been demonstrated that the HiCS has many satisfactory properties, including being easy to implement, only a unique parameter to be modulated, and having capacity for finding the local and global maximum. However, there are two unsolved problems in the previous work which include rigorous theoretical explanation and the treatment of high-dimensional optimization. In this paper, we will give the convergence analysis and related properties of this algorithm by introducing a new concept. Meanwhile, a new strategy will be proposed to sample the search surface to address high dimension optimization problems.

*huangyq@xtu.edu.cn.
**kaijiang@xtu.edu.cn.

In the following, we will briefly introduce the HiCS algorithm and prove its finite-step convergence in Sec. 2. The algorithm implementation is presented in Sec. 3. In particular, the new sampling strategy using the regular simplex is also given in this section. The numerical experiments including high dimensional optimization problems are showcased in Sec. 4. Finally the conclusion and discussions are given in Sec. 5.

2. Algorithm description and convergence analysis

Before going further, a short introduction of the HiCS method is necessary. We consider an unconstrained optimization problem

$$\min_{x\in\Omega\subset\mathbb{R}^d} f(x), \tag{1}$$

where the objective function $f(x):\mathbb{R}^d\to\mathbb{R}$. Let ρ be the search radius, $O(x_k,\rho)=\{x:\|x-x_k\|=\rho\}$ be the search surface at the k-th iteration with radius ρ. $\|\cdot\|$ is the common norm in $\mathbb{R}^d$ space. $U(x_k,\rho)$ is the neighbourhood of x_k with radius of ρ. To illustrate the algorithm more accurately, an useful concept of the suspected extreme point is introduced.

Definition 1. *For a given objective function $f(x)$ and a positive constant $\rho>0$, $\tilde{x}$ is a suspected extreme point if $f(\tilde{x})<f(x)$ or $f(\tilde{x})>f(x)$, for each $x\in O(\tilde{x},\rho)$. If $f(\tilde{x})<f(x)$ for all $x\in O(\tilde{x},\rho)$, $\tilde{x}$ is the suspected minimum point (SMP).*

Certainly, $\tilde{x}$ is a SMP if $\tilde{x}$ is a minimizer in the neighborhood of $U(\tilde{x},\rho)$. The opposite is not always true. With these notations, the HiCS algorithm can be presented as the Algorithm 1.

Algorithm 1 Hill-Climbing method with a stick (HiCS)

1: **Initialization:** Choose x_0 and ρ.
2: **For** $k=0,1,2,\ldots$
3: Find $\bar{x}=\operatorname{argmin}_{y\in O(x_k,\,\rho)} f(y)$.
4: If $f(\bar{x})<f(x_k)$, then set $x_{k+1}=\bar{x}$.
5: Otherwise, declare that a SMP is found, and end the iteration.

It is evident that the approximation error of the HiCS algorithm can be measured by the distance between the SMP and a minimum. When the HiCS converges, the approximation error is smaller than the search radius ρ. From our experience, the HiCS approach usually terminates in finite steps. It is a satisfactory property. In what follows, we will prove the finite-step convergence with some mild conditions.

Theorem 1 (Finite-step convergence)**.** *Suppose that objective function $f(x)$ is continuous and the search domain Ω is a compact set. If there are not two SMPs x_* and x^* satisfying $\|x_*-x^*\|=\rho$ and $f(x^*)=f(x_*)=\alpha$. Then Algorithm 1 converges in finite steps.*

Proof. Assume that the HiCS method generates an infinite pair sequence $\{x_n,f(x_n)\}_{x=0}^{\infty}$. From these assumptions, it is obvious $f(x)$ is bounded. The decreasing sequence $\{f(x_n)\}_{n=0}^{\infty}$ converges, and the bounded $\{x_n\}_{n=0}^{\infty}$ has a convergent subsequence $\{x_{n_k}\}_{k=0}^{\infty}$. Assume that $f(x_n)\to\alpha$ and $x_{n_k}\to x^*$. Obviously x^* is a SMP.

In accordance with the subsequence $\{x_{n_k}\}_{k=0}^{\infty}$, we can always choose an another bounded subsequence $\{x_{n_k-1}\}_{k=0}^{\infty}\subset\{x_n\}$ satisfying $\|x_{n_k-1}-x_{n_k}\|=\rho$. Due to the boundedness of iteration sequence, $\{x_{n_k-1}\}_{k=0}^{\infty}$ has a convergent subsequence $\{x_{n_m}\}_{m=0}^{\infty}$. Let $x_{n_m} \to x_*$ when $m \to \infty$. x_* is also a SMP. From the $\{x_{n_m}\}$, we can find a subsequence $\{x_{n_m+1}\} \subset \{x_{n_k}\}$ which satisfies $\|x_{n_m} - x_{n_m+1}\| = \rho$, and $x_{n_m+1} \to x^*$ $(m \to \infty)$. Obviously, $\|x^* - x_*\| = \rho$, and $f(x^*) = f(x_*) = \alpha$ which clearly contradicts the assumption. □

3. Algorithm implementation

As mentioned above, the HiCS algorithm can converge in finite steps with mild assumptions for a given search radius ρ. In numerical implementation, the search surface $O(x_k, \rho)$ in each iteration shall be discretized. Without a priori information of the objective function, a reasonable discretization principle for $O(x_k, \rho)$ should include symmetric and uniform distribution, and as few discretization points as possible. Our previous results have demonstrated that the uniformly distributed discretization method based on the spherical coordinate can be used to find the SMP [4]. The generated discretization points are as large as $2m^{d-1}$ in each iteration, m is the number of refinement, d is the dimensions of optimization problems. It restricts the application to high-dimensional problems. To overcome this limitation, it needs to develop a new strategy to discretize $O(x_k, \rho)$ with less discretization points but still satisfying these properties. A reasonable requirement is that the discretization points should be linear or quasi-linear growth as the problem's dimension increases. In this work, we will use the regular simplex and its rotations to discretize the search surface $O(x_k, \rho)$. The computational complexity grows linearly as the dimension of optimization problems increases.

The d-dimension regular simplex is a congruent polytope of $\mathbb{R}^d$ with a set of points $\{a_1, \ldots, a_d, a_{d+1}\}$, and all pairwise distances 1. Its Cartesian coordinates can be obtained from the following two properties:

1. For a regular simplex, the distances of its vertices $\{a_1, \ldots, a_d, a_{d+1}\}$ to its center are equal.
2. The angle subtended by any two vertices of the d-dimension simplex through its center is $\arccos(-1/d)$.

In particular, the above two properties can be implemented through the Algorithm 2.

Algorithm 2 Generate d-D regular simplex coordinates

Given a $d \times (d+1)$-order zero matrix $x(1:d, 1:d+1)$
Let $x(:,1) = (1, 0, \ldots, 0)$
for $i = 2 : 1 : d$ **do**
  $x(i,i) = \sqrt{1 - \sum_{k=1}^{i-1} [x(k,i)]^2}$
  for $j = i+1 : 1 : d+1$ **do**
    $x(i,j) = -\frac{1}{x(i,i)}\Big[\frac{1}{d} + x(1:i-1, i)^T \cdot x(1:i-1, j)\Big]$
  end for
end for
Output the column vectors, and let $a_j = x(:,j)$, $j = 1, 2, \ldots, d+1$

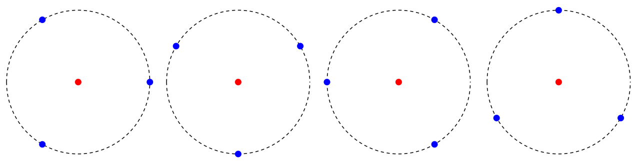
If the HiCS method does not find a better state at the initial given regular simplex, more points can be added to discretize $O(x_k, \rho)$. Here we rotate the regular simplex to obtain more discretization points. For a given rotation angle $\theta = (\theta_1, \theta_2, \ldots, \theta_d)$, the rotation matrix $\boldsymbol{Q}$ is given as

$$
\begin{aligned}
\boldsymbol{Q} = \prod_{i=2}^{d-1} &
\begin{array}{c}
\begin{matrix} & & & & i & & & & \end{matrix} \\
\begin{pmatrix}
1 & & & & \vdots & & & & \\
 & \ddots & & & \vdots & & & & \\
 & & 1 & & \vdots & & & & \\
 & & & \cos\theta_i & 0 & -\sin\theta_i & & & \\
 & & & 0 & 1 & 0 & & & \\
 & & & \sin\theta_i & 0 & \cos\theta_i & & & \\
 & & & & & & 1 & & \\
 & & & & & & & \ddots & \\
 & & & & & & & & 1
\end{pmatrix}
\end{array} \\
& \begin{pmatrix}
\cos\theta_1 & -\sin\theta_1 & 0 & & \\
\sin\theta_1 & \cos\theta_1 & 0 & & \\
0 & 0 & 1 & & \\
 & & & \ddots & \\
 & & & & 1
\end{pmatrix}
\begin{pmatrix}
1 & & & & \\
 & \ddots & & & \\
 & & 1 & 0 & 0 \\
 & & 0 & \cos\theta_d & -\sin\theta_d \\
 & & 0 & \sin\theta_d & \cos\theta_d
\end{pmatrix}.
\end{aligned}
\tag{2}
$$

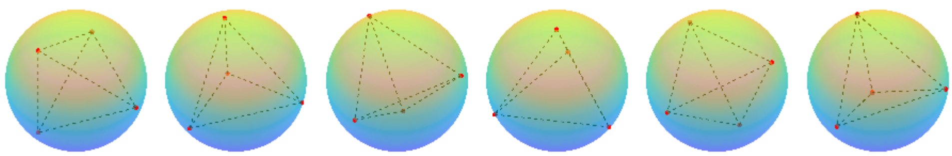
Then vertices of new simplex are

$$a_j = \boldsymbol{Q} a_j + x_k. \tag{3}$$

When without a priori knowledge of objective function, the uniform distribution of these regular



(a) 2D regular simplexes



(b) 3D regular simplexes

Figure 1: The first regular simplexes of sampling the search set $O(x, \rho)$.

simplexes is another principle to be obeyed. Standard schematic simplexes of 2D and 3D cases are given in Fig. 1. It should be noted that there are also other strategies to discretize $O(x_k, \rho)$.

For example, the new adding discretized points depend on the known information of objective functions.

To save computational amount, we choose a dynamic refinement strategy to discretize the search surface and compare function values in practice. Based on the dynamic refinement strategy, we propose the executable HiCS, see Algorithm 3. The computational amount is not larger than $m_{\max}(d+1)$ in each iteration which is linearly dependent on the dimension of optimization problems, $m_{\max}$ is the maximum number of rotation. Whence, it allows us to treat high-dimensional optimization problems.

Algorithm 3 Executable HiCS

Input x_0, ρ, and $m_{\max}$
for $k = 0, 1, 2, \cdots$ **do**
    Set $m = 0$
    if $m \le m_{\max}$ **then**
        Discretize $O(x_k, \rho)$ to obtain $O_h^m(x_k, \rho)$
        if $\exists x_j \in O_h^m(x_k, \rho)$, s.t. $f(x_j) < f(x_k)$ **then**
            Set $x_{k+1} = x_j$, and $m = m_{\max} + 1$
        else
            Set $m = m + 1$
        end if
    else
        Declare that find a SMP, end program
    end if
end for

It is evident that once the HiCS converges, the search space is shrunk to a ball with the radius ρ, and more significantly, the convergent ball contains a SMP. We will demonstrate this by several numerical experiments in Sec. 4. It is also noted that the convergent result provides a good initial value for other optimization approaches, including directional search and model-based algorithms.

We can also adjust the search radius ρ in HiCS method to improve the approximation precision as done in our previous work [4]. Algorithm 4 gives the process by adaptively changing ρ when Algorithm 3 fails to find $f(\bar{x}) < f(x_k)$, $\bar{x} \in O(x_k, \rho)$ for a fixed ρ. The approximation distance between convergent point and a SMP is improved when Algorithm 4 converges when $\eta < 1$. Certainly, the search surface can be expanded by setting control factor $\eta > 1$ if required. In fact, the Algorithm 4 can be restarted by fixed k iterations or by other criterions with different search radius.

Algorithm 4 Adaptive HiCS

Input $x_0, \rho, m_{\max}, \varepsilon$ and $\eta > 0$
if $\rho > \varepsilon$ **then**
  for $k = 0, 1, 2, \cdots$ **do**
    Set $m = 0$
    if $m \leq m_{\max}$ **then**
      Discretize $O(x_k, \rho)$ to obtain $O_h^m(x_k, \rho)$
      if $\exists x_j \in O_h^m(x_k, \rho)$, s.t. $f(x_j) < f(x_k)$ **then**
        Set $x_{k+1} = x_j$, and $m = m_{\max} + 1$
      else
        Set $m = m + 1$
      end if
    else
      Set $\rho = \eta\rho$
    end if
    Set $k = k + 1$
  end for
end if

4. Numerical results

In this section, we choose two kinds of high dimensional optimization functions, including the unimodal Gaussian function, multimodal problems, to demonstrate the performance of our proposed algorithm. These objective functions are all differentiable. However, it is emphasized that the HiCS method can be applied to non-differentiable problems [4]. In Algorithm 3, the discretized points of search set in each iteration are $m(n+1)$, n is the dimension of objective function. If not specified, the maximum number of rotation $m = 32$.

4.1. The unimodal problem: Gaussian function

The first objective function is the unimodal Gaussian problem

$$f(x) = -20 \exp\left(-\sum_{j=1}^{d} x_j^2\right), \tag{4}$$

which has one minimum 0 with $f(0) = -20$. The objective function is differentiable in $\mathbb{R}^d$, however, it quickly diffuses out towards zero out of the upside-down "bell".

We first investigate the convergent property of HiCS method for 10 dimensional Gaussian function using 30 experiments. In the set of experiments, the search radius ρ is fixed as 0.3, start points are all randomly generated in the space $[-1, 1]^{10}$. For each experiment, the HiCS method indeed converges and captures a neighbourhood of the peak 0 in finite iterations as Theorem 1 predicted. Fig. 2 gives the required iterations for convergence in the 30 numerical experiments. In these 30 runs, the average iterations of convergence is 20.5, while the maximum is 27, and the minimum is 9.

Then we decrease the search radius ρ to 0.1 to observe the behavior of the HiCS method in 30 numerical tests. The initial values are also randomly generated in the same region. The

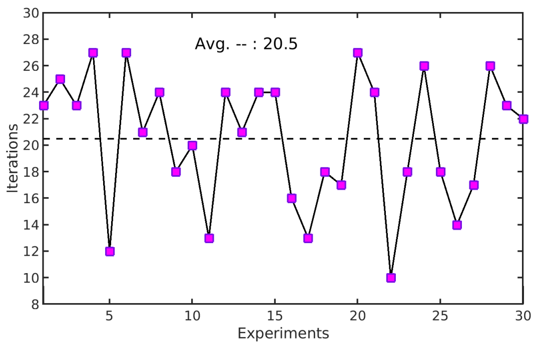


Figure 2: The required iteration steps of the HiCS algorithm for the Gaussian function (4) in 30 runs with randomly generated start points in the space $[-1, 1]^{10}$, and $\rho = 0.3$. The flat dashed line shows the average.

required iterations for convergence is given in Fig. 3. In these 30 runs, the average iterations of convergence is 77.2, while the maximum is 121, and the minimum is 54. From these results, the HiCS approach converges in finite iterations. Meanwhile, it is obvious that the value of ρ affects the number of iterations.

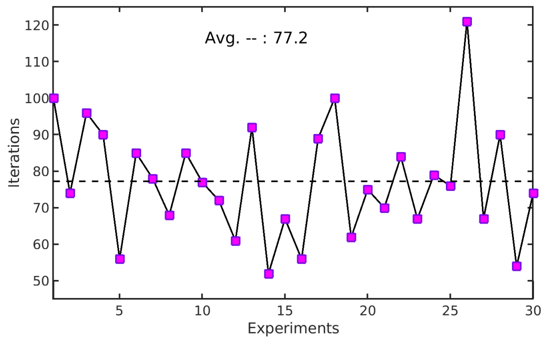


Figure 3: The required iteration steps of the HiCS algorithm for the Gaussian function (4) in 30 runs with randomly generated start points in the space $[-1, 1]^{10}$, and $\rho = 0.1$. The flat dashed line shows the average.

In the following, we apply the adaptive HiCS algorithm to 1000 dimensional Gaussian function. The initial value is randomly generated in domain $[-1000, 1000]^{1000}$, the initial search radius $\rho_0 = 2.0$, and control factor $\eta = (\sqrt{5} - 1)/2$. Fig. 4 presents the iteration process. The left image in Fig. 4 gives the difference between $f(x_k)$ and $f(0) = -20$. The right one in Fig. 4 plots the changes of search radius ρ and ℓ^2-distance between the iterator and the global minimizer $x^* = 0$, where $\|x\|_{\ell^2} = \left((\sum_{i=1}^{d} x_i^2)/d\right)^{1/2}$. From these results, it can be found that the HiCS is convergent for each ρ. Based on the iteration, the adaptive HiCS method can approximate the global minimum by decreasing the search radius ρ. Meanwhile, during the iteration, the global minimizer is always in the search neighbourhood.

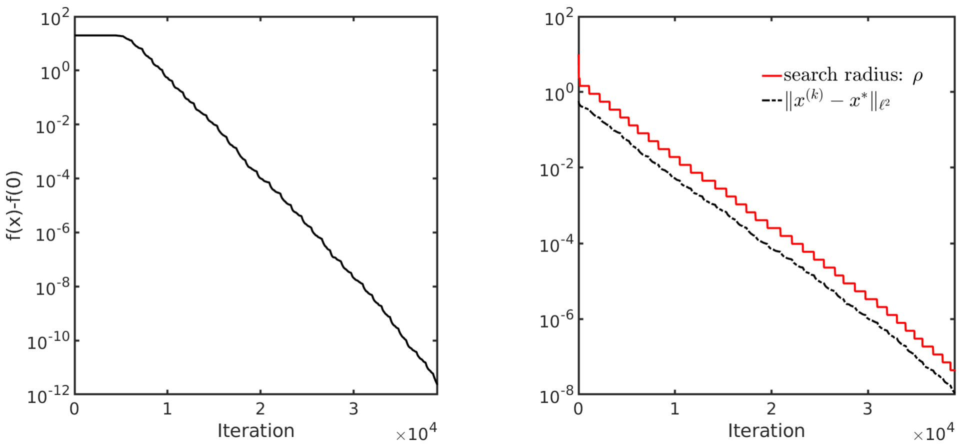


Figure 4: The iteration process of the adaptive HiCS method to 1000 dimensional Gaussian function. Start point is randomly generated in the space $[-1000, 1000]^{1000}$, $\rho = 2.0$ and control factor $\eta = (\sqrt{5}-1)/2$. The left plot is the energy difference, and the right one is the search radius and the ℓ^2 distance between the current iterator and the global minimizer $x^* = 0$.

4.2. The multimodal problems: Ackley and Arwhead functions

The second test objective function is the Ackley function [5] which is a widely used benchmark function for testing optimization algorithms. The expression of the Ackley function can be written as

$$f(x) = -20 \cdot \exp\left(-\frac{1}{5} \cdot \sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(2\pi x_i)\right) + 20 + e, \tag{5}$$

where n is the dimension. Ackley function has many local minima and a unique global minimum of 0 with $f(0) = 0$, which poses a risk for optimization algorithms to be trapped into one of local minima, such as the traditional hill-climbing method [6]. Our previous result has shown that the HiCS method can capture different local minimizer and the global minimizer for 2 dimensional problem through the choice of different ρ [4]. In this subsection, we will apply the improved HiCS algorithm to higher dimensional Ackley function. In the following simulation, the control factor $\eta = (\sqrt{5} - 1)/2$.

Table 1: The successful number N_s of capturing the global minimizer for each different initial search radius ρ_0 when applying the adaptive HiCS method to 100 dimensional Ackley function from 100 time numerical experiments. The initial values are randomly generated in $[-10, 10]^{100}$.

ρ_0	2.0	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2
N_s	98	99	97	73	93	100	99	84	76	57
ρ_0	0.1	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01
N_s	75	79	72	69	84	86	52	0	0	0

We first take 100 dimensional Ackley function as an example to test the performance of our proposed algorithm for finding minimizers. We run adaptive HiCS method 100 times for each different initial search radius ρ_0 from 0.01 to 2.0. The start points are all randomly generated in $[-10, 10]^{100}$. The convergent criterion is the search radius smaller than 10^{-10}. Tab. 1 gives the successful number N_s of capturing the global minimizer. When the algorithm is successful, the distance between the convergent iterator and the global minimizer is smaller than the search radius $\rho < 10^{-10}$. From these results, it is easy to find that our method is able to approximate the global minimizer. The value of ρ_0 heavily affects the probability of obtaining the global minimizer. When $\rho_0 > 0.04$, the adaptive HiCS can find the global minimizer with high probability. When ρ_0 is about 0.04, the successful probability is falling quickly to about 50%. As ρ_0 decreases to smaller than 0.03, the HiCS could not find the global minimizer. In addition, it should be pointed out that these so-called unsuccessful experiments have obtained other local minimizers.

We continue to apply the adaptive HiCS method to 2500 dimensional Ackley function. The initial search radius is $\rho_0 = 3.5$, and the initial position is generated randomly in $[-10, 10]^{2500}$. The iteration process is presented in Fig. 5. For such a high dimensional optimization problem, the iteration behavior is similar to previous numerical experiments. When $\rho = 3.5$, the HiCS method costs 90 steps to achieve convergence. By further shrinking search radius, the adaptive HiCS can capture global minimizer. As one can see from Fig 5, the global minimizer always locates in the search neighbourhood in this case. It demonstrates that the HiCS has the capacity of hupping the local basin even for such a high dimensional problem.

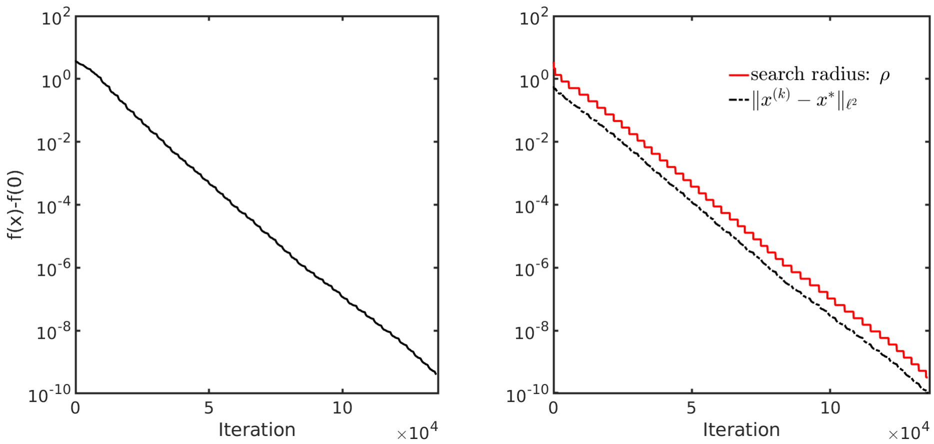


Figure 5: The iteration process of the adaptive HiCS method to 2500 dimensional Ackley function with initial search radius $\rho_0 = 3.5$. Start point is randomly generated in the space $[-10, 10]^{2500}$.

The last benchmark example is the Arwhead function, which has been also used by Powell to test the NEWUOA derivative-free method [7]. The expression of the Arwhead function is

$$f(x) = \sum_{i=1}^{d-1} [(x_i^2 + x_n^2)^2 - 4x_i + 3]. \tag{6}$$

The least value of f is zero, which occurs when the minimizer x^* take the values $x_j = 1$, $j = 1, 2, \ldots, d-1$ and $x_d = 0$. We directly apply the adaptive HiCS method ($\eta = 0.5$) to 1000 dimensional Arwhead function. The starting vector is given by $x_j^{(0)} = 1$, $j = 1, 2, \ldots, d$, as Powell done in Ref. [7] The initial search radius $\rho_0 = 3$ and $\eta = (\sqrt{5} - 1)/2$.

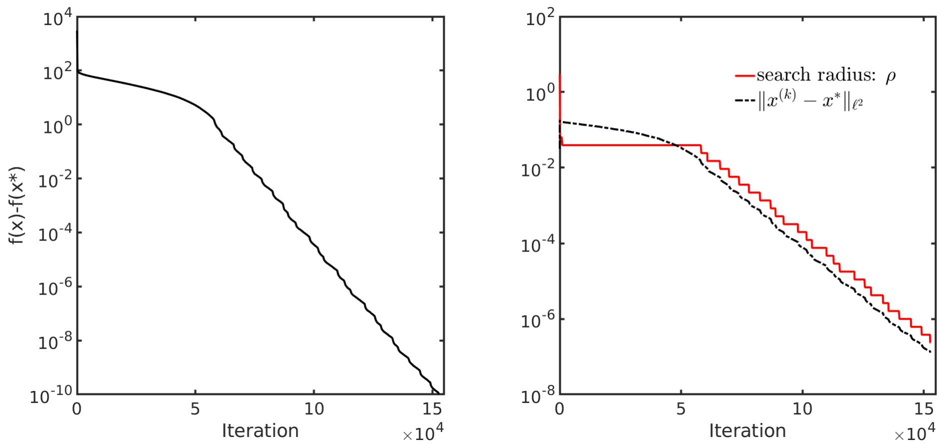


Figure 6: The iteration process of the adaptive HiCS method ($\eta = (\sqrt{5}-1)/2$) to the 1000 dimensional Arwhead function.

Fig. 6 gives the iteration process of applying the adaptive HiCS algorithm to 1000 dimensional Arwhead function. Obviously, the sequences of function values and iterators both approximate the global minimum and the global minimizer, respectively. The function value always decreases as the proposed algorithm indicates. While the distance $\|x^{(k)} - x^*\|_{\ell^2}$ demonstrates more interesting phenomena. At the beginning, the search radius ρ is larger than the distance which means the global minimizer x^* is in the search neighbourhood. Then when the distance is about 1.67×10^{-1}, the ρ is smaller than the distance which indicates x^* is not in the search neighbourhood. It means that the iterator locates in the valley of a local minimizer. However, as iteration evolves, the HiCS algorithm can jump out of the trap of the local energy well, and again contains the global minimizer in the search region.

5. Discussion

Inspired by the hill-climbing behavior of the blind, we has proposed a new derivative-free method to unconstrained optimization problems in our previous work [4]. In this paper, we establish a rigorous mathematical theory of the HiCS algorithm which theoretically guarantees the finite-step convergence under mild conditions. Numerical results also have demonstrated the satisfactory property. In practice, the computational complexity of the HiCS algorithm mainly depends on the discretized strategy on search boundary. In our previous work, the number of discretized points increases exponentially with the dimension of problems. It restricts the application to the high-dimensional optimization. In this work, to deal with high-dimensional problems, we proposed a new strategy of simplex discretization method to save computational amount. Using the simplex method, the number of function valuations is linearly dependent on the dimension of problems, which allows us to solve high dimensional optimization problems. Finally we demonstrate the efficiency of our proposed algorithm through solving several higher dimensional benchmark problems.

Acknowledgments

[1] W. Y. Sun and Y. Yuan, Optimization theory and methods: nonlinear programming, New York: Springer, 2006.

[2] J. Nocedal and S. J. Wright, Numerical optimization, Berlin: Springer-Verlag, 2nd ed., 2006.

[3] A. R. Conn, K. Scheinberg and L. N. Vicente, Introduction to derivative-free optimization, Philadelphia: SIAM, 2009.

[4] Y. Huang and K. Jiang, Hill-Climbing Algorithm with a Stick for Unconstrained Optimization Problems, Adv. Appl. Math. Mech., 2017, 9: 307–323.

[5] J. M. Dieterich and B. Hartke, Empirical review of standard benchmark functions using evolutionary global optimization, Appl. Math. 2012, 3: 1552–1564.

[6] T. Bäck, Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms, Oxford University Press, 1996.

[7] M. J. D. Powell, The NEWUOA software for unconstrained optimization without derivative, in Large-Scale Nonlinear Optimization , eds. G. Di Pillo and M. Roma, Springer (New York), 2006, 255 297.