ANALYSIS OF DIFFERENT NORMS AND CORRESPONDING LIPSCHITZ CONSTANTS FOR GLOBAL OPTIMIZATION

Remigijus Paulavičius[1], Julius Žilinskas[2]

[1]*Vilnius Pedagogical University, Studentų g. 39, LT-08106 Vilnius, Lithuania*
[2]*Institute of Mathematics and Informatics, Akademijos g. 4, LT-08663 Vilnius, Lithuania*
E-mail: [1]r.paulavicius@vpu.lt, [2]julius.zilinskas@mii.lt

Abstract. The paper discusses how the used norm and corresponding Lipschitz constant influence the speed of algorithms for global optimization. For this reason Lipschitz constants corresponding to different norms were estimated. Different test functions for global optimization were solved using branch-and-bound algorithm for Lipschitz optimization with different norms. Experiments have shown that the best results are achieved when combination of extreme (infinite and first) and sometimes Euclidean norms is used.

Keywords: global optimization, branch-and-bound algorithms, Lipschitz optimization, different norms, Lipschitz constant.

1. Introduction

Global optimization is considered in this paper. Mathematically the problem is formulated as

$$f^* = \max_{x \in D} f(x),$$

where an objective function $f(x)$, $f : \Re^n \to \Re$, is a nonlinear function of continuous variables, $D \subseteq \Re^n$ is a feasible region, n is number of variables. Besides the global optimum f^* one or all global optimizers $x^* : f(x^*) = f^*$ should be found.

Branch and bound is a technique to solve global optimization problems. Branch-and-bound algorithms divide the feasible region into subregions and detect subregions which cannot contain global optimizer evaluating bounds for the optimum over considered subregions. Optimization stops when global optimizers are bracketed in small subregions guaranteeing the required accuracy.

Lipschitz optimization is one of the most deeply investigated subjects of global optimization. It is based on the assumption that the slope of an objective function is bounded [1]. The advantages and disadvantages of Lipschitz global optimization methods are discussed in [1, 2]. A function $f : D \to \Re$, $D \subseteq \Re^n$, is said to be Lipschitz if it satisfies the condition

$$\left| f(x) - f(y) \right| \le L \left\| x - y \right\|, \forall x, y \in D , \qquad (1)$$

where $L > 0$ is a constant called Lipschitz constant, D is compact and $\| \; \|$ denotes the norm. Euclidean norm is used most often, but other norms could also be considered. In this work we investigate how the used different norm and corresponding Lipschitz constant influence the speed of algorithms for global optimization.

2. Branch and bound with simplicial partitions for Lipschitz optimization

The general n-dimensional simplex-based branch and bound algorithm for Lipschitz optimization is proposed in [3]. The rules of selection, covering, branching and bounding are justified by the results of experimental investigation.

An n-dimensional simplex is the convex hull of a set of $n + 1$ affinely independent points in n-dimensional space. In one-dimensional space a simplex is a line segment, in two-dimensional space it is a triangle, in three-dimensional space it is a tetrahedron. A simplex is a polyhedron in n-dimensional space which has a minimal number of vertexes $(n + 1)$. Therefore, if bounds for the optimum over a sub-region defined by polyhedron are estimated using function values at all vertexes of the polyhedron, a simplex

subregion requires the smallest number of function evaluations to estimate bounds.

Usually, a feasible region in Lipschitz optimization is defined by a hyper-rectangle – intervals of variables. To use simplicial partitions, the feasible region should be covered by simplexes. Experiments in [3] have shown that the most preferable initial covering is face to face vertex triangulation – partitioning of the feasible region into finitely many n-simplexes, the vertexes of which are also the vertexes of the feasible region.

There are several ways of dividing the simplex into subsimplexes. Experiments in [3] have shown that the most preferable partitioning is subdivision of simplex into two by a hyper-plane passing through the middle point of the longest edge and the vertexes which do not belong to the longest edge.

In Lipschitz optimization the upper bound for the optimum is found exploiting Lipschitz condition (1):

$$f(x) \leq f(y) + L\|x - y\|.$$

It is suggested in [3] to estimate the bounds for the optimum over the simplex using function values at one or more vertexes. The lower bound for the optimum is the largest value of the function at the vertex:

$$LB(I) = \max_{x_v \in I} f(x_v),$$

where x_v is a vertex of the simplex I. The upper bound for the optimum

$$UB(I) = \min_{x_v} \left(f(x_v) + L \max_{x \in I} \|x - x_v\| \right).$$

In this work the values of function at all the vertexes of the simplex are used. The branch-and-bound algorithm may be represented by the following pseudo-code:

$LB = -\infty$
$UB = \infty$
While $(UB - LB > \Sigma)$ **AND** $(D \neq \varnothing)$
    $UB = \infty$
    Take a simplex I from D
    $LB = \max (LB, \max_{x_v \in I} f(x_v))$
    For all the vertexes x_v of simplex I
    $UB = \mathbf{min} (UB, \min \left(f(x_v) + L \mathbf{\max}_{x \in I} \|x - x_v\| \right))$
    If $(UB \geq LB)$ divide simplex I into two.

3. Norms and corresponding Lipschitz constants

Although in Lipschitz optimization different norms can be considered to calculate distances:

$$\|x\|_q = \left(\sum_{i=1}^{n} |x_i|^q \right)^{1/q}, \quad x \in \mathfrak{R}^n.$$

Euclidean norm $(q = 2)$ is used most often. The efficient algorithms for Lipschitz optimization exist for one-dimensional case. In one-dimensional case, all norms are equal. Evaluated bounds in a multidimensional case depend on the norm used.

The estimate of Lipschitz constant should be avail-able to evaluate bounds exploiting Lipschitz condition. The value of Lipschitz constant depends on the used norm.

Theorem 1. For Lipschitz function $f(x), f : \mathfrak{R}^n \to \mathfrak{R}$,

$$|f(x) - f(y)| \leq L_p \|x - y\|_q , \qquad (2)$$

where $L_p = \sup \left\{ \|\nabla f(x)\|_p : x \in D \right\}$ is Lipschitz constant,

$\nabla f(x) = \left(\dfrac{\partial f}{\partial x_1}, \ldots, \dfrac{\partial f}{\partial x_n} \right)$ is gradient of the function $f(x)$, and $1/p + 1/q = 1, 1 \leq p, q \leq \infty.$

To make the proof clearer let us assume that $n = 2$. Then for $\forall x, y \in D$, we get

$$|f(x) - f(y)| =$$
$$|f(x_1, x_2) - f(y_1, x_2) + f(y_1, x_2) - f(y_1, y_2)|$$

(using inequality of triangle)

$$\leq \frac{|f(x_1, x_2) - f(y_1, x_2)|}{|x_1 - y_1|} \cdot |x_1 - y_1| +$$
$$\frac{|f(y_1, x_2) - f(y_1, y_2)|}{|x_2 - y_2|} \cdot |x_2 - y_2| =$$

(from middle values theorem)

$$\sum_{i=1}^{2} |f_{x_i}{}'(c_i)| |x_i - y_i|$$

(using Hölder's inequality)

$$\leq \left(\sum_{i=1}^{2} |f_{x_i}{}'(c_i)|^p \right)^{1/p} \left(\sum_{i=1}^{2} |x_i - y_i|^q \right)^{1/q},$$

where $1 \leq p, q \leq \infty, 1/p + 1/q = 1, c = (c_1, c_2) \in D.$ Since

$$\|x - y\|_q = \left(\sum_{i=1}^{2} |x_i - y_i|^q \right)^{1/q},$$

$$\|\nabla f(c)\|_p = \left(\sum_{i=1}^{2} |f_{x_i}{}'(c_i)|^p \right)^{1/p},$$

$$\left| f\left(x \right) - f\left(y \right) \right| \le \left\| \nabla f\left(c \right) \right\|_p \left\| x - y \right\|_q \le$$

$$\sup \left\{ \left\| \nabla f\left(c \right) \right\|_p : c \in D \right\} \left\| x - y \right\|_q .$$

Denoting

$$L_p = \sup \left\{ \left\| \nabla f\left(c \right) \right\|_p : c \in D \right\},$$

we get (2).

In conclusion, the supremum of p norm of gradient of the function is Lipschitz constant for q norm in Lipschitz condition.

4. Experimental rezults

4.1. Estimation of Lipschitz constants

Various test functions for global optimization from [2] and [4] were used in our experiments. Lipschitz constants were estimated using Theorem 1. Therefore, points of the feasible regions, where the norm of gradient of the objective function is the largest, were found. The largest values of derivatives and Lipschitz constants for some norms are shown in Table 1. Test functions are numbered according to [2] and [4].

If the estimate of Lipschitz constant is smaller than its actual value, the global optimum may be missed. If the es-

timate of Lipschitz constant is greater than its actual value, the speed of optimization may be degraded. Therefore, the estimate of Lipschitz constant should be as accurate as possible, but not smaller than its actual value.

The estimates of Lipschitz constant of the second norm ($p = q = 2$) L_2 were compared to those given in [2] which are repeated in Table 1. Most of the estimates co-incide. However, the estimates of Lipschitz constants of test functions 20 and 21 are quite different. We believe that our estimates are correct. In the case of test function 20, the values of derivatives at the feasible point $x = 1.0$, $y = 1.0$, $z = 0$ are equal to 200 and, therefore,

$$L_2 = \sqrt{200^2 + 200^2 + 200^2} \approx 346.41.$$

Test function 21 was optimized successfuly with different norms and corresponding estimated Lipschitz constants. Therefore, there is no reason to doubt our estimates.

Test function 23 is three-dimensional generalization of two-dimensional test function 13. If $z = 0.25$ test function 23 becomes test function 13 with $L_2 = 988.8$ which is larger than 971.59.

4.2. Optimization of two-dimensional test functions

The speed of global optimization was estimated using the number of function evaluations criterion.

The results of experiments for $n = 2$ are shown in Table 2. Test functions are numbered according to [2]. No single

Table 1. The largest values of derivatives and estimated Lipschitz constants

Test function	$\left\| f_{x_1}{}'(c_1) \right\|$	$\left\| f_{x_2}{}'(c_2) \right\|$	$\left\| f_{x_3}{}'(c_3) \right\|$	L_1	L_2	L_2 [2]	L_∞
1. [2]	50.2665	0.0	–	50.2665	50.266	50.2665	50.2665
2. [2]	1.98	6.0	–	7.98	6.3183	6.3183	6.0
3. [2]	107.09	34.248	–	141.338	112.43	112.44	107.09
4. [2]	0.0	2.0	–	2.0	2	2	2.0
5. [2]	1.0	1.0	–	2.0	$\sqrt{2}$	$\sqrt{2}$	1.0
6. [2]	256.8	1028.0	–	1284.8	1059.59	1059.59	1028.0
7. [2]	2100.0	12608.0	–	14708.0	12781.7	12781.7	12608.0
8. [2]	59.0	63.0	–	122.0	86.3134	86.3134	63.0
9. [2]	2177520.0	461520.0	–	2639040	2225891.7	2225892	2177520.0
10. [2]	13.0	11.0	–	24.0	17.029	17.034	13.0
11. [2]	42.015	22.0	–	64.015	47.426	47.426	42.015
12. [2]	40.4	40.0	–	80.4	56.582	56.852	40.4
13. [2]	988.81	0.844	–	989.654	988.81	988.82	988.81
20. [2]	200.0	200.0	200.0	600.0	346.41	244.95	200.0
21. [2]	18.32	0.32	0.012	18.652	18.322	42.626	18.32
23. [2]	988.78	0.01	0.16	988.95	988.81	971.59	988.78
24. [2]	10.708	12.108	10.7	33.516	19.39	19.39	12.108
25. [2]	1.192	2.383	1.192	4.767	2.919	2.919	2.383
26. [2]	80.0	80.0	64.0	224.0	129.99	130.0	80.0
5. [4]	14408.0	16808.0	2400.0	33616.0	22267.9	–	16808.0

Table 2. Numbers of function evaluations for $n = 2$

Test function	ε	$L_{1,\infty}\|x-y\|_{\infty,1}$	$L_\infty\|x-y\|_1$	$L_3\|x-y\|_{1.5}$	$L_2\|x-y\|_2$	$L_1\|x-y\|_\infty$
1. [2]	0.355	1087	2154	1587	1432	1087
2. [2]	0.0446	266	393	362	353	312
3. [2]	11.9	7116	11044	9108	9046	11187
4. [2]	0.0141	53	56	76	82	95
5. [2]	0.1	102	102	157	165	199
6. [2]	44.9	1869	3021	2674	2164	2134
7. [2]	542.0	12086	12661	12486	12442	21304
8. [2]	3.66	571	587	688	774	1131
9. [2]	62900	15177	15402	15317	15388	28868
10. [2]	0.691	2464	2542	2731	3109	4577
11. [2]	0.335	5718	6978	6450	7054	9674
12. [2]	0.804	29605	29629	30195	37982	59146
13. [2]	6.92	22811	45856	37941	37941	22895

norm and corresponding Lipschitz constant is the best for all test functions. The best results were achieved using combination of two extreme norms

$$L_{1,\infty}\|x-y\|_{\infty,1} = \min\left\{L_1\|x-y\|_\infty, L_\infty\|x-y\|_1\right\}. \quad (3)$$

Let us investigate this case. In case $n = 2$ the feasible region is square. Let us assume that it is unit square $D = [0,1] \times [0,1]$. It is covered by two right-angled equilateral triangles (see Fig 1).

When the right-angled equilateral triangle is divided into two through the middle of the longest edge, two right-angled equilateral triangles are produced again. Since the upper bound of optimum over the triangle is estimated using the values of function at the vertexes and the maximum edges from the considered vertex, two cases are possible. Let us investigate the simplex ABC:

1) The distance from the vertex A to the vertex B and from the vertex A to the vertex C (see Fig 1) is equal to 1 with all norms:

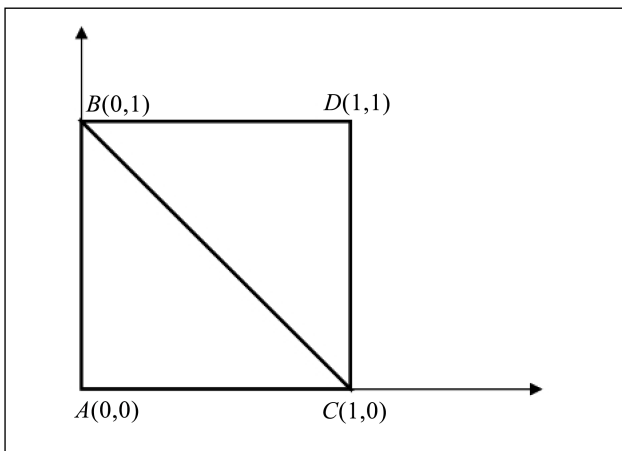**Fig 1.** Covering of square by simplexes (triangles)

$$\|A-B\|_p = \left(|0-0|^p + |0-1|^p\right)^{1/p} = 1,$$

$$\|A-C\|_p = \left(|0-1|^p + |0-0|^p\right)^{1/p} = 1.$$

Therefore, in this case the upper bound is the smallest when Lipschitz constant is the smallest, i e $L_\infty\|x-y\|_1$, as $\|A-B\|_1 = \ldots = \|A-B\|_\infty$, $\|A-C\|_1 = \ldots = \|A-C\|_\infty$, and $L_\infty \le L_p$, where $1 \le p \le \infty$.

2) In the case of the vertexes B and C the smallest upper bound is when $L_1\|x-y\|_\infty$ is used.

It is shown that the distance from the vertexes B and C to the vertex A is equal to 1, and the distance between the vertexes B and C depends on the norm used:

$$\|B-C\|_p = \left(|0-1|^p + |1-0|^p\right)^{1/p} = 2^{1/p}, \text{ when } p \ge 1,$$

$$\|B-C\|_\infty = \max\left\{|0-1|, |1-0|\right\} = 1.$$

Therefore,

$$L_\infty\|x-y\|_1 = \max\left\{\left|f_{x_1}'(c_1)\right|, \left|f_{x_2}'(c_2)\right|\right\} \cdot 2,$$

$$L_p\|x-y\|_q = \left(\left|f_{x_1}'(c_1)\right|^p + \left|f_{x_2}'(c_2)\right|^p\right)^{1/p} \cdot 2^{1/q},$$

where $p > 1, q > 1$ and $\dfrac{1}{p} + \dfrac{1}{q} = 1$. From the last equation $q = \dfrac{p}{p-1}$. Denoting

$$a = \left|f_{x_1}'(c_1)\right|, \ b = \left|f_{x_2}'(c_2)\right|,$$

$$\left(a^p + b^p\right)^{1/p} 2^{\frac{p-1}{p}} \ge (a+b).$$

In the case when $a = b$, the inequality becomes equality. In other cases the right-hand side is strictly greater than the left-hand side, i e

$$L_p \left\| x - x_v \right\|_q \geq L_1 \left\| x - x_v \right\|_\infty.$$

Therefore, both experimental (Table 2) and theoretical investigation shows that the combination of two extreme cases (3) gives the best bounds for branch and bound with right-angle equilateral triangle partitions.

When the combination (3) is used, the number of function evaluations is by 1 to 40 % smaller than that when Euclidean norm is used. On average, for all the test functions $n = 2$ used in our experiments, the combination (3) reduces the number of function evaluations by 22 %.

4.3. Optimization of three-dimensional test functions

The experimental results for $n = 3$ are shown in Table 3. Test functions are numbered according to [2] and [4]. No single norm and corresponding Lipschitz constant is the best for all the test functions. Combination of two extreme norms is not the best for all the test functions too. In this case the best results were achieved using combination of two extreme and Euclidean norms:

$$L_{1,\infty,2} \left\| x - x_v \right\|_{\infty,1,2} =$$
$$\min \left\{ L_1 \left\| x - x_v \right\|_\infty, L_\infty \left\| x - x_v \right\|_1, L_2 \left\| x - x_v \right\|_2 \right\}. \quad (4)$$

Let us investigate this case when $n = 3$. Let us as-sume that the feasible region is unit cube $D = [0,1] \times [0,1] \times [0,1]$. It is covered by four tetrahedrons (see Fig 2). Let us investigate the tetrahedron *ABCE*.

1) The distance from the vertex *A* to the vertex *B*, from the vertex *A* to the vertex *C* and from the vertex *A* to the vertex *E* (see Fig 2) is equal to 1 with all norms:

$$\left\| A - B \right\|_p = \left(\left| 0 - 0 \right|^p + \left| 0 - 1 \right|^p + \left| 1 - 1 \right|^p \right)^{1/p} = 1,$$

$$\left\| A - C \right\|_p = \left(\left| 0 - 1 \right|^p + \left| 0 - 0 \right|^p + \left| 1 - 1 \right|^p \right)^{1/p} = 1,$$

$$\left\| A - E \right\|_p = \left(\left| 0 - 0 \right|^p + \left| 0 - 0 \right|^p + \left| 1 - 0 \right|^p \right)^{1/p} = 1.$$

Therefore, in this case (similarly to the case when $n = 2$), the upper bound is the smallest when Lipschitz constant is the smallest, i e $L_\infty \left\| x - y \right\|_1$.

2) In the vertexes *B*, *C* and *E* when $n = 3$, the combination $L_1 \left\| x - y \right\|_\infty$ does not always give the best upper bound. Similarly to the case when $n = 2$, the largest distance from a vertex, say the vertex *E*, to the other vertexes is $2^{1/p}$:

$$\left\| E - B \right\|_p = \left(\left| 0 - 0 \right|^p + \left| 0 + 1 \right|^p + \left| 0 - 1 \right|^p \right)^{1/p} = 2^{1/p},$$
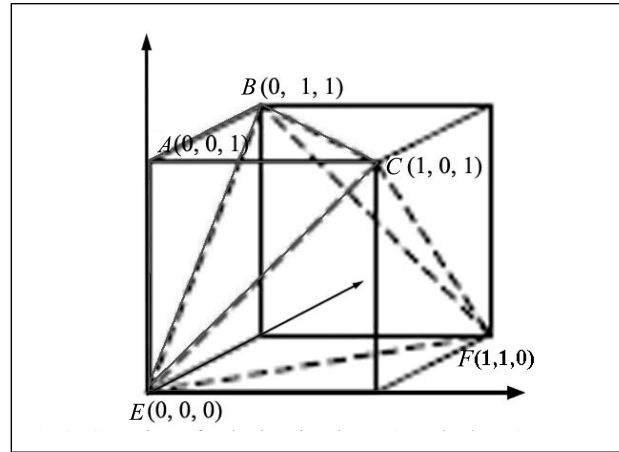
Fig 2. Covering of cube by simplexes (tetrahedrons)

$$\left\| E - C \right\|_p = \left(\left| 0 - 1 \right|^p + \left| 0 - 0 \right|^p + \left| 0 - 1 \right|^p \right)^{1/p} = 2^{1/p},$$

$$\left\| E - A \right\|_p = \left(\left| 0 - 0 \right|^p + \left| 0 - 0 \right|^p + \left| 0 - 1 \right|^p \right)^{1/p} = 1.$$

However, there is an additional third derivative $d = \left| f_{x_3}{}'(c_3) \right|$ for estimation of Lipschitz constant, and the inequality

$$\left(a^p + b^p + c^p \right) 2^{\frac{p-1}{p}} \geq \left(a + b + c \right),$$

is not always correct. The inequality is correct if one of the derivatives is equal to zero. Therefore, it is useful to include Euclidean norm in the combination (4).

Similarly, in the case of the tetrahedron *EBCF*, the maximal distance between vertexes is $2^{1/p}$:

$$\left\| E - B \right\|_p = \left(\left| 0 - 0 \right|^p + \left| 0 - 1 \right|^p + \left| 0 - 1 \right|^p \right)^{1/p} = 2^{1/p},$$

$$\left\| E - F \right\|_p = \left(\left| 0 - 1 \right|^p + \left| 0 - 1 \right|^p + \left| 0 - 1 \right|^p \right)^{1/p} = 2^{1/p},$$

$$\left\| E - C \right\|_p = \left(\left| 0 - 1 \right|^p + \left| 0 - 0 \right|^p + \left| 0 - 1 \right|^p \right)^{1/p} = 2^{1/p},$$

$$\left\| B - C \right\|_p = \left(\left| 0 - 1 \right|^p + \left| 1 - 0 \right|^p + \left| 1 - 1 \right|^p \right)^{1/p} = 2^{1/p},$$

$$\left\| B - F \right\|_p = \left(\left| 0 - 1 \right|^p + \left| 1 - 1 \right|^p + \left| 1 - 0 \right|^p \right)^{1/p} = 2^{1/p},$$

$$\left\| C - F \right\|_p = \left(\left| 1 - 1 \right|^p + \left| 0 - 1 \right|^p + \left| 1 - 0 \right|^p \right)^{1/p} = 2^{1/p}.$$

When the combination (4) is used, the number of function evaluations is by 8 to 58 % smaller than that when Euclidean norm is used. On average, for all the test functions $n = 3$ used in our experiments, the combination (4) reduces the number of function evaluations by 39 %.

Table 3. Number of function evaluations for $n = 3$

Test function	ε	$L_{1,\infty} \left\| x - y \right\|_{\infty,1}$	$L_{1,\infty,2} \left\| x - y \right\|_{\infty,1,2}$	$L_{\infty} \left\| x - y \right\|_1$	$L_2 \left\| x - y \right\|_2$	$L_1 \left\| x - y \right\|_{\infty}$
20. [2]	10.6	106877	75163	106877	178925	419088
21. [2]	0.369	8866	8866	46964	16754	8896
23. [2]	41.65	96673	96673	527719	179953	96673
24. [2]	3.36	69743	61492	69756	100175	245398
25. [2]	0.0506	47260	20335	47487	23148	82533
26. [2]	4.51	21598	18679	21598	44189	131994
5. [4].	5000.0	159559	103694	166157	117225	284069

5. Conclusions

When the upper bound of optimum is evaluated in simplex based branch and bound for Lipschitz optimization using the values of function at vertexes of simplexes, no single norm and corresponding Lipschitz constant is the best for all the test functions. When $n = 2$, the best results were achieved using combination of two extreme (infinite and first) norms, and in this case the number of function evaluations is by 22 % smaller that when Euclidean norm is used. When $n = 3$, the best results were achieved using combination of two extreme (infinite and first) and Euclidean norms, and in this case the number of function evaluations is by 39 % smaller than that when Euclidean norm is used alone.

6. Acknowledgements

References

1. Horst, R.; Pardalos, P. M. and Thoai, N. V. Introduction to Global Optimization. Kluwer Academic Publishers, 1995.

2. Hansen, P. and Jaumard, B. Lipshitz optimization. In: Horst, R. and Pardalos, P. (eds.) Handbook of Global Optimization. Kluwer Academic Publishers, 1995, p. 404–493.

3. Žilinskas, J. Optimization of Lipschitzian functions by simplex-based branch and bound. *Information Technology and Control*, Vol 14, No 1, 2000, p. 45–50.

4. Madsen, K. and Žilinskas, J. Testing Branch-and-Bound Methods for Global Optimization. Technical Report IMM-REP-2000-05, Technical University of Denmark, 2000.

SKIRTINGŲ NORMŲ IR JAS ATITINKANČIŲ LIPŠICO KONSTANTŲ ANALIZĖ GLOBALIAI OPTIMIZACIJAI

R. Paulavičius, J. Žilinskas

Santrauka

Šiame darbe ištirta, kaip įvairios normos ir jas atitinkančios Lipšico konstantos veikia globalios optimizacijos algoritmų greičius. Šiam tikslui buvo įvertintos įvairias normas atitinkančios Lipšico konstantos. Šakų ir rėžių algoritmas buvo naudojamas globalaus maksimumo paieškai. Eksperimento rezultatai parodė, kad geriausi rezultatai gaunami, kai Lipšico viršutiniam rėžiui įvertinti naudojamas kraštinių (begalinės ir pirmosios) normų junginys dvimačiam atvejui ir jų junginys su euklidine norma trimačiam atvejui.

Reikšminiai žodžiai: globali optimizacija, šakų ir rėžių algoritmas, Lipšico optimizacija, normos, Lipšico konstanta.

Remigijus PAULAVIČIUS. Doctoral student. Department of Systems Analysis, Institute of Mathematics and Informatics. First degree in Mathematics and Informatics, Vilnius Pedagogical University (2004). Master of Science (2006). Research interests: local and global optimization, parallel computing.

Jukius ŽILINSKAS. Senior research worker. Department of Systems Analysis, Institute of Mathematics and Informatics. Doctor of Science (informatics), Kaunas University of Technology (2002). Research interests: global optimization and parallel computing.