ANALYSIS OF DIFFERENT NORMS AND CORRESPONDING LIPSCHITZ CONSTANTS FOR GLOBAL OPTIMIZATION

Remigijus Paulavičius¹, Julius Žilinskas²

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

Received 15 June 2006; accepted 20 November 2006

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 \( f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \) is a nonlinear function of continuous variables, \( D \subseteq \mathbb{R}^n \) 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 sub-regions 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 \( f : D \rightarrow \mathbb{R}, D \subseteq \mathbb{R}^n \), is said to be Lipschitz if it satisfies the condition

\[ |f(x) - f(y)| \leq L \|x - y\| \forall x, y \in D , \quad (1) \]

where \( L > 0 \) is a constant called Lipschitz constant, \( D \) is compact and \( \| \cdot \| \) 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 vertices \( (n + 1) \). 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 sub-simplices. 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):

\[
|f(x) - f(y)| \leq L \|x - y\|.
\]

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:

\[
LB(I) = \max_{x_i \in I} f(x_i),
\]

where \( x_i \) is a vertex of the simplex \( I \). The upper bound for the optimum

\[
UB(I) = \min_{x_i \in I} \left( f(x_i) + L \max_{x_i \in I} \| x - x_i \| \right)
\]

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:

\[
\begin{align*}
\text{LB} &= -\infty \\
\text{UB} &= \infty \\
\text{While } & (\text{UB} - \text{LB} > \text{Sum}) \text{ AND } (D \neq \emptyset) \\
\text{UB} &= \infty \\
\text{Take a simplex } I \text{ from } D \\
\text{LB} &= \max \left( \text{LB}, \max_{x_i \in I} f(x_i) \right) \\
\text{For all the vertexes } x_i \text{ of simplex } I \\
\text{UB} &= \min \left( \text{UB}, \min \left( f(x_i) + L \max_{x_i \in I} \| x - x_i \| \right) \right) \\
\text{If } (\text{UB} \geq \text{LB}) \text{ divide simplex } I \text{ into two.}
\end{align*}
\]

3. Norms and corresponding Lipschitz constants

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

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

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 available to evaluate bounds exploiting Lipschitz condition. The value of Lipschitz constant depends on the used norm.

Theorem 1. For Lipschitz function \( f(x), f: \mathbb{R}^n \rightarrow \mathbb{R} \),

\[
|f(x) - f(y)| \leq L_p \|x - y\|,
\]

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

\[
\nabla f(x) = \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)
\]

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

\[
|f(x) - f(y)| \leq |f(x, y) - f(y, y)| + |f(y, y) - f(y, y)|
\]

(1)

(using inequality of triangle)

\[
|f(x, y) - f(y, y)| = |f(y, x) - f(y, y)| + |f(y, y) - f(y, y)|
\]

(2)

(from middle values theorem)

\[
\frac{1}{2} \sum_{i=1}^{n} f_{x_i}'(c_i) \|x_i - y_i\|
\]

(3)

(using Hölder’s inequality)

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

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

\[
\|x - y\|_p = \left( \sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p},
\]

\[
\|\nabla f(c)\|_p = \left( \sum_{i=1}^{n} |f_{x_i}'(c_i)|^p \right)^{1/p},
\]
\[
\left| f(x) - f(y) \right| \leq \left\| \nabla f(c) \right\|_p \left\| x - y \right\|_p
\]
\[
\sup_{c \in D} \left\| \nabla f(c) \right\|_p \leq L_p
\]

Denoting
\[
L_p = \sup_{c \in D} \left\| \nabla f(c) \right\|_p
\]
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 results

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 estimate 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 coincide. 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 successfully 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>
<thead>
<tr>
<th>Test function</th>
<th>$f_{x_1}'(c_1)$</th>
<th>$f_{x_2}'(c_2)$</th>
<th>$f_{x_3}'(c_3)$</th>
<th>$L_1$</th>
<th>$L_2$</th>
<th>$L_3$ [2]</th>
<th>$L_\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. [2]</td>
<td>50.2665</td>
<td>0.0</td>
<td>--</td>
<td>50.2665</td>
<td>50.266</td>
<td>50.266</td>
<td>50.266</td>
</tr>
<tr>
<td>3. [2]</td>
<td>107.09</td>
<td>34.248</td>
<td>--</td>
<td>141.338</td>
<td>112.43</td>
<td>112.44</td>
<td>107.09</td>
</tr>
<tr>
<td>4. [2]</td>
<td>0.0</td>
<td>2.0</td>
<td>--</td>
<td>2.0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5. [2]</td>
<td>1.0</td>
<td>1.0</td>
<td>--</td>
<td>$\sqrt{2}$</td>
<td>$\sqrt{2}$</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>6. [2]</td>
<td>256.8</td>
<td>1028.0</td>
<td>--</td>
<td>1284.8</td>
<td>1059.59</td>
<td>1059.59</td>
<td>1028.0</td>
</tr>
<tr>
<td>7. [2]</td>
<td>2100.0</td>
<td>12608.0</td>
<td>--</td>
<td>14708.0</td>
<td>12781.7</td>
<td>12781.7</td>
<td>12608.0</td>
</tr>
<tr>
<td>8. [2]</td>
<td>59.0</td>
<td>63.0</td>
<td>--</td>
<td>122.0</td>
<td>86.3134</td>
<td>86.3134</td>
<td>63.0</td>
</tr>
<tr>
<td>9. [2]</td>
<td>2177520.0</td>
<td>461520.0</td>
<td>--</td>
<td>2639040</td>
<td>2225891.7</td>
<td>2225892</td>
<td>2177520.0</td>
</tr>
<tr>
<td>10. [2]</td>
<td>13.0</td>
<td>11.0</td>
<td>--</td>
<td>24.0</td>
<td>17.029</td>
<td>17.034</td>
<td>13.0</td>
</tr>
<tr>
<td>11. [2]</td>
<td>42.015</td>
<td>22.0</td>
<td>--</td>
<td>64.015</td>
<td>47.426</td>
<td>47.426</td>
<td>42.015</td>
</tr>
<tr>
<td>12. [2]</td>
<td>40.4</td>
<td>40.0</td>
<td>--</td>
<td>80.4</td>
<td>56.582</td>
<td>56.582</td>
<td>40.4</td>
</tr>
<tr>
<td>13. [2]</td>
<td>988.81</td>
<td>0.844</td>
<td>--</td>
<td>989.654</td>
<td>988.81</td>
<td>988.82</td>
<td>988.81</td>
</tr>
<tr>
<td>14. [2]</td>
<td>200.0</td>
<td>200.0</td>
<td>200.0</td>
<td>600.0</td>
<td>346.41</td>
<td>244.95</td>
<td>200.0</td>
</tr>
<tr>
<td>15. [2]</td>
<td>18.32</td>
<td>0.32</td>
<td>0.012</td>
<td>18.652</td>
<td>18.322</td>
<td>42.620</td>
<td>18.32</td>
</tr>
<tr>
<td>16. [2]</td>
<td>988.78</td>
<td>0.01</td>
<td>0.16</td>
<td>988.95</td>
<td>988.81</td>
<td>971.59</td>
<td>988.78</td>
</tr>
<tr>
<td>19. [2]</td>
<td>80.0</td>
<td>80.0</td>
<td>64.0</td>
<td>224.0</td>
<td>129.99</td>
<td>130.0</td>
<td>80.0</td>
</tr>
<tr>
<td>5. [4]</td>
<td>14408.0</td>
<td>16808.0</td>
<td>2440.0</td>
<td>33616.0</td>
<td>22267.9</td>
<td>--</td>
<td>16808.0</td>
</tr>
</tbody>
</table>
norm and corresponding Lipschitz constant is the best for all test functions. The best results were achieved using combination of two extreme norms

\[ L_{\infty} \| x - y \|_{\infty} = \min \{ L_1 \| x - y \|_1, L_\infty \| x - y \|_\infty \}. \tag{3} \]

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:

\[ \| A - B \|_p = \left( | 0 - 0 |^p + | 1 - 0 |^p \right)^{1/p} = 1, \]
\[ \| A - C \|_\infty = \max \{ | 0 - 1 |, | 1 - 0 | \} = 1. \]

Therefore, in this case the upper bound is the smallest when Lipschitz constant is the smallest, i.e. \( L_{\infty} \| x - y \|_\infty \), as

\[ \| A - B \|_p = \ldots = \| A - B \|_\infty, \]
\[ \| A - C \|_\infty = \ldots = \| A - C \|_\infty, \]
and

\[ L_{\infty} \leq L_p, \]
where \( 1 \leq p \leq \infty \).

2) In the case of the vertexes \( B \) and \( C \) the smallest upper bound is when \( L_{\infty} \| x - y \|_\infty \) 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:

\[ \| B - C \|_p = \left( | 0 - 1 |^p + | 0 - 0 |^p \right)^{1/p} = 2^{1/p}, \]
when \( p \geq 1 \),
\[ \| B - C \|_\infty = \max \{ | 0 - 1 |, | 1 - 0 | \} = 1. \]

Therefore,

\[ L_{\infty} \| x - y \|_p = \max \left\{ \| f_{x_1}'(c_1) \|, \| f_{x_2}'(c_2) \| \right\} \cdot 2, \]
\[ L_p \| x - y \|_\infty = \left( \| f_{x_1}'(c_1) \| + \| f_{x_2}'(c_2) \| \right)^{1/p}, \]
where \( p > 1, q > 1 \) and \( \frac{1}{p} + \frac{1}{q} = 1 \). From the last equation

\[ q = \frac{p}{p - 1}. \]

Denoting

\[ a = \| f_{x_1}'(c_1) \|, \quad b = \| f_{x_2}'(c_2) \|, \]
\[ \left( a^p + b^p \right)^{1/p} \geq (a + b). \]
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.,

\[
L_p \| x - x_p \| \geq L_1 \| x - x_1 \|.
\]

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:

\[
L_{x,1,2} \| x - x_1 \|, L_{x,1,2} \| x - x_2 \|, L_{x,1,2} \| x - x_1 \|, (4)
\]

Let us investigate this case when \(n = 3\). Let us assume 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:

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

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

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

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., \(L_{x,1,2} \| x - y \|\).

2) In the vertices \(B, C\) and \(E\) when \(n = 3\), the combination \(L_{x,1,2} \| x - y \|\) 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 \(2^{1/p}\):

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

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

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

However, there is an additional third derivative \(d = \int f(x)dx\) for estimation of Lipschitz constant, and the inequality

\[
\left( a^p + b^p + c^p \right)^{1/p} \geq (a + b + c),
\]

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 \(2^{1/p}\):

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

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

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

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%.
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 than 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

The work of the second co-author is supported by Lithuanian State Science and Studies Foundation and the NATO Reintegration grant CBP.EAP.RIG.981300.

References