A Randomized Algorithm to Approximate the Star Discrepancy Based on Threshold Accepting

Michael Gnewuch\textsuperscript{1}

Institut für Informatik
Christian-Albrechts-Universität zu Kiel

August 29, 2011

Joint Work with
Magnus Wahlström & Carola Winzen
Max Planck Institute of Computer Science, Saarbrücken

\textsuperscript{1}Supported by the German Science Foundation (DFG) under grants GN-91/3 and GN-91/4
Star Discrepancy: Definition

\[ X \subset [0, 1]^d \text{ } n \text{-point set, } [0, y) := [0, y_1) \times \cdots \times [0, y_d) \text{ "test box".} \]

Local discrepancy: \( \delta(y) = \delta(y, X) = \text{vol}([0, y)) - \frac{1}{n} |X \cap [0, y)| \)

Star discrepancy: \( \text{disc}^*(X) = \sup_{y \in [0,1]^d} |\delta(y, X)| \)
Known Methods for Calculation

Elementary Method for Exact Calculation

**Simple Observation:** It suffices to consider $2(n + 1)^d$ test boxes to calculate the discrepancy.
Known Methods for Calculation

Elementary Method for Exact Calculation

For \( X = \{x^1, \ldots, x^n\} \subset [0, 1]^d \) put

\[
\Gamma_j(X) := \{x^i_j | i = 1, \ldots, n\} \cup \{1\}, \; j = 1, \ldots, d,
\]

\[
\Gamma(X) := \Gamma_1(X) \times \cdots \times \Gamma_d(X)
\]

Then \( \text{disc}^*(X) = \)

\[
\max_{y \in \Gamma(X)} \max \left\{ \text{vol}([0, y)) - \frac{1}{n} |[0, y) \cap X|, \frac{1}{n} |[0, y] \cap X| - \text{vol}([0, y)) \right\}
\]

Thus \( \text{disc}^*(X) \) can be calculated by considering at most \( 2(n + 1)^d \) test boxes.
Further Methods for Calculation

Improvements of the Elementary Method:

Exact formula for the star discrepancy in dimension $d = 1$ by Niederreiter ’72 ($d = 1$).

Faster methods than the elementary one by De Clerck ’86 ($d = 2$) and Bundschuh and Zhu ’93 ($d \geq 3$). Time to calculate the star discrepancy still $O(n^d)$.

Fastest algorithm to calculate the star discrepancy needs time $O(n^{1+d/2})$ [Dobkin, Eppstein, Mitchell ’96].
Known Methods for Calculation

**Observation**: Exact calculation of star discrepancy is discrete optimization problem.

**Bad news from Discrete Complexity Theory:**

**Theorem** [G., Srivastav, Winzen ‘09].
The calculation of the star discrepancy is $NP$-hard.

**Theorem** [Giannopoulos, Knauer, Wahlström, Werner ‘11].
Calculation of star discrepancy is $W[1]$-hard with respect to parameter $d$. 
Methods for Approximation

Algorithms with error guarantee

- Algorithms from [Thiémand '00, '01a] to approximate the star discrepancy up to some user-specified error $\delta$.
  Cost of the algorithms is $\geq \Omega(d\delta^{-d})$ [G.'08].

Algorithms based on optimization heuristics

- Algorithm from [Thiémand’01b] formulates problem as integer linear program (ILP) and relies on cutting plane and branch-and-bound techniques.
- Algorithm of Winker & Fang ’97 is a local search algorithm relying on the meta heuristic “Threshold Accepting”.
- Genetic algorithm of Shah ’10.
Randomized Approach

**Idea:** Calculate lower bound for star discrepancy by choosing test boxes randomly within a (refined) local search algorithm

**Algorithm of Winker & Fang ("Threshold Accepting")**

Threshold values $T_1 > T_2 > \cdots > T_I \geq 0$

Local neighborhood structure for $y \in \Gamma(X)$

$$N_k(y) \simeq \text{subgrid of } \Gamma(X) \text{ of cardinality } (2k + 1)^d \text{ with center } y,$$

with Laplace measure as probability measure.
Algorithm of Winker & Fang

For \( y \in [0, 1]^d \) put

\[
\delta(y) := \text{vol}([0, y]) - \frac{1}{n} |[0, y] \cap X|, \quad \bar{\delta}(y) := \frac{1}{n} |[0, y] \cap X| - \text{vol}([0, y]),
\]

and \( \delta^*(y) := \max\{\delta(y), \bar{\delta}(y)\} \).

Concrete Algorithm

Choose \( x \) randomly from \( \Gamma(X) \) and put \( x^* := x \).

For \( i = 1 \) to \( I \)

For \( j = 1 \) to \( J \)

Choose \( x \in N_k(x^*) \) randomly

If \( \delta^*(x^*) - \delta^*(x) \leq T_i \) then \( x^* := x \)

Return \( \delta^*(x^*) \).
Improving on the W&F-Algorithm

Modified Neighborhoods:

\[ C_k(x) := \text{conv}(N_k(x)), \]  
equipped with probability measure
\[ \mu_d := \bigotimes_{j=1}^{d} dy_j^{d-1} \lambda(dy_j), \]

\[ \lambda \text{ the Lebesgue measure on } \mathbb{R}. \]

After choosing \( y \in C_k(x) \), round each \( y_j \) up (down) to the next number in \( \Gamma_j(X) \) to get \( y^+ \in \Gamma(X) \) (\( y^- \in \Gamma(X) \)).

We want to maximize
\[ \hat{\delta}(y) := \max\{\delta(y^+), \overline{\delta}(y^-)\} \]
Modified Measure is Superior in High Dimension

**GLP-sets:** \( 0 < h_1 < h_2 < \cdots < h_d < n, \quad \exists j : \gcd(h_j, n) = 1 \)

\[ T := \{ t^1, \ldots, t^n \}, \quad t^i_j := \left\{ \frac{2ih_j - 1}{2n} \right\}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, d \]

Mean values of coordinates of optimal test boxes for randomly chosen GLP-sets:

\[
\begin{align*}
\text{d} = 4 & : 0.799743 \\
\text{d} = 5 & : 0.840825 \\
\text{d} = 6 & : 0.873523
\end{align*}
\]

Expectation of coordinates of randomly chosen \( y \) with respect to \( \mu_d \) is \( d/(d + 1) \):

\[
\begin{align*}
\text{d} = 4 & : 0.8 \\
\text{d} = 5 & : 0.8\overline{3} \\
\text{d} = 6 & : 0.857143
\end{align*}
\]
Further Improvement on the W&F-Algorithm

Procedures “snapping up” and “snapping down”:
Rounding $y^+$ and $y^-$ up and down to critical test boxes $y^{+,\text{sn}}$ and $y^{-,\text{sn}}$.
## Some Numerical Results

<table>
<thead>
<tr>
<th>Name</th>
<th>$d$</th>
<th>$n$</th>
<th>$\text{disc}^*(\cdot)$ found</th>
<th>TA improved</th>
<th>Winker &amp; Fang</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hits</td>
<td>Hits</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Best-of-10</td>
<td>Best-of-10</td>
</tr>
<tr>
<td>Faure 7</td>
<td>7</td>
<td>343</td>
<td>0.1298</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Faure 8</td>
<td>8</td>
<td>121</td>
<td>0.1702</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Faure 9</td>
<td>9</td>
<td>121</td>
<td>0.2121</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Faure 10</td>
<td>10</td>
<td>121</td>
<td>0.2574</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Faure 11</td>
<td>11</td>
<td>121</td>
<td>0.3010</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Faure 12</td>
<td>12</td>
<td>169</td>
<td>0.2718</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Sobol’ 50</td>
<td>50</td>
<td>2000</td>
<td>0.1030*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sobol’ 50</td>
<td>50</td>
<td>4000</td>
<td>0.0677*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Faure 50</td>
<td>50</td>
<td>2000</td>
<td>0.3112*</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Faure 50</td>
<td>50</td>
<td>4000</td>
<td>0.1979*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GLP 50</td>
<td>50</td>
<td>2000</td>
<td>0.1465*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GLP 50</td>
<td>50</td>
<td>4000</td>
<td>0.1205*</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table:** New instance comparisons. Discrepancy values marked with a star are lower bounds only (i.e., largest discrepancy found over all executions of algorithm variants). All data is computed using 100 trials of 100,000 iterations; reported is the average value of best-of-10 calls, and number of times (out of 100) that the optimum (or a value matching the largest known value) was found.
Further Numerical Results

<table>
<thead>
<tr>
<th>Class</th>
<th>$d$</th>
<th>$n$</th>
<th>$\text{disc}^*(\cdot)$</th>
<th>TA improved</th>
<th>Shah</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hits</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Best-of-10</td>
<td></td>
</tr>
<tr>
<td>Halton</td>
<td>5</td>
<td>50</td>
<td>0.1886</td>
<td>100</td>
<td>81</td>
</tr>
<tr>
<td>Halton</td>
<td>7</td>
<td>50</td>
<td>0.2678</td>
<td>100</td>
<td>22</td>
</tr>
<tr>
<td>Halton</td>
<td>7</td>
<td>100</td>
<td>0.1714</td>
<td>100</td>
<td>13</td>
</tr>
<tr>
<td>Halton</td>
<td>7</td>
<td>1000</td>
<td>0.0430</td>
<td>81</td>
<td>8(1)</td>
</tr>
<tr>
<td>Faure</td>
<td>10</td>
<td>50</td>
<td>0.4680</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td>Faure</td>
<td>10</td>
<td>100</td>
<td>0.2483</td>
<td>100</td>
<td>28</td>
</tr>
<tr>
<td>Faure</td>
<td>10</td>
<td>500</td>
<td>0.0717*</td>
<td>100</td>
<td>0(1)</td>
</tr>
</tbody>
</table>

Table: Comparison against point sets used by Shah. Reporting average value of best-of-10 calls, and number of times (out of 100) that the optimum was found; for Shah, reporting highest value found, and number of times (out of 100) this value was produced. The discrepancy value marked with a star is lower bound only (i.e., largest value found by any algorithm). Values marked (1) are recomputed using the same settings as in [Sha’10].