Constructing Optimal L_∞ Star Discrepancy Sets

François Clément[1], Carola Doerr[1], Kathrin Klamroth[2], Luís Paquete[3]

[1] Sorbonne Université, CNRS, LIP6, Paris, France
{francois.clement, carola.doerr}@lip6.fr

[2] University of Wuppertal, School of Mathematics and Natural Sciences, IMACM, Gaußstr. 20
42119 Wuppertal, Germany
klamroth@math.uni-wuppertal.de

[3] CISUC, Department of Informatics Engineering, University of Coimbra, Portugal
paquete@dei.uc.pt

Mots-clés : *Discrepancy, Non-linear programming, Computational geometry*

1 Introduction

Discrepancy measures are designed to quantify how closely a given set of points approximates the uniform distribution in a given space. Among the different discrepancy measures, one of the most important is the L_∞ star discrepancy. The L_∞ star discrepancy of a finite point set $P \subseteq [0,1)^d$ measures the worst absolute difference between the Lebesgue measure of a d-dimensional box anchored in $(0,\ldots,0)$ and the proportion of points that fall inside this box. The lower the discrepancy of a point set, the more uniformly distributed it is. Point sets and sequences of low L_∞ star discrepancy have been used in many different applications, from financial mathematics to computer vision. In particular, they are extensively used for Quasi-Monte Carlo integration [1] and design of experiments [8].

While designing asymptotically good low-discrepancy sequences has been extensively studied [6], finding optimal sets of n points in a given dimension d has long been seen as too difficult. Results are therefore sparse. White determined optimal point sets for $n \leq 6$ in dimension 2 in [9]. More recently, optimal point sets for $n = 1$ were obtained by Pillards, Cools, and Vandewoestyne in [7]. This was later extended to sets of size $n = 2$ by Larcher and Pillichshammer [4]. For a different measure, the periodic L_2 discrepancy, known to be easier to compute, Hinrichs and Oettershagen [3] obtained the optimal sets for $n \leq 16$.

Our contribution : In this paper, we provide non-linear programming formulations to compute optimal L_∞ star discrepancy point sets. Using these, we are able to compute optimal sets for $n \leq 21$ in dimension 2. Not only do these sets have a greatly lower discrepancy value, around 45% smaller on average, but they also present a very different structure to that of known sets. This is illustrated in Figure 1. We also show two theoretical results, one improving a lower bound for the discrepancy of small sets and the other showing that there is at least one optimal set whose points are in general position. These results help us refine our formulations. Though not presented in detail here, our models can easily be adapted to different discrepancy measures such as the multiple-corner discrepancy. All of our code and some figures are available at https ://github.com/frclement/OptiSets.

2 The L_∞ Star Discrepancy

The L_∞ star discrepancy of a point set represents the worst absolute difference between the volume of a box of the form $[0,q)$ and the proportion of points that fall inside this box. Formally, for a point set P of n points in $[0,1)^d$, the L_∞ star discrepancy of P, $d^*_\infty(P)$ is defined

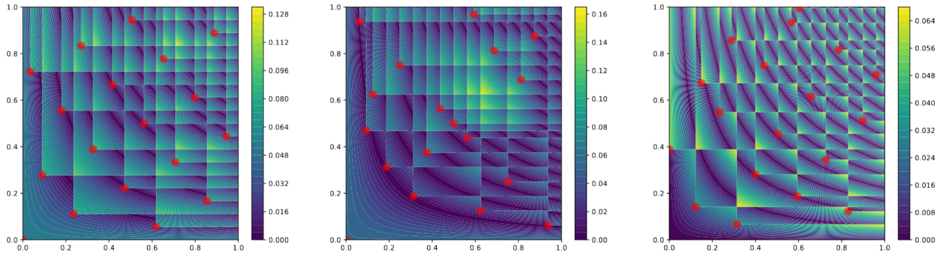

FIG. 1 – Left to right : Fibonacci, Sobol' and optimal sets' local discrepancies for $n = 18$.

by

$$d^*_\infty(P) := \sup_{q\in[0,1)^d} \left| \frac{D(q,P)}{|P|} - \lambda(q) \right|, \tag{1}$$

where $D(q,P)$ denotes the number of points of P in the box $[0,q)$ and $\lambda(q)$ the Lebesgue measure of the d-dimensional box $[0,q)$. For a given q, the discrepancy term is called *local discrepancy*. We will shorten L_∞ star discrepancy to discrepancy from now on. To define our models, we will require three more elements : the grid structure, a lower-bound on the discrepancy and a general position hypothesis for our points.

Grid structure : The key result enabling our models is that calculating the discrepancy of a point set can be treated as a discrete problem [5]. Indeed, for a point set $P = (x^{(1)}, \dots, x^{(n)}) \subseteq [0,1)^d$ the maximal value can only be obtained at a corner-point q lying on the grid $\Gamma(P) := \Gamma_1(P) \times \ldots \times \Gamma_d(P)$, where $\Gamma_1(P) := \{x_1^{(i)} : i = 1, \dots, n\} \cup \{1\}$. For any other corner-point q, we could either increase or decrease the volume of the box without changing the number of points inside and thus make the discrepancy worse. This gives us the following new equation for the discrepancy [2].

$$d^*_\infty(P) = \max\left\{ \max_{q\in\Gamma(P)} \frac{|P\cap[0,q]|}{|P|} - \lambda([0,q]),\ \max_{q\in\Gamma(P)} \lambda([0,q)) - \frac{|P\cap[0,q)|}{|P|} \right\}. \tag{2}$$

One will notice that the terms for the left maximum correspond to closed boxes while those for the right maximum correspond to open boxes. It is possible to further refine this grid and only consider *critical* boxes. For a box $[0,q)$ (or $[0,q]$) to be critical, it is necessary for there to be, for every $j \in \{1,\dots,d\}$, a point $x^{(i)} \in P$ such that $q_j = x_j^{(i)}$ and $x^{(i)} \le q$ coordinate-wise. In other terms, each of the outer edges of the critical box must have at least one point on it, or we could use the same reasoning as for non-grid corner-points.

A discrepancy lower-bound : Another crucial ingredient of our models is the following Theorem, previously proven by White for $d = 2$ [9], and that we generalize.

Theorem 1 *Let $n \ge 4$, $d \ge 2$, and let $P \subset [0,1)^d$ with $|P| = n$. Then $d^*_\infty(P) \ge 1/n$.*

Proof : Let n, d, P be as required.

— We first suppose that the points in P can be ordered such that $x^{(1)} \le x^{(2)} \le \ldots \le x^{(n)}$. For all the large open boxes reaching an outer edge of $[0,1]^d$ to have discrepancy smaller than $1/n$, each coordinate $x_j^{(i)}$ has to be smaller than i/n. Given this, in the best possible case, the smallest closed box containing $x^{(i)}$ has volume smaller than $(i/n)^d$ and contains i points. For $i = 2$ and $n \ge 4$, we then have a local discrepancy of the smallest closed box containing $x^{(i)}$ of more than $\frac{2}{n} - \frac{2^d}{n^d}$, which is at least $1/n$.

— If the points are not pairwise dominating each other, there exist x and y in P such that $x_1 > y_1$ and $x_2 < y_2$. The box $[0,q)$, where $q_1 = x_1$ and $q_2 = y_2$ and $q_j = \max(x_j, y_j)$ for $j \in \{3,\dots,d\}$, contains at least two fewer points than $[0,q]$ and has the same volume V_q. Let k be the number of points inside $[0,q)$. We then have $d^*_\infty(P) \ge \max\{|V_q - k/n|, |V_q - (k+2)/n|\} \ge 1/n$. One of the two boxes therefore has discrepancy at least $1/n$.

We have thus shown that in both cases, there exists a box with local discrepancy at least $1/n$, which concludes the proof. $\square$

General position hypothesis : To avoid problematic degenerate cases in our models, we require that no two points share a coordinate. To prove this general position hypothesis, we introduce *shifts*. While we only present *up*-shifts here, identical results for the different directions can be obtained in the same way.

Definition 1 *Consider an n-point set $P = \{x^{(1)}, \ldots, x^{(n)}\} \subseteq [0,1)^d$ and let $i \in \{1, \ldots, n\}$. The movement of a point $x^{(i)} \in P$ is called an* up-shift*, if the point $x^{(i)}$ is moved upwards in the second coordinate. An up-shift leads to a new n-point set up$(P, x^{(i)}, \delta) = P \setminus \{x^{(i)}\} \cup \{(x_1^{(i)}, x_2^{(i)} + \delta)\}$, with $0 < \delta \leq 1 - x_2^{(i)}$. An up-shift is called* admissible*, if*

$$\delta \leq \max \left\{ 0, \ \frac{1}{n} - \min_{k \neq i : x^{(k)} \leq x^{(i)}} x_2^{(i)} - x_2^{(k)} \right\}. \tag{3}$$

Theorem 2 *Let P^* be an optimal n-point set with $d_\infty^*(P^*) = f^*$. Then, an admissible up-shift never increases $d_\infty^*(P^*)$.*

Theorem 2 can be proven by considering the new grid boxes for which either the volume changes or the number of points inside changes. For all of these, either there exists a previous closed box with as many points inside and smaller volume, or there exists an *unchanged* open box whose volume is smaller by less than $1/n$ (by shift admissibility) and with one less point inside. In both cases, the new boxes have local discrepancy smaller than a previously existing box : the discrepancy of the set cannot have increased.

Not only does Theorem 2 give us a general position hypothesis, it also allows us to have lower bounds on the coordinates of each point. In particular, we can set the lowest coordinate in each dimension to f, the discrepancy of the set.

3 A Non-Linear Programming Approach

Continuous Model : Based on the above results, we can now introduce our first model to tackle this problem in two dimensions, Model (4). We represent the point set as a vector x_0 to x_{2n+1}, where $x^{(i)} = (x_{2i-1}, x_{2i})$ for $i \in \{1, \ldots, n\}$. The two extra elements, x_0 and x_{2n+1}, are dummy points so that we do not have to treat the boxes with a coordinate equal to 1 separately. These dummy points are handled in constraint (4h). Without loss of generality, we can suppose that the points are ordered according to the first coordinate. Model (4) then gives us the minimal discrepancy for a point set of n points when minimizing the objective f, and the associated variables x_i describe the point set.

$$\begin{aligned}
\min \ & f \\
\text{s.t. } & \frac{1}{n} \sum_{u=1}^{i} y_{uj} - x_{2i-1} x_{2j} \leq f + (1 - y_{ij}) && \forall i, j = 1, \ldots, n, \ j \leq i && \text{(4a)} \\
& \frac{-1}{n} \left(\sum_{u=0}^{i-1} y_{uj} - 1 \right) + x_{2i-1} x_{2j} \leq f + (1 - y_{ij}) && \forall i = 1, \ldots, n+1, \ j < i && \text{(4b)} \\
& x_{2i+1} - x_{2i-1} \geq \varepsilon && \forall i = 1, \ldots, n-1 && \text{(4c)} \\
& x_{2j} - x_{2i} \geq y_{ij} - 1 + \varepsilon && \forall i, j = 1, \ldots, n, \ i < j && \text{(4d)} \\
& x_{2j} - x_{2i} \leq y_{ij} && \forall i, j = 1, \ldots, n, \ i < j && \text{(4e)} \\
& y_{ij} = 1 - y_{ji} && \forall i, j = 1, \ldots, n, \ i > j && \text{(4f)} \\
& y_{ii} = 1 && \forall i = 1, \ldots, n && \text{(4g)} \\
& x_0 = 1; \ x_{2n+1} = 1; \ y_{0j} = 0 \, \forall j = 1, \ldots, n; \ y_{j0} = y_{(n+1)j} = 1 \, \forall j = 0, \ldots, n && && \text{(4h)} \\
& x_i \in (0, 1] \, \forall i = 1, \ldots, 2n, \ y_{ij} \in \{0, 1\} \, \forall i, j = 1, \ldots, n; \ f \geq 0.
\end{aligned}$$

As shown in (2), we need to verify that a number of terms are smaller than our global discrepancy f, the variable we would like to minimize. Each of these terms appears as a constraint of the form (4a) or (4b). For each of these, we need to count the number of points inside the box, and the volume of the box. We know that the top-right corner of these boxes is of the shape $(x_1^{(i)}, x_2^{(j)})$, therefore the volume is given by $x_{2i-1}x_j$. For the number of points, while we can track the points smaller in the first coordinate with our ordering, we require extra variables for the second coordinate. We introduce y_{ij} variables such that $y_{ij} = 1$ if and only if $x_{2i} \leq x_{2j}$. Constraints (4d) to (4g) allow for the definition of these variables. With these variables, the number of points inside the box $[(0,0),(x_{2i-1}, x_{2j})]$ is equal to $\sum_{u=1}^{i} y_{uj}$. Finally, we need to express these constraints only if the box is critical, which is equivalent to $y_{ij} = 1$ ($x^{(i)}$ and $x^{(j)}$ do not dominate each other). The last set of constraints that is required for the model is those in (4h), for the dummy points.

Using our previous results, it is possible to add extra constraints to limit the search space and, hopefully, help the solver. Constraint (4i) expresses Theorem 1. Constraints (4j) and (4k) are the most important ones obtained with Theorem 2. Using Theorem 2, we are also able to refine constraints (4c) and (4d) by taking $\varepsilon = 1/n$. Some extra constraints bounding the gaps between different coordinates can also be obtained, but are not listed here. Finally, some miscellaneous constraints such as transitivity of the y_{ij} variables could also be added.

$$f \geq \frac{1}{n} \qquad \text{valid if } n \geq 4 \tag{4i}$$

$$x_1 = f \tag{4j}$$

$$x_{2i} \geq f \qquad \forall i = 1, \ldots, n \tag{4k}$$

Assignment model : We are also able to introduce a second model, the assignment model M5, which, rather than optimizing directly the point set, optimizes the *underlying grid* and chooses a valid corresponding point set. While it does not perform better in 2 dimensions, it performs better in 3 dimensions and its main advantage is its simplicity. Indeed, we only need to define two vectors x and y which will define the possible coordinates, then pick among the a_{ij}, the intersection points on this grid. One a_{ij} must be picked on every line and column for this choice to generate the desired grid. This can be easily extended both to higher dimensions and to other discrepancy measures, such as the periodic, extreme or multiple-corner L_∞ discrepancies.

In particular, the multiple-corner discrepancy tackles one of the drawbacks of the L_∞ star discrepancy : all boxes being anchored in 0 can break the symmetry of the underlying space, which can be highly problematic for applications. To remedy this, the multiple-corner discrepancy considers the L_∞ star discrepancy from *every* corner of $[0,1)^d$ and not just 0. Given a point set $P \subseteq [0,1)^2$, the multiple-corner discrepancy in 2 dimensions is given by

$$D_\infty^{4cor}(P) := \sup_{q=(q_1,q_2)} \max\{D([0,q),P), D((q,1],P), D([0,q_1)\times(q_2,1],P), D((q_1,1]\times[0,q_2),P)\}. \tag{5}$$

For a given set P, this can be reformulated by considering the sets $P_2 := \{(y, 1-x) : (x,y) \in P\}$, $P_3 := \{(1-x, 1-y) : (x,y) \in P\}$ and $P_4 := \{(1-y, x) : (x,y) \in P\}$. We then have

$$D_\infty^{4cor}(P) = \max(d_\infty^*(P), d_\infty^*(P_2), d_\infty^*(P_3), d_\infty^*(P_4)). \tag{6}$$

With the assignment model, this problem is nearly the same as the initial star discrepancy one, the main change being 4 versions of the box constraints (the equivalents of constraints (4a) and (4b)).

4 Experimental results

We now present the results obtained by our models in dimension 2 for $n = 7$ to $n = 21$ (values for $n \leq 6$ were known previously in [9]). All the experiments were run with Gurobi

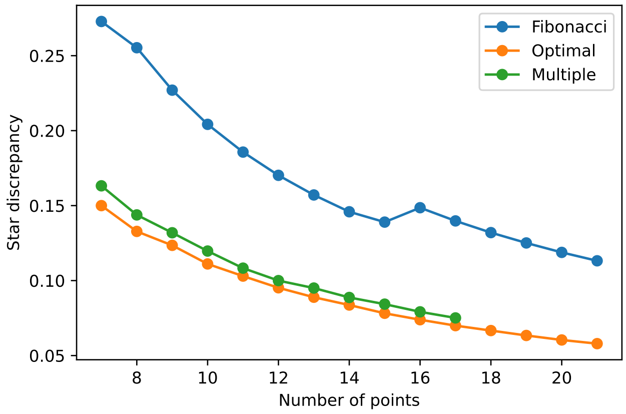

FIG. 2 – Star discrepancy of our optimal sets, compared with the Fibonacci set (the main 2 dimensional reference set in the community). The “multiple” line corresponds to *both* the regular L_∞ star discrepancy and the multiple-corner discrepancy of our optimal multiple-corner set.

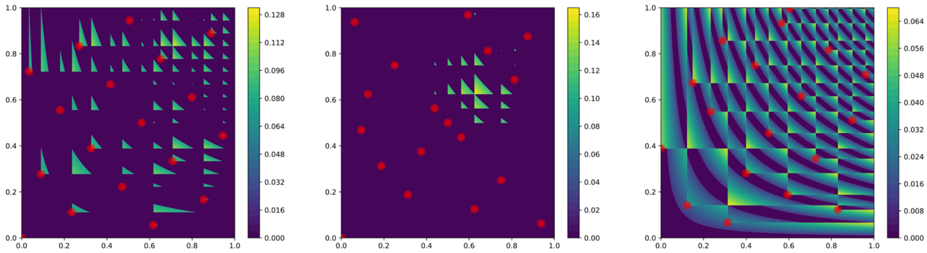

FIG. 3 – Left to right : Fibonacci, Sobol’ and optimal sets’ truncated local discrepancies for $n = 18$.

10.0.0 with an accuracy of 10^{-4}, using the Julia JuMP package. Experiments were run on a single machine of the MeSU cluster at Sorbonne Université, Intel Xeon CPU E5-2670 v3 with 24 cores. Figure 2 shows that our models provide point sets vastly outperforming the current reference for low-discrepancy sets, the Fibonacci set, with a nearly 50% improvement on average. The 21-point optimal set takes around 20 hours to compute.

We plot the *truncated* local discrepancy values over $[0,1]^2$ in Figures 3, with brighter colors indicating a worse discrepancy. “Triangles” whose corner is to the top-right correspond to open boxes with too few points whereas those with a corner to the lower-left correspond to closed boxes with too many points. Every local discrepancy value more than $1/n$ away from the maximal value has been set to 0, to see better where the worst discrepancy values are reached.

While in sets like Fibonacci or parts of the Sobol’ sequence only a few closed boxes give active constraints for the discrepancy, a much bigger set of boxes are very close to the maximal discrepancy value for our sets. For both Fibonacci and Sobol’ sets, only overfilled boxes appear, and this seems to be independent of n. However, for our sets, both types of triangles appear and quite often sharing a box corner.

The very large structural differences observed in these figures suggest that there remains a big improvement margin for low-discrepancy point sets.

Finally, we plot our optimal set for the multiple-corner discrepancy for $n = 16$ in Figure 4. Very interestingly, the *star* discrepancy value of the optimal *multiple-corner* point set is very close to that of our optimal set, as was observed in Figure 2. Furthermore, the multiple-corner discrepancy of our optimal multiple-corner set is very close to the *star* discrepancy value. This suggests that optimizing for the multiple-corner discrepancy could provide substantial improvements for applications, with a very small trade-off for the star discrepancy.

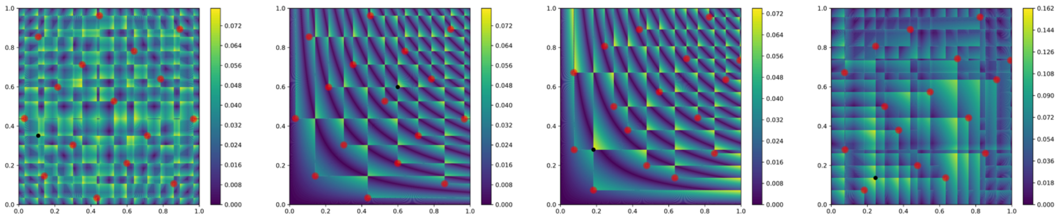

FIG. 4 – N=16. Left to right : local multiple-corner discrepancy of the optimal multiple-corner set, local star discrepancy of the optimal multiple-corner set, local star discrepancy of the optimal star discrepancy set and local multiple-corner discrepancies of the optimal star set.

5 Conclusion

We presented in this short paper models to compute optimal L_∞ star discrepancy sets, an open problem that has stumped the discrepancy community for over 50 years. These models are able to compute optimal sets for $n \leq 21$ in dimension 2 but are highly flexible : we have also done extensions to dimension 3, other discrepancy measures or simply adding a single point to an existing set. Perhaps more importantly than the discrepancy values, there is a clear structural difference between previously known sets and our optimal sets. Exploiting the inspiration from our optimal configurations is a natural next step. We hope this might provide insight into new construction methods, with the hope of tackling the great open question in discrepancy theory : what is the optimal asymptotic discrepancy order in dimensions higher than 2 ?

Références

[1] J. Dick and F. Pillichshammer. *Digital Nets and Sequences.* Cambridge University Press, Cambridge, 2010.

[2] C. Doerr, M. Gnewuch, and M. Wahlström. Calculation of discrepancy measures and applications in : W. Chen, A. Srivastav, G. Travaglini (eds.). *A Panorama of Discrepancy Theory, Springer*, pages 621–678, 2014.

[3] A. Hinrichs and J. Oettershagen. Optimal point sets for Quasi-Monte Carlo integration of bivariate periodic functions with bounded mixed derivatives. In *Monte Carlo and Quasi-Monte Carlo Methods*, 2014.

[4] G. Larcher and F. Pillichshammer. A note on optimal point distributions in $[0, 1)^s$. *Journal of Computational and Applied Mathematics*, 206(2) :977–985, 2007.

[5] H. Niederreiter. Discrepancy and convex programming. *Ann. Mat. Pura Appl.*, 93 :89–97, 1972.

[6] H. Niederreiter. Random number generation and Quasi-Monte Carlo methods. *SIAM CBMS NSF Regional Conf. Series in Applied Mathematics (SIAM, Philadelphia)*, 1992.

[7] T. Pillards, B. Vandewoestyne, and R. Cools. Minimizing the L_2 and L_∞ star discrepancies of a single point in the unit hypercube. *Journal of Computational and Applied Mathematics*, 197 :282–285, 12 2006.

[8] T.J. Santner, B.J. Williams, and W.I. Notz. *The Design and Analysis of Computer Experiments.* Springer Series in Statistics, Springer, 2003.

[9] B.E. White. On optimal extreme-discrepancy point sets in the square. *Numer. Math.*, 27 :157–164, 1976/1977.