

Sampling in CMA-ES: Low Numbers of Low Discrepancy Points

Jacob de Nobel^a, Diederick Vermetten^b, Thomas H. W. Bäck^c and Anna V. Kononova^d

LIACS, Leiden University, Leiden, Netherlands

{j.p.de.nobel, d.l.vermetten, t.h.w.baeck, a.kononova}@liacs.leidenuniv.nl


Keywords: Benchmarking, CMA-ES, Derandomization, Low Discrepancy Samples.


Abstract: The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is one of the most successful examples of a derandomized evolution strategy. However, it still relies on randomly sampling offspring, which can be done via a uniform distribution and subsequently transforming into the required Gaussian. Previous work has shown that replacing this uniform sampling with a low-discrepancy sampler, such as Halton or Sobol sequences, can improve performance over a wide set of problems. We show that iterating through small, fixed sets of low-discrepancy points can still perform better than the default uniform distribution. Moreover, using only 128 points throughout the search is sufficient to closely approximate the empirical performance of using the complete pseudorandom sequence up to dimensionality 40 on the BBOB benchmark. For lower dimensionalities (below 10), we find that using as little as 32 unique low discrepancy points performs similar or better than uniform sampling. In 2D, for which we have highly optimized low discrepancy samples available, we demonstrate that using these points yields the highest empirical performance and requires only 16 samples to improve over uniform sampling. Overall, we establish a clear relation between the L_2 discrepancy of the used point set and the empirical performance of the CMA-ES.


1 INTRODUCTION


Optimization techniques play a crucial role in various scientific and engineering applications. Exact methods systematically explore the parameter space but often suffer from inefficiency due to their exhaustive nature. For example, it has been shown that randomized search is superior to grid search for hyperparameter tuning (Bergstra and Bengio, 2012). This is because, especially in higher dimensions, a randomized process will provide better coverage of sample points in the domain than an exhaustive search, given a limited evaluation budget. While samples generated uniformly at random provide an improvement in exploring the search space, such samples can still be quite suboptimal in covering the domain. Given a limited number of samples, uniform samples can be distributed very unevenly (Halton, 1960). This notion of evenly spreading points across a given domain motivates the research into *Low-Discrepancy Sequences*. These are sequences of pseudo-randomly generated points that are designed to minimize the

gaps and clusters that often occur in uniform random sampling, providing more uniform coverage of the search space. Specifically, discrepancy measures are designed to measure how regularly a given point set is distributed in a given space (Clément et al., 2023b). While early work with low discrepancy point sets focuses on Monte Carlo integration (Halton, 1960; Sobol', 1967), they have subsequently been used in various domains, such as computer vision (Paulin et al., 2022) and financial modeling (Galanti and Jung, 1997). Low Discrepancy point sets have been used in the optimization domain to set up the Design of Experiments (DoE) within a constrained budget (Santner et al., 2003). One application of particular interest in our context is one-shot optimization, where low-discrepancy sequences have been shown to outperform more traditional uniform sampling (Bousquet et al., 2017). Moreover, random search using quasi-random points has been shown to outperform traditional random search (Niederreiter, 1992). In metaheuristics, randomized search is employed by many different algorithms, such as Evolution Strategies (ES) (Beyer, 2001). In ES, quasi-random point sets have been used as an alternative sampling strategy to pure random sampling (Teytaud and Gelly, 2007), specifically for the CMA-ES (Hansen and Os-

^a  <https://orcid.org/0000-0003-1169-1962>

^b  <https://orcid.org/0000-0003-3040-7162>

^c  <https://orcid.org/0000-0001-6768-1478>

^d  <https://orcid.org/0000-0002-4138-7024>

termeier, 2001). Simplified, the procedure of the CMA-ES can be divided into two steps, which are repeated until convergence:

- Sample λ points from a multivariate normal distribution $\mathcal{N}(\mathbf{m}, \mathbf{C})$.
- Adjust the parameters of the multivariate normal distribution to move towards the μ points with the highest fitness.

Modifying the sampling step to use points from a pseudo-random sequence (Teytaud, 2015) demonstrated increased performance and stability on benchmark functions. Moreover, this furthers *derandomization*, which aims to achieve self-adaptation without any independent stochastic variation of the strategy parameters (Bäck et al., 2023). This paper aims to extend this work by focusing on the number of samples drawn from a (pseudo) random sampling strategy specifically for the CMA-ES. Specifically, we investigate if repeatedly reusing small subsets of pseudo-random sequences in a deterministic manner can be an effective sampling strategy for the CMA-ES. We provide an analysis for several well-known low discrepancy point sets to investigate the relation between discrepancy and empirical performance on the BBOB benchmark functions.

2 PRELIMINARIES

2.1 Low-Discrepancy Sequences

The discrepancy of a set of points quantifies how regularly they are spaced in the domain. One of the most common discrepancy measures of a point set $P \subseteq [0, 1]^d$ is the L_∞ star discrepancy (Clément et al., 2023a), which is defined as follows:

$$d_\infty^*(P) = \sup_{q \in [0, 1]^d} \left| \frac{|P \cap [0, q]|}{|P|} - \lambda(q) \right| \quad (1)$$

Here, $\lambda(q)$ is the Lebesgue measure of the box $[0, q]$ and $d_\infty^*(P)$ measures the worst absolute difference between $\lambda(q)$ of a d -dimensional box anchored at the origin and the proportion of points that fall inside this box. Note that this measure should be minimized to evenly space points in the domain. Since the L_∞ star discrepancy can be computationally expensive (Clément et al., 2023b), we can also consider the L_2 star discrepancy, which is defined as follows (Zhou et al., 2013):

$$d_2^*(P) = \left(\int_{[0, 1]^d} \left| \frac{|P \cap [0, q]|}{|P|} - \lambda(q) \right|^2 dq \right)^{1/2} \quad (2)$$

Several pseudo-random sequences have lower star discrepancies than corresponding uniform sequences. These methods include Latin Hypercube Sampling (Loh, 1996), Jittered sampling (Pausinger and Steinerberger, 2016), and Hammersly sequences (Peart, 1982). Our work considers the Halton (Halton, 1960) and Sobol (Sobol', 1967) sequences.

2.2 Derandomization and CMA-ES

While Evolution Strategies (Beyer, 2001) depend on a random process to sample candidate solutions, the internal parameter update has been *derandomized* in state-of-the-art implementations of the algorithm. Derandomization ensures self-adaptation happens without any independent stochastic variation of the strategy parameters (Ostermeier et al., 1994). Effectively, this means the update of the strategy parameters is decoupled from the sampling of candidate solutions, moving away from the notion that ‘good solutions have good parameters’ of traditional ES. This allows modern ES, such as the CMA-ES, to be more robust and learn good strategy parameters while using relatively small population sizes (Hansen and Ostermeier, 2001). Within the CMA-ES, the sampling procedure is the only remaining source of stochasticity. At every generation, λ individuals are sampled from the d -dimensional Gaussian distribution $\mathcal{N}(\mathbf{m}, \sigma^2 \mathbf{C})$. Where \mathbf{m} is the mean of the sampling distribution, σ is the global step size, and \mathbf{C} is the covariance matrix. In practice, \mathbf{C} is spectrally decomposed into two matrices, \mathbf{B} and \mathbf{D} , representing the eigenvectors and inverse square root of the eigenvalues of \mathbf{C} , respectively. This allows the sampling of points in the CMA-ES to happen in a three-stage process:

1. $\mathbf{z}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2. $\mathbf{y}_k = \mathbf{B}\mathbf{D}\mathbf{z}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$
3. $\mathbf{x}_k = \mathbf{m} + \sigma\mathbf{y}_k \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$

Given this decomposition, the first step of the sampling process can be practically achieved by sampling from a uniform distribution $\mathbf{u} \sim \mathcal{U}(0, 1)^d$, and transforming each coordinate \mathbf{u}_i of the sample by:

$$\phi^{-1}(\mathbf{u}_i) = \sqrt{2}\text{erf}^{-1}(2\mathbf{u}_i - 1) \quad (3)$$

which is the inverse of the cumulative density function for a standard Gaussian distribution.

Since the sampling procedure in CMA-ES can be seen as sampling in $[0, 1]^d$, replacing the uniform sampling with a low-discrepancy sequence is a natural step. Previous work has shown that scrambled Halton sequences can improve performance over the standard sampling procedure on most considered benchmark problems (Teytaud and Gelly, 2007; Teytaud, 2015).

UNIFORM-16	-1.05	-1.40	-2.14	-3.75	-7.12
SOBOL-16	-1.40	-1.56	-2.35	-4.16	-7.62
HALTON-16	-1.39	-1.50	-2.16	-3.78	-7.21
OPT-16	-1.56	-1.54	-2.38	-4.26	-7.22
UNIFORM-32	-1.20	-1.55	-2.28	-3.84	-7.16
SOBOL-32	-1.67	-1.73	-2.40	-4.13	-6.99
HALTON-32	-1.64	-1.69	-2.30	-3.88	-7.26
OPT-32	-1.85	-1.75	-2.43	-4.15	-7.93
UNIFORM-64	-1.36	-1.70	-2.42	-3.97	-7.27
SOBOL-64	-1.93	-1.92	-2.52	-3.99	-7.17
HALTON-64	-1.90	-1.89	-2.45	-3.99	-7.34
OPT-64	-2.16	-1.77	-2.53	-4.26	-7.61
UNIFORM-128	-1.50	-1.85	-2.57	-4.10	-7.35
SOBOL-128	-2.18	-2.15	-2.68	-4.16	-7.44
HALTON-128	-2.15	-2.10	-2.61	-4.12	-7.40
OPT-128	-2.42	-2.09	-2.68	-4.16	-7.50
UNIFORM-256	-1.65	-2.00	-2.72	-4.23	-7.43
SOBOL-256	-2.48	-2.36	-2.83	-4.31	-7.58
HALTON-256	-2.42	-2.31	-2.78	-4.25	-7.45
OPT-256	-2.70	-2.31	-2.83	-4.36	-7.76
	2	5	10	20	40
	Dimensionality				

Figure 1: (Average) $\log_{10}(d_2^*)$ star discrepancy for the generated fixed-size point sets across all dimensionalities. Colors are (min-max) normalized on a per-dimensionality basis; darker colors indicate a worse (higher) d_2^* value.

3 METHODS

3.1 Point Set Generation

For our experiments, we modify how the CMA-ES samples from a normal distribution by exchanging the uniform number generator with a selection of points from a fixed point set stored in a cache. The cache is randomly permuted once, at the beginning of an optimization run, and cycled through in steps of size λ . The permutation is performed to break the bias that might be present in the ordering of the point set.

We consider four methods for generating point sets: Sobol, Halton, uniform, and ‘optimized’. Halton sequences are known to have unwanted correlations in higher dimensional spaces, and as such, we employ a scrambling method to prevent this (Braaten and Weller, 1979). For the Sobol sequences, its balance properties require that the number of points generated is equal to a power of 2, so we always round up our number of points sampled to the closest power of 2. Finally, our ‘optimized’ method for generating low-discrepancy point sets is split into two parts based on search space dimensionality. When $d = 2$, we use optimized Fibonacci sets (Clément et al., 2023a), considered among the best low-discrepancy point sets available. However, these are only available for 2D since generating optimized low-discrepancy

UNIFORM- ∞	0.73	0.58	0.54	0.51	0.48
SOBOL- ∞	0.77	0.59	0.54	0.51	0.48
HALTON- ∞	0.79	0.60	0.56	0.52	0.49
UNIFORM-16	0.55	0.37	0.33	0.25	0.18
SOBOL-16	0.71	0.53	0.50	0.34	0.25
HALTON-16	0.72	0.52	0.43	0.31	0.20
OPT-16	0.74	0.54	0.41	0.31	0.18
UNIFORM-32	0.66	0.50	0.44	0.37	0.28
SOBOL-32	0.75	0.58	0.54	0.48	0.36
HALTON-32	0.75	0.57	0.52	0.44	0.33
OPT-32	0.75	0.57	0.47	0.46	0.35
UNIFORM-64	0.71	0.56	0.51	0.46	0.39
SOBOL-64	0.76	0.59	0.55	0.51	0.43
HALTON-64	0.76	0.59	0.54	0.49	0.43
OPT-64	0.76	0.57	0.53	0.48	0.39
UNIFORM-128	0.73	0.57	0.52	0.48	0.44
SOBOL-128	0.77	0.60	0.55	0.50	0.48
HALTON-128	0.77	0.59	0.55	0.51	0.47
OPT-128	0.75	0.58	0.54	0.50	0.45
UNIFORM-256	0.74	0.58	0.53	0.50	0.46
SOBOL-256	0.78	0.59	0.55	0.50	0.48
HALTON-256	0.77	0.60	0.55	0.51	0.48
OPT-256	0.74	0.59	0.54	0.50	0.46
	2	5	10	20	40
	Dimensionality				

Figure 2: Average area under the EAF curve for each sampling method on the BBOB benchmark, grouped by dimension. Colors are (min-max) normalized on a per-dimensionality basis; darker colors indicate a worse (lower) EAF value.

point sets is a very hard computational problem. We use the improved Threshold Accepting subset selection heuristic from (Clément et al., 2024) for higher dimensionalities, using Sobol sequences as the base sets.

We create point sets of sizes $k \in \{16, 32, 64, 128, 256\}$ for each generation mechanism and calculate their L_2 star discrepancy¹². Figure 1 shows how the discrepancy changes with increasing dimensionality and size of the point sets. We observe that, as expected, the standard uniform sampling has a noticeably higher discrepancy than both Halton and Sobol sequences, with the optimized method usually having the lowest discrepancy. Note that for the optimized method, the relative difference in discrepancy is largest when $d = 2$ due to the different generation mechanism used for this dimensionality.

3.2 Experimental Setup

To gauge the impact of our derandomization, we run the Modular CMA-ES (de Nobel et al., 2021) with each of the generated point sets on the single-

¹We use L_2 instead of L_∞ star discrepancy for computational reasons.

²For the uniform points sets, we calculate the average over 100 samples.

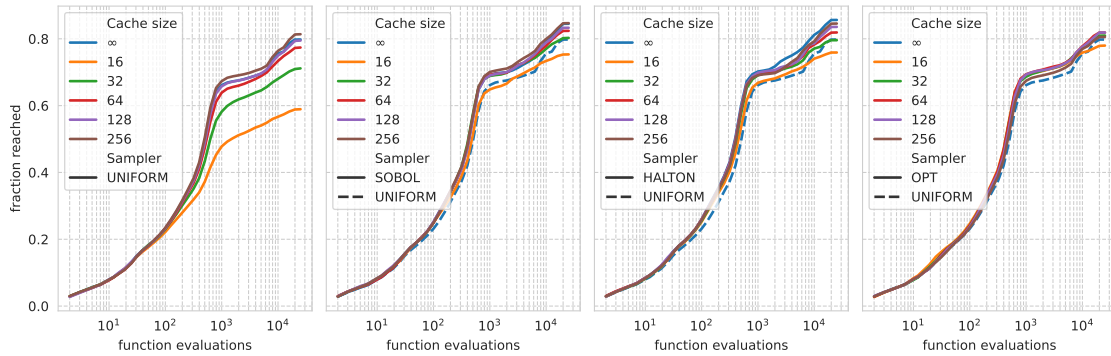


Figure 3: Empirical Attainment Function aggregated over all 24 BBOB functions for dimensionality 2. The methods are shown in color for each sampling strategy, with cache size indicating the number of points in the cache and ∞ indicating no caching; every sample is unique. From left to right, the subfigures show results using a uniform sampler, a Sobol sequence, a scrambled Halton sequence, and the optimized point sets. Note that in the three rightmost figures, the default CMA-ES sampling strategy, UNIFORM- ∞ , is included for comparison.

objective, noiseless BBOB suite (Hansen et al., 2009). This suite consists of 24 functions, of which we use 100 different instances (generated by transformations such as rotation and translation) in dimensionalities $d \in \{2, 5, 10, 20, 40\}$. We set our evaluation budget at $B = d \cdot 10000$. By using only a single run on each instance, we only need to generate a single point set for each (d, N) -pair, while the fact that we use 100 instances ensures a large enough sample size for each function. Additionally, we benchmark the default sampling mechanism, UNIFORM- ∞ , and sampling from arbitrarily long scrambled Halton and Sobol sequences, denoted by the ∞ suffix in our results. No caching is applied in these cases, and each generated sample in the optimization process is unique.

To measure performance, we consider the attainment-based cumulative distribution function (López-Ibáñez et al., 2024), with bounds 10^2 and 10^{-8} for the log-scaled precision values.

Reproducibility

All used point sets, the full experimental code, and data are in our reproducibility repository (de Nobel et al., 2024).

4 RESULTS

Figure 2 shows each sampling method’s normalized area under the EAF curve. Without caching, we observe that both low-discrepancy sequences outperform the uniform sampler, and overall, the scrambled Halton shows the highest empirical performance. Note that the differences tend to decrease with dimensionality. When comparing the results within each fixed point set size, the uniform sampler often per-

forms notably worse than any of the low-discrepancy sets. This effect is especially noticeable for smaller set sizes (lower values of k) but remains observable for all cache sizes. Interestingly, the optimized point sets only seem to bring performance benefits in low dimensionalities and using smaller cache sizes. For higher dimensionalities, they often perform similarly to uniform sampling. Overall, increasing the size of the cached point set increases performance, and using the complete sequence (∞) yields the highest performance for each method. Interestingly, point sets of only 64 to 128 samples are often enough to get very close in performance to using the complete sequence as a sampling strategy. Moreover, we find that such sequences still outperform the default sampling strategy of the CMA-ES, i.e., UNIFORM- ∞ , especially in lower dimensions. In 2D, we can observe that only using 16 optimized samples can outperform the default sampling strategy.

When looking at the performance of the algorithm variants over time, as visualized in Figure 3, we notice that the differences between the different sampling methods occur rather early in the search. While the used set does not have a noticeable impact on the initialization, soon after, the small sets with low discrepancy are seen to stagnate, resulting in the much worse anytime performance observed in Figure 2. By comparing the different sampler to the dashed line of the default uniform sequence, we note that all sets of size 32 and higher outperform this baseline at the end of the optimization. Additionally, we should note that the optimized sets perform best among the smallest sample sizes, but the Halton and Sobol sequences benefit more from increasing their size. Plots for other dimensionalities are available in the appendix. Finally, we show the relation between discrepancy and performance in Figure 4, where we observe a clear

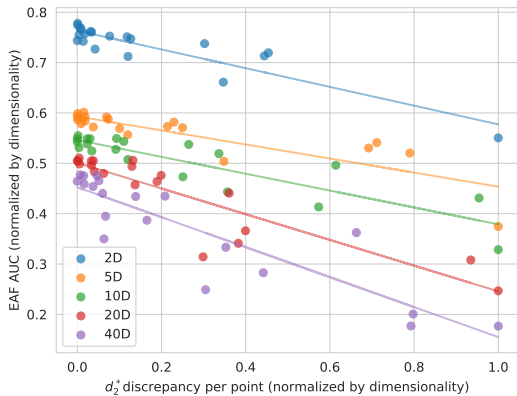


Figure 4: Average area under the Empirical Attainment Function over all BBOB functions, grouped by dimensionality vs. the L_2 star discrepancy, normalized by dimensionality and the number of points. Lines indicate a linear (least-squares) model for each dimensionality.

correlation between these two measures. This supports our previous observation that low discrepancy point sets are beneficial to the performance of the CMA-ES.

4.1 Multiples of λ

Since the CMA-ES only requires λ points at every iteration, a natural assumption would be that only λ unique points would be required for an effective sampling strategy, given that these points cover the domain sufficiently. In our previous experiments, we have used the default setting for λ , i.e., $4 + \lfloor 3 + \ln(d) \rfloor$, which is strictly less than 16, the smallest tested cache size k . This causes the sampler to cycle through the point set, causing variance between the samples used in each generation. Here, we investigate this effect using either a population size λ of 15 or 16. In the latter case, λ is a multiple of the cache size, and for $k = 16$, this ensures that every generation uses exactly the same samples. For $k < 16$, this causes a cyclic pattern, where every $\frac{k}{16}$ -th generation has the same samples. In figure 5, the empirical performance of this experiment is visualized. The figure shows that when using the default sampling strategy, i.e., UNIFORM- ∞ , using a $\lambda = 16$ is better than using $\lambda = 15$. This is reversed for a cache size $k = 16$, where there is no variance between the samples in subsequent generations. For larger k , this performance of $\lambda = 15$ becomes closer to $\lambda = 16$. This indicates there must be variation between generations, not only between the samples within each generation, which aligns with (Teytaud and Gelly, 2007). Intuitively, this makes sense, as the strategy parameters are being averaged over multiple generations, making diver-

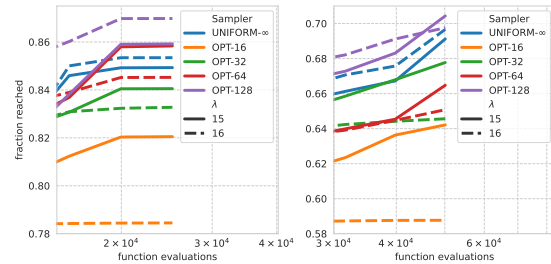


Figure 5: Empirical Attainment Function aggregated over all 24 BBOB functions for dimensionality 2 (left) and 5 (right), zoomed to final fraction reached. The default sampling strategy of the CMA-ES, UNIFORM- ∞ , is shown in comparison to using a cached sampling strategy, which uses the ‘OPT’ samples for a cache size $k \in \{16, 32, 64, 128\}$. The solid lines represent $\lambda = 15$ and the dashed lines $\lambda = 16$.

sity between subsequent populations important. Notably, though, using only 4-8 unique populations during the entire optimization procedure already yields higher performance than the UNIFORM- ∞ sampling strategy (for $d = 2$ and $d = 5$).

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have shown that replacing the uniform sampling with low discrepancy samples in the offspring generation of the CMA-ES can yield clear benefits in performance on commonly used benchmarks. While this matches previous observations on the advantage of low-discrepancy sequences (Teytaud, 2015), our experiments on repeatedly (re-)using small sets of points show that we don’t necessarily need to rely on generators when sampling with the CMA-ES. While we note that, in general, sampling from an arbitrarily long low discrepancy sequence is better than using a fixed point set, such sets remain competitive when using only 128 points. In fact, on the two-dimensional problems, where our L_∞ -optimized point sets can be considered state-of-the-art (Clément et al., 2023a), a cached set consisting of only 16 samples is enough to outperform the default sampling strategy of the CMA-ES.

However, our results in higher dimensionalities show that while there is a clear correlation between the discrepancy of a point set and the performance of the CMA-ES using it, this correlation is not perfect. While our optimized sets generated by improved Threshold Acceptance (Clément et al., 2024) have lower discrepancy than the corresponding Halton and Sobol sets, their anytime performance is slightly worse, indicating that there might be other as-

pects of the point sets we should take into account.

We have shown that not only does the discrepancy of samples within a single generation impact performance, but the diversity between subsequent generations also has an impact. Future work might focus on diving deeper into the relationship between point set size, discrepancy, diversity, and performance.

ACKNOWLEDGEMENTS

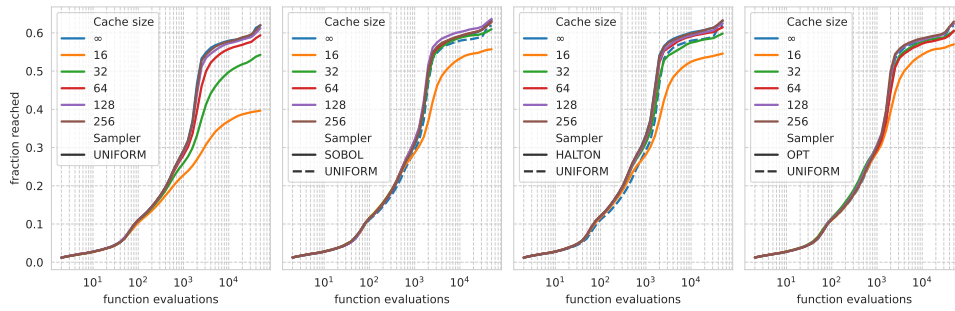
We want to thank François Clément and Carola Doerr for providing us with the optimized low-discrepancy point sets used in this paper.

REFERENCES

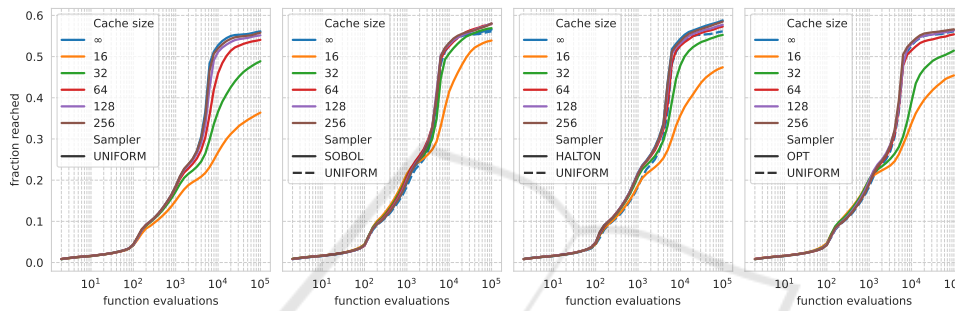
- Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2).
- Beyer, H. (2001). *The theory of evolution strategies*. Natural computing series. Springer.
- Bousquet, O., Gelly, S., Kurach, K., Teytaud, O., and Vincent, D. (2017). Critical hyper-parameters: No random, no cry. *arXiv preprint arXiv:1706.03200*.
- Braaten, E. and Weller, G. (1979). An improved low-discrepancy sequence for multidimensional quasi-monte carlo integration. *Journal of Computational Physics*, 33(2):249–258.
- Bäck, T. H. W., Kononova, A. V., van Stein, B., Wang, H., Antonov, K. A., Kalkreuth, R. T., de Nobel, J., Vermetten, D., de Winter, R., and Ye, F. (2023). Evolutionary Algorithms for Parameter Optimization—Thirty Years Later. *Evolutionary Computation*, 31(2):81–122.
- Clément, F., Doerr, C., Klamroth, K., and Paquete, L. (2023a). Constructing optimal l_∞ star discrepancy sets. *arXiv preprint arXiv:2311.17463*.
- Clément, F., Doerr, C., and Paquete, L. (2024). Heuristic approaches to obtain low-discrepancy point sets via subset selection. *Journal of Complexity*, 83:101852.
- Clément, F., Vermetten, D., De Nobel, J., Jesus, A. D., Paquete, L., and Doerr, C. (2023b). Computing star discrepancies with numerical black-box optimization algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1330–1338.
- de Nobel, J., Vermetten, D., Kononova, A. V., and Bäck, T. (2024). Reproducibility files and additional figures.
- de Nobel, J., Vermetten, D., Wang, H., Doerr, C., and Bäck, T. (2021). Tuning as a means of assessing the benefits of new ideas in interplay with existing algorithmic modules. In Krawiec, K., editor, *GECCO '21: Genetic and Evolutionary Computation Conference, Companion Volume, Lille, France, July 10-14, 2021*, pages 1375–1384. ACM.
- Galanti, S. and Jung, A. (1997). Low-discrepancy sequences: Monte carlo simulation of option prices. *The Journal of Derivatives*, 5(1):63–83.
- Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2:84–90.
- Hansen, N., Finck, S., Ros, R., and Auger, A. (2009). Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Research Report RR-6829, INRIA.
- Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195.
- Loh, W.-L. (1996). On latin hypercube sampling. *The annals of statistics*, 24(5):2058–2080.
- López-Ibáñez, M., Vermetten, D., Dréo, J., and Doerr, C. (2024). Using the empirical attainment function for analyzing single-objective black-box optimization algorithms. *CoRR*, abs/2404.02031.
- Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*. SIAM.
- Ostermeier, A., Gawelczyk, A., and Hansen, N. (1994). A derandomized approach to self-adaptation of evolution strategies. *Evolutionary Computation*, 2(4):369–380.
- Paulin, L., Bonneel, N., Coeurjoly, D., Iehl, J.-C., Keller, A., and Ostromoukhov, V. (2022). MatBuilder: Mastering sampling uniformity over projections. *ACM Transactions on Graphics (proceedings of SIGGRAPH)*.
- Pausinger, F. and Steinerberger, S. (2016). On the discrepancy of jittered sampling. *Journal of Complexity*, 33:199–216.
- Peart, P. (1982). The dispersion of the hammersley sequence in the unit square. *Monatshefte für Mathematik*, 94(3):249–261.
- Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer Series in Statistics, Springer.
- Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802.
- Teytaud, O. (2015). Quasi-random numbers improve the CMA-ES on the BBOB testbed. In Bonnevoy, S., Legrand, P., Monmarché, N., Lutton, E., and Schoenauer, M., editors, *Artificial Evolution - 12th International Conference, Evolution Artificielle, EA 2015, Lyon, France, October 26-28, 2015. Revised Selected Papers*, volume 9554 of *Lecture Notes in Computer Science*, pages 58–70. Springer.
- Teytaud, O. and Gelly, S. (2007). DCMA: yet another derandomization in covariance-matrix-adaptation. In Lipson, H., editor, *Genetic and Evolutionary Computation Conference, GECCO 2007, Proceedings, London, England, UK, July 7-11, 2007*, pages 955–963. ACM.
- Zhou, Y.-D., Fang, K.-T., and Ning, J.-H. (2013). Mixture discrepancy for quasi-random point sets. *Journal of Complexity*, 29(3-4):283–301.

APPENDIX

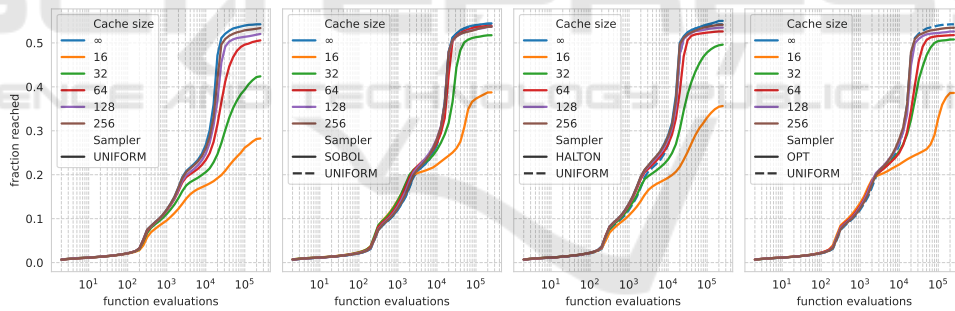
(a) Dimensionality 5



(b) Dimensionality 10



(c) Dimensionality 20



(d) Dimensionality 40

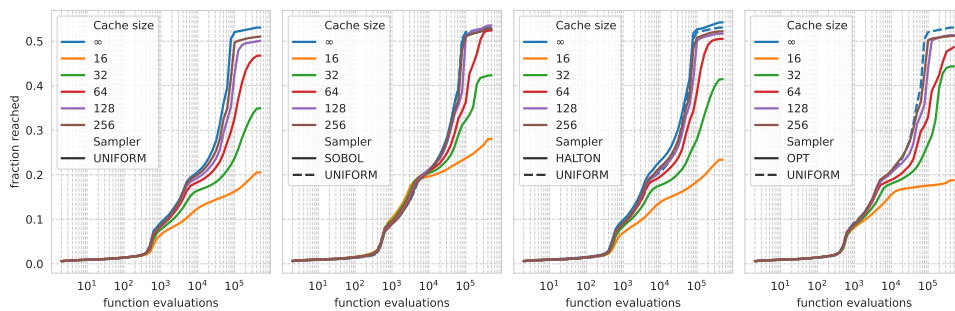


Figure 6: Empirical Attainment Function aggregated over all 24 BBOB functions for dimensionalities 5, 10, 20, and 40 (from top to bottom). The methods are shown in color for each sampling strategy, with cache size indicating the number of points in the cache and ∞ indicating no caching; every sample is unique. From left to right, the subfigures show results using a uniform sampler, a Sobol sequence, a scrambled Halton sequence, and the optimized point sets. Note that in the three rightmost figures, the default CMA-ES sampling strategy, UNIFORM- ∞ , is included for comparison.