

Recycling primes in Halton sequences: an optimization perspective

C. J. Price¹ and C. P. Price²,
Department of Mathematics and Statistics,
University of Canterbury,
Private Bag 4800,
Christchurch, New Zealand.

Abstract

An extension of the popular Halton sequence of quasi-random numbers is described. This extension allows the use of the same base for more than one dimension, and thus falls into the large class of sequences known as Niederreiter sequences. The property of distinct points being distinct in every element is retained, as is the asymptotic uniformity property. The re-use of bases (typically the smallest primes) means that the asymptotic uniformity property becomes meaningful sooner than with the standard Halton sequence. A new performance measure based on the minimal spanning tree is introduced, and its relevance to global optimization shown. The standard and prime recycling Halton sequences are compared using this performance measure, both with and without scrambling. These sequences are also compared using the distance from the closest of the first N members to randomly chosen points.

keywords: quasi random, scrambled Halton sequence, Niederreiter, optimization.

1 Introduction

Quasi-random numbers have been a topic of lively investigation for many years. Quasi-random numbers differ from pseudo-random numbers in important ways. Both are often used as substitutes for random numbers. However the former have properties that make them easily distinguishable from random numbers, whereas the latter are designed to mimic random numbers in every possible way. These distinguishing properties are often advantageous. For example quasi-random numbers are typically far more uniformly distributed than random numbers; a property of great value in search algorithms. Quasi random numbers have a variety of uses, including in numerical integration [17, 20], simulation [3], and optimization [1, 2, 24, 27, 32].

A number of types of quasi-random numbers exist including Halton [7], Sobol [27], and Faure sequences [6]. Such sequences can be generalized, leading to large classes of quasi-random sequences [18, 21, 22, 31]. The quality of these sequences is often measured using star discrepancy, which is especially relevant to numerical integration as the Koksma-Hlawka [8, 11] inequality relates the star discrepancy to a bound on the integration error. A related quantity,

¹email: chrisj.price@canterbury.ac.nz

²deceased

the L_2 star discrepancy is also popular due to its relative ease of calculation in comparison to the star discrepancy [16]. These measures are less relevant to optimization, and so a different measure based on minimal spanning trees is introduced.

Herein we use an archetypal optimization problem of the form

$$\min f(x) \quad \text{over } x \in [0, 1]^n \tag{1}$$

where the objective function f is continuous. Many other global optimization problems over bounded regions can be placed in this form by use of penalty or barrier functions, or by other means. Additionally many other conditions than continuity may apply to f . For clarity this work is written with (1) in mind, but it also applies to many other optimization problems. Points in a quasi random sequence can be used as sample points to explore the global features of the objective f by calculating f at each sample point. These sample points may be processed in any one of a myriad of ways, possibly leading to further global exploration, or local searches to refine minima that have been identified. Our interest lies in the effectiveness of the initial exploration, and its relationship with the locations of the sample points.

In this paper we modify the original Halton sequence so that primes can be used more than once, and compare the original and prime recycling sequences against one another. All sequences considered herein fall into the broad class of quasi random sequences known as Niederreiter sequences [18, 19]. Recycling Halton sequences can also be classified as Niederreiter-Halton sequences [9]. The extreme generality of Niederreiter sequences means it is much clearer to construct recycling sequences by generalizing the Halton sequence rather than specializing Niederreiter sequences. Our goal here is to identify and numerically test specific examples of prime recycling sequences in the context of global optimization.

Many techniques have been proposed to improve the behaviour of the original Halton sequences. These techniques can also be applied to prime recycling sequences. Indeed the scope for such variation is greater for the prime recycling sequences than the original one. We examine the effect on prime recycling sequences of what is perhaps the most popular of these techniques: scrambling.

Different performance measures are used than for numerical integration. We use two such measures; one based on the length of the minimal spanning tree between the first N points of a sequence, and the other a comparison between pairs of sequences as to which generates the closest point to a randomly chosen point.

2 Halton sequences and variants

The Halton sequence [7] $\{\phi(k)\}_{k=0}^\infty$ in $[0, 1]^n$ is defined componentwise. Each component has an associated base b_i , where the set of bases b_1, \dots, b_n are co-prime. The i^{th} component of the k^{th} point is generated by radix inversion in base b_i as follows. Let

$$k = d_m \cdots d_0 \quad \text{base } b_i$$

where d_j is the j^{th} digit of k in its base b_i expansion. The radix inversion of k gives the i^{th} element $\phi_i(k)$ which is

$$\phi_i(k) = 0.d_0 \cdots d_m \quad \text{base } b_i. \tag{2}$$

This may be written in the alternative form

$$\phi_i(k) = \sum_{j=0}^{\infty} \left\{ \left\lfloor \frac{k}{b_i^j} \right\rfloor \bmod b_i \right\} \frac{1}{b_i^{j+1}}.$$

Here $\lfloor x \rfloor$ is the ‘floor’ function defined by $\lfloor x \rfloor = \max\{s \in \mathbb{Z} : s \leq x\}$, where \mathbb{Z} is the integers.

A valued property (the Uniformity Property) of Halton sequences is that they are asymptotically uniformly distributed. The reason for this uniformity can be seen as follows. Along the i^{th} coordinate axis subdivide the interval $[0, 1)$ into b_i sub-intervals of equal length. The cross products of these sub-intervals subdivide $[0, 1)^n$ into $B = \prod_{i=1}^n b_i$ boxes. It can be shown [7] that for any B consecutive Halton sequence points, each box contains exactly one such point. If a point lies in a specific box, then $k \bmod b_i$ must take a specific value for each $i = 1, \dots, n$. The Chinese Remainder Theorem then determines $k \bmod B$ provided b_1, \dots, b_n are co-prime. This property, and its extensions to finer subdivisions of $[0, 1)^n$, ensure that Halton sequences are far more evenly distributed than random points. This property is only meaningful once k is of the order of B , or larger. The choice that minimizes B is when b_i is the i^{th} smallest prime. Even with this choice B grows rapidly as n increases.

For small n Halton sequences are very effective, however as dimension increases long periods of correlation between entries start to appear. For example, using $b_1 = 41$ and $b_2 = 43$ the first 41 points lie in the hyperplane $41x_1 = 43x_2$, and most of the next 41 points lie on a parallel hyperplane, and so on. Some pictures showing long periods of correlation between components of Halton sequences can be found in [5, 29]. This problem can be avoided for low dimensions by choosing b_i as the i^{th} prime [25]. For higher dimensions a variety of strategies have been proposed, including shuffling, random starts, leaped sequences, and scrambling.

Shuffling a Halton sequence [15] is done in two stages. First all N points in the Halton sequence are generated, and then for each $i = 1, \dots, n$ in turn, the i^{th} entries of all N Halton points are randomly permuted. A disadvantage of this is that N must be known in advance, which is not usually the case in optimization. Moreover, the uniformity property may be lost by this shuffling process. Numerical work in [26] shows that shuffled sequences perform significantly worse than other low discrepancy sequences.

The Halton sequence can be viewed as applying the von Neumann–Kakutani transformation in base b_i to the i^{th} element of $\phi(k)$, for $i = 1, \dots, n$, with the origin as the initial point [12]. Wang and Hickernell [30] and Ökten [23] generalize the Halton sequence by selecting a random initial point in $[0, 1)^n$ and repeatedly applying the von Neumann–Kakutani transformation to each element in base b_i . This preserves the uniformity of each individual Halton sequence. The random sampling also allows errors to be estimated in Monte Carlo integration.

Kocis and Whiten [10] introduce the Halton sequence leaped, which essentially uses every L^{th} point of a Halton sequence with bases b_1, \dots, b_n , where L is co-prime with every base b_i . Numerical integration experiments by these authors show the Halton sequence leaped is very effective in practice in a wide range of dimensions up to 400. The leaping strategy can be applied to sequences other than the Halton sequence. In particular [10] also investigate leaped Faure and Sobol sequences. Their results show that leaping is less effective for these sequences than for the Halton sequence.

2.1 Scrambled Halton Sequences

The most popular way of improving the Halton sequences seems to be scrambling. Braaten and Weller [4] introduced scrambling by replacing each d_j with $\pi_i(d_j)$, where π_i is a permutation on the set $\{0, \dots, b_i - 1\}$. These permutations were formed one element at a time by choosing each next element to minimize the L_2 star discrepancy of the permutation. This strategy has been shown to be effective in breaking up the long periods of correlation between

elements of ϕ [10, 29]. Vandewoestyne and Cools [29] showed using numerical experiments that one very effective option is the reversing permutation given by $\pi(0) = 0$ and $\pi(d) = b_i - d$ otherwise. Chi et. al. [5] and others [24] consider permutations of the form

$$\pi_i(d_j) = w_i d_j \pmod{b_i} \quad \text{where} \quad w_i \in [1, b_i - 1] \cap \mathbb{Z}. \quad (3)$$

The uniformity property is preserved for scramblings such as [4, 5], and also for leaped Halton sequences. More sophisticated permutations have been proposed, including reversing the binary digits of each d_j , and eliminating any values that are too large [10].

3 Recycling Primes

The uniformity property is useful in optimization as it means no large holes exist in the sample points. It is an asymptotic property, and only becomes meaningful once k is of the order of B , or larger. Unfortunately as n grows, B grows very rapidly even when b_i is the i^{th} smallest prime. One way to reduce B is to use the smallest primes more than once. This means the radix inversion process must be modified otherwise elements of $\phi(k)$ which share a base will have identical values. In this section a modified form of the radix inversion is given which possesses the uniformity property. Additionally it retains the property that different points differ in every element.

For the moment we consider a sequence constructed using single generic base b . We use d_j as the j^{th} digit in the base b expansion of k , viz.

$$k = d_m d_{m-1} \cdots d_1 d_0 \quad \text{base } b.$$

Multiple elements (ϕ_1, \dots, ϕ_S say) of ϕ can be constructed using the base b via

$$\phi_i = \sum_{j=0}^{\infty} \left(\frac{1}{b^{j+1}} \left(\left[\sum_{\ell=j}^{Sj+i-1} d_\ell \right] \pmod{b} \right) \right) \quad i = 1, \dots, S. \quad (4)$$

As an example consider the case when $n = S = 3$. Using δ_j^i to denote the j^{th} digit of ϕ_i (i.e. $\phi_i = 0.\delta_0^i \delta_1^i \dots$), the first two digits of ϕ_1 , ϕ_2 and ϕ_3 are given by

$$\begin{array}{llll} \delta_0^1 = d_0 & \delta_0^2 = d_0 + d_1 & \delta_0^3 = d_0 + d_1 + d_2 & \text{and} \\ \delta_1^1 = d_1 + \cdots + d_3 & \delta_1^2 = d_1 + \cdots + d_4 & \delta_1^3 = d_1 + \cdots + d_5 & \text{all mod } b. \end{array} \quad (5)$$

This example illustrates how the uniformity property is preserved. If dimensions 1 to 3 are evenly subdivided into b^2 intervals, their direct products give b^6 boxes in $[0, 1)^3$. Which box a point lies in is determined by $\delta_0^1, \dots, \delta_1^3$. Forward substitution in (5) gives a bijection between d_0, \dots, d_5 and $\delta_0^1, \dots, \delta_1^3$. Any b^6 consecutive points run through all possible combinations of d_0, \dots, d_5 , so any b^6 consecutive points have one point per box.

Recycling primes also creates subsequences of points which lie on a single hyperplane. For r and s in $\{1, \dots, S\}$, if $\phi_r = \phi_s$, then $\delta_j^r = \delta_j^s$ for $j = 0, 1, 2, \dots$. Using $s > r$ without loss of generality, we have

$$\delta_j^s - \delta_j^r = d_{Sj+r} + \cdots + d_{Sj+s-1} = 0 \quad \forall j = 0, 1, 2, \dots$$

The worst case is when $r = S - 1$ and $s = S$, in which case these equations become

$$d_{(j+1)S-1} = 0 \quad j = 0, 1, 2, \dots$$

These equations are satisfied for the first b^{S-1} points in a prime recycling Halton sequence, and also for later subsequences of the same length. This imposes an upper limit on how many times a prime can be recycled without resurrecting the problem of long periods of correlation between entries of ϕ . If larger primes are being used, smaller primes may be able to be recycled more than once without extending the maximum length of a subsequence of points on one hyperplane.

The corresponding period for which elements of ϕ with distinct bases are linearly related is the smaller of the two bases. For the same base used S times, it is b^{S-1} . One can clearly use each base twice without increasing the period for which entries are linearly related. In addition a base b can be used S times if b^{S-1} is not larger than the second largest base. Using the smallest primes as bases, the linear relation period for n dimensions is minimized by the first n bases in the list

$$\begin{array}{cccccccccccccccc} 2, & 3, & 2, & 5, & 3, & 2, & 7, & 5, & 11, & 7, & 2, & 3, & 13, & 11, & 17, & 13, \\ 2, & 19, & 17, & 23, & 19, & 29, & 23, & 5, & 3, & 31, & 29, & 37, & 31, & 2, & 41, & 37, \\ 43, & 41, & 47, & 43, & 53, & 47, & 7, & 59, & 53, & 61, & 59, & 67, & 61, & \dots \end{array}$$

We now revert to the full notation and consider all bases b_i at once. For theoretical purposes we sort the elements of ϕ so that all elements using the same base are contiguous. The largest index of elements using a base b_i is given by

$$T(i) = \max\{j \in 1, \dots, n : b_i = b_j\}.$$

Using $T(0) = 0$, the elements of ϕ using base b_i are those with index values from $T(i-1) + 1$ to $T(i)$ inclusive. Hence $T(i) - T(i-1) = S(i)$, which is the number of times base b_i is used.

3.1 Scrambling and Recycling

The original Halton sequence has been improved by a variety of scrambling techniques, and most of these can be directly employed in conjunction with recycling. Here we consider scrambling methods which permute each digit in each base b_i expansion of the point number k . Specifically the digits $\{0, \dots, b_i - 1\}$ are permuted using a permutation π on $\{0, \dots, b_i - 1\}$ which satisfies $\pi(0) = 0$ but is otherwise arbitrary. In [5] it is shown that permutations not satisfying $\pi(0) = 0$ lead to biased sequences because they alter the infinite sequence of trailing zeros that the radix inversion of every integer ends in. The permutation can vary between elements of ϕ , even if they share the same base, and from digit to digit. Hence the notation π_ℓ^i denotes the permutation applied to the ℓ^{th} digit d_ℓ^i of the expansion of k in base b_i . The formula for $\phi_i(k)$ becomes

$$\phi_i = \sum_{j=0}^{\infty} \left(\frac{1}{b_i^{j+1}} \left(\left[\sum_{\ell=j}^{S(i)j+i-T(i-1)-1} \pi_\ell^i(d_\ell^i) \right] \bmod b_i \right) \right). \quad (6)$$

3.2 Theoretical properties

The first result shows that distinct points in the sequence remain distinct after projection onto any subspace of \mathbb{R}^n spanned by one or more unit coordinate vectors. This is relevant to optimization because objective functions often have negligible variation with respect to some decision variables. Without this projection property, calculating an objective function

at many different points could yield the same function value many times over, leading to a massive waste of computational effort.

Proposition 1 *Any two distinct points in a scrambled prime-recycling Halton sequence are distinct in every coordinate.*

proof: Consider element i of $\phi(k)$:

$$\phi_i(k) = 0 \cdot \delta_0^i \delta_1^i \delta_2^i \dots \quad \text{base } b_i.$$

Equation (4) shows that

$$\delta_j^i = \pi_j^i(d_j^i) + \dots + \pi_{S(i)j+i-T(i-1)-1}^i \left(d_{S(i)j+i-T(i-1)-1}^i \right) \quad \text{mod } b_i. \quad (7)$$

For this Halton point there exists J such that $d_j^i = \pi_j^i(d_j^i) = 0$ for all $j > J$. Consider now element i of Halton point number \bar{k} , with the base b_i expansion

$$\bar{k} = \bar{d}_M \bar{d}_{M-1} \dots \bar{d}_1 \bar{d}_0 \quad \text{base } b_i$$

where the number of the zero leading digits of \bar{k} is sufficient to ensure $M > J$. Let $\phi_i(k) = \phi_i(\bar{k})$. Then back substitution using (7) with $j = M, \dots, J+1$ in that order shows that

$$\pi_M^i(\bar{d}_M) = \dots = \pi_{J+1}^i(\bar{d}_{J+1}) = 0$$

and hence

$$\bar{d}_M = \dots = \bar{d}_{J+1} = 0.$$

Continued back substitution with $j = J, \dots, 0$ determines $\pi_J^i(\bar{d}_J), \dots, \pi_0^i(\bar{d}_0)$ uniquely, which gives $\bar{d}_J, \dots, \bar{d}_0$ uniquely. Hence $k = \bar{k}$ giving the required result. \square

The next result shows that the scrambled prime recycling Halton sequences are asymptotically uniform across $[0, 1]^n$.

Proposition 2 *Let the unit hypercube be subdivided into boxes of equal size and shape, where each box is of the form*

$$\left[m_1 b_1^{-J_1}, (m_1 + 1) b_1^{-J_1} \right) \times \dots \times \left[m_n b_n^{-J_n}, (m_n + 1) b_n^{-J_n} \right),$$

for integers m_1, \dots, m_n satisfying $0 \leq m_i < b_i^{J_i}$ for all $i = 1, \dots, n$. Here each $J_i \in \mathbb{Z}$ is positive, and

$$\forall i, j \in \{1, \dots, n\}, \quad b_i = b_j \Rightarrow J_i = J_j.$$

Then any consecutive $\prod_{i=1}^n b_i^{J_i}$ points in the scrambled prime recycling Halton sequence have exactly one point in each box.

proof: We subdivide the entries of ϕ a collection of into sets, where each set has a single base b , and different sets have distinct bases. It is shown that the entries in each set can be handled together in a way that allows us to then apply the Chinese Remainder Theorem.

First we consider a single such set of entries of ϕ which correspond to a base b_i . Let the first J_i digits of $\phi_i(k)$ in base b_i be given by

$$\phi_i(k) = 0 \cdot \delta_0^i \delta_1^i \dots \delta_{J_i-1}^i \quad \text{base } b_i.$$

In what follows we suppress the ‘ i ’ superscript on d_j^i because d_j^i is independent of i on any range of i values for which b_i is constant. Equation (6) gives

$$\delta_j^i = \pi_j^i(d_j) + \cdots + \pi_{jS(i)+i-T(i-1)-1}^i(d_{jS(i)+i-T(i-1)-1}) \quad \text{base } b_i.$$

These equations can be solved for $\pi_0^i(d_0), \dots, \pi_{(j+1)S(i)-1}^i(d_{(j+1)S(i)-1})$ in that order by considering pairs of values (j, i) in increasing lexicographic order; that is to say (j_1, i_1) precedes (j_2, i_2) if either $j_1 < j_2$, or both $j_1 = j_2$ and $i_1 < i_2$. Once the permuted values of $d_0, \dots, d_{(j+1)S(i)-1}$ are known the original values of these digits are also determined uniquely. This gives the first $S(i)J_i$ digits of k , which may be expressed as

$$k = d_{S(i)J_i-1} \cdots d_0 \quad \text{mod } b^{S(i)J_i}. \quad (8)$$

Each set of entries of ϕ gives an equation of the form (8). Second, the Chinese Remainder Theorem may be applied to the collection of equations, each of the form (8), where each equation corresponds to a set of entries of ϕ . The Chinese Remainder Theorem shows that k is determined uniquely, $\text{mod } \prod_{i=1}^n b_i^{J_i}$, as required. \square

This result can be used to directly construct upper bounds on the L_∞ norm star discrepancy D^* following the approach taken by Halton [7] and others. We do not do so here because the difference between such bounds and the actual star discrepancy can be vast [25]. Also, such bounds are asymptotic and in practice one often uses few enough points that these asymptotic properties are not relevant. Finally, the star discrepancy is related to the quality of a quasi-random sequence in numerical integration via the Koksma-Hlawka inequality. Our interest lies primarily in optimization, and so the L_∞ star discrepancy is less relevant.

Proposition 2 introduces an interesting trade-off. Consider the Halton sequence with distinct b_i as the n smallest primes. Recycling primes by replacing the largest primes with smaller ones reduces the value of k at which the uniformity property becomes relevant, with the smallest such k occurring at $b_i = 2$ for all i . Initially recycling primes also reduces the long periods of correlation between entries of ϕ by removing the largest primes. There comes a point which these correlations occur between entries with the same base. Beyond this, one can only improve uniformity at the expense of worsening correlation.

4 Measuring Coverage

Sobol and Shukhman [28] show that the minimum distance ρ_N between two points in the first N members of a Sobol sequence is at least $\sqrt{n}/(2N)$. They numerically evaluate the minimum distances for Halton, Faure, and Sobol sequences in 3 and 10 dimensions. Their results show the three types of sequence are similar, with the minimum distances for these sequences being within a factor of 2 of one another. These numerical computations suggest that $\rho_N \approx 1/\sqrt[3]{N}$.

The minimum distance is not a particularly useful measure of quality in an optimization context. Consider the case when an objective function is sampled at N quasi-random points. Even if ρ_N is extremely small, there may only be a very small number of pairs of points which are very close, and hence only a small number of function evaluations are superfluous. A different quasi-random sequence may have a higher ρ_N , but many more pairs of points which are close enough to give approximately the same function value. The same defect may occur with larger groupings of points. For example in one dimension the points $\{\epsilon, 2\epsilon, \dots, 2N\epsilon\}$

are clearly inferior to N widely spaced pairs of points of the form $\{s/N, (s/N) + \epsilon\}$ for $s = 0, \dots, N - 1$, when $N\epsilon \ll 1$.

A quantity which more accurately represents the quality of a quasi random sequence is sought. Our approach is to use an existing simplification of stochastic models already used in global optimization.

A common strategy in global optimization is to model the unknown objective function by a stochastic process (see e.g. [32]). This can lead to very complicated conditional probability distributions as the number of known points grows. A simplifying assumption was introduced by Mockus [14], in which the distribution of values at a sample point is dependent only on the closest previous sample point. The distribution of function values at the new sample point is shown to be Gaussian [13], and its mean is taken to be the function value of the closest previous sample point. The variance of the new sample point's function value is a monotonically increasing function of the distance from it to the nearest known sample point. The distributions of the function value differences between sample points and their closest predecessors are assumed to be independent of one another. There is clearly much flexibility here. However without further knowledge of how an objective function has arisen, a natural approach is to consider the line segment between the new and closest previous sample points. An obvious assumption is that increments in the objective function are independent and identically distributed. This yields a variance proportional to the distance between these two sample points.

The quality of a set of sample points $Y = \{y_1, \dots, y_N\}$ is defined as the sum of standard deviations of the objective function values $f_r = f(y_r)$. The greater the total standard deviation, the more information is gained about the function. One sample point (hereafter y_1) is chosen as a 'first point'. This point has no nearest neighbor, and so the standard deviation of f_1 is not determined by the distance to any other sample point. Since the minimizer of the objective function f is invariant with respect to the addition of an arbitrary constant, the standard deviation of f_1 can be removed by choice of this arbitrary constant. The standard deviation of each f_r , $r > 1$, is determined by the distance between y_r and another sample point $y_{p(r)}$. A graph is constructed by drawing edges from $y_{p(r)}$ to y_r for each $r = 2, \dots, N$. This graph has N vertices and $N - 1$ edges. The only meaningful possibility is that the graph is a tree. If it were not, it would be disconnected and contain a cycle. If it were disconnected, the function values on two disjoint subgraphs would be invariant under translation of one subgraph's sample points relative to the other's. Also, if there exists a cycle then the random variables $f_r - f_{p(r)}$ on the edges of that cycle are no longer independent. The edges are selected to minimize the total standard deviation, and so this tree is the minimal spanning tree. The minimal spanning tree is independent of the choice of root y_1 , as required. If more than one minimal spanning tree exists, we choose one arbitrarily; our interest lies in the length of the tree, not the tree itself.

A simple observation as to why the standard deviation is used for the edge length rather than the variance is as follows. In one dimension if the variance is used as the edge length, the length of the minimal spanning tree is $\max(Y) - \min(Y)$ irrespective of the locations of the remaining $N - 2$ points. In contrast, with the standard deviation as the edge length, the length of the minimal spanning tree is maximized when the points are equally spaced between $\min(Y)$ and $\max(Y)$.

Table 1: Average (standard deviation) edge length of the minimal spanning trees for the first N points of the standard (std) and prime recycling (rec) Halton sequences. Neither sequence is scrambled.

N	5 dimensions		15 dimensions		25 dimensions		35 dimensions		45 dimensions	
	std	rec	std	rec	std	rec	std	rec	std	rec
25	0.71	0.69	0.95	1.05	0.95	1.18	0.95	1.21	0.95	1.26
50	0.66	0.64	0.99	1.04	1.01	1.19	1.01	1.26	1.01	1.32
100	0.60	0.59	0.95	1.00	1.05	1.18	1.05	1.27	1.05	1.33
200	0.54	0.54	0.92	0.97	1.04	1.18	1.06	1.28	1.08	1.34
400	0.50	0.51	0.90	0.95	1.02	1.16	1.05	1.27	1.08	1.34
800	0.47	0.47	0.88	0.92	1.01	1.13	1.04	1.27	1.06	1.34

4.1 Numerical coverage estimates

Sequences were compared in two ways. The first looked at the average edge lengths of the minimal spanning spanning trees for several standard and prime recycling Halton sequences. The second looked at which of a pair of sequences found the closest point to a randomly chosen point most often. In all numerical work, the prime recycling sequence used the primes in the order given in Section 3 as bases. It was found that the reversing permutation [29] made little difference on either the standard or the prime recycling sequence, and so only scrambling of the form (3) have been investigated further.

The minimal spanning tree length of the standard Halton sequence is compared to that of the prime recycling sequence (both without permutation) in Table 1. In 5 dimensions the two are comparable, except that for the first 50 points or so the standard sequence is superior. In contrast for 15, 25, 35, and 45 dimensions the prime recycling sequence is clearly superior, and the margin grows with increasing dimension. In higher dimensions the average separation between points initially grows with both types of sequence as N increases, and then eventually starts to fall. This is due to the fact that both sequences initially have correlations between elements with similar bases.

A similar comparison between the scrambled Halton, and scrambled prime recycling Halton sequences is given in Table 2. Both use multiplicative scrambling of the form (3), where the coefficients w_i are as listed in [5]. The recycling sequence alters the permutation from one use of a base b_i to the next. The permutation $\pi_j^i(d)$ of a digit d is formed as follows. First d is shifted upwards $i - T(i - 1) - 1$ times yielding a value d_{us} , where the upwards shift of d is $d + 1$, except that the upwards shift of $b_i - 1$ is 1 (not 0). The value of $\pi_j^i(d)$ is then set equal to $\pi_j^{T(i-1)+1}(d_{\text{us}})$. This makes the permutation for the first use of base b_i identical to the permutation used for b_i for the standard sequence.

The two scrambled sequences perform identically in lower dimensions, with the prime recycling one having a slight edge from 25 dimensions up. The margin between the sequences may appear very slight, but the numbers are averaged square roots of distances between points. In n dimensions raising the ratio of these numbers to the $(2n)^{\text{th}}$ power gives the ratio of volumes of hyperspheres closest to each point. In 45 dimensions the volume ratio is of the order of 5.

Pairwise comparisons between the various sequences are made on random cone functions. Each such cone function is of the form $\|x - c\|_2$, where the centre c is chosen randomly from a

Table 2: Average (standard deviation) edge length of the minimal spanning trees for first N points of the scrambled standard (std) and scrambled prime recycling (rec) Halton sequences. The recycling sequence uses shifted permutations for different instances of each prime.

N	5 dimensions		15 dimensions		25 dimensions		35 dimensions		45 dimensions	
	std	rec	std	rec	std	rec	std	rec	std	rec
25	0.68	0.69	1.07	1.07	1.23	1.25	1.36	1.39	1.46	1.49
50	0.64	0.65	1.03	1.04	1.22	1.24	1.34	1.37	1.43	1.48
100	0.60	0.60	1.00	1.00	1.19	1.20	1.33	1.35	1.42	1.46
200	0.56	0.55	0.98	0.97	1.18	1.18	1.31	1.33	1.41	1.45
400	0.52	0.50	0.95	0.94	1.16	1.16	1.30	1.32	1.40	1.43
800	0.47	0.47	0.93	0.92	1.14	1.14	1.28	1.30	1.39	1.42

uniform distribution on the unit hypercube. For each pair of sequences 400 such cone functions were minimized, and the number of times each sequence found the closest point to the centre c counted. The number of successes for each sequence follows a binomial distribution. If both sequences were equally likely to find the closest point to the cone’s minimizer, then the mean number of successes for each sequence would be 200, with a standard deviation of 10. Three pairs of sequences were compared: the standard and prime recycling sequences, both without scrambling; the scrambled standard and unscrambled prime recycling sequences; and scrambled standard sequence with the scrambled prime recycling sequence. The outcomes are presented in Table 3.

In 15 dimensions and above, recycling is clearly beneficial in the absence of scrambling. Indeed five of the ten trials showed the recycling sequence to be better by two or more standard deviations. The comparison between the two scrambled sequences shows a similar pattern, but not quite as strongly. Only four results reached two standard deviations in favour of recycling. The comparison between the scrambled standard sequence and the unscrambled recycling sequence is interesting. Essentially it shows a tie, with perhaps a suggestion that recycling is better at 25 dimensions and above. None of the results were outside the two standard deviation boundaries. In contrast, the scrambled standard sequence performs significantly better on the minimal spanning tree length test. The discrepancy between the two performance measures may be due to the fact that the cone functions are not produced using a stochastic process remotely like that envisioned by Mockus in [13, 14].

5 Conclusion

Specific instances of the Niederreiter-Halton sequence which allow bases to be re-used have been examined. Recycling bases retains the theoretical properties of the standard Halton sequences, and dramatically reduces the number of points needed before the uniformity property becomes meaningful. Strategies such as scrambling and leaping can be applied to recycling sequences. These strategies do not change the number of points needed before uniformity becomes meaningful. Hence prime recycling improves on the uniformity property for all such sequences. Recycling can also reduce the length of correlations between elements of a quasi-random sequence.

A new quality measure for quasi-random sequences relevant to optimization has been in-

Table 3: A comparison of the first N points of the standard and recycling Halton sequences on minimizing $\|x - c\|$ over 400 random centres c . The entries list the number of times the recycling sequence got the closest value to the centre for 400 centres drawn randomly from a uniform distribution on the unit hypercube.

n	no scrambling		only std scrambled		both scrambled	
	$N = 400$	$N = 800$	$N = 400$	$N = 800$	$N = 400$	$N = 800$
5	188	187	195	197	185	213
15	226	199	203	185	223	204
25	223	217	198	211	210	199
35	231	220	202	211	212	227
45	215	238	210	202	223	221

troduced. The standard and recycling sequences have been compared using this new measure, and on a large number of randomly generated cones. These show recycling significantly improves the standard Halton sequence. Scrambled standard and recycling sequences are also compared, and the recycling sequence is again better, although the margin is smaller.

There is scope for much further work in exploring the use of scrambling, leaping, and other strategies with recycling Halton sequences. The results for scrambled recycling sequences herein have simply copied the scrambling used for standard Halton sequences. A thorough exploration of possible permutations for recycling sequences is likely to yield scramblings that are significantly better on those sequences than scramblings designed for the standard sequence. Many papers have been written exploring these variations of the Halton sequence, and it is beyond the scope of a single paper to do so for the recycling sequences.

References

- [1] M. A. Abramson, C. Audet, J. E. Dennis Jr., and S. Le Digabel, OrthoMADS: a deterministic MADS instance with orthogonal directions, *SIAM J. Opt.*, **20** (2009), 948–966.
- [2] M. M. Ali and C. Storey, Topographical multi-level single linkage, *J. Global Optim.*, **5** (1994), pp. 349–358.
- [3] C. R. Bhat, Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences, *Transportation Research part B*, **37** (2003), 837–855.
- [4] E. Braaten and G. Weller, An improved low-discrepancy sequence for multi-dimensional quasi-Monte Carlo integration, *J. Comput. Phys.*, **33** (1979), 249–258.
- [5] H. Chi, and M. Mascagni, and T. Warnock, On the optimal Halton sequence, *Math. Computers in Simulation*, **70** (2005), 9-21.
- [6] H. Faure, Discr pance de suites associ es   un syst me e num ration (en dimension s), *Acta Arithmetica*, **41**, (1982), 337–351.
- [7] J. Halton, On the efficiency of certain quasi-random sequences in evaluating multi-dimensional integrals, *Numerische Mathematik*, **2** (1960), 84–90.

- [8] E. Hlawka, Funktionen von beschränkter variation in der theorie der Gleichverteilung, *Ann. Mat. Pura. Appl.* **54**, (1961), 352–333.
- [9] R. Hofer, On the distribution properties of Niederreiter-Halton sequences, *Journal of Number Theory*, **129** (2009), 451–463.
- [10] L. Kocis and W. J. Whiten, Computational investigations of low discrepancy sequences, *ACM Trans. Math. Software*, **23** (1997), 266–294.
- [11] J. F. Koksma, Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1, *Mathematica B (Zutphen)*, **11** (1942/43), 7–11.
- [12] J. P. Lambert, Quasi-Monte Carlo, low discrepancy sequences, and ergodic transformations, *J. Computational and App. Math.*, **12 & 13** (1985), 419–423.
- [13] J. Mockus, On Bayesian methods for seeking the extremum and their applications, *Information Processing 77*, North-Holland, 1977, 195–200.
- [14] J. Mockus, The simple Bayesian algorithm for multidimensional global optimization, in *Numerical techniques for stochastic systems*, F. Archetti and M. Cugiani (Eds.), pp. 369–377, North-Holland, 1980.
- [15] W. J. Morokoff and R. E. Caflish, Quasi-random sequences and their discrepancies, *SIAM J. Sci. Comput.*, **15** (1994), 1251–1279.
- [16] J. Matoušek, On the L_2 -discrepancy for anchored boxes, *J. Complexity*, **14** (1998), 527–556.
- [17] H. Niederreiter, Quasi-Monte Carlo methods and pseudo-random numbers, *Bulletin Amer. Math. Soc.*, **84** (1978), 957–1041.
- [18] H. Niederreiter, Point sets and sequences with small discrepancy, *Monatsch. Math.*, **104** (1987), 273–337.
- [19] H. Niederreiter, Low-discrepancy and low-dispersion sequences, *Journal of Number Theory*, **30** (1988), 51–70.
- [20] H. Niederreiter, Error bounds for quasi-Monte Carlo integration with uniform point sets, *J. Computational App. Math.*, **150** (2003), 283–292.
- [21] H. Niederreiter, Constructions of (t, m, s) nets and (t, s) sequences, *Finite Fields Appl.*, **11** (2005), 578–600.
- [22] H. Niederreiter and C.P. Xing, Low discrepancy sequences and global function fields with many rational places, *Finite Fields Appl.*, **2** (1996), 241–273.
- [23] G. Ökten, Generalized von Neumann-Kakutani transformation and random-start scrambled Halton sequences, *J. Complexity*, **25** (2009), 318–331.
- [24] C. J. Price, Non-linear semi-infinite programming, Ph.D. Thesis, 1992, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand.

- [25] Ch. Schlier, Discrepancy behaviour in the non-asymptotic regime, *App. Num. Math*, **50** (2004), 227–238.
- [26] Ch. Schlier, On scrambled Halton sequences, *App. Num. Math*, **58** (2008), 1467–1478.
- [27] I. M. Sobol, On the systematic search in a hypercube, *SIAM J. Num. Anal.*, **16** (1979), 790–793.
- [28] I. M. Sobol and B. V. Shukhman, Quasi-random points keep their distance, *Math. and Computers in Simulation*, **75** (2007), 80–86.
- [29] B. Vandewoestyne and R. Cools, Good permutations for deterministic scrambled Halton sequences in terms of L_2 -discrepancy, *J. Comp. App. Math.* **189** (2006), 341–361.
- [30] X. Wang and F. J. Hickernell, Randomized Halton sequences, *Math. Computer Modelling*, *32* (2000), 887–899.
- [31] S. Tezuka, Polynomial arithmetic analogue of Halton sequences, *ACM Trans. Model. Comput. Simul.* (1993), **3**, 99–107.
- [32] A. Torn and A. Žilinskas, Global Optimization, Lecture Notes in Computer Science, Vol. 350, Springer-Verlag, Berlin Heidelberg New York, 1989.