A direct inversion method for non-uniform quasi-random point sequences

Colas Schretter and Harald Niederreiter

Abstract. The inversion method is an effective approach for transforming uniform random points according to a given probability density function. In two dimensions, horizontal and vertical displacements are computed successively using a marginal and then all conditional density functions. When quasi-random low-discrepancy points are provided as input, spurious artifacts might appear if the density function is not separable. Therefore, this paper relies on combining intrinsic properties of the golden ratio sequence and the Hilbert space filling curve for generating non-uniform point sequences using a single step inversion method. Experiments show that this approach improves efficiency while avoiding artifacts for general discrete probability density functions.

Keywords. Quasi-random points, non-uniform distribution, inversion method, golden ratio sequence, van der Corput sequence.

2010 Mathematics Subject Classification. 11K31, 11K38, 11K45, 65C05, 65C10.

1 Introduction

In this paper, we consider the problem of generating a sequence of non-uniform self-avoiding points following a discrete probability density function given by a normalized two-dimensional gray-scale image. A solution to this classical problem has many potential practical applications since continuous density functions can be sampled for defining a suitable discrete approximation. Furthermore, the proposed method relies on very simple and fast one-dimensional generator sequences based on the golden ratio number and leverages properties of the Hilbert space filling curve that generalize naturally to arbitrary dimensions.

The volume of Devroye [3] is an impressive collection of non-uniform random variate generation methods that are adapted for generating sets and sequences of pseudo-random samples, given probability density functions (PDF) in either analytical or discrete tabular forms. Unfortunately, the documented inversion tech-
niques generate spurious clusters of points when deterministic quasi-random numbers are fed as input instead of pseudo-random coordinates.

In 1972, the work of Hlawka and Mück proposed a modified inversion solution for improving the discrepancy of transformed quasi-random point sets in the one-dimensional case [7] and a journal paper publication (written in German) addressed an extension to the multi-dimensional case [8]. The method improves discrepancy by constraining points to snap on a grid; hence, maximizing the minimum distance between adjacent points. More recent papers documented shortcomings of this extended technique [4,6]. We observed that the emergence of spurious clusters remains still a pending practical problem that is caused by successive transformations with the inversion method.

We considered the alias method [9,14] for its intrinsic efficiency, but it requires making a decision based on a random throw. That means that for inverting a one-dimensional probability density function, we need two quasi-random numbers per sample. Those two numbers have to be uncorrelated. The alias method disrupts the stratification of points anyway because of this second random choice. The goal of the alternative direct inversion method proposed in this work is generating better quality point sequences for low-discrepancy sampling of tabulated discrete probability density functions.

The remainder of this work is organized as follows. First, the classical successive inversions method is reviewed in Section 2 and a direct inversion construction is proposed in Section 3. Then, the golden ratio sequence is proposed as an alternative to the van der Corput sequence in Section 4. Experiments in Section 5 demonstrate an application for generating sequences of both non-uniform and uniform point samples. Finally, we summarize and plan future work in Section 6.

2 Successive inversions method

Given a discrete probability density function \( f(x, y) \) of \( N \times M \) picture elements, the inverse transform sampling can be used to warp a uniform point distribution from the marginal \( f_x \) and the cumulative marginal \( F_x \),

\[
f_x(x) = \sum_{y=1}^{M} f(x, y), \quad F_x(x) = \sum_{k=1}^{x} f_x(k)
\]

and all conditional distributions \( f(y|x) \) and cumulative distributions \( F(y|x) \),

\[
f(y|x) = \frac{f(x, y)}{f_x(x)}, \quad F(y|x) = \sum_{k=1}^{y} f(k|x).
\]

These one-dimensional probability and cumulated density functions are computed
from the normalized two-dimensional image $f$ such that

$$
\sum_{x=1}^{N} f_x(x) = \frac{1}{N} \sum_{y=1}^{M} f(y|x) = \sum_{x=1}^{N} \sum_{y=1}^{M} f(x, y) = 1.
$$

Given a random point $(u, v) \in [0, 1)^2$, a first numerical inversion procedure finds the transformed horizontal coordinate $x$ that satisfies $F_x(x) = u$. A second numerical inversion step finds the displaced vertical coordinate $y$ that satisfies $F(y|x) = v$. Since the cumulative distribution functions are monotonically increasing, they are trivially invertible with numerical methods. According to the book of Devroye, guide tables can be used for accelerating a line search technique [3, p. 96]. In practice, such hashing table acceleration yields very fast inversions in constant time on average [1].

While statistically equivalent, successive transformations are not commutative and an altogether different transformation is defined when transforming first the vertical coordinate from the vertical marginal distribution $f_y(y)$, then displacing horizontally the input point from the conditionals $f(x|y)$. Discontinuities in the input density function $f(x, y)$ induce seams between adjacent discrete conditional density functions and this might result in artifacts after the inversion when quasi-random numbers are used to generate uniformly distributed input points.

3 A direct inversion method

To alleviate the aforementioned problems with successive inversions, this section proposes an alternative approach in which the 2D density image is first transformed into a 1D signal with smooth and continuous cumulative density function. Then a single one-dimensional inversion step is applied and the transformed coordinate is mapped into a point in higher dimensions.

In the preprocessing step, the input 2D image is traversed along a Hilbert space filling curve as shown in Figure 1. Hilbert curves are fractal constructions defined by a locality-preserving mapping between one-dimensional and higher-dimensional spaces $H : \mathbb{R} \to \mathbb{R}^2$ and its inverse $H^{-1} : \mathbb{R}^2 \to \mathbb{R}$ such that $H(t) = (x, y)$ and $H^{-1}(x, y) = t$. Therefore, the 1D probability density function $f_H$ and its corresponding cumulative density function $F_H$ are given by

$$
f_H(t) = f(H(t)), \quad F_H(t) = \sum_{k=1}^{t} f_H(k)
$$

with $t \in [1, N \times M]$.

A Hilbert space filling curve defines an ordering on vertices of a discrete regular isotropic grid in such a way that the distance between successive points is always
constant. The natural proximity of points along the curve has been applied earlier for segmenting the integration domain in a non-square number of contiguous strata for stratified sampling [12].

The picture elements are ordered by a traversal following the path shown by the gradient from dark to bright in Figure 2. The traversal accumulates the value of successive pixels for producing a cumulative discrete density function stored in a 1D array that is shown in multiple lines on the right-most image in Figure 2. In this example, the input density function is a binary indicator function; therefore, the cumulative density contains constant plateaus connected by linear gradients, forming a continuous and smooth function.
A direct inversion method

For non-uniform sampling, a single coordinate \( u \in [0, 1) \) is generated, then the inversion method is applied with the line search method explained in the previous section. In a second step, the transformed 1D coordinate is mapped to a 2D point using the inverse Hilbert mapping. We use the non-recursive implementation of Lawder [10]. Such a non-recursive algorithm can be complicated to program, but very high precision can be reached with a sufficient order of recursion.

While improving speed was not our primary goal, the new method requires one unique inversion of a larger smooth one-dimensional conditional probability function (CDF). Smoothness and continuous properties adapt to compression and approximations for saving memory resources and speeding-up the specific computer implementation, like proposed in [1]. An interesting qualitative comparison of the alias and inversion methods on practical examples can be found in the aforementioned paper.

4 Golden ratio sequences

The previous section has presented a solution for casting the multidimensional successive inversions technique into a single one-dimensional inversion. This section proposes an approach for generating an infinite sequence of self-avoiding points, in analogy with the van der Corput sequence [5, 11]. When using such a method for generating input coordinates, non-uniform low-discrepancy point sequences can be obtained.

Each term in the golden ratio sequence with seed constant \( s \in [0, 1) \) is given by the fractional part \( G_s(i) \) of the sum between \( s \) and an integer multiple of the golden ratio

\[
\phi = \frac{1 + \sqrt{5}}{2} \approx 1.618034 \ldots
\]

More formally, golden ratio sequences are given by

\[
G_s(i) = \{s + i \cdot \phi\}, \quad \text{for all } i \geq 1,
\]

where \( \{t\} \) denotes the fractional part of the real number \( t \). Note that the conjugate golden section \( \tau = \phi - 1 \) can be used in place of \( \phi \) since only fractional parts are retained. Hence, computing \( G_s(i + 1) \) given \( G_s(i) \) only involves an addition and a test for integer overflow.

Figure 3 compares the first five terms of the van der Corput sequence in base 2 and the golden ratio sequence with \( s = 0 \). In fact, the van der Corput sequence always splits the largest interval between points in two and the ratio of the largest interval to the smallest is 2. Exceptionally, when the number of terms is a power of the base prime, all large intervals have been broken into small ones and the ratio is 1. Samples from the golden ratio sequence cover more evenly the unit inter-

\[
\phi = \frac{1 + \sqrt{5}}{2} \approx 1.618034 \ldots
\]
Figure 3. Comparison of the first five coordinates generated by the van der Corput sequence (top) and the golden ratio sequence (bottom). The latter spreads over a larger span, minimizing the maximum gap between the two closest coordinates.

The fractal Hilbert curve can be used for mapping one-dimensional coordinates to higher dimensions. However, strong aliasing artifacts appear when using low-discrepancy coordinates generated by the van der Corput sequence. Aliasing is due to the fact that both the recursive generation of the Hilbert curve and the construction of the van der Corput sequence rely on base-2 arithmetic. In contrast, the coordinates generated by the golden ratio sequence described in the following section do not align on any grid.

The golden ratio conjugate \( \tau \) is the “most irrational” among all irrational numbers in the interval \([0, 1)\) since all partial quotients in the continued fraction expansion of \( \tau \) assume the least possible value 1, see [2]. As a consequence, the variability of pairwise distances between points is minimized when using this constant \( \tau \) in fractional sequences. This key property is due to the “maximum irrationality” of the golden ratio. Two points in the golden ratio sequence in one dimension with similar fractional parts are projected to adjacent points on the Hilbert curve. Therefore, successive elements of a Hilbert curve sampled with a golden ratio sequence are likely mapped to well-separated points in higher dimensions without aliasing artifacts. To our knowledge, the golden ratio sequence has so far not yet been used in combination with the Hilbert space filling curve as a tool for the generation of non-uniform low-discrepancy point sequences.

5 Experiments

Two experiments have been conducted for comparing our proposed method with the result of transforming with the successive inversions methods a set of points
generated with the 2D Halton sequence in bases 2 and 3. In Figure 4, two different discrete probability density functions are represented by gray-scale images: one indicator function and a second density containing also a smooth ramp gradient, resolution lines and a Gaussian bell. The figure shows a side by side comparison between transformed Halton points and a transformed golden ratio sequence followed by a mapping into 2D according to the Hilbert space filling curve.

It can be observed that correlations yield structures in Halton points [13] and stratification artifacts are very visible in non-uniform low-discrepancy applications. In contrast, no apparent structures are observed with the direct inversion scheme using the one-dimensional golden ratio sequence. No repeated patterns are visible with the proposed method and the thin quasi-parallel lines seem to be more precisely resolved.

In a second experiment shown in Figure 5, a constant density image has been used in input; therefore, the inversion step can be skipped since the density profile is uniform. We can observe repeated patterns in the plot of 1000 Halton points. The points generated by the golden ratio sequences appear irregular while demonstrating comparable low-discrepancy properties.
C. Schretter and H. Niederreiter

Halton sequence in bases 2 and 3
Golden ratio seq. and Hilbert mapping

Figure 5. Comparison between the first 1000 points of the Halton sequence (left) and the Hilbert mapping of the first 1000 coordinates of the golden ratio sequence (right). Some repetitions in the Halton sequence are contoured by squares and circles. Similar repeated patterns are not observed with the proposed method.

6 Conclusion

This work proposed a novel solution for the problem of non-uniform low-discrepancy sampling. The approach generates first a low-discrepancy one-dimensional sequence using fractional parts of successive integer multiples of the golden ratio number. In a second step, coordinates are displaced along an unfolded Hilbert space filling curve with the classical numerical inversion method. Finally, coordinates are mapped to higher-dimensional points. The method requires only one step of inversion and is therefore computationally faster while avoiding artifacts observed with successive inversions.

Future work will investigate the properties of the generated point sets with respect to various discrepancy measures. It is also interesting to analyze the potential of such an approach when using generalizations of the Hilbert space filling curve in higher dimensions.

Bibliography


A direct inversion method


Received June 11, 2012; accepted November 15, 2012.

Author information

Colas Schretter, EECS, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139-4307, USA. E-mail: cschrett@mit.edu

Harald Niederreiter, RICAM, Austrian Academy of Sciences, Altenbergerstrasse 69, A-4040 Linz, Austria. E-mail: ghnied@gmail.com