THEORETICAL GUARANTEES FOR POISSON DISK SAMPLING USING PAIR CORRELATION FUNCTION

Bhavya Kailkhura⋆ Jayaraman J. Thiagarajan† Peer-Timo Bremer† Pramod K. Varshney⋆

* Department of EECS, Syracuse University  † Lawrence Livermore National Laboratory

ABSTRACT
In this paper, we study the problem of generating uniform random point samples on a domain of $d$-dimensional space based on a minimum distance criterion between point samples (Poisson-disk sampling or PDS). First, we formally define PDS via the pair correlation function (PCF) to quantitatively evaluate properties of the sampling process. Surprisingly, none of the existing PDS techniques satisfy both uniformity and minimum distance criterion, simultaneously. These approaches typically create an approximate PDS with high regularity, and inherently present high risk for sample aliasing. Our new formulation based on PCF introduces a new approach to evaluate PDS properties which leads to theoretical bounds on the size of a PDS in arbitrary dimensions as well as a faster algorithm to create better quality samplings than the current PDS approaches.

Index Terms— Poisson-disk sampling, dart throwing, multidimensional sampling, maximal sampling

1. INTRODUCTION
Exploratory analysis and inference in high dimensional parameter spaces is a ubiquitous problem in science and engineering and a wide variety of machine learning tools and optimization techniques are available. In the most generic formulation, one is interested in analyzing a high-dimensional function $f: \mathbb{D} \rightarrow \mathbb{R}$ defined on the $d$-dimensional domain $\mathbb{D}$. The common approach is to first create an initial sampling $X = \{x_i \in \mathbb{D}\}$ of $\mathbb{D}$, evaluate $f$ at all $x_i$, and approximate $f$ using only the resulting pairs $(x_i, f(x_i))$. However, it is well known that the effectiveness of this approach heavily depends on the quality of the initial sampling $X$. Furthermore, evaluating $f$ through simulations or experiments is typically expensive. Therefore, often the goal is to use as few samples as possible to provide as much information as possible.

Without prior knowledge of $f$, one objective when creating $X$ is that the sampling should be random to have an equal chance of finding features of interest, e.g., local minima in an optimization problem, anywhere in $\mathbb{D}$. The second objective is to cover all of $\mathbb{D}$ uniformly in order to guarantee that all sufficiently large features are found. The most common approach in practice is to use Latin Hypercube sampling [1] which generates samples that are typically neither random nor uniform. While easy to create, such designs, in general, require significantly more points to achieve, for example, the same regression error than more optimized sampling designs. Optimal sampling in this respect is referred to as Poisson Disk Sampling (PDS) [2–26] defined as a set of $x_i$’s that are randomly distributed (Objective 1) but no two samples are closer than a given minimum distance $r_{min}$ (Objective 2) (See Figure 1(a)). Unfortunately, currently there does not exist an algorithm to create a theoretically accurate PDS. In particular, existing techniques largely ignore the randomness objective and instead concentrate exclusively on the $r_{min}$ condition. This leads to the notion of a maximal-PDS (MPDS) in which no two samples are closer than $r_{min}$ and no more points in $\mathbb{D}$ can be added. However, as we will demonstrate later, existing techniques to generate MPDS actually violate the randomness constraint and in fact resulting samplings typically contain significantly more points than theoretically possible for a PDS of a given $r_{min}$. Furthermore, guaranteeing maximality requires expensive checks causing the resulting algorithms to be slow in moderate (2-5) and practically infeasible in high (7 and above) dimensions. Here we introduce a new approach to understand the nature of PDS which leads to theoretical bounds on the size of a PDS in arbitrary dimensions as well as a faster algorithm to create higher quality samplings than the current MPDS approaches.

First, we formally define a PDS via the pair correlation function (PCF). The PCF of a point measures the distribution of its distance to all other points (See Figure 1(b)). It has been shown that the degrees of freedom to characterize point distributions in the pair correlation space is low and can be directly linked to regularity [7]. Using this new PDS formulation, we formally pose an optimization problem for determining the upper bound on the sample budget $N$ (or $r_{min}$), that does not violate PDS requirements, for a given radius $r_{min}$ (or a sample budget $N$). These bounds hold for arbitrary dimensions and arbitrary parallelepipeds domains not just for the unit cube. A striking result is that the theoretical upper bound is significantly lower than the sample size of the corresponding MPDS and the difference increases significantly with increasing number of dimensions. In particular, we show that existing sequential algorithms to create a MPDS typically create an approximate PDS before the maximality constraint forces more and more samples to be added which only increase the regularity and thus the potential for aliasing. Paradoxically, due to the nature of these algorithms the later samples, added for maximality, are computationally more expensive to compute than the earlier samples. Therefore, knowing the theoretical upper bound on $N$ for a given $r_{min}$ leads to faster algorithms by simply terminating the sampling process early, once an approximate PDS has been found. Finally, we derive expressions for the power spectral density of a PDS and demonstrate that the current MPDS techniques can lead to undesired aliasing. In summary, by using PCFs to analyze PDSs we provide new theoretical insights into the sampling process and faster algorithms to create true PDS that are expected to perform significantly better in practice.

2. PRELIMINARIES

Poisson Disk Sampling: Poisson-disk sampling is a process that distributes uniform random point samples on a domain of $d$-dimensional space based on a minimum distance criterion between
3. DEFINING PDS USING PAIR CORRELATION FUNCTION

For Poisson processes, point locations are not correlated and, therefore, \( P(r) = \lambda dx \lambda dy \). This implies that for Poisson sampling \( G(r) = 1 \). Similarly, for PDS, due to the minimum distance constraint between the point sample pairs, we do not have any point samples in the region \( 0 < r < r_{\text{min}} \). Consequently, we define the PCF for Poisson disk sampling as a step function.

**Definition** Given the desired radius \( r_{\text{min}} \), Poisson disk sampling is defined in the PCF domain as

\[
G(r - r_{\text{min}}) = \begin{cases} 
0 & \text{if } r < r_{\text{min}} \\
1 & \text{if } r \geq r_{\text{min}}.
\end{cases}
\]

### 3.1. Power Spectral Density Derivation

In this section, we derive the power spectral density of the defined PDS based on Fourier properties. Let us denote the power spectral density by \( P(k) \) and PCF by \( G(r) \) respectively. We know that

\[
P(k) = 1 + \rho F(G(r) - 1)
\]

where \( \rho = N/V \), with \( V \) being volume of the sampling region, and \( F(.) \) denotes the \( d \)-dimensional Fourier transform. Note that, for the radially symmetric or isotropic functions, i.e., \( G(r) \) where \( r = ||r|| \), the above relationship simplifies to

\[
P(k) = 1 + \rho \left(2\pi\right)^d \frac{d}{2} \frac{d}{2} H_{\frac{d}{2} - 1} \left(r^\frac{d}{2} - 1 \left(G(r) - 1\right)\right) \quad (4)
\]

where

\[
H_v(f(r))(k) = \int_0^\infty r J_v(kr_f) r dr
\]

is the one dimensional Hankel transform of order \( v \) with \( J \) being the Bessel function. To derive the PSD of a step function, we first evaluate the Hankel transform of \( f(r) = (G(r) - 1) \) where \( G(r) \) is a step function.

\[
H_{\frac{d}{2} - 1} \left(r^\frac{d}{2} - 1 \left(G(r) - 1\right)\right) = \int_0^\infty r^\frac{d}{2} J_{\frac{d}{2} - 1}(kr) \left(G(r) - 1\right) dr
\]

Using this expression in (4),

\[
P(k) = 1 - \rho \left(\frac{2\pi r_{\text{min}}}{k}\right)^d \frac{d}{2} J_{\frac{d}{2}} \left(k r_{\text{min}}\right). \quad (5)
\]

4. THEORETICAL BOUNDS FOR PDS CREATION

Although the uniform randomness and disk constraints are sufficient to specify a valid PDS, there is an additional condition that characterizes a distribution as being maximal.

A maximal Poisson disk distribution is one where it is not possible to insert any further samples without violating the minimum distance constraint, i.e.,

\[
\forall x \in D, \exists x_i \in X : ||x - x_i|| < r_{\text{min}}.
\]
Our definition of PDS in the PCF domain enables us to obtain new insights into the sampling process.

In particular, (a) For a fixed $r_{\min}$, we obtain the number of point samples needed to make the Poisson disk sampling maximal in arbitrary dimension $d$. (b) For a fixed sampling budget $N$, we derive the maximum achievable $r_{\min}$ in arbitrary dimension $d$. Existing approaches for maximal-PDS [3] experimentally obtain the bounds for $N$, by sequentially adding samples until the disk condition is violated. Surprisingly, we find that none of those bounds are theoretically accurate, and more importantly such maximal-PDS samples can cause aliasing.

Before presenting our main results, we state the two necessary mathematical conditions that the PCF of a sampling pattern must satisfy to be realizable: (a) The pair correlation function of the sampling pattern must be non-negative, i.e., $G(r) \geq 0$, $\forall r$, and (b) The PSD of the sampling pattern must be non-negative, i.e., $P(k) \geq 0$, $\forall k$. Note that, these conditions limit the range of realizable PCFs.

**Finding Maximum $N$ for a fixed $r_{\min}$**: The problem of finding maximum number of point samples for PDS with a given disk radius $r_{\min}$ can be formalized as follows:

\[
\begin{align*}
\text{maximize} & \quad N \\
\text{subject to} & \quad P(k) \geq 0, \forall k \\
& \quad G(r - r_{\min}) \geq 0, \forall r
\end{align*}
\]

where $P(k) = 1 - \rho \left( \frac{2\pi r_{\min}}{k} \right)^{\frac{d}{2}} J_{\frac{d}{2}} (kr_{\min})$.

**Proposition 1** For a fixed Poisson disk radius $r_{\min}$, the maximum number of point samples needed for maximal Poisson disk sampling in the sampling region with volume $V$ is given by

\[
N = \frac{V}{\pi d^d r_{\min}^d}.
\]

**Proof** Using the definition of the step function, the constraint $G(r - r_{\min})$ is trivially satisfied. Note that, the constraint $P(k) \geq 0$, $\forall k$ is equivalent to $\min_k P(k) \geq 0$. In other words,

\[
\begin{align*}
& \min_k 1 - \rho \left( \frac{2\pi r_{\min}}{k} \right)^{\frac{d}{2}} J_{\frac{d}{2}} (kr_{\min}) \geq 0 \\
\iff & \quad \max_k \rho \left( \frac{2\pi r_{\min}}{k} \right)^{\frac{d}{2}} J_{\frac{d}{2}} (kr_{\min}) \leq 1 \\
\iff & \quad \rho \left( \frac{2\pi}{d} \right)^{\frac{d}{2}} r_{\min}^d \max_k \left( \frac{J_{\frac{d}{2}} (kr_{\min})}{(kr_{\min})^{\frac{d}{2}}} \right) \leq 1 \\
\iff & \quad \rho \left( \frac{2\pi}{d} \right)^{\frac{d}{2}} r_{\min}^d \frac{1}{2^{\frac{d}{2}} \Gamma \left( \frac{d}{2} + 1 \right)} \leq 1 \\
\iff & \quad N \leq V \left( d/2 + 1 \right) / \left( \pi (\pi^{\frac{d}{2}} r_{\min}^{\frac{d}{2}}) \right)
\end{align*}
\]

where, we have used the fact that $J_{\nu}(x) \approx (x/2)^\nu / \Gamma(\nu + 1)$.

Note that, for the 2-dimensional case, we have $J_0(\frac{kr_{\min}}{r_{\min}}) = jinc(kr_{\min})$.

Now using the fact that $jinc(x)$ has the maximum value equal to 1/2, for a fixed Poisson disk radius $r_{\min}$, the maximum number of point samples needed for maximal 2-d Poisson disk sampling is given by

\[
N = \frac{V}{\pi (r_{\min})^2},
\]
which again corroborates our bound in Proposition 1.

**Finding Maximum \( r_{\text{min}} \) for a fixed \( N \):** Alternately, we can also derive the bound for maximum Poisson disk radius with a fixed sampling budget \( N \) as follows:

\[
\begin{align*}
\text{maximize} & \quad r_{\text{min}} \\
\text{subject to} & \quad P(k) \geq 0, \quad \forall k \quad G(r - r_{\text{min}}) \geq 0, \quad \forall r
\end{align*}
\]

**Proposition 2** For a fixed sampling budget \( N \), the maximum Poisson disk radius \( r_{\text{min}} \) for Poisson disk sampling in the sampling region with volume \( V \) is given by

\[
r_{\text{min}} = \sqrt[3]{\frac{V(\frac{r}{2} + 1)}{\pi^2 N^2}}.
\]

**Proof** The proof is similar to the one in Proposition 1.

## 5. RESULTS AND DISCUSSION

### 5.1. Observations on PDS generation

We used the maximal-PDS algorithm proposed in \cite{3} to synthesize point samples for \( r_{\text{min}} = 0.01 \). We found that the algorithm terminated at \( N = 7054 \) samples, beyond which no more samples could be added without violating the disk condition. From Proposition 1, we find the upper bound \( N^* = 3185 \) for the proposed PDS distribution. In Figure 2, the sampling patterns generated at different iterations of this sequential algorithm are shown, along with their corresponding PCFs. In addition, we show Poisson samples (without the disk condition) and samples on the regular grid. Firstly, it can be seen that as we move away from Poisson sampling to regular grid sampling, the uniformity or randomness in the point sampling pattern decreases, which appears as a peak close to \( r_{\text{min}} \) in the PCF. Secondly, at the proposed sample bound (Fig. 2(b)), the PCF follows the ideal PDS definition, while compromising on the uniformity as the algorithm adds more samples. Surprisingly, none of the existing algorithms produce a theoretically correct PDS, the consequences of which will be discussed in the next section.

Since the proposed sample bound can potentially guarantee an accurate PDS, we can modify the MPDS algorithm \cite{3} to terminate earlier than required, i.e., when the sample size reaches \( N^* \). In Table 1 we compare the performance of the algorithm by measuring the run time (in seconds) to generate the proposed PDS in different dimensions \((d = 2, 3, 4)\), in comparison to the actual approach. It can be seen from Table 1 that later samples (after \( N^* \)), added for maximality in MPDS, are computationally more expensive to compute than the earlier samples. Furthermore, the run time performance deteriorates significantly as we increase the number of dimensions.

### 5.2. Empirical Analysis of Quality of Sampling Patterns

To assess the quality of the sampling patterns, one can analyze the spectral behavior of the sampling patterns. More specifically, the power spectral density of an ideal sampling pattern should satisfy the following two properties \cite{8}: (a) the spectrum should be close to zero for low frequencies which indicates the range of frequencies that can be represented with almost no aliasing, (b) the spectrum should be a constant for high frequencies or contain minimal amount of oscillations in the power spectrum which reduces the risk of aliasing. Based on these two criteria, we assess the quality of the PDS in the Fourier domain. To emulate the approximate PDS as observed with maximal-PDS algorithms, we model its PCF as

\[
G(r) = G(r - r_{\text{min}}) + a (G(r - r_{\text{min}}) - G(r - r_{\text{min}} - \delta))
\]

where \( G(r - r_{\text{min}}) \) is the step function, \( \delta \geq 0 \) and the peak height \( a \geq 1 \). We calculate the power spectral density (PSD) of the sampling patterns using Hankel transform relationship as given in (4).

In order to study the impact of compromising uniformity in the PCF, at the risk of aliasing, we evaluated the PSD of the approximate PDS with \( N = 195, r_{\text{min}} = 0.02, \delta = .005 \). Note that, we varied the PCF peak height \( a \) which reflects the behavior of existing sequential PDS algorithms. Surprisingly, as shown in Figure 3, increasing \( a \) results in both significantly higher low frequency aliasing and larger high frequency oscillations. As expected, the PSD of the ideal PCF proposed in this paper (or \( a = 1 \)) performs the best, i.e., the spectrum is close to zero for low frequencies and constant for high frequencies as required in an ideal sampling pattern.

Finally, we study the importance of choosing an appropriate \( r_{\text{min}} \) while generating PDS distributions. Though the general intuition is to choose the largest possible \( r_{\text{min}} \) for a given \( N \) (Proposition 2), our analysis reveals an interesting trade-off. In Figure 4, we plot PSD of the approximate PDS sampling pattern for \( N = 195 \) with varying PDS radius \( r_{\text{min}} \). For a fixed sample budget, as we increase the radius from a lower value towards its upper bound 0.04, we observe two contrasting changes to the PSD: 1) the spectrum tends to be close to zero at low frequencies and 2) significant increase in oscillations for high frequencies. Thus, there is a trade-off between low frequency aliasing and high frequency oscillations which can be controlled by varying \( r_{\text{min}} \) depending on the application of interest.

**Table 1.** The sample sizes and run time (in seconds) required to generate PDS in comparison to the approximate PDS \cite{3}

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Radius</th>
<th>Algorithm in [3]</th>
<th>Proposed PDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( N )</td>
<td>Time (s)</td>
<td>( N^* )</td>
</tr>
<tr>
<td>2</td>
<td>0.005</td>
<td>28098</td>
<td>14.34</td>
</tr>
<tr>
<td>2</td>
<td>0.007</td>
<td>14361</td>
<td>7.34</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>7054</td>
<td>3.51</td>
</tr>
<tr>
<td>3</td>
<td>0.03</td>
<td>28776</td>
<td>182.7</td>
</tr>
<tr>
<td>3</td>
<td>0.04</td>
<td>12384</td>
<td>17.55</td>
</tr>
<tr>
<td>4</td>
<td>0.1</td>
<td>10779</td>
<td>1154.59</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>849</td>
<td>18.22</td>
</tr>
</tbody>
</table>
6. REFERENCES


