Eurographics Symposium on Rendering 2026
I. Gkioulekas and A. Jarabo
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 45 (2026), Number 4

Optimized and Aligned Anisotropic Monte Carlo Sampling Patterns

Mirco Werner, Johannes Hanika, and Carsten Dachsbacher

Karlsruhe Institute of Technology, Germany

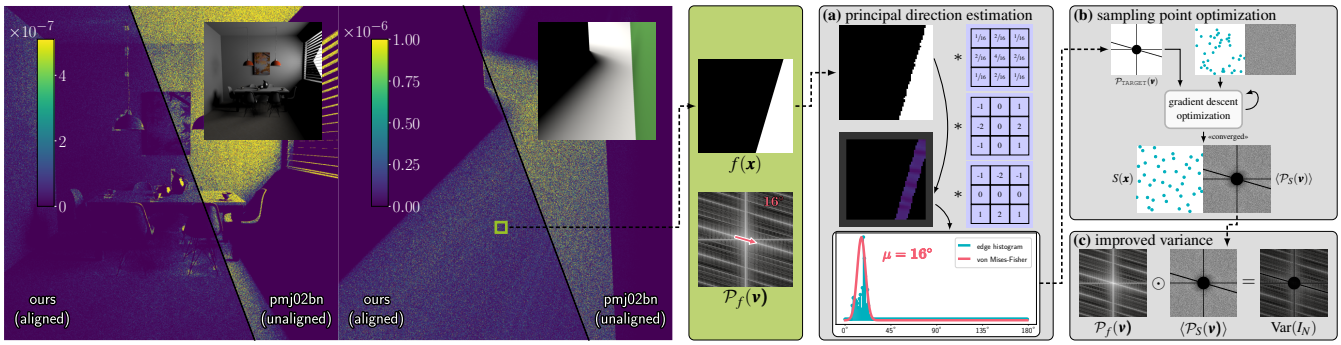


Figure 1: We improve the variance convergence rate of the Monte Carlo estimator in offline path tracing on two-dimensional discontinuous integrands. To that end, we (a) sparsely sample the integrand $f(\boldsymbol{x})$ in blocks of pixels to construct a histogram of occurring edge directions. We estimate principal directions in the power spectrum $\mathcal{P}_f(\boldsymbol{v})$ (red arrow) from the histogram, as principal directions in $\mathcal{P}_f(\boldsymbol{v})$ are perpendicular to edges in $f(\boldsymbol{x})$. (b) We optimize a set of sampling points $S(\boldsymbol{x})$ with an anisotropic spectrum that matches the estimated spectrum of the integrand. (c) Using our aligned points improves the convergence rate compared to existing unaligned anisotropic sampling patterns.

Abstract

Path tracing uses Monte Carlo integration to solve the rendering equation by evaluating the integrand at random sampling points. The convergence rate of the error can be significantly improved by using correlated instead of random sampling, especially on smooth integrands. However, on integrands with discontinuities due to, e.g., occlusion, the improvement is less pronounced. Prior work has shown that the variance of the estimator is equal to the product of the power spectrum of the integrand and the expected power spectrum of the sampling pattern. Discontinuous integrands have anisotropic power spectra that exhibit high energies along the directions of the discontinuities, which need to match the low-energy directions of the sampling pattern to reduce variance. However, existing anisotropic sampling patterns have at most two low-energy directions. Therefore, we propose an optimization-based algorithm to synthesize two-dimensional correlated sampling patterns with spectra that have more than two low-energy directions, leading to improved convergence behavior. Further, we propose a practical and sample-efficient algorithm that estimates the directions of discontinuities in the power spectra of two-dimensional integrands. We show that our algorithm can reliably estimate these directions, allowing us to align the low-energy directions of anisotropic correlated sampling patterns with the predicted directions. We demonstrate in an offline path tracer with light source sampling that our aligned sampling patterns improve the convergence rate on two-dimensional integrands with multiple discontinuities compared to existing anisotropic sampling patterns and thus reduce the error more quickly.

CCS Concepts

• ***Computing methodologies*** → ***Ray tracing;***

1. Introduction

Monte Carlo integration is a numerical method which computes integrals by evaluating the integrand stochastically at various random sampling points. For an unbiased estimator, the goal is to minimize the variance when using a certain number of samples to get an accurate estimate. In many applications, such as offline rendering, the integral is estimated using a large number of samples. For random sampling, the variance of the Monte Carlo estimator converges with a suboptimal rate of $\mathcal{O}(N^{-1})$. The convergence rate can be

significantly improved using correlated sampling [SOA*19]. Many methods focus on generating sampling patterns that introduce correlations into the sampling points by maximizing their distance to be evenly distributed in the domain, which already improves the variance and its convergence rate in the case of smooth integrands. However, these methods usually work independently of the actual integrand.

Prior work [Dur11, SK13, PSC*15, SSC*20] uses Fourier analysis to show that the variance of the Monte Carlo estimator can be expressed as the product of the power spectrum of the integrand and the expected power spectrum of the sampling pattern. This means that variance is reduced when frequencies with high energy in the power spectrum of the integrand match with frequencies with low energy in the expected power spectrum of the sampling pattern. The power spectra of typical integrands in rendering are usually highly anisotropic. Also, many sampling patterns used for Monte Carlo integration, e.g., multi-jittered [CSW94], Sobol [Sob67], or Halton [Hal64], have anisotropic spectra.

Existing spectra of anisotropic sampling patterns feature only two directions with low energy, often as (accidental) by-product of stratification with intervals. We propose an offline optimization-based technique that synthesizes two-dimensional sampling sets with anisotropic power spectra matching a provided target spectrum while ensuring stratification along *a set* of directions (Fig. 1 (b)). Our technique enables us to precompute sampling points that have more than two low-energy directions, matching the anisotropies in the spectra of typical discontinuous two-dimensional integrands that have more than two high-energy directions (cf. Fig. 2).

Singh et al. [SJ17] propose to align the anisotropic power spectrum of the sampling pattern with the anisotropic spectrum of the integrand and demonstrate an improved convergence rate, reducing the variance in offline rendering. Their method is based on *shearing* an existing point set with two low-energy directions in the spectrum, to align one of the two with the integrand. This method relies on Fourier analysis of the integrand, which is generally unknown before rendering.

We propose a practical and sample-efficient algorithm to perform Fourier analysis of two-dimensional discontinuous integrands: Our algorithm reliably estimates the principal directions in the anisotropic spectra of two-dimensional integrands from a set of sparse samples of the integrand (Fig. 1 (a)). To that end, we use edge detection to construct a histogram of occurring directions in the integrand. We propose to extract the principal directions from the noisy histogram by optimizing von Mises-Fisher distributions to minimize a mode-seeking reverse Kullback-Leibler divergence.

By aligning our sampling points with the estimated principal directions of the integrand, we can reduce variance in offline path tracing (Fig. 1 (c)). We evaluate our techniques on two-dimensional discontinuous integrands. A discontinuity carries energy in all frequencies. Monte Carlo integration will require an exceedingly large number of samples to resolve this. Aligning anisotropic point sets will improve *the rate* of convergence, i.e., the slope in a log/log plot. For a sufficiently large number of samples to resolve high-variance integrands, this will thus always beat unaligned methods.

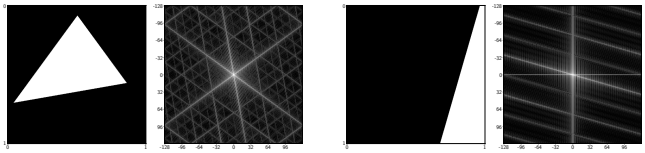


Figure 2: Two integrands with discontinuities and their respective power spectra. The power spectra show high-energy directions perpendicular to the orientation of the discontinuity. Due to the finite integration domain and as the Fourier transformation operates on a toroidally, the second integrand also has discontinuities along the horizontal and vertical axes that are visible in the power spectrum.

We show that our system improves the rendering error for a moderate number of samples already.

To summarize, our main contributions are:

- An offline optimization-based algorithm to synthesize two-dimensional anisotropic sampling sets whose anisotropic power spectra match the anisotropies in the spectra of typical discontinuous integrands (Section 4).
- A practical and sample-efficient algorithm that estimates the principal directions in the anisotropic power spectra of two-dimensional discontinuous integrands (Section 5).
- A reduced variance of the Monte Carlo estimator in offline path tracing, due to an improved convergence rate, by using our precomputed sampling points whose anisotropic power spectra, which have up to four low-energy directions, align with our estimated integrand spectra (Section 6).

In the next sections, we review Fourier analysis for Monte Carlo integration (Section 2) and provide an overview of related work (Section 3).

2. Background

Consider the multidimensional definite integral

$$I = \int_{[0,1)^d} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{1}$$

over the d-dimensional domain $[0,1)^d$ with $f : [0,1)^d \to \mathbb{R}$. The Monte Carlo estimator for Eq. 1 using N uniformly distributed ($p(\boldsymbol{x}) = 1$) random samples at the locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in [0,1)^d$ is

$$I_N = \frac{1}{N} \sum_{i=1}^{N} \frac{f(\boldsymbol{x}_i)}{p(\boldsymbol{x}_i)} \overset{p(\boldsymbol{x})=1}{=} \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}_i). \tag{2}$$

For random (uncorrelated) sampling, the variance of the Monte Carlo estimator converges to zero at a suboptimal rate of $\mathcal{O}(N^{-1})$. By carefully introducing correlations between sampling points, the convergence rate can be significantly improved. To that end, the sampling points are typically distributed more evenly in $[0,1)^d$ during construction, e.g., by maximizing the distance between neighboring sampling points (blue noise) or by subdividing the sampling domain into strata (stratified sampling).

Prior work [Dur11, SK13, PSC*15, SSC*20] has introduced

Fourier analysis for Monte Carlo integration to predict the variance of the estimator based on the correlations between the sampling points and the structure of the integrand. Therefore, the Monte Carlo estimator (Eq. 2) is first rewritten in continuous form as

$$I_N = \int_{[0,1)^d} S(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{3}$$

$S(\boldsymbol{x})$ is the sampling function and is defined as a sum of Dirac delta distributions with positive densities at the sampling locations:

$$S(\boldsymbol{x}) = \frac{1}{N} \sum_{i=1}^{N} \delta(\|\boldsymbol{x} - \boldsymbol{x}_i\|). \tag{4}$$

Then, for *homogeneous* sampling patterns, i.e., sampling patterns with statistical properties invariant to translation over the sampling domain [PSC*15], the variance can be expressed in the frequency domain as [SSC*20]

$$\mathrm{Var}(I_N) = \sum_{\mathbf{v} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \langle \mathcal{P}_S(\mathbf{v}) \rangle \, \mathcal{P}_f(\mathbf{v}), \tag{5}$$

using the power spectra $\mathcal{P}_\cdot(\mathbf{v}) = \mathcal{F}_\cdot(\mathbf{v}) \overline{\mathcal{F}_\cdot(\mathbf{v})} \in \mathbb{R}_{\geq 0}$ and Fourier coefficients $\mathcal{F}_\cdot(\mathbf{v})$ given by the respective Fourier transforms:

$$\mathcal{F}_f(\mathbf{v}) = \int_{[0,1)^d} f(\boldsymbol{x}) e^{-i2\pi \mathbf{v} \cdot \boldsymbol{x}} \mathrm{d}\boldsymbol{x} \quad \text{and} \quad \mathcal{F}_S(\mathbf{v}) = \frac{1}{N} \sum_{i=1}^{N} e^{-i2\pi \mathbf{v} \cdot \boldsymbol{x}_i}. \tag{6}$$

Eq. 5 shows that the variance of I_N is equal to the product of the expected power spectrum of the sampling pattern (as individual realizations of a sampling pattern contain randomness) and the power spectrum of the integrand summed over all integer frequencies except the zero frequency. As a consequence, variance is reduced when high-energy regions in the spectrum of the integrand align with low-energy regions in the spectrum of the sampling pattern (cf. Fig. 1 (c)). The convergence rate of the Monte Carlo estimator is determined by the direction with the highest energies in the product of the power spectrum of the integrand and the power spectrum of the sampling pattern [SJ17]. In the case of stratified sampling, variance is reduced if the number of strata M, and thus the number of evenly distributed samples, is increased along the high-energy directions of the integrand. If M is increased only by a constant factor, the variance is also improved only by a constant factor. Otherwise, if M depends on the number of points, an improved variance convergence is possible [SJ17].

The varying convergence behavior becomes most prominent on discontinuous integrands: These exhibit high energies at infinitely many frequencies along the corresponding direction in their power spectra (cf. Fig. 2, left). Due to the finite integration domain $[0,1)^2$, many integrands feature discontinuities along the horizontal and vertical axes (cf. Fig. 2, right). In practice, integrands often feature additional arbitrarily oriented discontinuities due to, e.g., occlusion, which leads to more than two high-energy directions. As existing anisotropic sampling patterns are only stratified along the horizontal and vertical axes, they have only two low energy directions (cf. Fig. 5) resulting in a poor convergence rate of only $\mathcal{O}(N^{-1.5})$ on discontinuous integrands compared to at least $\mathcal{O}(N^{-2})$ on smooth integrands.

In path tracing, a path $\bar{\mathbf{y}} \in \mathcal{P}$ is generated from a d-dimensional vector of (correlated) sampling points $\boldsymbol{x} \in [0,1)^d$ with uniform density using a mapping $t : [0,1)^d \to \mathcal{P}$ with $t(\boldsymbol{x}) = \bar{\mathbf{y}}$ and density $p_t(\bar{\mathbf{y}})$. This is expressed as [KSKAC02]

$$I_q = \int_{\mathcal{P}} f_q(\bar{\mathbf{y}}) \mathrm{d}\bar{\mathbf{y}} = \int_{[0,1)^d} f_q(t(\boldsymbol{x})) |J_t(\boldsymbol{x})| \mathrm{d}\boldsymbol{x} = \int_{[0,1)^d} \frac{f_q(t(\boldsymbol{x}))}{p_t(t(\boldsymbol{x}))} \mathrm{d}\boldsymbol{x}, \tag{7}$$

where $f_q(\bar{\mathbf{y}})$ is the measurement contribution function in pixel q and $|J_t(\boldsymbol{x})| = \frac{1}{p_t(t(\boldsymbol{x}))}$ the Jacobian determinant of the mapping t. The Monte Carlo estimator is

$$\hat{I}_{q,N} = \frac{1}{N} \sum_{i=1}^{N} \frac{f_q(t(\boldsymbol{x}_i))}{p_t(t(\boldsymbol{x}_i)) p(\boldsymbol{x}_i)} \overset{p(\boldsymbol{x})=1}{=} \frac{1}{N} \sum_{i=1}^{N} \frac{f_q(\bar{\mathbf{y}}_i)}{p_t(\bar{\mathbf{y}}_i)}. \tag{8}$$

Eq. 8 shows that by defining $f(\boldsymbol{x}) := \frac{f_q(t(\boldsymbol{x}))}{p(t(\boldsymbol{x}))}$ allows us to analyze the variance of the estimator in path tracing for the correlated sampling points and the (distorted) integrand $f(\boldsymbol{x})$ using Eq. 5. For the remainder of this paper, we will analyze and improve the variance convergence rate of the Monte Carlo estimator on two-dimensional discontinuous integrands, i.e., $d = 2$. In the context of path tracing, this means that we will evaluate our optimized and aligned sampling points on two-dimensional light transport problems, namely direct illumination, mapping two random numbers either to a point on an area light source (light source sampling) or to an outgoing direction (BSDF sampling), and depth of field, mapping two random numbers to a point on the aperture. We also present results for slightly smoothed discontinuities, which may occur due to, e.g., subpixel jittering.

3. Related Work

Correlated sampling and error analysis is an active research topic [SOA*19]. In this section, we provide an overview of related work.

Correlated Sampling and Point Synthesis Blue noise sampling patterns [Coo86, Mit91, Bri07, LD08, Fat11, dGBOD12, HSD13, PBC*20, SGSS22] introduce correlation by maximizing the distance between neighboring sampling points. Stratified sampling [Coo86, Shi91, CSW94, Ken13, KCSG18, JEK*19] subdivides the sampling domain into strata and places one sample in each stratum at a pseudo-random position. Recent work also proposes sampling patterns that combine blue noise with stratification [RRSG16, APC*16, ANHD17, PCX*18, CKK18]. Instead of using pseudo-random numbers, quasi-Monte Carlo sampling patterns [Hal64, Sob67] deterministically generate low-discrepancy sequences of sampling points. As quasi-Monte Carlo sequences are deterministic, they need to be randomized to be able to analyze and estimate the variance and to be unbiased. For this, toroidal shifts [CP76] or random digit scrambling [Bur20, Owe95] can be applied.

There also exist optimization-based sampling patterns that synthesize points with desired spectral properties given a target power spectrum as input: Leimkühler et al. [LSM*19] train the parameters of a sequence of filters that, after training, transform sets of points so that their spectra match the given power spectrum. The gradient-based optimization is driven by an objective function that determines the difference between the current spectrum of the points and the target spectrum. While this has the advantage of being able

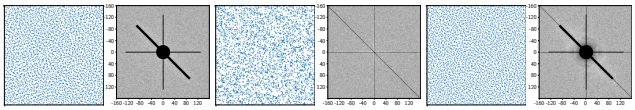
to train once and then quickly synthesize any number of realizations by starting with different random initialization, they report training times of several hours. Note that using their proposed architecture would require training one network for each anisotropy direction. We also introduce an offline gradient-based point synthesis technique. However, we directly optimize the sampling point locations, which allows us to obtain a realization of the sampling pattern with the desired low-energy directions faster. Additionally, we extend the optimization objective to ensure stratification along a set of directions, further reducing variance. Instead of optimizing based on an objective defined in the frequency domain, it is also possible to optimize the distribution of point location differences in the spatial domain [HSD13, ZHWW12]. Therefore, the input target power spectrum needs to be transformed to a differential distribution function [WW11] that only depends on location differences. During optimization, the differential distribution function has to be approximated from the current point locations. For our goal of suppressing frequencies along multiple directions, we found that these methods are harder to control and were only able to synthesize anisotropic points with less strongly suppressed frequency components compared to our method.

Fourier Analysis of Monte Carlo Integration The impact of different sampling patterns on the error and the convergence rate has been investigated by many researchers by analyzing correlated sampling patterns and integrands in the frequency domain [DHS*05,Dur11,SK13,PSC*15,SJ17,SSC*20]. Singh et al. [SJ17] analyze the variance convergence behavior of Monte Carlo estimators when using sampling patterns with anisotropic power spectra. They propose the general idea of first using Fourier analysis of the integrand to determine the directions with the highest energy, and subsequently using an anisotropic sampling pattern whose low-energy directions match the anisotropic spectrum of the integrand to improve the convergence rate. They demonstrate their idea on integrands that have high-energies mainly along one direction. By shearing an existing anisotropic point set with two low-energy directions, either horizontally or vertically, to match the high-energy direction, Singh et al. [SJ17] demonstrate an improved convergence rate. We start from the same general idea and show that using our optimized and aligned points with more than two low-energy directions further improves the convergence behavior on two-dimensional integrands with multiple discontinuities. Additionally, we propose a practical technique to perform Fourier analysis of two-dimensional discontinuous integrands to estimate directions with high energy, where no closed-form Fourier analysis is available. By this, our paper confirms the previously mostly theoretical analysis of Singh et al. [SJ17] and extends their idea to other rendering scenarios.

4. Point Synthesis With Anisotropic Power Spectra

The goal of our point synthesis algorithm is to determine the locations of a set of N uniformly distributed points $S \in [0,1)^{N\times 2}$ that introduce correlations in such a way that the power spectrum of S exhibits certain desired low-energy patterns (cf. Fig. 1 (b)). Mathematically, this is expressed as an optimization problem that determines the set of points S that minimizes a loss function $\mathcal{L}$:

$$S = \underset{S' \in [0,1)^{N\times 2}}{\arg\min}\ \mathcal{L}(S'). \tag{9}$$



(a) $\mathcal{L}_{\mathtt{spec}}(S)$ (b) $\mathcal{L}_{\mathtt{SW}}(S)$ (c) $\mathcal{L}_{\mathtt{spec}}(S)+c\mathcal{L}_{\mathtt{SW}}(S)$

Figure 3: Three synthesized point sets and their respective power spectra, optimized for different loss functions using our gradient-based optimization. (a) The power spectrum loss $\mathcal{L}_{\mathtt{spec}}(S)$ suppresses a finite set of frequencies. (b) The sliced Wasserstein loss $\mathcal{L}_{\mathtt{SW}}(S)$ ensures stratification for a given set of directions. (c) We combine both losses to synthesize points with both properties.

Our loss function is composed of two terms: A term comparing the power spectrum of S to a target power spectrum to suppress a set of frequencies $\mathcal{L}_{\mathtt{spec}}(S)$ (Section 4.1), and a term that ensures stratification along a predefined set of directions $\mathcal{L}_{\mathtt{SW}}(S)$ (Section 4.2). We explain both terms of our combined loss function in the following sections. They are combined as

$$\mathcal{L}(S) = \mathcal{L}_{\mathtt{spec}}(S) + c\mathcal{L}_{\mathtt{SW}}(S), \tag{10}$$

where the scalar $c \in \mathbb{R}_{>0}$ is used to balance the two terms. In Fig. 3, we show point sets and their respective power spectra when using our optimization with one (a, b) or both (c) losses.

4.1. Power Spectrum Loss

Similar to the work of Leimkühler et al. [LSM*19], the loss $\mathcal{L}_{\mathtt{spec}}(S)$ calculates the deviation of the power spectrum of the point set S from a target power spectrum. It is defined as the mean element-wise squared distance over the power spectrum $\mathcal{P}_S$ of S and the target power spectrum $\mathcal{P}_{\mathtt{TARGET}}$ evaluated at a finite set of frequencies $\Lambda \subset \mathbb{Z}^2$:

$$\mathcal{L}_{\mathtt{spec}}(S) = \frac{1}{|\Lambda|}\sum_{\mathbf{v}\in\Lambda} \|\mathcal{P}_S(\mathbf{v}) - \mathcal{P}_{\mathtt{TARGET}}(\mathbf{v})\|^2. \tag{11}$$

Since the computation of $\mathcal{P}_S$ is relatively expensive, we evaluate the loss only on frequency components that we want to suppress. In regions ignored in the optimization, the final power spectrum $\mathcal{P}_S$ will empirically have sufficient white noise properties (see Fig. 3), except in cases where the number of points is too small to realize the desired spectrum. Additionally, due to the symmetry of the power spectrum, the loss only has to be evaluated for half of the frequencies of interest. It is also conceivable to use a mask to weight the influence of the individual frequencies, e.g., to slowly fade out the low-energy region at the edges (at higher frequencies).

4.2. Sliced Wasserstein Loss

Our spectrum loss effectively suppresses frequencies over a finite length along the directions with a possible small extent along each of them (cf. the black stripes with finite length in the power spectrum in Fig. 3 (a)). This, however, does not necessarily ensure that

the number of samples along the directions is increased proportionally to the number of points in the set. Therefore, the goal of our second loss term is to ensure stratification for a given set of directions $\mathcal{A}$ (cf. the narrow black stripes with increased length in the power spectrum in Fig. 3 (b)). $\mathcal{A}$ will later correspond to the set of high-energy directions detected in the spectrum of the integrand.

As it is generally difficult to determine the target point locations that ensure stratification directly in multiple dimensions, we rely on optimizing the point locations individually in one-dimensional subspaces similar to prior work [PBC*20, HVCB21, SGSS22]. Individually, along each direction, the target locations of the (projected) points that ensure stratification can be calculated. Formally, we use the sliced Wasserstein distance that is defined as the expectation over one-dimensional optimal transport distances for projections of the points onto random directions $\boldsymbol{d}$:

$$\mathcal{L}_{\texttt{SW}}(S) = \mathbb{E}_{\boldsymbol{d}}\left[\mathcal{L}_{\texttt{SW1D}}(S,\boldsymbol{d})\right]. \tag{12}$$

As only one-dimensional optimal transport needs to be computed, this allows for efficient gradient-based optimization [HVCB21]. Determining the cost of the one-dimensional optimal transport is achieved by projecting the points in S onto the direction $\boldsymbol{d}$, sorting the projected positions, and computing the mean element-wise squared distance between the sorted projected positions and the target positions $D_{\text{TARGET}}(\boldsymbol{d})$ [HVCB21]:

$$\mathcal{L}_{\texttt{SW1D}}(S,\boldsymbol{d}) = \frac{1}{N}\|\texttt{sort}(S\cdot\boldsymbol{d}) - D_{\texttt{TARGET}}(\boldsymbol{d})\|^2. \tag{13}$$

In our case, where the set of directions $\mathcal{A}$ is small, we can compute the exact loss $\mathcal{L}_{\texttt{SW}}(S)$ during optimization:

$$\mathcal{L}_{\texttt{SW}}(S) = \frac{1}{|\mathcal{A}|}\sum_{\boldsymbol{d}\in\mathcal{A}}\mathcal{L}_{\texttt{SW1D}}(S,\boldsymbol{d}). \tag{14}$$

To ensure that the points still have uniform density in the unit square, the target positions $D_{\texttt{TARGET}}(\boldsymbol{d})$ must be distributed proportional to the height of the square projected onto the rotated direction $\boldsymbol{d}$ [Wol23]. Given the angle $\theta = \arccos(\boldsymbol{d}_0)$, the target positions $D_{\texttt{TARGET}}(\boldsymbol{d})$ are obtained by mapping equidistant points $u \in [-0.5, 0.5]$ using

$$F_\theta^{-1}(u) = \begin{cases} u\cos\theta, & \text{if } |u| < \frac{\cos\theta-\sin\theta}{2\cos\theta} \\ \frac{\cos\theta+\sin\theta}{2} - \sqrt{2\sin\theta\cos\theta\left(\frac{1}{2}-|u|\right)}, & \text{else if } u > 0 \\ \sqrt{2\sin\theta\cos\theta\left(\frac{1}{2}-|u|\right)} - \frac{\cos\theta+\sin\theta}{2}, & \text{else if } u < 0. \end{cases} \tag{15}$$

A complete derivation of $F_\theta^{-1}(u)$ can be found in the supplemental material.

While $\mathcal{L}_{\texttt{SW}}$ alone ensures correct marginal distributions of the points, it allows for clusters and holes to form in 2D (cf. Fig. 3 (b)). $\mathcal{L}_{\texttt{spec}}$ alone ensures evenly, blue-noise-like distributed points in 2D and suppresses a set of selected frequencies very effectively, but it cannot enforce directional stratification cheaply, and therefore fails to suppress increasingly higher frequencies (cf. Fig. 3 (a)).

5. Fourier Analysis of Two-Dimensional Discontinuous Integrands

To align our set of sampling points, we first identify directions in the power spectrum of the two-dimensional discontinuous integrand that exhibit high energy. We will refer to them as principal directions. In almost all two-dimensional discontinuous integrands, the spectrum contains high energies along the horizontal (0°) and vertical (90°) axes due to the boundary of the finite integration domain (cf. Fig. 2). Another cause might be occlusion boundaries of man-made structures that often align horizontally and vertically, too. Therefore, we always use 0° and 90° as the two principal directions that we, among other directions, want to align with and suppress frequencies during optimization. Additional principal directions in the spectrum are closely related to the directions with the highest local variation in the integrand, i.e., they are perpendicular to the edges (cf. Fig. 2). Our idea is to sparsely sample the integrand, apply an edge detection filter to obtain the orientation of the edges, bin the orientations in a histogram, and finally estimate the principal direction(s) from the histogram (cf. Fig. 1 (a)). As anisotropic sampling patterns can only feature a small number of low-energy directions, aligning is only meaningful if there is only a small number of straight discontinuities in the integrand. This is typically the case, for example, when sampling rectangular light sources with visibility blocked by a few occluders. We discuss how to detect cases where alignment is not useful in Section 5.2, and investigate an approach to reduce the distortion introduced by hemisphere sampling to keep edges straight when projected to $[0,1)^2$ in Section 6.3.

5.1. Direction Histogram

To create the direction histogram, we first sample the integrand on a regular grid and smooth the samples using a Gaussian filter. Then, we apply vertical and horizontal Sobel filters for gradient estimation, and bin the gradient magnitude into a histogram using the discretized direction of the gradient as a key, e.g., using one-degree bins. Of course, other common edge detection algorithms can be used as long as they provide the orientation of detected edges. We show an example histogram for an integrand with one edge in Fig. 1 (a) using 64^2 samples, and note that the featured mode in the histogram matches the principal direction in the power spectrum of the integrand closely. Before we estimate the principal directions from the histogram, we set the values around 0° and 90° to zero, as we always optimize our points to align with these two directions. We discuss the efficient cooperative construction of the histogram in blocks of pixels and the required number of samples for the direction histogram in Section 5.3.

5.2. Principal Direction Estimation

Given the direction histogram, we need to find the k most prominent principal direction(s), where $k+2$ is the number of directions for which we can optimize our point set ($+2$ since we always optimize for 0° and 90°). A naive solution uses the k maximum directions in the histogram. However, this solution is not robust against noise and can miss the correct direction by a few degrees or cover directions multiple times. As the principal directions should be as precise as

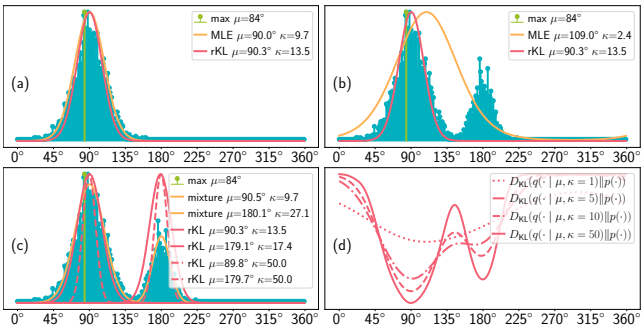


Figure 4: Comparison between different techniques to estimate the principal directions. The direction histogram is synthetically generated using samples from a mixture of vMF distributions with $(\mu = 90°, \kappa = 10)$, $(\mu = 180°, \kappa = 30)$, and weights $\frac{3}{4}$ and $\frac{1}{4}$. Our reverse KL optimization uses gradient descent to estimate both μ and κ. (a) When disabling one of the modes, MLE and our estimation show similar results. (b) With two modes, MLE shows a mode-averaging behavior while ours is mode-seeking. (c) To find multiple principal directions, a mixture with the correct number of components can be fitted, or multiple of our independently running estimations can be used. By using a fixed and large κ and optimizing only for μ (dashed), we reduce problems with overlapping modes that bias the mean location of the distribution. (d) The course of our objective w.r.t. μ varies for different values of κ.

possible to reduce variance, instead of using the maximum, we fit a set of von Mises-Fisher (vMF) distributions to the histogram. The density of a vMF distribution in two dimensions is given by

$$q(x \mid \mu, \kappa) = \frac{\exp(\kappa \cos(x-\mu))}{2\pi I_0(\kappa)}, \tag{16}$$

where $x, \mu \in [0, 2\pi]$, $\kappa > 0$, and $I_0(\kappa)$ is the modified Bessel function of the first kind.

Maximum Likelihood Estimation If the histogram features only a single mode, we can use maximum likelihood estimation (MLE) for estimating the parameters of the vMF (cf. Fig. 4 (a), yellow). However, this is limited to single-mode distributions: If there are multiple modes, MLE does not estimate the mean of either mode, but an average (cf. Fig. 4 (b), yellow). A mixture of vMFs can, in principle, fit multiple modes even when optimized via maximum likelihood (cf. Fig. 4 (c), yellow), but such optimization is expensive and prone to a poor fit in case of a wrong number of mixture components.

Reverse Kullback–Leibler Divergence The Kullback-Leibler (KL) divergence D_{KL} [KL51] measures the difference between two probability distributions q and p and is defined as

$$D_{KL}(q\|p) = \mathbb{E}_{x\sim q(\cdot)}\left[\log \frac{q(x)}{p(x)}\right], \tag{17}$$

with $D_{KL}(q\|p) \geq 0$ and $D_{KL}(q\|p) = 0 \Leftrightarrow q = p$. In our case, $q(x \mid \mu, \kappa)$ is the vMF distribution with unknown μ and κ, and $p(x)$ is the data distribution, i.e., the discrete normalized direction histogram. Minimizing $D_{KL}(q\|p)$ with respect to the parameters of q, forces q to cover only one of the modes of p [Bis06]. Otherwise, if q had high probability where p has low probability, i.e., between modes, $q(x)/p(x)$ would grow large. Consequently, the estimator has a mode-seeking behavior. In contrast, swapping the order of q and p in D_{KL} has a mode-averaging behavior. $D_{KL}(p\|q)$ is called forward KL divergence, while we minimize the reverse KL divergence $D_{KL}(q\|p)$ (Fig. 4 (b)). The reverse KL divergence is

$$\begin{aligned} &D_{KL}(q(\cdot \mid \mu,\kappa)\|p(\cdot)) \\ =&\mathbb{E}_{x\sim q(\cdot\mid\mu,\kappa)}\left[\log q(x \mid \mu,\kappa)\right] - \mathbb{E}_{x\sim q(\cdot\mid\mu,\kappa)}\left[\log p(x)\right] \\ =&\kappa\frac{I_1(\kappa)}{I_0(\kappa)} - \log(2\pi I_0(\kappa)) - \mathbb{E}_{x\sim q_d(\cdot\mid\mu,\kappa)}\left[\log p(x)\right]. \end{aligned} \tag{18}$$

In our case, the expectation simplifies to a sum since p, the direction histogram, is discrete. In Fig. 4 (d), we plot $D_{KL}(q\|p)$ for different values of κ and note that the local minima match with the desired locations of the vMFs. In general, minimizing $D_{KL}(q\|p)$ to obtain μ and κ involves gradient descent, requiring the gradients with respect to μ and κ. Please refer to our supplemental for the derivation. We show the result of fitting two vMFs by minimizing $D_{KL}(q\|p)$ with gradient descent in Fig. 4 (c) (red). By initializing the two vMFs near different local minima of the objective, i.e., at local maxima in the noisy direction histogram, and by using a small learning rate during gradient descent, they empirically converge to the respective local minimum.

In our case, we can apply two simplifications for minimizing $D_{KL}(q\|p)$: First, for our principal directions, we are only interested in the mean directions of the vMFs. This allows us to keep κ fixed and optimize $D_{KL}(q(\cdot \mid \mu,\kappa)\|p(\cdot))$ only for μ. By setting κ to some higher value, e.g., $\kappa \approx 50$, we can fit preferably high and narrow modes (cf. Fig. 4 (c), dashed). Second, by discretizing the values of μ, e.g., to integer angles, it becomes feasible to search for the minimum of the objective (Eq. 18) when keeping κ fixed. To that end, for the first vMF, we determine the highest value in the noisy direction histogram and search in a small window for the local minimum value of the objective. This process is repeated to determine the mean directions of the remaining $k-1$ vMFs. To find different local minima of the objective, we search near a different local maximum of the histogram by ensuring a minimum distance to previous vMFs.

In practice, not all integrands feature a few straight discontinuities. We need to detect such cases, as they are unsuitable for alignment, and revert to unaligned correlated sampling points. The concentration parameter κ of a vMF can be used as an indicator of how straight and pronounced the discontinuity is: A high κ indicates that the direction histogram contains only a few strongly pronounced directions, while a low κ indicates either many strongly pronounced directions, i.e., a curved discontinuity, or no discontinuities at all. As we do not optimize for κ in practice, we instead sum the normalized histogram values in a small window (we use $\pm 2°$) around μ, i.e., $\sum_{\mu-n}^{\mu+n} p(x_i)$, and compare it with a threshold to determine whether to use μ as a principal direction. This sum has a similar behavior as the actual unknown κ: κ measures how strongly the distribution is concentrated around the mean (the proposed principal direction). We can approximate the area under the curve, in a small window around the mean, by summing the noisy histogram entries. For a high κ, this area is larger compared to a low κ.

5.3. Practical Principal Direction Estimation in Blocks

Estimating the principal direction(s) requires a certain sampling resolution of the integrand for constructing the direction histogram. The order of magnitude at which reliable results can be seen is typically between 64^2 and 128^2 samples, depending on the structure of the integrand. It is unreasonable to take so many samples per pixel only for estimating the principal direction(s) instead of using them for the final Monte Carlo estimate directly. As different sampling patterns are used for principal direction estimation (unaligned points) and the final estimator (aligned points), combining both estimates has sub-optimal variance. We propose to compute shared principal directions in blocks of pixels. Using blocks of size 16×16 pixels requires each pixel to sample the integrand at only between $64^2/16^2 = 16$ and $128^2/16^2 = 64$ locations. When implementing the algorithm on the GPU, we launch one work group per block that cooperatively constructs the direction histogram in local memory and subsequently estimates the principal directions. Computing shared principal directions relies on structurally similar integrands. In a practical implementation, the sizes of the blocks should be adjusted adaptively such that pixels in a block share a structurally similar integrand, e.g., using G-buffer information.

5.4. Integration Into a Rendering System

Since our point synthesis algorithm takes several seconds to minutes, depending on the number of points, we precompute one point set for each combination of principal directions. To that end, we discretize the principal directions into 1° integer angles, i.e., in $\{0^\circ, 1^\circ, \ldots, 179^\circ\}$. When optimizing for three directions (horizontal, vertical, and one arbitrary direction), we need to precompute 46 point sets for principal directions between 0° and 45°. Using symmetries, the 46 point sets cover the entire 180° domain (see supplemental material). For four directions (horizontal, vertical, and two arbitrary directions), by exploiting symmetries, we need to precompute 4,141 sets, requiring about 136 MB in total for 2^{12} points per set.

Our final technique then works as follows (cf. Fig. 1): During runtime, we first estimate the shared principal direction(s) in blocks of pixels. Then, we select a precomputed point set with matching low-energy directions (currently up to four). If our principal direction algorithm does not find any other directions apart from the horizontal and vertical axes, e.g., on smooth integrands, we use an existing anisotropic sampling pattern. We decorrelate points across pixels using toroidal shifts [CP76], which makes it sufficient to precompute only one point set per principal direction combination.

6. Results

We show our optimized two-dimensional point sets with anisotropic power spectra with multiple low-energy directions and demonstrate an improved convergence rate on two-dimensional discontinuous integrands (Section 6.1). Further, we evaluate our principal direction estimation: We show that optimizing for and aligning with the most relevant estimated principal directions (horizontal, vertical, and up to two arbitrary directions) of two-dimensional integrands improves the variance in offline path tracing (Section 6.2). We also conduct additional experiments on smoothed discontinuities and for hemisphere sampling in Section 6.3.

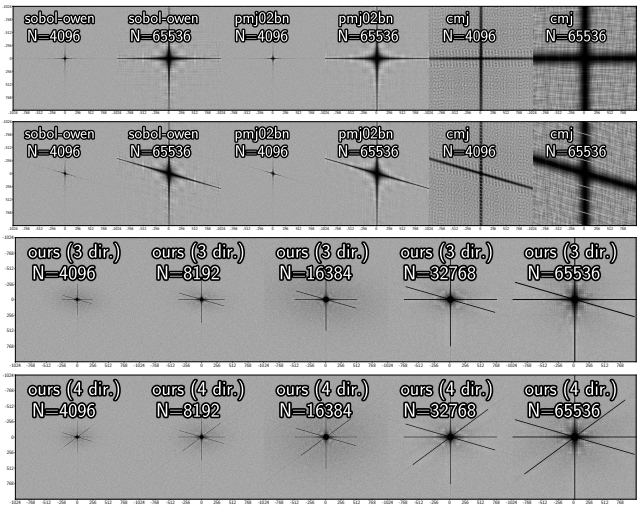


Figure 5: Expected power spectra computed on 1024 frequencies of existing (sheared) anisotropic sampling patterns and our optimized aligned sampling patterns with N points.

Optimization of the points using gradient descent is implemented using PyTorch [PGM*19] for automatic differentiation and Adam [KB17], a first-order optimizer. We implemented our technique for aligning anisotropic sampling points with the most relevant principal direction in an offline GPU path tracer. We provide the source code for the point optimization, the principal direction estimation, and the samplers in the supplemental material. Reported timings for the principal direction estimation and point optimization on the GPU are determined on an NVIDIA RTX 5080.

In addition to our optimized sampling sets, we evaluate the variance when using realizations of Owen-scrambled Sobol (sobol-owen) [Sob67, Owe95, Bur20] and progressive multi-jittered (0,2) with blue noise (pmj02bn) [CKK18] sampling sequences, along with correlated multi-jittered (cmj) [Ken13] sampling sets. The anisotropic expected power spectra of the sampling patterns are shown in Fig. 5. We homogenize the points using toroidal shifts [CP76] (cf. Section 2). The influence of homogenization on the convergence rate on discontinuous integrands is minor, as discussed by Singh et al. [SSC*20] and shown in our supplemental material.

6.1. Point Synthesis and Variance Convergence Rate

Fig. 5 shows the expected anisotropic power spectra of our synthesized sampling sets with 2^{12} to 2^{16} points. Our optimization effectively suppresses undesired frequencies and ensures 1D stratification along multiple directions. We can cover more than two directions, i.e., horizontal, vertical, and principal direction(s), compared to the two directions of existing anisotropic sampling patterns. When suppressing frequencies, we trade higher frequencies (the length of the black line segments) against angular range (the width of the black line segments): We want to avoid hairline anisotropies, so the point set will be robust against slight misalignment. The

more directions we want to optimize for in total, the less effectively we can suppress high frequencies along each direction (see supplemental material for convergence plots on integrands with an increasing number of discontinuities). In the limit of optimizing for an infinite number of directions, we would obtain an isotropic blue noise power spectrum.

The optimization time is mainly influenced by the calculation of the power spectrum loss, as it involves computing a complex exponential for each combination of frequency and point location (cf. Eq. 6). For our sampling set with 2^{12} points and 4,303 suppressed frequencies, optimization takes about 17 seconds. In contrast, optimizing 2^{16} points and suppressing 55,741 frequencies requires about 36 minutes. We refer to the supplemental material for a detailed listing with suppressed frequencies, optimization parameters, and optimization times. Of course, our technique is not limited to 2D sets of points with this specific type of spectrum, but also allows us to optimize points with arbitrarily suppressed frequencies.

Variance Convergence Rate In Fig. 6, we evaluate our point sets on various two-dimensional discontinuous integrands. We evaluate existing anisotropic sampling patterns that match only two high-energy directions (horizontal and vertical axes) and Singh et al.'s [SJ17] proposed shearing technique that still matches only two high-energy directions (one arbitrary direction and one remaining axis). These exhibit the well-known variance convergence rate of about $\mathcal{O}(N^{-1.5})$ on discontinuous integrands. In comparison, our points show an improved $\mathcal{O}(N^{-1.75})$ convergence rate. This means that our point optimization successfully increases the number of evenly distributed samples along all high-energy directions in the integrand proportionally to the number of points in the set, which leads to an improved variance convergence rate (cf. Section 2). On integrands with four high-energy directions, using our points that match with only three directions reduces variance slightly but again has a $\mathcal{O}(N^{-1.5})$ convergence rate. This result is consistent with the analysis of Singh et al. [SJ17] as the direction with the highest energy in the product of the power spectrum of the integrand and the power spectrum of the sampling pattern determines the convergence rate. During our experiments, we found that optimizing points on both the power spectrum loss as well as the sliced Wasserstein loss reduces variance the most, compared to using only one of them (see supplemental material for convergence plots).

6.2. Aligning Sampling Points in Offline Path Tracing

We evaluate our principal direction estimation algorithm and the influence of using our optimized point sets when aligning with the proposed principal directions on the convergence rate on four different scenes (Fig. 7). In the first three scenes, we sample a rectangular light source at the primary hit to compute direct illumination. The integrands in the BREAKFAST-ROOM scene exhibit high-frequency patterns caused by the window blinds. We rotate the light source by 45° to control the orientation of the pattern. The integrands in the CORNELL-BOX scene feature one or two discontinuities caused by occlusion of the light source by one of the small boxes. In the ATTIC scene, the integrands also mainly feature one or two discontinuities caused by the shape of the window blocking parts of the textured area light. In the last scene, DEPTH-OF-FIELD, we sample a point on a square aperture for depth of field rendering and deterministically connect to a point light at the primary hit. Discontinuities arise from different occlusion of the point light due to significantly different primary hits in a pixel that is out of focus.

Principal Direction Estimation In Fig. 7, we visualize the zero to two estimated principal directions (apart from the horizontal and vertical axes) in blocks of 16×16 pixels. Additionally, Fig. 9 shows the noisy direction histograms and our fitted vMFs as described in Section 5.2. While taking the direction with the maximum value in the histogram is susceptible to noise, our technique reliably estimates the principal directions near the noisy maxima and correctly determines scenarios where the integrand has only one or no principal direction at all. We use a total of 96^2 integrand samples per block, corresponding to 36 samples per pixel. We have experimented with taking fewer integrand samples per block, e.g., only 64^2, and have observed that the less accurate sampling of the discontinuities can be compensated for to a certain extent by using a smoothing kernel with a large support. This still works reliably on some integrands (cf. Fig. 1 (a), where we use 64^2 samples); however, we observe that the estimation on more complex integrands (high-frequency patterns or multiple discontinuities) deteriorates as too much information is missing about the features in the integrand. In our GPU implementation, this pre-processing step of estimating the shared principal directions in blocks with a screen resolution of 1280×1280 takes, dependent on the scene, no more than 20ms, which is acceptable for offline rendering.

Variance Convergence Rate We visualize the variance of the Monte Carlo estimator after 4,096 samples per pixel in Fig. 8. As existing anisotropic sampling patterns exhibit a similar convergence behavior on discontinuous integrands (cf. Fig. 6), we compare our aligned sampling sets with three and four low energy directions only to unaligned and sheared pmj02bn as a representative. We use our estimated principal directions (cf. Fig. 7) for alignment. Using our sampling sets that align with the horizontal, vertical, and up to two additional estimated principal directions improves variance in all scenes. This also shows that our estimated principal directions are sufficiently accurate. However, our point sets also allow for some misalignment as we suppress frequencies in a certain range along the principal directions during optimization. Please refer to our supplemental material for more details. Further, we observe that matching four instead of three directions reduces variance in regions where the differently oriented shadow edges overlap. Aligning existing anisotropic samplers using shearing has only advantages on the integrands with high-frequency pattern in BREAKFAST-ROOM, where the highest energies are along the direction of the pattern. The horizontal and vertical axes carry comparatively little energy, which means there is no advantage in covering them as well with our point sets. However, we observe that in the other scenes, shearing can even increase variance when the horizontal and vertical axes have slightly higher energies.

For a few pixel regions (marked in Fig. 7), we show detailed convergence plots when rendering the scenes with $2^8, 2^9, \ldots, 2^{12}$ samples per pixel in Fig. 9. There, we see a similar improved convergence rate when using our aligned points. Also, Fig. 9 shows visual comparisons between the different sampling patterns on lower

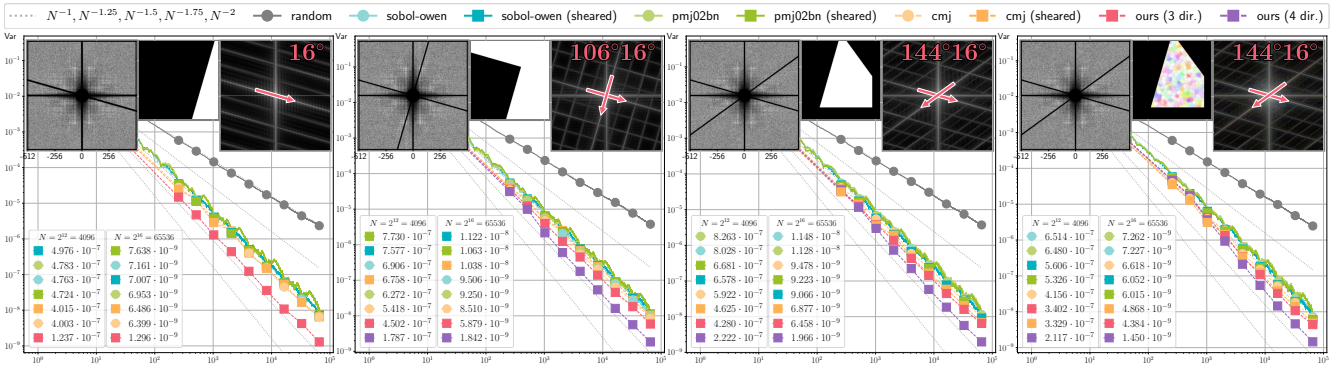


Figure 6: We compare the variance convergence rate on various discontinuous integrands of existing unaligned anisotropic sampling patterns (matching only horizontal and vertical), Singh et al.'s [SJ17] proposed shearing technique (matching one arbitrary direction and one remaining axis), and our optimized and aligned sampling sets with three and four low energy directions. While the existing patterns (both unaligned and sheared) have a convergence rate of about $\mathcal{O}(N^{-1.5})$, our points exhibit an improved $\mathcal{O}(N^{-1.75})$ convergence rate. When adding a smooth texture, the improved convergence rate only begins to occur once more samples are used, as this affects the distribution of energy across the spectrum. On top of each plot, we show the power spectrum of our aligned points, the integrand, and the spectrum of the integrand, including the principal directions that need to be matched (apart from the horizontal and vertical axes) to obtain an improved convergence rate.

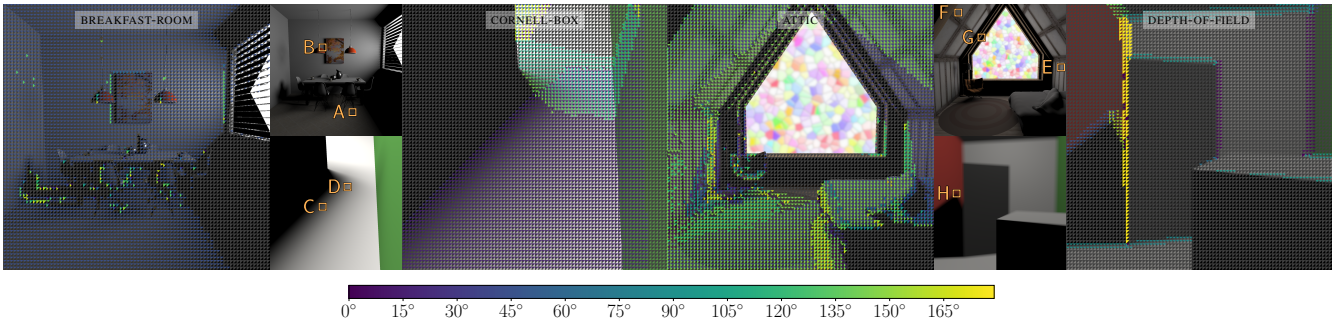


Figure 7: Principal direction estimation in scenes with discontinuous integrands. In each 16×16 pixel block, we visualize the angles of the zero, one, or two estimated principal directions (apart from the horizontal and vertical directions). Gray indicates that the integrand has zero or only one clear principal direction. Except for a small number of blocks, our algorithm reliably estimates the principal directions.

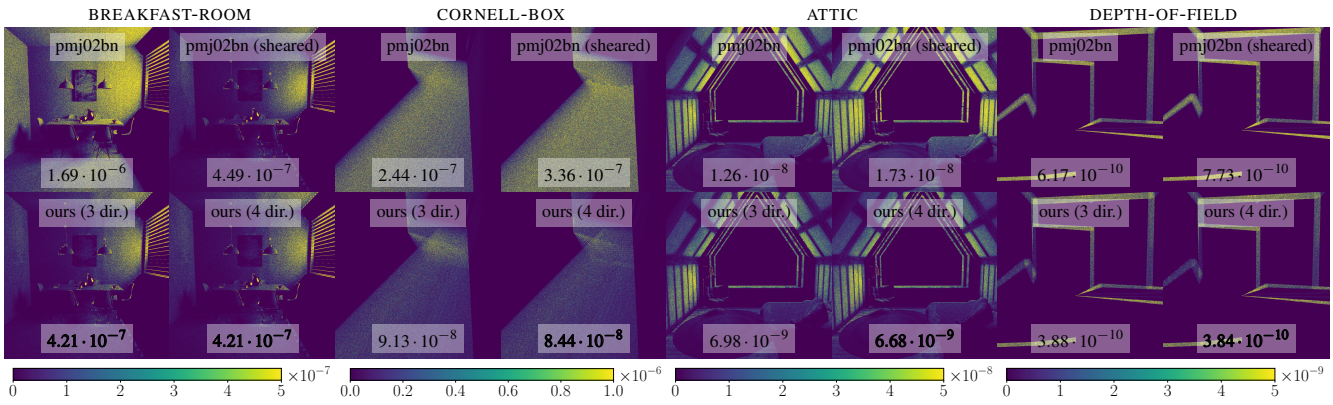


Figure 8: Variance after 4,096 samples per pixel. We compare unaligned pmj02bn (matching only horizontal and vertical), Singh et al.'s [SJ17] proposed shearing technique (matching one arbitrary direction and one remaining axis), and our aligned sampling sets with three and four low energy directions. Using our sampling sets that align with the horizontal, vertical, and up to two additional estimated principal directions improves variance in all scenes. Aligning existing anisotropic samplers using shearing has only advantages on the high-frequency integrands in BREAKFAST-ROOM; in the other scenes, shearing can even increase variance. In the CORNELL-BOX scene, matching four instead of three directions reduces variance in the corner, where the two shadow edges overlap. We report the variance calculated over the entire image.

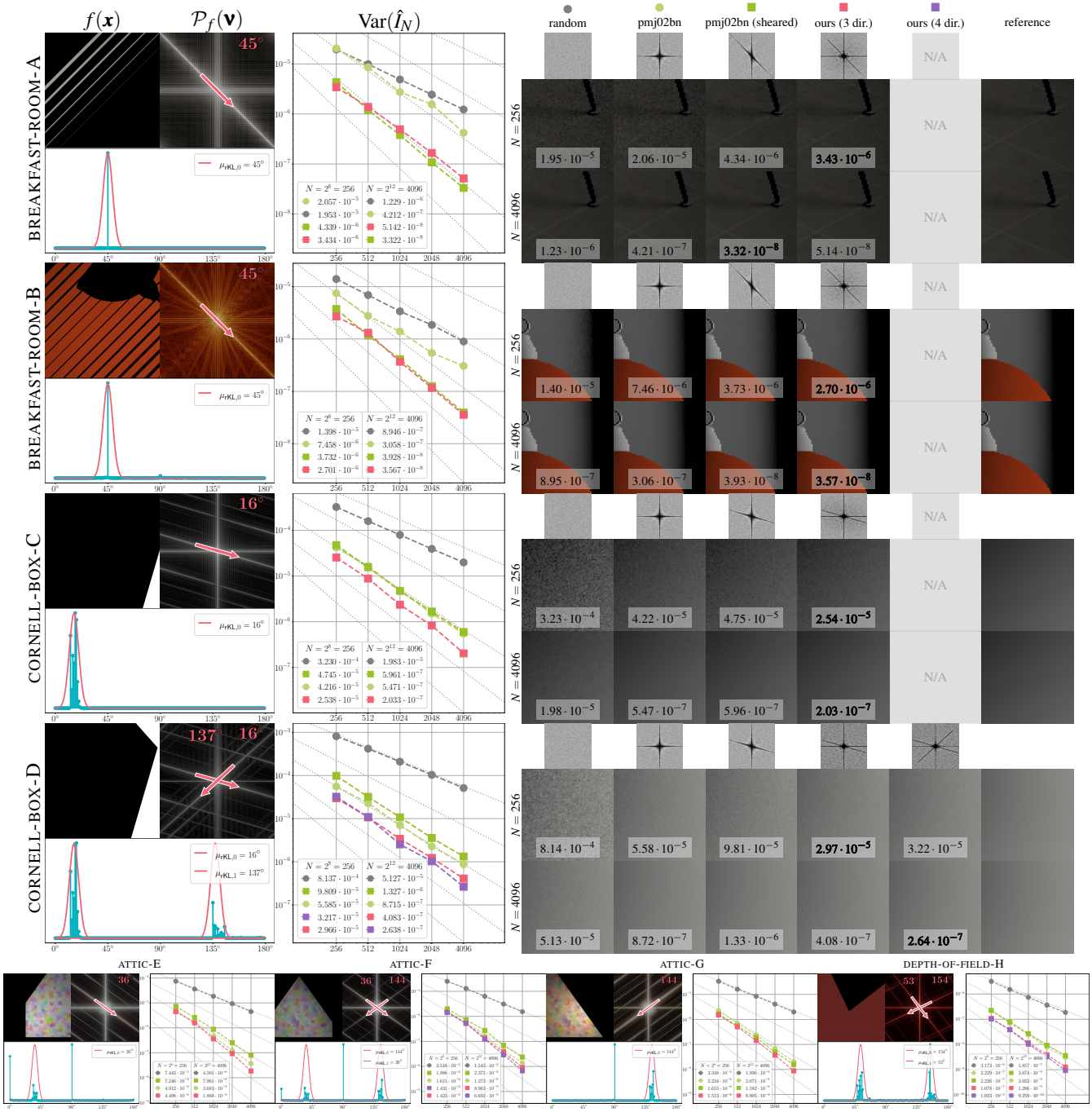


Figure 9: Variance evaluation in offline path tracing on discontinuous integrands. We show the integrand $f(\boldsymbol{x})$, its power spectrum $\mathcal{P}_f(\mathbf{v})$, the principal direction estimation, and the variance $\mathrm{Var}(I_N)$ after accumulating $2^8, 2^9, \ldots, 2^{12}$ samples per pixel. The variance is averaged over the block of pixels shown in Fig. 7. We compare unaligned pmj02bn, Singh et al.'s [SJ17] proposed shearing technique, and our aligned sampling sets with three and four low energy directions. In the BREAKFAST-ROOM scene, shearing pmj02bn and using our points to match our estimated principal direction, both significantly reduce variance and improve the convergence rate as most energy is along the direction of the high-frequency pattern. For the other integrands that feature only a few discontinuities, we observe a similar behavior as in Fig. 6 with our aligned points outperforming existing (sheared) points due to a better convergence rate. In the ATTIC scene, we notice that an improved convergence rate only begins to occur once more samples are used. This is mainly due to the influence of the geometry term and the texture of the light source, which affects the distribution of energy across the power spectrum. Since we improve the convergence *rate*, the difference between samplers increases with a higher number of samples. However, we can already perceive a difference in the noise in the insets with 256 samples per pixel (refer to the supplemental for $N = 2^9, \ldots, 2^{11}$). Please zoom in on the digital version to see all the details clearly.

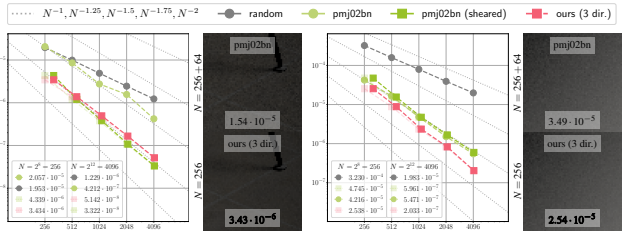


Figure 10: Equal time comparison. Due to the constant overhead for estimating the principal directions, we can take up to 64 additional samples with the unaligned samplers. We moved the curves of the aligned samplers 64 samples to the right in the convergence plots.

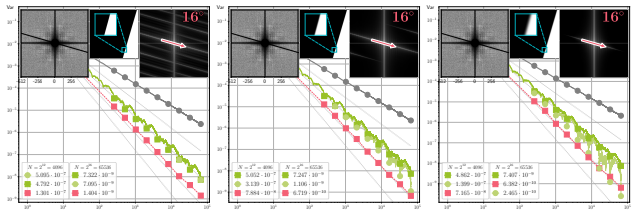


Figure 11: Variance convergence behavior on integrands with a discontinuity (left) and a smoothed discontinuity (center, right). For the smoothed integrands, we sample the step function on a 2048×2048 grid and filter the values with a Gaussian kernel ($\sigma = 2.83$ and $\sigma = 8.5$). In the smoothed case, the convergence rate of the estimator no longer depends on this direction. Existing unaligned anisotropic sampling patterns approach their asymptotic convergence rate as the number of samples increases. As our optimized and aligned points have a better initial convergence rate, it can still be beneficial to use them.

sample counts. Since we improve the convergence *rate*, the difference between samplers increases with higher number of samples. However, we can already perceive a difference in the noise with 256 samples per pixel.

As mentioned before, our principal direction estimation introduces a constant computational overhead of up to 20ms that is unnecessary when using unaligned samplers. Since this overhead is constant and we improve the convergence *rate*, our aligned samplers will eventually always surpass existing samplers. By a conservative estimate, we can already calculate a maximum of 64 additional samples using the unaligned samplers during this period. This means equal time comparisons cause the curve of aligned samplers to shift at a maximum of 64 samples to the right. In Fig. 10, we show this for two pixel regions from Fig. 9 and observe that even at just 256 samples per pixel with the aligned samplers (compared to $256+64$ samples for the unaligned samplers), a clear improvement when aligning is still noticeable.

6.3. Outlook

Smoothed Discontinuities So far, we have focused on integrands with additional arbitrarily oriented discontinuities due to, e.g., occlusion, making it necessary to align with these additional directions to improve the convergence rate. Here, we will investigate the influence of smoothing the discontinuity. This, for example, happens when applying subpixel jittering in path tracing. Subpixel jittering effectively moves the position of the discontinuity slightly over the pixel, leading to a slightly smoothed discontinuity in the integrand. In Fig. 11, we show the convergence plots on a discontinuous integrand and on the same integrand with slight smoothing applied. A smoothed discontinuity no longer needs an infinite number of frequencies to be reconstructed in the Fourier domain. We can see that directly in the power spectrum (Fig. 11), where high energies are only visible along the direction up to a certain frequency. This means that the *asymptotic* convergence rate of the estimator no longer depends on this arbitrarily oriented direction but only on the horizontal and vertical directions since there are still discontinuities at the boundaries. For existing unaligned anisotropic sampling patterns, however, it is visible that the estimator only approaches its asymptotic convergence rate as the number of samples increases, since the arbitrarily oriented smoothed discontinuity still has influence in the beginning. Our optimized points that match all three directions have the same convergence rate throughout. Optimizing for three directions naturally reduces the extent to which frequencies along the axes can be suppressed, therefore they do not reach the same asymptotic convergence rate as existing unaligned anisotropic sampling patterns. Still, it can be beneficial to use our aligned points on smoothed discontinuities, since we can reduce the variance more effectively up to a certain number of samples.

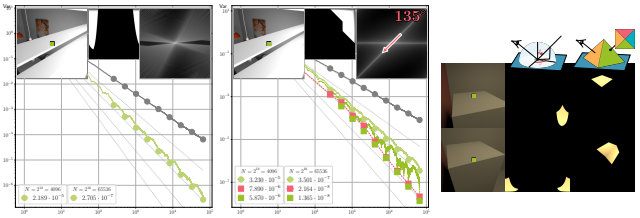


Figure 12: Hemispherical sampling: an octahedral (right plot) instead of an angular mapping (left plot) keeps edges in the integrand straight, and thus the energy in the power spectrum focused in one direction. This means our aligned sampling points significantly improve the variance. Right: examples of integrands using angular (center) or octahedral mapping (right). The energy is no longer spread across a range but into multiple discrete directions.

Hemispherical Sampling With Low Distortion Projection Until now, we have considered cases where only a few straight edges occur in the integrand. With hemispherical sampling, straight edges in the spherical domain are mapped to curved edges in the $[0,1)^2$ domain. The energy in the power spectrum is distributed across many directions, and aligning sampling points is generally impossible. To that end, we propose to use importance sampling techniques that distort straight edges as little as possible when mapping from the spherical to the $[0,1)^2$ domain. In Fig. 12, we experiment with an octahedron mapping [ED08] and show an improved convergence rate compared to standard hemisphere sampling.

7. Limitations and Future Work

Point Optimization Parameters The frequencies to suppress during optimization for a given number of points are currently determined empirically. The chosen frequencies are a tradeoff between suppressing higher frequencies and angular range to be robust against slight misalignment to a certain degree, and to match the spectra of a wide range of common integrands with different structures. For future work, it would be interesting to automate this process and determine the frequencies to suppress during optimization by automatically evaluating the variance on some representative test integrands.

Aligning with Horizontal and Vertical Directions There are cases where there are no discontinuities along the horizontal or vertical axes (e.g., Fig. 2 (left)). There, it can be beneficial to avoid aligning, if this allows us to align with other additional directions instead. In practice, however, most integrands have discontinuities along the horizontal and vertical, either due to the boundary of the finite integration domain or due to occlusion boundaries of man-made structures that often align horizontally or vertically (e.g., the ATTIC scene in Fig. 7). Nevertheless, in a future, optimized version of the technique, it may still be useful to detect such cases.

Runtime Cost for Principal Direction Estimation Taking an additional 16-64 samples per pixel for the principal direction estimation is, of course, a limitation (cf. Section 6.2). In the equal time comparison (Fig. 10), we can, however, show that our technique already leads to improvements at relatively low sample counts (256 spp). We believe that with further optimization, e.g., by dynamically adjusting the pixel block sizes and merging blocks with similar integrands, the number of additional samples per pixel can be reduced even further. Computing shared principal directions only in blocks of pixels that share a similar integrand would also make the estimation more reliable. Additionally, when rendering an animated sequence of frames in offline rendering, one could reuse estimated principal directions from previous frames where possible, amortizing the estimation cost over frames. It would also be interesting to investigate whether it is possible to optimize anisotropic point sets where the first few 16-64 points are always fixed and isotropically distributed so that we can use the samples for the principal direction estimation directly for the final Monte Carlo estimate.

Memory Overhead for Precomputed Point Sets The memory overhead currently grows impractically large to precompute sets with more than two free directions (cf. Section 5.4). This motivates future development of algorithms that allow on-the-fly construction of point sets with anisotropic power spectra with arbitrary low-energy directions. Ideally, these algorithms would generate points in a sampling sequence to allow for progressive rendering.

Higher Dimensional Transport Our technique achieves improved convergence in practical applications for two-dimensional integration problems with discontinuities (direct illumination with light or BSDF sampling, and DoF). We believe that our work represents a potential step toward practical improvement of the variance convergence on high-dimensional integrands with discontinuities by learning and aligning with principal directions in higher dimensions. A next step may be to use our method for aligned sampling at the first hit point, and unaligned sampling for higher dimensions in full global illumination path tracing, which could lead to improvements, depending on the structure of the integrand in higher dimensions. However, this requires further investigation and thorough evaluation.

8. Conclusion

We have introduced an optimization-based algorithm to synthesize sampling points with anisotropic spectra featuring more than two low-energy directions. Using our sampling points, we achieve an improved variance convergence rate of $\mathcal{O}(N^{-1.75})$ on integrands with a few discontinuities compared to $\mathcal{O}(N^{-1.5})$ of existing unaligned anisotropic sampling patterns that match only two directions. Furthermore, we have developed a technique to reliably estimate directions with high energy within the power spectrum of two-dimensional discontinuous integrands. By aligning our optimized points with the estimated directions, we can significantly reduce the variance in offline path tracing. We believe that our work is a step toward improved convergence rates on discontinuous integrands frequently occurring in rendering.

Acknowledgments

We thank Addis Dittebrandt for insightful discussions and the anonymous reviewers for their valuable feedback. The ATTIC scene is created by BlendSwap user *akashdlfps*. The BREAKFAST-ROOM scene is created by *Wig42* and the CORNELL-BOX by *Morgan McGuire*, both are available in McGuire's Computer Graphics Archive. This project has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 462649663.

References

[ANHD17] AHMED A. G. M., NIESE T., HUANG H., DEUSSEN O.: An adaptive point sampler on a regular lattice. *ACM Trans. Graph. 36*, 4 (2017). doi:10.1145/3072959.3073588. 3

[APC*16] AHMED A. G. M., PERRIER H., COEURJOLLY D., OSTROMOUKHOV V., GUO J., YAN D.-M., HUANG H., DEUSSEN O.: Low-discrepancy blue noise sampling. *ACM Trans. Graph. 35*, 6 (2016). doi:10.1145/2980179.2980218. 3

[Bis06] BISHOP C. M.: *Pattern recognition and machine learning*, vol. 4. Springer, 2006. 6

[Bri07] BRIDSON R.: Fast poisson disk sampling in arbitrary dimensions. In *SIGGRAPH Sketches* (2007), p. 22. doi:10.1145/1278780.1278807. 3

[Bur20] BURLEY B.: Practical hash-based owen scrambling. *Journal of Computer Graphics Techniques 10*, 4 (2020), 1–20. URL: http://jcgt.org/published/0009/04/01/. 3, 7

[CKK18] CHRISTENSEN P., KENSLER A., KILPATRICK C.: Progressive multi-jittered sample sequences. *Computer Graphics Forum 37*, 4 (2018), 21–33. doi:10.1111/cgf.13472. 3, 7

[Coo86] COOK R. L.: Stochastic sampling in computer graphics. *ACM Trans. Graph. 5*, 1 (1986), 51–72. doi:10.1145/7529.8927. 3

[CP76] CRANLEY R., PATTERSON T. N. L.: Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis 13*, 6 (1976), 904–914. doi:10.1137/0713071. 3, 7

[CSW94] CHIU K., SHIRLEY P., WANG C.: Multi-jittered sampling. In *Graphics Gems IV*. Academic Press, 1994, p. 370–374. 2, 3

[dGBOD12] DE GOES F., BREEDEN K., OSTROMOUKHOV V., DESBRUN M.: Blue noise through optimal transport. *ACM Trans. Graph. 31*, 6 (2012). doi:10.1145/2366145.2366190. 3

[DHS*05] DURAND F., HOLZSCHUCH N., SOLER C., CHAN E., SILLION F. X.: A frequency analysis of light transport. *ACM Trans. Graph. 24*, 3 (2005), 1115–1126. doi:10.1145/1073204.1073320. 4

[Dur11] DURAND F.: *A frequency analysis of monte-carlo and other numerical integration schemes*. Tech. Rep. MIT-CSAIL-TR-2011-052, MIT, 2011. URL: https://dspace.mit.edu/handle/1721.1/67677. 2, 4

[ED08] ENGELHARDT T., DACHSBACHER C.: Octahedron environment maps. In *Vision, Modeling, and Visualization* (2008), pp. 383–388. 11

[Fat11] FATTAL R.: Blue-noise point sampling using kernel density model. *ACM Trans. Graph. 30*, 4 (2011). doi:10.1145/2010324.1964943. 3

[Hal64] HALTON J. H.: Algorithm 247: Radical-inverse quasi-random point sequence. *Communications of the ACM 7*, 12 (1964), 701–702. doi:10.1145/355588.365104. 2, 3

[HSD13] HECK D., SCHLÖMER T., DEUSSEN O.: Blue noise sampling with controlled aliasing. *ACM Trans. Graph. 32*, 3 (2013), 1–12. doi:10.1145/2487228.2487233. 3, 4

[HVCB21] HEITZ E., VANHOEY K., CHAMBON T., BELCOUR L.: A sliced wasserstein loss for neural texture synthesis. In *Computer Vision and Pattern Recognition* (June 2021), pp. 9407–9415. doi:10.1109/CVPR46437.2021.00929. 5

[JEK*19] JAROSZ W., ENAYET A., KENSLER A., KILPATRICK C., CHRISTENSEN P.: Orthogonal array sampling for monte carlo rendering. *Computer Graphics Forum 38*, 4 (2019), 135–147. doi:10.1111/cgf.13777. 3

[KB17] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. In *International Conference on Learning Representations* (2017). URL: https://arxiv.org/abs/1412.6980. 7

[KCSG18] KULLA C., CONTY A., STEIN C., GRITZ L.: Sony pictures imageworks arnold. *ACM Trans. Graph. 37*, 3 (2018). doi:10.1145/3180495. 3

[Ken13] KENSLER A.: *Correlated multi-jittered sampling*. Tech. Rep. 13-01, Pixar Animation Studios, 2013. URL: https://graphics.pixar.com/library/MultiJitteredSampling/. 3, 7

[KL51] KULLBACK S., LEIBLER R. A.: On information and sufficiency. *The annals of mathematical statistics 22*, 1 (1951), 79–86. URL: https://www.jstor.org/stable/2236703. 6

[KSKAC02] KELEMEN C., SZIRMAY-KALOS L., ANTAL G., CSONKA F.: A simple and robust mutation strategy for the metropolis light transport algorithm. *Computer Graphics Forum 21*, 3 (2002), 531–540. doi:10.1111/1467-8659.t01-1-00703. 3

[LD08] LAGAE A., DUTRÉ P.: A comparison of methods for generating poisson disk distributions. *Computer Graphics Forum 27*, 1 (2008), 114–129. doi:https://doi.org/10.1111/j.1467-8659.2007.01100.x. 3

[LSM*19] LEIMKÜHLER T., SINGH G., MYSZKOWSKI K., SEIDEL H.-P., RITSCHEL T.: Deep point correlation design. *ACM Trans. Graph. 38*, 6 (2019). doi:10.1145/3355089.3356562. 3, 4

[Mit91] MITCHELL D. P.: Spectrally optimal sampling for distribution ray tracing. In *SIGGRAPH* (1991), Association for Computing Machinery, p. 157–164. doi:10.1145/122718.122736. 3

[Owe95] OWEN A. B.: Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (1995), pp. 299–317. doi:10.1007/978-1-4612-2552-2_19. 3, 7

[PBC*20] PAULIN L., BONNEEL N., COEURJOLLY D., IEHL J.-C., WEBANCK A., DESBRUN M., OSTROMOUKHOV V.: Sliced optimal transport sampling. *ACM Trans. Graph. 39*, 4 (2020). doi:10.1145/3386569.3392395. 3, 5

[PCX*18] PERRIER H., COEURJOLLY D., XIE F., PHARR M., HANRAHAN P., OSTROMOUKHOV V.: Sequences with low-discrepancy blue-noise 2-d projections. *Computer Graphics Forum 37*, 2 (2018), 339–353. doi:10.1111/cgf.13366. 3

[PGM*19] PASZKE A., GROSS S., MASSA F., LERER A., BRADBURY J., CHANAN G., KILLEEN T., LIN Z., GIMELSHEIN N., ANTIGA L., ET AL.: Pytorch: An imperative style, high-performance deep learning library. In *Neural Information Processing Systems* (2019), vol. 32. URL: https://arxiv.org/abs/1912.01703. 7

[PSC*15] PILLEBOUE A., SINGH G., COEURJOLLY D., KAZHDAN M., OSTROMOUKHOV V.: Variance analysis for monte carlo integration. *ACM Trans. Graph. 34*, 4 (2015). doi:10.1145/2766930. 2, 3, 4

[RRSG16] REINERT B., RITSCHEL T., SEIDEL H.-P., GEORGIEV I.: Projective blue-noise sampling. *Computer Graphics Forum 35*, 1 (2016), 285–295. doi:10.1111/cgf.12725. 3

[SGSS22] SALAÜN C., GEORGIEV I., SEIDEL H.-P., SINGH G.: Scalable multi-class sampling via filtered sliced optimal transport. *ACM Trans. Graph. 41*, 6 (2022). doi:10.1145/3550454.3555484. 3, 5

[Shi91] SHIRLEY P.: Discrepancy as a quality measure for sample distributions. In *Eurographics* (1991), vol. 91, pp. 183–194. 3

[SJ17] SINGH G., JAROSZ W.: Convergence analysis for anisotropic monte carlo sampling spectra. *ACM Trans. Graph. 36*, 4 (2017). doi:10.1145/3072959.3073656. 2, 3, 4, 8, 9, 10

[SK13] SUBR K., KAUTZ J.: Fourier analysis of stochastic sampling strategies for assessing bias and variance in integration. *ACM Trans. Graph. 32*, 4 (2013). doi:10.1145/2461912.2462013. 2, 4

[SOA*19] SINGH G., ÖZTIRELI C., AHMED A. G., COEURJOLLY D., SUBR K., DEUSSEN O., OSTROMOUKHOV V., RAMAMOORTHI R., JAROSZ W.: Analysis of sample correlations for monte carlo rendering. *Computer Graphics Forum 38*, 2 (2019), 473–491. doi:10.1111/cgf.13653. 2, 3

[Sob67] SOBOL' I.: On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics 7*, 4 (1967), 86–112. doi:10.1016/0041-5553(67)90144-9. 2, 3, 7

[SSC*20] SINGH G., SUBR K., COEURJOLLY D., OSTROMOUKHOV V., JAROSZ W.: Fourier analysis of correlated monte carlo importance sampling. *Computer Graphics Forum 39*, 1 (2020), 7–19. doi:10.1111/cgf.13613. 2, 3, 4, 7

[Wol23] WOLFE A.: Deriving the inverse CDF of a rotated square projected onto a line, 2023. URL: https://blog.demofox.org/2023/12/24/deriving-the-inverse-cdf-of-a-rotated-square-projected-onto-a-line/. 5

[WW11] WEI L.-Y., WANG R.: Differential domain analysis for non-uniform sampling. *ACM Trans. Graph. 30*, 4 (2011). doi:10.1145/2010324.1964945. 4

[ZHWW12] ZHOU Y., HUANG H., WEI L.-Y., WANG R.: Point sampling with general noise spectrum. *ACM Trans. Graph. 31*, 4 (2012). doi:10.1145/2185520.2185572. 4