Error Analysis of Common Sampling Strategies

Gurprit Singh
gsingh@mpi-inf.mpg.de
We start with random sampling [CLICK] which involves randomly generating samples in the integration domain, the noise level is significantly high. We can improve this sample distribution by...
We start with random sampling [CLICK] which involves randomly generating samples in the integration domain, the noise level is significantly high. We can improve this sample distribution by...
...dividing the domain in equal strata and placing the samples at the center of each stratum. The regularity has been known to cause some aliasing effects which can be easily avoided by randomly jittering...
...each sample independently in each stratum. This is called jittered sampling and as you might notice the noise level has already improved for the same sample count. We can further improve the uniformity of samples by, say,...
...generating a sample [CLICK] in the domain following a naive dart throwing approach where a radius is assigned to each sample and a new sample is only accepted if it falls outside the disk radius. This gives us Poisson disk samples...
...generating a sample [CLICK] in the domain following a naive dart throwing approach where a radius is assigned to each sample and a new sample is only accepted if it falls outside the disk radius. This gives us Poisson disk samples...
...generating a sample [CLICK] in the domain following a naive dart throwing approach where a radius is assigned to each sample and a new sample is only accepted if it falls outside the disk radius. This gives us Poisson disk samples...
... which is well distributed and the corresponding noise in the image has also gone down for the same number of samples. Another way to decrease variance is to keep...
...increase the sample count till the image becomes noise-free (converge). The rate at which this image converges depend on the underlying sampling pattern used. For example, with random samples [CLICK] the convergence rate is always $O(N^{-1})$, 4D jittered samples [CLICK], we would obtain a 4D convergence rate of $O(N^{-1.25})$ whereas with Poisson disk...
Variance Convergence Rate of Samplers

...increase the sample count till the image becomes noise-free (converge). The rate at which this image converges depend on the underlying sampling pattern used. For example, with random samples [CLICK] the convergence rate is always $O(N^{-1})$, 4D jittered samples [CLICK], we would obtain a 4D convergence rate of $O(N^{-1.25})$ whereas with Poisson disk...
...increase the sample count till the image becomes noise-free (converge). The rate at which this image converges depend on the underlying sampling pattern used. For example, with random samples [CLICK] the convergence rate is always $O(N^{-1})$, 4D jittered samples [CLICK], we would obtain a 4D convergence rate of $O(N^{1.25})$ whereas with Poisson disk...
increase the sample count till the image becomes noise-free (converge). The rate at which this image converges depend on the underlying sampling pattern used. For example, with random samples the convergence rate is always $O(N^{-1})$, 4D jittered samples, we would obtain a 4D convergence rate of $O(N^{-1.25})$ whereas with Poisson disk...
...increase the sample count till the image becomes noise-free (converge). The rate at which this image converges depend on the underlying sampling pattern used. For example, with random samples the convergence rate is always $O(N^{-1})$, 4D jittered samples, we would obtain a 4D convergence rate of $O(N^{-1.25})$ whereas with Poisson disk...
... we can obtain less noisy images at small sample count but [CLICK] as we increase the samples the convergence rate obtained is $O(N^{-1})$.

This reflects that if the sampling budget is small (which is the case in many interactive applications) it is best to use Poisson disk samplers, whereas for large sampling budget, for example in offline rendering for movie frames, we should consider samplers with good convergence.

These convergence rates can be empirically computed using the sample variance. We proposed a mathematical convergence tool in the Fourier domain, which allows to theoretically derive these convergence rates for blue noise samples. Now, lets try to understand what characteristics of these samples are affecting the quality of the rendered images.
... we can obtain less noisy images at small sample count but as we increase the samples the convergence rate obtained is $O(N^{-1})$.

This reflects that if the sampling budget is small (which is the case in many interactive applications) it is best to use Poisson disk samplers, whereas for large sampling budget, for example in offline rendering for movie frames, we should consider samplers with good convergence.

These convergence rates can be empirically computed using the sample variance. We proposed a mathematical convergence tool in the Fourier domain, which allows to theoretically derive these convergence rates for blue noise samples. Now, lets try to understand what characteristics of these samples are affecting the quality of the rendered images.
... we can obtain less noisy images at small sample count but [CLICK] as we increase the samples the convergence rate obtained is $O(N^{-1})$.

This reflects that if the sampling budget is small (which is the case in many interactive applications) it is best to use Poisson disk samplers, whereas for large sampling budget, for example in offline rendering for movie frames, we should consider samplers with good convergence.

These convergence rates can be empirically computed using the sample variance. We proposed a mathematical convergence tool in the Fourier domain, which allows to theoretically derive these convergence rates for blue noise samples. Now, let's try to understand what characteristics of these samples are affecting the quality of the rendered images.
We start by looking into error introduced by different stratification strategies.

We then analyze error due to blue noise distributions and how this idea is extended beyond to introduce discrepancy related measures.

In the end, we will briefly look at how to analyze importance sampling using the presented Fourier tools and how correlated samples could affect the convergence properties of these importance sampling strategies.
We start by looking into error introduced by different stratification strategies. We then analyze error due to blue noise distributions and how this idea is extended beyond to introduce discrepancy related measures. In the end, we will briefly look at how to analyze importance sampling using the presented Fourier tools and how correlated samples could affect the convergence properties of these importance sampling strategies.
We start by looking into error introduced by different stratification strategies.

[CLICK] We then analyze error due to blue noise distributions and how this idea is extended beyond to introduce discrepancy related measures.

[CLICK] In the end, we will briefly look at how to analyze importance sampling using the presented Fourier tools and how correlated samples could affect the convergence properties of these importance sampling strategies.
The regular artifacts due to regular grid sampling can be avoided by simply shifting...
... the samples randomly within the strata, which is called uniform jitter sampling.
... the samples randomly within the strata, which is called uniform jitter sampling.
Randomly jittered samples

Pauly et al. [2000]

Regular

Uniform jitter

Random jitter
Another approach to avoid these artifacts is to randomly generate samples within each stratum that is called random jittering.
We would like to understand how these two different strategies affect the error during rendering.
Here we show one example, where for Quad (square) area light source, random jittered sampling [CLICK] gives less noisy image compared to uniform jittering. Let’s try to understand why is that.
Here we show one example, where for Quad (square) area light source, random jittered sampling [CLICK] gives less noisy image compared to uniform jittering. Let’s try to understand why is that.
Here we are looking at a light source sampled using uniformly jittered samples. The orange part is occlude whereas the yellow part is visible from the light source. [CLICK] If we look at these samples near the discontinuity, uniform jittering shifts all samples on one side of the discontinuity creating some kind of positive correlation with respect to the discontinuity. However, with randomly jittered samples...
Here we are looking at a light source sampled using uniformly jittered samples. The orange part is occlude whereas the yellow part is visible from the light source. [CLICK] If we look at these samples near the discontinuity, uniform jittering shifts all samples on one side of the discontinuity creating some kind of positive correlation with respect to the discontinuity. However, with randomly jittered samples...
...if we look at the set of pixels with the discontinuity [CLICK], since the samples were generated randomly jittered, samples can be easily found on either side of the discontinuity. This decorrelates randomly jittered samples w.r.t. the discontinuity resulting in less noise.
...if we look at the set of pixels with the discontinuity [CLICK], since the samples were generated randomly jittered, samples can be easily found on either side of the discontinuity. This decorrelates randomly jittered samples w.r.t. the discontinuity resulting in less noise.
Now, if we change the shape of our light source to a Disk light, we see that [CLICK] uniform jitter is far better than random jittering. Furthermore,...
Now, if we change the shape of our light source to a Disk light, we see that uniform jitter is far better than random jittering. Furthermore,...
Polar mapping performs better for some samplers compared to concentric mapping observed by Andrew Kensler [2013]

It was observed that for circular light source, the way these samples are mapped also affect the quality. Surprisingly, polar mapping performs better than concentric mapping in some cases. Let's first look at these mappings.
Samples in a circular domain or Disk are generated by warping the samples from the square domain using, say, polar mapping, which ...
... which distributed samples in this manner.

[CLICK] If we look at the samples near the discontinuity, on the Disk [CLICK] they are placed on either side of the domain in a concentric ring. Here the strata are of uneven shape due to the distortion by the polar mapping which is not considered good, but it was reported that this might explain why polar mapping behaves better for correlated multi-jittered samples (by Kensler) then concentric mapping. In concentric mapping, ...
... which distributed samples in this manner.

[CLICK] If we look at the samples near the discontinuity, on the Disk [CLICK] they are placed on either side of the domain in a concentric ring. Here the strata are of uneven shape due to the distortion by the polar mapping which is not considered good, but it was reported that this might explain why polar mapping behaves better for correlated multi-jittered samples (by Kensler) then concentric mapping. In concentric mapping, ...
... which distributed samples in this manner. [CLICK] If we look at the samples near the discontinuity, on the Disk [CLICK] they are placed on either side of the domain in a concentric ring. Here the strata are of uneven shape due to the distortion by the polar mapping which is not considered good, but it was reported that this might explain why polar mapping behaves better for correlated multi-jittered samples (by Kensler) then concentric mapping. In concentric mapping, ...
... the strata have a very similar shape compared to the square strata and if we look again at these samples near the discontinuity, they remain [CLICK] in almost the same vicinity.
... the strata have a very similar shape compared to the square strata and
if we look again at these samples near the discontinuity, they remain in almost the same vicinity.
Cengiz analyzed this further and observed that with Disk light sources...
...sometimes uniform jitter is good and sometimes bad compared to random jittering. He proposed another variant named isotropic jittering.
Isotropic jitter = uniform jitter + random rotation

The idea is, [CLICK] you first randomly shift the samples and then, [CLICK] randomly rotate them.
Isotropic jitter = uniform jitter + random rotation

The idea is, [CLICK] you first randomly shift the samples and then, [CLICK] randomly rotate them.
The idea is,
[CLICK] you first randomly shift the samples and then,
[CLICK] randomly rotate them.
Rotated uniform jitter better for not too complex shadows

This showed improvements for not too complex shadows, especially when this rotation is done with some knowledge of the orientation of the occlusion boundaries.
This shows that these simple correlations introduced by different jittering variants can favorably affect the error.

[CLICK] Let's now look at how other correlations, namely blue noise samples, affect the error in Monte Carlo estimation.
This shows that these simple correlations introduced by different jittering variants can favorably affect the error.

Let's now look at how other correlations, namely blue noise samples, affect the error in Monte Carlo estimation.
Fourier tools are often used to understand the characteristics of sample correlations
Chapter 5. Popular sampling patterns

<table>
<thead>
<tr>
<th>Samples</th>
<th>Power spectrum</th>
<th>Radial mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Jitter</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Multi-jitter</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>N-rooks</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 5.6: Illustration of random and some stochastic grid-based sampling patterns with the corresponding Fourier expected power spectra and the corresponding radial mean of their expected power spectra.

For example, we can compute the expected power spectrum of random samples, which is a flat gray image. This spectrum can be [CLICK] radially averaged to get a 1D radial version of this spectrum. Here, [CLICK] the center of the 2D spectrum is the DC peak (or the zero frequency), which represents the starting frequency of the radially averaged spectrum.
For example, we can compute the expected power spectrum of random samples, which is a flat gray image. This spectrum can be [CLICK] radially averaged to get a 1D radial version of this spectrum. Here, [CLICK] the center of the 2D spectrum is the DC peak (or the zero frequency), which represents the starting frequency of the radially averaged spectrum.
For example, we can compute the expected power spectrum of random samples, which is a flat gray image. This spectrum can be [CLICK] radially averaged to get a 1D radial version of this spectrum. Here, [CLICK] the center of the 2D spectrum is the DC peak (or the zero frequency), which represents the starting frequency of the radially averaged spectrum.
For jittered samples, the spectrum changes and gets some dark region in the low frequency region around the DC peak which is also well captured in the radially averaged profile between the range [0,1]. Note that the horizontal axis in the radial profile represents a normalized frequency (m/sqrt[N] value for an m-th frequency).
For blue noise samplers, this energy free low frequency region becomes larger.
When we sample a given function \( f(x) \), the variance during Monte Carlo integration due to these samples is...
Variance in terms of power spectra

\[ \text{Var}(I_N) \propto f(\bar{x}) \]

... proportional to the product of the integrand power spectrum and the samples' expected power spectrum.

Fredo Durand [2011]
Pillebuoe et al. [2015]
For samples with isotropic power spectra, that is having same energy distribution for a given radial distribution, the corresponding variance expression can be simplified to...

\[ \text{Var}(I_N) \propto \]
Var\( (I_N) \propto \) 

For samples with isotropic power spectra, that is having same energy distribution for a given radial distribution, the corresponding variance expression can be simplified to...
Variance in terms of power spectra

For samples with isotropic power spectra, that is having same energy distribution for a given radial distribution, the corresponding variance expression can be simplified to...
For samples with isotropic power spectra, that is having same energy distribution for a given radial distribution, the corresponding variance expression can be simplified to...
Variance in terms of power spectra

\[ \text{Var}(I_N) \propto \text{Integrand power spectrum} \]

Pillebueo et al. [2015]

...the radial counterpart. By taking the radial average of the corresponding integrand spectrum...
Variance in terms of power spectra

\[ \text{Var}(I_N) \propto \text{Samples' expected power spectrum} \times \text{Integrand power spectrum} \]

...the corresponding variance can be represented as a function of these 1D radial profiles.
We used these radial profiles [CLICK] to derive convergence rates for different sampling patterns. The surprising result noted here is the convergence behavior of Poisson disk which belongs to the blue noise class with jittered samples, which has much better convergence.
Convergence rate depends on the low frequency region

\[ \text{Var}(I_N) \propto \text{Integrand power spectrum} \times \text{Samples' expected power spectrum} \]

<table>
<thead>
<tr>
<th>Samplers</th>
<th>Worst Case</th>
<th>Best Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>( \mathcal{O}(N^{-1}) )</td>
<td>( \mathcal{O}(N^{-1}) )</td>
</tr>
<tr>
<td>Poisson Disk</td>
<td>( \mathcal{O}(N^{-1}) )</td>
<td>( \mathcal{O}(N^{0}) )</td>
</tr>
<tr>
<td>Jitter</td>
<td>( \mathcal{O}(N^{-1.5}) )</td>
<td>( \mathcal{O}(N^{0}) )</td>
</tr>
<tr>
<td>CCVT</td>
<td>( \mathcal{O}(N^{-1.3}) )</td>
<td>( \mathcal{O}(N^{-2}) )</td>
</tr>
</tbody>
</table>

We used these radial profiles [CLICK] to derive convergence rates for different sampling patterns. The surprising result noted here is the convergence behavior of Poisson disk which belongs to the blue noise class with jittered samples, which has much better convergence.
This can be explained by [CLICK] looking at the low-frequency region of these radial profiles. [CLICK] For jittered samples, the radial profiles goes to zero near the DC (zero) frequency, whereas for Poisson disk samples the radial profile has an offset, which translates into it’s bad convergence.
Jittered samples converges faster than Poisson Disk

This can be explained by [CLICK] looking at the low-frequency region of these radial profiles. [CLICK] For jittered samples, the radial profiles goes to zero near the DC (zero) frequency, whereas for Poisson disk samples the radial profile has an offset, which translates into it’s bad convergence.
This can be explained by [CLICK] looking at the low-frequency region of these radial profiles. [CLICK] For jittered samples, the radial profiles goes to zero near the DC (zero) frequency, whereas for Poisson disk samples the radial profile has an offset, which translates into its bad convergence.
These convergence tools were developed for samplers with isotropic power spectra. However, the samplers used in practice have anisotropic energy distribution in their power spectrum. For example, ...
These convergence tools were developed for samplers with isotropic power spectra. However, the samplers used in practice have anisotropic energy distribution in their power spectrum. For example, ...
Convergence rate depends on the low frequency region

\[ \text{Var}(I_N) \propto \left| \frac{\text{Samples' expected power spectrum}}{\text{Integrand power spectrum}} \right| \]

<table>
<thead>
<tr>
<th>Samplers</th>
<th>Worst Case</th>
<th>Best Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>(O(N^{-1}))</td>
<td>(O(N^{-1}))</td>
</tr>
<tr>
<td>Poisson Disk</td>
<td>(O(N^{-1}))</td>
<td>(O(N^{-1}))</td>
</tr>
<tr>
<td>Jitter</td>
<td>(O(N^{-1.5}))</td>
<td>(O(N^{-2}))</td>
</tr>
<tr>
<td>CCVT</td>
<td>(O(N^{-1.3}))</td>
<td>(O(N^{-3}))</td>
</tr>
</tbody>
</table>

These convergence tools were developed for samplers with isotropic power spectra. However, the samplers used in practice have anisotropic energy distribution in their power spectrum. For example, ...
…a Latin hypercube sampler, that is constructed by first generating samples along a diagonal followed by randomly shuffling along the rows and then along the columns.
…a Latin hypercube sampler, that is constructed by first generating samples along a diagonal followed by randomly shuffling along the rows and then along the columns.
…a Latin hypercube sampler, that is constructed by first generating samples along a diagonal followed by randomly shuffling along the rows and then along the columns.
…a Latin hypercube sampler, that is constructed by first generating samples along a diagonal followed by randomly shuffling along the rows and then along the columns.
…a Latin hypercube sampler, that is constructed by first generating samples along a diagonal followed by randomly shuffling along the rows and then along the columns.
…a Latin hypercube sampler, that is constructed by first generating samples along a diagonal followed by randomly shuffling along the rows and then along the columns.
Latin Hypercube Sampler (N-rooks)

Shuffle columns

Slide after Wojciech Jarosz
Latin Hypercube Sampler (N-rooks)

Shuffle columns

Slide after Wojciech Jarosz
Latin Hypercube Sampler (N-rooks)

Slide after Wojciech Jarosz
As a result, the underlying power spectrum has an...
... anisotropic power spectrum with hairline structures visible as a dark cross in the middle. These hairline anisotropies are there due to the denser stratification along the X...
... anisotropic power spectrum with a dark cross in the middle. These hairline anisotropies are there due to the denser stratification along the X...
...and the Y-axis. It is also possible to directly obtain good 2D stratified samples...
...which has a power spectrum [CLICK] with a dark region around the center. Chiu and colleagues, found a better construction for these samples...
...which has a power spectrum [CLICK] with a dark region around the center. Chiu and colleagues, found a better construction for these samples...
Anisotropic Sampling Power Spectra

N-rooks / Latin Hypercube  
N-rooks Spectrum  
Multi-Jitter 
Multi-Jitter Spectrum

Chiu et al. [1993]

...to obtain denser stratification...
Anisotropic Sampling Power Spectra

N-rooks / Latin Hypercube  N-rooks Spectrum  Multi-jitter  Multi-Jitter Spectrum

Chiu et al. [1993]

...along the horizontal...
...and vertical axis, on top of 2D stratification, which results in multi-jittered samples with a hairline anisotropy along the canonical axes that is visible as a cross in the middle of it’s spectrum. The same ideas extend to...
Sampling in Higher Dimensions

...to higher dimensions. For example, in 4D...
…instead of directly sampling the full 4D space, Rob Cook in [1986] proposed to sample [CLICK] the lower 2D subspaces first, UV and XY here, and then randomly permute these 2D samples to form [CLICK] 4D tuples, which can then be used to evaluate an underlying 4D integrand. In practice…
...instead of directly sampling the full 4D space, Rob Cook in [1986] proposed to sample [CLICK] the lower 2D subspaces first, UV and XY here, and then randomly permute these 2D samples to form [CLICK] 4D tuples, which can then be used to evaluate an underlying 4D integrand. In practice...
...instead of directly sampling the full 4D space, Rob Cook in [1986] proposed to sample [CLICK] the lower 2D subspaces first, UV and XY here, and then randomly permute these 2D samples to form [CLICK] 4D tuples, which can then be used to evaluate an underlying 4D integrand. In practice...
…rendering systems tend to use jittered samples on these 2D subspaces. It is, however, beneficial to use more sophisticated samplers like Poisson disk which are more uniform. If we look at the power spectra for these underlying 2D projections…
...rendering systems tend to use jittered samples on these 2D subspaces. It is, however, beneficial to use more sophisticated samplers like Poisson disk which are more uniform. If we look at the power spectra for these underlying 2D projections...
...they are [CLICK] isotropic in nature. However, if look at the mixed 2D projections XU...
...they are [CLICK] isotropic in nature. However, if look at the mixed 2D projections XU...
... it has hairline anisotropy along the [CLICK] axes. [CLICK] The same is true for YV projections.
... it has hairline anisotropy along the [CLICK] axes. [CLICK] The same is true for YV projections.
Let's look at the power spectrum of an N-rooks expected power spectrum to understand the effect of these anisotropic structures on the variance convergence rate.
N-rooks expected spectrum has [CLICK] the same radial profile along the canonical axes, and [CLICK] a constant radial profile along all other directions. For the convergence rate, we only need…
N-rooks expected spectrum has [CLICK] the same radial profile along the canonical axes, and [CLICK] a constant radial profile along all other directions. For the convergence rate, we only need…
N-rooks expected spectrum has [CLICK] the same radial profile along the canonical axes, and [CLICK] a constant radial profile along all other directions. For the convergence rate, we only need…
N-rooks expected spectrum has [CLICK] the same radial profile along the canonical axes, and [CLICK] a constant radial profile along all other directions. For the convergence rate, we only need…
... one of the canonical direction (shown in orange) and one of the direction from the rest of the spectrum (shown in green) since the behavior is the same in all other directions. Now, depending on the integrands we can observe different convergence rates for the same sampler. For example,...
...for a step function with all the energy in the spectrum along the horizontal direction. Due to the dark hairline anisotropy [CLICK] present in the sampling spectrum, their [CLICK] product goes down very quickly, resulting in huge variance reduction and good asymptotic convergence. However, for the second pixel...
...for a step function with all the energy in the spectrum along the horizontal direction. Due to the dark hairline anisotropy [CLICK] present in the sampling spectrum, their [CLICK] product goes down very quickly, resulting in huge variance reduction and good asymptotic convergence. However, for the second pixel...
Var(\(I_N\)) = \(\sum_\Omega \langle P_{SN}(\nu) \rangle \times P_f(\nu) = \sum_\Omega \langle \rangle\) (N-rooks spectrum, Integrand spectrum)

...for a step function with all the energy in the spectrum along the horizontal direction. Due to the dark hairline anisotropy [CLICK] present in the sampling spectrum, their [CLICK] product goes down very quickly, resulting in huge variance reduction and good asymptotic convergence. However, for the second pixel...
\[ \text{Var}(I_N) = \sum_{\Omega} f(x) \times \text{N-rooks spectrum} \times \text{Integrand spectrum} = \sum_{\Omega} \text{Var}(\text{IN}) = X \]

... since the integrand spectrum has energy spread over all the directions, the [CLICK] hairline anisotropy of the sampling spectrum [CLICK] does not significantly reduce the product, resulting in higher variance. We further verified this.
\[
\text{Var}(I_N) = \sum_{\Omega} \langle P_{SN}(\nu) \rangle \times P_f(\nu) = \sum_{\Omega} \text{Var}(I_N)
\]

... since the integrand spectrum has energy spread over all the directions, the hairline anisotropy of the sampling spectrum does not significantly reduce the product, resulting in higher variance. We further verified this.
\[
\text{Var}(I_N) = \sum_{\Omega} \text{N-rooks spectrum} \times \text{Integrand spectrum} = \sum_{\Omega}
\]

... since the integrand spectrum has energy spread over all the directions, the [CLICK] hairline anisotropy of the sampling spectrum [CLICK] does not significantly reduce the product, resulting in higher variance. We further verified this.
...experimentally, where we plot variance with increasing sample count. This shows that if we can align the anisotropic structures of the sampling spectrum $P_s$ with that of the integrand spectrum $P_f$, we can gain huge variance reductions, as shown with the magenta curve. But in most scenarios...
...experimentally, where we plot variance with increasing sample count. This shows that if we can align the anisotropic structures of the sampling spectrum $P_s$ with that of the integrand spectrum $P_f$, we can gain huge variance reductions, as shown with the magenta curve. But in most scenarios...
...the underlying integrand spectrum has arbitrary orientation. If we choose to sample this function...
…with multi-jittered samples which has [CLICK] the following power spectrum, we won’t be able to benefit from these hairline anisotropic structures since they are axis-aligned. To solve this issue, we propose to shear…
...with multi-jittered samples which has [CLICK] the following power spectrum, we won’t be able to benefit from these hairline anisotropic structures since they are axis-aligned. To solve this issue, we propose to shear...
Shearing Multi-Jittered Samples

Sheared Samples

Sheared Spectrum

Integrand Spectrum

Singh and Jarosz [SIGGRAPH 2017]

...the samples in such a way that we can align the sampling spectrum with that of the integrand spectrum. Now, to show the improvements due to shearing, let's look at a rendering example.
We visualize the variance heatmap; For this, we render the same cornellbox scene multiple times generating 100 images with uncorrelated-multijittered samples followed by computing the variance of each pixel over these 100 images. The resulting variance heatmap is shown as a gray scale image where the brighter pixels mean high variance. After shearing the samples in the XU and YV subspaces…
We visualize the variance heatmap; for this, we render the same cornellbox scene multiple times generating 100 images with uncorrelated-multijittered samples followed by computing the variance of each pixel over these 100 images. The resulting variance heatmap is shown as a gray scale image where the brighter pixels mean high variance. After shearing the samples in the XU and YV subspaces…
We visualize the variance heatmap; For this, we render the same cornellbox scene [CLICK] multiple times generating 100 images with uncorrelated-multijittered samples followed by computing the variance of each pixel over these 100 images. The resulting variance heatmap [CLICK] is shown as a gray scale image where the brighter pixels mean high variance. After shearing the samples in the XU and YV subspaces...
...we observe dramatic variance reduction in pixels with no occluders and modest improvement at pixels with discontinuities. The problem here is that improvements come after a really large N. The major reason for this limited improvement is that existing samplers have only hairline anisotropic structures.
So far...

Blue noise samplers can have better convergence compared to stratified samples.

Denser stratification can lead to anisotropic spectra which improves convergence.
What properties we desire in a sampler?

Progressivity

High speed (millions of samples per second)

Extension to dimensions beyond 2D
(Spoke dart throwing, Mitchell [2018])
What properties we desire in a sampler?

- **Progressivity** ✓ (Ahmed et al. [2017], Christensen et al. [2018])

- High speed (millions of samples per second)

- Extension to dimensions beyond 2D
  (Spoke dart throwing, Mitchell [2018])
What properties we desire in a sampler?

- Progressivity ✓ (Ahmed et al. [2017], Christensen et al. [2018])
- High speed (millions of samples per second) ✓
- Extension to dimensions beyond 2D
  (Spoke dart throwing, Mitchell [2018])
What properties we desire in a sampler?

Progressivity \checkmark \quad (Ahmed et al. [2017], Christensen et al. [2018])

High speed (millions of samples per second) \checkmark

Extension to dimensions beyond 2D \times
\quad (Spoke dart throwing, Mitchell [2018])
Low-Discrepancy Sampling

**Deterministic** sets of points specially crafted to be evenly distributed (have low discrepancy).

Entire field of study called Quasi-Monte Carlo (QMC)
The Van der Corput Sequence

Table: Radical Inverse $\Phi_b$ in base 2

<table>
<thead>
<tr>
<th>$k$</th>
<th>Base 2</th>
<th>$\Phi_b$</th>
</tr>
</thead>
</table>

Subsequent points “fall into biggest holes”

best discrepancy for infinite sequence
The Van der Corput Sequence

Radical Inverse $\Phi_b$ in base 2

Subsequent points “fall into biggest holes”

<table>
<thead>
<tr>
<th>$k$</th>
<th>Base 2</th>
<th>$\Phi_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$0.1 = 1/2$</td>
</tr>
</tbody>
</table>

best discrepancy for infinite sequence
The Van der Corput Sequence

Radical Inverse $\Phi_b$ in base 2

Subsequent points “fall into biggest holes”

<table>
<thead>
<tr>
<th>$k$</th>
<th>Base 2</th>
<th>$\Phi_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.1 = 1/2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.01 = 1/4</td>
</tr>
</tbody>
</table>

best discrepancy for infinite sequence
The Van der Corput Sequence

Radical Inverse $\Phi_b$ in base 2

Subsequent points “fall into biggest holes”

<table>
<thead>
<tr>
<th>$k$</th>
<th>Base 2</th>
<th>$\Phi_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.1 = 1/2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.01 = 1/4</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>.11 = 3/4</td>
</tr>
</tbody>
</table>

best discrepancy for infinite sequence
The Van der Corput Sequence

Radical Inverse $\Phi_b$ in base 2
Subsequent points “fall into biggest holes”

<table>
<thead>
<tr>
<th>$k$</th>
<th>Base 2</th>
<th>$\Phi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.1 = 1/2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.01 = 1/4</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>.11 = 3/4</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>.001 = 1/8</td>
</tr>
</tbody>
</table>

best discrepancy for infinite sequence
The Van der Corput Sequence

Radical Inverse $\Phi_b$ in base 2

Subsequent points “fall into biggest holes”

<table>
<thead>
<tr>
<th>$k$</th>
<th>Base 2</th>
<th>$\Phi_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.1 = 1/2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.01 = 1/4</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>.11 = 3/4</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>.001 = 1/8</td>
</tr>
<tr>
<td>5</td>
<td>101</td>
<td>.101 = 5/8</td>
</tr>
</tbody>
</table>

best discrepancy for infinite sequence
The Van der Corput Sequence

Radical Inverse $\Phi_b$ in base 2

Subsequent points “fall into biggest holes”

<table>
<thead>
<tr>
<th>$k$</th>
<th>Base 2</th>
<th>$\Phi_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.1 = 1/2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.01 = 1/4</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>.11 = 3/4</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>.001 = 1/8</td>
</tr>
<tr>
<td>5</td>
<td>101</td>
<td>.101 = 5/8</td>
</tr>
<tr>
<td>6</td>
<td>110</td>
<td>.011 = 3/8</td>
</tr>
</tbody>
</table>

best discrepancy for infinite sequence
The Van der Corput Sequence

Radical Inverse $\Phi_b$ in base 2

Subsequent points “fall into biggest holes”

<table>
<thead>
<tr>
<th>$k$</th>
<th>Base 2</th>
<th>$\Phi_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.1 = 1/2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.01 = 1/4</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>.11 = 3/4</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>.001 = 1/8</td>
</tr>
<tr>
<td>5</td>
<td>101</td>
<td>.101 = 7/8</td>
</tr>
<tr>
<td>6</td>
<td>110</td>
<td>.011 = 3/8</td>
</tr>
<tr>
<td>7</td>
<td>111</td>
<td>.111 = 7/8</td>
</tr>
</tbody>
</table>

best discrepancy for infinite sequence
The Van der Corput Sequence

Radical Inverse $\Phi_b$ in base 2

Subsequent points “fall into biggest holes”

<table>
<thead>
<tr>
<th>$k$</th>
<th>Base 2</th>
<th>$\Phi_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.1 = 1/2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.01 = 1/4</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>.11 = 3/4</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>.001 = 1/8</td>
</tr>
<tr>
<td>5</td>
<td>101</td>
<td>.101 = 5/8</td>
</tr>
<tr>
<td>6</td>
<td>110</td>
<td>.011 = 3/8</td>
</tr>
<tr>
<td>7</td>
<td>111</td>
<td>.111 = 7/8</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

best discrepancy for infinite sequence
Halton and Hammersley Points

**Halton**: Radical inverse with different base for each dimension:

\[ \tilde{x}_k = (\Phi_2(k), \Phi_3(k), \Phi_5(k), \ldots, \Phi_{p_n}(k)) \]
Halton and Hammersley Points

**Halton**: Radical inverse with different base for each dimension:

\[ \bar{x}_k = (\Phi_2(k), \Phi_3(k), \Phi_5(k), \ldots, \Phi_{p^n}(k)) \]

- The bases should all be relatively prime.
Halton and Hammersley Points

**Halton:** Radical inverse with different base for each dimension:

\[ \vec{x}_k = (\Phi_2(k), \Phi_3(k), \Phi_5(k), \ldots, \Phi_{p_n}(k)) \]

- The bases should all be relatively prime.
- Incremental/progressive generation of samples
Halton and Hammersley Points

**Halton**: Radical inverse with different base for each dimension:

\[ \vec{x}_k = (\Phi_2(k), \Phi_3(k), \Phi_5(k), \ldots, \Phi_{p_n}(k)) \]

- The bases should all be relatively prime.
- Incremental/progressive generation of samples

**Hammersley**: Same as Halton, but first dimension is \( k/N \):

\[ \vec{x}_k = (k/N, \Phi_2(k), \Phi_3(k), \Phi_5(k), \ldots, \Phi_{p_n}(k)) \]
Halton and Hammersley Points

**Halton**: Radical inverse with different base for each dimension:

\[ \bar{x}_k = (\Phi_2(k), \Phi_3(k), \Phi_5(k), \ldots, \Phi_{p_n}(k)) \]

- The bases should all be relatively prime.
- Incremental/progressive generation of samples

**Hammersley**: Same as Halton, but first dimension is \( k/N \):

\[ \bar{x}_k = (k/N, \Phi_2(k), \Phi_3(k), \Phi_5(k), \ldots, \Phi_{p_n}(k)) \]

- Not incremental, need to know sample count, \( N \), in advance
The Hammersley Sequence

1 sample in each "elementary interval"
The Hammersley Sequence

1 sample in each "elementary interval"
The Hammersley Sequence

1 sample in each "elementary interval"
The Hammersley Sequence

1 sample in each "elementary interval"
The Hammersley Sequence

1 sample in each "elementary interval"
The Hammersley Sequence

1 sample in each "elementary interval"
Why do we need to scramble?

Halton Projection (29, 31)

These deterministic samplers can have voids in the domain when the sample count is not exactly dyadic. [CLICK] We can scramble these samples to fill the domain. Let’s see some rendering results.
Why do we need to scramble?

Halton Projection (29, 31)  Scrambled Halton Projection (29, 31)

These deterministic samplers can have voids in the domain when the sample count is not exactly dyadic. [CLICK] We can scramble these samples to fill the domain. Let’s see some rendering results.
Low discrepancy samplers show less noise compared to randomly jittered samples.
Monte Carlo (16 jittered samples)
Can we combine blue noise properties with low discrepancy?
Recently, couple of papers were published that incorporate certain stratification properties to different blue noise targets resulting in lowering their discrepancy properties.
Recently, couple of papers were published that incorporate certain stratification properties to different blue noise targets resulting in lowering their discrepancy properties.
Low-Discrepancy Blue Noise

The plots are shown where the discrepancy is comparable to Sobol.
The corresponding variance convergence is also comparable.
Perrier and colleagues took a step further and developed a smart scrambling strategy that introduces blue noise properties directly into the well known sobol sequences.
Perrier and colleagues took a step further and developed a smart scrambling strategy that introduces blue noise properties directly into the well known sobol sequences.
Perrier and colleagues took a step further and developed a smart scrambling strategy that introduces blue noise properties directly into the well known sobol sequences.
This also shows improvements in rendering quality near the edges of the scene.
In the end, we now briefly touch upon correlated importance sampling which is recently published at CGF and we are hoping to present it at EGSR this year.
In the end, we now briefly touch upon correlated importance sampling which is recently published at CGF and we are hoping to present it at EGSR this year.
In the end, we now briefly touch upon correlated importance sampling which is recently published at CGF and we are hoping to present it at EGSR this year.
Two common importance sampling (IS):
light IS: generate samples on the light source and generate shadow rays to the hit point. This results in a smooth looking underlying integrand at that hit point.
BSDF IS: When the visible hemisphere is sampled, the samples see the light source boundary as a discontinuity, making the underlying integrand $C_0$ discontinuous.
This directly reflected into the corresponding convergence rates. We focus on two pixels: Pixel P (the hit point is directly visible from the light source) and Pixel Q (partially occluded from the light source).
light IS for unoccluded hit points (Pixel P with no occluders) shows good convergence rate compared to partially occluded hit points. BSDF IS however does not improve any convergence behavior irrespective of whether the hit point is occluded or not.
Occluded pixels (no improvement in convergence)

Reference

Underlying pixel functions

Pixel P

Pixel Q

Singh et al. [2019]

More at EGSR ;)
Occluded pixels (no improvement in convergence)

More at EGSR ;)}
Future directions:
Might be interesting to generate samples with correlations that can have wider anisotropic structures.
Future research directions

Direct link between spatial and Fourier statistics needs further investigation

Progressive samplers in higher dimensions

Adapting sample correlations w.r.t. the underlying integrand in high dimensions
Acknowledgments

Some slides borrowed from Wojciech Jarosz and Kartic Subr

All the anonymous reviewers who helped shape this survey paper into its final form
Acknowledgments

Some slides borrowed from Wojciech Jarosz and Kartic Subr

All the anonymous reviewers who helped shape this survey paper into its final form