EUROGRAPHICS 2014 / Wojciech Jarosz and Pieter Peers
(Guest Editors)

Volume 33 (2014), Number 4

Error analysis of estimators that use combinations of stochastic sampling strategies for direct illumination

Kartic Subr[1], Derek Nowrouzezahrai[2], Wojciech Jarosz[1], Jan Kautz[3] and Kenny Mitchell[1]

[1]Disney Research, Zurich, [2] University of Montreal, [3] University College London

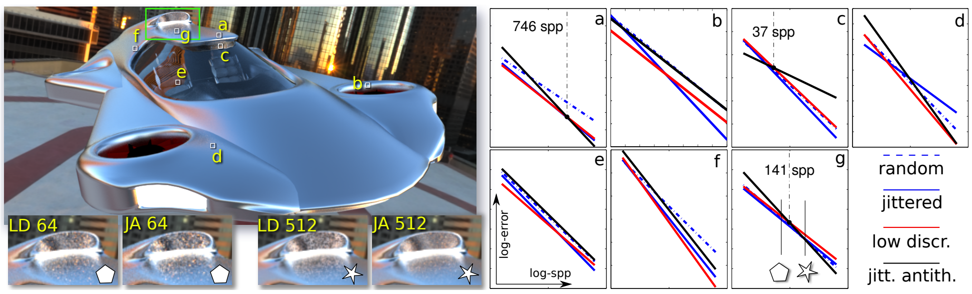

Figure 1: The "optimal" sampling strategy varies as a function of the sample count as well as spatially over pixels. A formal study of the variance of combinations of strategies is necessary to understand such behaviour. We compare combinations of four different sampling strategies with multiple importance sampling to render the image (left). The rates at which numerical integration errors decrease with increasing sample counts (per pixel; spp) are shown at 7 pixels (a-g). Our theoretical analysis provides insight into such behaviour, and motivates a new, jittered antithetic importance sampling, estimator (black) for rendering.

Abstract

We present a theoretical analysis of error of combinations of Monte Carlo estimators used in image synthesis. Importance sampling and multiple importance sampling are popular variance-reduction strategies. Unfortunately, neither strategy improves the rate of convergence of Monte Carlo integration. Jittered sampling (a type of stratified sampling), on the other hand is known to improve the convergence rate. Most rendering software optimistically combine importance sampling with jittered sampling, hoping to achieve both. We derive the exact error of the combination of multiple importance sampling with jittered sampling. In addition, we demonstrate a further benefit of introducing negative correlations (antithetic sampling) between estimates to the convergence rate. As with importance sampling, antithetic sampling is known to reduce error for certain classes of integrands without affecting the convergence rate. In this paper, our analysis and experiments reveal that importance and antithetic sampling, if used judiciously and in conjunction with jittered sampling, may improve convergence rates. We show the impact of such combinations of strategies on the convergence rate of estimators for direct illumination.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

The estimation of light energy recorded by each pixel on the sensor of a virtual camera involves numerical integration over an infinite set of light paths that arrive at the pixel. An *estimator* for this integral is typically obtained by averaging the radiance along representative *samples* of light paths. Error in this estimation manifests as variance (image noise) and bias (usually structured artifacts). Two sampling considerations crucially impact the extent of error — the number of samples and the strategy used to select the representative samples.

The *convergence rate* of an estimator describes how rapidly its error decreases as the number of samples is increased, and is an important consideration for assessing estimators. This has been extensively studied in statistics [Owe13] and for some estimators in computer graphics [Mit96, KPR12]. Convergence rate is typically estimated as the slope of the plot of error vs number of samples in log-log scale. Fig. 1 visualises the convergence plots at 8 different pixels (a-g), for four different estimators. The observed error and convergence rate depend on the particular sampling strategy used. There is a large body of literature [Vea97, KPR12] devoted to mitigating error. Some methods achieve this by tailoring the distribution of samples. For example, importance sampling (IS) is beneficial when the integrand is highly varying and such behavior is known a priori. Unfortunately, the benefit of IS is constant and diminishes as more samples are drawn, i.e. IS does not affect the convergence rate. A second strategy imposes positional constraints, e.g. partitioning constraints of stratified sampling and the minimum separation constraints (Poisson-disk, halton sampling, etc.). These methods are popular because they exhibit improved convergence rates.

The obvious solution, of combining some of the above strategies, is a standard feature in most rendering software [PH10, Jak10]. While such a combination is observed to improve the quality of rendered images in many cases, we are not aware of a formal analysis of the variance of these combined estimators. One of our contributions is that we derive the exact error (cf. Fig. 2) – the convergence rate as well as constant term. Our equations for variance of jittered importance sampling and jittered multiple importance sampling (MIS) reveal counter-intuitive results such as the possibility of increasing estimator variance by stratifying MIS.

A third strategy, antithetic sampling [HM56b] (AS) is a classical strategy in the statistics literature that reduces error by generating samples that are inversely correlated with each other. Although it boasts the ability to produce zero-error estimates for special cases (linear and skew-symmetric integrands), this method is no longer heavily applied [SP09] for two reasons: first, it can increase error for complex integrands, and it is not as easy to manipulate as importance sampling in these cases. As with IS, antithetic sampling also does not impact convergence rate. However, recent work has derived improved convergence rates for combinations of this approach with low-discrepancy sequences [Owe08].

Combining jittered, importance and antithetic sampling provides many benefits. Jittered sampling improves convergence for smooth integrands but suffers in strata that contain discontinuities [Mit96]. On the other hand, while IS does not improve convergence, it is useful for dealing with discontinuities that are known a priori. Combining JS with a judicious choice of importance function reduces error in strata with discontinuities. For maximum benefit from this combination, however, the integrand divided by the importance function needs to be constant. The advantage of including AS in the mix is that it is sufficient that the ratio is linear (rather than constant). Integrals within such strata will be estimated with zero-error due to AS. Thus, jittered antithetic importance sampling (JAIS) is an effective combination for improving convergence, and introduces a fresh perception of what constitutes a "good" importance function. Motivated by these observations, and from recent resurgence of AS, we derive exact errors for the combination of JS with IS and MIS and further qualitatively analyse their combination with AS. We demonstrate the validity of our analysis with estimators for direct illumination.

2. Related work and motivation

We broadly classify stochastic sampling strategies into distribution- and arrangement-based methods. The former, such as importance sampling, reduce error by tailoring the parent distributions of the samples. The latter set imposes constraints on sample placement e.g. by subdividing the sampling domain into sub-domains [Mit96], ensuring a minimum sample spacing [Coo86], or guaranteeing that a certain fraction of samples lie in specific subdomains [Shi91].

Distribution: Monte Carlo (MC) estimators using non-uniformly distributed samples normalise (divide) each evaluation by the probability density evaluated at that sampling location. IS tailors this distribution to have the normalized integrand evaluations remain as constant as possible across samples, and the reduction in error depends on this degree of constancy. Ideally, the normalized integrand evaluations are constant; for this, the distribution must be proportional to the integrand. On the other hand, for the distribution to be a valid probability density function (pdf) the proportionality constant is precisely the value of the integral being estimated. As such, the ideal importance function is of little practical value since it assumes knowledge of the quantity being estimated. In image synthesis, IS is a popular choice and has been applied to integrate distant illumination [ARBJ03], material BRDFs [LRR04], sums of distributions via multiple importance sampling (MIS) [Vea97], products of distributions [CJAMJ05] and participating media [GKH*13]. Pharr and Humphreys [PH10] provide a more comprehensive survey. The benefit of these approaches are typically verified by visual inspection, where the variance manifests as noise, and each of these algorithms exploits specific knowledge of the integrand from which heuristic importance functions are proposed. For a more general overview readers are referred to primers [Owe13, OZ00, Bis06, Hes03].

Arrangement for antithetic variates: Antithetic variates reduce variance using negatively correlated random variables [HM56b, HM56a]. For every sampling location, a corresponding "antithetic" location is generated. The hope is that the values of the integrand at these locations are negatively correlated. The estimator is unbiased so long as the distribution of the samples remains unchanged (see Sec. 6). It is possible that perfectly inversely correlated samples lead to positively correlated estimates at those locations. In such cases, the estimator's variance would increase. In image synthesis, these sampling decisions are performed dynamically and so we derive a simple mechanism, inspired by copula theory [SS83], to control the degree and sign of correlation between random samples of arbitrary dimensionality (Sec. 4). We also present equations that test for this pre-emptively (Sec. 6). Recent sampling techniques were thought to have superseded antithetic sampling, even to the point that it was believed to be redundant [SP09]. Our analysis and experiments with direct illumination show that it is beneficial to combine antithetic variables (in a complementary manner) with more recent techniques. We analyze the combination of stratified sampling and IS with antithetic variates (Sec. 4.5) and demonstrate a marked improvement with this scheme.

Arrangement for stratified sampling: Stratified sampling is a powerful divide-and-conquer variance reduction scheme designed to cope with integrands that exhibit piecewise homogeneous behavior. The domain is partitioned into subdomains and estimates of the integral within each of subdomain are computed and carefully combined to estimate the integral over the entire domain. Naïve MC algorithms exhibit $O(N^{-0.5})$ convergence; yet, with cubically stratified sampling, Haber [Hab70] confirmed that an improved $O(N^{-1-p/d})$ convergence is attainable for d-dimensional integrals when the integrand is p-differentiable. The practical application of stratified sampling in statistics fundamentally differs from jittered sampling used in graphics. Statisticians actively strive to seek a meaningful partitioning of the domain [Ney34]. Further, they might choose to tailor the importance function within each stratum [Owe13]. While their analyses are not directly applicable to image synthesis, we draw inspiration from such work to analyse a potentially useful combination for image synthesis. *Jittered sampling* is a special form of stratified sampling in which the subdomains all have equal area (usually a regular grid of cells) and sample allocation is equal across strata. It has been observed empirically [Mit96] (for anti-aliasing) and analytically [RAMN12] (for visibility integration) that jittered sampling leads to an estimator that converges at $O(N^{-1.5})$, as opposed to $O(N^{-0.5})$ (naïve MC). We start with a derivation for this in 1D, which provides insight into the benefit of combining antithetic variables and IS with stratification (see Sec. 4.5), and explain why it typically improves convergence in our experiments.

Arrangement for quasi-Monte Carlo: While random sampling guarantees that, on average, the proportion of points generated within any subset of the domain is proportional to the measure of the subset, no such guarantee holds for particular instances (say a subset) of samples. A quantitative measure of this desirable property is called *discrepancy* and was introduced to the graphics community by Shirley [Shi91]. We again point interested readers to several comprehensive references on low-discrepancy sampling [DP10, Lem10, KHN06, Nie92], including some specialized to rendering [Kel96, KPR12]. Intuitively, the goal of low-discrepancy sampling is to arrange samples uniformly in the domain so that the number of points in any chosen stratum is kept constant over instances of sampling patterns. This, in combination with deterministic sampling, allows QMC to exhibit improved convergence rates of $O(N^{-1+\varepsilon})$. Randomized QMC can be even lower, at $O(N^{-3/2+\varepsilon})$, however with moderate smoothness constraints on the integrand, for any $\varepsilon > 0$.

We exclusively analyse Monte Carlo estimators. It is possible that our insights into combining stratification with other variance reduction schemes extend to QMC methods; however, theoretical derivations of similar results for QMC are beyond the scope of our work since it is unclear how to extend our probabilistic analysis to the number theoretic proofs usually required for QMC methods. We perform empirical comparisons against popular QMC methods in image synthesis.

Copulas: Copulas capture complex inter-relationships between random variables. A copula parameterizes such interdependencies separately from the parameters of the random numbers' marginal distributions. We use Gaussian copulas as a tool to generate random variables that are uniformly distributed in $[0,1]$ with specified mutual correlations. Although the Gaussian copula has been criticised for its inability to cope with non-linear relationships between variables, it suffices for our purpose of variance analysis. In Sec. 4, we explain how the copulas may be useful for graphics applications. Interested readers are referred to standard textbooks [SS83, Nel06] for formal definitions.

Combining variance reduction schemes: In image synthesis, the combination of regular sampling with IS [KC08] and that of IS using multiple functions [VG95] have been explored. In statistics, low-discrepancy sampling has been combined with locally antithetic samples [Owe08] to achieve a convergence of $O(N^{-3/2-1/d+\varepsilon})$. More recently, in concurrent unpublished work in statistics [Owe13], the combination of jittered sampling with antithetic variables is explored.

In this paper, we derive exact errors for the combination of jittered sampling with importance sampling and MIS and further qualitatively analyse their combination with antithetic sampling.

Contributions: We make the following contributions:

- derivations of exact error for popular estimators such as IS, MIS, jittered IS and jittered MIS (Fig. 2);
- introduce jittered antithetic importance sampling (JAIS) and assess conditions that improve its convergence rate;
- introduce a mechanism to generate multidimensional samples with specified mutual correlations; and,
- verify the validity of JAIS for estimating direct illumination.

Insight: Our analysis confirms previously observed behavior and also reveals interesting properties:

- for certain N, jittering increases the error of IS;
- using MIS to sample from two distributions, such as lighting and BRDF, produces lower error than importance sampling with their averaged pdfs;
- the average pdf is no longer a better choice (compared to MIS) when combined with jittered sampling;
- for IS with AS, it is sufficient that the ratio of the integrand to the importance function is linear (rather than constant);
- and a judicious choice of importance functions can improve the convergence rate of jittered antithetic importance sampling.

name	symbol	form	variance
random	$\tilde{\mu}_{\mathrm{N}}$	$\int_{\mathcal{D}} f(x)\ \mathrm{d}x$	$\frac{1}{N}\left(\int_{\mathcal{D}} f^2(x)\ \mathrm{d}x - I^2\right)$
importance	$\tilde{\mu}_{\mathrm{isN}}$	$\int_{\mathcal{D}} \frac{f(x)}{g(x)} g(x)\ \mathrm{d}x$	$\frac{1}{N}\left(\int_{\mathcal{D}} \frac{f^2(x)}{g(x)}\ \mathrm{d}x - I^2\right)$
multiple imp.	$\tilde{\mu}_{\mathrm{misN}}$	$N_g \sum_{k=1}^{N_g} \int_{\mathcal{D}} \frac{f(x)}{\sum g_k(x)} g_k(x)\ \mathrm{d}x$	$\mathrm{V}\left(\tilde{\mu}_{\mathrm{avg.isN}}\right) + \frac{\kappa}{N_g N}$
jittered	$\tilde{\mu}_{\mathrm{jsN}}$	$\sum_{i=1}^{N} \int_{\mathcal{D}_i} f(x)\ \mathrm{d}x$	$\frac{1}{N}\mathrm{V}\left(\tilde{\mu}_{\mathrm{N}}\right) + \frac{2}{N^2}\sum_{i=1}^{N}\sum_{j<i} I_i I_j$
jittered imp.	$\tilde{\mu}_{\mathrm{jisN}}$	$\sum_{i=1}^{N} \int_{\mathcal{D}_i} \frac{f(x)}{g(x)} g(x)\ \mathrm{d}x$	$\frac{1}{N}\mathrm{V}\left(\tilde{\mu}_{\mathrm{isN}}\right) + \frac{2}{N^2}\sum_{i=1}^{N}\sum_{j<i} I_i I_j$
jittered MIS	$\tilde{\mu}_{\mathrm{jmisN}}$	$N_g \sum_{i=1}^{N}\sum_{k=1}^{N_g} \int_{\mathcal{D}_i} \frac{f(x)}{\sum g_k(x)} g_k(x)\ \mathrm{d}x$	$\frac{N_g}{N}\mathrm{V}\left(\tilde{\mu}_{\mathrm{avg.isN}}\right) + \frac{N_g^2}{N^2}\psi$

where

$$I = \int_{\mathcal{D}} f(x)\ \mathrm{d}x$$

$$I_i = \int_{\mathcal{D}_i} f(x)\ \mathrm{d}x$$

$$I_{\bar{g}i} = \int_{\mathcal{D}} \frac{f(x)}{\sum g_k(x)} g_i(x)\ \mathrm{d}x$$

$$I_{\bar{g}i}^{j} = \int_{\mathcal{D}_j} \frac{f(x)}{\sum g_k(x)} g_i(x)\ \mathrm{d}x$$

$$\kappa = \sum_{j=1}^{N_g}\sum_{j=1}^{N_g} I_{\bar{g}i} I_{\bar{g}j} - N_g(N_g-1)I^2$$

$$\psi = \sum_{i=1}^{i<\frac{N}{N_g}}\sum_{j\neq i}^{j<\frac{N}{N_g}} I_{\bar{g}i} I_{\bar{g}j} - \frac{N_g-1}{N_g} I^2 + \sum_{i=1}^{i<N_g}\sum_{j=1}^{j<\frac{N}{N_g}}\sum_{k\neq j}^{k<\frac{N}{N_g}} I_{\bar{g}i}^{j} I_{\bar{g}i}^{k}$$

Figure 2: A tabulation of our derivations of the exact expressions for variance for a number of estimators that are commonly used in graphics. $\tilde{\mu}_{\mathrm{N}}$, $\tilde{\mu}_{\mathrm{sN}}$, $\tilde{\mu}_{\mathrm{isN}}$, $\tilde{\mu}_{\mathrm{jisN}}$, $\tilde{\mu}_{\mathrm{misN}}$ and $\tilde{\mu}_{\mathrm{jmisN}}$ are N-sample MC estimators obtained by rewriting the integral of $f(x)$ over $\mathcal{D}$ as shown in the "form" column of the table. The variance, $\mathrm{V}\left(\tilde{\mu}_{\mathrm{avg.isN}}\right)$, of the N-sample average importance sampled estimator is simply $\mathrm{V}\left(\tilde{\mu}_{\mathrm{isN}}\right)$ with $g(x)$ set as the average of the individual pdfs. See supplemental material A for their derivations.

3. Exact error analysis

We derived the errors of popular estimators (summarised in Figure 2) for direct illumination: IS, MIS, jittered IS, and jittered MIS. Deriving these error rates, while involved, is a standard exercise in statistics. The full derivations can be found in supplemental material A. In the remainder of this section, we interpret the results and explain their significance.

Notation: We analyse Monte Carlo estimators for the multidimensional integral $I = \int_{\mathcal{D}} f(x)\mathrm{d}x,\ x \in \mathcal{D}$ where $f(x)$ is a real-valued function defined over the domain. We use $\tilde{\mu}$ to denote an estimator for I along with a suffix to identify the sampling strategy and sample count. We capitalize random variables ($\mathbf{X}$, $\mathbf{Y}_i$, etc.). For example, the primary importance sampling estimator is $\tilde{\mu}_{\mathrm{is},1} = f(\mathbf{X})/g(\mathbf{X}),\ \ \mathbf{X} \sim g(x)$ and the secondary estimator is $\tilde{\mu}_{\mathrm{is,N}} = \frac{1}{N}\sum f(\mathbf{X}_i)/g(\mathbf{X}_i),\ \ \mathbf{X}_i \sim g(x)$. Table 1 lists our notation for expectation and variance.

Jittered importance sampling: From the variance of the JIS estimator (Fig. 2), an upper bound is obtained as

$$\mathrm{V}\left(\tilde{\mu}_{\mathrm{jis,N}}\right) \ \leq\ \frac{\mathrm{V}\left(\tilde{\mu}_{\mathrm{is,N}}\right)}{N} + \frac{N-1}{N^3} I^2, \qquad (1)$$

since $I = \sum I_i$. This upper bound is achieved when the integral over each of the strata is equal: $I_i = I_j, \forall i, j$. Eq. 1 shows that the combination of jittered sampling with importance sampling leads to improved convergence with error $O(N^{-1})$ instead of $O(N^{-1/2})$ achieved by importance sampling. There is a constant additive error which also decays as $O(N^{-1})$. Interestingly, if the number of samples is chosen so that $N \leq I/\sqrt{\mathrm{V}\left(\tilde{\mu}_{\mathrm{is,N}}\right)}$, then $\mathrm{V}\left(\tilde{\mu}_{\mathrm{jis,N}}\right) > \mathrm{V}\left(\tilde{\mu}_{\mathrm{is,N}}\right)$, i.e. *it is possible that jittering increases the error of IS*.

Table 1: Notation used in this paper.

Symbol	Definition
$f(x)$	integrand
$g(x)$	sampling distribution
$\tilde{\mu}_{\mathrm{N}}$	estimate of I using N samples
$\tilde{\mu}_{\mathrm{is,N}}$	N-sample importance sampling estimator
$\tilde{\mu}_{\mathrm{is,1}}$	primary (1-sample) imp. samp. estimator
$\mathbf{X}$, $\mathbf{Y}$, $\mathbf{X}_i$, $\mathbf{Y}_i$	random variates
$\langle\phi(\mathbf{X})\rangle_g$	Exp. with $\mathbf{X} \sim g(x)$: $\int_{\mathcal{D}} \phi(x)\ g(x)\ \mathrm{d}x$
$\mathrm{V}\left(\phi(\mathbf{X})\right)_g$	Var.: $\int_{\mathcal{D}} \phi^2(x)g(x)\ \mathrm{d}x - \int_{\mathcal{D}} \phi(x)g(x)\ \mathrm{d}x$
$\langle\phi(\mathbf{X})\rangle$	$\langle\phi(\mathbf{X})\rangle_g$ with $g(x) = 1/\lvert\mathcal{D}\rvert$
$\mathrm{V}\left(\phi(\mathbf{X})\right)$	$\mathrm{V}\left(\phi(\mathbf{X})\right)_g$ with $g(x) = 1/\lvert\mathcal{D}\rvert$

MIS vs IS with average pdf: Consider the two-distribution MIS estimator. Since $2I = I_{\bar{g}1} + I_{\bar{g}2}$ (by definition), we can show that $\kappa \geq 0$ (κ is defined in Fig. 2) otherwise $(I_{\bar{g}1} - I_{\bar{g}2})^2 < 0$. Thus we can prove by contradiction that $\mathrm{V}\left(\tilde{\mu}_{\mathrm{mis}}\right) \leq \mathrm{V}\left(\tilde{\mu}_{\mathrm{avg,is}}\right)$. A similar argument may be constructed for the case involving more than two sampling distributions. We have shown that, even when it is possible to average two importance functions such as rotated environment lighting and BRDF, *it is preferable (lower variance) to use MIS with the balance heuristic*. Note that the expected value of the variance of the one-sampled MIS estimator is identical to the N-sampled MIS estimator with the balance heuristic since, on average, allocation is equal across the different distributions.

Jittered MIS vs JIS with average pdf: When jittered sampling is combined with MIS, the error term ψ (see Fig. 2) is dominated by the two positive terms. Thus, the variance of the jittered MIS estimator is worse than the jittered estimator with the averaged pdfs. So, where possible, *it is preferable to use the averaged importance functions with jittered samples in the canonical domain than using MIS with the jittered samples*.

4. Combining strategies with antithetic sampling

The variance of a MC estimator is obtained as the sum of the variances of uncorrelated estimates. If the estimates are correlated with each other, then the overall variances consists of an additional covariance term. If this additional term is negative (when correlations are negative), then there is an overall reduction in variance. Negative correlations are simple to introduce for simple (eg. linear) integrands. If the sampling strategy is not designed carefully, for general integrands, the covariance amongst estimates could very well be positive despite the samples being negatively correlated. To avoid this, here we propose the use of a mechanism to control the correlation between samples. The actual parameters provided to this mechanism will depend on the particular application.

4.1. Perfectly antithetic variables

When uniformly distributed random variates $\mathbf{X}$ and $\mathbf{Y}$ are correlated, their corresponding naïve estimators are also potentially correlated. The variance of the estimator $\tilde{\mu}_{\text{av},2} = (f(\mathbf{X}) + f(\mathbf{Y}))/2$ can easily be shown to be $(1+\rho)\mathrm{V}(\tilde{\mu}_2)$ where ρ is the correlation between $f(\mathbf{X})$ and $f(\mathbf{Y})$ and $\tilde{\mu}_2$ is the 2-sample naïve estimator. If $f(\mathbf{X})$ and $f(\mathbf{Y})$ are perfectly negatively correlated, then $\mathbf{X}$ and $\mathbf{Y}$ are said to be *antithetic samples* in the classical sense, and the estimator has zero variance. This is trivially achieved when the integrand is linear and $\mathbf{Y} = c - \mathbf{X}, \forall c > 0$.

AS yields a zero variance estimator when the samples are perfectly inversely correlated and the integrand is either linear or skew symmetric (about the center of the domain). Such cases are rare in image synthesis, where the integrand is often a discontinuous (due to visibility, texture, etc.) superposition of all-frequency functions (lighting, reflectance, texture, visibility, etc.). Under such conditions, it is possible that antithetic sampling would lead to an increase in variance. It is therefore necessary to be able to detect (Sec. 6) the benefit of AS as well as to control the extent to which the variables are antithetic for multidimensional domains.

4.2. Generating correlated variables in 1D: review

Consider a single sample $(\mathbf{X},\mathbf{Y}) \in \Re \times \Re$, or $\in \Re^2$ drawn according to a zero-centered 2D Normal distribution with covariance matrix $M = \begin{pmatrix} 1 & \sigma_{XY} \\ \sigma_{XY} & 1 \end{pmatrix}$. Note that $\mathbf{X}$ and $\mathbf{Y}$ are each (1D) normally distributed with zero mean and unit variance. Let $(\mathbf{U}_x,\mathbf{U}_y) \equiv (\Phi(\mathbf{X}),\Phi(\mathbf{Y})) \in [0,1]\times[0,1]$, where Φ is the cdf of the standard unit normal. By passing the random variates through the cdf, $\mathbf{U}_x$ and $\mathbf{U}_y$ are each uniformly distributed in $[0,1]$ but are still correlated by σ_{XY}. If $\sigma_{XY} = -1$ then the method is analogus to generating antithetic pairs using the popular transformation $(\mathbf{Z}, 1-\mathbf{Z})$ where $\mathbf{Z}$ is uniformly distributed in $[0,1]$.

4.3. Correlated multidimensional sampling

We propose the use of Gaussian copulas to adaptively control the degree to which the samples are antithetic. Consider the goal of generating two d-dimensional random variables $\mathbf{X}_1$ and $\mathbf{X}_2$ that form an antithetic pair where $\mathbf{X}_i = (\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, ..., \mathbf{x}_{i,d}),\ 1 < i \leq 2$. The linear dependencies (correlations) between all components can be fully described using a $2d \times 2d$ correlation matrix Σ, which is symmetric and positive semi-definite by definition. Then, to obtain a pair $(\mathbf{X}_1,\mathbf{X}_2)$ of generalized antithetic variables, we draw $[\mathbf{X}_1\ \mathbf{X}_2]^T$ as a single sample from the $2d$ dimensional zero-mean Normal distribution with Σ as its covariance matrix. Depending on the structure of this correlation matrix, we define three special classes of antithetic pairs.

Conservative multidimensional antithetic pairs: If the correlation matrix $\Sigma = \mathrm{I}_{2d\times 2d} + c$ where $c = \frac{-1}{(4d^2-2d)}$, all inter-dimensional correlations are equally and minimally negative while Σ remains positive semi-definite. If $(\mathbf{X}_1,\mathbf{X}_2)$ is generated using such a structure, we call it a conservative multidimensional antithetic pair.

Uniform multidimensional antithetic pairs: One problem with conservative multidimensional antithetic pairs is that the random variables $\mathbf{X}_1$ and $\mathbf{X}_2$ are not each uniformly distributed in $[0,1]^d$. Thus they lead to biased integrators. To ensure that each variable in the pair is uniformly distributed in $[0,1]^d$, it suffices that the two $d \times d$ blocks along the leading diagonal of Σ are identity matrices. That is, $\Sigma = [\mathrm{I}_{d\times d}\ \ C; C^T\ \ \mathrm{I}_{d\times d}]$ where C is any matrix so long as Σ remains positive semi-definite.

Simple multidimensional antithetic pairs: A special case of uniform antithetic pairs is when C is a diagonal matrix. In this case, $\Sigma_{i,d+j} = \mathrm{Corr}(\mathbf{x}_{1,i},\mathbf{x}_{2,j})$ and $\Sigma_{i,i} = 1,\ i,j \leq 2d$. If $\mathrm{Corr}(\mathbf{x}_{1,i},\mathbf{x}_{2,j}) = 1$, this is equivalent to generating the anithetic pair $(\mathbf{X}_1, 1-\mathbf{X}_1)$. We call such pairs *simple antithetic pairs* since there are no mutual correlations across dimensions.

Since uniform antithetic pairs (and hence simple pairs) do not alter the distribution of the individual multidimensional samples, they lead to unbiased estimators. However, the complex correlations between them affect the variances of the estimators. If the correlations are negative, and are preserved by evaluation of the integrand, then there is a variance reduction.

4.4. Antithetic importance sampling

Two possibilities: Consider a 1D canonical domain $[0,1]$. To generate $\mathbf{X} \in [0,1], \mathbf{X} \sim g(x)$, a typical approach in rendering is to use the transformation $\mathbf{X} = G^{-1}(\mathbf{U})$, where $\mathbf{U} \in [0,1]$ is uniformly distributed in the canonical domain, and G^{-1} is the inverse of the cdf of $g(x)$. There are two obvious ways to generate $\mathbf{Y}$ for an antithetic pair $(\mathbf{X},\mathbf{Y})$. Either $\mathbf{Y}_x = 1-\mathbf{X}$, which is a reflection of $\mathbf{X}$ about the center or $\mathbf{Y}_u = G^{-1}(1-\mathbf{U})$, which corresponds to reflecting $\mathbf{U}$ first. These two options result in generally different results, as explained below.

The estimators: From symmetry, $\mathbf{Y}_x \sim g(1-x)$ and the two-sample antithetic importance sampling estimator resulting from $(\mathbf{X},\mathbf{Y}_x)$ is $\tilde{\mu}_{\text{ais},2} = (f_g(\mathbf{X}) + f_g(1-\mathbf{X}))$ where we use the shorthand notation $f_g(x) = f(x)/g(x)$. Since both samples are divided by their respective probability densities, this estimator is unbiased. However, $\mathbf{Y}_u \sim h(x)$ where the pdf $h(x)$ obeys the cdf $H^{-1}(x) = G^{-1}(1-x)$. Thus, determining the correct weight for the antithetic sample to be combined without introducing a bias is tricky, except when $g(x)$ is constant (and hence G^{-1} is linear). Henceforth, in this paper, we use $(\mathbf{X}, 1-\mathbf{X})$ as the antithetic pair for importance sampling.

Convergence is unaffected: The antithetic sampling primary estimator uses two samples, and its variance is reduced due to negative correlation between the two samples. The corresponding secondary estimator's variance is lowered by a proportional factor regardless of the number of primary estimates used. Antithetic sampling potentially reduces the overall RMSE, but cannot improve convergence [Owe08]. It might be misleading to imagine that if a negative correlation was induced between every pair of N secondary estimates, that the convergence could be improved. However, this is not possible since the covariance matrix (between the variables) needs to remain positive definite.

4.5. Qualitative analysis of AS with jittered sampling:

In theory, under certain smoothness assumptions, the variance of the stratified sampling estimator (in 1D) is

$$\mathrm{V}(\tilde{\mu}_{\text{sN}}) \;=\; \frac{|\mathcal{D}|^4}{N^4}\sum_{i=1}^{N}(f'(\mathbf{X}_i))^2 \;\approx\; \frac{|\mathcal{D}|^3}{N^3}\int_{\mathcal{D}}(f'(x))^2\;\mathrm{d}x$$

which means that RMSE convergence for stratified sampling is $O(N^{-3/2})$ in 1D domains, provided the integrand is somewhat linear within each stratum and square integrable without sharp variation. Although Owen et al. [Owe13] derive diminishing returns for higher dimensions, this convergence rate has been observed to hold in image synthesis applications for 2D antialiasing [Mit96] and 3D visibility integration [RAMN12]. This motivates us to explore its use for estimating direct illumination.

Jittered antithetic sampling: Antithetic sampling provides a zero-variance estimator for linear integrands. Combining stratification with antithetic sampling yields an estimator whose only source of variance is in regions where the integrand varies sharply (high gradient). However, as noted previously, it is possible that antithetic sampling increases variance. It is here that the condition derived in Appendix A proves useful. If the condition is not satisfied, then the antithetic sample should be drawn as an uncorrelated sample. This strategy ensures that, in the worst case, stratified antithetic sampling degenerates to standard stratified sampling.

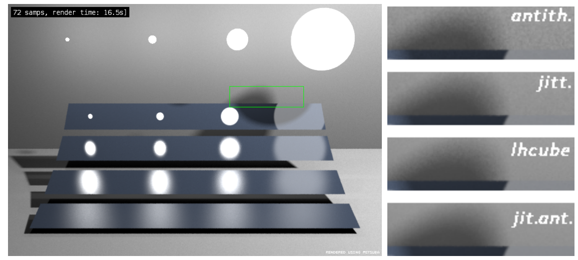

Figure 4: An inset from Veach's scene is shown rendered with four different strategies: antithetic, jittered antithetic, latin hypercube and jittered antithetic sampling. Combining stratification and antithetic sampling reduces error (visible noise) and improves convergence (see Fig. 3).

Jittered antithetic importance sampling: Stratified antithetic sampling provides a benefit when the integrand $f(x)$ is piece-wise (within the strata) linear or skew-symmetric. The inclusion of an importance function to generate the samples requires instead that $f_g(x) = f(x)/g(x)$ is piece-wise linear or skew-symmetric. Since importance functions are typically designed to warp the integrand to a constant, it can be expected that stratified antithetic importance sampling will have low error, in general. Again, if the condition (derived in Appendix B) is not satisfied, the correlation between the antithetic samples must be removed so that it degenerates to stratified sampling.

Variance reduction and convergence: The M samples in the jittered sampling estimator are subtly correlated. As M increases, the sizes of each of the strata reduce, thereby reducing the variance within each stratum. In addition, the integrand becomes increasingly close to linear or even constant within each stratum. These three combined effects result in improved convergence. When importance sampling or MIS is combined with stratification, the error is reduced in strata that are piece-wise linear or skew-symmetric. We observe empirically and in examples of image synthesis that the fraction of strata that this occurs is significant enough that there is an improvement in convergence. *This is an exciting new observation, that standard techniques which only improve error can be combined with stratification to further improve the convergence rate of stratified sampling.*

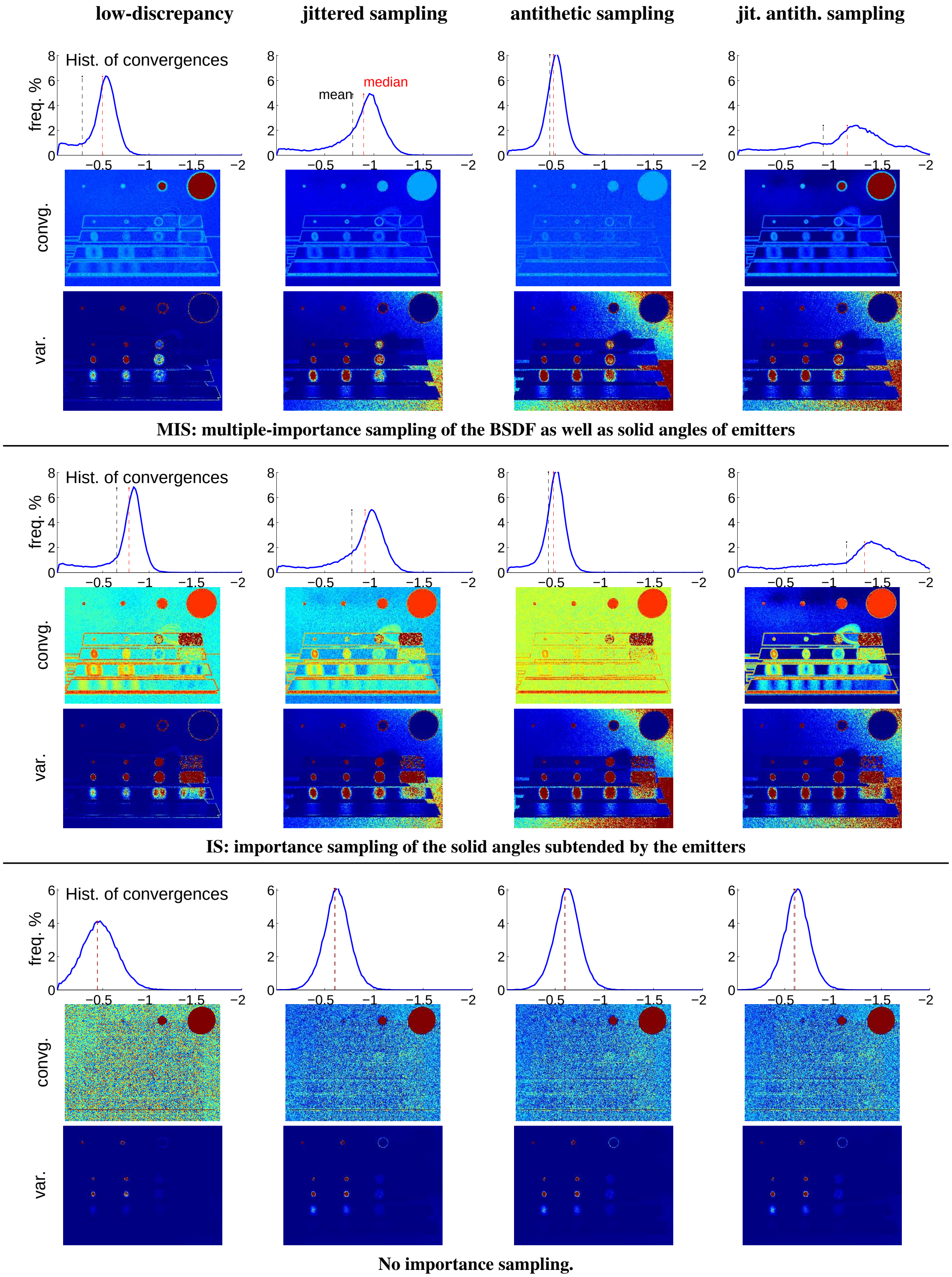

Figure 3: Antithetic sampling consistently improves convergence rates in the presence of any form of importance sampling, all at negligible cost. Measured convergence rates and errors for direct illumination in Veach's scene with MIS (top set of three rows), IS of the emitters (middle set) and without importance sampling (bottom set) are shown. The convergence rates are best with MIS, and approach naïve Monte Carlo when no importance sampling is used (bottom set).

5. Results

We verified our theoretical results and qualitative analyses in estimators for direct illumination, including visibility. For each estimator, we obtained the per-pixel variance by rendering hundreds of images. We computed the convergence rates, again per-pixel, by measuring the slopes of the log-log plots of variances against sample counts (spp).

Experiments: Fig. 1 visualises the convergence plots at 8 different pixels (a-g), for four different estimators. The ranking of estimators, based on their errors, depends on the number of samples used. e.g. Although the black line in 1c has a lower error when fewer than 37 samples (per pixel) are used, the red and blue lines have lower errors for larger sampling counts. At pixel c, JAIS is a better choice interactive or real-time rendering (where the sample budget is low) while jittered sampling or low-discrepancy sequences are better for high-quality rendering (with higher sampling budgets).

Fig. 3 visualises these convergence rates as well as the variances for Veach's reflectors scene, as a heat map. Four estimators are compared. Two of them are state-of-the-art (low-discrepancy and jittered sampling) while the other two — antithetic and jittered antithetic — are new to rendering. Blue indicates a low (rapid convergence or low error) value and red indicates a high value (slow convergence or high error). We repeated the experiment for three scenarios: With MIS, with IS of the emitters, and without importance sampling. Since each method performs best for a characteristically chosen number of samples (e.g. perfect square for jittered sampling), we measured the convergence rates using carefully chosen numbers of samples per pixel and random samples for the remainder to make sample counts match. Fig. 4 visualises the manifestation of error as unpleasant noise. Finally, Fig. 5 compares per-pixel convergence rates (slope of the convergence plot) and errors (Y-intercepts) of JAIS with jittered sampling. The renders show a Copper object with a Beckmann's distribution (alpha=0.1) and lit by an environment , using one-sample MIS with path tracing.

Interpretation of results: Fig. 3 clearly illustrates that JAIS is a winning strategy. When combined with standard importance sampling (whether of the subtended solid angle of emitters or an MIS of emitters with surface BRDF sampling), jittered antithetic sampling consistently improves the convergence rate. In fact, the results shows that over half the pixels have a convergence rate of better than $O(N^{-1.2})$, with some pixels reaching a convergence rate of $O(N^{-2})$ — significantly better than that achieved by low-discrepency sampling, jittered sampling, or antithetic sampling in isolation. Moreover, even without any importance sampling, including antithetic sampling does not decrease convergence rates. As such, we advocate the use of antithetic sampling in all such circumstances. In Fig. 5, we see that JAIS enjoys a better convergence (black pixels in second column) than jittered sampling in many pixels but does show slightly higher error in many pixels (white pixels in third column). That is, jittered

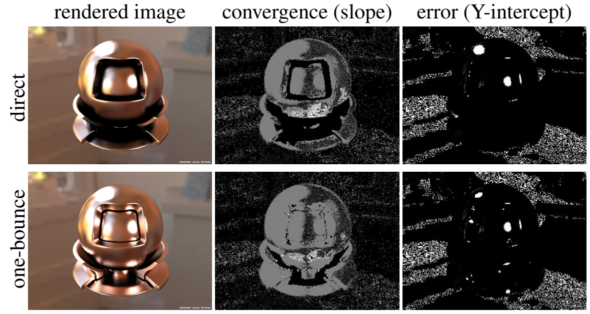

Figure 5: A comparison of JAIS with jittered sampling. For the same scene (first column), convergence (second column) and error (third column) are shown. The top row considers only direct illumination, while the bottom row includes one-bounce indirect illumination.

sampling is a better choice for low-sample counts, but when it is affordable to invest in more samples, the benefit of the improved convergence makes JAIS a better choice.

6. Discussion

Jittered antithetic IS: Our experiments (fig. 3) indicate that good importance functions are crucial for the improved convergence of jittered antithetic sampling. With judiciously chosen importance functions, our jittered antithetic importance sampling estimator converges more rapidly than existing strategies. In fig. 3 (top row) the histogram over pixels shows that our estimator improves the convergence of more than half the pixels, when compared to low-discrepancy sampling. Our analysis provides the fresh insight, that when used with antithetic sampling, it is sufficient that the importance function warp the integrand to a linear function, instead of the traditional goal of warping to a constant function.

Bias in antithetic IS, with arbitrary correlations: Perfect antithetic samples may be viewed as drawn from $g_2(x) = g(1-x)$. If the samples are generated with a different correlation (see supplemental material B for visualization), then the distribution of the antithetic sample is no longer $g_2(x)$. If the estimator does not account for this modified sampling distribution it will be biased. The correlation parameter allows control of this bias. At zero, the estimator is unbiased, but has additional variance, because it amounts to averaging the estimates due to importance sampling and naïve MC.

Sampling domain and distortion: We derive all our results for the d-dimensional canonical domain $[0,1]^d$. Rendering algorithms make different interpretations of the integration, e.g. for direct illumination, the sampling domains of interest are emitter surface area, local hemisphere, emitter's projected solid angle and the material's reflectance distribution. All of these are 2D domains, and apply different transformations to $[0,1]^2$. The distortions introduced by these transformations further influence the actual error and the convergence rates of the jittered antithetic importance sampling estimator. However such analysis needs to be targeted specially at each case. In this paper, we sample the canonical $[0,1]^2$ domain without including such parametrizations.

Comparing stochastic integrators for rendering: Even for a single pixel, the determination of the best choice of sampling strategy depends on the available budget for samples. For real-time and interactive ray-tracing purposes, this budget is potentially small, and lower error is more important than steeper convergence. For high-quality offline rendering, convergence is potentially more important to consider.

Product importance sampling: In rendering, the importance function is often a product [BGH05] of two functions, such as environment lighting and reflectance distributions. Our derivation for the variance of importance sampling encompasses this case. The arrangement of samples, from such product distributions, depends on the specific algorithm used. This makes it impossible to derive a general theoretical result for the variance of jittered sampling from product functions.

7. Conclusion and future avenues

We have derived the variance of combinations of unbiased estimators popularly used in rendering, in closed form. We also motivated the use of negatively correlated estimates for variance reduction. We found that certain arrangement strategies, such as jittered sampling, when combined with antithetic sampling can lead to improved convergence rates. We reported that antithetic sampling possibly leads to increased variance for certain integrands. To avoid this, we proposed a sampling strategy to control the degree of correlation between samples. We found that our new jittered antithetic importance sampling strategy leads to an improved convergence rate at many pixels while estimating direct illumination.

Further research is required to adapt our observations for high-dimensional integrals. JAIS provides little benefit for estimating global illumination in the modified (glossy) Cornell box scene in fig. 6 (top row). We compared convergence rates of the aforementioned four estimators while estimating a high-dimensional integration for global illumination (fig. 6). We plotted histograms of convergence rates over pixels using path tracing as well as photon-mapping with final gather. Path tracing with infinite bounces, for global illumination, corresponds to an integral of a highly non-linear function over an infinite-dimensional domain. In this case, the stratification for jittered sampling results in large strata with distorted footprints in path space. Performing antithetic or importance sampling in this domain is hopeless.

One possibility is to express the high-dimensional integral alternatively as a low-dimensional integration of a smooth function by using photon mapping with final gather. Doing so, improves the convergence rate of the JAIS estimator in fig. 6 (top row) to be comparable with jittered sampling and low-discrepancy sampling. This is one of the drawbacks of the naïve combination of stratification with antithetic sampling and importance sampling. We believe that this analysis will inspire sophisticated combinations of antithetic and importance sampling for estimation of global illumination.

Acknowledgements: Kartic Subr was funded through FI-Content, a European (EU-FI PPP) project. Thanks to Herminio Nieves, http://oigaitnas.deviantart.com, for the HN48 Flying Car model, and Blochi, http://www.hdrlabs.com/sibl/archive.html, for the Helipad Golden Hour environment map used in Fig. 1.

References

[ARBJ03] AGARWAL S., RAMAMOORTHI R., BELONGIE S., JENSEN H. W.: Structured Importance Sampling of Environment Maps. *ACM Transactions on Graphics 22*, 3 (2003), 605–612. 2

[BGH05] BURKE D., GHOSH A., HEIDRICH W.: Bidirectional importance sampling for direct illumination. In *Eurographics Conf. on Rendering Techniques* (2005), EGSR'05, pp. 147–156. 9

[Bis06] BISHOP C. M.: *Pattern Recognition and Machine Learning*. Springer, 2006. 2

[CJAMJ05] CLARBERG P., JAROSZ W., AKENINE-MÖLLER T., JENSEN H. W.: Wavelet Importance Sampling: Efficiently Evaluating Products of Complex Functions. *ACM Transactions on Graphics 24*, 3 (2005). 2

[Coo86] COOK R. L.: Stochastic Sampling in Computer Graphics. *ACM Transactions on Graphics 5*, 1 (Jan. 1986), 51–72. 2

[DP10] DICK J., PILLICHSHAMMER F.: *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, New York, NY, USA, 2010. 3

[GKH*13] GEORGIEV I., KŘIVÁNEK J., HACHISUKA T., NOWROUZEZAHRAI D., JAROSZ W.: Joint Importance Sampling of Low-order Volumetric Scattering. *ACM Transactions on Graphics 32*, 6 (2013), 164:1–164:14. 2

[Hab70] HABER, S.: Numerical Evaluation of Multiple Integrals. *SIAM Review 12* (1970), 481–526. 3

[Hes03] HESTERBERG T. C.: *Advances in Importance Sampling*. PhD thesis, Stanford University, 2003. 2

[HM56a] HAMMERSLEY J. M., MAULDON J. G.: General Principles of Antithetic Variates. *Proc. Cambridge Philos. Soc. 52* (1956), 476–481. 2

[HM56b] HAMMERSLEY J. M., MORTON K. W.: A New Monte Carlo Technique: Antithetic Variates. *Proc. Cambridge Philos. Soc. 52* (1956), 449–475. 2

[Jak10] JAKOB W.: Mitsuba Renderer, 2010. http://www.mitsuba-renderer.org. 2

[KC08] KŘIVÁNEK J., COLBERT M.: Real-time Shading with Filtered Importance Sampling. *Computer Graphics Forum (EGSR '08) 27*, 4 (2008), 1147–1154. 3

[Kel96] KELLER A.: Quasi-Monte Carlo methods in computer graphics: the global illumination problem. *Lectures in Applied Mathematics. 32* (1996), 455–470. 3

[KHN06] KELLER A., HEINRICH S., NIEDERREITER H.: *Monte Carlo and Quasi-Monte Carlo methods*. Springer, 2006. 3

[KPR12] KELLER A., PREMOZE S., RAAB M.: Advanced (Quasi) Monte Carlo Methods for Image Synthesis. In *ACM SIGGRAPH 2012 Courses* (2012), pp. 21:1–21:46. 1, 3

[Lem10] LEMIEUX C.: *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics. Springer, 2010. 3

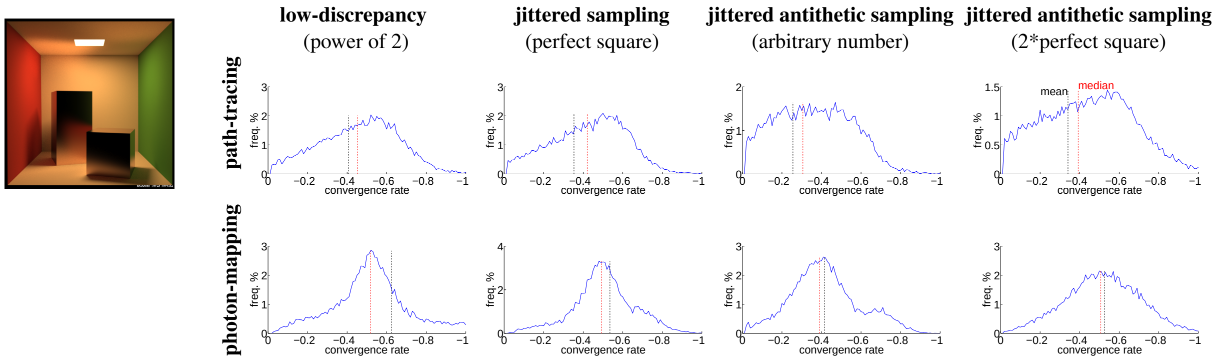

Figure 6: Limitation: Estimating global illumination in a modified (glossy) Cornell Box. *Top Row*: Path traced; simple antithetic assumptions are not satisfied in high-dimensional domains so the mean convergence is slightly lower than low-discrepancy sampling. *Bottom Row* Photon mapping with final gather; if the high-dimensional integral is interpreted as a nested integral then the global illumination component is viewed as a smooth low-dimensional function. In this case JAIS is comparable to state of the art but not obviously better. Further research is needed for reaping benefit with higher-dimensional integrals.

[LRR04] Lawrence J., Rusinkiewicz S., Ramamoorthi R.: Efficient BRDF Importance Sampling Using a Factored Representation. *ACM Transactions on Graphics* (2004), 496–505. 2

[Mit96] Mitchell D. P.: Consequences of Stratified Sampling in Graphics. In *SIGGRAPH* (1996), pp. 277–280. 1, 2, 3, 6

[Nel06] Nelsen R. B.: *An Introduction to Copulas*. Springer Series in Statistics. Springer, 2006. 3

[Ney34] Neyman J.: On the two different aspects of the representative method: the method of stratified sampling and of purposive selection. *J. of the Royal Stat. Society 97*, 4 (1934), 558–625. 3

[Nie92] Niederreiter H.: *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, 1992. 3

[Owe08] Owen A. B.: Local Antithetic Sampling with Scrambled Nets. *The Annals of Statistics 36*, 5 (2008), 2319–2343. 2, 3, 6

[Owe13] Owen A. B.: *Monte Carlo Theory, Methods and Examples*. 2013. To be published, see http://statweb.stanford.edu/ owen/mc/. 1, 2, 3, 6

[OZ00] Owen A., Zhou Y.: Safe and Effective Importance Sampling. *Journal of the American Statistical Association 95*, 449 (2000), 135–143. 2

[PH10] Pharr M., Humphreys G.: *Physically Based Rendering, Second Edition: From Theory To Implementation*, 2nd ed. Morgan Kaufmann Publishers Inc., 2010. 2

[RAMN12] Ramamoorthi R., Anderson J., Meyer M., Nowrouzezahrai D.: A Theory of Monte Carlo Visibility Sampling. *ACM Trans. Graph. 31*, 5 (2012), 121:1–121:16. 3, 6

[Shi91] Shirley P.: Discrepancy as a Quality Measure for Sampling Distributions. In *Eurographics* ('91), pp. 183–194. 2, 3

[SP09] Saliby E., Paul R. J.: A farewell to the use of antithetic variates in Monte Carlo simulation. *JORS 60*, 7 (2009), 1026–1035. 2

[SS83] Schweizer B., Sklar A.: *Probabilistic metric spaces*. North-Holland Series in Probability and Applied Mathematics. 1983, ch. 6. 2, 3

[Vea97] Veach E.: *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, 1997. 1, 2

[VG95] Veach E., Guibas L. J.: Optimally combining sampling techniques for monte carlo rendering. In *22Nd Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1995), SIGGRAPH '95, ACM, pp. 419–428. 3

Appendix A: Test for antithetic sampling

If $\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = \sigma_{XY}$, then using the Delta method, we can derive up to first order that $\mathrm{Cov}(f(\mathbf{X}), f(\mathbf{Y})) \approx_1 \left(f'(\mu)\right)^2 \sigma_{XY}$, where $\langle \mathbf{X} \rangle = \langle \mathbf{Y} \rangle = \mu$. When the function is linear, the negative correlation is preserved and its magnitude depends on the variance of the function. This is not very useful for non-linear functions, so we derive the second order approximation $\mathrm{Cov}(f(\mathbf{X}), f(\mathbf{Y})) \approx_2 f(\mu) f''(\mu)(\sigma_{XX} + \sigma_{YY})/2 + \left(f'(\mu)\right)^2 \sigma_{XY}$ from which we derive

$$\boxed{\sigma_{XY} < -\frac{f(\mu) f''(\mu)}{\left(f'(\mu)\right)^2}} \tag{2}$$

for the negative correlation between $\mathbf{X}$ and $\mathbf{Y}$ to remain negative after the function is evaluated at these locations. Such an analysis is possibly inaccurate for non-linear functions over large domains since the integrand cannot be approximated well using a low-order polynomial.

Appendix B: Antithetic importance sampling

Treating f_g as the integrand, the variance of the antithetic importance sampled estimator is easily derived as $\mathrm{V}\left(\tilde{\mu}_{\text{ais},2}\right) = \mathrm{V}\left(\tilde{\mu}_{\text{is},2}\right) + \mathrm{Cov}\left(f_g(\mathbf{X}), f_g(1-\mathbf{X})\right)/2$. Since this is similar to the case without importance sampling, we now substitute $f_g(x)$ in place of $f(x)$ in Eq. 2 to obtain

$$\boxed{\sigma_{XY} < -\frac{f(\mu)}{g(\mu)} \frac{f''(\mu) g(\mu) - f(\mu) g''(\mu)}{f'(\mu) g(\mu) - f(\mu) g'(\mu)} + 2 g(\mu) g'(\mu)} \tag{3}$$

as the condition for antithetic importance sampling to result in a variance reduction. The above rules of thumb (Eq. 2 and Eq. 3) may be used to control the correlation between the random variables to obtain a variance reduction while using antithetic variates. If the tests report an increase in variance, then the correlation between the samples may be set to zero, to default to the naïve MC estimator.