MIS Compensation: Optimizing Sampling Techniques in Multiple Importance Sampling

ONDŘEJ KARLÍK, Chaos Czech a. s.
MARTIN ŠIK, Chaos Czech a. s.
PETR VÉVODA, Charles University, Prague and Chaos Czech a. s.
TOMÁŠ SKŘIVAN, IST Austria
JAROSLAV KŘIVÁNEK, Charles University, Prague and Chaos Czech a. s.

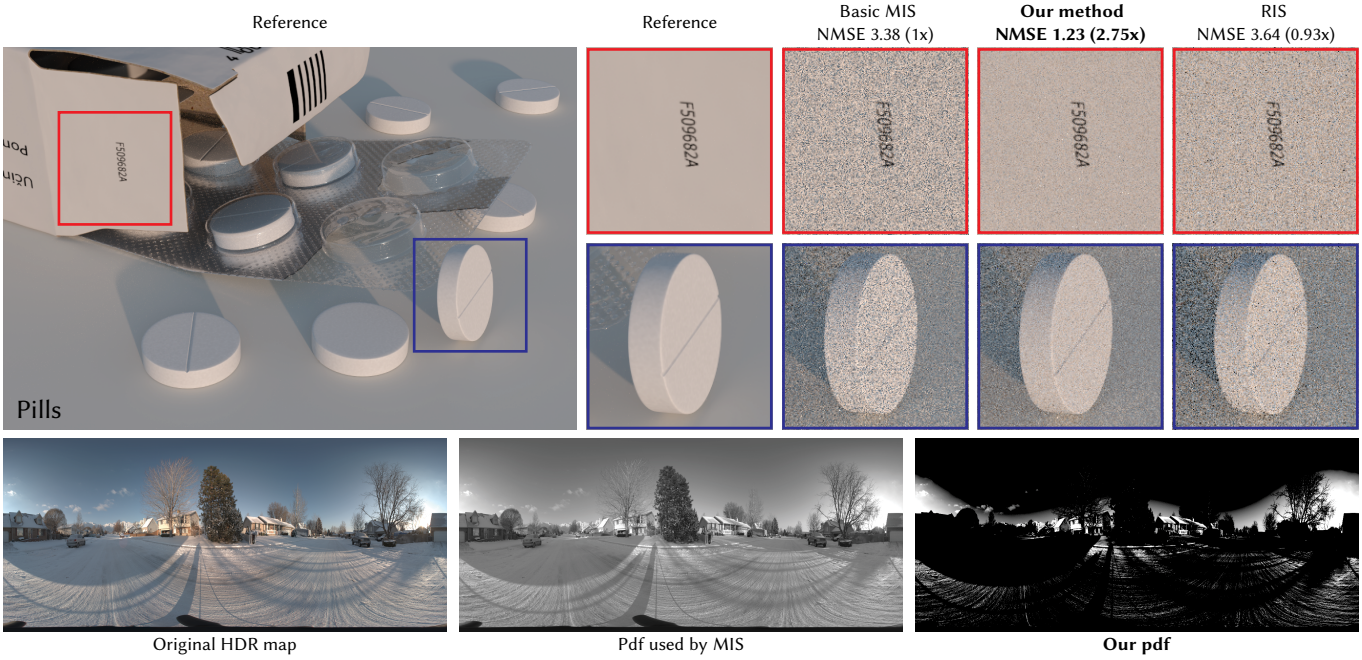


Fig. 1. Equal-time (5 s) comparison of basic multiple importance sampling with the balance heuristic (Basic MIS), resampled importance sampling (RIS) and our (normal-independent) method applied to image-based lighting computation. While RIS performs similarly to Basic MIS here, our method achieves 2.75× lower normalized mean square error (NMSE) by redefining the pdf of one of the sampling techniques, while taking into account that MIS is being applied. The pdfs are shown in the bottom row. The pdf optimization presented in this paper is general and can be applied in any MIS estimator.

Multiple importance sampling (MIS) has become an indispensable tool in Monte Carlo rendering, widely accepted as a near-optimal solution for combining different sampling techniques. But an MIS combination, using the common balance or power heuristics, often results in an overly defensive estimator, leading to high variance. We show that by generalizing the MIS framework, variance can be substantially reduced. Specifically, we optimize one of the combined sampling techniques so as to decrease the overall variance of the resulting MIS estimator. We apply the approach to the computation of direct illumination due to an HDR environment map and to the computation of global illumination using a path guiding algorithm. The implementation can be as simple as subtracting a constant value from the tabulated sampling density done entirely in a preprocessing step. This produces a consistent noise reduction in all our tests with no negative influence on run time, no artifacts or bias, and no failure cases.

CCS Concepts: • **Computing methodologies** → **Rendering**; **Ray tracing**.

Additional Key Words and Phrases: multiple importance sampling, combined estimators, light transport, physically-based rendering

ACM Reference Format:
Ondřej Karlík, Martin Šik, Petr Vévoda, Tomáš Skřivan, and Jaroslav Křivánek. 2019. MIS Compensation: Optimizing Sampling Techniques in Multiple Importance Sampling. *ACM Trans. Graph.* 38, 6, Article 151 (November 2019), 12 pages. https://doi.org/10.1145/3355089.3356565

Authors' addresses: Ondřej Karlík, Chaos Czech a. s. Prague; Martin Šik, Chaos Czech a. s. Prague; Petr Vévoda, Charles University, Prague, Chaos Czech a. s. Prague; Tomáš Skřivan, IST Austria, Vienna; Jaroslav Křivánek, Charles University, Prague, Chaos Czech a. s. Prague.

0730-0301/2019/11-ART151 $15.00
https://doi.org/10.1145/3355089.3356565

1 INTRODUCTION

Multiple importance sampling (MIS) has become an essential tool in Monte Carlo rendering, since it provides a simple yet robust means for combining sampling techniques [Veach and Guibas 1995]. Not only has MIS enabled the development of advanced bidirectional rendering algorithms [Georgiev et al. 2012b; Hachisuka et al. 2012; Křivánek et al. 2014], it is also an indispensable tool for robust combination of the fundamental building blocks of any production path tracer, e.g. direct illumination sampling from area lights and environment maps [Pharr et al. 2016], subsurface scattering [King et al. 2013], free-flight sampling in participating media [Kulla and Fajardo 2012], etc. As such, MIS has been one of the enabling technologies behind the "path tracing revolution" in the computer animation and VFX industries [Keller et al. 2015; Křivánek et al. 2018]. While MIS had originally been introduced in computer graphics, it has made an impact in other fields as well [Cornuet et al. 2012; He and Owen 2014; Owen and Zhou 2000].

The original MIS formulation [Veach and Guibas 1995] bases its near-optimality arguments on the assumption that the sampling densities and the sample count for the combined sampling techniques are given upfront and fixed. The degree of freedom left for optimization are the weighting functions used to combine samples from the individual techniques. The balance heuristic weighting has become a popular choice for its simplicity and tight optimality bounds when only non-negative weighting functions are considered [Veach and Guibas 1995]. Truly optimal MIS weighting functions were proposed recently [Kondapaneni et al. 2019]. Other works investigate the sample allocation among the sampling techniques [He and Owen 2014; Sbert and Havran 2017; Vorba et al. 2019].

To the best of our knowledge, no previous work has addressed designing the sampling densities themselves in the context of MIS. We show that such a modification of the MIS framework provides a previously unexploited potential for variance reduction. We consider an MIS combination of several sampling techniques, but unlike in previous work, we allow the probability density function (pdf) of one of the techniques to take any shape. Subsequently, we find the pdf of that 'free' technique that decreases variance of the resulting MIS estimator. This optimization can be done upfront, before the sampling starts, and does not rely on any adaptive updates [Cappé et al. 2008]. In contrast to previous research, such as various product-sampling methods [Herholz et al. 2016; Rousselle et al. 2008; Talbot et al. 2005], we do *not* devise a specialized sampling technique that would make the free pdf match the integrand closer. Our method can actually do the opposite, but in such a way that the overall variance of the entire MIS estimator is still decreased. We call our approach *MIS compensation*, because its net result is, so to say, sharpening of the 'free' technique's pdf, that effectively compensates for the pdf averaging induced by the balance heuristic.

While our general result has various potential applications, we verify its usefulness on two problems. First, the image-based lighting, i.e. computation of direct illumination due to a high-dynamic-range (HDR) environment map on a surface with an arbitrary bidirectional reflectance distribution function (BRDF). The traditional approach would combine sampling techniques with pdfs proportional to the individual factors, i.e. the HDR map and the BRDF. We show that this approach is overly defensive, and use our *MIS compensation* to redefine the HDR map sampling density in a way that reduces overall variance. The improved sampling technique is defined by a simple analytic formula used to modify the tabulated pdf in a preprocessing step. Its simplest form can be used in any renderer relying on MIS for HDR map sampling without any modifications of the sampling routines themselves. Our method has been deployed in a production renderer, yielding a consistent performance improvement.

Second, we apply our method to *path guiding* [Vorba et al. 2019] in the path tracing algorithm. Path guiding combines direction sampling proportional to the BRDF with sampling according to a learned guiding density. By optimizing the guiding density we improve the performance of the algorithm by Müller et al. [2017].

2 RELATED WORK

Multiple importance sampling (MIS) [Veach and Guibas 1995] generalizes the idea of sampling from a mixture of distributions [Hesterberg 1995; Torrie and Valleau 1977]. The MIS framework represents a family of estimators, parametrized by weighting functions used to combine samples from different sampling techniques. One example is the balance heuristic; Veach and Guibas show that no estimator in the MIS family (with non-negative weighting functions [Kondapaneni et al. 2019]) can have significantly lower variance. Despite this result, they develop alternative weighting functions, such as the power or cutoff heuristics, that perform better when one of the sampling techniques is a good match to the integrand. However, performance of any weighting functions, including the optimal ones [Kondapaneni et al. 2019], is limited by the fixed sampling densities. We lift this limitation and instead of modifying the weighting functions, we optimize directly one of the sampling densities.

Numerous other works have investigated possible improvements of MIS. Owen and Zhou [2000] show that using a mixture of the sampling pdfs as a control variate allows reducing the variance bound in MIS. Fan et al. [2006] applied this work in rendering and Kondapaneni et al. [2019] linked it to optimal weighting functions for the MIS family. Elvira et al. [2015; 2016] analyze random vs. deterministic assignment of samples among sampling techniques and propose a clustering of techniques to reduce computational overhead.

Sample allocation in MIS. A widely studied direction is sample allocation among the sampling techniques. Pajot et al. [2011] propose optimized sample allocation based on a heuristically defined *representativity*. Subsequent works cast optimal sample allocation as a minimization of variance or of a related functional. A closed form solution is yet to be found, so a variety of work-around ideas have been proposed instead. Lu et al. [2013] approximate the variance using a truncated Taylor expansion, yielding sample allocation in a closed form. He and Owen [2014] prove convexity of variance in the number of samples, and employ convex optimization to determine optimal sample allocation. Havran and Sbert [2014] and Sbert et al. [2016] show that the optimal sample allocation must equalize the second moment of the weighted estimates corresponding to the individual sampling techniques. Sbert and Havran [2017] follow up with an approximate solution based on variance estimates and Sbert et al. [2018] introduce new balance heuristic variants better than the balance heuristic with equally allocated samples. Cappé et al. [2008]

and Lai et al. [2015] optimize mixture densities in the population Monte Carlo framework.

The above methods all require pilot samples or an adaptive sampling scheme, which could involve a significant intervention into a rendering system, limit its interactivity, and be susceptible to sampling errors. In contrast, our method optimizes the sampling density without requiring any MC samples to be taken up front.

Image-based lighting. MIS combination of BRDF and HDR environment map sampling is a standard approach to image-based lighting in Monte Carlo renderers [Pharr et al. 2016]. Whereas the sampling pdf for the HDR map is usually proportional to its luminance, we show how to modify the pdf so as to reduce variance.

Other ideas on image-based lighting include stratification of the environment map [Agarwal et al. 2003] and sampling from the *product* of the BRDF and the environment map. Resampled importance sampling approximates product sampling at the cost of extra samples [Burke et al. 2005; Talbot et al. 2005]. Hierarchical methods build the product distribution on the fly [Clarberg and Akenine-Möller 2008b; Clarberg et al. 2005; Jarosz et al. 2009; Rousselle et al. 2008]. While the product sampling methods redefine the sampling procedure itself and often come with a substantial computational and memory overhead, our *MIS compensation* does not require any modification to a standard MIS-based environment map sampler.

Path guiding. Learning incident illumination in a scene for importance sampling has become known as *path guiding* [Vorba et al. 2019]. These methods learn the guiding density either through regression modelling [Lafortune and Willems 1995; Müller et al. 2017] or density estimation from particles obtained in a preprocessing step [Herholz et al. 2016, 2019; Jensen 1995; Vorba et al. 2014]. In our second application we use MIS compensation to optimize the guiding density in the Müller et al.'s [2017] method.

3 REVIEW OF MULTIPLE IMPORTANCE SAMPLING

Consider a nonzero integral $F = \int_\mathcal{X} f(x)\,\mathrm{d}x$ of a non-negative function $f(x)$. An unbiased Monte Carlo estimator $\langle F\rangle$ for the integral can be constructed by taking n random variables (samples) X_i, $i = 1, \ldots, n$ generated by a *sampling technique* with the probability density function (pdf) $p(x)$, and setting $\langle F\rangle = \frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i)}{p(X_i)}$. Variance of the estimator depends on the number of samples and on how 'similar' the pdf $p(x)$ is to the integrand $f(x)$. Whenever p is exactly proportional to f up to a normalization constant, i.e. $p(x) = f(x)/F$, the estimator has zero variance [Veach and Guibas 1995].

Finding a single pdf that closely approximates f under all circumstances may be difficult, but we may have multiple sampling techniques $t \in \mathcal{T}$, each with a pdf $p_t(x)$ approximating some important feature of f (e.g. BRDF sampling and HDR map sampling). Multiple importance sampling (MIS) provides a general framework for constructing a combined estimator given multiple sampling techniques. Suppose we take $n_t = c_t n$ independent samples $X_{t,i}$ from each technique, where c_t denote the respective fractions of the total sample count. The *multi-sample combined MIS estimator (ms)* then reads

$$\langle F\rangle_{\mathrm{ms}} = \sum_{t\in\mathcal{T}} \frac{1}{n_t}\sum_{i=1}^{n_t} w_t(X_{t,i})\frac{f(X_{t,i})}{p_t(X_{t,i})}. \tag{1}$$

In the *one-sample (os)* variant, a single sample X_t is drawn from the technique $p_t(x)$ chosen with probability c_t:

$$\langle F\rangle_{\mathrm{os}} = w_t(X_t)\frac{f(X_t)}{c_t p_t(X_t)}. \tag{2}$$

The weighting functions $w_t(x)$ provide a degree of freedom that can be used to optimize the estimator's variance. The *balance heuristic*,

$$\hat{w}_t(x) = \frac{c_t p_t(x)}{\sum_{t'\in\mathcal{T}} c_{t'} p_{t'}(x)}, \tag{3}$$

is the optimal choice for $w_t(x)$ in the one-sample variant and a provably good choice in the multi-sample variant in the sense that no other set of *non-negative* weighting functions can yield a much lower variance [Veach and Guibas 1995]. Negative weights can further decrease variance beyond the guarantee for the balance heuristic given by Veach and Guibas. The optimal weighting functions for the multi-sample variant, minimizing its variance, are in fact often negative [Kondapaneni et al. 2019]. However, performance of any weighting functions, including the optimal ones, is *limited by the fixed set of sampling densities*. We lift this restriction and open up a new, previously unexplored opportunity for reducing variance.

Plugging the balance heuristic (3) into the general MIS estimators (1) and (2) yields the combined estimators in the form

$$\langle F\rangle_{\mathrm{ms}} = \frac{1}{n}\sum_{t\in\mathcal{T}}\sum_{i=1}^{n_t}\frac{f(X_{t,j})}{p_{\mathrm{eff}}(X_{t,i})}, \quad \langle F\rangle_{\mathrm{os}} = \frac{f(X_t)}{p_{\mathrm{eff}}(X_t)}, \tag{4}$$

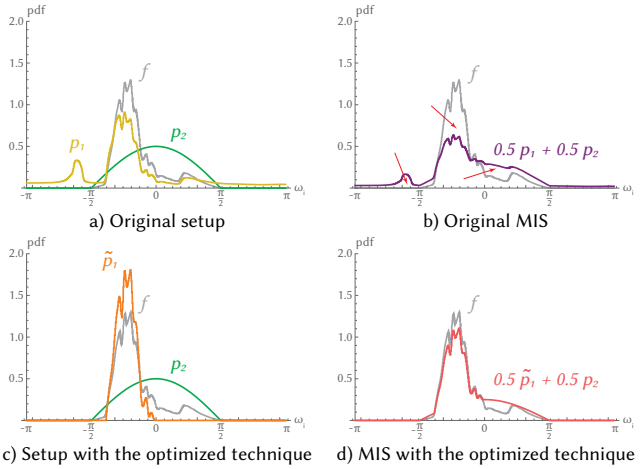
where $p_{\mathrm{eff}}(x) = \sum_{t'\in\mathcal{T}} c_{t'} p_{t'}(x)$. This shows that the use of the balance heuristic effectively corresponds to a regular MC estimator with samples drawn from a *mixture density* $p_{\mathrm{eff}}(x)$. Note that this effective mixture density is known only for the balance heuristic and cannot be readily expressed for other weighting functions.

4 MIS COMPENSATION

While the balance heuristic is near-optimal within the MIS family, the MIS combination of a given, fixed set of sampling techniques may not yield a particularly good estimator. As shown in Fig. 2 the effective mixture density may be too defensive, which causes the high-value areas of the integrand to be undersampled and the low-value areas to be oversampled. We significantly improve the result by directly optimizing one of the sampling techniques. Its pdf is 'sharpened', i.e. has pronounced peaks and deep valleys, so as to compensate for the effect of pdf averaging induced by the balance heuristic. Hence the method name, *MIS compensation*.

Problem statement. Consider an MIS combination of a given set $\mathcal{T}$ of sampling techniques $t \in \mathcal{T}$ with sample fractions $c_t > 0$. One of the sampling techniques, τ, is designated as *free*. The objective is to find the free pdf $p_\tau(x)$ that will minimize variance of the combined estimator with the balance heuristic, Eq. (4).

MIS-compensated solution. In an ideal case, both the multi- and one-sample MIS estimators from Eq. (4) should have zero variance. A necessary condition for this to happen is that $p_{\mathrm{eff}}(x)$ is exactly proportional to the integrand $f(x)$, i.e., $p_{\mathrm{eff}}(x) = f(x)/F$. Assuming the balance heuristic, it holds that $p_{\mathrm{eff}}(x) = q(x) + c_\tau p_\tau(x)$, where we have separated pdfs of all the fixed techniques under a single



a) Original setup b) Original MIS

c) Setup with the optimized technique d) MIS with the optimized technique

Fig. 2. An example motivating our work. a) We would like to integrate function f (gray). We cannot sample it directly but we are able to sample according to the two factors forming f: the techniques p_1 (yellow) and p_2 (green). b) Since it is difficult to predict which of the techniques p_1, p_2 is better, we take equal number of samples from both of them and combine them using the balance heuristic. We can see that the effective mixture density (purple) is too defensive and causes the high-value areas to be undersampled and the low-value areas to be oversampled (see the red arrows). c) We redefine technique p_1, taking into account the fact that some parts of the integrand are already well-enough sampled by technique p_2, yielding a new technique $\tilde{p}_1$ (orange). d) Combining the optimized $\tilde{p}_1$ with p_2 yields an effective mixture density (red) that is much closer to the integrand.

term $q(x) = \sum_{t \in \mathcal{T} \setminus \{\tau\}} c_t p_t(x)$. Simple algebra then yields

$$p_\tau(x) = \frac{f(x)}{c_\tau F} - \frac{q(x)}{c_\tau}. \tag{5}$$

This definition, however, does not ensure that $p_\tau(x)$ is a pdf, i.e. that it integrates to one and $p_\tau(x) \geq 0$. To achieve a valid pdf, we clamp negative values and re-normalize to obtain the final result

$$\boxed{\tilde{p}_\tau(x) = 1/b \ \max\{0, p_\tau(x)\},} \tag{6}$$

where the normalization factor is $b = \int_X \max\{0, p_\tau(x)\} \, \mathrm{d}x$.

Discussion. MIS compensation can be seen as minimizing the overlap between the free pdf and the remaining fixed pdfs, because sampling a part of the integrand that is already well sampled by some other technique in the MIS combination is a waste of effort. The best choice of the free pdf is therefore the ideal pdf minus all the fixed pdfs, i.e. Eq. (5).

Although the *MIS-compensated* solution requires knowledge of the target integral F, this does limit its use in practice. A suitable approximation is the key, and we present several practical solutions in Sections 6 and 8.

The *MIS-compensated* pdf maintains unbiasedness: The MIS estimators from Eq. (4) are unbiased if $p_{\text{eff}}(x) > 0$ whenever $f(x) > 0$. For every x such that $f(x) > 0$ the *MIS-compensated* pdf $\tilde{p}_\tau(x)$ can be zero only if $q(x)$ is nonzero. Conversely, should $q(x)$ be zero, $\tilde{p}_\tau(x)$ must be nonzero, as the subtraction in Eq. (5) vanishes.

The *MIS-compensated* pdf as derived thus far provides no guarantee of the resulting $\tilde{p}_\tau(x)$ being a variance minimizer, because of the ex post use of the max operator and the re-normalization. In the following section we rigorously derive a truly optimal free pdf that provably minimizes variance of the one-sample estimator under the problem statement above. Furthermore, we show that the *MIS-compensated* pdf $\tilde{p}_\tau$ is often close to the optimal result.

Since our practical applications are based solely on the *MIS-compensated* solution, readers may prefer to skip the theoretical discussion in the following section and continue directly to Sec. 6.

5 OPTIMALITY DISCUSSION

To assess the quality of our *MIS-compensated* solution, we first derive the *optimal* solution to the minimization problem stated in Sec. 4. We then discuss the relation of the two solutions.

5.1 Optimal solution

To find the pdf $p^*_\tau(x)$ of the free technique $\tau \in \mathcal{T}$ that provably minimizes variance of the one-sample estimator (Eq. (4)), we write the variance as $J[p] - F^2$, where the functional $J[p]$ is the estimator's second moment. Since the integral F does not depend on the pdf, all we need is to minimize $J[p]$ with respect to p:

$$p^*_\tau(x) = \arg\min_p J[p], \qquad J[p] = \int_X \frac{f^2(x)}{q(x) + c_\tau p(x)} \, \mathrm{d}x. \tag{7}$$

To avoid the ad hoc use of the max operator and normalization *after* the minimization, we subject the minimization to the constraints

$$p^*_\tau(x) \geq 0, \qquad \text{and} \qquad \int_X p^*_\tau(x) \, \mathrm{d}x = 1. \tag{8}$$

In Appendix A we use the Karush-Kuhn-Tucker (KKT) conditions to derive the *optimal* pdf $p^*_\tau(x)$ in the form

$$\boxed{\begin{aligned} p^\pm_\tau(x) &= \frac{f(x)}{\sqrt{c_\tau \lambda}} - \frac{q(x)}{c_\tau} \\ p^*_\tau(x) &= \max\{0, p^\pm_\tau(x)\} \end{aligned}} \tag{9}$$

where the Lagrange multiplier λ ensures normalization.

Due to the presence of the max operator in Eq. (9), there is no analytical formula for computing λ, so one has to resort to an iterative method to find just the right value of λ that ensures a normalized pdf $p^*_\tau(x)$. This makes the *optimal* pdf rather impractical. But should the *MIS-compensated* solution $\tilde{p}_\tau(x)$ be close-enough to the *optimal* one, it could be used instead. We show that this is often indeed the case, and sometimes the two solutions are even equal.

5.2 Bounding the MIS-compensated solution

To demonstrate the similarity between the *optimal* pdf $p^*_\tau(x)$ and the *MIS-compensated* pdf $\tilde{p}_\tau(x)$, we first derive the following bounds of the Lagrange multiplier λ

$$c_\tau F^2 \leq \lambda \leq (1/c_\tau) F^2. \tag{10}$$

Derivation of this result in Appendix B also shows that if $p^\pm_\tau(x)$ (Eq. (9)) is non-negative, λ is in fact equal to its lower bound $c_\tau F^2$. In that case $\tilde{p}_\tau(x) = p^*_\tau(x)$. This suggests that the *MIS-compensated* pdf is often not far from the *optimal* one. In the next section we verify this result on an example.

For the cases where $\tilde{p}_\tau(x)$ differs from $p^*_\tau(x)$ one may be interested in how the resulting MIS estimator variance increases if we use $\tilde{p}_\tau(x)$. We prove in Appendix C that the estimator's second moment can increase at most by a factor $1/c_\tau$ over using the *optimal* pdf. Since this bound might not be tight enough, we have searched for the actual worst case that would yield the highest increase of the estimator's variance when using our *MIS-compensated* pdf instead of the *optimal* one. In our experiments the variance has never increased more than 1.6 times (see the supplemental material for details).

To summarize, we have shown that while the *MIS-compensated* pdf $\tilde{p}_\tau(x)$ may differ from the *optimal* pdf $p^*_\tau(x)$, they are often similar. Since computing the *optimal* pdf is expensive (requires repeated integration of $p^*_\tau(x)$ for different λ values until the integral is one), we recommend using the *MIS-compensated* pdf instead (requires just one integration, the target integral F, which can be often efficiently approximated). We follow this recommendation in our applications discussed in the next sections.

6 APPLICATION I: IMAGE-BASED LIGHTING

Here we apply the *MIS-compensated* solution derived in Sec. 4 to the image-based lighting (IBL) problem. We compute unoccluded[1] direct illumination from a high-dynamic-range (HDR) map on a surface with an arbitrary bidirectional reflectance distribution function (BRDF), see Fig. 3. This translates to the following integral

$$L_{\text{dir}}(\mathbf{x}, \omega_\text{o}) = \int_{\mathcal{H}(\mathbf{n})} L_\text{I}(\omega_\text{i})\, \rho(\mathbf{x}, \omega_\text{o}, \omega_\text{i}) |\omega_\text{i} \cdot \mathbf{n}|_+ \,\text{d}\omega_\text{i}, \qquad (11)$$

where $\mathbf{x}$ is the position on the surface, $\mathbf{n}$ is the surface normal at $\mathbf{x}$, ω_o is the outgoing (view) direction, and $\mathcal{H}(\mathbf{n})$ is the hemisphere centered on the normal $\mathbf{n}$. The HDR map emission coming from direction ω_i is given by $L_\text{I}(\omega_\text{i})$, and ρ denotes the surface BRDF. To simplify the notation we have defined $|\omega_\text{i} \cdot \mathbf{n}|_+ = \max\{0, \omega_\text{i} \cdot \mathbf{n}\}$ (i.e. the positive part of the cosine of the angle between ω_i and $\mathbf{n}$).

Monte Carlo estimation of $L_{\text{dir}}(\mathbf{x}, \omega_\text{o})$ usually relies on two techniques: sampling of the HDR map and sampling of the product of the BRDF and $|\omega_\text{i} \cdot \mathbf{n}|_+$. These techniques are then combined using MIS with either the multi- or one-sample estimator, Eq. (4).

To reduce the estimator's variance, we optimize one of the sampling techniques as discussed in the previous sections. Sampling from the HDR map is usually implemented using a tabulated pdf $p_\text{I}(\omega_\text{i})$, while sampling of the BRDF-cosine product is an analytical formula derived from a pdf $p_\rho(\omega_\text{i}|\omega_\text{o}, \mathbf{x})$ (which generally depends on the outgoing direction and surface position). Since modifying a tabulated pdf is simple in practice, we choose to optimize $p_\text{I}(\omega_\text{i})$.

Note that optimizing both of the pdfs at once would yield the obvious zero-variance pdf with no practical recipe how to use it. With more sampling techniques it could be possible to optimize more than one, but we are not aware of a practical setting where this would be useful. Therefore, we always optimize just one pdf and we choose the one that is the simplest to modify.

6.1 General MIS-compensated solution

We get the *MIS-compensated* pdf $\tilde{p}_\text{I}$ for the IBL problem by plugging the integrand from Eq. (11) and the BRDF sampling technique $p_\rho(\omega_\text{i}|\omega_\text{o}, \mathbf{x})$ into Eq. (6). This yields

$$\tilde{p}_\text{I}(\omega_\text{i}|\omega_\text{o}, \mathbf{x}) = \frac{1}{b} \max\left\{0, \frac{f_\text{I}(\mathbf{x}, \omega_\text{o}, \omega_\text{i})}{c_\text{I} L_{\text{dir}}(\mathbf{x}, \omega_\text{o})} - \frac{1 - c_\text{I}}{c_\text{I}} p_\rho(\omega_\text{i}|\omega_\text{o}, \mathbf{x})\right\}, \qquad (12)$$

where we have simplified the notation by defining $f_\text{I}(\mathbf{x}, \omega_\text{o}, \omega_\text{i}) = L_\text{I}(\omega_\text{i})\, \rho(\mathbf{x}, \omega_\text{o}, \omega_\text{i}) |\omega_\text{i} \cdot \mathbf{n}|_+$. The normalization constant b then ensures that $\tilde{p}_\text{I}(\omega_\text{i}|\omega_\text{o}, \mathbf{x})$ integrates to one.

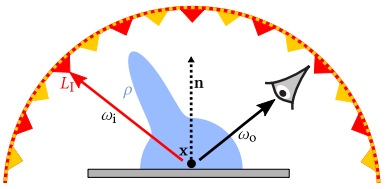


Fig. 3. Image-based lighting concerns itself with the computation of reflected radiance $L_{\text{dir}}(\mathbf{x}, \omega_\text{o})$ at a point $\mathbf{x}$ on a surface with the BRDF ρ given an outgoing (view) direction ω_o. Illumination given by the HDR map $L_\text{I}(\omega_\text{i})$ is modulated by the BRDF ρ and the cosine of the angle between the incoming direction ω_i and the surface normal $\mathbf{n}$.

6.2 Practical solution

The solution in Eq. (12) is difficult to use in practice, because the pdf $\tilde{p}_\text{I}(\omega_\text{i}|\omega_\text{o}, \mathbf{x})$ depends on both ω_o and $\mathbf{x}$, and we would need to tabulate the pdf for each view direction and surface position. We now present a practical, albeit approximate, solution.

Normal-dependent pdf (nd). First, we assume a Lambertian BRDF with unit albedo, $\rho \equiv 1/\pi$, yielding

$$p^{\text{nd}}_\text{I}(\omega_\text{i}|\mathbf{n}) = \frac{1}{b_{\text{nd}}} \max\left\{0, \frac{f_{\text{nd}}(\omega_\text{i}, \mathbf{n})}{c_\text{I} \int_{\mathcal{H}(\mathbf{n})} f_{\text{nd}}(\omega, \mathbf{n})\,\text{d}\omega} - \frac{1 - c_\text{I}}{c_\text{I}} \frac{|\omega_\text{i} \cdot \mathbf{n}|_+}{\pi}\right\}, \qquad (13)$$

where we have defined $f_{\text{nd}}(\omega_\text{i}, \mathbf{n}) = L_\text{I}(\omega_\text{i})|\omega_\text{i} \cdot \mathbf{n}|_+/\pi$. Since the pdf now depends only on the surface normal $\mathbf{n}$ it can be easily precomputed (including the integral in the denominator) for a number of normal directions. This is already a solution applicable in practice, although it requires additional memory for storing the precomputed pdfs and possible changes to a particular renderer implementation.

Normal-independent pdf (ni). To obtain an even more convenient result that can be readily applied in existing renderers, we remove the dependence on $\mathbf{n}$ by averaging $p^{\text{nd}}_\text{I}(\omega_\text{i}|\mathbf{n})$ over all possible normal directions (see Appendix D). This yields our second approximation

$$p^{\text{ni}}_\text{I}(\omega_\text{i}) = \frac{1}{b_{\text{ni}}} \max\left\{0,\, L_\text{I}(\omega_\text{i}) - 2(1 - c_\text{I})\bar{L}_\text{I}\right\}, \qquad (14)$$

where $\bar{L}_\text{I}$ is the mean HDR map luminance and b_{ni} ensures that $p^{\text{ni}}_\text{I}(\omega_\text{i})$ integrates to one. This solution is as simple as subtracting a constant value from the tabulated HDR map pdf and re-normalizing.

In the rest of this section, we compare the derived formulas on a simple example. Sec. 7 then analyzes them in a Monte Carlo renderer.

6.3 Empirical test

We now empirically investigate the relation of our practical solutions (the normal-dependent p^{nd}_I, Eq. (13), and the normal-independent p^{ni}_I, Eq. (14)) to the general MIS-compensated solution $\tilde{p}_\text{I}$, Eq. (12),

[1] While we ignore occlusion in our derivations, it will likely influence variance of our IBL estimators in practice.

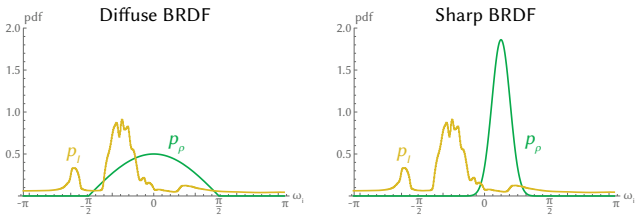


Fig. 4. Original pdfs used in our flatland setup. The BRDF sampling techniques use pdfs proportional to the product of the BRDF and the cosine term $|\omega_i \cdot \mathbf{n}|_+$. The diffuse BDRF (left) corresponds to $1/\pi$, while the sharp BRDF (right) corresponds to a normalized Phong lobe [Phong 1975] with exponent 20 shifted by $\frac{\pi}{8}$ away from the normal.

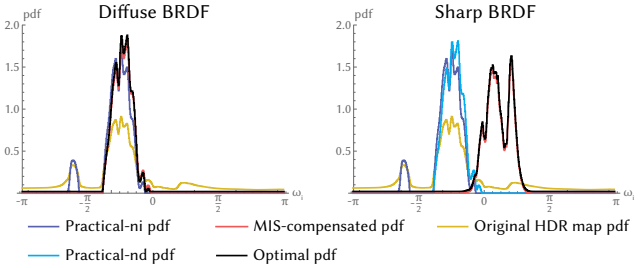


Fig. 5. Comparison of the *practical normal-independent (ni)* pdf p_I^{ni} (Eq. 14), the *practical normal-dependent (nd)* pdf p_I^{nd} (Eq. 13), the general *MIS-compensated* pdf $\tilde{p}_I$ (Eq. 12), the *optimal* pdf p_I^* (Eq. 9), and the original unmodified HDR map pdf p_I. The two cases (left, right) differ by the BRDF (see Fig. 4 for their definition). This figure demonstrates that the *optimal* and *MIS-compensated* pdfs are almost identical, even though the *MIS-compensated* pdf only uses the approximate formula Eq. (6). Note that for the diffuse BRDF, the *practical-nd* pdf is the same as the *MIS-compensated* pdf.

and also to the provably optimal solution p_I^*, Eq. (9). We consider a simple flatland setup as depicted in Fig. 3 with a 1D HDR map and two different BRDFs. The pdfs of the corresponding unmodified sampling strategies are shown in Fig. 4.

To compute the pdf for the optimal solution, we need to find the Lagrange multiplier λ, making p_I^* integrate to one. Since the norm of the pdf is a monotonous function of λ, the value normalizing the pdf can be found using a simple iterative bisection root-finder within 100 iterations.

Pdf shape comparison. Fig. 5 shows comparison of the *practical normal-independent*, *practical normal-dependent*, general *MIS-compensated*, and *optimal* pdfs, for the two different BRDFs.

A first important observation is that the general *MIS-compensated* and the *optimal* pdfs are virtually identical. We have tested more different combinations of BRDFs and HDR maps and we have not been able to find any case where the *MIS-compensated* and *optimal* pdf would appreciably differ (the difference, measured as normalized MSE, was less than 10^{-5}). This is additional evidence that the *MIS-compensated* pdf is a reliable approximation of the *optimal* result.

On the other hand, the two *practical* pdfs are similar to the general *MIS-compensated* pdf only for the diffuse BRDF, since a diffuse BRDF is an assumption used in their derivation. The two versions of the *practical* pdf, the normal-dependent (*practical-nd*) and the normal-independent (*practical-ni*), are similar, but *practical-ni* has some non-zero values for angles outside of the $[-\frac{\pi}{2}, \frac{\pi}{2}]$ interval.

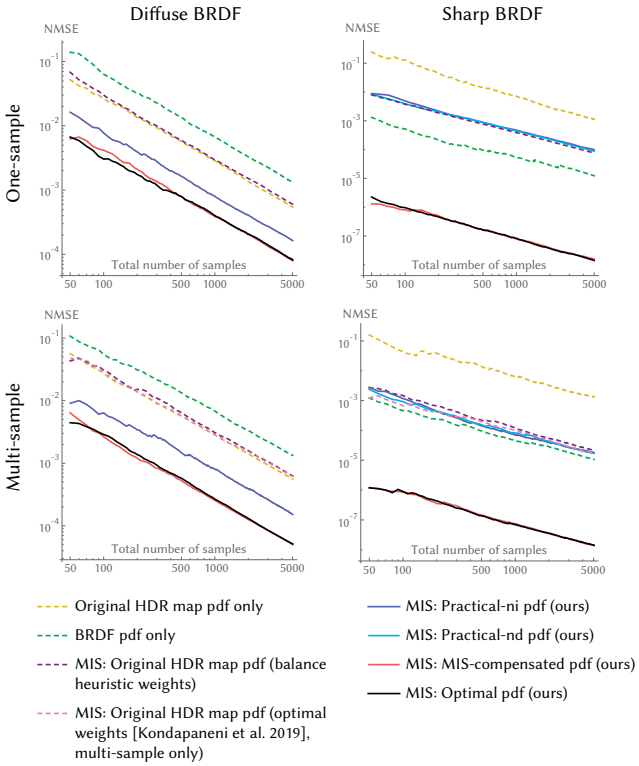


Fig. 6. Log-log normalized mean square error plots for different estimators. Two estimators use only one sampling technique each (Original HDR map and BRDF, respectively), while the other estimators use MIS. The Original HDR map pdf corresponds to the unmodified pdf based on HDR map values, while the *practical-ni*, *practical-nd*, *MIS-compensated* and *optimal* pdfs modify the original pdf as described in the text. The two columns (left, right) differ by the BRDF used (see Fig. 4). The first row corresponds to the one-sample estimator, the second row to the multi-sample estimator. For the one-sample version, only the balance heuristic is used (since it is optimal), while multi-sample additionally shows the result for the optimal MIS weights [Kondapaneni et al. 2019]. Note that for the diffuse BRDF, the *practical-nd* pdf is the same as the *MIS-compensated* pdf.

Estimator error comparison. Fig. 6 shows error plots when using the different pdfs to estimate image-based lighting Eq. (11) in the flatland setup (Fig. 3). We show results for the two different BRDFs and for the one-sample and multi-sample version (Eq. (4)) of the estimators (additional measurements are provided in the supplemental material). We first discuss the one-sample case.

For the diffuse BRDF, using MIS with the original pdf is roughly as good as sampling just from the HDR map. By using our *practical-ni* pdf, we decrease the normalized MSE 3.7 times. The *practical-nd*, *MIS-compensated* or *optimal* pdf reduces the MSE 7.5 times. In the sharp BRDF case, using any of our *practical* pdfs does not improve over baseline MIS, but, importantly, *it does not make the result any worse*. This shows that we can use even the simple *practical-ni* pdf without risking any negative impact. Using the *MIS-compensated* or *optimal* pdf for the sharp BRDF reduces error 847 times compared to sampling the BRDF and 4809 times compared to MIS without optimized pdfs. This provides an incentive for future work on finding better practical approximations of the *MIS-compensated* pdf.

Results for the multi-sample estimators are similar to the one-sample case. The biggest difference is visible for the sharp BRDF, where the various MIS combinations perform better than in the one-sample case, and are close to sampling from the BRDF only. This is in line with the known fact that multi-sample estimators have lower variance than their one-sample counterparts [Veach and Guibas 1995]. In the multi-sample case the balance heuristic is no longer the optimal choice for the MIS weighting functions, therefore we show also an MIS combination using the recently derived optimal weights [Kondapaneni et al. 2019] (not to be confounded with our *optimal* pdf discussed in Sec. 5). In our setup the optimal weighting provides only a minor improvement over the balance heuristic because no linear combination of the sampling pdfs is a good approximation to the integrand [Kondapaneni et al. 2019].

Nonetheless, the important point is that performance of any weighting, even the optimal one, is still limited by the use of a pre-defined, fixed set of sampling densities. In contrast, using the *MIS-compensated* pdf reduces the variance much more and suggests that optimizing sampling densities offers an additional opportunity for variance reduction over optimizing the MIS weighting functions.

7 RESULTS I: IMAGE-BASED LIGHTING

In this section we experimentally verify our theoretical results on a mix of production and synthetic scenes. We implement our *practical* method in a production path tracer[2], where we use it as a part of next event estimation carried out at each path vertex. All materials use a GGX BRDF with a Fresnel-modulated Lambertian diffuse term [Hill et al. 2016]. We render all results on a machine with an Intel® Core® i7-5820K CPU (3.3 GHz, 6 cores, 12 threads) and 64 GB of RAM.

One-sample vs. multi-sample. We test both one-sample and multi-sample estimators (Eq. (4)). For the multi-sample version the same number of samples from the HDR map pdf and the BRDF pdf are taken. For each sample in the one-sample version, one of the two pdfs is first chosen at random with equal probability. The one-sample estimator produces a slightly higher error in all our tests, otherwise the results for both versions are very similar. Therefore, here we only present the multi-sample version, while results for the one-sample estimator are provided in the supplemental material.

MIS heuristics. For the multi-sample estimator the balance heuristic weighting functions may not be the optimal choice. Therefore, we also implemented the *power heuristic* with power factor $\beta = 2$, the *cutoff heuristic* with cutoff threshold $q_{max} = 0.1$, and the *maximum heuristic* [Veach and Guibas 1995]. We then compared all four in all our scenes using the original HDR map pdf (i.e. without our method involved) with 160 samples per pixel. The balance heuristic gave us the best results, with the power, cutoff, and maximum heuristics producing, respectively, 3.14%, 0.03%, and 41.27% increase in the normalized mean squared error (NMSE), on average (complete measured data are provided in the supplemental material). Overall differences were visually and numerically almost negligible, except the maximum heuristic which performed much worse, in accordance with Veach and Guibas [1995] observation. As a result, we decided to compare our method only to the best performing balance heuristic. The recently introduced optimal weights [Kondapaneni et al. 2019] were not compared as their implementation is available only for computing direct illumination at the first bounce (so we could use it in our empirical tests in Sec. 6.3, but not in a path tracer).

[2]We use Corona Renderer (https://corona-renderer.com)

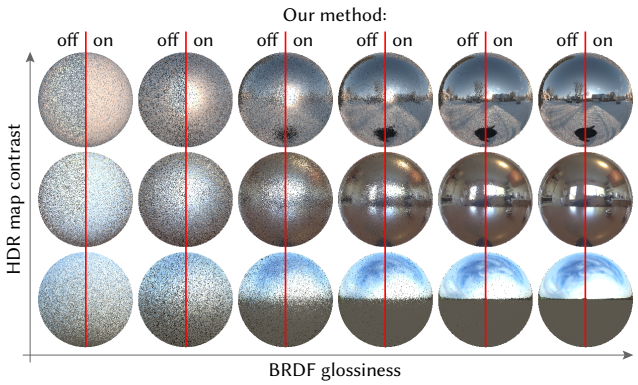


Fig. 7. Equal-sample (1 sample per pixel) comparison of image-based lighting on spheres with increasing BRDF glossiness and HDR map contrast. The left half of each sphere is rendered with regular MIS, and the right half with our *practical normal-independent* solution. The most visible improvement is achieved for a diffuse BRDF with a high-contrast HDR map (top-left corner), while in no case does our method make the result worse.

7.1 Normal-independent solution

Implementation. We implement our *practical normal-independent* solution (*practical-ni*) by simply modifying the tabulated pdf for sampling from the HDR map according to Eq. (14). This is done once in the preprocessing step, no other modifications of the path tracer are required. As a result, the method has virtually no overhead and marginal impact on the time per sample. To verify this, we have rendered all our scenes with 160 samples each, observing that on average our method increased the rendering time by 0.23%, including the time needed to modify the pdf (see the supplemental material for details). This means that *equal-sample* comparisons are also *equal-time* comparison for the *practical-ni* method.

Synthetic test. To systematically investigate the factors influencing the method's performance, we organize test cases into a two-dimensional space with BRDF glossiness and HDR map contrast as its axes. The results in Fig. 7 show that the method performs best for diffuse BRDFs and high-contrast HDR maps. The corresponding sphere is rendered with 8.26× difference in the NMSE. Our method makes only a marginal difference for highly glossy BRDFs, which render almost noise-free in the first place. A remarkable thing to note is that *our method never makes any of the results worse*, despite the assumption of a diffuse BRDF in our derivation.

Real production scenes. Next, we test our method in three real production scenes. *Pills*, a still-life scene shown in Fig. 1, *Room*, a typical interior architectural visualization scene lit by an exterior HDR map, combined with a weaker interior light, and *Car*, an automotive model with a glossy material on a diffuse background lit by a low-contrast map (8 bits per pixel, non-HDR photograph). The latter two are shown in Fig. 8. For all three scenes equal-time comparisons are presented, after 5 seconds for the *Pills* and *Car* scenes and 50 seconds for the *Room* scene.

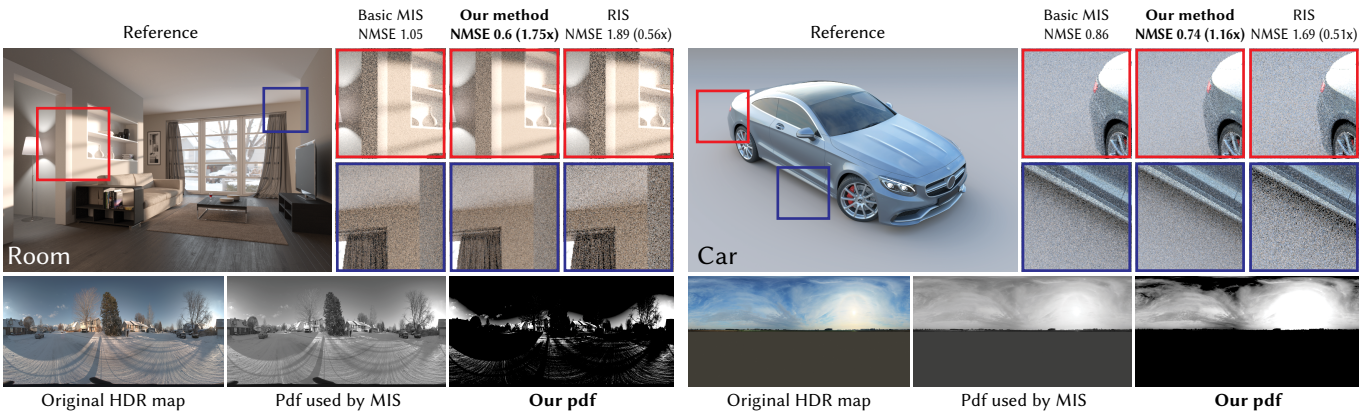


Fig. 8. Equal-time comparison of basic MIS with the balance heuristic (Basic MIS), resampled importance sampling (RIS) and our *practical normal-independent* solution applied to image-based lighting computation in real scenes. In the *Room* scene (50 s) our method visibly improves the noise level. The *Car* scene (5 s) on the other hand uses a low contrast non-HDR environment map and, subsequently, the differences are less visible in the side-by-side comparison. The bottom row shows the corresponding HDR maps and the used pdfs. The RIS method performs worse than Basic MIS in both scenes.

Our method consistently decreases the normalized MSE (NMSE) in all cases. Specifically, the error is 2.75× lower in the *Pills* scene, 1.75× in the *Room* scene, and 1.16× lower in the *Car* scene. These error reductions correspond to respective reductions of rendering time needed to achieve the same error. The key factor influencing performance of our method is the HDR map contrast – it performs best in the *Pills* scene featuring high contrast HDR map lighting. Our method has slightly lower benefit in the *Room* scene. We believe this is because there are additional variance sources, namely significant global illumination and shadowing. The *Car* scene is designed as a case in which our method does not provide a significant benefit due to the uniform lighting lacking contrast. Even there, we do not observe any NMSE increase; our method still improves convergence, albeit in a more subtle amount. From these results, as well as from the results of synthetic scenes, and our previous hands-on experience, we conclude that our *practical normal-independent* method is safe to be applied in any renderer for all scenes where image-based lighting is computed. See the supplemental material for a full comparison.

Comparison with an alternative IBL method. Besides MIS combination with the original HDR map pdf, we compare our method also with another IBL approach – resampled importance sampling (RIS) [Talbot et al. 2005]. Instead of using MIS, RIS approximates a sampling density proportional to the product of the HDR map and BRDF densities and draws samples directly from it. It first takes a set of M initial samples from one density, computes their pdf values in the other density, and then picks one sample from the set proportionally to these values. With increasing number of initial samples M, the resulting density approaches the product, but also the overhead of creating and storing the samples increases.

In our implementation we take the initial samples from the original HDR map pdf and then resample according to the BRDF pdf. We have found $M = 8$ to perform best for a given amount of time. This number is quite low since the overhead grows rapidly (for $M = 8$ it is already 60% on average, see the supplemental material for measurements). As a result, the RIS method does not provide any improvement over the original MIS combination.

We do not include hierarchical product sampling methods [Clarberg and Akenine-Möller 2008b; Jarosz et al. 2009; Rousselle et al. 2008]. These are difficult to apply in production rendering (our target application), because BRDFs can be given by shaders and cannot be easily represented in a wavelet or spherical harmonic basis.

7.2 Normal-dependent solution

Implementation. We implement our *practical normal-dependent* solution (*practical-nd*) as well. We select 512 directions uniformly covering the sphere and for each of them we precompute one tabulated pdf for sampling from the HDR map according to Eq. (13). For a given shading point we then first pick a direction closest to the normal at that point and then use the associated stored pdf. This incurs 10% overhead on average in our scenes (see the supplemental material for measurements).

To conserve memory, the precomputed tabulated pdfs for different normal directions have a fairly low resolution of 32×16. These pdfs are used to sample a region of the environment map. Inside the region we then sample only according to the *normal-independent practical-ni* pdf. This hybrid approach requires only about 1 MB memory in addition to storing the HDR map, while providing an effective approximation for sampling from *practical-nd*.

Results. Fig. 9 shows an equal-time comparison (5 s) in the *Pills* scene. To understand how much variance reduction comes just from including the normal dependency (without the MIS compensation) we tested also a method that for each normal direction only premultiplies the original HDR map pdf by the corresponding diffuse BRDF, i.e. computes Eq. (13) without the subtraction. Such a method mainly avoids sampling back-facing bright regions in the HDR map. It achieves 1.59× lower variance than the original MIS but still 1.73× higher than the *normal-independent practical-ni* pdf.

We can see that our MIS compensation, even in the simplest possible form (*practical-ni*) still offers better results than including the normal dependency alone. The full *normal-dependent practical-nd* method then performs the best, decreasing the variance 3.28× in comparison with original MIS, and 1.19× in comparison with

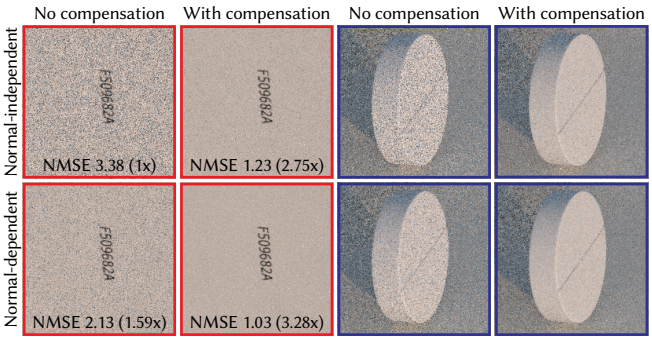


Fig. 9. Equal-time (5 s) comparison of the effect of MIS compensation versus including the normal dependency in the *Pills* scene. Just including the normal dependency (bottom left) improves over the basic MIS (top left) but the normal-independent MIS compensation (top right) is actually better. Combining both the normal dependency and the MIS compensation then yields the best result (bottom right).

practical-ni, yet at the cost of a slightly more complicated implementation. We provide a full comparison in the supplemental material.

Note that the not-compensated normal-dependent method corresponds to approximate product sampling on diffuse surfaces. Yet, adding the MIS-compensation further improves the result. This is because we compute MIS with the BRDF sampling and the compensation avoids wasting the product samples on directions already well sampled according to the BRDF. Sampling using the not-compensated method only (without MIS) could perform better on diffuse surfaces but would fail on highly glossy ones.

8 APPLICATION II: PATH GUIDING

Next we apply the *MIS-compensated* solution to path guiding [Vorba et al. 2019]. Instead of unoccluded direct illumination in the IBL problem we now compute full global illumination, i.e. the integral

$$L(\mathbf{x}, \omega_\mathrm{o}) = \int_{\mathcal{H}(\mathbf{n})} L_\mathrm{G}(\omega_\mathrm{i}, \mathbf{x})\, \rho(\mathbf{x}, \omega_\mathrm{o}, \omega_\mathrm{i}) |\omega_\mathrm{i} \cdot \mathbf{n}|_+ \, \mathrm{d}\omega_\mathrm{i}. \qquad (15)$$

The only difference from the IBL integral in Eq. (11) is that the HDR map emission $L_\mathrm{I}(\omega_\mathrm{i})$ is replaced by total radiance $L_\mathrm{G}(\omega_\mathrm{i}, \mathbf{x})$ arriving at the surface position $\mathbf{x}$ from the direction ω_i. Note that while the former does not depend on the surface position the latter does.

Path guiding approaches usually combine sampling of the BRDF-cosine product and sampling proportionally to a learned approximation of $L_\mathrm{G}(\omega_\mathrm{i}, \mathbf{x})$ using the one-sample MIS estimator (Eq. (4)). We choose the method of Müller et al. [2017] since it uses a practical discrete approximation of $L_\mathrm{G}(\omega_\mathrm{i}, \mathbf{x})$ stored in a 5D spatio-directional tree structure. A pdf $p_\mathrm{G}(\omega_\mathrm{i}|\mathbf{x})$ of the corresponding sampling technique is therefore again tabulated and suitable for being optimized by our method while the BRDF sampling pdf $p_\rho(\omega_\mathrm{i}|\omega_\mathrm{o}, \mathbf{x})$ is held fixed. Note that this time the tabulation is done not only for directions ω_i but also for surface positions $\mathbf{x}$.

By applying the exactly same steps as in the IBL problem in Sec. 6.1 and 6.2, it is easy to derive the *MIS-compensated* pdf $\tilde{p}_\mathrm{G}$, *practical normal-dependent* pdf p_G^nd and *practical normal-independent* pdf p_G^ni. Since the first two would require additional tabulation for outgoing directions ω_o and normals $\mathbf{n}$, respectively (i.e. 7D table instead of 5D), we focus solely on the *normal-independent* solution

$$p_\mathrm{G}^\mathrm{ni}(\omega_\mathrm{i}|\mathbf{x}) = \frac{1}{b_\mathrm{ni}} \max\left\{0,\, L_\mathrm{G}(\omega_\mathrm{i}, \mathbf{x}) - 2(1 - c_\mathrm{G})\bar{L}_\mathrm{G}(\mathbf{x})\right\}. \qquad (16)$$

Here $\bar{L}_\mathrm{G}(\mathbf{x})$ is the mean radiance arriving from all directions to the surface position $\mathbf{x}$ and b_ni ensures that $p_\mathrm{G}^\mathrm{ni}(\omega_\mathrm{i}|\mathbf{x})$ integrates to one.

When sampling the direction ω_i at the surface position $\mathbf{x}$, Müller et al. [2017] first descend a binary tree dividing the 3D spatial domain to find a leaf box containing $\mathbf{x}$. It stores a quad tree dividing the 2D directional domain. Each of its leaf quads Q contains an estimate of the total incident radiance $\hat{L}_\mathrm{G}(Q, \mathbf{x})$ arriving from all directions represented by Q to the leaf box containing $\mathbf{x}$. One of the leaf quads Q_i is then selected proportionally to these estimates and the represented directions are sampled uniformly to obtain ω_i. The *practical normal-independent* pdf for sampling the leaf quads is obtained by integrating $p_\mathrm{G}^\mathrm{ni}(\omega_\mathrm{i}|\mathbf{x})$ over all directions ω_i in Q_i yielding

$$p_\mathrm{G}^\mathrm{ni}(Q_\mathrm{i}|\mathbf{x}) = \frac{1}{b'_\mathrm{ni}} \max\left\{0,\, \hat{L}_\mathrm{G}(Q_\mathrm{i}, \mathbf{x}) - 2(1 - c_\mathrm{G})|Q_\mathrm{i}| \sum_k \hat{L}_\mathrm{G}(Q_k, \mathbf{x})\right\}, \qquad (17)$$

where $|Q_\mathrm{i}|$ denotes the area of the leaf quad Q_i.

9 RESULTS II: PATH GUIDING

Now we test our path guiding solution in rendering. We use the latest Mitsuba implementation of the Müller et al.'s method [2017] provided by its authors. It runs in iterations: samples taken in one iteration update data in the 5D tree structure which then guides sampling in the next iteration. To apply our MIS compensation, we perform one pass over all quad trees in the structure at the beginning of every iteration and modify the quad pdfs according to Eq. (17).

Our MIS compensation results in better guiding pdfs: Not only is variance reduced, but paths are generally shorter as they reach a light source sooner and the path tracing time decreases (we observe 7% rendering time decrease, see the supplemental material).

Fig. 10 shows an equal-time comparison (150 s) in two scenes with complex indirect illumination. The *Kitchen* scene contains various glossy materials and is lit by sunlight entering through a window, reflected off a glass tabletop. The *Pool* scene features difficult specular-diffuse-specular light transport (e.g. refracted caustics) inside the pool. While the Müller et al.'s algorithm performs well in these scenes, our solution can improve its performance even further. It optimizes the guiding pdf so it omits the directions that are already well sampled by the BRDF technique and focuses on the most important regions only. This leads to a visible noise reduction, specifically the error is decreased 1.38× in the *Kitchen* scene and 1.6× in the *Pool* scene. The improvement is not as high as in the IBL problem because global illumination at a point has typically lower contrast than direct illumination. In summary, the modification is simple to implement, it decreases both variance and time per sample, and we have not observed any failure cases. A full comparison is provided in the supplemental material.

10 LIMITATIONS AND FUTURE WORK

Number of samples. We use equal sample allocation among the sampling techniques in all our tests. Optimized sample allocation [He

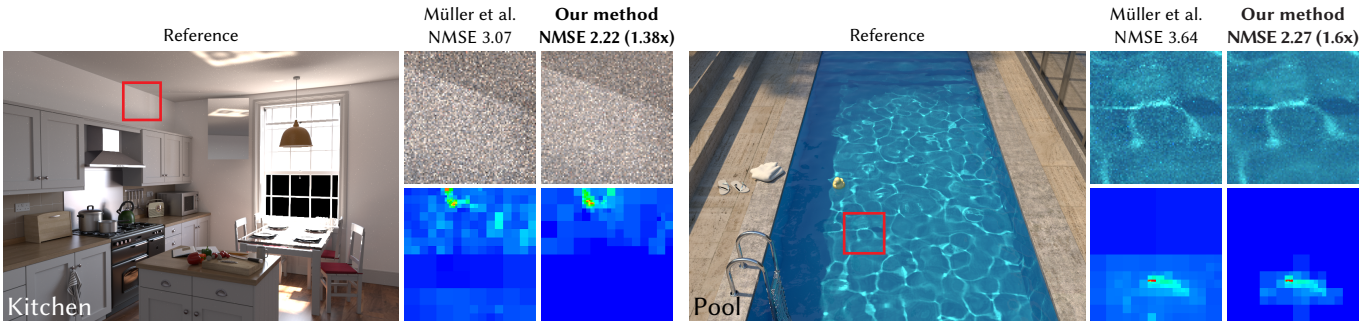


Fig. 10. Equal-time (150 s) comparison of the path guiding method by Müller et al. [2017] and our *practical normal-independent* solution applied on top of it. The bottom row shows false color visualizations of the guiding pdf over directions on a sphere as used at a surface point in the middle of the corresponding inset. By optimizing the pdf our method achieves visible noise reduction.

and Owen 2014; Lu et al. 2013; Pajot et al. 2011; Sbert and Havran 2017; Sbert et al. 2019; Vorba et al. 2019] may yield further improvement. Joint optimization of both the free pdf and the sample allocation presents an interesting open problem.

Occlusion. Our derivation for the IBL problem ignores occlusion. Including it would require a position-dependent pdf and learning occlusion in the scene. An interesting idea for future research is to account for occlusion via Control Variate approaches [Belcour et al. 2018; Clarberg and Akenine-Möller 2008a].

Better approximation. While our current approach significantly reduces the estimator variance, it is based on a crude approximation of the *MIS-compensated* pdf derived in Sec. 4. The example in Sec. 6.3 shows that the *MIS-compensated* pdf can achieve orders of magnitude better variance reduction than the currently used *practical* pdf. This is due to the *practical* pdf ignoring the BRDF and the viewing direction. Incorporating some of these features should result in a more efficient estimator.

Analytic pdfs. While optimization of tabulated pdfs is straightforward, the same is not true for analytic pdfs. Successful application of our optimization to analytic pdfs may, therefore, involve some non-trivial approximations. A possible direction could use finite function spaces as a pdf representation and optimize directly the respective coefficients.

Applications. The general pdf optimization derived in Sec. 4 can be applied to a wide range of other problems relying on MIS, such as free-flight sampling in heterogeneous media [Kulla and Fajardo 2012], light sampling [Georgiev et al. 2012a; Vévoda et al. 2018], or path connections sampling [Popov et al. 2015]. All those methods feature empirically constructed pdfs combined with one or more defensive techniques and should lend themselves well to our method.

11 CONCLUSION

We extend the arsenal of tools available for variance reduction in combined estimators by optimizing one of the sampling densities itself. Given a multiple importance sampling estimator with the balance heuristic weighting functions, we derive an MIS-compensated pdf for one of the sampling techniques that reduces the overall variance of the MIS estimator. We show that this solution is bounded by the provably optimal, yet more expensive pdf.

While we apply our MIS compensation to image-based lighting and path guiding, yielding significant variance reduction, the technique is more general and can be used in other applications, potentially even outside of computer graphics. Thanks to its simplicity, efficiency, and robustness, we have been able to use it in a production renderer without encountering any failure cases.

ACKNOWLEDGMENTS

This work was supported by the Czech Science Foundation Grant 19-07626S and the Charles University Grant SVV-2017-260452.

REFERENCES

Sameer Agarwal, Ravi Ramamoorthi, Serge Belongie, and Henrik Wann Jensen. 2003. Structured importance sampling of environment maps. *ACM Trans. Graph.* 22, 3 (July 2003), 605–612.

Laurent Belcour, Guofu Xie, Christophe Hery, Mark Meyer, Wojciech Jarosz, and Derek Nowrouzezahrai. 2018. Integrating Clipped Spherical Harmonics Expansions. *ACM Trans. Graph.* 37, 2, Article 19 (July 2018).

David Burke, Abhijeet Ghosh, and Wolfgang Heidrich. 2005. Bidirectional Importance Sampling for Direct Illumination. In *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques (EGSR '05)*. 147–156.

Olivier Cappé, Randal Douc, Arnaud Guillin, Jean-Michel Marin, and Christian P. Robert. 2008. Adaptive importance sampling in general mixture classes. *Statistics and Computing* 18, 4 (December 2008), 447–459.

Petrik Clarberg and Tomas Akenine-Möller. 2008a. Exploiting Visibility Correlation in Direct Illumination. In *Eurographics Symposium on Rendering (EGSR '08)*.

Petrik Clarberg and Tomas Akenine-Möller. 2008b. Practical Product Importance Sampling for Direct Illumination. *Comput. Graph. Forum* 27, 2 (April 2008).

Petrik Clarberg, Wojciech Jarosz, Tomas Akenine-Möller, and Henrik Wann Jensen. 2005. Wavelet Importance Sampling: Efficiently Evaluating Products of Complex Functions. *ACM Trans. Graph.* 24, 3 (July 2005), 1166–1175.

Jean-Marie Cornuet, Jean-Michel Marin, Antonietta Mira, and Christian P. Robert. 2012. Adaptive multiple importance sampling. *Scandinavian Journal of Stats.* 39, 4 (2012).

Víctor Elvira, Luca Martino, David Luengo, and Mónica F. Bugallo. 2015. Generalized multiple importance sampling. arXiv:1511.03095.

Víctor Elvira, Luca Martino, David Luengo, and Mónica F. Bugallo. 2016. Heretical multiple importance sampling. *IEEE Signal Processing Letters* 23, 10 (October 2016).

Shaohua Fan, Stephen Chenney, Bo Hu, Kam Wah Tsui, and Yu Chi Lai. 2006. Optimizing control variate estimators for rendering. *Comput. Graph. Forum (EUROGRAPHICS '06)* 25, 3 (2006), 351–357.

Iliyan Georgiev, Jaroslav Křivánek, Stefan Popov, and Philipp Slusallek. 2012a. Importance caching for complex illumination. *Comput. Graph. Forum (EUROGRAPHICS '12)* 31, 2 (May 2012), 701–710.

Iliyan Georgiev, Jaroslav Křivánek, Tomáš Davidovič, and Philipp Slusallek. 2012b. Light transport simulation with vertex connection and merging. *ACM Trans. Graph. (SIGGRAPH Asia '12)* 31, 6 (November 2012).

Toshiya Hachisuka, Jacopo Pantaleoni, and Henrik Wann Jensen. 2012. A path space extension for robust light transport simulation. *ACM Trans. Graph. (SIGGRAPH Asia '12)* 31, 6 (November 2012).

Vlastimil Havran and Mateu Sbert. 2014. Optimal Combination of Techniques in Multiple Importance Sampling. In *Proceedings of the 13th ACM SIGGRAPH International*

Conference on Virtual-Reality Continuum and Its Applications in Industry. 141–150.
Hera Y. He and Art B. Owen. 2014. Optimal mixture weights in multiple importance sampling. arXiv:1411.3954.
Sebastian Herholz, Oskar Elek, Jiří Vorba, Hendrik Lensch, and Jaroslav Křivánek. 2016. Product importance sampling for light transport path guiding. *Comput. Graph. Forum* 35, 4 (July 2016), 67–77.
Sebastian Herholz, Yangyang Zhao, Oskar Elek, Derek Nowrouzezahrai, Hendrik P. A. Lensch, and Jaroslav Křivánek. 2019. Volume Path Guiding Based on Zero-Variance Random Walk Theory. *ACM Trans. Graph.* 38, 3 (2019).
Tim Hesterberg. 1995. Weighted average importance sampling and defensive mixture distributions. *Technometrics* 37, 2 (May 1995), 185–194.
Stephen Hill, Stephen McAuley, Cyril Jover, Sébastien Lachambre, Angelo Pesce, and Xian-Chun Wu. 2016. Physically based shading in theory and practice. In *ACM SIGGRAPH 2016 Courses (SIGGRAPH '16)*. Article 21.
Wojciech Jarosz, Nathan A. Carr, and Henrik Wann Jensen. 2009. Importance sampling spherical harmonics. *Comput. Graph. Forum* 28, 2 (April 2009), 577–586.
Henrik Wann Jensen. 1995. Importance driven path tracing using the photon map. In *Rendering Techniques*. 326–335.
Alexander Keller, Luca Fascione, Marcos Fajardo, Iliyan Georgiev, Per Christensen, Johannes Hanika, Christian Eisenacher, and Gregory Nichols. 2015. The path tracing revolution in the movie industry. In *ACM SIGGRAPH 2015 Courses*. Article 24.
Alan King, Christopher Kulla, Alejandro Conty, and Marcos Fajardo. 2013. BSSRDF importance sampling. In *ACM SIGGRAPH 2013 Talks (SIGGRAPH '13)*. Article 48.
Ivo Kondapaneni, Petr Vévoda, Pascal Grittmann, Tomáš Skřivan, Philipp Slusallek, and Jaroslav Křivánek. 2019. Optimal Multiple Importance Sampling. *ACM Trans. Graph. (SIGGRAPH 2019)* 38, 4 (July 2019).
Jaroslav Křivánek, Iliyan Georgiev, Toshiya Hachisuka, Petr Vévoda, Martin Šik, Derek Nowrouzezahrai, and Wojciech Jarosz. 2014. Unifying points, beams, and paths in volumetric light transport simulation. *ACM Trans. Graph.* 33, 4 (August 2014).
Jaroslav Křivánek, Ondřej Karlík, Vladimir Koylazov, Henrik Wann Jensen, Thomas Ludwig, and Christophe Chevallier. 2018. Realistic Rendering in Architecture and Product Visualization. In *ACM SIGGRAPH 2018 Courses (SIGGRAPH '18)*. Article 10.
Christopher Kulla and Marcos Fajardo. 2012. Importance sampling techniques for path tracing in participating media. *Comput. Graph. Forum* 31, 4 (June 2012), 1519–1528.
Eric P. Lafortune and Yves D. Willems. 1995. A 5D Tree to Reduce the Variance of Monte Carlo Ray Tracing. In *Proceedings of the 6th Eurographics Workshop on Rendering*.
Yu-Chi Lai, Hsuan-Ting Chou, Kuo-Wei Chen, and Shaohua Fan. 2015. Robust and efficient adaptive direct lighting estimation. *The Visual Computer* 31, 1 (2015).
Heqi Lu, Romain Pacanowski, and Xavier Granier. 2013. Second-order approximation for variance reduction in multiple importance sampling. *Comput. Graph. Forum* 32, 7 (October 2013).
Thomas Müller, Markus Gross, and Jan Novák. 2017. Practical path guiding for efficient light-transport simulation. *Comput. Graph. Forum (EGSR '17)* 36, 4 (June 2017).
Art Owen and Yi Zhou. 2000. Safe and effective importance sampling. *J. Amer. Statist. Assoc.* 95, 449 (March 2000), 135–143.
Anthony Pajot, Loic Barthe, Mathias Paulin, and Pierre Poulin. 2011. Representativity for robust and adaptive multiple importance sampling. *IEEE Transactions on Visualization and Computer Graphics* 17, 8 (August 2011), 1108–1121.
Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2016. *Physically Based Rendering, Third Edition: From Theory to Implementation* (3rd ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
Bui Tuong Phong. 1975. Illumination for computer generated pictures. *Commun. ACM* 18, 6 (June 1975), 311–317.
Stefan Popov, Ravi Ramamoorthi, Frédo Durand, and George Drettakis. 2015. Probabilistic connections for bidirectional path tracing. *Comput. Graph. Forum* 34, 4 (2015).
Fabrice Rousselle, Petrik Clarberg, Luc Leblanc, Victor Ostromoukhov, and Pierre Poulin. 2008. Efficient product sampling using hierarchical thresholding. *The Visual Computer* 24, 7 (July 2008), 465–474.
Andrzej Ruszczynski. 2006. *Nonlinear Optimization*. Princeton University Press.
Mateu Sbert and Vlastimil Havran. 2017. Adaptive multiple importance sampling for general functions. *The Visual Computer* 33, 6 (June 2017), 845–855.
Mateu Sbert, Vlastimil Havran, and Laszlo Szirmay-Kalos. 2016. Variance Analysis of Multi-sample and One-sample Multiple Importance Sampling. *Comput. Graph. Forum* 35, 7 (October 2016), 451–460.
Mateu Sbert, Vlastimil Havran, and László Szirmay-Kalos. 2018. Multiple importance sampling revisited: breaking the bounds. *EURASIP Journal on Advances in Signal Processing* 2018, 1 (February 2018).
Mateu Sbert, Vlastimil Havran, and László Szirmay-Kalos. 2019. Optimal Deterministic Mixture Sampling. In *Eurographics 2019, Short Papers*.
Justin Talbot, David Cline, and Parris K. Egbert. 2005. Importance resampling for global illumination. In *Eurographics Symposium on Rendering*. 139–146.
Glenn Torrie and John Valleau. 1977. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* 23, 2 (1977).
Eric Veach and Leonidas J. Guibas. 1995. Optimally combining sampling techniques for Monte Carlo rendering. In *SIGGRAPH '95*. 419–428.
Petr Vévoda, Ivo Kondapaneni, and Jaroslav Křivánek. 2018. Bayesian Online Regression for Adaptive Direct Illumination Sampling. *ACM Trans. Graph.* 37, 4 (2018). https://doi.org/10.1145/3197517.3201340
Jiří Vorba, Johannes Hanika, Sebastian Herholz, Thomas Müller, Jaroslav Křivánek, and Alexander Keller. 2019. Path Guiding in Production. In *ACM SIGGRAPH 2019 Courses*. https://doi.org/10.1145/3305366.3328091
Jiří Vorba, Ondřej Karlík, Martin Šik, Tobias Ritschel, and Jaroslav Křivánek. 2014. On-line Learning of Parametric Mixture Models for Light Transport Simulation. *ACM Trans. Graph. (SIGGRAPH '14)* 33, 4 (July 2014).

A DERIVATION OF THE PROVABLY OPTIMAL PDF

Here we discuss minimization of the functional $J[p]$ from Eq. (7) under the constraints Eq. (8). We solve the problem by satisfying the Karush-Kuhn-Tucker (KKT) necessary conditions, which generalize Lagrange multipliers to inequality contraints [Ruszczynski 2006]. We start by rewriting the constraints as *primal feasibility conditions*

$$k_1[p] = \int_{\mathcal{X}} p(x)\,\mathrm{d}x - 1 = 0, \qquad k_2[p] = -p(x) \leq 0. \tag{18}$$

Given KKT multipliers $\lambda \in \mathbb{R}$ and $\mu(x) \in \mathbb{R} \to \mathbb{R}$, we must satisfy the *stationarity condition*

$$\lambda \nabla k_1[p] + \langle \mu | \nabla k_2[p] \rangle = -\nabla J[p], \tag{19}$$

where $\langle v|u\rangle = \int_{\mathcal{X}} v(x)u(x)\,\mathrm{d}x$ is the standard L^2 inner product. The gradient $\nabla J[p]$ can be computed from the directional derivative as

$$\begin{aligned}\langle \nabla J[p] | \mathrm{d}p \rangle &= \frac{\mathrm{d}}{\mathrm{d}p}\int_{\mathcal{X}} \frac{f^2(x)}{q(x)+c_\tau p(x)}\,\mathrm{d}x = \int_{\mathcal{X}} \mathrm{d}p \frac{-c_\tau f^2(x)}{(q(x)+c_\tau p(x))^2}\mathrm{d}x \\ \nabla J[p] &= \frac{-c_\tau f^2(x)}{(q(x)+c_\tau p(x))^2}.\end{aligned} \tag{20}$$

Similarly, we compute $\nabla k_1[p] = 1$ and $\langle \mu | \nabla k_2[p] \rangle = -\mu(x)$. Plugging these results into the stationarity condition, Eq. (19), we get $\lambda - \mu(x) = -\frac{-c_\tau f^2(x)}{(q(x)+c_\tau p(x))^2}$, which yields the solution

$$p(x) = \frac{f(x)}{\sqrt{c_\tau(\lambda - \mu(x))}} - \frac{q(x)}{c_\tau}. \tag{21}$$

Finally, due to the *complementary slackness* condition $\mu(x)p(x) = 0$, we know that $\mu(x)$ must be equal to zero whenever $p(x) > 0$, and therefore we can write the solution as

$$p(x) = \max\left\{0, \frac{f(x)}{\sqrt{c_\tau \lambda}} - \frac{q(x)}{c_\tau}\right\}. \tag{22}$$

The KKT multiplier λ now remains the only unknown, but we know that it must ensure $k_1[p] = 0$ (i.e. the resulting pdf must integrate to 1) and thus we may find λ using an iterative root-finding method.

B DERIVATION OF BOUNDS ON λ

Here we derive bounds of the Lagrange multiplier λ from the *optimal* pdf solution in Eq. (9). We start by defining the subspace $\mathcal{X}_\tau \subseteq \mathcal{X}$

$$\mathcal{X}_\tau = \{x \in \mathcal{X} | p^{\pm}_\tau(x) > 0\}. \tag{23}$$

Given $\mathcal{X}_\tau$ the integration constraint from Eq. (8) can be rewritten using $p^{\pm}_\tau(x)$ defined in Eq. (9) instead of $p^*_\tau(x)$, which eliminates the

max operator and allows us to bound λ from above as follows

$$
\begin{aligned}
1 &= \int_{\mathcal{X}_\tau} p_\tau^{\pm}(x)\,\mathrm{d}x = \int_{\mathcal{X}_\tau} \frac{f(x)}{\sqrt{c_\tau \lambda}} - \frac{q(x)}{c_\tau}\,\mathrm{d}x \\
\sqrt{\lambda} &= \int_{\mathcal{X}_\tau} \frac{f(x)}{\sqrt{c_\tau}}\,\mathrm{d}x - \frac{\sqrt{\lambda}}{c_\tau}\int_{\mathcal{X}_\tau} q(x)\,\mathrm{d}x \\
\sqrt{\lambda} &= \frac{\int_{\mathcal{X}_\tau} \frac{f(x)}{\sqrt{c_\tau}}\,\mathrm{d}x}{1 + \frac{1}{c_\tau}\int_{\mathcal{X}_\tau} q(x)\,\mathrm{d}x} \\
\lambda &= c_\tau \left(\frac{\int_{\mathcal{X}_\tau} f(x)\,\mathrm{d}x}{c_\tau + \int_{\mathcal{X}_\tau} q(x)\,\mathrm{d}x}\right)^2 \le c_\tau \left(\frac{\int_{\mathcal{X}} f(x)\,\mathrm{d}x}{c_\tau}\right)^2 = \frac{F^2}{c_\tau}. \qquad (24)
\end{aligned}
$$

Before deriving the lower bound, let us realize that the following inequality holds directly from the definition of $\mathcal{X}_\tau$, since for every $x \in \mathcal{X} \setminus \mathcal{X}_\tau$ the pdf $p_\tau^{\pm}(x)$ is not positive

$$
\int_{\mathcal{X}\setminus\mathcal{X}_\tau} \frac{f(x)}{\sqrt{c_\tau\lambda}}\,\mathrm{d}x \le \int_{\mathcal{X}\setminus\mathcal{X}_\tau} \frac{q(x)}{c_\tau}\,\mathrm{d}x. \qquad (25)
$$

Note that if $\mathcal{X}$ equals $\mathcal{X}_\tau$ the above is an equality, since we integrate over empty set. Now we can derive the lower bound on λ from $F\sqrt{c_\tau}$

$$
\begin{aligned}
F\sqrt{c_\tau} &= \int_{\mathcal{X}_\tau} f(x)\sqrt{c_\tau}\,\mathrm{d}x + \int_{\mathcal{X}\setminus\mathcal{X}_\tau} f(x)\sqrt{c_\tau}\,\mathrm{d}x \\
&\le \int_{\mathcal{X}_\tau} f(x)\sqrt{c_\tau}\,\mathrm{d}x + \int_{\mathcal{X}\setminus\mathcal{X}_\tau} q(x)\sqrt{\lambda}\,\mathrm{d}x \\
&= \int_{\mathcal{X}_\tau} f(x)\sqrt{c_\tau}\,\mathrm{d}x + \sqrt{\lambda}\left(1 - c_\tau - \int_{\mathcal{X}_\tau} q(x)\,\mathrm{d}x\right) \\
&= \int_{\mathcal{X}_\tau} \left(f(x)\sqrt{c_\tau} - \sqrt{\lambda}q(x)\right)\,\mathrm{d}x + \sqrt{\lambda}(1 - c_\tau) \\
&= \sqrt{\lambda}c_\tau \int_{\mathcal{X}_\tau} \left(\frac{f(x)}{\sqrt{c_\tau\lambda}} - \frac{q(x)}{c_\tau}\right)\,\mathrm{d}x + \sqrt{\lambda}(1 - c_\tau) \\
&= \sqrt{\lambda}c_\tau + \sqrt{\lambda}(1 - c_\tau) = \sqrt{\lambda}, \qquad (26)
\end{aligned}
$$

i.e. $F^2 c_\tau \le \lambda$. We have used Eq. (25) and $\int_{\mathcal{X}} q(x)\,\mathrm{d}x = 1 - c_\tau$. Note that if $\mathcal{X}$ equals $\mathcal{X}_\tau$, Eq. (25) becomes equality and λ equals $F^2 c_\tau$.

C BOUNDING ESTIMATOR'S SECOND MOMENT

Here we derive the upper bound of one sample estimator's second moment (Eq. (7)) for the *MIS-compensated* pdf. We make several observations: given the bounds on λ (Eq.(10)), we can write $\sqrt{c_\tau\lambda}$ as

$$
\sqrt{c_\tau\lambda} = kF, \qquad c_\tau \le k \le 1. \qquad (27)
$$

We observe that normalization constant b of $\tilde{p}_\tau(x)$, Eq. (6), can be bounded from above as follows

$$
b = \int_{\mathcal{X}} \max\{0, \frac{f(x)}{c_\tau F} - \frac{q(x)}{c_\tau}\}\,\mathrm{d}x \le \int_{\mathcal{X}} \frac{f(x)}{c_\tau F}\,\mathrm{d}x = \frac{1}{c_\tau} \qquad (28)
$$

and also from below (where we use the lower bound $F^2 c_\tau \le \lambda$)

$$
\begin{aligned}
b &\ge \int_{\mathcal{X}_\tau} \max\{0, \frac{f(x)}{c_\tau F} - \frac{q(x)}{c_\tau}\}\,\mathrm{d}x \\
&\ge \int_{\mathcal{X}_\tau} \max\{0, \frac{f(x)}{\sqrt{c_\tau\lambda}} - \frac{q(x)}{c_\tau}\}\,\mathrm{d}x = 1. \qquad (29)
\end{aligned}
$$

We first prove that the effective pdf p_{eff} is greater or equal in $\mathcal{X} \setminus \mathcal{X}_\tau$ if we use the *MIS-compensated* pdf instead of the *optimal* pdf

$$
p_{\text{eff}}(x) = q(x) + c_\tau\tilde{p}_\tau(x) \ge q(x) = q(x) + c_\tau p_\tau^*(x), \qquad (30)
$$

where we have used the fact that $\tilde{p}_\tau(x) \ge 0$ and $p_\tau^*(x) = 0$ in $\mathcal{X} \setminus \mathcal{X}_\tau$. Now we bound the same effective pdf p_{eff} in $\mathcal{X}_\tau$

$$
\begin{aligned}
q(x) + c_\tau\tilde{p}_\tau(x) &= q(x) + \frac{c_\tau}{b}\left(\frac{f(x)}{c_\tau F} - \frac{q(x)}{c_\tau}\right) \\
&= \frac{k}{bc_\tau}\left(\frac{bc_\tau - c_\tau + k}{k}q(x) + c_\tau p_\tau^*(x)\right) \\
&\ge \frac{k}{bc_\tau}\left(q(x) + c_\tau p_\tau^*(x)\right) \\
&\ge c_\tau\left(q(x) + c_\tau p_\tau^*(x)\right), \qquad (31)
\end{aligned}
$$

where we have used Eq. (27), the fact that $bc_\tau - c_\tau \ge 0$ due to the lower bound of b in Eq. (29), and $k/(bc_\tau) \ge c_\tau$ due to lower bound of k in Eq. (27) and upper bound of b in Eq. (28).

If we now put Eqs. (30) and (31) together we get the following inequality for all $x \in \mathcal{X}$: $q(x) + c_\tau\tilde{p}_\tau(x) \ge c_\tau(q(x) + c_\tau p_\tau^*(x))$. From this we immediately get the inequality for the second moment of the one-sample estimator (Eq. (7)), $J[\tilde{p}_\tau] \le \frac{1}{c_\tau} J[p_\tau^*]$.

D DERIVING PRACTICAL HDR MAP PDF

Here we derive the final pdf for practical sampling from an HDR map, Eq. (14), by integrating the pdf $p_{\text{I}}^{\text{nd}}(\omega_{\text{i}}|\mathbf{n})$ defined in Eq. (13) over all normal directions with positive cosine term $|\omega_{\text{i}} \cdot \mathbf{n}|_+$.

$$
\int_{\mathcal{H}(\omega_{\text{i}})} 1/b_{\text{nd}} \max\{0, \overbrace{\frac{f_{\text{nd}}(\omega_{\text{i}}, \mathbf{n})}{c_{\text{I}}\int_{\mathcal{H}(\mathbf{n})} f_{\text{nd}}(\omega, \mathbf{n})\,\mathrm{d}\omega}}^{P_1(\omega_{\text{i}},\mathbf{n})} - \frac{1 - c_{\text{I}}}{c_{\text{I}}}\overbrace{\frac{|\omega_{\text{i}} \cdot \mathbf{n}|_+}{\pi}}^{P_2(\omega_{\text{i}},\mathbf{n})}\,\mathrm{d}\mathbf{n}.
$$

Since the above integral cannot be evaluated analytically, we move the integration inside the max operator and integrate P_1 and P_2 from the above equation separately. We can see that $\int_{\mathcal{H}(\omega_{\text{i}})} P_2(\omega_{\text{i}}, \mathbf{n})\,\mathrm{d}\mathbf{n} = 1$ and we approximate integration of $P_1(\omega_{\text{i}}, \mathbf{n})$ as

$$
\begin{aligned}
\int_{\mathcal{H}(\omega_{\text{i}})} P_1(\omega_{\text{i}}, \mathbf{n})\,\mathrm{d}\mathbf{n} &= \int_{\mathcal{H}(\omega_{\text{i}})} \frac{f_{\text{nd}}(\omega_{\text{i}}, \mathbf{n})}{c_{\text{I}}\int_{\mathcal{H}(\mathbf{n})} f_{\text{nd}}(\omega, \mathbf{n})\,\mathrm{d}\omega}\,\mathrm{d}\mathbf{n} \\
&\approx \int_{\mathcal{H}(\omega_{\text{i}})} \frac{f_{\text{nd}}(\omega_{\text{i}}, \mathbf{n})}{c_{\text{I}}\int_{\mathcal{H}(\mathbf{n})} L_{\text{I}}(\omega)/\pi\,\mathrm{d}\omega}\,\mathrm{d}\mathbf{n} \\
&\approx \int_{\mathcal{H}(\omega_{\text{i}})} \frac{L_{\text{I}}(\omega_{\text{i}})|\omega_{\text{i}} \cdot \mathbf{n}|_+/\pi}{2\bar{L}_{\text{I}}c_{\text{I}}}\,\mathrm{d}\mathbf{n} \\
&= \frac{L_{\text{I}}(\omega_{\text{i}})}{2\bar{L}_{\text{I}}c_{\text{I}}}\int_{\mathcal{H}(\omega_{\text{i}})} |\omega_{\text{i}} \cdot \mathbf{n}|_+/\pi\,\mathrm{d}\mathbf{n} = \frac{L_{\text{I}}(\omega_{\text{i}})}{2\bar{L}_{\text{I}}c_{\text{I}}}, \qquad (32)
\end{aligned}
$$

where we have used that the mean $\bar{L}_{\text{I}}$ of the tabulated illumination L_{I} is roughly equal to $1/(2\pi)\int_{\mathcal{H}(\mathbf{n})} L_{\text{I}}(\omega)\,\mathrm{d}\omega$. We can now put the above results together and we get

$$
\int_{\mathcal{H}(\omega_{\text{i}})} p_{\text{I}}^{\text{nd}}(\omega_{\text{i}}|\mathbf{n})\,\mathrm{d}\mathbf{n} \approx 1/b_{\text{nd}} \max\{0, \frac{L_{\text{I}}(\omega_{\text{i}})}{2\bar{L}_{\text{I}}c_{\text{I}}} - \frac{1 - c_{\text{I}}}{c_{\text{I}}}\}. \qquad (33)
$$

Finally, we can remove the factor $1/c_{\text{I}}$, multiply by $2\bar{L}_{\text{I}}$, and normalize the pdf to get the final result, Eq. (14).