Unbiased and consistent rendering using biased estimators

ZACKARY MISSO, Dartmouth College, USA
BENEDIKT BITTERLI, Dartmouth College, USA and NVIDIA, USA
ILIYAN GEORGIEV, Autodesk, United Kingdom
WOJCIECH JAROSZ, Dartmouth College, USA

We introduce a general framework for transforming biased estimators into unbiased and consistent estimators for the same quantity. We show how several existing unbiased and consistent estimation strategies in rendering are special cases of this framework, and are part of a broader debiasing principle. We provide a recipe for constructing estimators using our generalized framework and demonstrate its applicability by developing novel unbiased forms of transmittance estimation, photon mapping, and finite differences.

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

Additional Key Words and Phrases: Monte Carlo, infinite series, Taylor series

ACM Reference Format:
Zackary Misso, Benedikt Bitterli, Iliyan Georgiev, and Wojciech Jarosz. 2022. Unbiased and consistent rendering using biased estimators. *ACM Trans. Graph.* 41, 4, Article 48 (July 2022), 13 pages. https://doi.org/10.1145/3528223.3530160

1 INTRODUCTION

From estimating the amount of radiance reaching a camera sensor, to estimating how much light transmits through a participating medium, there are countless situations in graphics which require estimating intricate integrals. While we have developed a large arsenal of unbiased estimation techniques, situations still arise where we must fall back on biased formulations.

We consider problems where we need to compute some finite quantity I, but we only have a biased estimator $\langle I(k)\rangle$ with a *controllable* amount of bias—dictated by some parameter k—at our disposal. By adjusting the bias parameter towards some limit (e.g. $k \to \infty$) the estimator's expected value $I(k)$ approaches the correct answer:

$$I = \lim_{k\to\infty} I(k). \tag{1}$$

The bias parameter k could be continuous or discrete; for example, a discrete k could represent the maximum path length in a path tracer, while a continuous k could correspond to the step size in ray marching or the kernel radius in photon mapping. We assume $I(\infty)$ cannot be estimated directly e.g. due to computational constraints, and only biased estimates of I for finite values of k are available.

When faced with a situation like Eq. (1), one existing approach in graphics is to at least form a *consistent* estimator in which both bias and variance are guaranteed to vanish, though only in the limit of infinite work. This can be done for certain classes of problems, like progressive photon mapping approaches [Hachisuka et al. 2010; Hachisuka and Jensen 2009; Jarosz et al. 2011b; Knaus and Zwicker 2011], by averaging realizations of the biased estimate while carefully controlling the bias parameter k to ensure convergence.

In this paper we instead introduce and extend a recent set of techniques from the statistics literature for entirely *debiasing* biased estimators. Our framework can provide unbiased solutions to problems ranging from estimating transmittance in non-exponential media, to computing unbiased derivatives, to creating unbiased versions of entire global illumination algorithms like (progressive) photon mapping.

To debias a chosen problem, we decompose the unbiased quantity

$$I = I(k) + \underbrace{\sum_{j=k}^{\infty} \Delta_j}_{B(k)} \tag{2}$$

into a biased value $I(k)$ and an infinite sum representing the bias $B(k) := I - I(k)$. We then form unbiased estimators for both terms.

There are many different ways to transform a problem into an infinite series in the form of Eq. (2), all differing by their construction of Δ_j. In the following two sections we will discuss two such transformations and their utilization in prior work for isolated problems in graphics. The first one involves eliminating the bias via a Taylor series (Sec. 2), and the second does so via an initial-value integral or a corresponding discrete telescoping-series transformation (Sec. 3). In Sec. 4 we will discuss how to formulate estimators for Eq. (2) which encompasses both the Taylor and telescoping formulations.

By carefully constructing estimators so that $B(k)$ vanishes in the limit, we can formulate unbiased solutions for all problems in the form of Eq. (2). However, not all such solutions can be achieved with both finite variance and finite work, despite their guaranteed eventual convergence by the law of large numbers. For those cases, we show how to still formulate unbiased estimators (Sec. 4.4) and, alternatively, derive consistent progressive estimators (Sec. 4.4).

We provide a general recipe for creating unbiased estimates out of biased estimators (Recipe 1). In Sec. 5 we show how our recipe can be applied to estimate a variety of problems, including novel unbiased non-exponential transmittance [Bitterli et al. 2018] estimation, unbiased variants of (progressive) photon-mapping, and unbiased (for smooth functions) and progressive (for discontinuous functions) finite difference estimators that can be used as ground truth when developing differentiable rendering techniques. We provide full implementations on GitHub [Misso et al. 2022].

Authors' addresses: Zackary Misso, zackary.t.misso.gr@dartmouth.edu, Dartmouth College, USA; Benedikt Bitterli, benedikt.bitterli@gmail.com, Dartmouth College, USA and NVIDIA, USA; Iliyan Georgiev, iliyan.georgiev@autodesk.com, Autodesk, United Kingdom; Wojciech Jarosz, wojciech.k.jarosz@dartmouth.edu, Dartmouth College, USA.

2 TAYLOR-SERIES DEBIASING

One common problem that arises in graphics is the need to form estimators for integrals F which are nested inside some non-linear function g:

$$I \coloneqq g(F) = g\left(\int_\Omega f(x)\,\mathrm{d}x\right). \tag{3}$$

For instance, g could be the reciprocal function when estimating a normalization factor, or the exponential function when estimating volumetric transmittance. Naively applying Monte Carlo estimation to the integral F will generally yield a biased estimate if the function g is non-affine. Outside of graphics, these types of problems are referred to as functions of expectations [Blanchet et al. 2015].

Assuming $g(F)$ is finite and has a convergent Taylor series about some expansion point α, we can express Eq. (3) as:

$$I \coloneqq g(F) = \underbrace{\sum_{j=0}^{k-1}\Delta_j^{\text{tay}}}_{I(k)} + \underbrace{\sum_{j=k}^{\infty}\Delta_j^{\text{tay}}}_{B(k)}, \text{ with } \Delta_j^{\text{tay}} = g^{(j)}(\alpha)\frac{(F-\alpha)^j}{j!}, \tag{4}$$

where $g^{(j)}(\alpha)$ denotes the j^{th} derivative of g evaluated at α. For any fixed bias parameter k, this decomposes I into a biased formulation $I(k)$ consisting of the first k terms of the Taylor series, and the remaining terms of the Taylor series constituting the bias $B(k)$. For instance, when $k = 1$, $I(k)$ becomes just the evaluation of g at the expansion point, $g(\alpha)$, which will generally be a biased evaluation of I unless we happen to choose an expansion point, $\alpha = F$, equal to the integral we want to estimate. The infinite sum then corrects for this bias (see Fig. 1).

The Taylor expansion effectively separates the evaluation of g from the eventual estimation of F, thus removing the problematic non-linearities. While $(F-\alpha)^j$ is still a non-linear operation, we can estimate it in an unbiased manner by treating it as the product of j independent evaluations of $(F-\alpha)$ [Blanchet et al. 2015; Georgiev et al. 2019].

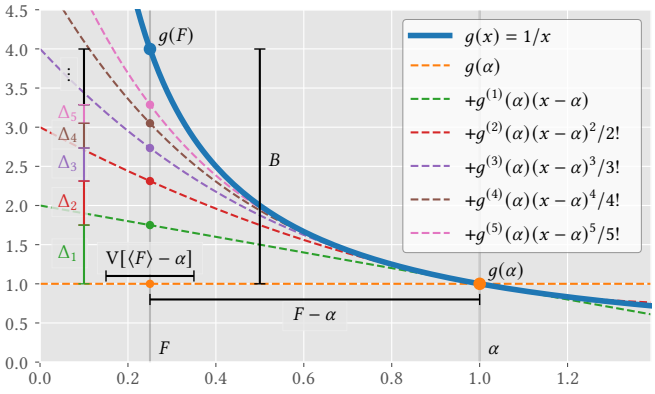


Fig. 1. Equation (4) enables the unbiased estimation of $g(F)$ using only unbiased estimates $\langle F\rangle$ of F and the analytic derivatives of g. Starting with $I(1) = g(\alpha)$, it progressively adds terms (vertical colored bars labeled Δ_j on the left) of a Taylor series expansion of $g(x)$ (dashed curves), each evaluated at a stochastic distance $\langle F\rangle - \alpha$.

2.1 Prior work

The Taylor series has not yet been introduced to graphics as a general debiasing technique; however, formulations like Eq. (4) have appeared for problems like transmittance and reciprocal estimation, though they initially arose via other mathematical manipulations. We will show how these specific instances can be easily derived directly from Eq. (4).

Reciprocal estimation. We sometimes need to estimate the reciprocal of an integral,

$$g(F) \coloneqq \frac{1}{F}. \tag{5}$$

For instance, this happens in both photon mapping [Qin et al. 2015] and differentiable rendering [Bangaru et al. 2020] where F is a Monte Carlo sampling probability (density) that is not available in closed form, so its reciprocal must be estimated numerically. Applying Eq. (4) to Eq. (5), choosing $k = 1$, and performing some elementary simplifications provides

$$g(F) = \frac{1}{\alpha} + \sum_{j=1}^{\infty}\Delta_j^{\text{recip}}, \quad \text{where} \quad \Delta_j^{\text{recip}} = \alpha^{-j-1}(\alpha - F)^j, \tag{6}$$

which can be readily estimated in an unbiased way provided $\alpha \neq 0$.

Unbiased reciprocal estimation was first introduced to graphics by Qin et al. [2015] where they imported the technique directly from the nuclear engineering literature [Booth 2007]. Booth originally derived his formulation by applying control variates to Eq. (5), then effectively taking a Maclaurin series expansion (Taylor series about $\alpha = 0$) by pattern matching with the well-known analytical solution of the geometric series.

Applying control variates then taking a Maclaurin series expansion is equivalent to taking the Taylor series expansion where α is set to the control. Thus, our formulation (6) generalizes the control function Booth introduced in his derivation. The Bernoulli trial estimation technique later introduced by Qin et al. [2015]—which has also been used for specular manifold sampling [Zeltner et al. 2020] and path connections in refractive media [Pediredla et al. 2020]—is a special case of a prefix sum estimator (Sec. 4) for Eq. (6) when estimating a reciprocal probability (density). We show that it is also a special case of the pseries-cumulative estimator [Georgiev et al. 2019] adapted for reciprocal estimation within our full implementation [Misso et al. 2022].

Exponential-transmittance estimation. Estimating the proportion of light which passes through a participating medium is a core component in simulating volumetric transport. The classical transmittance is defined as

$$g(F) \coloneqq \mathrm{e}^{-F}, \tag{7}$$

where $F = \tau \coloneqq \int_a^b \mu(x)\,\mathrm{d}x$ is the optical depth integral over the medium extinction.

Applying Eq. (4) to (7), setting $k = 1$ and performing some elementary simplifications allows us to express transmittance as

$$g(F) = \mathrm{e}^{-\alpha} + \sum_{j=1}^{\infty}\Delta_j^{\text{exp}}, \quad \text{where} \quad \Delta_j^{\text{exp}} = \mathrm{e}^{-\alpha}\frac{(\alpha - F)^j}{j!}, \tag{8}$$

which is amenable to unbiased estimation.

Both Georgiev et al. [2019] and Jonsson et al. [2020] introduced formulations for exponential transmittance equivalent to Eq. (8), however, both of these were obtained in a similar fashion to Booth's [2007] derivations for the reciprocal: they first applied control variates directly to Eq. (7) to introduce a majorant, and then applied a power (Maclaurin) series expansion to the resulting exponential. One of the major insights from Georgiev et al.'s derivations was that the majorant used by all prior null-collision based tracking estimators corresponds to a simple application of control variates.

In contrast, we have arrived at the same formulation as a direct consequence of applying the Taylor series at a user-chosen expansion point α. Our derivation shows that the majorant in null-scattering methods is simply the Taylor series expansion point.

Kettunen et al. [2021] developed a prefix-sum based estimator for classical transmittance which exploits the expansion point used in the Taylor expansion of the exponential function along with a few other techniques for variance reduction. Their final estimator had a form similar to Eq. (4). They pointed out that if the first term in their formulation was separated from the infinite sum (i.e. $k = 1$), then their estimator debiases an instance of ray-marching. Our derivation shows that *all* prior unbiased transmittance estimators which are derived from a Taylor series expansion can be thought of as a form of debiased estimation. We will discuss why this is also true for the Volterra based estimators [Georgiev et al. 2019] in Sec. 3.3.

Outside of graphics. Blanchet et al. [2015] introduced the idea of the Taylor series as a generalized means of formulating unbiased estimators for functions of expectations. Dauchet et al. [2018] mentions this idea too in the context of handling non-linearities with Monte Carlo. Outside of these instances, the Taylor expansion has been utilized throughout statistics to reduce bias of general estimators of differentiable functions [Jiao and Han 2020; Withers 1987].

3 INITIAL-VALUE & TELESCOPING-SERIES DEBIASING

While the Taylor expansion is effective in situations where we deal with smooth functions of expectations, unfortunately not all problems can be represented this way. Either the function does not have a convergent Taylor series, or a problem may simply not fit within the narrow scope of a function of expectations (3), even if it is an instance of Eq. (1). To handle this, we first derive a continuous formulation, and then show how once it is discretized, it simplifies to a debiasing technique developed recently within the statistics literature.

3.1 Continuous formulation

For any k, we can define the relationship between the limit I and the expectation $I(k)$ of a biased estimate by manipulating Eq. (2),

$$I = I(k) + B(k) = I(k) + I(\infty) - I(k). \tag{9}$$

Assuming that k is a continuous parameter and that the derivative of I w.r.t. k exists, we can rewrite Eq. (9) as an initial value problem:

$$I = I(k) + \int_k^\infty \frac{\mathrm{d}}{\mathrm{d}x} I(x)\,\mathrm{d}x. \tag{10}$$

This formulation is a continuous version of Eq. (2) for debiasing a biased quantity, by integrating how the bias changes as we change the variable k dictating the bias.

3.2 Discrete formulation

Many of the problems defined by Eq. (1) involve k's which are discrete instead of continuous. We can obtain a discrete analog to Eq. (10) by splitting the integral into a sum of unit-length integrals,

$$I = I(k) + \sum_{j=k}^{\infty} \int_j^{j+1} \frac{\mathrm{d}}{\mathrm{d}x} I(x)\,\mathrm{d}x. \tag{11}$$

By replacing the integral over the derivative with its anti-derivative, we arrive at

$$I = I(k) + \sum_{j=k}^{\infty} \Delta_j^{\text{tele}}, \quad \text{where} \quad \Delta_j^{\text{tele}} = I(j+1) - I(j). \tag{12}$$

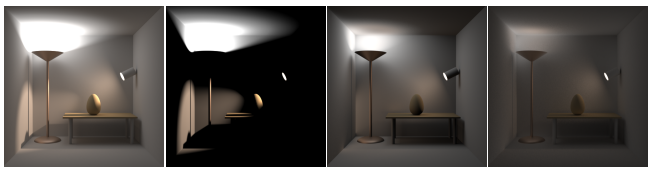
This is an infinite telescoping sum of readily estimatable expectations, similar to Eq. (4). Note, that any finite sum up to k of Δ_j^{tele}, simplifies to $I(k)$.

3.3 Prior work

Path tracing. Path tracing computes an unbiased estimate of an image by accumulating the contribution from all length paths. If we define $I(j+1) - I(j)$ to represent the difference between $(j+1)$-length transport and j-length transport, i.e. the contribution of only the $(j+1)$-st bounce (see Fig. 2), we can interpret path tracing as an instance of a telescoping sum (12) which compensates for the bias of direct illumination by evaluating each additional path length up to infinity. We further show that Russian roulette termination in path tracing becomes a single-term estimation strategy (Sec. 4) applied to Eq. (12) in the case of naive path-tracing with no next-event estimation and no scattering on emissive surfaces. Otherwise, it becomes an instance of a prefix-sum based estimation strategy.

Volterra transmittance. Georgiev et al. [2019] introduced an integral formulation for estimating transmittance in a classical medium which they derived by starting from a reduced form of the differential radiative transfer equation.

If we define $I := \lim_{k\to b} I(k)$ where b is the distance to the end of a path segment, and we plug in the definition of exponential



$\sum_{j=0}^{20}\left[I(j+1) - I(j)\right]$ $\quad I(1) - I(0)$ $\quad I(2) - I(1)$ $\quad I(3) - I(2)$

Fig. 2. Illustration of path tracing as an instance of Eq. (12). Each term in the summation is the difference in contribution between $(j+1)$- and j-length paths, simplifying to a sum over the contribution of various bounces.

transmittance $I = \mathrm{Tr}(k,b) := e^{-\int_k^b \mu(y)\,\mathrm{d}y}$ into Eq. (10), we obtain

$$\begin{aligned}\mathrm{Tr}(k,b) &= \mathrm{Tr}(k,k) + \int_k^b \frac{\mathrm{d}}{\mathrm{d}x}\,\mathrm{Tr}(x,b)\,\mathrm{d}x \\ &= 1 - \int_k^b \mu(x)\,\mathrm{Tr}(x,b)\,\mathrm{d}x.\end{aligned} \tag{13}$$

This is the same as Georgiev et al.'s [2019] Volterra formulation, except instead of starting from the radiative transfer equation, which is not easily generalized to non-exponential transmittance functions, we derived via Eq. (10). This representation was used by Vicini et al. [2021] to formulate a non-exponential transmittance model which allows for learning improved volumetric scene representations.

Outside of classical transmittance estimation, we found the application of Eq. (10) to perform worse than Eq. (12) due to the derivatives in the continuous case having unbounded variance. For this reason, we will choose to focus on the discrete formulation instead, however, everything we will cover generalizes nicely to the continuous case as well.

Reciprocal estimation. Bangaru et al. [2020] introduced an unbiased differentiable rendering algorithm which utilized the divergence theorem to handle discontinuous integrands. Their formulations introduced a discontinuous warp field which they make continuous through convolution with a smoothing kernel whose normalization factor is scene-dependent and must be estimated.

This amounts to formulating unbiased estimates for the reciprocal of a random variable for which either Eq. (4) or Eq. (12) could be used. Bangaru et al. [2020] utilized the estimator proposed by McLeish [2011] which is an instance of Eq. (12).

Outside of graphics. Within the statistical literature, McLeish [2011] and Rhee and Glynn [2012] independently introduced the first estimation schemes for Eq. (12). Rhee and Glynn focused on the specific problem of creating unbiased estimators for stochastic differential equations (SDEs) by adding randomized truncation (akin to Russian roulette) to multi-level Monte Carlo [Giles 2008; Heinrich 1998; Keller 2001]. McLeish [2011] introduced a more general estimator using a prefix-sum that we'll discuss in Sec. 4.

Rhee [2013] analyzed the optimal design decisions in creating estimators for Eq. (12) and later Blanchet et al. [2015] equated this with the optimal design decisions for the Taylor expansion approach in Eq. (4). Rhee and Glynn [2015] later used these findings to introduce practical estimators for SDEs while Blanchet et al. [2015] introduced a variance reduction technique to improve the convergence rate of a certain class of estimators. More recently, Moka et al. [2019] has looked into the special case of formulating non-negative estimators for the reciprocal.

4 ESTIMATION AND CONVERGENCE

Up until now, all our formulations have been defined with respect to expected values, but to use them in practice we need to turn them into estimators. In this section we develop a recipe (Recipe 1) for creating debiased and consistent estimators for any problem in the form of Eq. (2). Assuming we have already chosen a valid representation for Δ_j by using Eq. (4) or Eq. (12), all that remains to complete step ① is to construct a primary and secondary estimator.

Recipe 1. A recipe for creating debiased and consistent estimators for problems in the form of Eq. (2).

① **Construct estimators:** Choose Δ_j representation respecting Eq. (1), build primary (Sec. 4.1) and secondary (Sec. 4.2) estimators.
② **Determine rates:** Determine asymptotic rates of Δ_j, variance $\mathrm{V}[\langle\Delta_j\rangle]$, and cost $\mathrm{C}[\langle\Delta_j\rangle]$, as $j \to \infty$.
③ **Derive PMF:** Insert rates into Eq. (18); check if resulting PMF $p(j)$ can be normalized. If not, skip to step ⑤.
④ **Verify finite work-normalized variance:** Insert $p(j)$ into estimator $\langle I\rangle_1$ and confirm the work $\mathrm{C}[\langle I\rangle_1]$ is finite. If yes, work-normalized variance is finite, done.
⑤ **Build unbiased infinite-variance estimator:** Fix $p(j)$ to something reasonably normalizable, set the growth rate of $\mathrm{C}[\langle k_j\rangle]$ sufficiently high while still remaining finite to decrease the rate at which $\mathrm{V}[\langle k_j\rangle]$ approaches infinity. Evaluate secondary estimator. If noise is acceptable, done.
⑥ **Build consistent progressive estimator:** Convert the secondary estimator into a consistent progressive estimator, and reparameterize Δ_k such that $\mathrm{V}[\langle\Delta_k\rangle]$ behaves asymptotically strictly sublinearly w.r.t. k, done.

4.1 Primary estimators

Based off the nomenclature introduced by Georgiev et al. [2019], we consider two approaches for building primary estimators for Eq. (2).

Single-term estimator. One approach is to randomly choose one term j of the bias sum (2) according to some probability mass function (PMF) $p(j)$, then combine it with the biased estimate $\langle I(k)\rangle$:

$$\langle I\rangle_1^k = \langle I(k)\rangle + \langle B(k)\rangle = \langle I(k)\rangle + \frac{\langle\Delta_j\rangle}{p(j)}. \tag{14}$$

Prefix-sum estimator. Alternatively, having sampled a term j from $p(j)$ we could evaluate the prefix of all terms up to j:

$$\langle I\rangle_1^k = \langle I(k)\rangle + \langle B(k)\rangle = \langle I(k)\rangle + \sum_{i=k}^{j} \frac{\langle\Delta_i\rangle}{P(j \geq i)}, \tag{15}$$

where $P(j \geq i)$ is the cumulative mass function (CMF) giving the probability for sampling $j \geq i$.

Note that we could create analogous estimators for the continuous formulation (10) by making $p(j)$ a probability density function, and using a "prefix integral" instead of a sum in Eq. (15).

4.2 Secondary estimator

To ensure that our approach converges to the sought value I, we form a secondary estimator,

$$\langle I\rangle_N = \frac{1}{N}\sum_{i=1}^{N} \langle I\rangle_1^{k_i}, \tag{16}$$

which averages N independent instances of the primary estimator. Typically $k_i = k$ will be a constant, meaning the secondary estimator Eq. (16) will take the average of N independent and identical evaluations of the primary estimator. However, no matter how k_i is assigned, Eq. (16) will always be unbiased because the primary estimator is unbiased for any starting k. For example, an unbiased

progressive estimator could be constructed from Eq. (16) by setting $k_i = i$. This completes step ①.

4.3 Estimation under finite variance

Steps ②, ③, and ④ involve determining if we can formulate a primary estimator with finite variance, and if so, how to do so as optimally as practically possible.

Since our primary estimators have the correct expected value by construction, then as long as they have finite variance, the central limit theorem tells us that the secondary estimator $\langle I \rangle_N$ will converge to I at an asymptotic rate of $\mathrm{V}[\langle I \rangle_N] = O(1/N)$. The constant factor in this asymptotic rate, however, will depend heavily on the construction of $\langle \Delta_j \rangle$ and the choice of $p(j)$. When a primary estimator has finite variance, the ideal strategy would be to choose $\langle \Delta_j \rangle$ and $p(j)$ to minimize its work-normalized variance:

$$\arg\min_{p(j)} \mathrm{V}[\langle I \rangle_1]\, \mathrm{C}[\langle I \rangle_1], \tag{17}$$

where $\mathrm{C}[\langle I \rangle_1]$ is its expected cost. Rhee [2013] derived expressions for the work-normalized variance optimal $p(j)$. Unfortunately, this is primarily of theoretical interest since it requires knowledge of not only $\mathrm{V}[\langle \Delta_j \rangle / p(j)]$ and $\mathrm{C}[\langle \Delta_j \rangle / p(j)]$ (which are unfortunately not available for most problems in graphics), it also requires knowledge of I, the very quantity we are trying to estimate.

A more pragmatic approach would be to derive a $p(j)$ which simply minimizes variance, $\mathrm{V}[\langle \Delta_j \rangle / p(j)]$, and then check whether the resulting estimator also has finite work-normalized variance. In Appendix A, we prove that for both continuous and discrete forms of single-term estimation (14), the variance-optimal $p(j)$ is

$$p(j) \propto \sqrt{\mathrm{V}[\langle \Delta_j \rangle] + (\Delta_j)^2}. \tag{18}$$

This matches the variance-optimal $p(j)$ that Rhee [2013] derived for the case of the discrete telescoping series (12). Later, Blanchet et al. [2015] showed that this $p(j)$ also applies to the general Taylor series formulation (4).

Compared to Eq. (17), it is more feasible to apply Eq. (18) since even though we may not know $\mathrm{V}[\langle \Delta_j \rangle]$ or Δ_j exactly, we can often obtain their asymptotic rates ② and use them to derive ③ an asymptotically optimal $p(j)$. If $p(j)$ is not normalizable then our estimator will not have finite work-normalized variance. If $p(j)$ is normalizable, we need to additionally verify ④ that $\mathrm{C}[\langle \Delta_j \rangle / p(j)]$ is finite. If not, then modifying the estimator to have finite work for practical implementations might still result in infinite variance.

4.4 Estimation under infinite variance

Since our primary estimators have the correct expected value by construction, the weak law of large numbers guarantees that the secondary estimator (16) will converge in probability to I as N increases. Remarkably, this is the case even if the variance of the primary estimator is infinite, as we will see for the case of debiased photon mapping; however, the convergence rate may be slower than $O(1/N)$ since the central limit theorem does not apply. We demonstrate this in Fig. 3 for the case of estimating the integral of a simple function with a controllable amount of variance. Even in cases where the primary estimator has infinite variance, a secondary estimator will converge in expectation, but at a slower rate compared to when variance is finite.

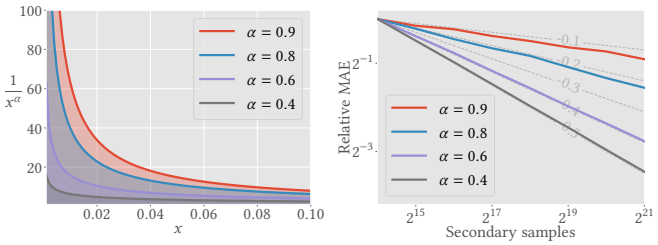


Fig. 3. The integral $\int_0^1 \frac{1}{x^\alpha}\,\mathrm{d}x$ (left) provides a useful test case for finite vs. infinite variance primary estimators. The integral takes on a closed-form value of $(1-\alpha)^{-1}$ for any $\alpha \in (0,1)$, but the variance of a primary estimator using uniform sampling is finite only for $\alpha < 0.5$. For $\alpha \geq 0.5$ the variance is infinite, leading to slower convergence of a secondary estimator (right).

We can choose to formulate either an unbiased estimator ⑤, or formulate a consistent estimator ⑥. The consistent estimator will be biased, but the unbiased estimator will exhibit higher variance.

Unbiased estimation. When utilizing an unbiased primary estimator in the case of infinite variance, the convergence rate is heavily influenced by the rate at which $\mathrm{V}[\langle \Delta_j \rangle / p(j)]$ approaches infinity. We could create practical estimators by choosing the estimators' parameters such that $\mathrm{V}[\langle \Delta_j \rangle / p(j)]$ approaches infinity as slowly as reasonably possible.

Selecting an optimal parameter configuration for such an estimator may be challenging, so we instead take a more pragmatic approach: For many applications $\mathrm{V}[\langle \Delta_j \rangle]$ is inversely related with $\mathrm{C}[\langle \Delta_j \rangle]$; for example, in photon mapping, tracing more photons reduces its variance. By increasing the rate at which $\mathrm{C}[\langle \Delta_j \rangle]$ grows with j, we also reduce the rate at which $\mathrm{V}[\langle \Delta_j \rangle]$ increases with j. This way we may achieve workable convergence rates for debiased estimators while still maintaining finite cost $\mathrm{C}[\langle I \rangle_1]$.

Consistent estimation. Although debiasing is still possible in the case of infinite variance, the noise which arises from unbiased estimators with slower convergence may not be acceptable in all applications. In these cases, we can formulate an alternative consistent progressive estimator, similar to progressive photon mapping, which trades noise for bias. In Appendix E, we show that a consistent progressive estimator can be derived directly from Eq. (16) by choosing $k_i = i$ and never evaluating $B(k)$.

This consistent progressive formulation is identical to the one used by progressive photon mapping, so we can utilize the analysis of prior work to construct efficient consistent estimators. First, we take our estimator from ⑤ and reparameterize k_i such that the primary estimators' variance $V(k_i) = \mathrm{V}[\langle I(k_i) \rangle]$ behaves asymptotically strictly sublinearly w.r.t. i. That is, $V(k_i)$ is allowed to increase, but not faster than i itself, which ensures that the variance of $\langle I \rangle_N^{\mathrm{c}}$ vanishes. The parameterization that maximizes that estimator's error convergence rate is the one that achieves $O(V(k_i)) = kO\big(B^2(k_i)\big)$.

Once we have constructed a consistent progressive estimator, we have completed ⑥, and now we can apply our recipe to formulate unbiased estimators for novel problems.

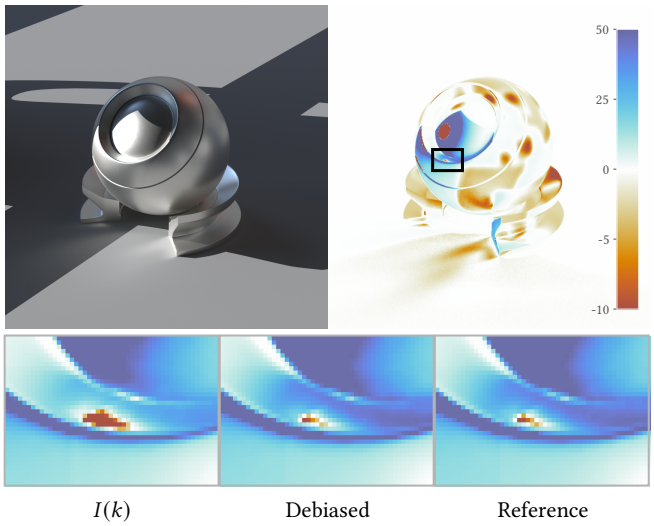


Fig. 4. Debiased finite differences with respect to the roughness of a metallic object in a scene where the main light source is partially obstructed. Our method debiases the initial $I(k)\,(h = 0.005)$ to match the reference (right) which was rendered using $(h = 0.0001)$.

5 APPLICATIONS

In this section, we show applications of our theory to common problems in graphics. We show several novel applications, namely unbiased/progressive finite differences (Sec. 5.1), unbiased non-exponential transmittance (Sec. 5.2), as well as unbiased photon mapping and unbiased progressive photon mapping (Sec. 5.3). We implemented all methods in the PBRT renderer [Pharr et al. 2016].

5.1 Finite differences

As an example of directly applying Recipe 1 to estimate a novel problem, consider computing the derivative of an integral $F(x) = \int_\Omega f(x,y)\,\mathrm{d}y$ with respect to the parameter x:

$$I := \frac{\partial F(x)}{\partial x} = \frac{\partial}{\partial x}\int_\Omega f(x,y)\,\mathrm{d}y. \tag{19}$$

For example, the function $f(x,y)$ could represent the contribution of a light path y in a scene parameterized by x. Note that the domain of integration Ω may depend on x.

Devising ways to estimate Eq. (19) efficiently and accurately is the main goal of differentiable rendering. Even though finite differences are not a practical approach for estimating high-dimensional derivatives in an inverse rendering context, they remain a common baseline for validating new differentiable rendering techniques, thanks to their generality and ease of implementation. Unfortunately, since finite differences are biased, such "ground truth"-baseline comparisons remain imperfect. Following all steps outlined in Recipe 1, we will derive both unbiased and consistent alternatives to finite differences, which could serve as more accurate ground-truth techniques for validation of future differentiable rendering methods.

We first denote the (biased) finite difference approximation of Eq. (19) as $I(k)$,

$$\frac{\partial F(x)}{\partial x} \approx I(k) := \frac{1}{h_k}\left(\int_{\Omega_1} f(x+h_k, y_1)\,\mathrm{d}y_1 - \int_{\Omega_2} f(x,y_2)\,\mathrm{d}y_2\right), \tag{20}$$

where k maps to a finite-difference step size h_k, and $\Omega_1 \neq \Omega_2$ if the integration domain depends on x. By the definition of a derivative, this formulation satisfies Eq. (1) as long as $h_k \to 0$ as $k \to \infty$; we use $h_k \propto 2^{-k}$.

We can now formulate unbiased finite differences using our telescoping-series formulation (12) as

$$I = I(k) + \underbrace{\sum_{j=k}^{\infty} \Delta_j^{\mathrm{FD}}}_{B(k)}, \quad \text{where} \quad \Delta_j^{\mathrm{FD}} = I(j+1) - I(j). \tag{21}$$

We use standard Monte Carlo rendering to form $\langle I(j)\rangle$ where we choose to maximally correlate the two integral estimates, and we use a single-term estimate of Eq. (21). This completes step ①.

Unbiased estimation. For the cases where $f(x,y)$ is finite, continuously differentiable with respect to x, and Ω does not depend on x, we show in Appendix D.1 how to form an estimator $\langle \Delta_j^{\mathrm{FD}} \rangle$ with finite variance and derive the asymptotic rates for Δ_j and $\mathrm{V}[\langle \Delta_j \rangle]$, completing step ②. Inserting these formulations into Eq. (18), we arrive at the variance optimal $p(j) \propto 2^{-j}$ to utilize in our single-term primary estimator, completing step ③. Since the cost of evaluating finite differences is constant, we have arrived at an unbiased estimator with finite work-normalized variance ④. Figure 4 shows an example of differentiating with respect to material roughness. Since Ω is not dependent on x for this case, we choose to fully correlate paths by always performing BSDF importance sampling using the same BSDF and evaluate all integral estimates simultaneously.

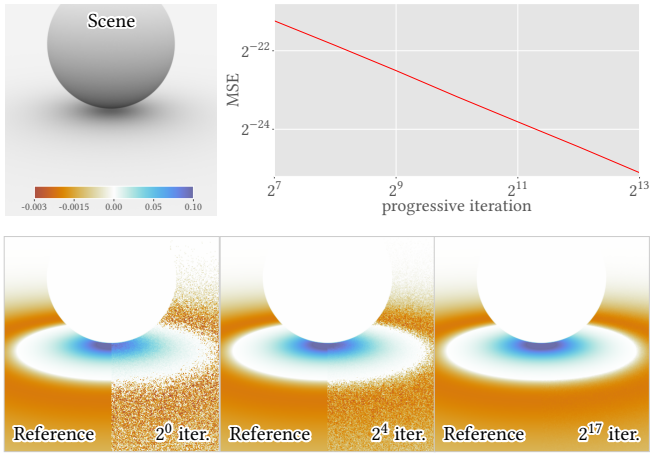


Fig. 5. We show the convergence of our progressive finite differences for evaluating the derivative of ambient occlusion on a plane by a sphere as the sphere moves upwards, whose analytic ground truth derivative is known.

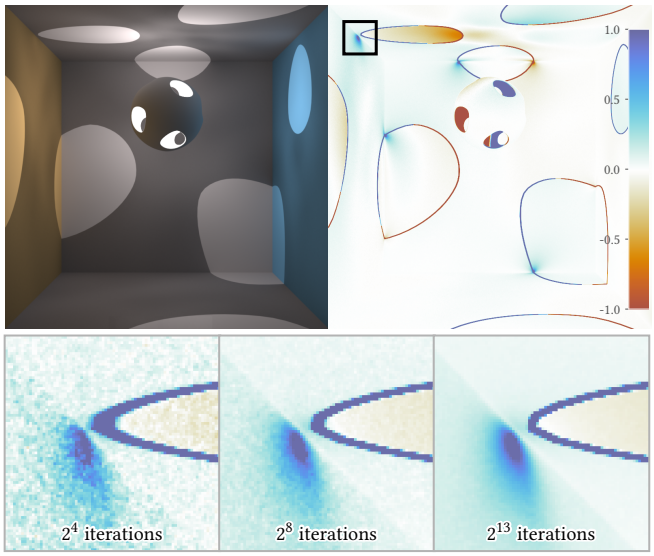


Fig. 6. Convergence of progressive finite differences w.r.t. moving a point light source to the right in a scene with visibility discontinuities. Each progressive iteration uses 16 samples/pixel.

Consistent estimation. If $f(x, y)$ is not continuous with respect to x (for example, when x impacts visibility), then the variance of the primary estimator, $\mathrm{V}[\langle I\rangle_1^k]$, can become infinite. For this case, we skip to step ⑤. Through experimentation we found the unbiased estimators for the infinite variance case to be too noisy in practice, so instead we continue to step ⑥ and formulate a consistent estimator. For the case where Ω does not depend on x, we show in Appendix D.2 that finite differences are biased because they blur the true derivatives with a box kernel. This allows us to draw an analogy to progressive photon beams to form a progressive finite difference estimator with the same theoretical convergence guarantees for the ideal case. The estimators will still converge for the discontinuous case, however their convergence rates will be slower. In Fig. 5, we verify that this progressive estimator is consistent on an ambient occlusion scene with a close-form ground truth; in Fig. 6, we show a more complex example in a scene with obvious discontinuities, showing how progressive iterations of our estimator reduce the blurring of finite differences in much the same way as in progressive photon beams.

5.2 Non-exponential transmittance

At its core, computing transmittance in volumetric rendering involves estimating a function of expectations (3) $I := g(F)$ where F is the optical depth integral along a ray. Estimating F numerically in heterogeneous media leads to bias when passed through the transmittance function g. In classical radiative transport, the transmittance is simply the exponential, $g(F) := \mathrm{e}^{-F}$, but recent non-exponential formulations [Bitterli et al. 2018; Jarabo et al. 2018] allow for using any monotonically decreasing function g that starts at $g(0) = 1$.

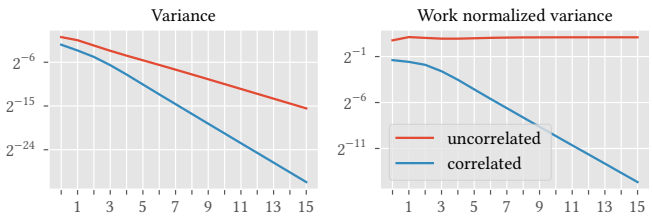


Fig. 7. Correlating samples when estimating $\langle \Delta_j^{\mathrm{RM}}\rangle$ can have a dramatic impact on convergence. Here $I := \mathrm{e}^{\int -f(x)\,\mathrm{d}x}$ where $f(x)$ is some random noise and the integral is estimated with a 2^j-sample Monte Carlo estimator. The graph in red shows the variance (left) and work-normalized variance (right) with uncorrelated evaluation $\langle \Delta_j\rangle = \langle I(j+1)\rangle - \langle I(j)\rangle$, while in blue we show the improvement when evaluating $\langle \Delta_j\rangle = \langle I(j+1) - I(j)\rangle$ using Blanchet and Glynn's [2015] technique for correlating samples.

Unfortunately, estimation of such non-exponential transmittance still lags behind traditional transport because most available techniques [Georgiev et al. 2019; Kettunen et al. 2021; Kutz et al. 2017; Novák et al. 2014] rely on the exponential assumption. This leaves only raymarching (biased) or regular tracking (slow) as possible techniques for general forms of transmittance g. Luckily, since transmittance is a function of expectations, it is readily debiasable using our framework with both Eqs. (4) and (12).

Telescoping series for general transmittance. The telescoping series formation (12) allows us—for the first time—to form unbiased estimators for general transmittance functions as

$$I := g(F) = \underbrace{I(k)}_{\text{biased transmittance estimate}} + \overbrace{\sum_{j=k}^{\infty} \Delta_j^{\mathrm{RM}}}^{\text{bias } B(k)}, \quad \text{with} \quad \Delta_j^{\mathrm{RM}} = I(j+1) - I(j). \tag{22}$$

We use $I(k)$ to denote the biased evaluation of transmittance g where we use ray marching with a monotonically decreasing sequence of step sizes to estimate F (though any numerical estimation technique for F is admissible as long as it ensures consistency (1) in the limit as $k \to \infty$).

We form a single-term estimator $\langle I\rangle_1^k$ of Eq. (22) using a geometric PMF $p(j) = r(1-r)^j$ with $r = 0.65$, as recommended by Blanchet and Glynn [2015] for applying Eq. (12) to smooth functions of expectations. We set the ray marching step size to $s_j \propto 2^{-j}$. Doubling the number of ray marching steps this way allows for dramatic sample reuse and variance reduction when estimating $\langle \Delta_j^{\mathrm{RM}}\rangle$ since we can evaluate both $\langle I(j+1)\rangle$ and $\langle I(j)\rangle$ *simultaneously with a shared set of ray marching samples.* In fact, we can reduce variance further [Blanchet and Glynn 2015] by splitting the N samples from $\langle I(j+1)\rangle$ into two groups and averaging two separate estimators for $\langle I(j)\rangle$ with $N/2$ samples each. We show in Fig. 7 that these measures have a profound impact on not just the variance $\mathrm{V}[\langle \Delta_j^{\mathrm{RM}}\rangle]$, but also its asymptotic convergence rate.

In Fig. 8 we demonstrate this estimator on the WDAS cloud scene using the two-parameter Davis and Mineev-Weinstein transmittance

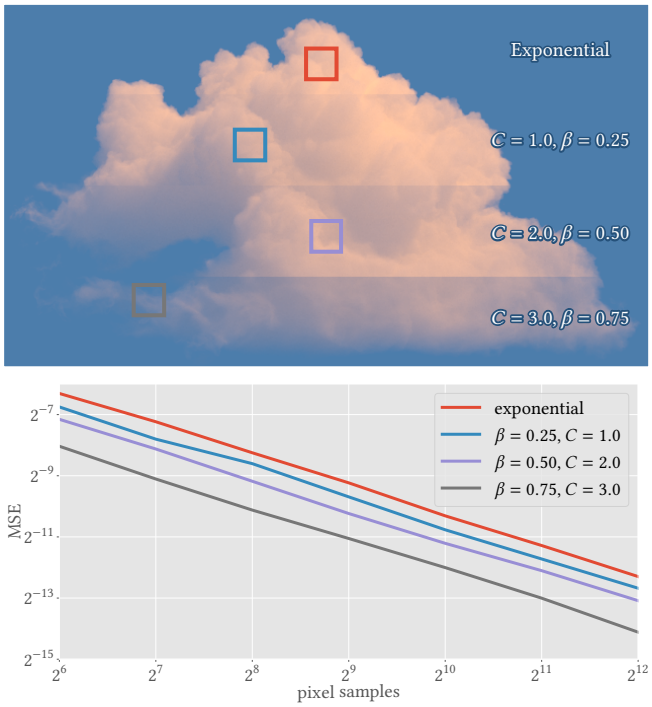


Fig. 8. Illustration of our telescoping-based debiased ray-marching (22). The split image shows classical transmittance as well as non-exponential transmittance with three different choices for C and β. The plot shows the convergence of four select pixels towards a regular-tracked ground truth.

model,

$$g(F) \coloneqq \left(1 + F^{\beta} C^{1+\beta}\right)^{-\frac{F^{1-\beta}}{C^{1+\beta}}}, \tag{23}$$

for a variety of settings of β and C which together control the shape of the transmittance curve. We compare our results to regular tracking, confirming that our estimator is unbiased, and that the variance $\mathrm{V}[\langle I \rangle_N]$ falls off at the expected $O(1/N)$ convergence rate.

Taylor series for pink-noise transmittance. We can also apply the Taylor series formulation to Eq. (23). For arbitrary values of β, we found that the complexity of the derivative terms in the Taylor expansion grows exponentially for increasing values of j. However, for the special case of "pink-noise" ($\beta = 1$), Eq. (23) simplifies to

$$g(F) \coloneqq \left(1 + FC^2\right)^{-\frac{1}{C^2}}, \tag{24}$$

which admits a compact Taylor series formulation (4):

$$I \coloneqq g(F) = \underbrace{\sum_{j=0}^{k-1} \Delta_j^{\text{pink}}}_{I(k)} + \underbrace{\sum_{j=k}^{\infty} \Delta_j^{\text{pink}}}_{B(k)}, \qquad \text{where}$$
$$\Delta_j^{\text{pink}} = \frac{(\alpha - F)^j}{j!} \left(1 + \alpha C^2\right)^{\frac{-1}{C^2} - j} \prod_{i=1}^{j-1} \left(1 + iC^2\right). \tag{25}$$

We create a prefix-sum primary estimator of Eq. (25) by defining the PMF $p(j)$ indirectly via Russian roulette probabilities:

$$P_{\text{acc}}(j) = \frac{1 + (j-1)C^2}{j}, \quad p(j) = (1 - P_{\text{acc}}(j)) \prod_{i=0}^{j-1} P_{\text{acc}}(i). \tag{26}$$

Since this is independent of both the expansion point α and any optical depth evaluation, $\langle F \rangle$, we can take advantage of all the same variance reduction techniques that Kettunen et al. [2021] introduced for exponential transmittance. We use uniformly jittered regular samples with a fixed step-size (0.2) to estimate $\langle F \rangle$. We set $k = 3$, and choose to evaluate the bias only a portion ($P = 0.1$) of the time which reduces the cost while slightly increasing the variance of the estimator.

In Fig. 9 we compare both our Taylor series (25) and telescoping series (22) estimators against prior techniques for unbiased estimation of both pink-noise- and exponential-transmittance functions. Bitterli et al. [2018] showed that ratio tracking (or any existing exponential transmittance estimator) can be reused for pink noise by first sampling a multiplicative factor γ from a gamma distribution, and then estimating exponential transmittance in a medium with its density scaled by γ. We also apply this trick to Kettunen et al. [2021]'s more recent unbiased ray marching technique, to allow it to handle both exponential and pink-noise transmittance. As shown in Fig. 9, our estimators consistently outperform these prior methods, sometimes reducing RMSE by a factor of 10× at equal cost.

In our full implementation [Misso et al. 2022] we additionally provide a novel estimator akin to Georgiev et al.'s [2019] pseries-cumulative estimator but for pink-noise and include more comparisons in canonical pink-noise and exponential media for all our estimators.

5.3 Photon mapping

Photon mapping is the most practical algorithm for rendering complex caustic effects. It emits photons from light sources which are then used to compute the radiance at a point in the scene via density estimation. However, density estimation yields a result with bias dependent on the radius r of the density kernel used; the bias vanishes as the radius approaches zero. Note that while an unbiased formulation of photon mapping exists [Qin et al. 2015], it does not handle purely specular caustics.

We denote the (biased) expectation of the photon mapping radiance estimate as $I(k)$, where k maps to a monotonically decreasing sequence of radii r_k used for density estimation. We can create a telescoping-series formulation (12) for the problem as

$$I = I(k) + \underbrace{\sum_{j=k}^{\infty} \Delta_j^{\text{PM}}}_{B(k)}, \quad \text{where} \quad \Delta_j^{\text{PM}} = I(j+1) - I(j), \tag{27}$$

which fully debiases photon mapping, even in the presence of caustics. In essence, Eq. (27) represents the bias as consecutive differences of kernels with monotonically decreasing radii (Fig. 10).

By utilizing the known convergence rates of progressive photon mapping, we show in Appendix C that it is not possible for both

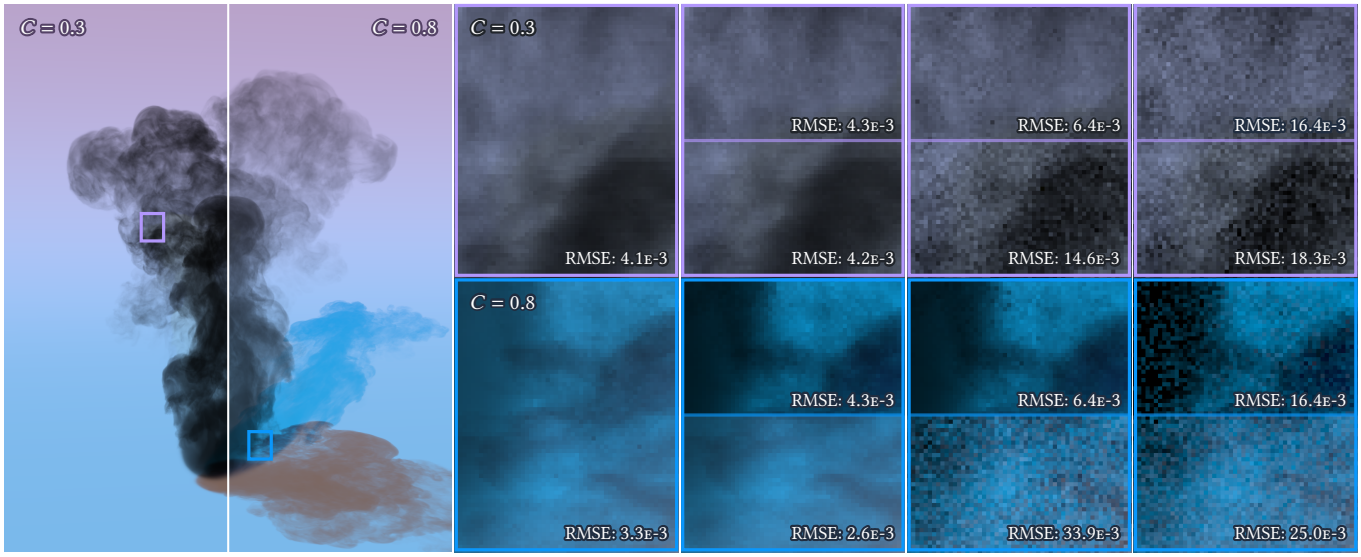


Fig. 9. Comparing a variety of transmittance estimators for a non-exponential medium with the Davis and Mineev-Weinstein transmittance model Eq. (23) for two different values of C for an equal number of extinction calls (1.2E8 for $C = 0.3$, 0.8, and 1.0E8 for exponential). The three estimators on the right can be used as estimators for exponential transmittance, so we also include inset comparisons for an exponential medium at the top of their respective insets. RMSE values are measured across the entire image.

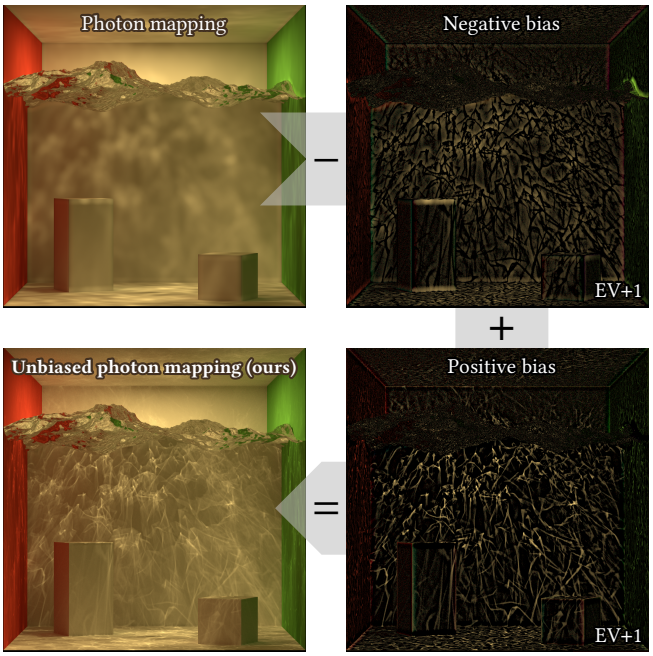


Fig. 10. Decomposition of our unbiased photon mapping into the sum of a biased estimate and a bias term. We separate the bias into positive and negative parts for ease of visualization.

③ and ④ to hold, i.e., we cannot debias Eq. (27) with finite work-normalized variance. However, unbiased secondary estimators still converge by the law of large numbers.

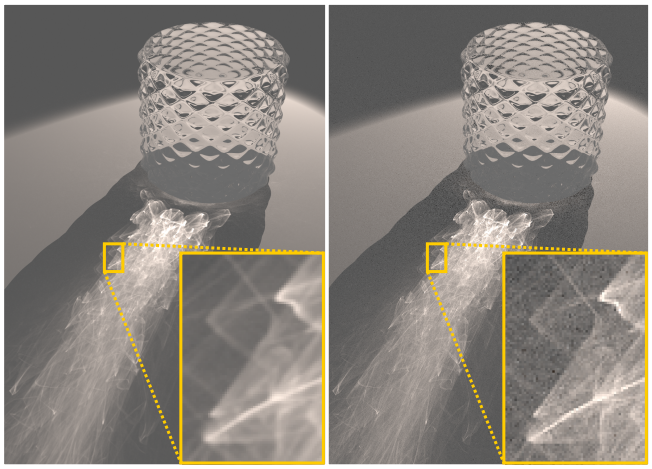
Unbiased photon mapping. We formulate unbiased photon mapping via a single-term primary estimator (14), and a secondary estimator (16) with $k_i = k$ for ⑤. We use a cone kernel and fully correlate all kernel evaluations, in essence by estimating $\langle \Delta_j^{\mathrm{PM}} \rangle := \langle I(j+1) - I(j) \rangle$ instead of $\langle I(j+1) \rangle - \langle I(j) \rangle$. We use $p(j) \propto \mathcal{O}(j^{\alpha-2})$, and set the number of global photons to $\mathcal{O}(j^{1-c\alpha})$ where $c = 1.001$ and $\alpha = 2/3$. Compared to prior progressive photon mapping methods, our debiased photon mapping trades bias (i.e., blurring) for noise (Fig. 11).

Unbiased progressive photon mapping. We can also debias progressive photon mapping by making a small change in the secondary estimator (16): setting $k_i = i$ instead of having it constant to complete ⑤. This is equivalent to progressive photon mapping, with an additional estimation of the bias of each progressive iteration. If we instead continued from ⑤ and performed ⑥, we would end up with prior work's consistent progressive photon mapping.

In Fig. 12, we compare convergence rates of our debiased methods and prior progressive photon mapping in terms of emitted photons.

6 CONCLUSION

We have introduced a general framework for formulating unbiased and consistent estimators from biased ones as well as provided a general recipe for applying our framework to novel problems. We have shown how specializations of these techniques have already appeared in graphics for reciprocal and classical transmittance estimation, and how to apply them to a broader class of problems. We have introduced unbiased progressive estimation and have shown how prior biased but consistent formulations, i.e. progressive photon mapping, can be obtained from our unbiased formulation as a means



Progressive photon mapping **Unbiased photon mapping (ours)**

Fig. 11. Compared to progressive photon mapping (with $\alpha = 2/3$) at equal number (3.3E10) of emitted photons, our unbiased photon mapping trades blurring (i.e., bias) for noise.

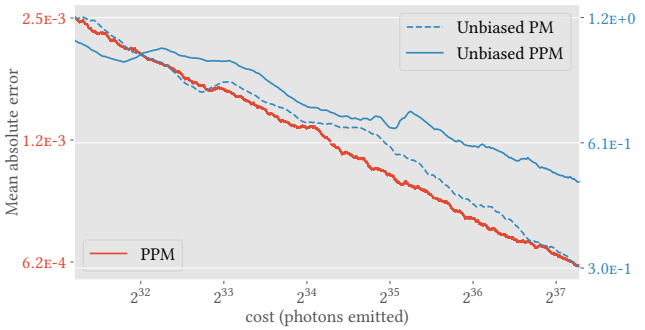


Fig. 12. We show the convergence of the mean absolute difference of our unbiased photon mapping and unbiased progressive photon mapping against progressive photon mapping for a canonical scene containing only a plane and a point light source while measured from a single location for equal cost. The convergence rates between progressive photon mapping and our unbiased photon mapping are similar.

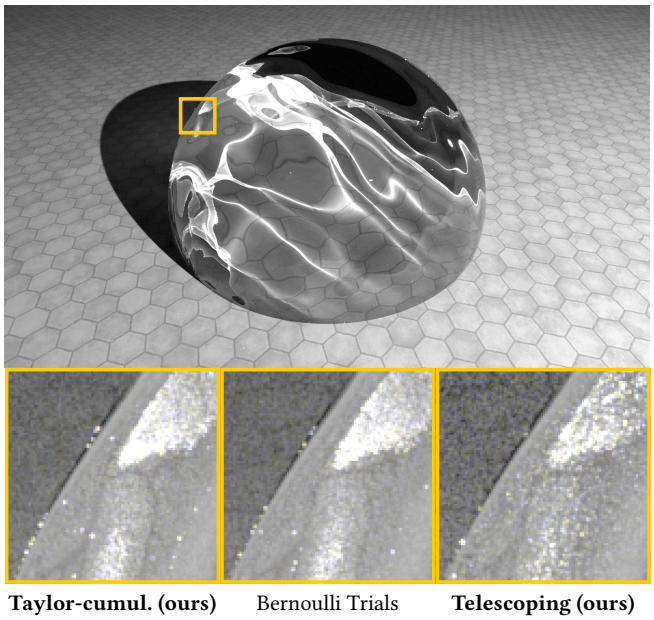
of trading noise for bias when unbiased estimation leads to infinite variance. Using our framework, we have introduced a debiased ray-marching approach which is the first unbiased estimator to support general non-exponential transmittance. We have introduced the first unbiased formulation for photon mapping and progressive photon mapping which supports caustics. We have also shown how to make finite differences unbiased for smooth functions, and formulated a consistent estimator in the presence of discontinuities.

Discussion and future work

The primary estimators we have used rely on either a Taylor series (4) or telescoping series (12) expansion, so a reasonable question to ask is which one should be used when they both apply? We have found this to be extremely problem-dependent. For example, some of the variance reduction techniques introduced by Kettunen et al. [2021] for transmittance estimation cannot be utilized for reciprocal estimation due to its singularity, causing the optimal design decisions for reciprocal estimation to differ from problems akin to transmittance. While an extensive comparative analysis could prove invaluable in guiding such design decisions in future work, there are some clear trade-offs between the two algorithms that are already apparent from our investigations.

The telescoping series performs best when implemented as a global algorithm which combines evaluations of $I(k)$ as a post process. This requires re-structuring rendering algorithms to integrate multiple values separately and correlate entire paths for maximum performance. The Taylor series is applicable only to smooth functions of expectations, making it less general, but more self-contained and easier to maintain, compared to the telescoping series. By utilizing relatively large initial k's and each formulations' respective variance reduction techniques, both formulations can reach comparably low work-normalized-variances.

In addition to looking into optimal estimation configurations for specific applications, there are ample opportunities for exploring



Taylor-cumul. (ours) Bernoulli Trials **Telescoping (ours)**

Fig. 13. Preliminary equal-time (10 minutes) results showing a potential application in formulating unbiased estimation for specular manifold sampling [Zeltner et al. 2020] (i.e. reciprocal estimation). Our Taylor-series estimator (left), based on the pseries-cumulative estimator estimator of Georgiev et al. [2019], performs similarly to Bernoulli trials. Making our telescoping-series estimator (right) competitive requires further investigation.

the full potential of our framework. Most notably, the current state of the art for estimating the reciprocal of a probability is a Bernoulli trial estimator (Fig. 13). While we found some of our estimators can match the performance of Bernoulli trials, we believe further work could use our framework to derive even more efficient estimators.

Many algorithms in graphics use Russian roulette to probabilistically account for an infinite series in finite time. This is just an instance of Eq. (12)'s telescoping debiasing. Russian roulette, however, does not typically choose a selection PMF a priori. The effective PMF

is rather a consequence of each Russian roulette decision, which can notably adapt to local information on the fly. So far, our applications of Eq. (12) have only explored specifying a single global PMF. In the case of path tracing truncation, this would be like specifying the probability of sampling specific length paths without knowing the reflectivity of the surfaces encountered along the paths. Exploring and constructing estimators which exploit local information in Russian roulette and weight windows [Vorba and Křivánek 2016] could be a promising avenue for future work.

The Volterra formulation for transmittance has been shown to encompass all of the prior tracking-based transmittance estimators, which have analogous free-flight distance sampling routines [Georgiev et al. 2019]. Originally, this formulation was derived in a manner which was tied to the definition of classical exponential transmittance. Reformulating this directly as an initial value problem (9,13) opens the possibility to more easily derive unbiased free-flight distance sampling routines for non-exponential media.

ACKNOWLEDGMENTS

The cloud model in Fig. 8 is from Walt Disney Animation Studios. The scenes for Fig. 2, Fig. 4 and Fig. 10 were based off of scenes from Bitterli [2016]. This work was generously supported by NSF awards 1812796 and 1844538, a Neukom Institute CompX faculty grant, and a Facebook PhD fellowship.

REFERENCES

Sai Praveen Bangaru, Tzu-Mao Li, and Frédo Durand. 2020. Unbiased Warped-Area Sampling for Differentiable Rendering. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 39, 6 (Nov. 2020), 245:1–245:18. https://doi.org/10/gj32sk

Benedikt Bitterli. 2016. Rendering resources. https://benedikt-bitterli.me/resources/.

Benedikt Bitterli, Srinath Ravichandran, Thomas Müller, Magnus Wrenninge, Jan Novák, Steve Marschner, and Wojciech Jarosz. 2018. A Radiative Transfer Framework for Non-Exponential Media. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 37, 6 (Nov. 2018), 225:1–225:17. https://doi.org/10/gfz2cm

Jose H. Blanchet, Nan Chen, and Peter W. Glynn. 2015. Unbiased Monte Carlo Computation of Smooth Functions of Expectations via Taylor Expansions. In *2015 Winter Simulation Conference (WSC)*. 360–367. https://doi.org/10/ggr94p

Jose H. Blanchet and Peter W. Glynn. 2015. Unbiased Monte Carlo for Optimization and Functions of Expectations via Multi-Level Randomization. In *2015 Winter Simulation Conference (WSC)*. 3656–3667. https://doi.org/10/ggr94h

T. E. Booth. 2007. Unbiased Monte Carlo Estimation of the Reciprocal of an Integral. *Nuclear Science and Engineering* 156, 3 (July 2007), 403–407. https://doi.org/10/gfzq76

Jérémi Dauchet, Jean-Jacques Bezian, Stéphane Blanco, Cyril Caliot, Julien Charon, Christophe Coustet, Mouna El Hafi, Vincent Eymet, Olivier Farges, Vincent Forest, Richard Fournier, Mathieu Galtier, Jacques Gautrais, Anaïs Khuong, Lionel Pelissier, Benjamin Piaud, Maxime Roger, Guillaume Terrée, and Sebastian Weitz. 2018. Addressing Nonlinearities in Monte Carlo. *Scientific Reports* 8, 1 (Dec. 2018), 13302. https://doi.org/10/gd76pc

Anthony B. Davis and Mark B. Mineev-Weinstein. 2011. Radiation Propagation in Random Media: From Positive to Negative Correlations in High-Frequency Fluctuations. *Journal of Quantitative Spectroscopy and Radiative Transfer* 112, 4 (March 2011), 632–645. https://doi.org/10/fgxnc4

Iliyan Georgiev, Jaroslav Křivánek, Tomáš Davidovič, and Philipp Slusallek. 2012. Light Transport Simulation with Vertex Connection and Merging. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 31, 6 (Nov. 2012), 192:1–192:10. https://doi.org/10/gbb6q7

Iliyan Georgiev, Zackary Misso, Toshiya Hachisuka, Derek Nowrouzezahrai, Jaroslav Křivánek, and Wojciech Jarosz. 2019. Integral Formulations of Volumetric Transmittance. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 38, 6 (Nov. 2019), 154:1–154:17. https://doi.org/10/dffn

Michael B. Giles. 2008. Multilevel Monte Carlo Path Simulation. *Operations Research* 56, 3 (June 2008), 607–617. https://doi.org/10/fmrpfp

Toshiya Hachisuka, Wojciech Jarosz, and Henrik Wann Jensen. 2010. A Progressive Error Estimation Framework for Photon Density Estimation. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 29, 6 (Dec. 2010), 144:1–144:12. https://doi.org/10/dmttfg

Toshiya Hachisuka and Henrik Wann Jensen. 2009. Stochastic Progressive Photon Mapping. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 28, 5 (Dec. 2009), 130:1–130:8. https://doi.org/10/d8xxn3

S. Heinrich. 1998. A Multilevel Version of the Method of Dependent Tests. In *Proc. of the 3rd St. Petersburg Workshop on Simulation*. St. Petersburg University Press, 31–35.

Adrian Jarabo, Carlos Aliaga, and Diego Gutierrez. 2018. A Radiative Transfer Framework for Spatially-Correlated Materials. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 37, 4 (July 2018), 83:1–83:13. https://doi.org/10/gd52qq

Wojciech Jarosz, Derek Nowrouzezahrai, Iman Sadeghi, and Henrik Wann Jensen. 2011a. A Comprehensive Theory of Volumetric Radiance Estimation Using Photon Points and Beams. *ACM Transactions on Graphics* 30, 1 (Jan. 2011), 5:1–5:19. https://doi.org/10/fcdh2f

Wojciech Jarosz, Derek Nowrouzezahrai, Robert Thomas, Peter-Pike Sloan, and Matthias Zwicker. 2011b. Progressive Photon Beams. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 30, 6 (Dec. 2011), 181:1–181:12. https://doi.org/10/fn5xzj

Henrik Wann Jensen. 2001. *Realistic Image Synthesis Using Photon Mapping*. AK Peters, Ltd., Natick, MA, USA.

Jiantao Jiao and Yanjun Han. 2020. Bias Correction with Jackknife, Bootstrap, and Taylor Series. *IEEE Transactions on Information Theory* 66, 7 (July 2020), 4392–4418. https://doi.org/10/gj32t9

Daniel Jonsson, Joel Kronander, Jonas Unger, Thomas B. Schon, and Magnus Wrenninge. 2020. Direct Transmittance Estimation in Heterogeneous Participating Media Using Approximated Taylor Expansions. *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. https://doi.org/10/gj32rc

A. Keller. 2001. Hierarchical Monte Carlo Image Synthesis. *Mathematics and Computers in Simulation* 55, 1-3 (2001), 79–92.

Markus Kettunen, Eugene d'Eon, Jacopo Pantaleoni, and Jan Novak. 2021. An Unbiased Ray-Marching Transmittance Estimator. (Feb. 2021). arXiv:2102.10294 [cs.GR]

Claude Knaus and Matthias Zwicker. 2011. Progressive Photon Mapping: A Probabilistic Approach. *ACM Transactions on Graphics* 30, 3 (May 2011), 25:1–25:13. https://doi.org/10/bcw2ph

Peter Kutz, Ralf Habel, Yining Karl Li, and Jan Novák. 2017. Spectral and Decomposition Tracking for Rendering Heterogeneous Volumes. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 36, 4 (July 2017), 111:1–111:16. https://doi.org/10/gbxjxg

Don McLeish. 2011. A General Method for Debiasing a Monte Carlo Estimator. *Monte Carlo Methods and Applications* 17, 4 (Jan. 2011). https://doi.org/10/c2fpdx

Zackary Misso, Benedikt Bitterli, Iliyan Georgiev, and Wojciech Jarosz. 2022. *Unbiased and consistent rendering using biased estimators supplemental code and data.* https://doi.org/10.5281/zenodo.6515009

Sarat Babu Moka, Dirk P. Kroese, and Sandeep Juneja. 2019. Unbiased Estimation of the Reciprocal Mean for Non-Negative Random Variables. In *2019 Winter Simulation Conference (WSC)*. 404–415. https://doi.org/10/ggr944

Jan Novák, Andrew Selle, and Wojciech Jarosz. 2014. Residual Ratio Tracking for Estimating Attenuation in Participating Media. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 33, 6 (Nov. 2014), 179:1–179:11. https://doi.org/10/f6r2nq

Adithya Pediredla, Yasin Karimi Chalmiani, Matteo Giuseppe Scopelliti, Maysamreza Chamanzar, Srinivasa Narasimhan, and Ioannis Gkioulekas. 2020. Path Tracing Estimators for Refractive Radiative Transfer. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 39, 6 (Nov. 2020), 241:1–241:15. https://doi.org/10/gj4msh

Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2016. *Physically Based Rendering: From Theory to Implementation* (3rd ed.). Morgan Kaufmann, Cambridge, MA.

Hao Qin, Xin Sun, Qiming Hou, Baining Guo, and Kun Zhou. 2015. Unbiased Photon Gathering for Light Transport Simulation. *ACM Transactions on Graphics* 34, 6 (Oct. 2015). https://doi.org/10/f7wrc6

Chang-han Rhee. 2013. *Unbiased Estimation with Biased Samplers*. Ph. D. Dissertation. Stanford University.

Chang-han Rhee and Peter W. Glynn. 2012. A New Approach to Unbiased Estimation for SDE's. (July 2012). arXiv:1207.2452 [q-fin.CP]

Chang-Han Rhee and Peter W. Glynn. 2015. Unbiased Estimation with Square Root Convergence for SDE Models. *Operations Research* 63, 5 (Oct. 2015), 1026–1043. https://doi.org/10/f7xf6m

Delio Vicini, Wenzel Jakob, and Anton Kaplanyan. 2021. A Non-Exponential Transmittance Model for Volumetric Scene Representations. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 40, 4 (Aug. 2021), 136:1–136:16. https://doi.org/10/hshg

Jiří Vorba and Jaroslav Křivánek. 2016. Adjoint-Driven Russian Roulette and Splitting in Light Transport Simulation. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 35, 4 (July 2016), 42:1–42:11. https://doi.org/10/f89mcz

Christopher S. Withers. 1987. Bias Reduction by Taylor Series. *Communications in Statistics - Theory and Methods* 16, 8 (Jan. 1987), 2369–2383. https://doi.org/10/bmqbg4

Tizian Zeltner, Iliyan Georgiev, and Wenzel Jakob. 2020. Specular Manifold Sampling for Rendering High-Frequency Caustics and Glints. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 39, 4 (July 2020). https://doi.org/10/gg8xc8

A VARIANCE-OPTIMAL SINGLE-TERM ESTIMATION

Here we derive the variance-optimal PDF/PMF $p(x)$ for the integral estimator in Eq. (14) in the continuous and discrete case.

Continuous case. Since the estimator is unbiased, minimizing its variance is equivalent to minimizing its second moment,

$$\mathrm{E}\left[\left(\frac{\langle\Delta_j\rangle}{p(j)}\right)^2\right] = \int_k^\infty \frac{\mathrm{E}\left[\langle\Delta_x\rangle^2\right]}{p(x)^2} p(x)\,\mathrm{d}x = \int_k^\infty \frac{\mathrm{E}\left[\langle\Delta_x\rangle^2\right]}{p(x)}\,\mathrm{d}x, \quad (28)$$

under the condition that $p(x)$ is normalized, i.e., $\int_k^\infty p(x)\,\mathrm{d}x = 1$. This problem can be solved using the Euler-Lagrange equation from the calculus of variations as follows. We first define

$$\mathcal{L}(x,\lambda,p) = \frac{\mathrm{E}\left[\langle\Delta_x\rangle^2\right]}{p(x)} + \lambda p(x), \quad (29)$$

where λ is the (yet unknown) Lagrange multiplier. We then write the Euler-Lagrange equation for q, which we solve for p:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p} - \frac{\mathrm{d}}{\mathrm{d}x}\underbrace{\left(\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p'}\right)}_{=0} = 0 \quad \Leftrightarrow \quad -\frac{\mathrm{E}\left[\langle\Delta_x\rangle^2\right]}{p^2(x)} + \lambda = 0 \quad (30)$$

$$\Leftrightarrow \quad p(x) = \sqrt{\frac{\mathrm{E}\left[\langle\Delta_x\rangle^2\right]}{\lambda}}. \quad (31)$$

Since $p(x)$ must be normalized, λ is its normalization constant. Finally, since $\mathrm{V}[\langle\Delta_x\rangle] = \mathrm{E}\left[\langle\Delta_x\rangle^2\right] - (\Delta_x)^2$, we minimize variance when

$$p(x) \propto \sqrt{\mathrm{V}[\langle\Delta_x\rangle] + (\Delta_x)^2}. \quad (32)$$

Discrete case. In this case the second moment takes the form

$$\mathrm{E}\left[\left(\frac{\langle\Delta_j\rangle}{p(j)}\right)^2\right] = \sum_{j=k}^\infty \frac{\mathrm{E}\left[\langle\Delta_j\rangle^2\right]}{p(j)}. \quad (33)$$

To find the minimum of this sum w.r.t. $p(j)$, we use the method of Lagrange multipliers. We define a new minimization problem

$$\mathcal{L}(\lambda,p) = \sum_{j=k}^\infty \frac{\mathrm{E}\left[\langle\Delta_j\rangle^2\right]}{p(j)} - \lambda\left(1 - \sum_{j=k}^\infty p(j)\right) \quad (34)$$

and find its local extrema by solving for roots of its gradient. The sum in the parentheses encodes the PMF normalization condition. Solving for $p(j)$ yields

$$\frac{\partial\mathcal{L}}{\partial p(j)} = 0 \quad \Leftrightarrow \quad -\frac{\mathrm{E}\left[\langle\Delta_j\rangle^2\right]}{p(j)^2} + \lambda = 0 \quad (35)$$

$$\Leftrightarrow \quad p(j) = \sqrt{\frac{\mathrm{E}\left[\langle\Delta_j\rangle^2\right]}{\lambda}} \quad \Leftrightarrow \quad p(j) \propto \sqrt{\mathrm{V}[\langle\Delta_j\rangle] + \left(\Delta_j\right)^2}, \quad (36)$$

where in Eq. (36) λ is a normalization constant, arriving at an optimal PMF analogous to the one for the continuous case above.

B CONDITIONS FOR CONSISTENCY

For a secondary estimator to be consistent, we need its MSE to converge:

$$\lim_{N\to\infty} \mathrm{MSE}[\langle I\rangle_N] = \lim_{N\to\infty}\left(\mathrm{V}[\langle I\rangle_N] + \mathrm{B}[\langle I\rangle_N]^2\right) = 0. \quad (37)$$

For this we need both the bias and the variance to converge:

$$\lim_{N\to\infty} \mathrm{V}[\langle I\rangle_N] = \lim_{N\to\infty} \frac{1}{N^2}\sum_{i=1}^N \underbrace{\mathrm{V}[\langle I(k_i)\rangle]}_{V(k_i)} = \frac{O(V(k_N))}{N} = 0, \quad (38)$$

$$\lim_{N\to\infty} \mathrm{B}[\langle I\rangle_N] = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \underbrace{\mathrm{B}[\langle I(k_i)\rangle]}_{B(k_N)} = O(B(k_N)) = 0, \quad (39)$$

This result tells us that consistency is achieved when the primary estimator's bias $B(k)$ vanishes, at any rate, while its variance is even allowed to increase, though sublinearly. Among all parameterizations for k_N that meet these two conditions, the one that maximizes the MSE convergence rate is the one that achieves $O(\mathrm{V}[\langle I\rangle_N]) = O\left(\mathrm{B}[\langle I\rangle_N]^2\right)$, i.e. $O(V(k_N)) = NO\left(B^2(k_N)\right)$.

C DEBIASING PHOTON MAPPING

In the following we describe the steps to debias photon mapping. We begin with the biased but consistent base estimator $\langle I(k)\rangle$ which traces M_k photons and performs density estimation with a kernel of radius r_k. For suitably chosen M_k and r_k, this estimator converges to I as $k \to \infty$ [Jensen 2001].

We form a single-term telescoping-series estimator (14) assuming perfectly correlated evaluation of $\langle\Delta_j\rangle = \langle I(j+1)\rangle - \langle I(j)\rangle$, i.e. both evaluations of $\langle I(j)\rangle$ and $\langle I(j+1)\rangle$ share the same set of photons. If both estimators trace different numbers of photons (i.e. $M_{j+1} \neq M_j$), we ignore unshared photons for the rest of this analysis, giving a lower bound on the variance and bias. Under this assumption, evaluation of the two estimators differs only in the kernel radius used (r_j and r_{j+1}), and we can express the difference of the two estimators as a single photon mapping estimator that uses the difference of the two kernels for density estimation.

This allows us to use the analysis of Knaus and Zwicker [2011] as a basis to derive the expectation and variance of $\langle\Delta_j\rangle$ which are needed for deriving a variance-optimal PMF for sampling the telescoping series. Repeating the derivation of Knaus and Zwicker using the modified difference kernel yields the following asymptotic relationships:

$$\mathrm{V}[\langle\Delta_j\rangle] \propto \frac{1}{M_j}\int \left(k(r_{j+1},x) - k(r_j,x)\right)^2 \mathrm{d}x, \quad (40)$$

$$\Delta_j \propto r_j^2, \quad (41)$$

where $k(r,x)$ evaluates the density-estimation kernel of radius r at location x.

The asymptotics above depend on the relationship of the number of photons M_k and the kernel radius r_k with respect to k, as well as the specific choice of kernel. Knaus and Zwicker [2011] and Georgiev et al. [2012] show that $r_j = \sqrt{j^{\alpha-1}}$ with $0 < \alpha < 1$ yields consistent progressive estimation. This leads to the following asymptotics for the expectation and the variance:

$$\Delta_j = O(j^{\alpha-2}) \quad (42)$$

$$\mathrm{V}[\langle\Delta_j\rangle] = M_j^{-1}O(j^{-\alpha}), \quad \text{(constant kernel)} \quad (43)$$

$$\mathrm{V}[\langle\Delta_j\rangle] = M_j^{-1}O(j^{-\alpha-1}). \quad \text{(cone kernel)} \quad (44)$$

Given the asymptotics, we can now test whether debiasing yields finite work-normalized variance. First, the optimal PMF

$$p(j) \propto \sqrt{\mathrm{V}[\langle\Delta_j\rangle] + \left(\Delta_j\right)^2} = \sqrt{M_j^{-1}O(j^{-\alpha-1}) + O(j^{2\alpha-4})} \quad (45)$$

must be normalizable (using the best-case, cone-kernel variance). Second, the expected work

$$\mathrm{C}[\langle I\rangle_1] \propto \sum_{j=1}^{\infty} M_j \cdot p(j) \propto \sum_{j=1}^{\infty} \sqrt{M_j O(j^{-\alpha-1}) + M_j^2 O(j^{2\alpha-4})} \quad (46)$$

must be finite (here, using the number of emitted photons as an estimate of the cost). Sadly, there is no choice of M_j and α that can satisfy both these conditions. The estimator is still debiasable, but with infinite work-normalized variance, i.e., with reduced secondary-estimator convergence rate.

D IDEAL-CASE FINITE-DIFFERENCE ANALYSIS

D.1 Expectation & variance of finite-difference deltas

To form an estimator of Δ_j^{FD} (21), we first need to estimate the integrals in Eq. (20). In the ideal case, we can perfectly correlate evaluations of $f(x,y)$ through the use of the same samples for y, so for brevity we will simplify $f(x,y)$ to $f(x)$ in the following derivations. This results in

$$\langle\Delta_j^{\mathrm{FD}}\rangle = \frac{\langle f(x+h_{j+1})\rangle - \langle f(x)\rangle}{h_{j+1}} - \frac{\langle f(x+h_j)\rangle - \langle f(x)\rangle}{h_j}. \quad (47)$$

If $f(x)$ is continuously differentiable with respect to x, we have

$$f(x+h) = f(x) + h\frac{\partial}{\partial x}f(x) + h^2\frac{\partial^2}{\partial x^2}f(x) + O(h^3). \quad (48)$$

Inserting Eq. (48) into (47) and canceling terms gives

$$\langle\Delta_j^{\mathrm{FD}}\rangle = (h_{j+1} - h_j)\left\langle\frac{\partial^2}{\partial x^2}f(x)\right\rangle + O(h_{j+1}^2) - O(h_j^2), \quad (49)$$

and since $h_{j+1} < h_j$ by construction, we have

$$\Delta_j^{\mathrm{FD}} \propto (h_{j+1} - h_j), \quad \text{and} \quad \mathrm{V}[\langle\Delta_j^{\mathrm{FD}}\rangle] \propto \left(h_{j+1} - h_j\right)^2. \quad (50)$$

D.2 Bias & variance rates of consistent finite differences

For the ideal case where all evaluations of $f(x,y)$ can be evaluated using the same samples for y Eq. (47), estimating $I(k)$ in Eq. (20) is equivalent to estimating the true derivative convolved with a box kernel (for brevity, we use $f(x)$ since we evaluate each $f(x,y)$ using a shared sample y):

$$\langle I(k)\rangle = \frac{\langle f(x+h_k) - f(x)\rangle}{h_k} = \frac{1}{h_k}\int_x^{x+h_k}\left\langle\tfrac{\partial}{\partial x}f(t)\right\rangle \mathrm{d}t \quad (51)$$

$$= \int_{-\infty}^{\infty} K(x-t)\left\langle\tfrac{\partial}{\partial x}f(t)\right\rangle \mathrm{d}t, \qquad \text{where} \quad (52)$$

$$K(\tau) = \begin{cases} \frac{1}{h_k} & 0 < \tau < h_k \\ 0 & \text{otherwise} \end{cases} \qquad \text{is a normalized box kernel.} \quad (53)$$

If we replace $\left\langle\frac{\partial}{\partial x}f(t)\right\rangle$ with its expectation, this shows that the bias in finite difference estimators results from a 1D blur (along the direction of differentiation), analogously to how photon beams [Jarosz et al. 2011a] return a 1D blurred version of volumetric radiance.

Given this equivalence, a direct Monte Carlo estimator of Eq. (20) will have the same bias and variance convergence rates as derived by Jarosz et al. [2011b] for photon beams when the parameter of differentiation does not modify the path space, with the finite difference step size h_k standing in for the blur kernel width:

$$\mathrm{B}[\langle I(k)\rangle] = O(h_k) \quad \text{and} \quad \mathrm{V}[\langle I(k)\rangle] = O\left(h_k^{-1}\right). \quad (54)$$

This implies that, just like in progressive photon mapping, we can form a progressive finite difference estimator by averaging multiple finite difference renderings while reducing the step size h_k according to $h_k = h_0\sqrt{k^{a-1}}$. This is guaranteed to be a consistent estimator for parameter $a \in (0,1)$, and provides its optimal MSE convergence rate with $a = 2/3$.

More generally, by the law of large numbers progressive estimators will still converge even if all y samples cannot be correlated, however, the convergence rate will be slower then our derived ideal rates Eq. (54).

E DERIVING CONSISTENT PROGRESSIVE ESTIMATION FROM UNBIASED ESTIMATION

We first construct an unbiased progressive secondary estimator Eq. (16) with $k_i = i$, and replace the primary estimator with its expectation:

$$I = \frac{1}{N}\sum_{k=0}^{N}\left(I(k) + \sum_{j=k}^{\infty}\Delta_j\right) \quad (55)$$

$$= \frac{1}{N}\sum_{k=0}^{N} I(k) + \sum_{k=0}^{N}\sum_{j=k}^{\infty}\frac{\Delta_j}{N}. \quad (56)$$

From here it is easy to see that since all Δ_j's are finite by construction ($\Delta_\infty = 0$), the right (i.e. bias) term in Eq. (56) converges to zero:

$$\lim_{N\to\infty}\sum_{k=0}^{N}\sum_{j=k}^{\infty}\frac{\Delta_j}{N} = \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N}\sum_{j=k}^{\infty}\Delta_j = \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N} a_k = 0. \quad (57)$$

Choosing to never evaluate the bias yields a biased but consistent progressive formulation:

$$I = \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N} I(k). \quad (58)$$