

Importance Sampling BRDF Derivatives

YASH BELHE, University of California San Diego, USA

BING XU, University of California San Diego, USA

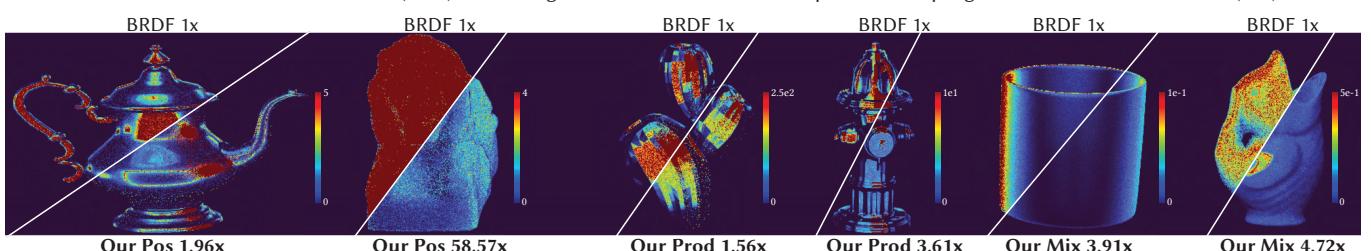
SAI PRAVEEN BANGARU, MIT CSAIL, USA

RAVI RAMAMOORTHI, University of California San Diego, USA

TZU-MAO LI, University of California San Diego, USA



Our techniques can importance sample the derivatives of a wide variety of materials, some of which are shown above. The insets indicate the regions where our estimators have lower standard deviation (**blue**) and the regions where standard BRDF importance sampling has lower standard deviation (**red**).



Standard deviation of derivative estimators: BRDF Importance Sampling (top left diagonal) vs our three decompositions, Positivization (Pos), Product Decomposition (Prod), and Mixture Decomposition (Mix) (bottom right diagonal). Numbers indicate improvement in gradient estimation, **higher is better**.

Fig. 1. We propose new importance sampling techniques for sampling derivatives of BRDFs, and they achieve significant variance reduction in the estimated derivatives. Our techniques work better because they correctly deal with real-valued BRDF derivatives, for which BRDF importance sampling from forward rendering is not well suited. Our techniques are general and apply to a wide variety of BRDF derivatives, which was not possible by previous work in differentiable rendering [Zeltner et al. 2021; Zhang et al. 2021a]. 3D models courtesy of Turbosquid users id_inc (teapot), Evilordus (lion), Adrian Kulawik (hydrant), 3d_molier International (cactus), cgaustria (fish vase).

We propose a set of techniques to efficiently importance sample the derivatives of a wide range of BRDF models. In differentiable rendering, BRDFs are replaced by their differential BRDF counterparts which are real-valued and can have negative values. This leads to a new source of variance arising from their change in sign. Real-valued functions cannot be perfectly importance sampled by a positive-valued PDF, and the direct application of BRDF sampling leads to high variance. Previous attempts at antithetic sampling only addressed the derivative with the roughness parameter of isotropic microfacet BRDFs. Our work generalizes BRDF derivative sampling to anisotropic microfacet models, mixture BRDFs, Oren-Nayar, Hanrahan-Krueger, among other analytic BRDFs.

Our method first decomposes the real-valued differential BRDF into a sum of single-signed functions, eliminating variance from a change in sign.

Next, we importance sample each of the resulting single-signed functions separately. The first decomposition, positivization, partitions the real-valued function based on its sign, and is effective at variance reduction when applicable. However, it requires analytic knowledge of the roots of the differential BRDF, and for it to be analytically integrable too. Our key insight is that the single-signed functions can have overlapping support, which significantly broadens the ways we can decompose a real-valued function. Our product and mixture decompositions exploit this property, and they allow us to support several BRDF derivatives that positivization could not handle. For a wide variety of BRDF derivatives, our method significantly reduces the variance (up to 58x in some cases) at equal computation cost and enables better recovery of spatially varying textures through gradient-descent-based inverse rendering.

ACM Reference Format:

Yash Belhe, Bing Xu, Sai Praveen Bangaru, Ravi Ramamoorthi, and Tzu-Mao Li. 2024. Importance Sampling BRDF Derivatives. *ACM Trans. Graph.* 1, 1 (February 2024), 23 pages. <https://doi.org/10.1145/nmnnnn.nmnnnn>

Authors' addresses: Yash Belhe, University of California San Diego, USA, ybelhe@ucsd.edu; Bing Xu, University of California San Diego, USA, b4xu@ucsd.edu; Sai Praveen Bangaru, MIT CSAIL, USA, sbangaru@mit.edu; Ravi Ramamoorthi, University of California San Diego, USA, ravir@ucsd.edu; Tzu-Mao Li, University of California San Diego, USA, tzli@ucsd.edu.

© 2024 Association for Computing Machinery.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *ACM Transactions on Graphics*, <https://doi.org/10.1145/nmnnnn.nmnnnn>.

1 INTRODUCTION

BRDF importance sampling is an essential variance reduction technique for Monte Carlo forward rendering. However, there is no

simple counterpart for differentiable rendering. Taking the derivative of a BRDF with respect to one of its parameters transforms it into a real-valued differential BRDF. The differential BRDF can have a very different shape from the BRDF, and can also take negative values. Previous attempts at tackling this problem [Zeltner et al. 2021] are limited to the roughness derivatives of isotropic GGX (Trowbridge-Reitz) and Beckmann BRDFs, and cannot handle even their anisotropic counterparts. Another method [Zhang et al. 2021a] was developed primarily for odd functions with symmetric positive and negative lobes, and can produce substantially higher variance when the derivative is close to an even function. We propose effective importance sampling of derivatives of not only anisotropic GGX and Beckmann BRDFs, but also a wide variety of other analytic BRDF models like Ashikhmin-Shirley, Oren-Nayar, Hanrahan-Krueger, Mixture BRDFs, and ABC models. Fig. 1 demonstrates the benefits of our method on several BRDFs compared to BRDF sampling.

Importance sampling a real-valued function leads to unique challenges. Its variance has two sources, *a) its sign, and b) its shape*. Our idea is to decompose the function into a sum of single-signed functions, which we call **single-signed decompositions**. A single-signed function is either positive everywhere or negative everywhere in its domain. Single-signed functions, by definition, have no sign variance. Importance sampling these functions eliminates their shape variance.

A classical strategy, positivization [Owen and Zhou 2000], is a special case of our single-signed decomposition. It has positive and negative parts with *non-overlapping support*, which in turn requires a) analytic knowledge of the roots and b) analytic integrability of the BRDF derivative up to the roots, which is possible only for certain BRDF derivatives. To sidestep these issues due to a partition of the domain, we introduce the product and mixture decompositions for which *we allow the positive and negative parts to overlap*. In fact, we ensure that both the positive and the negative parts have support over the entire hemisphere. This enables analytic integrability and significantly expands upon the set of BRDF derivatives we can handle. *Our main contributions are three single-signed decompositions and the corresponding importance sampling PDFs of a large set of BRDF derivatives, see Table 1.*

Positivization. First, we introduce a simple decomposition called positivization (Sec. 4.2), which *partitions a real-valued function about its roots* into a positive and a negative function. We show that Zeltner et al.’s [2021] antithetic sampling is a special case of positivization, and positivization provides an explanation of the efficiency of their approach. When applicable, positivization leads to significant variance reduction. For example, it can be applied for sampling the isotropic GGX, Beckmann and Hanrahan-Krueger BRDF derivatives. However, others like anisotropic GGX, Beckmann, Ashikhmin-Shirley (Sec. 4.2.2) are not analytically integrable up to their roots, and the derivatives with mixture weights (Sec. 6) do not have analytic roots. Positivization cannot handle these derivatives. Zeltner et al.’s antithetic sampling inherits these limitations too.

Product Decomposition. Second, we propose a *novel* product decomposition (Sec. 5). Our key observation is that after differentiation,

many BRDF derivatives can be decomposed into single-signed functions by *separating the terms that result from the derivative product rule*. Product decomposition does *not* require knowledge of the roots for the decomposition and only requires the resulting single-signed functions to be analytically integrable. Product decomposition can importance sample the derivatives of anisotropic GGX, Beckmann, Ashikhmin-Shirley, and more.

Mixture Decomposition. Finally, we introduce mixture decomposition (Sec. 6). Derivatives of BRDFs with linear combination coefficients, e.g., mixture weights of a layered BRDF, result in real-valued functions whose roots cannot be found analytically in most cases. Our mixture decomposition exploits the fact that this derivative is the difference between two positive-valued terms. Separating them results in a single-signed decomposition, and the two terms can then be importance sampled separately. Mixture decomposition handles the derivatives of Oren-Nayar and mixture weights of *Uber* BRDFs such as the Disney BRDF or Autodesk Standard Surface.

It is likely that several other BRDF derivatives not surveyed in this paper can also be dealt with by one of our three decompositions, and we provide a recipe for handling them in Sec. 7. We provide a library of importance sampling PDFs for the derivatives of all the BRDF models discussed in this work in Table 1.

2 RELATED WORK

Our work connects two areas in rendering research, differentiable rendering and BRDF sampling.

2.1 Differentiable Rendering

Derivatives in rendering. Computing derivatives or gradients of light transport has a long history. Earlier work focused on accelerating light transport using derivatives [Arvo 1994; Ramamoorthi et al. 2007; Ward and Heckbert 1992]. Approximate differentiable renderers [de La Gorce et al. 2011; Kato et al. 2018; Laine et al. 2020; Liu et al. 2020; Loper and Black 2014] have been used for many computer vision tasks, and light transport derivatives have been used for recovering scattering coefficients [Gkioulekas et al. 2013; Khungurn et al. 2015].

Background on differentiable rendering. Much of the current interest in Monte Carlo differentiable rendering was started by Li et al. [2018], who introduced an edge sampling approach to correctly handle discontinuities in both primary and secondary visibility. As shown by them and subsequent work [Zhang et al. 2019], the derivative of the rendering equation is made up of an interior integral which handles continuous function variation, and a boundary integral which encapsulates discontinuities.

Follow up work [Bangaru et al. 2020; Loubet et al. 2019; Xu et al. 2023a; Yan et al. 2022; Yu et al. 2022; Zhang et al. 2020, 2023] focused on accurately computing the boundary integral. Some other recent work focused on reducing memory requirements [Nimier-David et al. 2020; Vicini et al. 2021], and building automatic differentiation systems and compilers [Bangaru et al. 2023; Jakob et al. 2022; Nimier-David et al. 2019]. Efforts have been made to handle different light transport phenomena [Wu et al. 2021; Yi et al. 2021; Zhang et al. 2019, 2021b]. Some recent work has investigated the possibility to leverage

Table 1. LIST OF SUPPORTED MATERIAL DERIVATIVES. The first column lists the name of the BRDF, and the second column lists the corresponding parameter whose derivative we can importance sample. The third column lists the type of single-signed decomposition applied (Positivization, Product Decomposition, Mixture Decomposition). The fourth column lists the section number in the Appendix (with links) with the relevant sampling PDFs. Please refer to the original papers for definitions of the parameters.

Material	Param.	SSD	PDFs
Isotropic GGX [1975; 2007]	α	Pos.	A.1.1
Isotropic Beckmann [1987]	α	Pos.	A.1.2
Blinn Phong (Minnaert) [1977; 1941]	n	Pos.	A.1.3
Henyey-Greenstein (Hanrahan-Krueger) [1993; 1941]	g	Pos.	A.1.4
Anisotropic GGX [1975; 2007]	α_x, α_y	Prod.	A.2.1
Anisotropic Beckmann (Ward) [1987; 1992]	α_x, α_y	Prod.	A.2.2
Ashikhmin-Shirley [2001]	n_u, n_v	Prod.	A.2.3
Isotropic ABC [Löw et al. 2012]	B, C	Prod.	A.2.4
Isotropic Hemi-EPD [2009]	κ	Prod.	A.2.5
Burley Diffuse Reflectance [2015]	d	Prod.	A.2.6
Mixture Model (e.g., Autodesk, Disney BRDF) [2012; 2019]	w	Mix.	A.3.1
Oren-Nayar [1994]	σ	Mix.	A.3.2
Microcylinder [2013]	k_d	Mix.	A.3.3

coherency between gradient descent iterations in inverse rendering to reduce variance [Chang et al. 2023; Nicolet et al. 2023; Wang et al. 2023]. Much of the recent inverse rendering work has started to incorporate differentiable rendering components [Azinović et al. 2019; Che et al. 2020; Deschaintre et al. 2018; Luan et al. 2021; Nimier-David et al. 2021, 2022; Sun et al. 2023; Wu et al. 2023]. These efforts are complementary to our work.

Zeltner et al. [2021] show that directly importance sampling a BRDF’s derivative leads to a *detached derivative* with only an interior term, and no boundary term. They also show that reparameterization before differentiation leads to a different *attached derivative* with not only an interior term but an additional boundary term too. The boundary term requires careful handling for unbiased estimates and extra auxiliary rays at each shading point to estimate it too (4 to 64 extra rays as per Bangaru et al. [2020]). As a result, attached estimators are not suitable for the low sample budget within which we aim to operate. Our estimators fall under the detached derivative regime, which does not require these extra auxiliary rays, which makes them suitable for low sample budget derivative estimation.

2.2 BRDFs and Importance Sampling

Our work supports importance sampling the derivatives of a wide variety of analytic BRDF models. Table 1 shows a *comprehensive list of the supported BRDF derivatives and their importance sampling PDFs*. The code for the sampling routines is included in supplementary material.

Importance Sampling BRDFs. Importance sampling according to the BRDF [Pharr et al. 2016] is a fundamental variance reduction technique used in Monte Carlo forward rendering. While essential, it was initially limited to Phong-like BRDFs [Lafortune et al.

1997; Phong 1975] and Ward [1992]. Lawrence et al. [2004] introduced a non-negative matrix based factorization to efficiently fit analytic and measured BRDFs for sampling. Walter et al. [2007] introduced the GGX BRDF [Trowbridge and Reitz 1975] along with its importance sampling routines. Follow-up works have correctly accounted for the shadowing and masking terms to sample microfacet BRDFs [Heitz 2017, 2018; Heitz and d’Eon 2014; Jakob 2014].

Data-Driven BRDFs. Apart from analytic BRDFs, data-driven measured BRDFs [Dupuy and Jakob 2018; Matusik et al. 2003] and Neural BRDFs [Fan et al. 2022; Kuznetsov et al. 2021, 2022; Sztrajman et al. 2021; Xu et al. 2023b] are another common class of BRDF models that can model a wide variety of materials. However, both these Neural BRDFs and non-analytic measured BRDFs have a very large number of parameters, and it is unclear which parameters one might want to differentiate and importance sample with respect to. Hence, we do not consider either of these classes of BRDFs in our work and focus instead on common analytic BRDF models.

3 BACKGROUND

For the sake of simplicity, we begin our discussion by focusing on the direct lighting setting, and extend it to indirect lighting in Sec. 9. The reflected radiance L_r , at a shading point \mathbf{y} , in the direction ω_o , is given by the reflection equation [Cohen and Wallace 1993],

$$L_r(\mathbf{y}, \omega_o; \alpha) = \int f(\mathbf{y}, \omega_i, \omega_o; \alpha) L_i(\mathbf{y}, \omega_i) d\omega_i. \quad (1)$$

Here, f is the cosine-weighted BRDF at \mathbf{y} , and α is a scalar BRDF parameter that controls f . In practice, α is the vector of all BRDF parameters in a given scene. However, for ease of exposition, we assume α is scalar-valued, with the results for the other parameters following similarly. For example, α could be the roughness of an isotropic GGX BRDF. Since we are dealing with only direct lighting, the incident radiance L_i does not depend upon α . Differentiating the expression for the reflected radiance with α , we get

$$\partial_\alpha L_r(\mathbf{y}, \omega_o; \alpha) = \int \partial_\alpha f(\mathbf{y}, \omega_i, \omega_o; \alpha) L_i(\mathbf{y}, \omega_i) d\omega_i. \quad (2)$$

Since the parameter α does not affect discontinuities, we do not have an additional boundary integral [Li et al. 2018].

In forward rendering, BRDF sampling aims to minimize the variance of the BRDF f in the reflection equation, Eqn. (1). Similarly, our goal is to *minimize the variance of the differential BRDF $\partial_\alpha f$* in differentiable rendering, i.e. the variance of the estimator for

$$I(\omega_o; \alpha) = \int \partial_\alpha f(\omega_i, \omega_o; \alpha) d\omega_i. \quad (3)$$

We drop the spatial coordinate \mathbf{y} , without loss of generality, for simplicity. We deal with the incident radiance L_i using light source sampling. The estimators for $\partial_\alpha f$ and L_i can be combined using Multiple Importance Sampling [Veach and Guibas 1995]. We finally want to compute $\partial_\alpha L_r$ so the final estimator must always include multiplication by L_i . All following integrals in the paper omit L_i to stress that we only focus on $\partial_\alpha f$.

3.1 Previous Work on Variance Reduction for Differentiable Rendering

3.1.1 Detached & Antithetic Sampling. Zeltner et al. [2021] noticed that standard BRDF sampling using a PDF $p \propto f$ for the differential BRDF $\partial_\alpha f$ leads to high variance since $\partial_\alpha f$ and f can be very different functions. They instead construct a PDF $p \propto |\partial_\alpha f|$, called the differential detached PDF, which matches $\partial_\alpha f$ in shape. This eliminates variance from the shape of $\partial_\alpha f$, i.e., the sample weights $\partial_\alpha f/p$ are constant in magnitude. There is, however, additional *sign variance* resulting from the mismatch in the sign between the positive-valued p and the real-valued integrand $\partial_\alpha f$ resulting in sample weights $\partial_\alpha f/p$ that change sign.

To deal with sign variance, Zeltner et al. [2021] applied antithetic sampling. (3) We show that Zeltner et al.’s method is a special case of another technique called positivization [Owen and Zhou 2000].

We show in Sec. 4.2 and Appendix B that positivization provides a theoretical grounding of antithetic sampling: the effectiveness mainly comes from the stratification (separating the real-valued function into a positive and a negative function). The *major drawback of antithetic sampling is its inapplicability to several BRDF derivatives*, due to the lack of closed forms of root finding and integration, which we discuss in Sec. 4.2.2.

Positivization has also been applied by concurrent work in differentiable rendering to the *target function* q in resampled importance sampling (RIS) [Talbot et al. 2005] by Chang et al. [2023] (see Eqn. 11 in their paper). Our work is orthogonal to their method: They use BRDF importance sampling as their *candidate distribution* p . The variance of RIS is a linear combination of $\text{Var}(f/p)$ and $\text{Var}(f/q)$ (see Eqn 4.2 in Talbot et al. [2005]’s work); positivized RIS reduces the variance of the second term by positivizing the target function q into q_+, q_- . Our method of positivizing the differential BRDF (using p_+, p_- instead of p) reduces the variance of the first term.

3.1.2 Antithetic Sampling of Odd Derivatives. Zhang et al. [2021a] introduce another antithetic-sampling-based method to deal with the derivative of the GGX Normal Distribution Function, $D(\omega_h)$ with the half vector ω_h . They exploit the fact that the derivative $\partial_{\omega_h} D(\omega_h)$ is odd about the local shading normal, i.e.

$$\partial_{\omega_h} D([\omega_{h,x}, \omega_{h,y}, \omega_{h,z}]) = -\partial_{\omega_h} D([- \omega_{h,x}, -\omega_{h,y}, \omega_{h,z}]). \quad (4)$$

Their estimator for Eqn. (3) requires two antithetic samples $\omega_{i,1}$ and $\omega_{i,2}$, and is given by

$$I \approx \frac{\partial_\alpha f(\omega_{i,1}) + \partial_\alpha f(\omega_{i,2})}{p(\omega_{i,1}) + p(\omega_{i,2})}. \quad (5)$$

Here, and going forward, we drop ω_o and α from the function arguments of $I(\omega_o, \alpha)$ and $f(\omega_i, \omega_o, \alpha)$ for brevity. This method works well for the odd derivative with ω_h . However, for non-odd derivatives, there are no variance reduction guarantees. Furthermore, several BRDF derivatives are even, e.g., roughness of GGX, Beckmann, and Zhang et al.’s method increases variance in these cases.

Additionally, Eqn. (5) is not in the standard importance sampling form of $\partial_\alpha f/p$ due to the presence of a sum in the numerator and denominator. Hence, it is unclear how to use it in conjunction with multiple importance sampling.

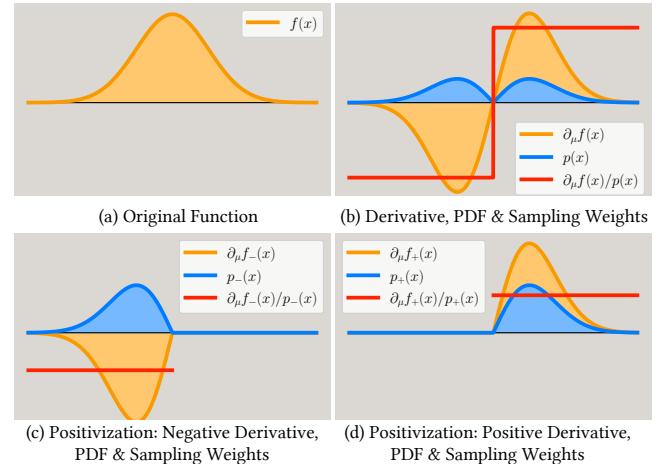


Fig. 2. SIGN VARIANCE AND POSITIVIZATION. Differentiating the positive original function (a, yellow) results in a real-valued derivative (b, yellow). (b) Although the derivative (b, yellow) $\partial_\mu f$ and the PDF (b, blue) p match in shape, i.e., $p \propto |\partial_\mu f|$, the sample weights $\partial_\mu f/p$ (b, red) are non-constant due to a mismatch between their signs, causing sign variance. Positivization [Owen and Zhou 2000] splits the derivative into its positive and negative (c, d, yellow) parts. Since both parts are either purely non-negative or non-positive, they can be perfectly importance sampled by constructing PDFs p_+ and p_- (c, d, blue). The resulting sampling weights (c, d, red), are constant and the corresponding estimator has zero variance.

4 SINGLE-SIGNED DECOMPOSITIONS

In this section, we describe the concept of *sign variance* in real-valued integrals, and then show how our first decomposition, *positivization*, can handle this source of variance for some BRDF derivatives. Positivization requires a) analytic knowledge of roots and b) analytic integrability of the BRDF derivative, which limits its applicability. In Sec. 5, we present a *novel product decomposition* that exploits the single-signed nature of the terms resulting from the product rule for derivatives, for the correct handling of sign variance. It significantly expands the set of BRDF derivatives we can handle. In Sec. 6, we present a *novel mixture decomposition* that exploits the fact that derivatives with mixture weights are a difference of two positive functions, to decompose them into single-signed functions, allowing us to importance sample even more BRDF derivatives. Finally, we describe a general recipe to handle other BRDF derivatives not surveyed in this paper in Sec. 7.

4.1 Sign Variance

We introduce sign variance through the following representative 1D example, showing the real-valued derivative $\partial_\mu f$ of a normal distribution $f(x; \mu, \sigma)$ with its mean μ , as shown in Fig. 2 (a,b):

$$I = \int_{-\infty}^{\infty} \partial_\mu f(x; \mu, \sigma) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^3} (x - \mu) e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx. \quad (6)$$

For $x < \mu$, the integrand $\partial_\mu f$ is negative, and for $x > \mu$, it is positive. The importance sampling strategy using a single PDF $p \propto |\partial_\mu f|$ no longer has zero variance [Owen and Zhou 2000]. This is due to the sign variance, i.e., the positive-valued PDF p cannot match the sign

of the real-valued integrand $\partial_\alpha f$ over the entire domain, leading to non-constant sample weights, see Fig. 2 (b).

4.2 Positivization

It is possible to construct an estimator for any real-valued integrand $\partial_\alpha f$, e.g., Eqn. (6), which has zero variance. By partitioning $\partial_\alpha f$ into its positive $\partial_\alpha f_+$ and negative $\partial_\alpha f_-$ parts,

$$\begin{aligned}\partial_\alpha f_+(x) &= \max(\partial_\alpha f(x), 0), \quad \partial_\alpha f_-(x) = \min(\partial_\alpha f(x), 0) \\ \partial_\alpha f(x) &= \partial_\alpha f_+(x) + \partial_\alpha f_-(x),\end{aligned}\quad (7)$$

we are left with two functions that are single-signed by definition. They can be *perfectly importance sampled* if we can construct the following two PDFs, $p_-(x) \propto \partial_\alpha f_-(x)$ and $p_+(x) \propto \partial_\alpha f_+(x)$, see Fig. 2(c,d). The resulting estimator is

$$\begin{aligned}I &= \int \partial_\alpha f(x) dx = \int \partial_\alpha f_+(x) dx + \int \partial_\alpha f_-(x) dx \\ &\approx \frac{\partial_\alpha f_+(X_+)}{p_+(X_+)} + \frac{\partial_\alpha f_-(X_-)}{p_-(X_-)},\end{aligned}\quad (8)$$

where $X_+ \sim p_+$ and $X_- \sim p_-$.

This technique is called positivization [Owen and Zhou 2000], and we apply it to importance sampling BRDF derivatives. The zero-variance claim is only with regard to the variance arising from the differential BRDF $\partial_\alpha f$. The derivative of the reflection equation, see Eqn. (2), is a product of the differential BRDF and the lighting, and as a result, it will still have variance from the lighting.

In Appendix B, we show how Zeltner et al.’s approach can be understood as a special case of positivization with correlated samples X_+, X_- ; for positivization, these samples are independent. Through an empirical study, we have found that the majority of the variance reduction of antithetic sampling comes from the implicit splitting of $\partial_\alpha f$ into positive and negative lobes ($\partial_\alpha f_+$ and $\partial_\alpha f_-$), instead of the negative correlation between samples, see Fig. 3.

4.2.1 Positivization of Isotropic GGX. Positivization, and by extension antithetic sampling, is very effective at reducing variance for BRDF derivatives when p_+ and p_- can be constructed. For this, p_+ needs to be analytically integrated over the region where $\partial_\alpha f > 0$ to obtain the necessary PDF and CDF required for sampling, see Fig. 4 for the overall pipeline to construct them. This step faces two challenges a) the roots, which define the region where $\partial_\alpha f > 0$, do not have a closed-form expression for some BRDF derivatives, and b) $\partial_\alpha f$ is not analytically integrable over the region where it is positive, for others. A similar argument follows for p_- too.

Some BRDF derivatives, like isotropic microfacet GGX and Beckmann can be handled by positivization. These microfacet models are given by the following equation,

$$f(\omega_i, \omega_o) = \frac{F(\omega_i, \omega_o, \eta)G(\omega_i, \omega_o)D(\omega_h)}{4 \cos \theta_o}, \quad (9)$$

where F is the Fresnel term, G is the shadowing and masking term, and D is the normal distribution function for the specific BRDF. The unit vector ω_h is halfway between ω_i and ω_o , and its spherical coordinates are θ_h, ϕ_h .

The derivative of the isotropic GGX BRDF with its roughness α has two components, $\partial_\alpha D(\omega_h)$ and $\partial_\alpha G(\omega_i, \omega_o)$. However, as noted

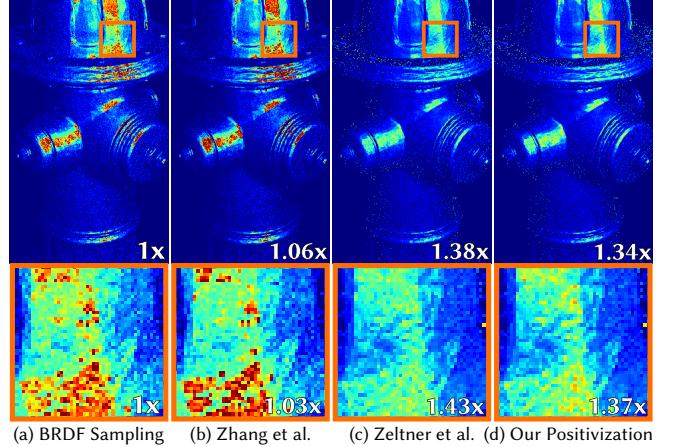


Fig. 3. Comparison between BRDF Sampling, Zeltner et al. [2021], Zhang et al. [2021a], and our Positivization, for the derivative of an isotropic GGX BRDF with its roughness α . The object in the scene is a fire hydrant lit by two area lights. BRDF sampling is unable to correctly handle the sign or shape variance of the differential GGX BRDF and has high variance. Zhang et al.’s method is unsuitable for even derivatives like roughness and produces a high variance estimator as well. Zeltner et al.’s method is a special case of positivization and both of them have very similar variance reduction properties. We have found that positivization and Zeltner et al.’s method have similar performance across several scenes. Hydrant model courtesy of Turbosquid user Adrian Kulawik.

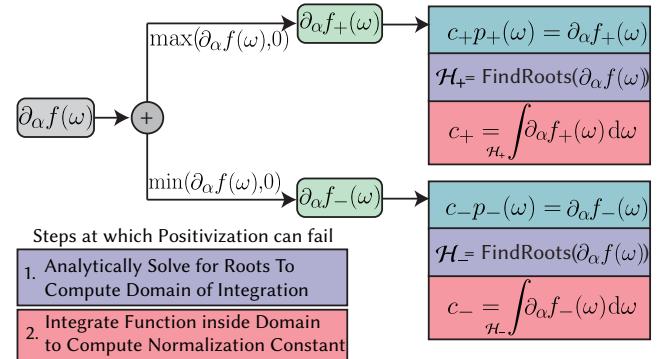


Fig. 4. IMPORTANCE SAMPLING PDF CONSTRUCTION FOR POSITIVIZATION. Positivization splits f about its roots into single-signed f_+, f_- . For PDF construction, we first solve for f ’s roots, which split the domain \mathcal{H} based on f ’s sign into \mathcal{H}_+ and \mathcal{H}_- . Next, we integrate inside the domain to compute the normalization constants. There are two potential roadblocks for positivization i) no analytic form of the roots of f , so \mathcal{H}_+ and \mathcal{H}_- cannot be computed, see Fig. 6, and ii) the positivized functions are not analytically integrable over $\mathcal{H}_+/\mathcal{H}_-$, see Fig. 5.

by previous work [Zeltner et al. 2021; Zhang et al. 2021a], the $\partial_\alpha G$ term only has a minor effect on the overall derivative. Hence, we

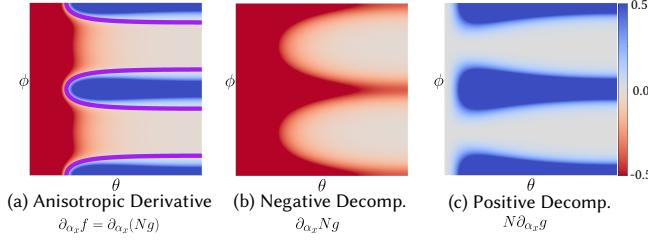


Fig. 5. COMPLICATED ROOTS AND PRODUCT DECOMPOSITION OF THE ANISOTROPIC GGX BRDF DERIVATIVE. (a) The derivative $\partial_{\alpha_x} f$ has complicated roots shown in (a), purple curve. Positivization requires analytical integration over the domain defined by these roots, which we were unable to perform. Our product decomposition separates the two terms resulting from the product rule for differentiation, $\partial_{\alpha_x} f = \partial_{\alpha_x} Ng + N \partial_{\alpha_x} g$, which are both single-signed as shown in (b,c). Product decomposition does not require integration up to the complicated roots, which enables easy PDF construction for both the positive and negative decompositions.

focus on the $\partial_{\alpha} D$ term, which is given by

$$D(\omega_h) = \frac{1}{\pi \alpha^2 \left(\frac{\sin^2 \theta_h}{\alpha^2} + \cos^2 \theta_h \right)^2} \quad (10)$$

$$\partial_{\alpha} D(\omega_h) = \frac{2 \cos^2 \theta_h (\tan^2 \theta_h - \alpha^2)}{\pi \alpha^5 \left(\frac{\sin^2 \theta_h}{\alpha^2} + \cos^2 \theta_h \right)^3}. \quad (11)$$

Its roots have an analytic form and are $\tan \theta_h = \alpha$ for all ϕ_h . Additionally, the derivative $\partial_{\alpha} D$ is analytically integrable over both the positive and negative regions, and so both conditions to apply positivization are met. Hence, the importance sampling PDFs (and CDFs) can be obtained for this derivative.

Apart from the derivatives of the isotropic GGX roughness, positivization can also be used for the derivative of Beckmann BRDFs with their roughness, and Hanrahan-Krueger BRDF with the scattering parameter g of its Henyey-Greenstein phase function.

4.2.2 Inapplicability of Positivization to Anisotropic GGX. For many BRDFs' derivatives, however, one of the two conditions fails, which precludes the use of positivization (and antithetic sampling) for them. For example, consider the derivative of $D(\omega_h)$ of an *anisotropic* GGX BRDF with its roughness α_x ,

$$D(\omega_h) = \frac{1}{\pi \alpha_x \alpha_y \left(\frac{\sin^2 \theta_h \cos^2 \phi_h}{\alpha_x^2} + \frac{\sin^2 \theta_h \sin^2 \phi_h}{\alpha_y^2} + \cos^2 \theta_h \right)^2}$$

$$\partial_{\alpha_x} D(\omega_h) = \frac{\cos^2 \theta_h (3 \tan^2 \theta_h \cos^2 \phi_h - \tan^2 \theta_h \sin^2 \phi_h \alpha_x^2 / \alpha_y^2 - \alpha_x^2)}{\pi \alpha_x^4 \alpha_y \left(\frac{\sin^2 \theta_h \cos^2 \phi_h}{\alpha_x^2} + \frac{\sin^2 \theta_h \sin^2 \phi_h}{\alpha_y^2} + \cos^2 \theta_h \right)^3}. \quad (12)$$

Its roots are the set of (θ_h, ϕ_h) such that the expression $3 \tan^2 \theta_h \cos^2 \phi_h - \tan^2 \theta_h \sin^2 \phi_h \alpha_x^2 / \alpha_y^2 - \alpha_x^2 = 0$, and are shown in Fig. 5 (a), purple curve. However, we were unable to *analytically integrate* the derivative $\partial_{\alpha_x} D$ over the positive and negative strata, see Fig. 5 (a), red and blue regions, which prevented us from applying

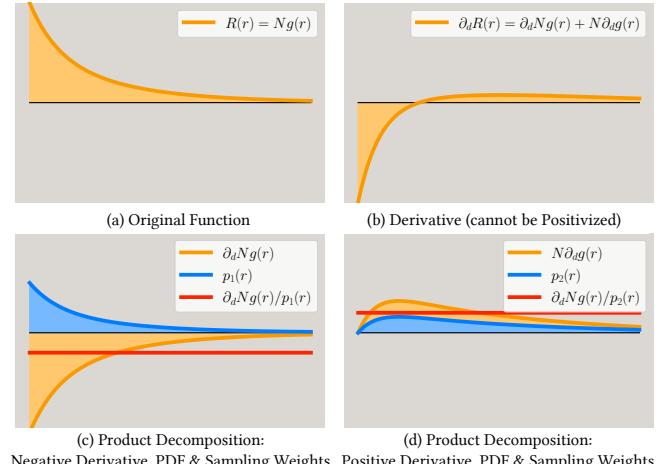


Fig. 6. PRODUCT DECOMPOSITION AND INAPPLICABILITY OF POSITIVIZATION. The BSSRDF profile $R(r, d) = N(d)g(r, d)$ from Burley [2015] (a), is the product of a shape function $g(r, d)$ and normalization term $N(d)$. r is the spatial coordinate, and d is a parameter that controls its width and height. Differentiating with respect to d gives rise to the real-valued derivative $\partial_d R(r, d)$ (b). The derivative's root is given by $3e^{-2r/3d} = (3d - r)/(r - d)$, which has no analytic solution and renders positivization inapplicable to this derivative. However, the derivative can be written as $\partial_d R(r, d) = \partial_d N(d)g(r, d) + N(d)\partial_d g(r, d)$ due to the product rule for derivatives. The first term (c, yellow) is purely negative and the second term (d, yellow) is purely positive. These single-signed functions have no sign variance and can be perfectly importance sampled with the PDFs p_1, p_2 (c,d, blue) leading to constant sample weights (c,d, red).

positivization to this derivative. We were also unable to find closed-form expressions for the roots of the derivatives of other materials like the diffuse BSSRDF from Burley [2015] (Fig. 6) and the isotropic microfacet ABC BRDF, which prevented us from positivizing them.

It is common in BRDF importance sampling to numerically invert a CDF using binary search or Newton iterations, and we will do this with some of our product and mixture decomposition CDFs. However, for positivization, taking a purely numerical approach is not practical. Numerically approximating non-analytic roots and non-analytically integrable PDFs requires storing a high dimensional representation (for e.g. 6D $\omega_i, \omega_o, \alpha_x, \alpha_y$ for anisotropic GGX). Storing such a high dimensional histogram (piecewise approximation) can be infeasible.

Discussion. Positivization is a specific single-signed decomposition that decomposes the real-valued function into a positive and a negative function *with non-overlapping supports*. As a result, it requires root-finding and analytic integration over complicated domains defined by these roots. In the following two sections, we discuss two novel decompositions for which the positive and negative functions are defined over simple domains of integration like a plane or hemisphere, with overlapping support. As a result, they *do not require root finding, or integration over complicated domains*, and as a consequence can handle a broader class of derivatives.

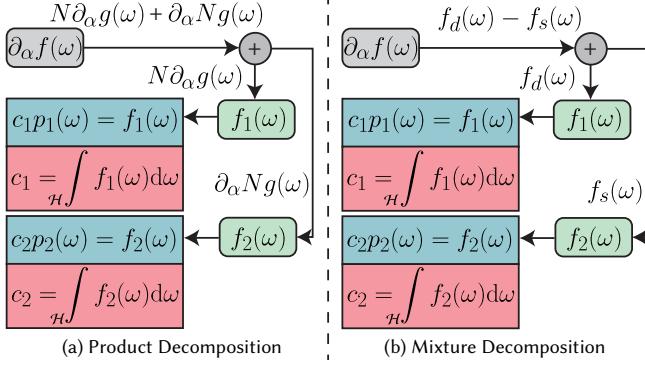


Fig. 7. IMPORTANCE SAMPLING PDF CONSTRUCTION FOR PRODUCT AND MIXTURE DECOMPOSITION. Differentiating the positive-valued functions $f(\omega, \alpha) = N(\alpha)g(\omega, \alpha)$ gives us $\partial_\alpha f(\omega, \alpha) = \partial_\alpha N g(\omega) + N \partial_\alpha g(\omega)$ for product decomposition, and differentiating $f(\omega, \alpha) = \alpha f_d(\omega) + (1 - \alpha) f_s(\omega)$ gives us $\partial_\alpha f(\omega, \alpha) = f_d(\omega) - f_s(\omega)$ for mixture decomposition. For both decompositions, the two terms $f_1(\omega)$ and $f_2(\omega)$ are single-signed. Neither of the decompositions requires complicated root finding and integration up to the roots, unlike positivization, see Fig. 4. Instead, the integration domain is simply the entire hemisphere \mathcal{H} . This makes PDF construction for these decompositions possible for several BRDF derivatives where positivization was not applicable.

5 PRODUCT DECOMPOSITION

Our first new decomposition is product decomposition. It can handle the derivatives of anisotropic microfacet BRDFs, diffuse BSSRDFs, and the isotropic ABC BRDF that positivization could not handle. The key idea we exploit is that after differentiating any of these materials, they split up into two terms following the product rule. Both of these are single-signed, have no sign variance, and are analytically integrable over their simple domains of integration (hemisphere or plane).

Several BRDFs (or normal distribution functions) are of the form

$$f(\omega_h, \alpha) = N(\alpha)g(\omega_h, \alpha), \quad (13)$$

where $g(\omega_h, \alpha)$ is a non-negative shape function, which determines the overall shape of the BRDF over all ω_h , at the parameter value α . $N(\alpha)$ is a directionally constant (independent of ω_h) normalization term that ensures f integrates to 1. Differentiating f with α , we get,

$$\partial_\alpha f(\omega_h, \alpha) = \partial_\alpha N(\alpha)g(\omega_h, \alpha) + N(\alpha)\partial_\alpha g(\omega_h, \alpha). \quad (14)$$

Because N and $\partial_\alpha N$ are directionally constant, the variance in the two terms above comes from g and $\partial_\alpha g$ respectively. The first term above is single-signed because $g \geq 0$. The second term with $\partial_\alpha g$ can potentially be real-valued. However, we have found it to be single-signed for several common BRDFs. For example, for the anisotropic

GGX normal distribution function $D(\omega_h)$, we have,

$$\begin{aligned} D(\omega_h, \alpha_x, \alpha_y) &= N(\alpha_x, \alpha_y)g(\omega_h, \alpha_x, \alpha_y) \\ N(\alpha_x, \alpha_y) &= (\pi\alpha_x\alpha_y)^{-1} \\ g(\omega_h, \alpha_x, \alpha_y) &= \left(\frac{\sin^2 \theta_h \cos^2 \phi_h}{\alpha_x^2} + \frac{\sin^2 \theta_h \sin^2 \phi_h}{\alpha_y^2} + \cos^2 \theta_h \right)^{-2} \\ \partial_\alpha g(\omega_h, \alpha_x, \alpha_y) &= 4g(\omega_h, \alpha_x, \alpha_y)^{3/2} \sin^2 \theta_h \cos^2 \phi_h \alpha_x^{-3}, \end{aligned} \quad (15)$$

where $\partial_\alpha g$ is single-signed, see Fig. 5. Additionally, $\partial_\alpha g$ is also analytically integrable over its hemispherical domain.

Let us provide some geometric intuition for why the shape derivative $\partial_\alpha g$ is often single-signed. For our BRDFs, the parameter α often controls the variance of the distribution, e.g., α_x, α_y for GGX, Beckmann, n_u, n_v for Ashikhmin-Shirley. For all of these, the variance α stretches g horizontally, and increases (or decreases) its value at all locations, making its derivative single-signed. On the other hand, α stretches $N(\alpha)$ vertically to negate the increase (or decrease) in area due to g , and ensure it integrates to 1.

We construct importance sampling PDFs for the two single-signed terms separately, with PDFs $p_1 \propto g$ and $p_2 \propto \partial_\alpha g$,

$$\begin{aligned} I &= \int \partial_\alpha f(\omega_h) d\omega_h \\ &= \int \partial_\alpha N g(\omega_h) d\omega_h + \int N \partial_\alpha g(\omega_h) d\omega_h \\ &\approx \frac{\partial_\alpha N g(\omega_{h,1})}{p_1(\omega_{h,1})} + \frac{N \partial_\alpha g(\omega_{h,2})}{p_2(\omega_{h,2})}. \end{aligned} \quad (16)$$

Fig. 7 (a) describes the pipeline to generate importance sampling PDFs for product decomposition. Product decomposition can handle the derivatives of anisotropic GGX, Beckmann, Ashikhmin-Shirley which are not analytically integrable over the positivized domains, and Burley's diffuse BSSRDF and the isotropic ABC BRDF, which have no closed-form solution for the roots. However, they all have single-signed $\partial_\alpha g$ which is analytically integrable.

Note that the product rule in and of itself does not guarantee a single-signed decomposition. For example, the product of the microfacet distribution (D) and geometric terms (G) does not lead to a single-signed decomposition for the derivative with α_x (or α_y). This is because both $\partial_{\alpha_x} D$ and the $\partial_{\alpha_x} G$ terms are real-valued. The decomposition $D = Ng$ is one of the many product decompositions, but the only one we found to preserve the single-signed property.

6 MIXTURE DECOMPOSITION

Our second new decomposition further expands the set of BRDF derivatives we can handle. Consider, for example, a BRDF made up of a diffuse f_d and specular f_s lobe with scalar mixture weights k_d and $1 - k_d$ respectively.

$$\begin{aligned} f(\omega_i, \omega_o) &= k_d f_d(\omega_i, \omega_o) + (1 - k_d) f_s(\omega_i, \omega_o) \\ \partial_{k_d} f(\omega_i, \omega_o) &= f_d(\omega_i, \omega_o) - f_s(\omega_i, \omega_o). \end{aligned} \quad (17)$$

The derivative with the mixture weight k_d is positive when the diffuse lobe contribution is higher than the specular lobe and negative otherwise. In general, this derivative is very hard to positivize,

because f_d and f_s can be arbitrary BRDFs, and so the roots of $f_d - f_s$ are unlikely to have a simple analytic form.

However, we can once again decompose this derivative into single-signed functions with overlapping support; we refer to this as the mixture decomposition. Since f_d and f_s are non-negative valued BRDFs, they are single-signed, and can be importance sampled separately with appropriate PDFs p_d and p_s .

$$\begin{aligned} I &= \int \partial_{k_d} f(\omega_i) d\omega_i = \int f_d(\omega_i) d\omega_i - \int f_s(\omega_i) d\omega_i \\ &\approx \frac{f_d(\omega_{i,d})}{p_d(\omega_{i,d})} - \frac{f_s(\omega_{i,s})}{p_s(\omega_{i,s})}. \end{aligned} \quad (18)$$

Mixture weights show up in all *Uber* BRDFs, like the Autodesk Standard Surface, Disney BRDF, etc., and our mixture decomposition can be applied to all of them.

Mixture decomposition is also applicable to the derivative of BRDFs that are not explicitly mixture models, but internally are made up of different lobes, with parametric weights. For example, the Oren-Nayar BRDF, which is a linear combination of two terms. Here, the positive weights $A(\sigma), B(\sigma)$ depend upon the roughness σ of the BRDF.

$$\begin{aligned} f(\omega_o, \omega_i) &= A(\sigma) \frac{\rho}{\pi} \cos \theta_i \\ &\quad + B(\sigma) \frac{\rho}{\pi} \max(0, \cos(\phi_i - \phi_o)) \sin \alpha \tan \beta \cos \theta_i, \end{aligned} \quad (19)$$

where $\alpha = \max(\theta_i, \theta_o)$, $\beta = \min(\theta_i, \theta_o)$. Once again, since both terms of the BRDF above are positive, the real-valued derivative with σ is simply the sum of a positive and a negative term,

$$\begin{aligned} \partial_\sigma f(\omega_o, \omega_i) &= \partial_\sigma A(\sigma) \frac{\rho}{\pi} \cos \theta_i \\ &\quad + \partial_\sigma B(\sigma) \frac{\rho}{\pi} \max(0, \cos(\phi_i - \phi_o)) \sin \alpha \tan \beta \cos \theta_i, \end{aligned} \quad (20)$$

with the sign of the term decided by the sign of $\partial_\sigma A$ and $\partial_\sigma B$. Importance sampling the first term is simply cosine-hemispherical sampling, and we provide an importance sampling PDF for the second term in Appendix A.3.2. Besides Oren-Nayar, the microcylinder BRDF [Sadeghi et al. 2013] is also a mixture model with weights $k_d, 1 - k_d$, where k_d is the isotropic scattering coefficient, and can be handled by mixture decomposition as well.

7 RECIPE FOR IMPORTANCE SAMPLING BRDF DERIVATIVES

We now present a *recipe* to importance sample BRDF derivatives based on the key ideas introduced in the previous sections.

Step 1, Positivization. Given a real-valued BRDF derivative $\partial_\alpha f$, check if it can be positivized. For positivization to be applicable, $\partial_\alpha f$ should have analytic roots. Compute the normalization constants for the solid angle PDFs $p_+(\omega_i) \propto \max(\partial_\alpha f(\omega_i), 0)$, $p_-(\omega_i) \propto \min(\partial_\alpha f(\omega_i), 0)$, and their marginal and conditional counterparts $p_+(\phi_i), p_-(\phi_i), p_+(\theta_i|\phi_i), p_-(\theta_i|\phi_i)$, if they are analytically integrable. See Fig. 4 for the PDF generation, and Eqn. (8) for the estimator.

Step 2, Try Product or Mixture Decomposition. If positivization is inapplicable for either reason (no analytic roots or lack of analytic

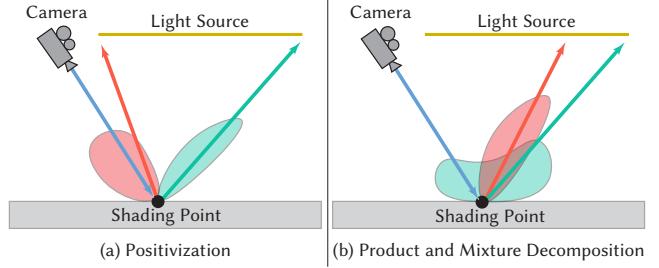


Fig. 8. Positivization, Product and Mixture Decomposition for direct illumination. All three techniques send out two shadow rays corresponding to two different sampling techniques at each shading point, shown by red and green arrows. For positivization, the PDFs for these sampling techniques have non-overlapping support, shown by the red and green lobes. For mixture and product decomposition, however, the corresponding PDFs may have overlapping support.

integrability), either product or mixture decomposition might still be applicable.

Step 2.1, Product Decomposition. If the original BRDF is of the form $N(\alpha)g(\omega_i, \alpha)$, where α appears in a directionally invariant (independent of ω_i) normalization term $N(\alpha)$ and an unnormalized shape function $g(\omega_i, \alpha)$, product decomposition may be applicable. First check if $\partial_\alpha g$ is single-signed, i.e., it has a constant sign for all ω_i , and is analytically integrable. If these conditions hold, product decomposition is applicable. Construct a PDF $p_2(\omega_i) \propto \partial_\alpha g$ and compute the normalization terms for it and its conditional and marginal counterparts. The other PDF $p_1(\omega_i) \propto g$ is simply the BRDF sampling PDF. See Fig. 7 for the PDF generation, and Eqn. (16), for the estimator.

Step 2.2, Mixture Decomposition. If instead the parameter α appears in the form of linear combination weights either explicitly as a mixture model between two BRDFs, or implicitly as a mixture between two lobes that form a single BRDF, mixture decomposition is likely applicable here. In this case, simply use the PDFs and sampling strategies most suitable for the two mixture lobes if they are available (e.g., visible normal distribution function sampling for a GGX lobe), or construct PDFs $p_1(\omega_i) \propto f_1(\omega_i)$, $p_2(\omega_i) \propto f_2(\omega_i)$ for the two lobes, where f_1, f_2 are the two lobes. See Fig. 7 for the PDF generation, and Eqn. (18) for the estimator.

Fig. 8 depicts the estimators for all three of our decompositions for direct illumination. They all require two shadow rays at the shading point, corresponding to the positive and negative lobes of the corresponding decomposition.

Although we have not found examples that require it, our three decompositions can also be interleaved with one another for complicated BRDF derivatives. For example, it is possible that for some BRDF derivatives, the derivative of the shape function from the product rule $\partial_\alpha g$ could be real-valued. It could then further be positivized to eliminate sign variance.

Forward Rendering Sampling Technique Reuse. Both product and mixture decomposition reuse BRDF sampling developed for forward rendering as one (or both) of the techniques for differential

BRDF sampling. For product decomposition, this corresponds to $p_1 \propto g$. For mixture decomposition, perfect importance sampling can be achieved by *only* employing two standard BRDF sampling techniques from forward rendering in some cases. BRDF sampling when used directly to estimate for $\partial_\alpha f$ suffers from sign and shape variance, however, when paired with the right decomposition, it can correctly handle the shape variance of one of the terms.

Multiple Importance Sampling. For the product and mixture decompositions, the positive and negative decomposition PDFs can have overlapping support (for positivization they are necessarily non-overlapping). As a result, the samples generated for one decomposition can be shared with the other using Multiple Importance Sampling. Also, all three of our decompositions reduce the variance from the differential BRDF $\partial_\alpha f$, and can be used in conjunction with light source sampling via Multiple Importance Sampling to reduce the lighting, L_i 's variance.

8 RESULTS

We organize our results into two subsections. First, we demonstrate that our decompositions do reduce variance in practice for a number of BRDF derivatives under a wide variety of lighting conditions in Sec. 8.1. Next, we demonstrate that lower variance in gradients indeed does enable better spatially-varying texture recovery in inverse rendering, in Sec. 8.2.

Implementation Details. We implemented all the different decompositions and BRDFs on our own CPU-based differentiable renderer, using the Embree [Wald et al. 2014] library for ray tracing. At each shading point, all three of our decompositions require two shadow rays, see Fig. 8. To have a fair comparison with BRDF sampling, we shoot out two shadow rays at each shading point for it too, which ensures an equal-ray comparison with our method. Since our sampling routines have similar timings to BRDF importance sampling (Table 2), all comparisons are also equal time and we report the timings for each experiment in insets. All our standard deviation comparison images are computed by taking the square root of the averaged squared error of the gradient images, which were each generated at 9 samples per pixel over 50 runs. The variance improvement can be easily computed by squaring all the standard deviation improvement numbers in the insets. Additionally, we also provide difference images to show the specific regions where our estimators outperform BRDF sampling in Fig. 29.

Sample Generation Timings. We report the average times (over 1,000,000 runs) to generate samples using our routines and baselines in Table 2. Our routines involve CDF inversion, which we perform using a combination of Newton iterations and bisection search as described by Yuksel et al. [2022] (maximum 64 iterations, 10^{-6} CDF error tolerance). Our sampling routines only add a small overhead in the overall path-tracing pipeline, see timings in Figs. 10 to 13 and 18 to 23.

Convergence Rates. Our estimators are intended to be used within inverse rendering pipelines. This necessitates low sample counts per pixel (1 to 16) for fast iteration times. For these sample budgets,

Table 2. SAMPLE GENERATION TIMINGS FOR OUR SINGLE SIGNED DECOMPOSITIONS (SSD). The timings are to generate a single sample, averaged over 1,000,000 runs. The first two rows include timings for BRDF importance sampling for cosine and GGX BRDFs for reference; our routines add negligible overhead in the overall run time in a differentiable path tracer, see Figs. 10 to 13 and 18 to 23.

Material	SSD	Time	Method
Cosine	Fwd.	$0.64\mu s$	Analytic
GGX [Heitz 2018]	Fwd.	$2.25\mu s$	Analytic
Anisotropic GGX (ours)	Prod.	$5.97\mu s$	CDF Inv.
Anisotropic Beckmann (ours)	Prod.	$6.32\mu s$	CDF Inv.
Anisotropic Ashikhmin-Shirley (ours)	Prod.	$4.93\mu s$	CDF Inv.
Oren-Nayar (ours)	Mix.	$1.97\mu s$	CDF Inv.
Hanrahan-Krueger (ours)	Pos.	$9.04\mu s$	CDF Inv.
Isotropic GGX [Zeltner et al. 2021]	Pos.	$0.91\mu s$	Analytic

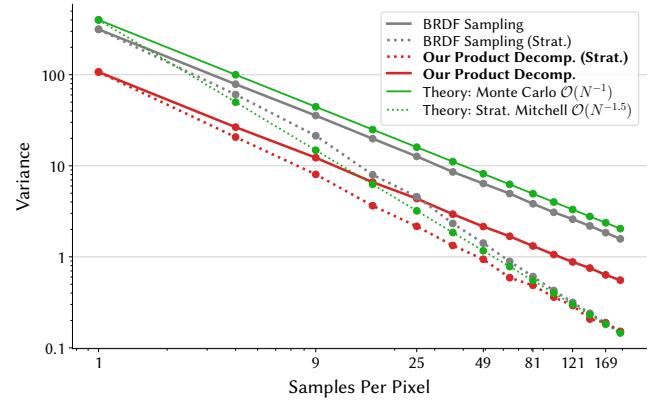


Fig. 9. CONVERGENCE RATE FOR PRODUCT DECOMPOSITION. Product Decomposition improves upon BRDF Sampling in the low sample regime (1-16 samples) with and without stratification. With stratification, at high sample counts, both estimators converge to the same rate since the variance is dominated by discontinuities in the integrand [Mitchell 1996]. Additionally, in this regime, as the strata get smaller, most of them become single-signed; in effect, this reduces (but does not eliminate) sign variance for BRDF sampling. Inverse rendering operates in the low sample count regime to enable fast iteration times, so the variance reduction provided by estimators (even with stratification) plays an important role here.

our estimators outperform baselines both without stratification and with it (Fig. 9, all other results in the paper are with stratification).

Without stratification, the estimators show the typical $O(N^{-1})$ convergence. Stratification improves the convergence rate. Interestingly, with stratification, standard BRDF importance sampling performs as well as our methods at high sample counts (100+).

Mitchell [1996] provides a possible explanation: at high sample counts with stratification, variance is dominated by visibility discontinuities in the integrand, so these pixels have convergence rates of $O(N^{-1.5})$. Better importance sampling improves the convergence rate for the continuous differential BRDF, noticeable at low sample counts. At higher sample counts, visibility dominates the variance which none of the techniques handle. Also, BRDF sampling (which

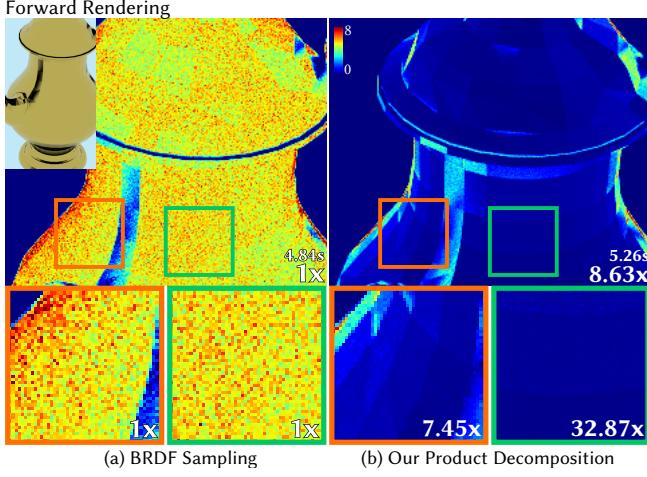


Fig. 10. PRODUCT DECOMPOSITION V.S. BRDF SAMPLING UNDER CONSTANT ILLUMINATION. We show the estimated standard deviation of derivatives. Numbers indicate the relative improvement, **higher is better**. The scene contains an anisotropic Beckmann BRDF under constant environment illumination. Under constant illumination, the BRDF derivative is the main source of variance. Our product decomposition correctly handles both the sign and shape variance, because of which we see an overall $8.63\times$ reduction in standard deviation compared to BRDF sampling. 3D model courtesy of Turbosquid user id_inc.

does not handle sign variance) benefits from stratification at high sample counts since smaller strata are increasingly single-signed. Nevertheless, our methods are able to provide significant variance reduction (by nearly an order of magnitude) in the low sample regime, most relevant to us. This phenomenon is a general observation for importance sampling, not limited to differentiable rendering.

8.1 Derivative Comparison

8.1.1 Positivization. First, we compare positivization with BRDF sampling for the derivative of two BRDFs in Fig. 1. The scene is lit by two area lights. The isotropic GGX teapot (with $\alpha = 0.02$) is differentiated with its roughness α , and the Hanrahan-Krueger (with $g = -0.9$) lion is differentiated with its Henyey-Greenstein parameter for anisotropy g . The Henyey-Greenstein phase function at $g = -0.9$ is highly back-scattering and is very badly importance sampled by regular BRDF sampling, which cannot correctly account for the highly peaked and signed nature of the derivative. Since positivization is correctly able to handle both sign and shape related variance, we see significant standard deviation reduction of $1.96\times$ and $58.57\times$ for the teapot and lion respectively.

8.1.2 Product Decomposition. Next, we compare product decomposition with BRDF sampling for the derivative of an anisotropic Beckmann BRDF with its roughness α_x , lit under constant environment illumination in Fig. 10. Positivization (and by extension Zeltner et al.) cannot handle this derivative, see Sec. 4.2.2, and Zhang et al.’s method fails for even derivatives like this one, see Fig. 26. Constant illumination eliminates variance from lighting and only

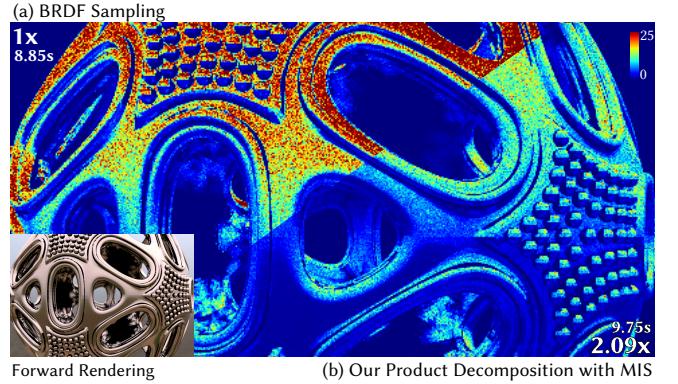


Fig. 11. PRODUCT DECOMPOSITION FOR THE ANISOTROPIC BECKMANN BRDF. In this scene, we demonstrate that product decomposition (with MIS across the two techniques) can significantly reduce the standard deviation ($2.09\times$) for the estimation of the anisotropic roughness derivative $\partial_\alpha D$ under (non-constant) environment illumination and complex geometry. The object exhibits inter-occlusions due to its highly non-convex geometry leading to additional visibility variance. 3D model courtesy of Turbosquid user Sonic_art.

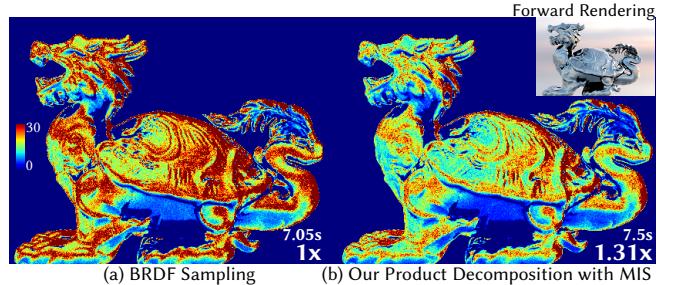


Fig. 12. PRODUCT DECOMPOSITION FOR THE ANISOTROPIC GGX BRDF. Our product decomposition (with multiple importance sampling across the two techniques) is also effective in reducing the standard deviation of the derivative of the anisotropic roughness for the GGX BRDF ($1.31\times$). Here, we showcase a scene lit under environment lighting where the object exhibits intricate geometry with several inter-occlusions. 3D model courtesy of Turbosquid user Shef2oo.

keeps variance from the BRDF derivative and visibility. Since product decomposition can correctly handle both the sign and shape variance of the BRDF derivative, it has an overall $8.63\times$ reduction in standard deviation, whereas BRDF sampling fails because it cannot handle either source of variance. In most regions (Fig. 10 see right inset), the derivative of the normal distribution function $\partial_\alpha D$ is the major source of BRDF derivative variance; we eliminate it and see a big improvement of $32.7\times$. However, in the grazing angle regions (Fig. 10 see left inset), the derivative of the shadowing function $\partial_\alpha G$ dominates. Here, our improvement is still significant ($7.45\times$), but relatively less pronounced, since our sampling strategy minimizes $\partial_\alpha D$ ’s variance.

Now, we change the lighting to realistic environment illumination (see insets for forward renderings), and the geometry to objects

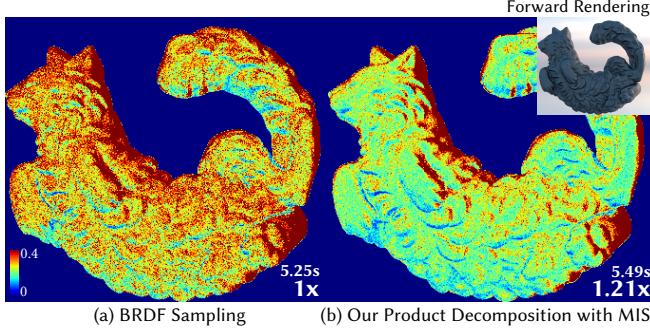


Fig. 13. PRODUCT DECOMPOSITION FOR THE ASHIKHMIN-SHIRLEY BRDF. We also demonstrate the benefits of our product decomposition in estimating the derivatives with the anisotropic exponent n_u for the Ashikhmin-Shirley BRDF; it reduces standard deviation by $1.21\times$. The cat sculpture is quite non-convex and the lighting is environment illumination which leads to additional sources of variance (from visibility and lighting) apart from the differential BRDF. 3D model courtesy of Turbosquid user Skazok.

with intricate geometry in Figs. 11 to 13; we estimate the derivative of the anisotropic Beckmann, GGX and Ashikhmin-Shirley BRDFs in these experiments respectively. Apart from BRDF derivative variance, these scenes have two other major sources of variance, lighting and visibility. When the variance is significant from other sources too, we have found that sharing samples between the positive and negative decomposition is beneficial, see Sec. 7, Multiple Importance Sampling (MIS). In all three experiments, we have found that our product decomposition is able to outperform BRDF importance sampling and achieve lower standard deviation in gradient estimation $2.09\times$, $1.31\times$ and $1.21\times$ respectively.

We show two more examples of product decomposition in Fig. 1, for anisotropic GGX and Beckmann BRDF derivatives, which achieve standard deviation reductions of $1.56\times$ and $3.61\times$ respectively. The insets in the top row of Fig. 1 show the regions where our decomposition has lower variance than BRDF sampling in blue. Product decomposition outperforms BRDF sampling in almost all regions.

We provide additional comparisons with Zhang et al. for the anisotropic roughness derivatives of the GGX BRDF for the scenes modelled in Figs. 10 and 12, in fig. 26; their estimator has higher standard deviation than BRDF importance sampling. We emphasize that this behavior is expected, since they designed their estimator to estimate odd derivatives (for example, the derivative with the half vector), but the roughness derivative is even.

8.1.3 Mixture Decomposition. Finally, we compare BRDF sampling with Mixture Decomposition to estimate the derivative of a mixture model with its mixture weight for the fish-shaped pot in Fig. 1. The mixture model is a linear combination of a Lambertian diffuse lobe, and a GGX specular lobe and the lighting is two area lights. Mixture decomposition can reduce the standard deviation by $4.72\times$, because it correctly handles shape and sign variance, unlike BRDF sampling.

Fig. 1 also shows an example of an Oren-Nayar pot, and its derivative with the roughness σ . BRDF sampling here is simply cosine hemispherical sampling, and works quite well in the central regions of the pot, because the cosine lobe is dominant in the non-grazing

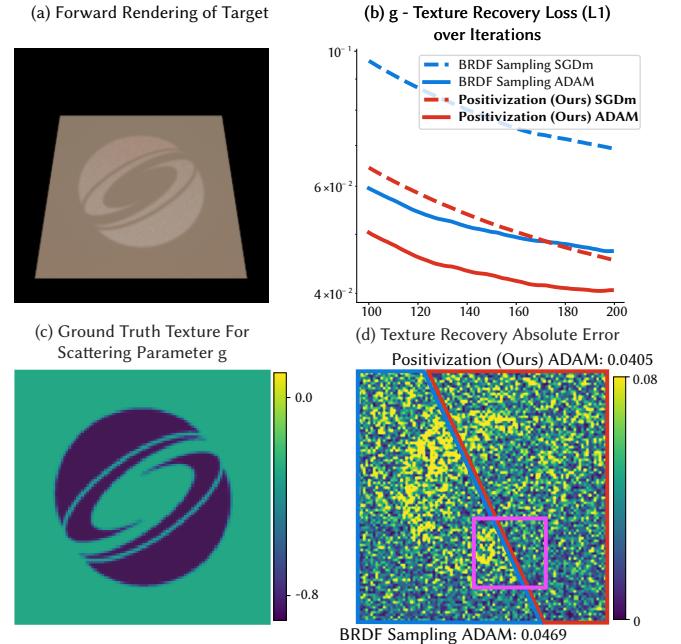


Fig. 14. INVERSE RENDERING OF THE SCATTERING PARAMETER g OF A HANRAHAN-KRUEGER BRDF. (a) forward rendering of the target. Our positivization has a lower texture recovery error than BRDF sampling under two setups, one using the ADAM optimizer and the other using SGD with momentum (SGDm) (b). For the SGDm experiments, we manually tune the learning rate to the highest possible value without the loss diverging. Positivization admits a higher learning rate ($lr = 1$) since its gradients have lower variance than BRDF sampling ($lr = 0.5$), see Fig. 28 for the final SGDm results. Similar to [Nimier-David et al. 2022], we find that the difference between our improved estimator (positivization) and the baseline estimator (BRDF sampling) is more pronounced when SGDm is used. (c) shows the ground truth texture, with an inset of the recovered texture using positivization. (d) shows the error images for positivization and BRDF sampling, both using the ADAM optimizer. BRDF sampling is unable to recover the texture in the highly backscattering logo region (see purple box). However, positivization can handle the logo region well.

angle regions, see Eqn. (19). However, in the grazing angle regions towards the edges of the pot where the correction term is more dominant, BRDF sampling breaks down and has high variance. On the other hand, our mixture decomposition with MIS correctly accounts for the derivative of both terms with regard to their sign and shape variance, and can achieve low variance in *all* regions of the pot, and leads to a $3.91\times$ reduction in standard deviation.

8.2 Inverse Rendering Comparison

We demonstrate the benefits of correctly handling sign variance in gradients, for gradient-descent-based inverse rendering. We apply inverse rendering to the task of spatially varying texture recovery, and evaluate the effectiveness of all three of our decompositions on it. Our results for positivization are presented in Fig. 14, product decomposition in Fig. 15, and mixture decomposition in Fig. 16. All

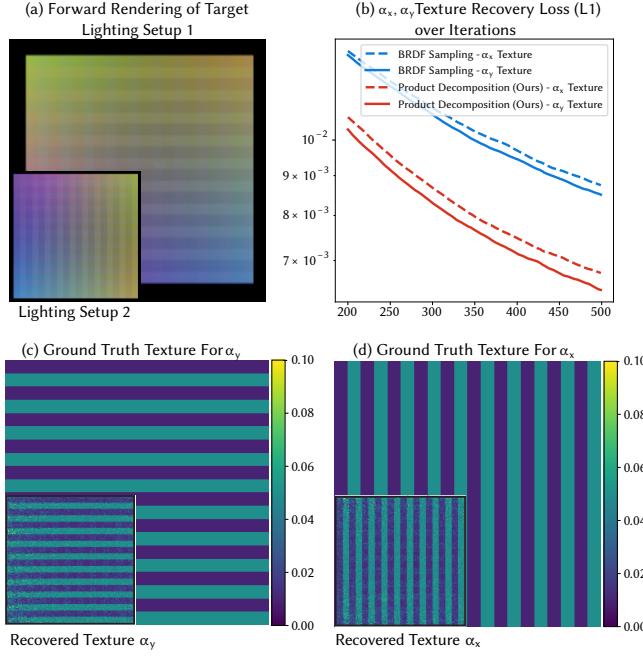


Fig. 15. INVERSE RENDERING OF THE ROUGHNESS OF AN ANISOTROPIC BECKMANN PLATE UNDER A PHOTOMETRIC STEREO SETUP. The forward rendering of the target under the two lighting setups is shown in (a) and its inset. The texture recovery loss (b) demonstrates that our product decomposition achieves lower texture recovery error than BRDF sampling for both the α_x and α_y textures. (c) and (d) show the ground truth textures for α_x and α_y , and the insets show our recovered textures.

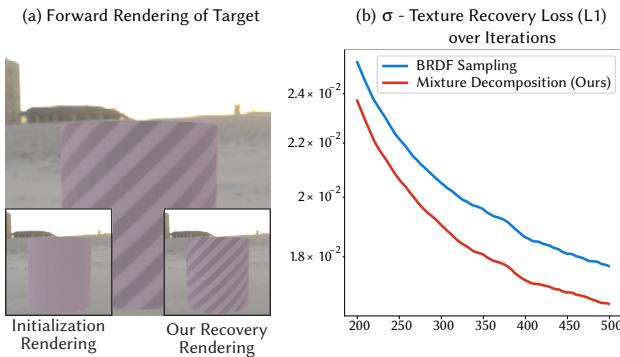


Fig. 16. INVERSE RENDERING OF THE ROUGHNESS σ OF AN OREN-NAYAR BRDF. (a) forward rendering of the target, insets show a rendering of the initialization and our final recovered texture. (b) Our mixture decomposition correctly deals with sign variance and has lower variance in gradients, and as a result has lower texture recovery error.

our inverse rendering results use 4 samples per pixel for both forward and gradient rendering at each optimization iteration. We use the ADAM optimizer [Kingma and Ba 2015] and the respective loss graphs show the mean absolute texture recovery error (L_1) after some initial iterations. In Fig. 14, we show that using the ADAM

optimizer instead of SGD with momentum (SGDm) reduces the impact of gradient outliers, similar to Nimier-David et al.’s observation [2022]. Nonetheless, our estimators outperform the baselines in all setups.

For positivization (Fig. 14), we recover the (spatially varying) scattering parameter g of a Hanrahan-Krueger BRDF with the semi-infinite depth assumption, lit by a single area light. The ground truth texture consists of a slightly back-scattering background region with $g = -0.3$, and a highly back-scattering logo region with $g = -0.9$, see Fig. 14 (c). We initialized the scattering parameter g with a random negative number. Positivization consistently has lower texture recover error compared to BRDF sampling, especially in the highly back-scattering region.

For product decomposition (Fig. 15), we optimize the spatially varying anisotropic roughness textures (α_x and α_y) of a Beckmann BRDF under a photometric stereo setup under two illumination conditions. The two lighting conditions are rotated versions of the same environment map. Starting from a random initialization for both textures, product decomposition’s correct handling of the sign variance leads to a gradient estimator with lower overall variance, and consequently ensures lower texture recovery error across all iterations, as shown in Fig. 15 (b). The final recovery is displayed in Fig. 15 (c),(d).

Our product decomposition computes the gradients for both roughness values using three samples at each shading point combined using multiple importance sampling (one each from $p_1, p_{2,x}, p_{2,y}$). To ensure an equal-ray budget, we use three samples for BRDF sampling at each shading point too.

For mixture decomposition in Fig. 16, we recover the spatially varying roughness of an Oren-Nayar BRDF under environment map illumination. Once again, mixture decomposition benefits from lowered variance in gradients, and can recover a texture with lower error than BRDF sampling at an equal ray-triangle intersection budget, see Fig. 16 (b).

9 GLOBAL ILLUMINATION

We now describe how to importance sample BRDF derivatives under multiple bounce global illumination. The recursive rendering equation [Kajiya 1986] (ignoring emission) is given by a generalization of Eqn. (1),

$$L_r(\mathbf{y}, \omega_o; \alpha) = \int f(\mathbf{y}, \omega_i, \omega_o; \alpha) L_r(z, -\omega_i; \alpha) d\omega_i, \quad (21)$$

where we have substituted the incoming radiance $L_i(\mathbf{y}, \omega_i)$, with the outgoing/reflected radiance $L_r(z, -\omega_i; \alpha)$, and $z = \text{rayTrace}(\mathbf{y}, \omega_i)$ is the first intersection point from \mathbf{y} in the direction ω_i . The recursive call of L_r is a function of the BRDF parameter α because upon unrolling the recursion, it may be a function of an α dependent BRDF. Differentiating this expression, we get,

$$\partial_\alpha L_r(\mathbf{y}, \omega_o; \alpha) = \int \partial_\alpha f(\mathbf{y}, \omega_i, \omega_o; \alpha) L_r(z, -\omega_i; \alpha) d\omega_i \quad (22)$$

$$+ \int f(\mathbf{y}, \omega_i, \omega_o; \alpha) \partial_\alpha L_r(z, -\omega_i; \alpha) d\omega_i, \quad (23)$$

which recursively describes how differential radiance is reflected. The two integrals (Eqn. (22) and (23)) can be importance sampled

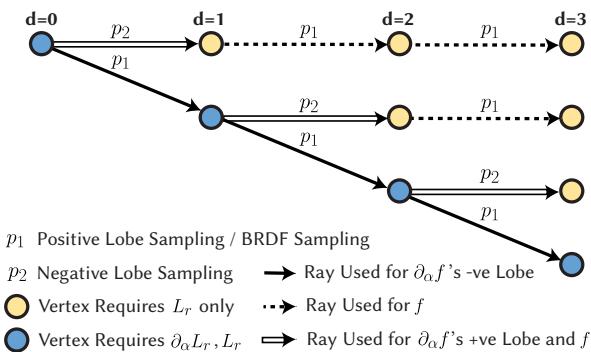


Fig. 17. BRANCHING FOR GLOBAL ILLUMINATION WITH MAX DEPTH=3 FOR PRODUCT & MIXTURE DECOMPOSITION. Starting from a single vertex at $d = 0$, estimating $\partial_\alpha L_r$ requires recursive estimation of both L_r and $\partial_\alpha L_r$ via Eqns. (22) and (23). For product & mixture decomposition, since one new branch is created at every vertex, the total number of rays used here is quadratic i.e. $O(d^2)$, in the maximum depth d . The sampling technique p_1 is the same for BRDF sampling and the positive lobe of the BRDF derivative, and p_2 is the sampling technique for the negative lobe. For positivization, two new branches are created at each vertex corresponding to p_+, p_- , however, once created, they only require evaluation of L_r , and do not further branch out, so the complexity is still $O(d^2)$.

separately. We have seen how to importance sample Eqn. (22) by applying different BRDF derivative decompositions in Sections 4.2, 5, and 6. Irrespective of the decomposition required, this requires two evaluations of L_i corresponding to the positive and negative lobes and is done by regular path tracing (similar to the standard splitting approach [Arvo and Kirk 1990]). To importance sample Eqn. (23), we follow standard BRDF sampling and continue the same recursive importance sampling of $\partial_\alpha L_r$ at the next shading point.

This means that we need three samples at each shading point, one each for BRDF, positive lobe and negative lobe importance sampling. Fortunately, for product and mixture decomposition, we can reduce this to two samples at each shading point. For product decomposition, as we saw in Sec. 5, one of either the positive or negative lobe decomposition PDFs is the same as BRDF sampling, and can share a sample with it. For mixture decomposition, BRDF sampling can be simulated by randomly choosing a sample from either the positive or negative lobes with the probability equal to the mixture weight of the BRDF sampling strategy.

Branching Complexity and Comparison with BRDF sampling. Even though we use two samples to estimate Eqn. (22), the total number of rays required to estimate $\partial_\alpha L_r$ for a maximum depth d is quadratic i.e., $O(d^2)$, instead of exponential, see Fig. 17, whereas it is $O(d)$ for BRDF sampling. This is because we only apply splitting when estimating Eqn. (22), which recurses on L_r , and we do not split when estimating Eqn. (23). The recursive call of L_r in Eqn. (22) does not require splitting, which prevents exponential branching.

Variance Reduction under Global Illumination. All of our results are subject to an equal-ray budget setup to ensure fair comparisons. We also include wall clock times that reflects this.

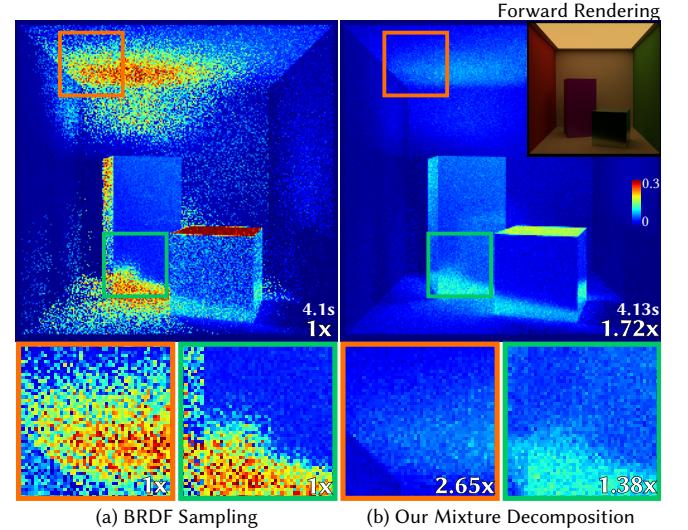


Fig. 18. Comparison between BRDF Sampling and Mixture Decomposition under 1 bounce Global Illumination and equal ray-triangle intersection budget. The mixture BRDF is given by $f = wf_d + (1 - w)f_s$ with $w = 0.1$ for the left box and $w = 0.9$ for the right box. We estimate the derivative with the weight w . f_d is a Lambertian diffuse lobe, and f_s is an isotropic GGX lobe with $\alpha = 0.05$. Mixture decomposition leads to lower variance in most regions since it correctly deals with sign variance. However, in some regions, e.g., the green inset, the lighting, and visibility variance is more significant, and our improvements in this region are less prominent.

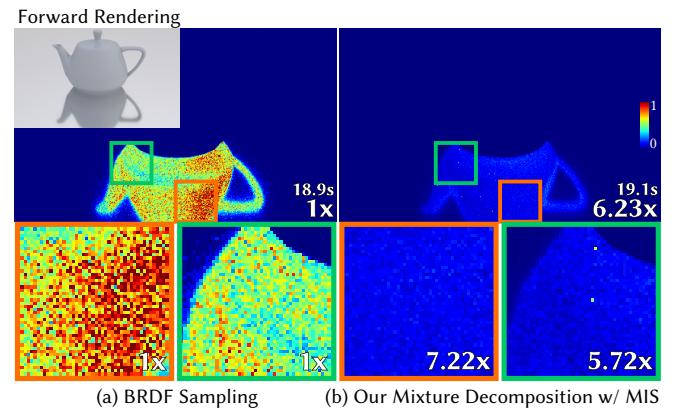


Fig. 19. DERIVATIVE OF GLOSSY REFLECTION. We estimate the derivative of a glossy reflection of a teapot with the teapot's mixture weight (100 ray budget). The teapot is modelled by a BRDF with a GGX lobe ($\alpha = 0.05$) and a diffuse lobe, with the weight of the GGX lobe $w = 0.1$ and the glossy plate by a GGX BRDF with $\alpha = 0.01$. Our estimator produces derivatives with significantly lower variance by performing mixture decomposition (with MIS across the two techniques) at the second path vertex (on the teapot) which reduces both sign and shape variance of the differential BRDF. 3D model courtesy of Benedikt Bitterli.

We have found that for one-bounce global illumination, our mixture decomposition can reduce standard deviation by 1.55×, see Fig. 18. Next, we show results for gradient estimation in two

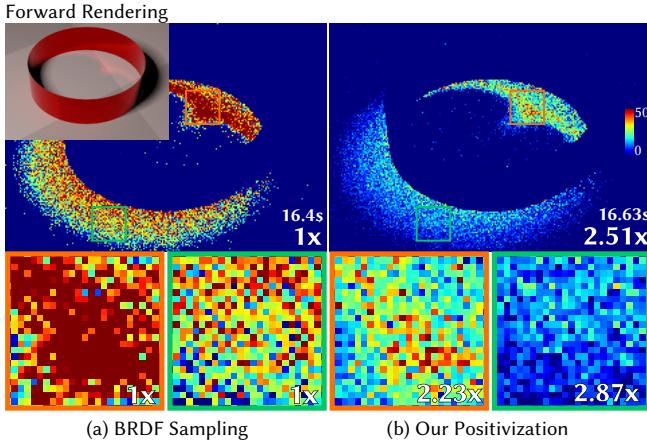


Fig. 20. CAUSTIC DERIVATIVE. We estimate the derivative of a caustic pattern formed on a diffuse surface by a metallic cylinder modelled by an isotropic GGX BRDF with $\alpha = 0.01$ (100 ray budget). Despite the fact that positivization branches out into three paths at the second path vertex (on the cylinder), it is able to produce much cleaner gradients with lower variance since it can correctly account for the differential BRDF’s sign and shape variance on the cylinder.

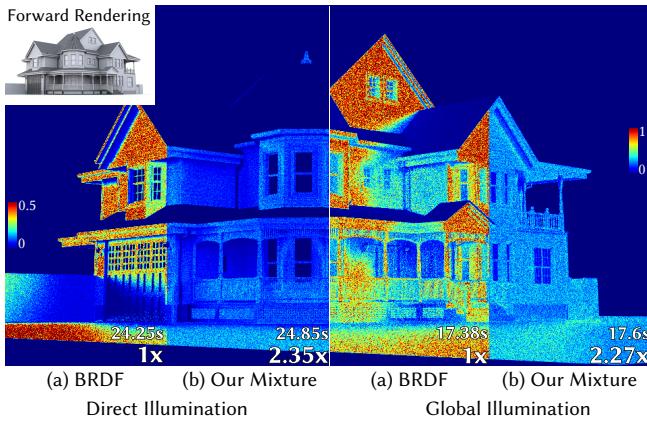


Fig. 21. MIXTURE WEIGHT DERIVATIVE ESTIMATION FOR AN OUTDOOR SCENE UNDER DIRECT AND ONE BOUNCE GLOBAL ILLUMINATION. All surfaces (except the roof) are modelled by a mixture BRDF (between a diffuse and a GGX lobe with the weight of the specular lobe $w = 0.9$), we estimate the derivative with w . Both our direct and global illumination mixture decomposition estimators (with MIS across the two techniques) significantly reduce the standard deviation of the estimated gradients as compared to BRDF sampling at an equal computation budget (9 samples for DI and 25 rays for GI). 3D model courtesy of Blendswap user MrChimp2313.

typical global illumination setups, glossy reflections and caustics. In Fig. 19, we compute the derivative of a glossy reflection with respect to the mixture weight of the teapot that produces the reflection. Once again, our estimator produces gradients with much lower standard deviation, $6.23\times$ in this setting due to our better importance sampling strategy (mixture decomposition) at the second bounce on the teapot. In Fig. 20, we compute the derivative of a caustic pattern

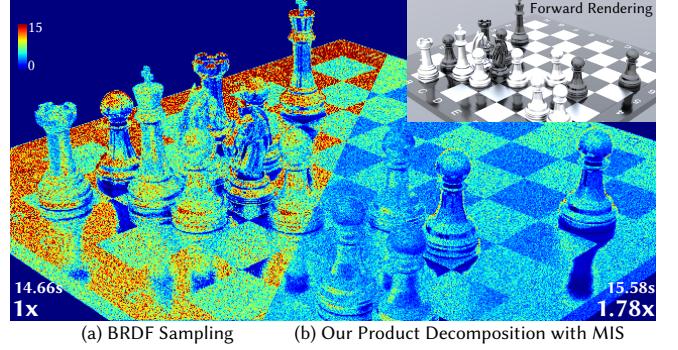


Fig. 22. PRODUCT DECOMPOSITION UNDER GLOBAL ILLUMINATION FOR A GLOSSY CHESSBOARD. Our product decomposition can efficiently estimate the anisotropic roughness derivative of a chess set modelled by a Beckmann distribution under one-bounce global illumination (25 ray budget). This scene has more complex visibility due to inter-occlusions and also some global illumination effects like glossy reflections of the chess pieces onto the chessboard; our decomposition handles all these effects better than BRDF sampling. 3D model courtesy of Turbosquid user Vadim Manoli. Chess position is from Magnus Carlsen vs Dommaraju Gukesh at the 11th Norway Blitz (2023), move 54.

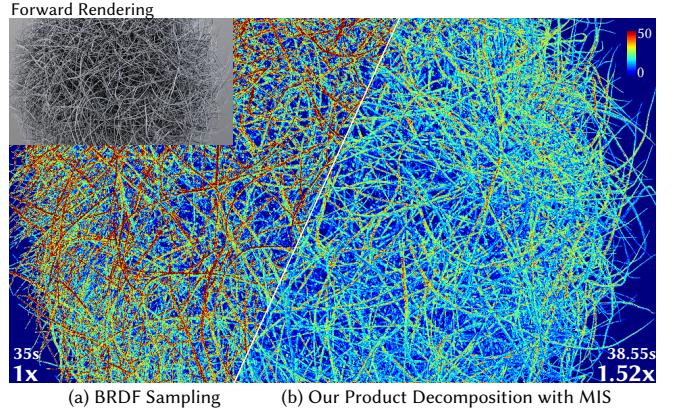


Fig. 23. PRODUCT DECOMPOSITION UNDER GLOBAL ILLUMINATION AND VERY COMPLEX VISIBILITY. To stress-test visibility variance (due to inter-occlusions) we estimate the derivative of an anisotropic metallic hairball (Beckmann with $\alpha_x = 0.01, \alpha_y = 0.02$, 25 ray budget). Under this challenging condition too, we still achieve a significant improvement in the standard deviation, $1.52\times$ over regular BRDF sampling due to better handling of sign and shape variance of the differential BRDF. 3D model courtesy of NVIDIA Research.

formed on a diffuse surface with respect to the isotropic roughness of a metallic cylinder that produces the caustic. This setup uses the positivization estimator with a higher branching factor; nonetheless, it is still able to significantly reduce standard deviation (by $2.51\times$) due to correct handling of the BRDF derivative at the second bounce on the cylinders surface.

We also demonstrate that our estimators work well even in the presence of more detailed geometry and complicated visibility in Fig. 21. We estimate the derivative with the mixture weight under

both direct and one bounce global illumination. Our mixture decomposition reduces the standard deviation of the estimated gradients by $2.35\times$ and $2.27\times$ under the two settings respectively as compared to BRDF importance sampling.

Finally, we show the effectiveness of the product decomposition under one bounce global illumination with complex visibility, glossy reflections and other effects intertwined in a single scene in Figs. 22 and 23. In Fig. 22, we estimate the anisotropic roughness derivative of a chess set modelled by a Beckmann BRDF which includes effects such as inter-occlusion among the pieces and glossy reflections on the chessboard. Product decomposition reduces the standard deviation by $1.78\times$ on this scene. In Fig. 23, we consider the case of very complex visibility. We estimate the anisotropic roughness derivative of a metallic hairball. Even under this challenging condition, we are able to better handle sign and shape variance of the differential BRDF resulting in a $1.52\times$ reduction in standard deviation over BRDF sampling. We provide difference images which show the specific regions of improvement for our techniques over BRDF sampling in Fig. 29. Additionally, we also show the effect of longer path lengths in Table 3.

Inverse rendering under global illumination. Positivization has the highest branching factor of our three decompositions. Applying it at a path vertex which requires a derivative results in the original path branching out into three sub paths, one of which requires recursive derivative computation. It is conceivable that this added computation may worsen positivization’s performance as compared to BRDF sampling, since BRDF sampling does not induce any branching. Experimentally, we have found that despite the additional branching, positivization is able to reduce variance and better recover textures under a single bounce of global illumination as compared to BRDF importance sampling at an equal computation budget, see Fig. 24. In the setup, we optimize the GGX roughness of a plate initialized uniformly with $\alpha = 0.05$ to match a target with a checkerboard pattern with $\alpha = 0.005, 0.02$. Light is incident upon the plate *only* indirectly via a reflection off a reflector, also modelled as a GGX plate with a fixed roughness of $\alpha = 0.01$, see Fig. 24 (b). The reduced variance due to better handling of the sign and shape variance by positivization leads to a better recovery of the texture. This is despite using the ADAM optimizer which ameliorates the issues of higher variance estimators [Nimier-David et al. 2022].

Our next inverse rendering experiment under global illumination recovers the anisotropic roughness of a GGX BRDF Fig. 25. The scene consists of two dragons ($\alpha_x = 0.005, \alpha_y = 0.007$) only viewed indirectly through a glossy reflection. The initialization starts with $\alpha_x = 0.5, \alpha_y = 0.007$ and we optimize for the correct α_x value. Our product decomposition is able to better handle the variance of the differential BRDF at the second bounce (on the dragons surface), which leads to faster convergence.

10 LIMITATIONS AND FUTURE WORK

Determining the number of samples for each decomposed component. For all three decompositions, our current implementation applies a two-sample estimator which uses one sample per component. It is possible that a different estimator can be more efficient in some cases. For example, when the two components have different

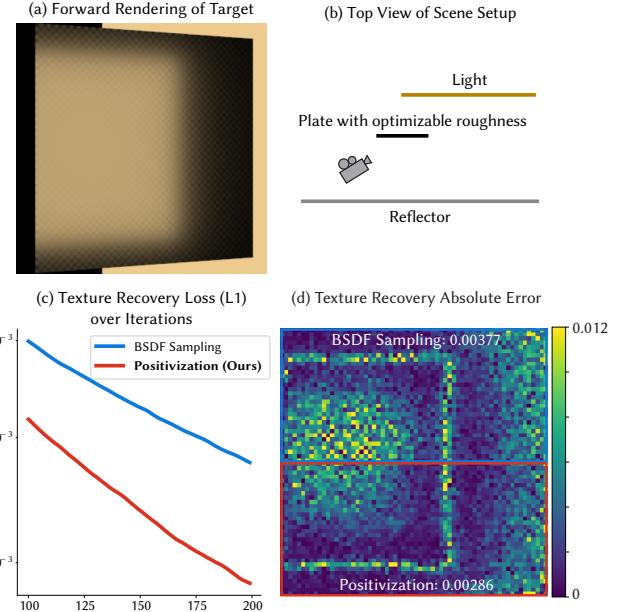


Fig. 24. INVERSE RENDERING OF THE ROUGHNESS OF AN ISOTROPIC GGX PLATE UNDER GLOBAL ILLUMINATION. The isotropic GGX plate shown in (a) can only receive light from the source via a reflector, setup shown in (b). We optimize the roughness of the GGX plate and find that our positivization is able to better recover the ground truth checkerboard texture as depicted by the loss curve (c) and the error in the recovered texture (d). Despite the extra branching (into three paths) introduced by positivization at the first bounce, the overall variance is reduced as compared to BRDF sampling (at an equal compute budget) which leads to a better recovery.

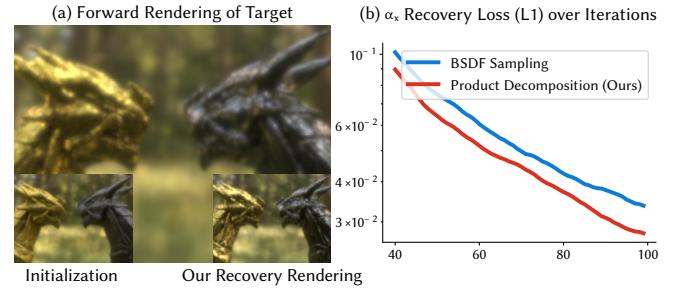


Fig. 25. INVERSE RENDERING OF ANISOTROPIC ROUGHNESS VIEWED ONLY THROUGH GLOSSY REFLECTION. (a) Our scene consists of glossy reflections of two dragons $\alpha_x = 0.005, \alpha_y = 0.007$ formed by a metal plate with $\alpha = 0.001$, all modelled by GGX distributions. We recover α_x for both the dragons, starting with an initial value of $\alpha_x = 0.5$, see inset for forward rendering of initialization. (b) Product decomposition (with MIS across the two lobes) achieves lower recovery error under this setting, the graph shows the sum of errors for the recovered α_x for the two dragons. Better handling of the differential BRDF’s sign and shape variance lets us recover the roughness values better at an equal compute budget, see our recovery in the inset in (a). 3D model courtesy of Turbosquid user techunit.

areas (i.e., $\int \partial_\alpha f_1 \neq \int \partial_\alpha f_2$ for components f_1 and f_2), it might be useful to adjust the number of samples according to the area of the

component (we show in Appendix C that microfacet normal distribution functions always have components with equal area). Research in allocating budgets for multiple importance sampling can likely help in our case as well [Grittmann et al. 2022; He and Owen 2014; Sbert et al. 2018]. Our estimator that always samples all components belongs to the *deterministic mixture* scheme [Owen 2013]. An alternative is a *random mixture*, which randomly chooses one component. We opt for deterministic mixtures since they consistently outperform random mixtures in our direct lighting experiments (due to the stratification effect, similar to standard MIS v.s. *one-sample* MIS). For global illumination, random mixtures are the same as applying Russian Roulette to keep only one of the two branches, and can be more computationally convenient in some cases since they omit the need for quadratic branching.

Branching and Global Illumination. Our adoption of deterministic mixtures requires path splitting for global illumination. While the branching complexity is quadratic instead of exponential (same as a bidirectional path tracer), it can add undesired overheads. There are several ways to reduce the branching, 1) deterministically using only BRDF sampling or using random mixtures instead of deterministic mixtures after a certain recursion depth, 3) using path reconnection similar to Zhang et al.’s approach [2020], to reconnect the branches back to a single primary path. Figuring out an effective strategy to deal with branching is crucial for applying our method to estimate derivatives in volumetric rendering and subsurface scattering with long paths; this is an exciting avenue for future research.

Multiple Parameter Optimization. Prior work [Zeltner et al. 2021] and ours only deal with estimating the derivative of a single material parameter (such as roughness, mixture weights, etc.) and constructs specialized sampling schemes for each of them. However, in inverse rendering, we are often interested in recovering several parameters simultaneously. For example, Uber shaders [Burley 2012; Georgiev et al. 2019] have a large number of tunable parameters. Naively applying our decompositions for each of these parameters independently can lead to very large branching factors. Developing sampling strategies that efficiently estimate the derivatives of all the parameters simultaneously is still an open research problem.

Unsupported BRDFs. Our work covers a large variety of popular analytic surface reflection models, see Table 1. However, our method currently does not support data driven BRDFs or Neural BRDFs. In both cases, there are a very large number of tunable parameters, and it is unclear as to which parameters one might want to differentiate or importance sample.

Better Optimization Schemes. Ultimately, for inverse rendering, the optimization is both ill-posed and non-convex. Recently, we have seen some work [Xing et al. 2022] which takes a step in this direction. We believe the study of efficient estimators of the derivatives is largely orthogonal and equally crucial.

11 CONCLUSION

Our importance sampling techniques provide a fundamental component for future differentiable rendering work, enabling correct handling of sign and shape variance of differential BRDFs. BRDF

sampling is widely used in forward rendering to deal with a variety of light transport phenomena; this includes unidirectional, bidirectional and gradient domain path tracing, Metropolis light transport, path guiding, photon mapping, etc. Similarly, as the need to deal with the differentials of more complicated light transport phenomena arises, we will need to develop differential counterparts of these algorithms and we believe that our method will be well suited to serve as a fundamental building block for them. Our product and mixture decompositions can also potentially have use outside of graphics for importance sampling real-valued functions.

ACKNOWLEDGEMENTS

This work was funded in part by NSF grants 2105806 and 2212085, gifts from Adobe, Google and Qualcomm, the Ronald L. Graham Chair and the UC San Diego Center for Visual Computing. We thank Tizian Zeltner for providing additional information about antithetic sampling, and Alexandr Kuznetsov, Xuanda Yang and Wesley Chang for proofreading our paper.

REFERENCES

- James Arvo. 1994. The Irradiance Jacobian for Partially Occluded Polyhedral Sources. In *SIGGRAPH*, 343–350.
- James Arvo and David Kirk. 1990. Particle Transport and Image Synthesis. *Comput. Graph. (Proc. SIGGRAPH)* 24, 4 (1990), 63–66.
- Michael Ashikhmin and Peter Shirley. 2001. An Anisotropic Phong BRDF Model. *J. Graph. Tools* 5, 2 (2001), 25–32.
- Dejan Azinović, Tzu-Mao Li, Anton Kaplanyan, and Matthias Nießner. 2019. Inverse Path Tracing for Joint Material and Lighting Estimation. In *Computer Vision and Pattern Recognition*. 2447–2456.
- Sai Praveen Bangaru, Tzu-Mao Li, and Frédo Durand. 2020. Unbiased Warped-Area Sampling for Differentiable Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 39, 6 (2020), 245:1–245:18.
- Sai Praveen Bangaru, Lifan Wu, Tzu-Mao Li, Jacob Munkberg, Gilbert Bernstein, Jonathan Ragan-Kelley, Frédo Durand, Aaron Lefohn, and Yong He. 2023. SLANG.D: Fast, Modular and Differentiable Shader Programming. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 42, 6, Article 264 (2023).
- Petr Beckmann and Andre Spizzichino. 1987. *The scattering of electromagnetic waves from rough surfaces*.
- James F. Blinn. 1977. Models of Light Reflection for Computer Synthesized Pictures. *Comput. Graph. (Proc. SIGGRAPH)* 11, 2 (1977), 192–198.
- Brent Burley. 2012. Physically-based shading at Disney. In *SIGGRAPH Course Notes. Practical physically-based shading in film and game production*. 1–7.
- Brent Burley. 2015. Extending the Disney BRDF to a BSDF with integrated subsurface scattering. In *SIGGRAPH Course Notes. Physically Based Shading in Theory and Practice*. 1–9.
- Wesley Chang, Venkataram Sivaram, Derek Nowrouzezahrai, Toshiya Hachisuka, Ravi Ramamoorthi, and Tzu-Mao Li. 2023. Parameter-space ReSTIR for Differentiable and Inverse Rendering. In *SIGGRAPH Conference Proceedings*. Article 18, 10 pages.
- Chengqian Che, Fujun Luan, Shuang Zhao, Kavita Bala, and Ioannis Gkioulekas. 2020. Towards learning-based inverse subsurface scattering. In *International Conference on Computational Photography (ICCP)*.
- Michael Cohen and John Wallace. 1993. *Radiosity and Realistic Image Synthesis*. Morgan Kaufmann.
- Martin de La Gorce, David J Fleet, and Nikos Paragios. 2011. Model-based 3D hand pose estimation from monocular video. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 9 (2011), 1793–1805.
- Valentin Deschaintre, Miika Aittala, Frédo Durand, George Drettakis, and Adrien Bousseau. 2018. Single-image SVBRDF Capture with a Rendering-aware Deep Network. *ACM Trans. Graph. (Proc. SIGGRAPH)* 37, 4 (2018), 128:1–128:15.
- Jonathan Dupuy and Wenzel Jakob. 2018. An Adaptive Parameterization for Efficient Material Acquisition and Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 37, 6 (2018), 274:1–274:18.
- Jiahui Fan, Beibei Wang, Milos Hasan, Jian Yang, and Ling-Qi Yan. 2022. Neural Layered BRDFs. In *SIGGRAPH Conference Proceedings*. Article 4, 8 pages.
- Iliyan Georgiev, Jamie Portsmouth, Zap Andersson, Adrien Herubel, Alan King, Shinji Ogaki, and Frederic Servant. 2019. Autodesk Standard Surface.
- Ioannis Gkioulekas, Shuang Zhao, Kavita Bala, Todd Zickler, and Anat Levin. 2013. Inverse Volume Rendering with Material Dictionaries. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 32, 6 (2013), 162:1–162:13.

- Pascal Grittmann, Ömercan Yazici, Iliyan Georgiev, and Philipp Slusallek. 2022. Efficiency-Aware Multiple Importance Sampling for Bidirectional Rendering Algorithms. *ACM Trans. Graph. (Proc. SIGGRAPH)* 41, 4, Article 80 (2022).
- Pat Hanrahan and Wolfgang Krueger. 1993. Reflection from Layered Surfaces Due to Subsurface Scattering. In *SIGGRAPH*. 165–174.
- Hera Y He and Art B Owen. 2014. Optimal mixture weights in multiple importance sampling. *arXiv preprint arXiv:1411.3954* (2014).
- Eric Heitz. 2017. *A Simpler and Exact Sampling Routine for the GGX Distribution of Visible Normals*. Research Report. Unity Technologies.
- Eric Heitz. 2018. Sampling the GGX Distribution of Visible Normals. *Journal of Computer Graphics Techniques* 7, 4 (2018), 1–13.
- Eric Heitz and Eugene d’Eon. 2014. Importance Sampling Microfacet-Based BSDFs using the Distribution of Visible Normals. *Comput. Graph. Forum (Proc. EGSR)* 33, 4 (2014), 103–112.
- L. G. Henyey and J. L. Greenstein. 1941. Diffuse radiation in the Galaxy. *The Astrophysical Journal* 93 (1941), 70–83.
- Wenzel Jakob. 2014. *An Improved Visible Normal Sampling Routine for the Beckmann Distribution*. Technical Report. Cornell University.
- Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, and Delio Vicini. 2022. Dr.Jit: A Just-In-Time Compiler for Differentiable Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH)* 41, 4 (2022).
- James T. Kajiya. 1986. The Rendering Equation. *Comput. Graph. (Proc. SIGGRAPH)* 20, 4 (1986), 143–150.
- Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. 2018. Neural 3D Mesh Renderer. In *Computer Vision and Pattern Recognition*. 3907–3916.
- Pramook Khungurn, Daniel Schroeder, Shuang Zhao, Kavita Bala, and Steve Marschner. 2015. Matching Real Fabrics with Micro-Appearance Models. *ACM Transactions on Graphics* 35, 1 (2015), 1:1–1:26.
- Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*.
- Alexandr Kuznetsov, Krishna Mullia, Zexiang Xu, Miloš Hašan, and Ravi Ramamoorthi. 2021. NeuMIP: Multi-Resolution Neural Materials. *ACM Trans. Graph. (Proc. SIGGRAPH)* 40, 4, Article 175 (2021), 13 pages.
- Alexandr Kuznetsov, Xuezhang Wang, Krishna Mullia, Fujun Luan, Zexiang Xu, Miloš Hašan, and Ravi Ramamoorthi. 2022. Rendering Neural Materials on Curved Surfaces. In *SIGGRAPH Conference Proceedings*. 9 pages.
- Eric P. F. Laforte, Sing-Cheong Foo, Kenneth E. Torrance, and Donald P. Greenberg. 1997. Non-Linear Approximation of Reflectance Functions. In *SIGGRAPH*. 117–126.
- Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. 2020. Modular Primitives for High-Performance Differentiable Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 39, 6 (2020).
- Jason Lawrence, Szymon Rusinkiewicz, and Ravi Ramamoorthi. 2004. Efficient BRDF Importance Sampling Using a Factored Representation. *ACM Trans. Graph. (Proc. SIGGRAPH)* (2004), 496–505.
- Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. 2018. Differentiable Monte Carlo Ray Tracing through Edge Sampling. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 37, 6 (2018), 222:1–222:11.
- Shichen Liu, Tianyu Li, Weikai Chen, and Hao Li. 2020. A general differentiable mesh renderer for image-based 3D reasoning. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 1 (2020), 50–62.
- Matthew W. Loper and Michael J. Black. 2014. OpenDR: An Approximate Differentiable Renderer. In *European Conference on Computer Vision*, Vol. 8695. 154–169.
- Guillaume Loubet, Nicolas Holzschuch, and Wenzel Jakob. 2019. Reparameterizing discontinuous integrands for differentiable rendering. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 38, 6 (2019).
- Joakim Löw, Joel Kronander, Anders Ynnerman, and Jonas Unger. 2012. BRDF Models for Accurate and Efficient Rendering of Glossy Surfaces. *ACM Transactions on Graphics* 31, 1, Article 9 (2012), 14 pages.
- Fujun Luan, Shuang Zhao, Kavita Bala, and Zhao Dong. 2021. Unified Shape and SVBRDF Recovery using Differentiable Monte Carlo Rendering. *Comput. Graph. Forum (Proc. EGSR)* 40, 4 (2021), 101–113.
- Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. 2003. A Data-Driven Reflectance Model. *ACM Trans. Graph. (Proc. SIGGRAPH)* 22, 3 (2003), 759–769.
- M. Minnaert. 1941. The reciprocity principle in lunar photometry. *The Astrophysical Journal* 93 (1941), 403–410.
- Don P. Mitchell. 1996. Consequences of Stratified Sampling in Graphics. In *SIGGRAPH*. 277–280.
- Baptiste Nicolet, Fabrice Rousselle, Jan Novák, Alexander Keller, Wenzel Jakob, and Thomas Müller. 2023. Recursive Control Variates for Inverse Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH)* 42, 4 (2023).
- Merlin Nimmer-David, Zhao Dong, Wenzel Jakob, and Anton Kaplanyan. 2021. Material and Lighting Reconstruction for Complex Indoor Scenes with Texture-space Differentiable Rendering. In *Eurographics Symposium on Rendering - DL-only Track*.
- Merlin Nimmer-David, Thomas Müller, Alexander Keller, and Wenzel Jakob. 2022. Unbiased Inverse Volume Rendering with Differential Trackers. *ACM Trans. Graph. (Proc. SIGGRAPH)* 41, 4 (2022), 44:1–44:20.
- Merlin Nimmer-David, Sébastien Speierer, Benoît Ruiz, and Wenzel Jakob. 2020. Radiative Backpropagation: An Adjoint Method for Lightning-Fast Differentiable Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH)* 39, 4 (2020).
- Merlin Nimmer-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. 2019. Mitsuba 2: A retargetable forward and inverse renderer. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 38, 6 (2019), 1–17.
- Ko Nishino. 2009. Directional statistics BRDF model. In *International Conference on Computer Vision*. 476–483.
- Michael Oren and Shree K. Nayar. 1994. Generalization of Lambert’s Reflectance Model. In *SIGGRAPH*. 239–246.
- Art Owen and Yi Zhou. 2000. Safe and Effective Importance Sampling. *J. Amer. Statist. Assoc.* 95, 449 (2000), 135–143.
- Art B. Owen. 2013. *Monte Carlo theory, methods and examples*. Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2016. *Physically Based Rendering: From Theory to Implementation* (3rd ed.). Morgan Kaufmann Publishers Inc. 1266 pages.
- Bui Tuong Phong. 1975. Illumination for Computer Generated Pictures. *Commun. ACM* 18, 6 (1975), 311–317.
- Ravi Ramamoorthi, Dhruv Mahajan, and Peter Belhumeur. 2007. A First-Order Analysis of Lighting, Shading, and Shadows. *ACM Transactions on Graphics* 26, 1 (2007), 21 pages.
- Iman Sadeghi, Oleg Bisker, Joachim De Decker, and Henrik Wann Jensen. 2013. A Practical Microcylinder Appearance Model for Cloth Rendering. *ACM Transactions on Graphics* 32, 2, Article 14 (2013), 12 pages.
- Mateu Sbert, Vlastimil Havran, and Laszlo Szirmay-Kalos. 2018. Multiple importance sampling revisited: breaking the bounds. *EUROSIP Journal on Advances in Signal Processing* 2018, 1 (2018), 1–15.
- Cheng Sun, Guangyan Cai, Zhengqin Li, Kai Yan, Cheng Zhang, Carl Marshall, Jian-Bin Huang, Shuang Zhao, and Zhao Dong. 2023. Neural-PBIR Reconstruction of Shape, Material, and Illumination. In *International Conference on Computer Vision*. 18046–18056.
- Alejandro Sztajman, Gilles Rainer, Tobias Ritschel, and Tim Weyrich. 2021. Neural BRDF representation and importance sampling. *Comput. Graph. Forum* 40, 6 (2021), 332–346.
- Justin F. Talbot, David Cline, and Parris Egbert. 2005. Importance Resampling for Global Illumination. *Rendering Techniques (Proc. EGSR)*, 139–146.
- TS Trowbridge and Karl P Reitz. 1975. Average irregularity representation of a rough surface for ray reflection. *J. Opt. Soc. Am.* 65, 5 (1975), 531–536.
- Eric Veach and Leonidas J. Guibas. 1995. Optimally Combining Sampling Techniques for Monte Carlo Rendering. In *SIGGRAPH*. 419–428.
- Delio Vicini, Sébastien Speierer, and Wenzel Jakob. 2021. Path Replay Backpropagation: Differentiating Light Paths using Constant Memory and Linear Time. *ACM Trans. Graph. (Proc. SIGGRAPH)* 40, 4 (2021), 108:1–108:14.
- Ingo Wald, Sven Woop, Carsten Benthin, Gregory S Johnson, and Manfred Ernst. 2014. Embree: a kernel framework for efficient CPU ray tracing. *ACM Trans. Graph. (Proc. SIGGRAPH)* 33, 4 (2014), 143:1–143:8.
- Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. 2007. Microfacet models for refraction through rough surfaces. *Rendering Techniques (Proc. EGSR)* (2007), 195–206.
- Yu-Chen Wang, Chris Wyman, Lifan Wu, and Shuang Zhao. 2023. Amortizing Samples in Physics-Based Inverse Rendering using ReSTIR. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 42, 6 (2023).
- Greg Ward and Paul Heckbert. 1992. Irradiance Gradients. In *Eurographics Workshop on Rendering*. 85–98.
- Gregory J. Ward. 1992. Measuring and modeling anisotropic reflection. *Comput. Graph. (Proc. SIGGRAPH)* 26, 2 (1992), 265–272.
- Lifan Wu, Guangyan Cai, Ravi Ramamoorthi, and Shuang Zhao. 2021. Differentiable Time-Gated Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 40, 6 (2021), 287:1–287:16.
- Liwen Wu, Rui Zhu, Mustafa B Yaldiz, Yiniao Zhu, Hong Cai, Janarbek Matai, Fatih Porikli, Tzu-Mao Li, Mammoohan Chandraker, and Ravi Ramamoorthi. 2023. Factorized Inverse Path Tracing for Efficient and Accurate Material-Lighting Estimation. In *International Conference on Computer Vision*. 3848–3858.
- Jiankai Xing, Fujun Luan, Ling-Qi Yan, Xuejun Hu, Houde Qian, and Kun Xu. 2022. Differentiable Rendering using RGBXY Derivatives and Optimal Transport. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 41, 6, Article 189 (2022), 13 pages.
- Bing Xu, Liwen Wu, Milos Hasan, Fujun Luan, Iliyan Georgiev, Zexiang Xu, and Ravi Ramamoorthi. 2023b. NeuSample: Importance Sampling for Neural Materials. In *SIGGRAPH Conference Proceedings*. Article 41.
- Peiyu Xu, Sai Bangaru, Tzu-Mao Li, and Shuang Zhao. 2023a. Warped-Area Reparameterization of Differential Path Integrals. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 42, 6 (2023).
- Kai Yan, Christoph Lassner, Brian Budge, Zhao Dong, and Shuang Zhao. 2022. Efficient Estimation of Boundary Integrals for Path-Space Differentiable Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH)* 41, 4 (2022), 123:1–123:13.

- Shinyoung Yi, Dongguan Kim, Kiseok Choi, Adrian Jarabo, Diego Gutierrez, and Min H. Kim. 2021. Differentiable Transient Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 40, 6 (2021).
- Zihan Yu, Cheng Zhang, Derek Nowrouzezahrai, Zhao Dong, and Shuang Zhao. 2022. Efficient Differentiation of Pixel Reconstruction Filters for Path-Space Differentiable Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 41, 6 (2022), 1–16.
- Cem Yuksel. 2022. High-Performance Polynomial Root Finding for Graphics. *ACM Comput. Graph. Interact. Tech. (Proc. HPG)* 5, 3 (2022), 7:1–7:15.
- Tizian Zeltner, Sébastien Speirer, Iliyan Georgiev, and Wenzel Jakob. 2021. Monte Carlo Estimators for Differential Light Transport. *ACM Trans. Graph. (Proc. SIGGRAPH)* 40, 4 (2021).
- Cheng Zhang, Zhao Dong, Michael Doggett, and Shuang Zhao. 2021a. Antithetic Sampling for Monte Carlo Differentiable Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH)* 40, 4 (2021), 77:1–77:12.
- Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. 2020. Path-Space Differentiable Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH)* 39, 4 (2020), 143:1–143:19.
- Cheng Zhang, Lifan Wu, Changxi Zheng, Ioannis Gkioulekas, Ravi Ramamoorthi, and Shuang Zhao. 2019. A Differential Theory of Radiative Transfer. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 38, 6 (2019), 227:1–227:16.
- Cheng Zhang, Zihan Yu, and Shuang Zhao. 2021b. Path-Space Differentiable Rendering of Participating Media. *ACM Trans. Graph. (Proc. SIGGRAPH)* 40, 4 (2021), 76:1–76:15.
- Ziyi Zhang, Nicolas Roussel, and Wenzel Jakob. 2023. Projective Sampling for Differentiable Rendering of Geometry. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 42, 6 (2023).

A BRDF DERIVATIVE IMPORTANCE SAMPLING PDFS AND CDFS

All PDFs and CDFs are in solid angle coordinates, and do not include multiplication by $\sin \theta$ for change of variables to spherical coordinates. PDFs may be defined in either ω_i or ω_h space, depending on the BRDF. The PDFs defined in ω_h space must finally be transformed to ω_i space, and while doing so must include the appropriate jacobian $4\omega_o \cdot \omega_h$. The PDFs are denoted by p and their corresponding CDFs are P . In the cases where CDFs are provided instead of inverse transform sampling routines, CDF inversion is done numerically.

A.1 Positivization

These are all isotropic BRDFs, and sampling for the azimuthal angle ϕ is uniform sampling. We introduce PDFs and sampling routines for Blinn-Phong and Hanrahan-Krueger derivatives that have not been discussed in past literature to the best of our knowledge.

Importance sampling routines for the derivatives of isotropic GGX and Beckmann were first introduced by Zeltner et al. [2021] in Appendix A of their paper, and we do not repeat them here. However, they do not provide explicit formulae for the PDFs p_+, p_- that we need for positivization. These PDFs have a different normalization by a factor of 2 than the PDF p they use, so we define the PDFs p_+, p_- here.

A.1.1 Isotropic GGX.

$$\begin{aligned} r(\theta_h) &= \frac{8\alpha^2 \sec^3 \theta_h (\tan^2 \theta_h - \alpha^2)}{(\tan^2 \theta_h + \alpha^2)^3} \\ p_{\alpha,-}(\theta_h) &= -\min(r(\theta_h), 0) \\ p_{\alpha,+}(\theta_h) &= \max(r(\theta_h), 0) \end{aligned} \quad (24)$$

A.1.2 Isotropic Beckmann.

$$\begin{aligned} r(\theta_h) &= \frac{4e^{1-tan^2 \theta_h/\alpha^2} \sec^3 \theta_h (\tan^2 \theta_h - \alpha^2)}{\alpha^4} \\ p_{\alpha,-}(\theta_h) &= -\min(r(\theta_h), 0) \\ p_{\alpha,+}(\theta_h) &= \max(r(\theta_h), 0) \end{aligned} \quad (25)$$

A.1.3 Blinn-Phong (Minnaert).

$$\begin{aligned} r(\theta_h) &= e(n+2) \cos^{n+1} \theta_h ((n+2) \log \cos \theta_h + 1) \\ p_{n,-}(\theta_h) &= -\min(r(\theta_h), 0) \\ p_{n,+}(\theta_h) &= \max(r(\theta_h), 0) \\ P_{n,+}(\theta_h) &= -e(n+2) \cos^{n+2} \theta_h \log \cos \theta_h \\ P_{n,-}(\theta_h) &= 1 - e(n+2) \cos^{n+2} \theta_h \log \cos \theta_h \end{aligned} \quad (26)$$

For the Minnaert BRDF, the sampling routines are the same as above, but defined in θ_i space instead of θ_h .

A.1.4 Henyey-Greenstein (Hanrahan-Krueger).

$$\begin{aligned} C &= \frac{3^{3/2} g^2 (1 - g^2)}{(3 + g^2)^{3/2} - 3^{3/2} (1 - g^2)} \\ r(\theta_i) &= Cg^2 \frac{(g^2 + 3) \cos \theta_i + g(g^2 - 5)}{(g^2 - 2g \cos \theta_i + 1)^{5/2}} \\ p_{g,-}(\theta_i) &= -\min(r(\theta_i), 0) \\ p_{g,+}(\theta_i) &= \max(r(\theta_i), 0) \\ P_{g,-}(\theta_i) &= \begin{cases} (1 - C) - C \left[\frac{3g^2 + 1 - g(g^2 + 3) \cos \theta_i}{(g^2 - 2g \cos \theta_i + 1)^{3/2}} \right], & \text{if } p_{g,-}(\theta_i) > 0 \\ 1, & \text{otherwise} \end{cases} \\ P_{g,+}(\theta_i) &= \begin{cases} C \left[\frac{3g^2 + 1 - g(g^2 + 3) \cos \theta_i}{(g^2 - 2g \cos \theta_i + 1)^{3/2}} \right] - 1, & \text{if } p_{g,+}(\theta_i) > 0 \\ 1, & \text{otherwise} \end{cases} \end{aligned} \quad (27)$$

A.2 Product Decomposition

For product decomposition, there are two sampling PDFs. The first is $p_1 \propto g$, which is just regular BRDF sampling (e.g. visible normal distribution function sampling for GGX/ Beckmann); we do not repeat them here. We provide importance sampling PDFs and CDFs for $\partial_\alpha g$.

For Anisotropic GGX and Beckmann, we provide the PDFs and importance sampling routines for $\partial_{\alpha_x} g$ with one of the directional parameters α_x . The corresponding PDFs and CDFs for the other directional parameter α_y can be obtained by swapping α_x with α_y and $\cos \phi_h$ with $\sin \phi_h$. We do the same for Ashikhmin-Shirley too, except the directional parameters are n_u, n_v in this case.

For the three BRDFs above, the CDF for ϕ_h generates an azimuthal angle in the range $[0, \pi/2]$. ϕ_h is mirror symmetric about $\pi/2$ and has a period of π , which is used to transform ϕ_h to the range $[0, 2\pi]$ (and the jacobian needs to account for this via a division by 4 as well). The CDF for θ_h generates an elevation angle in $[0, \pi/2]$.

A.2.1 Anisotropic GGX. Derivative with α_x .

$$\begin{aligned} a(\phi_h) &= \frac{\cos^2 \phi_h}{\alpha_x^2} + \frac{\sin^2 \phi_h}{\alpha_y^2} \\ g(\theta_h, \phi_h) &= \left(a(\phi_h) \sin^2 \theta_h + \cos^2 \theta_h \right)^{-2} \\ p_{\alpha_x}(\phi_h) &= \frac{4 \cos^2 \phi_h}{\pi \alpha_x^3 \alpha_y a(\phi_h)^2} \\ p_{\alpha_x}(\theta_h | \phi_h) &= \frac{4a(\phi_h)^2 \tan^2 \theta_h \sec^3 \theta_h}{(\tan^2 \theta_h a(\phi_h) + 1)^3} \\ P_{\alpha_x}(\phi_h) &= \frac{2}{\pi} \left[\tan^{-1} \left(\frac{\alpha_x}{\alpha_y} \tan \phi_h \right) \right. \\ &\quad \left. + \frac{\alpha_y \alpha_x \sin(2\phi_h)}{\alpha_x^2 + \alpha_y^2 + (\alpha_y^2 - \alpha_x^2) \cos(2\phi_h)} \right] \\ P_{\alpha_x}(\theta_h | \phi_h) &= \frac{a(\phi_h)^2}{a(\phi_h)^2 - 1} \\ &\quad - \frac{a(\phi_h)^2 ((1 - a(\phi_h)) \cos(4\theta_h) + a(\phi_h) + 3)}{4(a(\phi_h)^2 - 1) ((a(\phi_h) - 1) \sin^2 \theta_h + 1)^2} \end{aligned} \quad (28)$$

A.2.2 Anisotropic Beckmann (Ward). Derivative with α_x .

The importance sampling PDFs and CDFs for the anisotropic Beckmann and Ward BRDFs are the same since the shape functions g for both the BRDFs (and their derivatives) take on a similar functional form. The PDF $p_{\alpha_x}(\phi_h)$ and CDF $P_{\alpha_x}(\phi_h)$ for them is the same as GGX, see Eqn. (28). Also see Eqn. (28) for the definition of $a(\phi_h)$.

$$\begin{aligned} g(\theta_h, \phi_h) &= \sec^3 \theta_h e^{-a(\phi_h) \tan^2 \theta_h} \\ p_{\alpha_x}(\theta_h | \phi_h) &= 2a(\phi_h)^2 \tan^2 \theta_h \sec^3 \theta_h e^{-a(\phi_h) \tan^2 \theta_h} \\ P_{\alpha_x}(\theta_h | \phi_h) &= 1 - (1 + a(\phi_h) \tan^2 \theta_h) e^{-a(\phi_h) \tan^2 \theta_h} \end{aligned} \quad (29)$$

A.2.3 Ashikhmin-Shirley. Derivative with n_u .

$$\begin{aligned} a(\phi_h) &= n_u \cos^2 \phi_h + n_v \sin^2 \phi_h \\ g(\theta_h, \phi_h) &= \cos \theta_h^{a(\phi_h)} \\ p_{n_u}(\phi_h) &= \frac{4(n_u + 1)^{3/2} \sqrt{n_v + 1} \cos^2 \phi_h}{\pi(1 + a(\phi_h))^2} \\ p_{n_u}(\theta_h | \phi_h) &= -\log \cos \theta_h (1 + a(\phi_h))^2 \cos \theta_h^{a(\phi_h)} \\ P_{n_u}(\phi_h) &= \frac{2}{\pi} \left[\tan^{-1} \left(\sqrt{\frac{n_v + 1}{n_u + 1}} \tan \phi_h \right) \right. \\ &\quad \left. + \frac{\sqrt{(n_u + 1)(n_v + 1)} \sin(2\phi_h)}{n_u + n_v + 2 + (n_u - n_v) \cos(2\phi_h)} \right] \\ P_{n_u}(\theta_h | \phi_h) &= 1 - (1 - (a(\phi_h) + 1) \log \cos \theta_h) \cos \theta_h^{a(\phi_h)+1} \end{aligned} \quad (30)$$

A.2.4 Microfacet ABC. The ABC Microfacet BRDF is an isotropic microfacet BRDF, and so the sampling for ϕ_h is uniform. The parameter A does not play a role in the microfacet BRDF (it is canceled out by the normalization constant), so we ignore it, and only consider the derivatives with the parameters B, C .

$$\begin{aligned} g(\theta_h) &= (1 + B(1 - \cos(\theta_h)))^{-C} \\ p_B(\theta_h) &= \frac{B^2 C (C - 1)(B + 1)^C (\cos \theta_h - 1)(1 + B(1 - \cos \theta_h))^{-1-C}}{(1 + BC - (B + 1)^C)} \\ p_C(\theta_h) &= \frac{B(C - 1)^2}{1 - (1 + B)^{1-C} ((C - 1) \log(B + 1) + 1)} \frac{\log(1 + B(1 - \cos \theta_h))}{(1 + B(1 - \cos \theta_h))^C} \\ P_B(\theta_h) &= \frac{(B + 1)^C (1 + B(1 - \cos \theta_h))^{-C} (1 + BC(1 - \cos \theta_h)) - (B + 1)^C}{(1 + BC - (B + 1)^C)} \\ P_C(\theta_h) &= \frac{1 - (1 + B(1 - \cos \theta_h))^{1-C} ((C - 1) \log(1 + B(1 - \cos \theta_h)) + 1)}{1 - (B + 1)^{1-C} ((C - 1) \log(1 + B) + 1)} \end{aligned} \quad (31)$$

A.2.5 Hemi-EPD. The Hemi-EPD microfacet BRDF is another isotropic BRDF, so ϕ_h is importance sampled using uniform sampling. Γ is the incomplete gamma function.

$$\begin{aligned} g(\theta_h) &= e^{\kappa \cos^\gamma \theta_h} - 1 \\ p_\kappa(\theta_h) &= \frac{\gamma \kappa (-\kappa)^{1/\gamma}}{\Gamma(1 + 1/\gamma, 0) - \Gamma(1 + 1/\gamma, -\kappa)} \cos^\gamma(\theta_h) e^{\kappa \cos^\gamma \theta_h} \\ P_\kappa(\theta_h) &= \frac{\Gamma(1 + 1/\gamma, -\kappa \cos^\gamma \theta_h) - \Gamma(1 + 1/\gamma, -\kappa)}{\Gamma(1 + 1/\gamma, 0) - \Gamma(1 + 1/\gamma, -\kappa)} \end{aligned} \quad (32)$$

A.2.6 Burley Diffuse BSSRDF. This BSSRDF is defined over an infinite plane, and is radially symmetric. The polar angle ϕ is sampled uniformly. We provide an importance sampling routine to sample the radial distance $r \in [0, \infty]$, for the derivative with the parameter d that controls both its height and width. Once again, a jacobian for multiplication with r is required here.

$$\begin{aligned} g(r, d) &= \frac{e^{-r/d} + e^{-r/3d}}{r} \\ p_d(r) &= \frac{e^{-r/d} + e^{-r/3d}/3}{4d^2} \\ P_d(r) &= 1 - \frac{e^{-r/d}(r + d)}{4d} - \frac{e^{-r/3d}(3d + r)}{4d} \end{aligned} \quad (33)$$

A.3 Mixture Decomposition

A.3.1 Mixture Model. We are interested in differentiating a mixture model f , given by

$$\begin{aligned} f(\omega_i, \omega_o) &= wf_1(\omega_i, \omega_o) + (1 - w)f_2(\omega_i, \omega_o) \\ \partial_w f(\omega_i, \omega_o) &= f_1(\omega_i, \omega_o) - f_2(\omega_i, \omega_o), \end{aligned} \quad (34)$$

with its parameter w . Here, f_1 and f_2 are the two lobes of the BRDF. The importance sampling scheme for the two terms of the derivative $\partial_w f$ are simply the BRDF importance sampling techniques for f_1 and f_2 respectively.

A.3.2 Oren-Nayar. We are interested in differentiating the roughness σ . The PDFs once again are in solid angle coordinates, not in spherical coordinates. The first term of Eqn. (20) requires standard cosine hemispherical sampling and we provide an importance sampling routine for the second term. Here, $p_2(\theta_i)$ is made up of two terms depending on whether $\theta_i < \theta_o$, and they have weights $A'_{21}, 1 - A'_{21}$ respectively. For ϕ_i , an exact inverse transform sampling routine is available.

$$\begin{aligned} A_{21} &= \frac{1}{2} \sin(\theta_o)(\theta_o - \sin(\theta_o) \cos(\theta_o)) \\ A_{22} &= \frac{1}{3} \tan(\theta_o)(1 - \sin^3(\theta_o)) \\ T_2 &= A_{21} + A_{22}, \quad A'_{21} = A_{21}/T_2, \\ p_2(\theta_i) &= \begin{cases} A'_{21} \frac{\sin(\theta_i)}{0.5(\theta_i - \sin(\theta_i) \cos(\theta_i))}, & \text{if } \theta_i < \theta_o \\ (1 - A'_{21}) \frac{3 \sin^2(\theta_i) \cos(\theta_i)}{1 - \sin^3(\theta_o)}, & \text{otherwise} \end{cases} \\ p_2(\phi_i) &= 0.5 \max(0, \cos(\phi_o - \phi_i)) \\ P_2(\theta_i) &= \begin{cases} A'_{21} \frac{\theta_i - \sin(\theta_i) \cos(\theta_i)}{\theta_o - \sin(\theta_o) \cos(\theta_o)}, & \text{if } \theta_i < \theta_o \\ A'_{21} + (1 - A'_{21}) \frac{\sin^2(\theta_i) - \sin^3(\theta_o)}{1.0 - \sin^3(\theta_o)}, & \text{otherwise} \end{cases} \\ \phi_i &= \begin{cases} \phi_o - \sin^{-1}(2u), & \text{if } u < 0.5 \\ \phi_o + \sin^{-1}(2u - 1), & \text{otherwise} \end{cases} \end{aligned} \quad (35)$$

A.3.3 Microcylinder. We want to importance sample the derivative of the BRDF corresponding to the volumetric scattering component $f_{r,v}$ in the original paper's notation, with the linear combination

weight k_d . This BRDF does *not* include cosine foreshortening.

$$\begin{aligned} f(\omega_i, \omega_o) &= F \frac{(1 - k_d)g(\theta_h; \gamma_v) + k_d A}{\cos \theta_i + \cos \theta_o} \\ \partial_{k_d} f(\omega_i, \omega_o) &= F \frac{1}{\cos \theta_i + \cos \theta_o} A - F \frac{g(\gamma_v, \theta_h)}{\cos \theta_i + \cos \theta_o} A, \end{aligned} \quad (36)$$

where F is the Fresnel term, A is the albedo, and g is a Gaussian with width γ_v . The first term is importance sampled using cosine hemispherical sampling, which is also the importance sampling technique used for this BRDF in forward rendering. The second term is importance sampled using inverse transform sampling for the Gaussian.

B ZELTNER ET AL.’S ANTITHETIC SAMPLING IS A SPECIAL CASE OF POSITIVIZATION

Zeltner et al.’s [2021] antithetic sampling involves generating *paired and correlated* samples for the positive and negative lobes of the BRDF derivative $\partial_\alpha f$ in two separate passes, one pass for each lobe, and then *averages* out the final result.

The correlation is induced by using the same random number generator state across the two passes. The only difference between the two passes are that the first one uses a flag to trigger sampling from the positive lobe p_+ of the PDF $p = wp_+ + (1 - w)p_-$, and the second one triggers sampling from the negative lobe p_- . Here, w is the relative area of the positive lobe of $\partial_\alpha f$, given by $|\int \partial_\alpha f_+| / (|\int \partial_\alpha f_+| + |\int \partial_\alpha f_-|)$ and is equal to 0.5 for the BRDF derivatives they consider, see Appendix Sec. C.

Their estimator for the integrand $\partial_\alpha f$ is given by,

$$I = \frac{1}{2} \left(\frac{\partial_\alpha f(X_+)}{p(X_+)} + \frac{\partial_\alpha f(X_-)}{p(X_-)} \right), \quad (37)$$

where the samples are drawn from $X_+ \sim p_+$ and $X_- \sim p_-$, and the factor of 1/2 comes from the fact that they average the result of the two passes. We can further simplify Eqn. (37), to bring it in a form similar to the positivization estimator in Eqn. (8), by noticing that $\partial_\alpha f(X) = \partial_\alpha f_+(X)$ when $X \sim p_+$ and similarly for p_- too, which gives us

$$I = \frac{\partial_\alpha f_+(X_+(u))}{p_+(X_+(u))} + \frac{\partial_\alpha f_-(X_-(u))}{p_-(X_-(u))}. \quad (38)$$

The only difference between the estimator above and the positivization estimator is that the samples $X_-(u)$ and $X_+(u)$ are correlated because they use the same uniform random number u , whereas they are uncorrelated for positivization because positivization does not impose any such restriction. Thus, antithetic sampling is a special case of *positivization with correlated random numbers*.

Positivization (with uncorrelated random numbers) achieves its variance reduction due to the stratification of the real-valued function into positive and negative functions, and we have experimentally verified that antithetic sampling (with correlated random numbers) consistently has similar variance reduction as positivization. As a result, antithetic sampling’s variance reduction can be explained by the *implicit stratification of $\partial_\alpha f$ into positive and negative lobes*. See Fig. 3 for an example of the variance reduction.

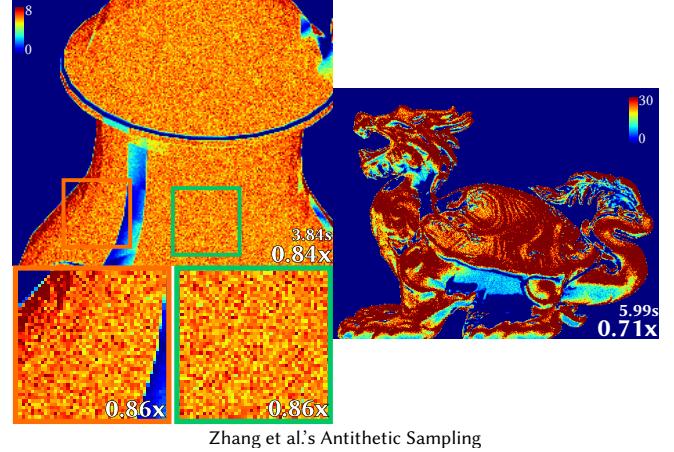


Fig. 26. ADDITIONAL COMPARISONS OF ZHANG ET AL. [2021A]. WITH BRDF SAMPLING FOR FIGS. 10 AND 12. Zhang et al. propose a method to estimate an *odd derivative* of the GGX BRDFs (the derivative with the half vector) by using antithetic sampling. However, their method is not directly applicable to many BRDF derivatives which can often be *even functions*. Take for example, the isotropic or anisotropic roughness derivative of a GGX BRDF. Applying an estimator tailored to odd integrands to an even integrand naturally leads to estimates with higher variance than the base sampling technique (BRDF sampling). We see this in the figure above as the variance is increased in both scenes (inset numbers are less than 1) as compared to BRDF sampling. Our product decomposition performs better than BRDF sampling; thus, it also outperforms Zhang et al.

C MICROFACET BRDF DERIVATIVES INTEGRATE TO ZERO

Previous work [Zeltner et al. 2021] has noted that the derivative of the normal distribution function of the isotropic GGX (and Beckmann) BRDF with its roughness parameter has positive and negative lobes with equal area. Here, we prove that this observation extends to *all the derivatives of all microfacet normal distribution functions*.

The projected area of a microfacet BRDF’s normal distribution function D always integrates to 1 i.e a constant,

$$\int D(\omega_h, \alpha) \cos \theta_h d\omega_h = 1 \quad (39)$$

As a result, its derivative with any parameter α integrates to 0,

$$\int \partial_\alpha D(\omega_h, \alpha) \cos \theta_h d\omega_h = 0 \quad (40)$$

which means that the *positive and negative lobes of $\partial_\alpha D \cos \theta_h$ have equal area*. Since we generally construct microfacet derivative sampling PDFs proportional to the derivative of the projected normal distribution function, the sampling PDFs (irrespective of the decomposition) for the positive and negative lobes of the derivative must *necessarily* have equal area.

D ADDITIONAL RESULTS

We present some additional comparisons and results in this section. First, we apply Zhang et al.’s [2021a] method to estimating the anisotropic roughness derivative of GGX BRDFs of the scenes

Table 3. DERIVATIVE ESTIMATION COMPARISON UNDER 3 BOUNCE GLOBAL ILLUMINATION I.E. DEPTH=5. The scenes are Figs. 21 to 23, respectively. The numbers indicate standard deviation improvement (higher is better) for our decompositions over BRDF sampling. As the maximum length of light paths increases (depth=3 vs depth=5), applying our decompositions at each path vertex is less effective due to extra branching. Figuring out which path vertices one should perform the decomposition at (or path reconnection after splitting) will likely ameliorate these issues, and is an exciting area for future research in differentiable rendering; Section 10 discusses this in further detail. Nonetheless, even with splitting at each path vertex, our estimators show improvement over BRDF sampling at depth=5; Azinović et al. [2019] argue that depth=4 is sufficient for inverse recovery of materials and lighting under diffuse global illumination.

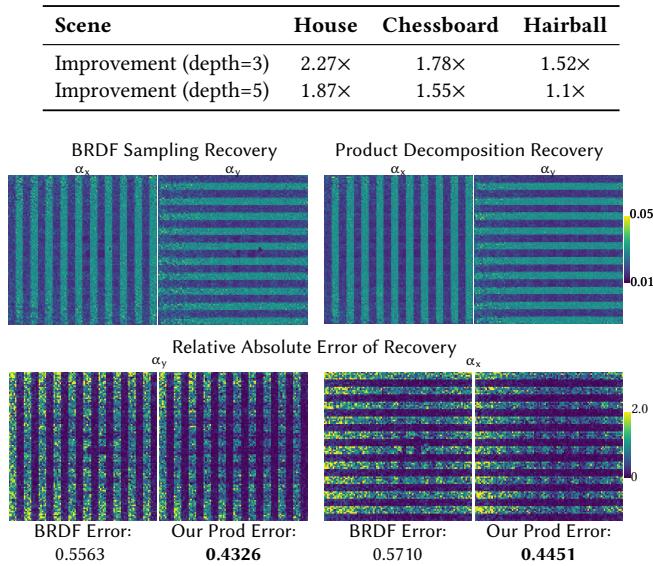


Fig. 27. TEXTURE RECOVERY COMPARISON FOR FIG. 15. Product decomposition is able to better recover textures due to its lower variance, this is especially noticeable in the highly specular regions (with low roughness values). It is hard to visually compare the differences between recovered textures, since they are often subtle. However, product decomposition’s lower variance leads to a smoother texture recovery (see bottom row) and lower error.

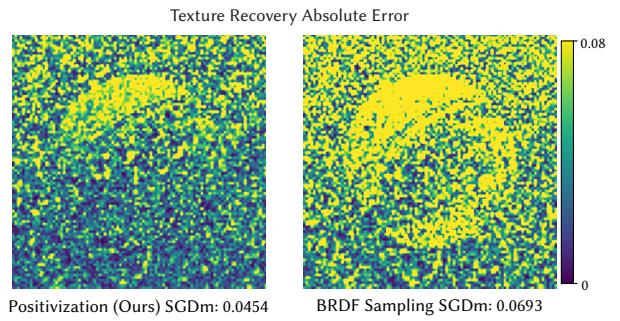


Fig. 28. INVERSE RENDERING OF THE SCATTERING PARAMETER g OF A HANRAHAN-KRUEGER BRDF WITH SGDM. Using stochastic gradient descent with momentum (SGDm) instead of ADAM exposes the higher variance of BRDF sampling as compared to positivization in the form of higher inverse rendering error in the recovered texture. The recovered texture with the ADAM optimizer is shown in Fig. 14.

depicted Figs. 10, 12 and 13 in Fig. 26. Second, we show the recovered textures using the SGDm optimizer in the experiment outlined in Fig. 14 in Fig. 28. Third, we show the recovered textures by both our method and BRDF sampling in Fig. 15 in Fig. 27. Next, we show the improvement regions of our estimators vs BRDF sampling for all the standard deviation estimation experiments in Fig. 29. Finally, we discuss the effect of larger path lengths in Table 3.

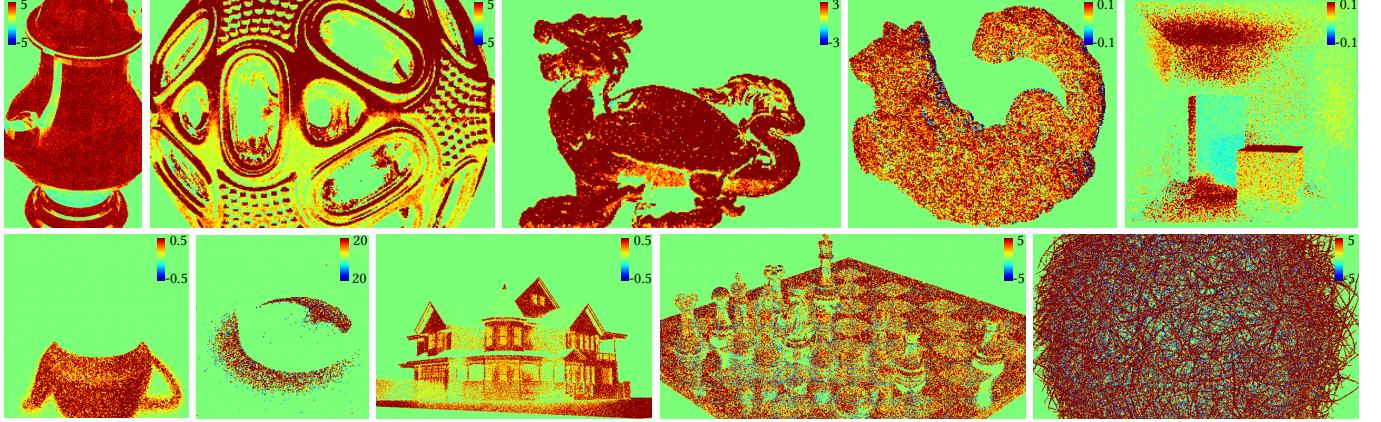


Fig. 29. IMPROVEMENT REGIONS FOR DERIVATIVE ESTIMATION. For each of the standard deviation images for derivatives estimators in Figs. 10 to 13 and 18 to 23, we compute the difference between the standard deviation of BRDF sampling and our techniques (positivization/ product decomposition/ mixture decomposition). Positive numbers indicate regions (shown in red) where our estimators have **better performance** (i.e. lower standard deviation) than BRDF sampling. Negative numbers indicate regions (shown in blue) where our estimators have **worse performance**. As it can be seen, we are better almost everywhere across all experiments.