A Constructive Theory of Sampling for Image Synthesis using Reproducing Kernel Bases

Christian Lessig*   Mathieu Desbrun   Eugene Fiume

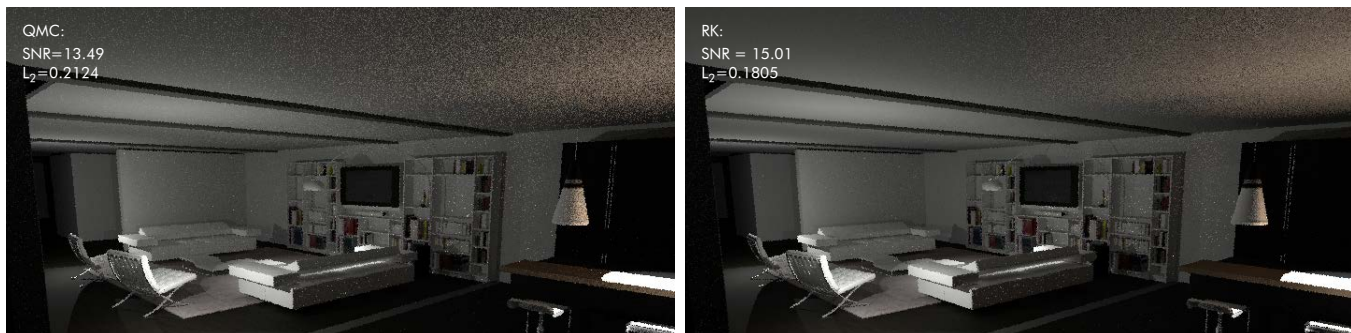DGP | University of Toronto   CMS | Caltech   DGP | University of Toronto

Figure 1: *Lighting in a Villa interior obtained from a photon map. For each shading point in the scene, the incoming light energy density is approximated using 64 hemispherical samples. The left image is computed using classical final gathering using a low-discrepancy sequence. For the right image the same samples were employed but with an integration rule derived using our theory of sampling, providing substantially reduced visual noise and lower numerical error rates.*

Abstract

Sampling a scene by tracing rays and reconstructing an image from such pointwise samples is fundamental to computer graphics. To improve the efficacy of these computations, we propose an alternative theory of sampling. In contrast to traditional formulations for image synthesis, which appeal to nonconstructive Dirac deltas, our theory employs constructive reproducing kernels for the correspondence between continuous functions and pointwise samples. Conceptually, this allows us to obtain a common mathematical formulation of almost all existing numerical techniques for image synthesis. Practically, it enables novel sampling based numerical techniques designed for light transport that provide considerably improved performance per sample. We exemplify the practical benefits of our formulation with three applications: pointwise transport of color spectra, projection of the light energy density into spherical harmonics, and approximation of the shading equation from a photon map. Experimental results verify the utility of our sampling formulation, with lower numerical error rates and enhanced visual quality compared to existing techniques.

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Display Algorithms I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Radiosity;

Keywords: sampling, light transport simulation, reproducing kernel Hilbert space

1 Introduction

Existing computational techniques for image synthesis rely extensively on pointwise sampling: path tracing computes the transport

---

*e-mail:lessig@dgp.toronto.edu

over a finite number of ray paths, photon mapping interpolates discrete "photons", and Metropolis light transport mutates existing pointwise paths, to name a few examples. Motivated by this importance of pointwise techniques, we present an alternative theory of sampling for image synthesis. In contrast to existing approaches that rely on nonconstructive sampling operators based on Dirac deltas, we instead employ constructive reproducing kernels to relate a continuous function to pointwise samples. This enables pointwise numerical techniques to be designed for light transport and devised for the finite sample counts and non-equidistant distributions employed in practice. Our example applications demonstrate that this can provide considerably improved visual fidelity and lower numerical error rates. Through its functional analytic foundation, our formulation also enables us to employ Galerkin projection to obtain sampling based algorithms such as path tracing and photon mapping. This overcomes the dichotomy between radiosity-based and ray tracing techniques that exists in the current literature.

The foundation of our formulation of sampling are reproducing kernels. These provide a theoretically insightful yet numerically practical conceptualization of the correspondence between pointwise samples and continuous functions. Reproducing kernels $k_\lambda(x)$ have the sifting property $\delta_\lambda[f] = \langle k_\lambda(x), f(x) \rangle = f(\lambda)$ of the Dirac delta δ_λ but are constructive Hilbert space functions. This enables us to employ kernels $k_{\lambda_i}(x)$ "located" at a set of points λ_i as basis functions. The basis expansion of an arbitrary signal $f(x)$ in such a reproducing kernel basis takes the form

$$f(x) = \sum_{i=1}^{m} \langle f(y), k_{\lambda_i}(y) \rangle \, \tilde{k}_i(x) = \sum_{i=1}^{m} f(\lambda_i) \, \tilde{k}_i(x).$$

This equation is at the heart of our work: pointwise samples $f(\lambda_i)$ provide the basis function coefficients—without computing inner products—and exact reconstruction from only these values $f(\lambda_i)$ is possible using well defined dual basis functions $\tilde{k}_i(x)$.

Our theory of sampling encompasses many classical results on the subject. For example, the Shannon sampling theorem arises as an orthonormal reproducing kernel basis for Fourier-bandlimited functions; the weights of quadrature rules, such as Gauss-Legendre, are obtained by integrating the dual functions $\tilde{k}_i(x)$; and reconstruction kernels, such as those used on the image plane, are dual kernels $\tilde{k}_i(x)$ for appropriate function spaces. However, with our formu-
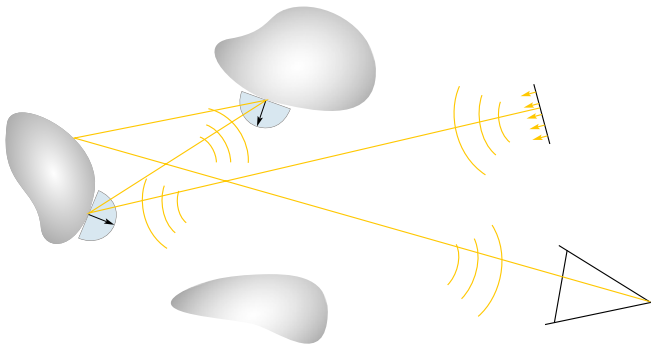
Figure 2: *Importance of point sampling for image synthesis. The light energy density is a continuous field whose propagation through space is described by a partial differential equation. Numerically, only the propagation along individual trajectories can be evaluated. This is equivalently to a pointwise solution of the partial differential equation using the method of characteristics. All solutions of the continuous equation have to be reconstructed from this pointwise information alone.*

lation, existing pointwise techniques become instances of a more versatile construction that provides a recipe to develop numerically practical sampling, quadrature, and interpolation formulas for arbitrary domains and all function spaces admitting a reproducing kernel. This is crucial for light transport simulation where signals are defined over arbitrary manifolds and where classical approximation spaces are ill-suited due to frequently occurring discontinuities. The requirement that a function space must admit a reproducing kernel might appear as limitation. Moreover, any finite function space, and hence any function space relevant for numerical computations, does admit a reproducing kernel. Nonetheless, the question of good approximation spaces and convergence to the infinite dimensional limit requires further investigation.

Our formulation of sampling also provides a novel quality measure for the distribution of sampling points which we call rk-discrepancy. The measure depends on the domain and function space under consideration. Together with its numerical computability, this enables the use of computer-based optimization techniques to find nearly optimal points–with respect to rk-discrepancy–for a given application.

We demonstrate the constructive nature and practicality of our formulation of sampling with three example applications, as follows:

i) Almost all techniques for image synthesis represent the color spectrum of light using pointwise samples. Peercy [1993] showed that this can be inefficient and that continuous basis functions can reduce the number of samples that is required. We employ reproducing kernel bases to ensure a well defined continuous representation of color with only pointwise samples.

ii) Spherical harmonics are used in a variety of techniques for image synthesis. For these, the spherical harmonics basis function coefficients of signals such as the light energy density have to be determined. We develop a quadrature rule for spherical harmonics basis projection and numerically optimize sampling locations on the sphere to obtain sequences that are nearly optimal for the problem.

iii) Photon mapping with final gathering is one of the most popular techniques for image synthesis. An image is thereby generated by determining the outgoing light energy density towards the camera using (Quasi) Monte Carlo integration in the hemisphere above each scene point. Based on the eigendecomposition of the scattering operator, we develop an optimal quadrature rule for this problem and demonstrate how importance sampling can be realized within our formulation.

For all three applications, we provide experimental results that demonstrate that our formulation significantly increases the efficacy attained per sample. These improvements result from tailoring the techniques to the function spaces that naturally arise and from intrinsically constructing them for the finite sample counts to which we are restricted in practice.

The remainder of the paper is structured as follows. The mathematical formulation of our constructive theory of sampling is developed in Sec. 2. In Sec. 3, we will discuss how the theory provides a unified formulation for a wide range of existing light transport simulation techniques. Subsequently in Sec. 4, we develop our three example applications and present experimental results. The paper concludes in Sec. 5 with an outlook on other possible applications of our formulation of sampling for image synthesis.

2 From Samples to Functions

In this section, we develop the mathematical formulation of our theory of sampling. To facilitate intuition, its concepts are exemplified for a 1-dimensional function class in Example 1 (boxed, on the next page) and the reader is encouraged to use the supplementary source code to gain some practical understanding. A more pedagogical exposition of our approach to sampling is presented in the supplementary tutorial. Readers unfamiliar with finite dimensional Hilbert spaces and their connection to numerical linear algebra should also consult this material or a standard text such as [Daubechies 1992; Mallat 2009].

2.1 Motivation

In computer graphics, the Dirac delta $\delta_\lambda[f] = f(\lambda)$ is commonly used to conceptualize the correspondence between a continuous function $f : X \to \mathbb{R}$ and pointwise samples $f(\lambda)$ of it. However, the Dirac delta $\delta_\lambda[\cdot]$ is a distribution in the sense of Schwartz and provides no information about a function f beyond its pointwise value $f(\lambda)$. This makes it ill-suited for the exact reconstruction of f from only a *finite* set of pointwise samples $f(\lambda_i)$, which is what we are interested in and restricted to in practice. Reconstruction from finitely many "samples" f_i is well defined using basis expansions. However, the f_i are then basis function coefficients defined by continuous inner products $\langle f, \tilde{\phi}_i \rangle = f_i$ whose numerical evaluation often requires costly approximations. Fortunately, for regular enough Hilbert spaces there exist functions $k_\lambda(x)$ such that the inner product is equivalent to point evaluation,

$$\langle f(y), k_\lambda(y) \rangle = f(\lambda) = \delta_\lambda[f]. \tag{1}$$

See Example 1.b for some concrete reproducing kernels. When such kernels $k_{\lambda_i}(x)$ are employed as basis functions, the samples $f(\lambda_i)$ immediately provide basis function coefficients—without the need to evaluate inner products. As will become clear in the remainder of the section, the realization of point evaluation as a function $k_\lambda(x)$ and not a distribution such as the Dirac delta is the essential difference between classical approaches to sampling and our work.

Some insight into the working principle of reproducing kernels $k_\lambda(x)$ can be obtained when we consider the Hilbert space basis expansion of a function $f(x)$ in an orthonormal basis $\{\phi_i\}_{i=1}^n$:

$$f(x) = \sum_{i=1}^n f_i \, \phi_i(x) = \sum_{i=1}^n \big\langle f(y), \phi_i(y) \big\rangle \phi_i(x). \tag{2a}$$

Exploiting linearity of the inner product $\langle \, , \rangle$ and that $f(y)$ does not depend on the index i we obtain for the right hand side

$$f(x) = \Big\langle f(y), \sum_{i=1}^n \phi_i(y) \, \phi_i(x) \Big\rangle. \tag{2b}$$

Comparing Eq. 2b to Eq. 1 we see that

$$k_x(y) = k(x, y) = \sum_{i=1}^n \phi_i(y) \, \phi_i(x). \tag{3}$$

Hence, the reproducing kernel acts by projecting a function $f(y)$ onto the basis functions $\phi_i(y)$, implicitly yielding coefficients f_i, and then reconstructing it at x using the $\phi_i(x)$. The result is the pointwise value $f(x)$. The reproduction of $f(x)$ can hence also be understood as the well known idempotence of projection operators. Eq. 2 also shows that the reproducing property only holds for $f \in \mathcal{H}(X)$ and that otherwise $\tilde{f}(x) = \langle f(y), k_x(y)\rangle$ where $\tilde{f}$ is the L_2 projection of f onto $\mathcal{H}(X)$.

Example (Shannon sampling theorem (preliminaries)). *The reproducing kernel for the space $\Omega_B(\mathbb{R})$ of B-Fourier bandlimited functions is the sinc_B-kernel. Its representation in the Fourier domain is the box function $\chi_B(\xi)$. A convolution of the Fourier representation $\hat{f}(\xi)$ of an arbitrary signal $f(x)$ with $\chi_B(\xi)$ bandlimits $\hat{f}(\xi)$, which is the projection operation associated with reproducing kernels.*

2.2 Reproducing Kernel Bases

We will now employ reproducing kernels to construct basis representations whose basis function coefficients are pointwise samples.

Let $\mathcal{H}(X)$ be a separable reproducing kernel Hilbert space of dimension n over a set X with reproducing kernel $k_\lambda(x)$, for example the space spanned by the first N Legendre polynomials in Example 1. Furthermore, let $\Lambda = \{\lambda_i\}_{i=1}^m$ with $m \geq n$ be a set of locations $\lambda_i \in X$. The set Λ then generates m reproducing kernel functions $k_i(x) \equiv k_{\lambda_i}(x)$ (Example 1.b). Moreover, for a suitable choice of locations λ_i, see Sec. 2.3, the kernels $k_i(x)$ span $\mathcal{H}(X)$ (Example 1.c). The Hilbert space biorthogonality condition $\langle k_i(x), \tilde{k}_j(x)\rangle = \delta_{ij}$ then also defines dual kernel functions $\tilde{k}_j(x)$ (Example 1.d). With both the $k_i(x)$ and the $\tilde{k}_i(x)$, any function $f(x) \in \mathcal{H}$ can be represented by the basis expansion

$$f(x) = \sum_{i=1}^m \langle f(y), k_i(y)\rangle \, \tilde{k}_i(x). \tag{4a}$$

By the reproducing property of the $k_i(y)$ the equations equals

$$f(x) = \sum_{i=1}^m f(\lambda_i) \, \tilde{k}_i(x). \tag{4b}$$

Eq. 4b is at the heart of our work: it provides a Hilbert space basis expansion for which the basis function coefficients are sampled function values $f(\lambda_i)$. We can hence benefit from the convenience of basis expansions while working directly with the available pointwise information $f(\lambda_i)$. The crucial benefits of this combination will be the subject of the remainder of the paper. The basis pair

$$\left(\{k_i(x)\}_{i=1}^m, \{\tilde{k}_i(x)\}_{i=1}^m\right) \tag{5}$$

will be called a *reproducing kernel basis*.

Example (Shannon sampling theorem (cont'd)). *The Shannon sampling theorem can be stated as*

$$f(x) = \sum_{i=-\infty}^{\infty} f(i\,B) \, \mathrm{sinc}_B(x-i).$$

The above formula holds only if $f \in \Omega_B(\mathbb{R})$ and the famous Nyquist condition is satisfied From our point of view, the sinc_B-functions form an orthonormal reproducing kernel basis. Hence, the primary and dual kernel functions coincide and we have that

$$f(x) = \sum_{i=-\infty}^{\infty} \langle f(y), \, \mathrm{sinc}_B(y-i)\rangle \, \mathrm{sinc}_B(x-i).$$

The evaluation of the above inner product is made trivial by the reproducing property of the $\mathrm{sinc}_B(y-i)$.

In Eq. 5, the $k_i(x)$ only form a basis for $\mathcal{H}(X)$ when $m = n$. When $m > n$ they form an overcomplete frame (see the supplementary tutorial for a discussion of frames). In the following, we will continue to refer to Eq. 5 as a reproducing kernel basis even for $m > n$ and only distinguish bases and frames when there is, for our purposes, a material difference.

Example 1 (Legendre Polynomials). Let $\mathcal{H}_{\leq 4}$ be the space spanned by the first five Legendre polynomials $P_l(x)$ over $X = [-1.0, 1.0]$:


a.)
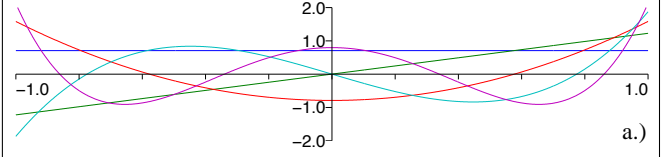
The reproducing kernel for $\mathcal{H}_{\leq 4}$ can be constructed using Eq. 3. Three example kernel functions for different locations λ are (the cross marks the reproducing point λ):


b.)
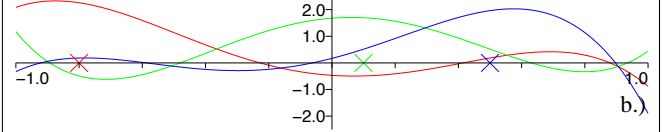
It can be observed that the reproducing kernel functions have local maxima close to the locations λ but their shape has no simple description and varies with varying λ. For five random points λ_i in $X = [-1.0, 1.0]$ the kernel functions $k_i(x)$ span $\mathcal{H}_{\leq 4}$ and hence provide a reproducing kernel basis for the space. When we choose five well separated locations we obtain the following reproducing kernel basis (crosses mark again the reproducing points λ_i):


c.)
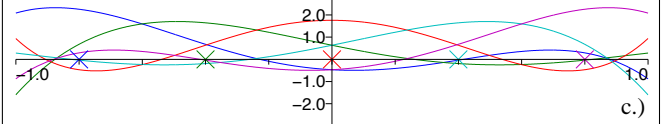
The corresponding dual kernel functions $\tilde{k}_i(x)$, defined by the biorthogonality condition $\langle k_i(x), \tilde{k}_j(x)\rangle = \delta_{ij}$, are (color used to indicate correspondence with the $k_i(x)$ in the above figure):

d.)
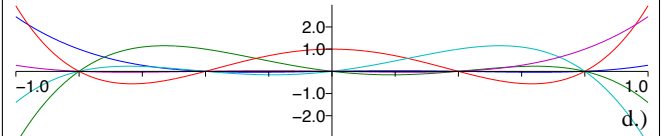
In contrast to the results in the above figure, for clustered locations λ_i the dual basis functions oscillate wildly, cf. the supplementary tutorial. As an example for a numerical technique we can construct a quadrature rule for $\mathcal{H}_{\leq 4}$ using Eq. 10. Randomly chosen locations yield for the quadrature weights w_i (w_i shown vertically at quadrature nodes, dual kernel functions $\tilde{k}_i(x)$ in the background):

e.)
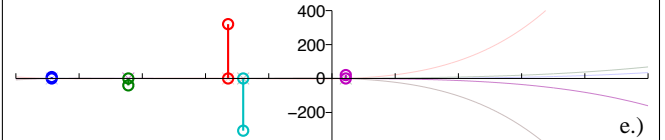
When we numerically optimize the sampling locations λ_i with the rk-discrepancy $\mu(\Lambda)$ as energy function, cf. Sec. 2.3, then the λ_i become more equally distributed and one obtains more uniform and better scaled quadrature weights:

f.)
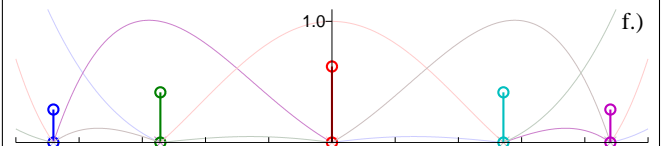
The rk-discrepancies before and after optimization are $\mu(\Lambda_{\mathrm{init}}) = 1362.9$ and $\mu(\Lambda_{\mathrm{opt}}) = 1.4286$, respectively. However, both quadrature rules are valid and for the first one the large weights compensate for the skewed distribution of the locations λ_i. We will see in Sec. 4 that much of the behaviour observed for $\mathcal{H}_{\leq 4}$ generalizes to the more complex settings encountered in light transport.

Kernel and Sampling Matrices   Functions in a Hilbert space are represented numerically by their basis function coefficients. When we consider Eq. 3 from this perspective, it provides the expansion of the reproducing kernel function $k_x(y) = k(x, y)$ with respect to basis functions $\phi_i(x)$ with coefficients $\phi_i(y)$. The m functions that form a reproducing kernel basis $\{k_i(x)\}_{i=1}^m$ are thus conveniently represented by the $m \times n$ matrix

$$K_\phi(\Lambda) = \begin{pmatrix} \phi_1(\lambda_1) & \cdots & \phi_n(\lambda_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\lambda_m) & \cdots & \phi_n(\lambda_m) \end{pmatrix} \in \mathbb{R}^{m \times n} \quad (6)$$

whose i^{th} row is given by the basis function coefficients $\phi_i(\lambda_j)$ of the ith kernel basis function $k_i(x) \equiv k_{\lambda_i}(x)$. We call $K \equiv K_\phi(\Lambda)$ the *kernel matrix*. The matrix is a special case of a basis matrix, see the supplementary tutorial, and it hence provides the change of basis from the orthonormal reference basis $\{\phi_i\}_{i=1}^n$ to the kernel basis. This can also be inferred directly by considering the scalar products of the rows of K with a coefficient vector $(f_1, \cdots, f_n)$. For the jth row one obtains the reconstruction formula $f(\lambda_j) = \sum_{i=1}^n f_i \, \phi_i(\lambda_j)$ and the resulting pointwise value $f(\lambda_j)$ is nothing but the jth basis function coefficients of the kernel basis. For most applications, the kernel matrix provides the numerically relevant representation of a reproducing kernel basis. Importantly, this representation can typically be determined to very high precision since it only requires the evaluation of the reference basis functions $\phi_i(x)$ at the λ_j.

Example (Vandermonde matrix). *When $\phi_i(x) = x^i$, namely when the reference basis functions are the monomials on $\mathbb{R}$, then the kernel matrix becomes the Vandermonde matrix.*

Example (Shannon sampling theorem (cont'd)). *By the zero crossings of the sinc_B-function, the infinite kernel matrix for the Shannon sampling theorem is*

$$K(\mathbb{Z}) = S(\mathbb{Z}) = \delta_{ij}. \quad (7)$$

That $K(\mathbb{Z})$ is diagonal immediately establishes that the Shannon sampling theorem employs an orthogonal reproducing kernel basis.

In the basis representation, the biorthogonality condition $\langle k_i(x), \tilde{k}_j(x) \rangle = \delta_{ij}$ that defines the dual kernel functions $\tilde{k}_i(x)$ is given by

$$KS \equiv KK^{-1} = \text{Id}. \quad (8)$$

We call the inverse of the kernel matrix the *sampling matrix* $S \equiv S_\phi(\Lambda) = K_\phi^{-1}(\Lambda)$ and denote its elements by s_{ij}. The columns of S contain the basis function coefficients of the dual kernel functions $\tilde{k}_i(x)$ with respect to $\{\phi_i\}_{i=1}^n$. Inverting the kernel matrix hence provides a means to explicitly construct the $\tilde{k}_i(x)$. It follows immediately from Eq. 8 that S provides the change of basis from the kernel basis to the orthonormal reference basis,

$$f_i = \sum_{j=1}^m s_{ij} \, f(\lambda_j). \quad (9)$$

The above reconstruction formula will be employed extensively in the following.

Example (Vandermonde matrix (cont'd)). *For the kernel functions $k_i(x)$ defined by the Vandermonde matrix, a closed form expression for the dual kernel functions $\tilde{k}_i(x)$ is given by the Lagrange interpolation polynomials. As shown in the accompanying Mathematica files, their basis expansion in the monomial basis can be obtained by inverting the Vandermonde matrix.*

As an example for how to utilize reproducing kernel bases for the development of numerical techniques that only employ pointwise information, let us consider the construction of quadrature rules. Representing the integrand in a reproducing kernel basis we obtain

$$\int_X f(x) \, dx = \int_X \Big(\sum_{i=1}^m f(\lambda_i) \, \tilde{k}_i(x) \Big) dx \quad (10a)$$

and exploiting linearity yields

$$\int_X f(x) \, dx = \sum_{i=1}^m f(\lambda_i) \int_X \tilde{k}_i(x) \, dx. \quad (10b)$$

By defining weights w_i as $w_i = \int_X \tilde{k}_i(x) \, dx$ we have

$$\int_X f(x) \, dx = \sum_{i=1}^m w_i \, f(\lambda_i). \quad (10c)$$

The right hand side is the prototype for a quadrature rule.

Example (Gauss-Legendre quadrature). *Gauss-Legendre quadrature is obtained from Eq. 10 when we employ a reproducing kernel basis for the space $\mathcal{P}_{N-1}$ spanned by the first N Legendre polynomials, cf. Example 1, with the zero crossings of the $N + 1$ Legendre polynomial $P_N(x)$ as sampling locations. See Example 36 in the accompanying tutorial.*

Example (Monte Carlo integration). *Monte Carlo integration is obtained from Eq. 10 as a quadrature rule for the space spanned by a characteristic basis, that is the basis formed by the characteristic or indicator functions induced by a partition $\{X_i\}_{i=1}^n$ of the domain X, see Sec. 1 in the supplementary material for details. The standard Monte Carlo estimator results from a uniform partition where all elements X_i have the same size while a partition whose elements are proportional to a given density $p(x)$ yields the Monte Carlo estimator with "importance sampling."*

2.3   Choice of Sampling Locations

So far we postulated the existence of a set $\Lambda = \{\lambda_i\}_{i=1}^m$ of locations λ_i such that the resulting reproducing kernel functions $k_i(x) = k_{\lambda_i}(x)$ span the space $\mathcal{H}(X)$. When mere existence is of interest, then almost any set of locations λ_i can typically be employed. In particular, for globally supported basis functions $\phi_i(x)$ any choice of distinct locations λ_i is suitable, up to a set of measure zero [Lessig 2012]. For example, one can employ low discrepancy sequences, cf. Sec. 4. For locally supported basis functions the situation is more complicated, but at least one sample point is needed in the support of every basis function and the distribution of the sampling points should reflect their local density; see our discussion on local reconstruction in Sec. 4 in the supplementary material.

For numerical computations, a qualitative characterization of the admissibility of sampling locations is not sufficient. Instead, we require the ability to quantitatively distinguish "good" and "bad" sampling sequences. Eq. 9, which describes the reconstruction of orthonormal basis function coefficients f_i from pointwise samples $f(\lambda_j)$, is equivalent to solving the linear system $f(\Lambda) = K\bar{f}(\phi)$. The accuracy with which the f_i can be obtained is hence determined by the condition number $\text{cond}(K)$ of the kernel matrix K. A quality measure $\mu(\Lambda)$ for the distribution of sampling locations can therefore be defined as

$$\mu(\Lambda) \equiv \mu_{\mathcal{H}(X)}(\Lambda) = \text{cond}(K_\phi(\Lambda)). \quad (11)$$

We will refer to $\mu(\Lambda)$ as *rk-discrepancy*, in analogy to classical discrepancy [Niederreiter 1992; Novak and Woźniakowski 2010; Dick and Pillichshammer 2010] that also measures the suitability of pointwise locations for numerical computations. With respect to rk-discrepancy, an optimal sampling sequence satisfies $\mu(\Lambda) = 1$ and the measure $\mu(\Lambda)$ increases as the quality of the locations decreases. When $\mu(\Lambda) = \infty$ a sequence is not admissible and the associated kernel functions do not span $\mathcal{H}(X)$. Since $\mu(\Lambda)$ can be determined numerically, one can employ it as objective function for the numerical optimization of locations. In contrast to most existing quality criteria for sampling distributions, $\mu(\Lambda)$ is thereby sensitive to the domain X and the function space $\mathcal{H}(X)$. As we will see in Sec. 4, this enables the construction of sampling sequences that are designed for the application at hand and which therefore provide

better performance than generic "well-distributed" sequences. Example 1.e) and 1.f) show the effect of optimization on quadrature weights for the simple example space considered there.

Aside from rk-discrepancy, other notions of optimal sampling sequences have been considered before in computer graphics. Classical L_p-discrepancy was introduced by Shirley [1991]; see [Keller 2006] for the state-of-the-art on the subject for image synthesis. However, L_p-discrepancy is restricted to functions in the Sobolev space $W^{1,1}([0,1]^d)$ over the d-dimensional hypercube and only very recent work, such as those by Basu and Owen [2014], considers functions over arbitrary domains. It hence does not allow easy tuning of sampling sequences to an application. To our knowledge, our notion of rk-discrepancy is mathematically not related to L_p-discrepancy. Signal specific sampling sequences were constructed for importance sampling [Clarberg et al. 2005; Jarosz et al. 2009]. However, these lack a rigorous theoretical basis, in particular for reconstructing signals. Mitchell [1987; 1991] studied optimal sampling and reconstruction using Fourier theory. His viewpoint only yields qualitative insights and also does not provide the possibility to tune sequences to an application. Similar restrictions apply to recent work by Nehab and Hoppe [2014].

Rather than using numerical optimization, we can also employ oversampling to improve the rk-discrepancy $\mu(\Lambda)$. Oversampling corresponds to increasing the redundancy of the frame, for which it is well known that this increases robustness [Mallat 2009]. For reproducing kernel frames this results in smaller reconstruction error from samples. As we will see in Sec. 4, often a combination of oversampling and optimization performs best in practice.

2.4 Non-bandlimited Input

The foregoing discussion did not consider an important aspect of any computational scheme that only relies on point sampling: insight into a continuous function from solely pointwise information is only possible when assumptions about the function are made. In our formulation, the reproducing property $f(x) = \langle f(y), k_x(y) \rangle$ only holds when $f(x)$ is an element in the Hilbert space $\mathcal{H}(X)$ associated with the reproducing kernel $k_\lambda(x)$. For light transport, suitable function spaces $\mathcal{H}(X)$ have not been determined. Nonetheless, it is clear that signals such as the light energy density will typically not be completely contained in the finite dimensional spaces that are spanned by a finite number of samples. For practical signals, the reproducing property will hence only be approximately satisfied and working with only pointwise samples will incur error.

An advantage of the functional analytic foundations of our formulation of sampling is that a practical characterization of the error resulting from non-ideal input signals is possible, cf. [Nashed and Walter 1991]. Assume that $\{\phi_i\}_{i=1}^p$ is an orthonormal basis for a Hilbert space $\mathcal{G}$ and that the subset $\{\phi_i(x)\}_{i=1}^n$ is an orthonormal basis for a Hilbert space $\mathcal{H} \subset \mathcal{G}$. Then for $f \in \mathcal{G}$, the error $\mathrm{err}_k(f)$ in the kth basis function coefficient f_k with respect to $\{\phi_i(x)\}_{i=1}^n$, when reconstructed from only pointwise samples $f(\lambda_i)$ using Eq. 9, is bounded by

$$|\mathrm{err}_k(f)| \leq \|\hat{f}\| \, \|\gamma_k\|. \tag{12}$$

The first term on the right hand side is the norm of the signal component $\hat{f} \in \mathcal{G} \setminus \mathcal{H}$ and the vector $\gamma_k = (\gamma_{n+1}^k, \cdots, \gamma_p^k)$ is given by $\gamma_i^k = \sum_{j=1}^m \phi_i(\lambda_j) \, s_{kj}$ where the s_{kj} are the elements of the sampling matrix. Intuitively, γ_i^k can be understood as the aliasing error of the $\phi_i(x)$, with $i > n$, when projected onto $\mathcal{H}$ using the sampling matrix. Note that Eq. 12 has the form of the Koksma-Hlawka inequality with a signal dependent term $\|\hat{f}\|$ and a term $\|\gamma_k\|$ that depends on the distribution of sampling points. See the supplementary material for a more detailed discussion of the above formula and experimental results.

Example (Gauss-Legendre quadrature (cont'd)). *As discussed previously, Gauss-Legendre quadrature of order N is designed for the space $\mathcal{P}_{N-1}$ spanned by the first N Legendre polynomials. Hence, one would expect that the rule is exact for functions in this space. However, the rule uses the roots of the $N+1$ Legendre polynomial $P_N(x)$ as sampling locations. For $\mathcal{G} = \mathcal{P}_N$, this leads to γ_k being the zero vector so that in fact all functions in $\mathcal{P}_N$ can be integrated exactly. See Example 36 in the supplementary tutorial for details.*

2.5 Discussion

Our constructive theory of sampling employs reproducing kernels $k_\lambda(x)$ to relate a continuous function $f : X \to \mathbb{R}$ to pointwise samples $f(\lambda)$ of it. A one-to-one correspondence is obtained using reproducing kernel bases (or frames) for which the set of samples $\{f(\lambda_i)\}_{i=1}^m$ provides the basis function coefficients. Such representations are possible since reproducing kernels $k_{\lambda_i}(x)$ are Hilbert space functions and not Schwartz distributions as Dirac deltas δ_{λ_i}. As we will see in the following, the Hilbert space setting of reproducing kernel bases is also crucial for the numerical practicality of our formulation and it ensures the existence of a recipe to translate from continuous formulas to practical pointwise techniques.

Our formulation of sampling subsumes many classical pointwise techniques, for example the Shannon sampling theorem, quadrature rules, and polynomial interpolation. These classical realizations of reproducing kernel bases are defined over the real line and for traditional function spaces. Additionally, they often rely on continuous symmetries and a uniform set of sampling locations. In contrast, our formulation enables the development of pointwise numerical techniques for arbitrary domains X and arbitrary functions spaces admitting a reproducing kernel for any admissible non-uniform sampling distributions. This generality is vital for applications to light transport where functions are defined over complex manifolds and signals are discontinuous.

Crucial to the effectiveness of our formulation of sampling is the choice of Hilbert space that is spanned by the reproducing kernel functions. As shown in Sec. 2.4, only when it provides a suitable approximation space (possibly up to a small residual) is it possible to effectively work with only pointwise samples. At first sight, an additional restriction of our theory is that the approximation space has to be a reproducing kernel Hilbert space. However, any finite function space, and hence any space that is possibly relevant for numerical computations, admits a reproducing kernel. Thus, although reproducing kernel Hilbert spaces are only a small subset of all Hilbert spaces over a fixed set X, all numerically relevant spaces are contained in this subset. A question that requires further investigation is convergence as the number of samples goes to infinity. We believe that existing work on meshless finite element methods, which we discuss in more detail in Sec. 3 in the supplementary material, might prove useful for this purpose.

A naïve application of our formulation of sampling is effective only for function spaces of low to moderate dimension since otherwise the construction and inversion of the sampling matrix becomes computationally too expensive. This is highly problematic for applications to light transport where it is not uncommon to have hundreds of millions of samples per image. One possibility to circumvent this restriction is to employ the characteristic basis. The kernel matrix is then diagonal and inversion is trivial. Additionally, for the characteristic basis inversion is a purely local problem, which becomes important in particular as the dimension increases. As discussed previously and detailed in Sec. 1 in the supplementary material, the characteristic basis is the setting of Monte Carlo integration. The prevalence of Monte Carlo integration in current rendering practice can hence also be explained from the point of view of our theory. Our theory, however, also suggests novel ways to generalize and optimize Monte Carlo techniques, for example by choosing other

bases that yield diagonal or almost diagonal kernel matrices. One such alternative is sketched in Sec. 4 in the supplementary material where we exploit that for any sufficiently locally supported basis the kernel matrix is almost diagonal. In the example, we employ wavelets over the interval as locally supported basis. By adaptively choosing the local depth of the wavelet tree this enables us to represent piecewise smooth functions with discontinuities in a space of essentially optimal dimensionality, cf. [Mallat 2009], while only requiring point samples as input.

Reproducing kernels have recently found applications in various areas of computational science, such as mesh-less finite elements, machine learning, and Quasi Monte Carlo integration. Since the literature is vast, we relegate a detailed discussion to Sec. 3 in the supplementary material. Nonetheless, an important distinction between existing approaches and our work is the use of reproducing kernel bases. These ensure an easy translation of continuous formulas to pointwise numerical techniques, which is of particular importance to an applied field such as computer graphics.

3 Light Transport Techniques Revisited

In this section, we employ reproducing kernel bases together with Petrov-Galerkin projection to develop a common theoretical basis for existing techniques for light transport simulation, including ones typically considered from a probabilistic perspective such as path tracing and photon mapping. We will begin by recalling the essentials of light transport theory in Sec. 3.1 and introducing reproducing kernel Galerkin projection in Sec. 3.2.

3.1 Light Transport Theory

Globally, light transport is described by the balance equation

$$\hat{\ell} = \ell_0 + T\hat{\ell} \tag{13}$$

where ℓ_0 is the emitted light energy density and $\hat{\ell}$ the steady state light energy density to be determined, cf. [Kajiya 1986; Arvo 1995a; Veach 1997; Lessig 2012]. Both ℓ_0 and $\hat{\ell}$ are defined over the tensor product $\mathcal{M}^2 \equiv \mathcal{M} \times \mathcal{M}$ of the scene manifold $\mathcal{M}$ and describe light energy density $\ell(\bar{x} \to x)$ transported from $\bar{x} \in \mathcal{M}$ to $x \in \mathcal{M}$. The operator T in Eq. 13 is an integral operator with kernel

$$t(\bar{x} \to x \to \bar{\bar{x}}) = \rho(\bar{x} \to x \to \bar{\bar{x}}) \, G(\bar{x} \leftrightarrow x) \tag{14}$$

which is formally defined over the double tensor product $\mathcal{M}^2 \times \mathcal{M}^2$ and but for which we additionally require that path edges $\bar{x} \to x$ and $x \to \bar{\bar{x}}$ meet in a common point x for the kernel to be nonzero. The operator T encodes transport in space, described through the so called geometry term $G(\bar{x} \leftrightarrow x)$ [Veach 1997], and surface scattering, given by the scattering kernel $\rho(\bar{x} \to x \to \bar{\bar{x}})$. The kernel is more commonly parametrized in local spherical coordinates at a point $x \in \mathcal{M}$. It then describes scattering as part of the shading equation

$$\bar{\ell}_x(\bar{\omega}) = \int_{H_x^2} \rho_x(\omega, \bar{\omega}) \, \ell_x(\omega) \, (n_x \cdot \omega) \, d\omega \tag{15}$$

where $\omega, \bar{\omega} \in S_x^2$ are directions. We will often write $\bar{\rho}_x(\omega, \bar{\omega}) = \rho_x(\omega, \bar{\omega})(n_x \cdot \omega)$. Since the transport operator is coercive for the physical scenes of interest to us and satisfies $\|T\| < 1$, Eq. 13 can be formally solved using the Neumann series [Kajiya 1986]

$$\hat{\ell} = \sum_{k=0}^{\infty} T^k \ell_0 = \ell_0 + T\ell_0 + T^2\ell_0 + \cdots . \tag{16}$$

Each application of T on the right hand side of Eq. 16 corresponds to one bounce of light propagation in the scene. T^k is hence defined over the $(k+1)$-fold tensor product $\mathcal{M}^{k+1}$. A point in $\mathcal{M}^k$ is known as a path $p^k = (x_0 \to \cdots \to x_k)$ and the unit-amount of light that is transported along a path p^k is given by the throughput

$$\pi(p^k) = t(x_0 \to x_1 \to x_2) \cdots t(x_{k-2} \to x_{k-1} \to x_k). \tag{17}$$

The space of all paths of length k is denoted by P^k and by the Neumann series the path space for a scene is $P = \bigcup_{k=0}^{\infty} P^k$.

3.2 Reproducing Kernel Galerkin Projection

Galerkin projection [Galerkin 1915; Petrov 1940] transforms an infinite-dimensional operator equation

$$A f = g, \tag{18}$$

into a finite or infinite-dimensional matrix-vector equation, see for example [Trenogin 2002]. For our purposes, we will assume that A maps a Hilbert space H into itself and furthermore that H is a reproducing kernel Hilbert space with orthonormal reproducing kernel basis $\{k_i\}$ defined over locations $\Lambda = \{\lambda_i\}$. Carrying out Galerkin projection of Eq. 18 then yields

$$\langle A f, k_j \rangle = \langle g, k_j \rangle \tag{19a}$$

$$\left\langle A\left(\sum_i f(\lambda_i) \, k_i\right), k_j \right\rangle = \langle g, k_j \rangle \tag{19b}$$

and by linearity

$$\sum_i f(\lambda_i) \, \langle A \, k_i, k_j \rangle = \langle g, k_j \rangle . \tag{19c}$$

Exploiting now the reproducing kernel property of the k_i we obtain

$$\sum_i f(\lambda_i) \, A(\lambda_i, \lambda_j) = g(\lambda_j). \tag{19d}$$

Hence, when Galerkin projection is performed with a reproducing kernel basis then the resulting matrix-vector equation only depends on pointwise values of the functions and operator. This is to be contrasted with Galerkin projection for an arbitrary basis where the finite representations of the input signal and operator are obtained by the evaluation of inner products. These can typically only be approximated and, as the experience with radiosity shows, in particular the approximation of the operator can be highly non-trivial. In the following, it will be convenient to call a reproducing kernel basis tight when it is either an orthonormal reproducing kernel basis, as we used above in the derivation in Eq. 19, or a normalized tight (or "snug" [Daubechies 1992]) reproducing kernel frame; see the supplementary material for a discussion of tight frames. Tightness is the essential condition for Eq. 19 to hold (possibly up to a constant).

3.3 Reproducing Kernel Galerkin Projection Light Transport

In the following, we will derive pointwise image synthesis techniques using reproducing kernel Galerkin projection.

Distributed Ray Tracing was introduced by Cook et al. [1984] to account for non-specular, global light transport effects. The technique approximates the shading equation in Eq. 15 using a Monte Carlo estimator [Cook 1986; Kajiya 1986] of the form

$$\bar{\ell}_x(\bar{\omega}) \approx \frac{1}{n} \sum_{i=1}^{n} \ell_x(\lambda_i) \, \bar{\rho}_x(\lambda_i, \bar{\omega}) \tag{20}$$

where the $\lambda_i \in H_x^2$ are uniformly distributed random directions in the hemisphere H_x^2 above $x \in \mathcal{M}$. Let us assume that the incoming light energy density $\ell_x(\omega)$ can be represented in a tight reproducing kernel basis $\{k_i(\omega)\}_{i=1}^n$ over sampling points $\Lambda = \{\lambda_i\}_{i=1}^m$ as

$$\ell_x(\omega) = \sum_{i=1}^{n} \ell_x(\lambda_i) \, k_i(\omega). \tag{21}$$

Performing Galerkin projection by inserting this representation of $\ell_x(\omega)$ into the shading equation in Eq. 15 we obtain

$$\bar{\ell}_x(\bar{\omega}) = \sum_{i=1}^{n} \ell_x(\lambda_i) \int_{H_x^2} k_i(\omega) \, \bar{\rho}_x(\omega, \bar{\omega}) \, d\omega \tag{22a}$$

and by the reproducing property of the $k_i(\omega)$ we have

$$\bar{\ell}_x(\bar{\omega}) = \sum_{i=1}^{n} \ell_x(\lambda_i) \, \bar{\rho}_x(\lambda_i, \bar{\omega}). \tag{22b}$$

Eq. 22b, obtained using reproducing kernel Galerkin projection, is, up to a factor of $1/n$, equivalent to Eq. 20. The factor of $1/n$ can be obtained by using a uniform characteristic reproducing kernel basis for $\{k_i\}_{i=1}^n$, as one would expect from our previous discussion of Monte Carlo integration.

Path Tracing was introduced by Kajiya as the first algorithm that could solve the complete light transport problem [Kajiya 1986]. The algorithm approximates the Neumann series in Eq. 16 using the Monte Carlo estimator

$$\ell\,|_{\text{cam}} \approx \frac{1}{n} \sum_{k=0}^{\infty} \sum_{p_i^k} \ell_0(x_0^i \to x_1^i)\,\pi(p_i^k) \qquad (23)$$

where $\ell\,|_{\text{cam}}$ is the light energy density reaching the camera plane and the p_i^k are uniformly distributed random paths with throughput $\pi(p_i^k)$, cf. Eq. 17.

Assuming we have a tight reproducing kernel basis that is defined over path edges $\lambda_i = (\bar{x}_i \to x_i)$ as sampling "points", a representation of the light energy density $\ell(\bar{x} \to x)$ is given by

$$\ell(\bar{x} \to x) = \sum_{i=1}^{m} \ell(\lambda_i)\,k_i(\bar{x} \to x) = \sum_{i=1}^{m} \ell(\bar{x}_i \to x_i)\,k_i(\bar{x} \to x).$$

We will furthermore assume that for every edge $\lambda_i = (\bar{x}_i \to x_i)$ there is only one other edge λ_j that continues from the endpoint x_i and satisfies $\lambda_j = (\bar{x}_j = x_i \to x_j)$; this condition will ensure that the samples connect to paths as used in path tracing. Performing Galerkin projection with the reproducing kernel representation of $\ell(\bar{x} \to x)$, a bounce $\ell^{k+1} = T\ell^k$ in the Neumann series in Eq. 16 is given by

$$\ell^{k+1}(x \to \bar{\bar{x}}) = \int_{\mathcal{M}} \ell^k(\bar{x} \to x)\,t(\bar{x} \to x \to \bar{\bar{x}})\,dA$$
$$= \sum_{i=1}^{m} \ell(\lambda_i) \int_{M} k_i(\bar{x} \to x)\,t(\bar{x} \to x \to \bar{\bar{x}})\,dA.$$

By the reproducing property of the $k_i(\bar{x} \to x)$ this equals

$$\ell^{k+1}(x \to \bar{\bar{x}}) = \sum_{i=1}^{m} \ell(\lambda_i)\,t(\lambda_i, x \to \bar{\bar{x}}).$$

Carrying out also the second part of the Galerkin method and projecting the left hand side onto the basis function $k_j(x \to \bar{\bar{x}})$, we obtain, after the usual steps and exploiting the reproducing kernel property, that

$$\ell^{k+1}(\lambda_j) = \sum_{i=1}^{m} t(\lambda_i, \lambda_j)\,\ell^k(\lambda_i) \qquad (25)$$

where, by its locality, the transport kernel $t(\lambda_i, \lambda_j)$ is nonzero if and only if sample points "line up" and for $\lambda_i = (x_1^i \to x_2^i)$ and $\lambda_j = (x_1^j \to x_2^j)$ it holds that $x_2^i = x_1^j$. With Eq. 25, we obtain for the next iterate $\ell^{k+2} = T^2\ell^k = T(T\ell^k) = T\ell^{k+1}$ in the Neumann series that

$$\ell^{k+2}(\lambda_a) = \sum_{i=1}^{m} t(\lambda_i, \lambda_j)\,\ell^{k+1}(\lambda_i) \qquad (26a)$$
$$= \sum_{i=1}^{m} t(\lambda_j, \lambda_a)\,t(\lambda_i, \lambda_j)\,\ell^k(\lambda_i) \qquad (26b)$$

where, by our assumptions on the edges, $\ell^{k+2}(\lambda_j)$ is nonzero only if λ_i, λ_j, and λ_a line up. Eq. 26 holds analogously for all following iterates of T^k. Inserting the definitions of a path and path throughput into Eq. 26 we immediately obtain

$$T^k\ell_0 = \sum_{p_i^k} \ell_0(\lambda_i)\,\pi(p_i^k). \qquad (27)$$

By the Neumann series in Eq. 16, the sought after steady state light energy density $\hat{\ell}$ is thus given by

$$\hat{\ell} = \sum_{k=0}^{\infty} \sum_{p_i^k} \ell_0(\lambda_0^i)\,\pi(p_i^k). \qquad (28)$$

where λ_j^i is the jth edge of the ith path p_i^k. Eq. 28, obtained using reproducing kernel Galerkin projection, is, again up to a factor of $1/n$ equivalent to Eq. 23, and again the factor can be recovered when we employ a characteristic reproducing kernel basis.

Photon Mapping was introduced by Jensen in the mid-1990s [Jensen and Christensen 1995; Jensen 1995] and it is the most popular of a larger class of algorithms based on density estimation, e.g. [Shirley et al. 1995; Walter et al. 1997]. Photon mapping represents the incoming light energy density ℓ in a scene by a set of pointwise samples $\ell(\lambda_i) = \ell(x_i, \omega_i)$ called photons. The continuous density $\ell(x, \omega)$ is reconstructed from the photons by

$$\ell(x, \omega) = \sum_{i=1}^{m} \ell(\lambda_i)\,\hat{p}(\lambda_i) \qquad (29a)$$
$$= \sum_{i=1}^{m} \ell(\lambda_i) \left(\frac{1}{m\,h} \sum_{j=1}^{m} k\left(\frac{|x_i - x_j|}{h} \right) \right) \qquad (29b)$$

where $\hat{p}(\lambda_i)$ is a density estimate for some suitable kernel $k(x, y) = k(x - y)$ and kernel width h. The above estimate is typically employed for the incoming light energy density in final gathering, which is used for actual image generation.

In the supplementary material we show how density estimation can be interpreted using reproducing kernel Galerkin projection. When the light energy density is represented in a tight reproducing kernel basis, then the density estimate $\ell(x, \omega)$ can be understood as the image under a Green's function $G(x, y)$:

$$\ell(x, \omega) = \int \ell(z)\,G(z, \bar{z})\,dz = \frac{n}{m} \sum_{i=1}^{m} \ell(\lambda_i)\,G(\lambda_i, y) \qquad (30)$$

where $z, \bar{z} \in S^2\mathcal{M}$. Eq. 29 results from Eq. 30 for special choices of $G(\lambda_i, y)$ such as the heat kernel, i.e. a scaled Gaussian. Green's functions are the reproducing kernel for the time flow of an associated partial differential equation [Aronszajn 1950]. This provides an alternative explanation for the smoothing that occurs when density estimation is employed. A reproducing kernel formulation of final gathering is presented in Sec. 4.3 where we also employ it for numerical computations.

Metropolis Light Transport was introduced by Veach and Guibas for the simulation of light transport in scenes that are difficult to handle with other approaches [Veach and Guibas 1997]. The technique constructs a continuous Markov chain whose steady state distribution is proportional to the steady state light energy density $\hat{\ell}$. Traversing this Markov chain many times then generates a sample distribution that is approximately proportional to $\hat{\ell}$.

For a continuous Markov chain with transition kernel $\pi(p, q) : P \times P \to [0, 1]$ defined over path space P, the steady state distribution $\hat{\ell}(p)$ is characterized by

$$\hat{\ell}(p) = \int_{P} \pi(p, q)\,\hat{\ell}(q)\,dq, \qquad (31)$$

that is, $\hat{\ell}(p)$ is the eigenfunction of the above integral operator associated with the largest eigenvalue $\lambda = 1$. When $\hat{\ell}$ is represented in a tight reproducing kernel basis, then Galerkin projection of Eq. 31 yields, in full analogy to the other examples we have already seen,

$$\hat{\ell}(\lambda_j) = \sum_{i=1}^{m} \hat{\ell}(\lambda_i)\,\pi(\lambda_i, \lambda_j). \qquad (32)$$

From our derivation it is not apparent that Eq. 32 is a faithful discretization of the Markov chain, that is that $\Pi = \{\pi(\lambda_i, \lambda_j)\}$ forms a stochastic matrix. Representing $\pi(p, q)$, as a function of p, in the reproducing kernel basis we obtain

$$1 = \int_{P} \pi(p, q)\,dp = \int_{P} \sum_{i=1}^{m} \pi(\lambda_i, q)\,k_i(p)\,dp \qquad (33a)$$
$$= \sum_{i=1}^{m} \pi(\lambda_i, q) \int_{P} k_i(p)\,dp. \qquad (33b)$$

Thus, P is a stochastic matrix when $\int_P k_i(x)\,dx = 1$. This is satisfied for the characteristic basis, with slight modifications in the above argument to accommodate for the fact that the basis is orthogonal but not orthonormal, cf. Sec. 1 in the supplementary material. Eq. 32 also immediately implies that $\hat{\ell} = (\hat{\ell}(\lambda_1), \cdots, \hat{\ell}(\lambda_n))$ is an eigenvector of Π and since

$$\hat{\ell}(x) = \sum_{i=1}^{m} \hat{\ell}(\lambda_i)\,k_i(x) \tag{34}$$

it is in fact a faithful finite representation of the sought after continuous steady state distribution. Now taking a frequentist approach and replacing the stationary distribution $f(x)$ with a distribution of samples proportional to it yields the Metropolis-Hastings algorithm. The Hasting's acceptance step is thereby justified by the balance condition which implies that the sample distribution has to be proportional to the steady state distribution.

3.4 Discussion

In this section, we showed that a wide range of existing sampling-based image synthesis techniques can be obtained using reproducing kernel Galerkin projection. For light transport, Galerkin projection was used previously for diffuse scenes with classical finite elements [Nishita and Nakamae 1984; Goral et al. 1984; Heckbert and Winget 1991] and higher order basis functions [Zatz 1993; Gortler et al. 1993; Schröder et al. 1993]. Christensen [1996; 1997] also explored its applicability for glossy environments. However, these techniques require the computation of basis function coefficients using inner products, which is typically expensive and can often only be approximated. In contrast, with reproducing kernel Galerkin projection one obtains techniques in which functions and operators are represented by their pointwise values. These can readily be determined numerically. The ansatz additionally provides a functional analytic alternative to the traditional probabilistic interpretation of techniques such as path tracing or Metropolis light transport, opening up new directions to improve and extend them, and to develop error formulas.

The main limitations of our formulation are the assumptions that a tight reproducing kernel basis exists for the domain of a technique and that all signals lie in the associated function space. Different examples of tight reproducing kernel bases exist for $\mathbb{R}^n$, for example the sinc-basis of the Shannon sampling theorem. On general domains, however, the only example we are aware of are characteristic bases. Using the characteristic basis is consistent with our formulation of Monte Carlo integration as a quadrature rule and it connects the classical, probabilistic interpretation of sampling-based techniques with our functional analytic one. Additionally, the characteristic basis becomes dense in L_2 as the number of partitions increases so that any signal can be approximated arbitrarily well for a sufficiently large number of partitions. For most signals, however, the order of approximation achieved with the characteristic basis is far from optimal. The formulation presented in this section enables the development of pointwise numerical techniques with bases that are better adapted to the properties of the light energy density and hence yield faster convergence rates; our example in Sec. 4.3 can be seen as a first realization of this idea. Next to ideal tight reproducing kernel bases one can often employ oversampling together with the numerical optimization of sampling locations to obtain frames that behave in practice like their ideal counterparts. The application of such representations for light transport provides an interesting direction for future work. Another avenue for future works is to employ existing results on meshless finite element methods to study convergence of reproducing kernel Galerkin projection for light transport.

4 From Theory to Practice

In this section, we will employ our constructive theory of sampling to develop three applications: the representation and transport of color information using pointwise samples, the projection of the angular part of the light energy density into spherical harmonics, and the approximation of the local shading equation from a photon map.

For all three applications we will employ the same recipe to develop a pointwise numerical technique:

1. Choose a reference basis that is well suited for the approximation of the signals of interest.

2. Construct a reproducing kernel basis for the space spanned by the reference basis by choosing sampling locations. If the locations are pre-determined, optimize the locations to minimize rk-discrepancy.

3. Employ reproducing kernel Galerkin projection to map the continuous description to a pointwise formula.

4. Implement the technique using the kernel and sampling matrices of the reproducing kernel basis.

For each application, some adjustments to the above recipe are necessary. Nonetheless, following the four steps is typically sufficient for the remaining details to become transparent.

4.1 Representation of Color Information

Most techniques for light transport simulation represent color using pointwise samples of the continuous frequency spectrum, for example using three samples $\lambda_{\{r,g,b\}}$ that correspond to red, green, and blue, cf. Fig. 4. A notable exception is the technique proposed by Peercy [1993] where color $\ell(\lambda)$ as a function of wavelength λ is represented using basis functions $\gamma_i(\lambda)$,

$$\ell(\lambda) = \sum_{i=1}^{n} \langle \ell(\bar{\lambda}), \gamma_i(\bar{\lambda}) \rangle\,\gamma_i(\lambda). = \sum_{i=1}^{n} \ell_i\,\gamma_i(\lambda). \tag{35}$$

In this formulation, surface interactions are modelled by

$$\bar{\ell}_j = \sum_{i=1}^{n} R_{ji}\,\ell_i \tag{36}$$

where ℓ_i and $\bar{\ell}_j$ are the basis function coefficients before and after scattering, respectively, and the R_{ij} are determined using Galerkin projection of the reflection model. Peercy demonstrated that his technique can lead to substantially improved results, in particular when PCA is used to determine a suitable basis $\{\gamma_i\}_{i=1}^{n}$. Using our formulation of sampling, we can obtain a technique that only uses point samples at discrete wavelengths λ_i, as is common in existing rendering systems, but properly represents the continuous spectrum. Using the PCA basis proposed by Peercy, we construct a reproducing kernel basis for the same space using Eq. 3. Then $\ell(\lambda)$ is given by

$$\ell(\lambda) = \sum_{i=1}^{m} \ell(\lambda_i)\,\tilde{k}_i(\lambda). \tag{37}$$

With $K \equiv K_\gamma(\Lambda)$ being the kernel matrix for the reproducing kernel basis, Eq. 36 becomes

$$\bar{\ell}(\lambda_j) = \sum_{i=1}^{m} (KRK^{-1})_{ji}\,\ell(\lambda_i). = \sum_{i=1}^{m} R(\lambda_j, \lambda_i)\,\ell(\lambda_i) \tag{38}$$

We hence have the following algorithm:

Experimental Results In Fig. 3 we compare classical point sampling with Peercy's linear model and our technique. The results demonstrate that our technique leads to substantially less color distortion than traditional point sampling while using the same pointwise

Figure 3: *Experimental results for the representation of color information using (from left to right) point sampling with four spectral samples, Peercy's linear model with two basis functions, our technique with two basis functions and four samples, a reference solution with 50 spectral samples.*

Precomputation:
– Construct PCA basis and choose sampling locations $\Lambda = \{\lambda_i\}$.
– Construct kernel matrix $K \equiv K(\Lambda)$.
– Compute $R_b(\lambda_j, \lambda_i) = K R K^{-1}$ for every shading model R_b.

At every bounce: compute $\ell(\lambda_j)$ using Eq. 38.

information only. The figures were generated with a basic implementation of the above algorithm in pbrt [Pharr and Humphreys 2010]. Additional results are provided in the supplementary material.

4.2 Computation of Spherical Harmonics Coefficients

In various areas of image synthesis, such as precomputed radiance transfer [Sloan et al. 2002; Lehtinen 2007], volume rendering [Kajiya and von Herzen 1984], shading [Sillion et al. 1991; Westin et al. 1992], and irradiance probes [Greger et al. 1998; Pharr and Humphreys 2010], spherical harmonics $y_{lm}(\omega)$ are used to represent signals defined over the sphere. A typical example, and the one we will consider in the following, is the representation of the incoming light energy density $\ell_x(\omega)$ in the basis One then has to compute coefficients

$$\ell_{lm} = \langle \ell_x(\omega), y_{lm}(\omega) \rangle = \int_{S^2} \ell_x(\omega)\, y_{lm}(\omega)\, d\omega \qquad (39)$$

from typically only pointwise values $\ell_x(\lambda_j)$ of the light energy density for directions $\lambda_j \in S^2$. In the literature, Monte Carlo integration is typically used to determine the ℓ_{lm}. Instead, we formulate the problem as a change of basis from a reproducing kernel basis for the space $\mathcal{H}_{\leq L}$ spanned by all spherical harmonics up to band L to the Legendre spherical harmonics basis for the space. Using Eq. 9, the $i \equiv (l, m)$-th spherical harmonics coefficient is then given by

$$\ell_i = \sum_{j=1}^{m} s_{ij}\, \ell_x(\lambda_j) \qquad (40)$$

where the s_{ij} are the elements of the sampling matrix $S_{y_{lm}}(\Lambda)$. We hence have the following algorithm:

Precomputation:
– Choose sampling directions $\Lambda = \{\lambda_i\}$ and optimize locations.
– Construct $K_{y_{lm}}(\Lambda)$ and $S_{y_{lm}}(\Lambda) = K_{y_{lm}}^{-1}(\Lambda)$.

At every sampling point x in the scene:
– Trace rays for the directions λ_i and determine incoming $\ell_x(\lambda_i)$.
– Reconstruct spherical harmonics coefficients ℓ_{lm} using Eq. 40.

For an orthonormal or tight reproducing kernel basis the sampling matrix $S_{y_{lm}}(\Lambda)$ is given by $K_{y_{lm}}(\Lambda)^T$. Eq. 40 then becomes

$$\ell_i = \sum_{j=1}^{m} y_i(\lambda_j)\, \ell(\lambda_j). \qquad (41)$$

Up to a constant, this is the Monte Carlo estimator for the basis function coefficients ℓ_i. The entries s_{ij} of the sampling matrix S for a non-orthogonal kernel basis can hence be interpreted as correcting for non-ideal sampling locations. Note that this correction incurs almost no additional cost, in particular when compared to the costs for tracing rays.

Experimental Results We developed two implementations to experimentally evaluate the reconstruction of spherical harmonics coefficients using the above algorithm. We employed Matlab to perform synthetic experiments with precise control of the inputs and parameters and we modified the existing irradiance probe implementation in pbrt [Pharr and Humphreys 2010] to understand the behavior for real-world applications. As sampling directions we used well known low discrepancy sequences [Niederreiter 1992] and the spiral points proposed by Saff and Kuijlaars [1997]. Sample point optimization as discussed in Sec. 2.3 was performed using an existing L-BFGS algorithm [Zhu et al. 1997]. Fig. 5 and Fig. 6 show that our technique can lead to substantial performance improvements compared to classical (Quasi) Monte Carlo integration. The figure also demonstrates that the numerical optimization of sampling points provides significant performance benefits, both for synthetic tests and real world applications, and that it is useful to combine it with oversampling. A visual comparison of unoptimized and optimized sampling locations is shown on the right in Fig. 6 and in the accompanying video.

4.3 Final Gathering

Photon mapping [Jensen 1995; Jensen and Christensen 1995; Jensen 2001] with final gathering is a widely used techniques for image synthesis. Classically, final gathering approximates the outgoing light energy density $\bar{\ell}_x(\bar{\omega})$ towards the camera given by

$$\bar{\ell}_x(\bar{\omega}) = \int_{H_x^2} \bar{\rho}_x(\omega, \bar{\omega})\, \ell_x(\omega)\, d\omega \qquad (42a)$$

using a Monte Carlo or Quasi Monte Carlo approximation

$$\bar{\ell}_x(\bar{\omega}) \approx \frac{1}{n} \sum_{i=1}^{n} \bar{\rho}_x(\lambda_i, \bar{\omega})\, \ell_x(\lambda_i). \qquad (42b)$$

The integral kernel $\bar{\rho}_x(\omega, \bar{\omega})$ is the cosine-weighted scattering kernel at x and $\ell_x(\omega)$ is the incoming light energy density estimated from the photon map for directions $\lambda_i \in H_x^2$.

Our alternative technique for final gathering is based on the operator formulation $\bar{\ell} = P\ell$ of the shading equation [Arvo 1995b]. The shading operator P is of Hilbert-Schmidt type [Lessig 2012] and hence has a well defined eigen-decomposition consisting of distinct eigenvalues σ_i and eigenfunctions φ_i. The incoming light energy density ℓ_x can thus be applied very efficiently to P by representing ℓ_x in the eigenbasis $\{\varphi_i\}_{i=1}^{n}$ and multiplying the basis function coefficients ℓ_i with the eigenvalues σ_i. For numerical computations, only a finite subset of the in general infinitely many σ_i can be employed. The optimal approximation using k eigenvalues is obtained when the k largest ones are used and $\ell_x(\omega)$ is represented using the associated eigenfunctions [Mallat 2009]. To implement the approach, we hence
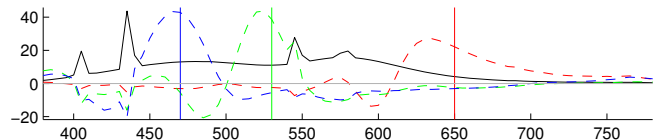

Figure 4: *Spectrum of CIE fluorescent light illuminant F1 (black), typical point samples for blue, green, and red (vertical lines), and reproducing kernel basis functions for Peercy's PCA basis (dotted) for these locations.*
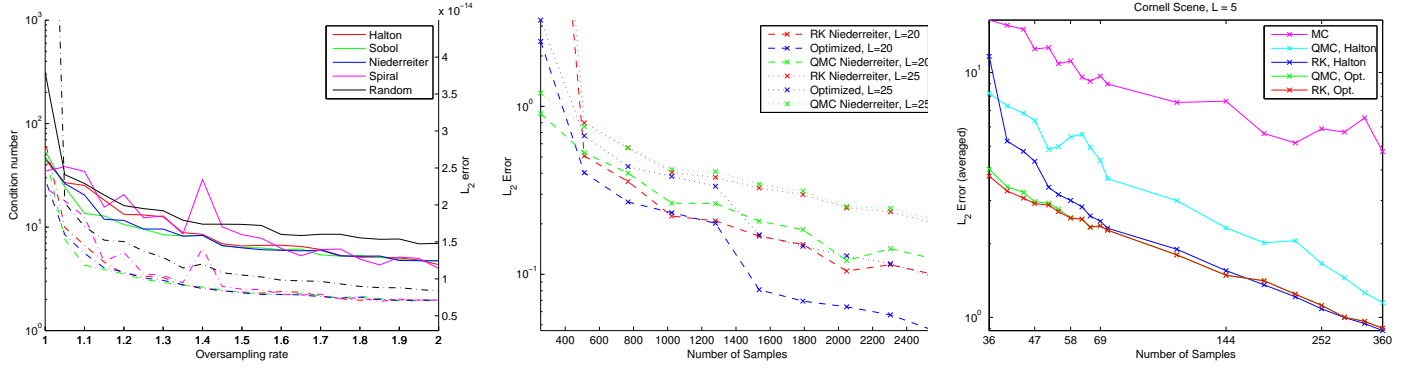
Figure 5: *Experimental results for the projection of signals into the space $H_{\leq L}$ spanned by the first L spherical harmonics bands. Left: To demonstrate the effect of oversampling, we reconstruct an ideal input signal from different sampling sequences mapped from $[0, 2\pi] \times [0, \pi]$ onto the sphere using an area preserving mapping. As expected, for all sequences the reconstruction error decreases as the oversampling rate, and hence the number of samples, increases. The plot also demonstrates that random points provide consistently higher error rates than well distributed ones. Shown in the graph as dashed lines are the condition numbers of the kernel matrices for the different sampling sets. These correlate well with the reconstruction error. Middle: To demonstrate the robustness of our approach, we reconstruct spherical harmonics coefficients for $\mathcal{H}_{\leq 15}$ from a large number of non-ideal input signals with bandwidth $L = 20$ and $L = 25$. Reported are average error rates. Our technique (RK Niederreiter, i.e. Niederreiter points mapped to the sphere used with our technique for reconstruction) outperforms quasi Monte Carlo integration (QMC Niederreiter) in both cases although the performance advantage becomes smaller as L increases. Point locations optimized for our technique (blue curves) attain considerably lower error rates. Right: Reconstruction error for spherical harmonics coefficients for the incoming light energy density at points in the Cornell box scene; shown is the error averaged over a large number of scene points. Continued in Fig. 6.*

need the first k basis function coefficients ℓ_i of the incoming light energy $\ell_x(\omega)$ in the eigenbasis. Since only pointwise values $\ell_x(\lambda_i)$ are available, we will employ our theory to efficiently determine the coefficients ℓ_i.

Derivation By performing Galerkin projection of the shading equation using the first k eigenfunctions φ_i we obtain

$$\bar{\ell}_x(\bar{\omega}) \approx P\,\ell = \sum_{i=1}^{k} (\sigma_i\,\ell_i)\,\varphi_i(\bar{\omega}) \tag{43}$$

and we know that this is the optimal linear k-term approximation [Mallat 2009] of the continuous shading equation. To determine the outgoing light energy density $\bar{\ell}(\bar{\omega})$ using Eq. 43, we need the expansion coefficients $\ell_i = \langle \ell_x(\omega), \varphi_i(\omega) \rangle$ with respect to the eigenbasis $\{\varphi_i\}_{i=1}^{k}$. However, the only information about $\ell_x(\omega)$ that is immediately available are samples $\ell(\lambda_i) \equiv \ell_x(\lambda_i)$. We compute the coefficients ℓ_i from only this information by constructing a reproducing kernel basis for the space spanned by the first k eigenfunctions for the given sampling directions $\Lambda = \{\lambda_j\}_{j=1}^{m}$. The reproducing kernels for the basis can be obtained using Eq. 3 and the corresponding dual kernel functions $\tilde{k}_i(x)$ are defined by the biorthogonality condition. We then have

$$\ell_x(\omega) \approx \sum_{j=1}^{m} \ell(\lambda_j)\,\tilde{k}_j(\omega) \tag{44}$$

and by Eq. 9 the coefficients ℓ_i are given by

$$\ell_i \approx \sum_{j=1}^{m} s_{ij}\,\ell(\lambda_j) \tag{45}$$

where the s_{ij} are the elements of the sampling matrix. Inserting into Eq. 43 we obtain

$$\bar{\ell}(\bar{\omega}) \approx \sum_{i=1}^{k} \left(\sigma_i \sum_{j=1}^{m} \ell(\lambda_j)\,s_{ij} \right) \varphi_i(\bar{\omega}). \tag{46a}$$

Reordering the summation over i and j yields

$$\bar{\ell}(\bar{\omega}) \approx \sum_{j=1}^{m} \ell(\lambda_j) \underbrace{\sum_{i=1}^{k} \sigma_i\,s_{ij}\,\varphi_i(\bar{\omega})}_{w_j(\bar{\omega})} \tag{46b}$$

and with the weights $w_j(\bar{\omega})$ we have

$$\bar{\ell}(\bar{\omega}) \approx \sum_{j=1}^{m} w_j(\bar{\omega})\,\ell(\lambda_j). \tag{46c}$$

Eq. 46c provides a quadrature rule for final gathering that determines the outgoing light energy density towards the camera using an optimal k-term approximation of the scattering operator P. Before we turn to its implementation, let us consider how to perform importance sampling for the quadrature rule.

The weights $w_j(\bar{\omega})$ in Eq. 46c are associated with the sampling directions λ_j and evidently when $|w_j(\bar{\omega})|$ is small then the associated sample $\ell(\lambda_j)$ will have a small contribution to the outgoing light energy density $\bar{\ell}(\bar{\omega})$. Hence, "important" directions λ_j are those where the weight $w_j(\bar{\omega})$ is large. For example, the figure on the right shows the weights $w_j(\bar{\omega})$ for a Phong BRDF with the red cross marking the direction of specular reflection for the outgoing direction $\bar{\omega}$. To generate importance sampled directions we therefore begin with a larger number $M > m$ of sampling points λ_j than is desired, compute the $w_j(\bar{\omega})$ for these points, and retain only the m directions λ_j with largest weights.

Numerical Realization To implement our technique, we have to determine the eigenfunctions for an arbitrary scattering operator P. Except for special cases, such as in [Mahajan et al. 2008], no analytic form exists and the eigenfunctions have to be determined numerically. We employ again Galerkin projection to obtain a continuous representation of the eigenfunctions that allows us to easily evaluate them in any direction $\bar{\omega}$. Using Eq. 19 yields

$$\hat{P}_{ij} = \langle \psi_i(\omega) \mid \bar{\rho}(\omega, \bar{\omega}) \mid \psi_j(\bar{\omega}) \rangle \tag{47}$$

where ψ_i is a suitable, finite dimensional reference basis. For its numerical convenience and since the scattering functions of interest to us are smooth, and hence can be approximated efficiently
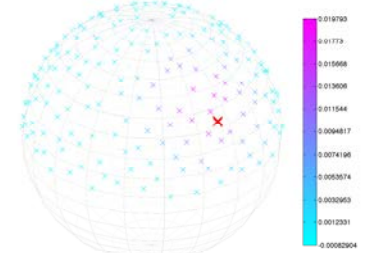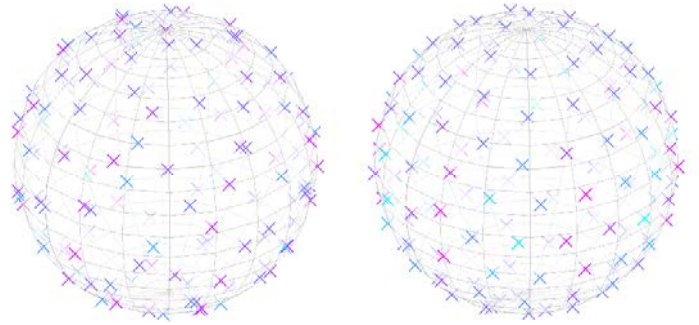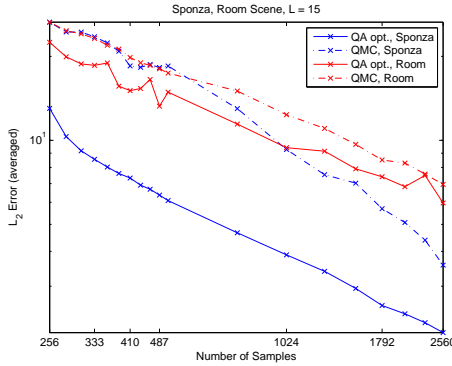
Figure 6: *Continuation of Fig. 5.* Left*: Reconstruction error for spherical harmonics coefficients for the Sponza scene and a room interior. The lighting in the room scene is more complex than in the Sponza scene and hence the advantage of our technique designed for signals in $\mathcal{H}_{\leq 15}$ is smaller.* Right*: Niederreiter sequence on the sphere S^2 before (left) and after optimization (right) for the space $\mathcal{H}_{\leq 10}$ with 10% oversampling. Shown are also the quadrature weights for the projection onto one spherical harmonics basis function (violet = large, cyan = small). Note how, in accordance with our theory, the optimization increases the uniformity of the points on S^2 and homogenizes the quadrature weight. An animation of the optimization process is shown in the accompanying video.*

using a smooth basis [Mallat 2009], we employ spherical harmonics $y_i(\omega) \equiv y_{lm}(\omega)$ up to some band L as reference basis. With the matrix $\hat{P}$ formed by the $\hat{P}_{ij}$ we numerically compute its eigen-decomposition and obtain a representation of the eigenfunctions in the spherical harmonics domain. Note that since we included the co-sine term in our definition of the scattering kernel, $\hat{P}$ is not symmetric and hence the singular value decomposition $\hat{P} = \hat{U}_l \hat{\Sigma} \hat{U}_r^T$ must be computed to obtain the eigen-decompositon. This also requires distinguishing between left and right eigenfunctions φ_i^l and φ_i^r, respectively, in Eq. 45 and Eq. 46. Given the eigen-decomposition, we can compute the quadrature weights $w_j(\bar{\omega})$ in Eq. 46c by constructing the kernel matrix

$$K = \begin{pmatrix} \varphi_1^r(\lambda_1) & \cdots & \varphi_n^r(\lambda_1) \\ \vdots & \ddots & \vdots \\ \varphi_1^r(\lambda_m) & \cdots & \varphi_n^r(\lambda_m) \end{pmatrix} \qquad (48)$$

and solving for the elements s_{ij} of its inverse. The weights are then

$$w_j(\bar{\omega}) = \sum_{i=1}^{k} \sigma_i \, s_{ij} \, \varphi_i(\bar{\omega}) \qquad (49a)$$

$$= \sum_{i=1}^{k} \sigma_i \, s_{ij} \sum_{j=1}^{n} \varphi_j^{l,i} \, y_j(\bar{\omega}) \qquad (49b)$$

where $\varphi_j^{l,i}$ is the basis function coefficients of the left eigenfunction φ_i^l with respect to the jth spherical harmonic; equivalently $\varphi_j^{l,i}$ is the $(j, i)^{\text{th}}$ element of the matrix $\hat{U}_l$. We thus have the following algorithm to determine the outgoing light energy density $\bar{\ell}_x(\bar{\omega})$ for all shading point $x \in \mathcal{M}$:

Precomputation:
– For every scattering model ρ compute $\hat{P}_{ij}$ using Eq. 47 and compute $\hat{P} = \hat{U}_l \hat{\Sigma} \hat{U}_r^T$.

At every shading point x in the scene:
– Trace rays for directions $\Lambda = \{\lambda_i(x)\}$ and determine $\ell_x(\lambda_i)$.
– Construct kernel matrix $K_x(\Lambda)$ in Eq. 48 and $S_x(\Lambda) = K_x(\Lambda)^{-1}$.
– Compute quadrature weights using Eq. 49.
– Evaluate outgoing light energy density $\bar{\ell}_x(\bar{\omega})$ using Eq. 46c.

Implementation Details We briefly describe important aspects of our reference implementation of the above final gathering technique in pbrt [Pharr and Humphreys 2010]; additional details are provided in the supplementary material.

In the previous discussion, we considered one scattering function. The scattering kernels used in most rendering systems are linear

combinations of elementary scattering functions,

$$\rho_x(\omega, \bar{\omega}) = \sum_{i=1}^{r} \alpha_i(x) \, \rho_x^i(\omega, \bar{\omega}). \qquad (50)$$

where the $\rho_x^i(\omega, \bar{\omega})$ are, for example, the Lambertian, Phong, or Microfacet scattering model and the spatially varying linear weights $\alpha_i(x)$ are obtained from a texture. A naïve implementation of our technique would require the computation of the eigen-decomposition of $\rho_x(\omega, \bar{\omega})$ for every point x on the scene manifold that is sampled from the camera. This is infeasible even for simple scenes. Fortunately, both Eq. 46c and Eq. 50 are linear. We can hence determine the eigen-decomposition and quadrature weights $w_j^i(\bar{\omega})$ for each model $\rho_x^i(\omega, \bar{\omega})$ separately and independent of x and then linearly combine them,

$$w_j(\bar{\omega}) = \sum_{i=1}^{r} \alpha_i(x) \, w_j^i(\bar{\omega}). \qquad (51)$$

In addition to the spatial variations of the linear weights $\alpha_i(x)$ one sometimes also varies a nonlinear parameter. A classical example is the modulation of the Phong exponent. This can currently not be accommodated by our technique.

As sampling directions λ_i in the hemisphere above each point x we currently use quasi-random points generated by pbrt. This requires computing weights for each x, which takes a substantial amount of time. However, it allows for better comparisons to classical final gathering, which uses the same locations, and it also simplifies the implementation. Alternatives will be discussed shortly.

Experimental Evaluation We evaluated our technique for final gathering using three scenes available for pbrt: the classical Sponza scene, a room with a Killeroo, and a Villa interior. For the Sponza scene we used a Phong shading model with exponent 8.0 and not a purely Lambertian shading since the eigenspace for the latter one is 1-dimensional and our technique then exactly reproduces classical importance sampling with respect to the cosine term. As
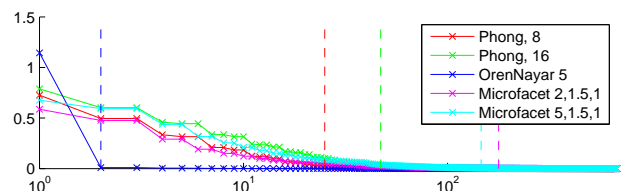

Figure 7: *Eigenvalue distribution for different scattering functions. The vertical lines are at 95% of the spectral energy.*
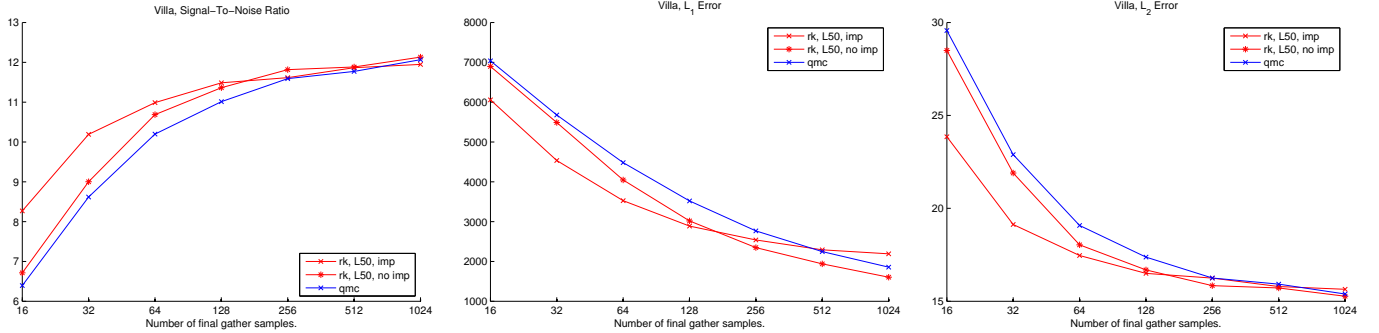
Figure 8: *Signal-to-Noise ratio, L_1 error, and L_2 error (from left to right) for our technique with importance sampling (full), without importance sampling (dashed), and QMC with importance sampling (dash-dot, full circles) for the Villa scene (see Fig. 9 for the corresponding images).*

reference, we employed pbrt's implementation of final gathering with importance sampling. To avoid distortion due to the image filter, all images were generated with one sample per-pixel and with a per-pixel box filter. Error rates as a function of gather samples are reported in Fig. 8 and Fig. 9 for the Villa scene and in the supplementary material for the other two scenes. For each scene and sample count we report the best result from using {85%, 85%, 90%, 92.5%, 95%, 97.5%, 99%, 99.5%} of the total spectral energy.

The experimental results demonstrate that for the Villa and the Killeroo scene our technique provides substantially lower error rates and signal-to-noise ratios than Quasi Monte Carlo integration and the latter technique only asymptotically attains our performance. The lower error rates also correspond to substantially reduced visual noise. Visible in the images is a small distortion and slight banding artifacts. These result from the finite truncation of the spectrum and the fact that the eigenfunctions in general do not align with a visually pleasing subdivision of the domain of the shading operator. We experimentally confirmed that the possible negativity of $\bar{\ell}(\bar{\omega})$, see below, does not substantially contribute to the distortion. The graphs also show that importance sampling is beneficial mainly for low sampling densities. For the Sponza scene the improvements are more modest than for the other two scenes and in fact there we attain higher L_p error rates than Quasi Monte Carlo integration. This results mainly from the very diffuse light scattering in the scene and the hence very low effective dimension of the shading operator, cf. Fig. 7. The effect of the spherical harmonics bandwidth is shown in the supplementary material and the accompanying video.

Limitations and Future Work   With the use of spherical harmonics and the singular value decomposition, we cannot guarantee that the outgoing light energy density is positive. In future work it would be interesting to develop a positive basis for the hemisphere and to explore if the eigen-decomposition can be replaced by a positive matrix factorization, cf. [Donoho and Stodden 2004]. Instead of developing a positive basis, improvements can also be expected from using spherical Slepian functions [Simons 2010; Lessig and Fiume 2010], localized for the hemisphere, which are readily available. Selecting the L_2-optimal k-term approximation is not always visually most pleasing. One can hence envision selecting k basis functions based on visual reconstruction quality, cf. [De Witt et al. 2012]. This might alleviate the slight banding artifacts that are present for example in the Killeroo scene. In our implementation most computation time is spent on determining the quadrature weights at each location $x \in \mathcal{M}$. Since this involves inverting a matrix, our current implementation is substantially slower than classical final gathering. This can be avoided by precomputing a set of sampling sequences over the hemisphere and randomly selecting one of them at runtime or by applying a random rotation to one or more precomputed sequences. Applying a random rotation is interesting in particular with our

choice of spherical harmonics as reference basis since then the diagonal representation of the shading operator is closed under rotation and efficient algorithms for the rotation exist [Pinchon and Hoggan 2007; Lessig et al. 2012]. A precomputation of sequences would also enable the use of optimized sampling locations which provided significant benefits in Sec. 4.2. Another limitation of our technique is that it requires a linear parameter dependence. For a varying Phong parameter, one possibility might be to take the logarithm before computing the eigen-decomposition and exponentiating afterwards. In general, one could envision parameter dependent eigen- or tensor decompositions.

4.4   Discussion

In this section, we demonstrated the versatility of our constructive formulation of sampling for light transport with three example applications: pointwise sampling of color information, projection of the angular part of the light energy density into spherical harmonics, and approximation of the shading equation for final gathering. Another application of reproducing kernel bases has appeared elsewhere [Lessig et al. 2012]. In contrast to existing approaches to sampling, our formulation is inherently designed for a finite number of samples and it can be adapted to a specific setting, for example by choosing a suitable function space or by optimizing sampling locations. We believe that this explains our improved experimental results. The source code for all applications is available online.[1]

5   Outlook

In this paper, we developed a constructive theory of sampling for light transport simulation. In contrast to traditional formulations that are based on the Dirac delta, our theory employs reproducing kernels to conceptualize the correspondence between continuous signals $f(x)$ and their pointwise samples $f(\lambda)$. Since reproducing kernels are Hilbert space functions and not Schwartz distributions, they can be employed as basis functions. The resulting reproducing kernel bases provide the numerical convenience of Hilbert space representations while only requiring pointwise samples $f(\lambda_i)$ as basis function coefficients. With this:

- We enable the construction of sampling theorems, quadrature rules, and interpolation schemes for arbitrary domains and approximation spaces. This is crucial for light transport where signals are defined over complex manifolds and the spaces classically used in approximation theory are ill-suited. Additionally, we can tailor techniques to the finite sample counts and non-equidistant sample distributions employed in practice. This provides precise condition for their applicability, for example when they yield perfect reconstruction, and enables the derivation of error formulas.

---

[1]www.dgp.toronto.edu/people/lessig/cosalt/.

Figure 9: *Experimental results for final gathering for our technique (top row) and classical final gathering (bottom row). The columns correspond to 16, 32, 64, 128, 256, 512, and 1024 samples in the hemisphere, respectively. See Fig. 8 for the corresponding error rates. Results for other scenes are presented in the supplementary material.*

- We overcome the dichotomy between radiosity-based and ray tracing techniques that persists in the literature. By introducing reproducing kernel Galerkin projection we are able to also derive sampling-based techniques such as path tracing and Metropolis light transport using a functional ansatz. This also provides an alternative to the more common probabilistic formulation of these techniques.

- We introduce rk-discrepancy as a novel quality measure for the distribution of sampling points. The measure can be determined numerically, which enables us to optimize sampling locations for an application.

We demonstrated the practical utility of our formulation of sampling for image synthesis with three applications: the pointwise representation of color information, the projection of the angular part of the light energy density into spherical harmonics, and the approximation of the shading equation. For all applications, our techniques provide significant visual and numerical improvements over existing approaches for an equal number of samples. Nonetheless, the present work should be understood as a first step and much theoretical and algorithmic development needs to be done to make our theory truly practical.

Central to our constructive formulation of sampling is its functional analytic foundation. This enabled in Sec. 3 the derivation of almost all existing light transport simulation techniques from the common ansatz of Galerkin projection. It was also vital for our final gathering application in Sec. 4.3 by enabling the use of the spectral theory of Hilbert-Schmidt operators for a technique that directly employs point samples. A classical problem of functional analytic techniques is the curse of dimensionality, that is the exponential dependence of the computational complexity on the dimension of the domain [Traub and Werschulz 1999]. For light transport, which is defined over a six or even higher dimensional space, it is one of the main motivation for the use of Monte Carlo methods. Over the past few years, other techniques to break the curse of dimensionality have been developed, such as sparse and multigrid methods [Bungartz and Griebel 2004] and neuronal networks [Donoho 2000]. In the future, we would like to combine these results with reproducing kernel bases to obtain novel, sampling-based image synthesis techniques that do not suffer from the curse of dimensionality. A functional analytic formulation is a prerequisite for quantifying the optimality of computations and to analyse error. For light transport, these objectives were considered by Arvo [1994] but received little attention since. Currently missing for such analyses is a characterization of the function spaces of light transport, analogous to what is available for images [Mallat 2009]. We consider such a characterization to be an important objective for future research.

Acknowledgements

References

ARONSZAJN, N. 1950. Theory of Reproducing Kernels. *Transactions of the American Mathematical Society 68*, 3, 337–404.

ARVO, J., TORRANCE, K., AND SMITS, B. 1994. A Framework for the Analysis of Error in Global Illumination Algorithms. In *SIGGRAPH '94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, ACM, New York, NY, USA, 75–84.

ARVO, J. 1995. *Analytic Methods for Simulated Light Transport*. Ph.d. thesis, Yale University.

ARVO, J. 1995. The Role of Functional Analysis in Global Illumination. In *Rendering Techniques '95*, P. M. Hanrahan and W. Purgathofer, Eds. Springer-Verlag, New York, 115–126.

BASU, K., AND OWEN, A. B. 2014. Low discrepancy constructions in the triangle.

BUNGARTZ, H.-J., AND GRIEBEL, M. 2004. Sparse Grids. *Acta Numerica 13* (May), 147–269.

CHRISTENSEN, P. H., STOLLNITZ, E. J., SALESIN, D., AND DEROSE, T. 1996. Global Illumination of Glossy Environments Using Wavelets and Importance. *ACM Trans. Graph. 15*, 37–71.

CHRISTENSEN, P. H., LISCHINSKI, D., STOLLNITZ, E. J., AND SALESIN, D. H. 1997. Clustering for Glossy Global Illumination. *ACM Transactions on Graphics (TOG) 16*, 1.

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

COOK, R. L., PORTER, T., AND CARPENTER, L. 1984. Distributed Ray Tracing. *ACM SIGGRAPH Computer Graphics 18*, 3.

COOK, R. L. 1986. Stochastic Sampling in Computer Graphics. *ACM Trans. Graph. 5*, 51–72.

DAUBECHIES, I. 1992. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

DE WITT, T., LESSIG, C., AND FIUME, E. 2012. Fluid Simulation Using Laplacian Eigenfunctions. *ACM Transactions on Graphics 31*, 1 (Jan.), 1–11.

DICK, J., AND PILLICHSHAMMER, F. 2010. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press.

DONOHO, D., AND STODDEN, V. 2004. Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts? In *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge, MA, S. Thrun, L. Saul, and B. Schölkopf, Eds.

DONOHO, D. L. 2000. High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality. In *Mathematical Challenges of the 21st Century*, American Mathematical Society, Los Angeles, CA, USA.

GALERKIN, B. G. 1915. On Electrical Circuits for the Approximate Solution of the Laplace Equation. *Vestnik Inzh. 19*, 897–908.

GORAL, C. M., TORRANCE, K. E., GREENBERG, C. P., AND BATTAILE, B. 1984. Modeling the Interaction of Light between Diffuse Surfaces. *SIGGRAPH Comput. Graph. 18*, 213–222.

GORTLER, S. J., SCHRÖDER, P., COHEN, M. F., AND HANRAHAN, P. 1993. Wavelet Radiosity. In *Computer Graphics Proceedings, Annual Conference Series, 1993 (ACM SIGGRAPH '93 Proceedings)*, 221–230.

GREGER, G., SHIRLEY, P., HUBBARD, P. M., AND GREENBERG, D. P. 1998. The Irradiance Volume. *IEEE Computer Graphics and Applications 18*, 2, 32–43.

HECKBERT, P., AND WINGET, J. M. 1991. Finite Element Methods for Global Illumination. Technical Report UCB/CSD 91/643, University of California at Berkeley, Berkeley, CA.

JAROSZ, W., CARR, N. A., AND JENSEN, H. W. 2009. Importance Sampling Spherical Harmonics. *Computer Graphics Forum 28*, 2, 577–586.

JENSEN, H. W., AND CHRISTENSEN, N. J. 1995. Photon Maps in Bidirectional Monte Carlo Ray Tracing of Complex Objects. *Computers & Graphics 19*, 2, 215–224.

JENSEN, H. W. 1995. Importance Driven Path Tracing using the Photon Map. In *Rendering Techniques '95*, Springer-Verlag, P. Hanrahan and W. Purgathofer, Eds., 326–335.

JENSEN, H. W. 2001. *Realistic Image Synthesis using Photon Mapping*. A. K. Peters, Ltd., Natick, MA, USA.

KAJIYA, J. T., AND VON HERZEN, B. P. 1984. Ray Tracing Volume Densities. *International Conference on Computer Graphics and Interactive Techniques 18*, 3.

KAJIYA, J. T. 1986. The Rendering Equation. *ACM SIGGRAPH Computer Graphics 20*, 4.

KELLER, A. 2006. *Myths of Computer Graphics*. Springer, 217–243.

LEHTINEN, J. 2007. A Framework for Precomputed and Captured Light Transport. *ACM Trans. Graph. 26*, 13.

LESSIG, C., AND FIUME, E. 2010. On the Effective Dimension of Light Transport. *Computer Graphics Forum 29*, 4 (Aug.), 1399–1403.

LESSIG, C., DE WITT, T., AND FIUME, E. 2012. Efficient and Accurate Rotation of Finite Spherical Harmonics Expansions. *Journal of Computational Physics 231*, 2 (Jan.), 243–250.

LESSIG, C. 2012. *Modern Foundations of Light Transport Simulation*. Ph.d. thesis, University of Toronto, Toronto.

MAHAJAN, D., TSENG, Y.-T., AND RAMAMOORTHI, R. 2008. An Analysis of the In-Out BRDF Factorization for View-Dependent Relighting. *Comput. Graph. Forum (Proceedings of EGSR) 27*, 1137–1145.

MALLAT, S. G. 2009. *A Wavelet Tour of Signal Processing: The Sparse Way*, third ed. Academic Press.

MITCHELL, D. P. 1987. Generating Antialiased Images at Low Sampling Densities. In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, 65–72.

MITCHELL, D. P. 1991. Spectrally Optimal Sampling for Distribution Ray Tracing. *SIGGRAPH Comput. Graph. 25*, 157–164.

NASHED, M. Z., AND WALTER, G. 1991. General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces. *Mathematics of Control, Signals, and Systems 4*, 4, 363–390.

NEHAB, D., AND HOPPE, H. 2014. A Fresh Look at Generalized Sampling. *Foundations and Trends in Computer Graphics and Vision 8*, 1, 1–84.

NIEDERREITER, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, PA, USA.

NISHITA, T., AND NAKAMAE, E. 1984. Calculation of Interreflections and its Representation Method. In *1984 Annual Conference of the Japanese Illumination Engineering Society*, vol. 63.

NOVAK, E., AND WOŹNIAKOWSKI, H. 2010. *Tractability of Multivariate Problems: Standard Information for Functionals*, vol. II of *EMS Tracts in Mathematics*. European Mathematical Society Publishing House.

PEERCY, M. S. 1993. Linear Color Representations for Full Speed Spectral Rendering. In *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, 191–198.

PETROV, G. I. 1940. Application of the Method of Galerkin to a Problem Involving the Stationary Flow of a Viscous Fluid. *Prikl. Matem. Mekh. 4*, 3.

PHARR, M., AND HUMPHREYS, G. 2010. *Physically Based Rendering: From Theory to Implementation*, second ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

PINCHON, D., AND HOGGAN, P. E. 2007. Rotation Matrices for Real Spherical Harmonics: General Rotations of Atomic Orbitals in Space-Fixed Axes. *Journal of Physics A: Mathematical and Theoretical 40*, 1597–1610.

SAFF, E., AND KUIJLAARS, A. 1997. Distributing many points on a sphere. *The Mathematical Intelligencer 19*, 1, 5–11.

SCHRÖDER, P., GORTLER, S. J., COHEN, M. F., AND HANRAHAN, P. 1993. Wavelet Projections for Radiosity. In *Fourth Eurographics Workshop on Rendering*, 105–114.

SHIRLEY, P., WADE, B., HUBBARD, P. M., ZARESKI, D., WALTER, B., AND GREENBERG, D. P. 1995. Global Illumination via Density-Estimation. In *Proceedings of 6th Workshop on Rendering*, Springer, 219–230.

SHIRLEY, P. 1991. Discrepancy as a Quality Measure for Sample Distributions. In *Proceedings of Eurographics 1991*.

SILLION, F., ARVO, J., WESTIN, S., AND GREENBERG, D. P. 1991. A Global Illumination Solution for General Reflectance Distributions. In *Proceedings of ACM SIGGRAPH 1991*, ACM Press, New York, NY, USA, 187–196.

SIMONS, F. J. 2010. Slepian Functions and Their Use in Signal Estimation and Spectral Analysis. In *Handbook of Geomathematics*, W. Freeden, Z. M. Nashed, and T. Sonar, Eds.

SLOAN, P.-P., KAUTZ, J., AND SNYDER, J. 2002. Precomputed Radiance Transfer for Real-Time Rendering in Dynamic, Low-Frequency Lighting Environments. In *Proceedings of ACM SIGGRAPH 2002*, ACM Press, New York, NY, USA, 527–536.

TRAUB, J. F., AND WERSCHULZ, A. G. 1999. *Complexity and Information*. Cambridge University Press, New York, NY, USA.

TRENOGIN, V. A. 2002. Galerkin Method. In *Encyclopedia of Mathematics*, M. Hazewinkel, Ed. Kluwer Academic Publishers.

VEACH, E., AND GUIBAS, L. J. 1997. Metropolis Light Transport. In *SIGGRAPH '97: Proceedings of the 24th annual conference on computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 65–76.

VEACH, E. 1997. *Robust Monte Carlo Methods for Light Transport Simulation*. Ph.d. thesis, Stanford University.

WALTER, B., HUBBARD, P. M., SHIRLEY, P., AND GREENBERG, D. P. 1997. Global Illumination using Local Linear Density Estimation. *ACM Trans. Graph. 16*, 217–259.

WESTIN, S. H., ARVO, J., AND TORRANCE, K. E. 1992. Predicting Reflectance Functions from Complex Surfaces. In *Proceedings of ACM SIGGRAPH 1992*, ACM Press, New York, NY, USA, 255–264.

ZATZ, H. R. 1993. Galerkin Radiosity: A Higher Order Solution Method for Global Illumination. In *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA.

ZHU, C., BYRD, R. H., LU, P., AND NOCEDAL, J. 1997. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Soft. (TOMS) 23*, 4.