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

Supplementary Material

Christian Lessig* Mathieu Desbrun Eugene Fiume

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

In the following, we provide further details on the material presented in the paper: "A Constructive Theory of Sampling for Image Synthesis using Reproducing Kernel Bases". A tutorial on the mathematical formulation of our theory, including necessary functional analytic background, is provided in a separate document.

1 Monte Carlo Integration and Reproducing Kernels

In this section, we will show how reproducing kernels enable a functional analytic interpretation of Monte Carlo integration. We will begin by introducing the concept of a characteristic basis and its connection to reproducing kernels.

1.1 Characteristic Bases

Let X be a set, for example a compact subset of $\mathbb{R}^d$. The *characteristic function* $\chi_Y : X \to \mathbb{R}$ for a subset $Y \subset X$ is

$$\chi_Y(x) = \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{otherwise} \end{cases}. \tag{1}$$

A *partition of* X is a collection of disjoint subsets X_i, $i = 1 \cdots n$, of X such that their union forms again X, that is

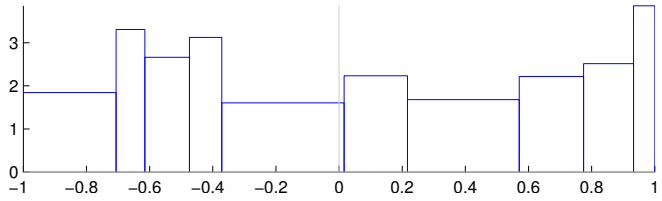
$$X = \bigcup_{i=1}^{n} X_i. \tag{2}$$

With each X_i we can naturally associate the characteristic function $\chi_i(x)$ that is nonzero only over this domain X_i. A *characteristic basis* is then the collection $\{\chi_i(x)\}_{i=1}^n$ of all $\chi_i(x)$ associated with a partition $\{X_i(x)\}_{i=1}^n$. The basis spans the space $\text{span}(\chi_i(x)) = \mathcal{H}_\chi(X) \subset L_2(X)$ and since it is a closed subspace of $L_2(X)$ it is a Hilbert space equipped with the L_2 inner product. Note that $\mathcal{H}_\chi(X)$, $\{\chi_i(x)\}_{i=1}^n$ and $\{X_i(x)\}_{i=1}^n$ are all in a one-to-one correspondence.

Instead of unit height characteristic basis functions $\chi_i(x)$, we will in the following usually work with their normalized siblings

$$\bar{\chi}_i(x) = \frac{1}{\sqrt{|X_i|}} \chi_i(x) \tag{3}$$

where $|X_i|$ denotes the area of X_i. The basis $\{\bar{\chi}_i(x)\}_{i=1}^n$ is an orthonormal characteristic basis for $\mathcal{H}_\chi(X)$. For example, for $X = [-1, 1]$ such a basis is given by:



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

By Eq. 3 in the paper, the reproducing kernel for $\mathcal{H}_\chi(X)$ at $\lambda \in X_j$ is given by

$$k_\lambda(x) = \sum_{i=1}^{n} \bar{\chi}_i(\lambda)\, \bar{\chi}_i(x). \tag{4a}$$

Since the product $\bar{\chi}_i(\lambda)\, \bar{\chi}_i(x)$ vanishes unless $\lambda \in X_j$ the sum collapses and we have

$$k_\lambda(x) = \bar{\chi}_j(\lambda)\, \bar{\chi}_j(x). \tag{4b}$$

By the definition of the normalized characteristic functions we moreover obtain

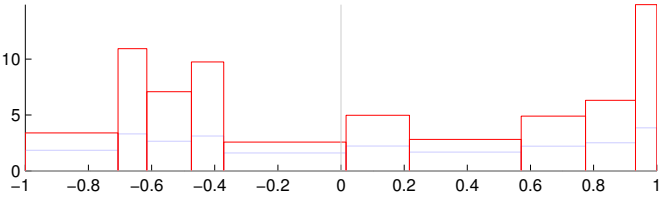
$$k_\lambda(x) = \frac{1}{|X_j|} \chi_j(\lambda) \tag{4c}$$

Hence, the reproducing kernel $k_\lambda(x)$ for $\mathcal{H}_\chi(X)$ coincides with the characteristic function $\chi_i(x)$ up to a constant.

From Eq. 4 it follows that a reproducing kernel basis for $\mathcal{H}_\chi$ can be formed by choosing one location λ_i in each X_i. The reproducing kernel basis functions in Eq. 4c are then orthogonal, since their support is disjoint, but they are not orthonormal. Moreover, we cannot normalize the $k_i(x)$ since they would then lose the reproducing property; this is an instance where the general wisdom that every orthogonal basis can be carried over to an orthonormal basis is not true, or at least it would destroy the reproducing property that is crucial for us. From the biorthogonality condition $\langle k_i(x), \tilde{k}_j(x) \rangle = \delta_{ij}$ it follows that the dual kernel functions for the reproducing kernel basis $\{k_i = 1/|X_i| \chi_x\}_{i=1}^n$ are given by $\tilde{k}_i(x) = \chi_i(x)$, that is by unnormalized characteristic functions. The basis pair for a characteristic reproducing kernel basis is hence

$$\left(\left\{ k_i(x) = \frac{1}{|X_i|} \chi_i(x) \right\}, \left\{ \tilde{k}_i(x) = \chi_i(x) \right\} \right). \tag{5}$$

For example, for the orthonormal characteristic basis that we considered before, the associated reproducing kernel basis is given by (reproducing kernels in red):



An important property of characteristic reproducing kernel bases is that these can be constructed and are practical for arbitrary domains X, including manifolds and high-dimensional spaces. Equivalently, the kernel matrix for a characteristic reproducing kernel basis is a diagonal matrix that is trivially orthogonal and maximally localized.

1.2 Monte Carlo Integration Revisited

With the foregoing definition of characteristic reproducing kernel bases we are prepared to show how Monte Carlo integration can be obtained as a quadrature rule. Let $\{\chi_i\}_{i=1}^n$ be a *uniform* characteristic basis over a partition for the set $X = [a, b] \subset \mathbb{R}$ whose elements X_i have equal volume $|X_i| = |X|/n$. With locations $\Lambda = \{\lambda_i\}$ such that each X_i contains exactly one sampling point

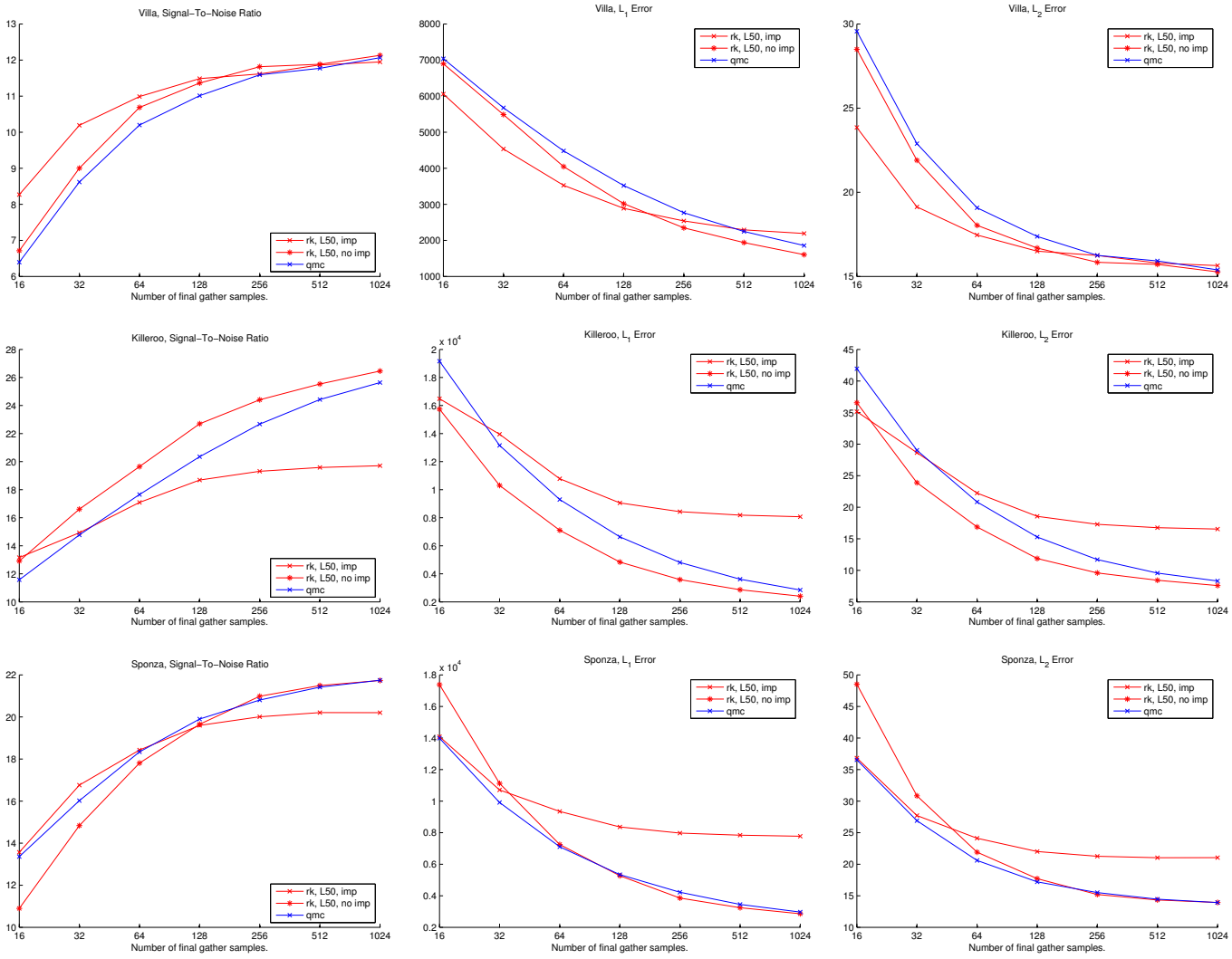


Figure 1: *Experimental results for final gathering: Signal-to-Noise ration, L_2 error, and L_1 error (from left to right) for the Villa, Killeroo and Sponza scenes. The graphs for the Villa scene are those also presented in the paper.*

λ_i we obtain a characteristic reproducing kernel basis for the space $\mathcal{H}_\chi([0,1])$ with dual kernel functions $\tilde{k}_i(x) = \chi_i(x)$. A quadrature rule associated with this reproducing kernel basis can be obtained using Eq. 10 in the paper. Since the characteristic basis is uniform we obtain for the quadrature weights

$$w_i = \int_X \tilde{k}_i(x)\,dx = \int_X \chi_i(x)\,dx = |\chi_i| = \frac{|X|}{n}. \tag{6}$$

The quadrature rule for the space $\mathcal{H}_\chi([0,1])$ is thus

$$\int f(x)\,dx = \sum_{i=1}^{n} w_i\, f(\lambda_i) = \frac{|X|}{n} \sum_{i=1}^{n} f(\lambda_i) \tag{7a}$$

$$= \frac{b-a}{n} \sum_{i=1}^{n} f(\lambda_i) \tag{7b}$$

Eq. 7b formally coincides with the standard Monte Carlo estimator for uniformly distributed sampling locations. From the definition of a probability (or measure) space, the result that Monte Carlo integration arises as a quadrature rule for the space spanned by characteristic functions is by no means surprising, cf. [Rudin 1987]. Also note that $\{\chi_i\}_{i=1}^n$ becomes dense in $L_2([a,b])$ as the number of partitions goes to infinity and hence asymptotically the quadrature rule converges for any $f \in L_2(X) \cap L_1(X)$.

For samples drawn from an arbitrary probability distribution function $p(x)$ the requirement of one sample per unit height characteristic basis functions implies that the partition elements X_i can no longer have equal size but need to have the form $X_i = [x_i, x_{i+1}]$ for suitable interval bounds $x_i \in [a,b]$. Choosing the x_i such that in the support of every $\chi_i(x)$ is on average one sample is then equivalent to

$$n\,P([x_i, x_{i+1}]) = n \int_{x_i}^{x_{i+1}} p(x)\,dx = 1 \tag{8}$$

which with $y \in [x_i, x_{i+1}]$ is to zero$^{\text{th}}$ order

$$p(y)\,(x_{i+1} - x_i) = 1/n. \tag{9}$$

With λ_i being the samples in the support of χ_i one thus has $|\chi_i| = x_{i+1} - x_i = 1/(n\,p(\lambda_i))$. The quadrature rule with samples distributed according to $p(x)$ is therefore

$$\int f(\bar{x})\,d\bar{x} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(\lambda_i)}{p(\lambda_i)} \tag{10}$$

Eq. 10 formally coincides with the standard Monte Carlo estimator for importance sampling.

Figure 2: *Experimental results for final gathering using our technique (top row for each scene) and classical final gathering (bottom row for each scene). The columns correspond to 16, 32, 64, 128, 256, 512, and 1024 samples in the hemisphere, respectively. See Fig. 1 for the corresponding error rates.*

2 Final Gathering

In this section, we present additional implementation details and experimental results on our final gathering application in Sec. 4.3 in the main paper.

2.1 Implementation Details

Galerkin Projection of Shading Operator To implement our technique for final gathering, we have to compute the Galerkin projection

$$\hat{P}_{ij} = \left\langle y_i(\omega) \mid \rho(\omega, \bar{\omega}) \mid y_j(\bar{\omega}) \right\rangle$$

of the shading operation P into a finite spherical harmonics space, cf. Eq. 4.14 in the paper. The two inner products that have to be computed to obtain $\hat{P}$ cannot be determined analytically. We hence first compute a pointwise representation $\check{P} = \{\rho(\lambda_i, \lambda_j)\} \in \mathbb{R}^{N \times N}$ which is determined by evaluating the scattering kernel for a large number of directions $\lambda_i \in S^2$. Since the matrix $\check{P}$ provides a representation of P over samples, we can employ our theory to obtain an approximation of $\hat{P}$ through

$$\hat{P} \approx S_y^L(\Lambda)\, \check{P}\, S_y^L(\Lambda),$$

cf. Eq. 2.8 in the paper. $S_y^L(\Lambda)$ is the spherical harmonics sampling matrix for the locations Λ and maximum spherical harmonics band L. The cardinality of Λ is chosen such that one has a sufficient oversampling rate. With $\hat{P}$ we compute the singular value decomposition $\hat{P} = \hat{U}\,\hat{\Sigma}\,\hat{V}^T$ using the Eigen C++ library.[1] The Galerkin projection has to be computed once and can then be reused whenever the same scattering function is used again. We implement this by writing the Galerkin projection and its singular value decomposition to disk and reloading it when necessary.

Importance Sampling Importance sampling requires the computation of weights that never directly contribute to image generation but are only employed to estimate the relative importance of sampling locations. We can hence employ a less accurate approximation for this weight computation. For spherical harmonics this is easily realized by using a smaller bandwidth L for this computation. The effect of this on image quality for the Villa scene is shown in the following table (L refers here to the maximal bandwidth used for importance sampling):

L	1	5	10	20	25	50
SNR	14.95	15.06	15.08	15.06	15.07	15.07
L_2 err	7.01	6.90	6.88	6.90	6.89	6.89
Time	189.2	190.3	202.5	245.6	270.5	528.0

2.2 Experimental Evaluation

The dependence of signal-to-noise ratio, L_1 error, and L_2 error on the number of gather samples is shown in Fig. 1. In Fig. 3 we show the dependence of the error on the maximum spherical harmonics band used to represent the shading operator and on the number of eigenvalues used in the approximation.

[1] `http://eigen.tuxfamily.org/`

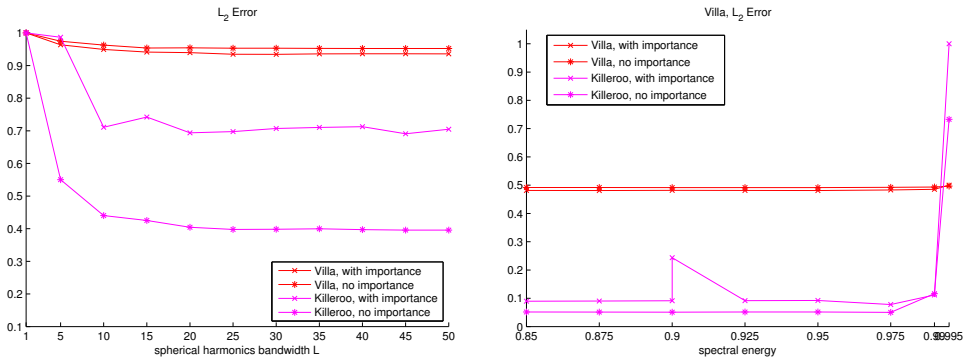


Figure 3: *Experimental results for final gathering. L_2 error for the Villa and Killeroo scenes as a function of one varying parameter; left, spherical harmonics bandwidth; right, percentage of spectral energy. Signal-to-Noise ratio and L_1 are qualitatively equivalent and hence not shown. For the spherical harmonics bandwidth we see that $L = 25$ or $L = 30$ would suffice. The spectral energy decrease up to 0.95% and afterwards it increase again since then the oversampling rate becomes too small for the number of samples that was used (here 128).*

3 Reproducing Kernels in the Literature

To our knowledge, our formulation of sampling-based numerical techniques based on reproducing kernel bases, with an emphasis on finite reproducing kernel Hilbert spaces and practical techniques, did not appear before in the literature. Related to our presentation is work on general sampling theorems, see for example [Nashed and Walter 1991; Unser 2000; Zayed 1993; Benedetto and Ferreira 2001]. In these works reproducing kernels are also employed for the correspondence between continuous functions and pointwise samples. However, typically more theoretical questions are considered, for example whether an infinite dimensional function space admits a set of sampling. Related to our formulation of sampling are so called kernel methods, see for example the survey articles [Schaback and Wendland 2006] and [Fasshauer 2011]. However, since the work is not based on reproducing kernel bases it is harder to employ in practice. Many applications of reproducing kernels can also be found in machine learning. Best known are probably kernel methods [Schölkopf and Smola 2002; Berlinet and Thomas-Agnan 2004] which exploit that reproducing kernels satisfy $\langle k_x(z), k_y(z)\rangle = k(x, y)$. This enables to evaluate high-dimensional inner products $\langle k_x(z), k_y(z)$ by evaluating reproducing kernel functions $k(x, y)$. More closely related to our use of reproducing kernels than kernel methods is work by Smale and collaborators on the foundations of learning, see [Cucker and Smale 2002; Poggio and Smale 2003; Smale and Zhou 2004; Smale and Zhou 2005; Smale and Zhou 2007] and reference therein. There, reproducing kernels are also employed for the conceptualization of point samples. Yet another area where reproducing kernel have become popular in recent years are meshless finite element methods, see [Babuška et al. 2003; Li and Liu 2004; Griebel and Schweitzer 2007] and reference therein. These methods are closely related to the reproducing kernel Galerkin projection in Sec. 3.2 in the paper. We hope that this connection proves useful both for the development of numerical techniques for image synthesis and for relaxing some of the assumptions currently required in our interpretation of light transport techniques in Sec. 3 in the paper. The numerical optimization of sampling points for quadrature and spherical harmonics transforms has been explored before by Sloan and Womersley [2004; 2009]. However, in our experiments the energy function used in their work for optimization was less efficient than rk-discrepancy. Using reproducing kernels, Gräf and Potts [2011] recently showed that discrepancy and attraction-repulsion schemes for point optimization are related, and they also explored an interesting connection to stippling [Gräf et al. 2011]. In recent years, reproducing kernels began to play a prominent role in the Quasi Monte Carlo literature [Dick and Pillichshammer 2010]. For example, classical L_2 discrepancy is the worst case error for equal weight (QMC) quadrature for H^s, as a reproducing kernel Hilbert space, see again [Dick and Pillichshammer 2010, Chapter 2] or [Novak and Woźniakowski 2010, Chapter 9]. With this interpretation, reproducing kernels enable one to generalize discrepancy to integration for functions in other spaces.

4 Large Sample Counts and Adaptive Sampling

On the image plane, millions of samples are typically used to reconstruct an image. Constructing and possibly inverting the kernel matrix for such a large number of samples is impossible. At the same time, the image intensity is a signal with locally varying regularity. The globally supported reproducing kernel functions we considered in the three applications in the paper are hence ill-suited for approximating and representing the signal [Mallat 2009]. Fortunately, the questions of how to effectively construct and invert an arbitrarily large kernel matrix and how to determine locally supported reconstruction filters that adapt to the local signal regularity are two sides of the same medal and can be addressed simultaneously. To emphasize the idea, we consider in the following functions over $[a, b] \subset \mathbb{R}$. We will employ reproducing kernels for a space spanned by locally supported basis functions and then locally invert the kernel matrix by considering only a subset of its rows and columns to determine reconstruction filters.

Adaptivity and Local Inversion A class of bases well suited for the representation of signals with locally varying regularity, for example smooth sections interrupted by singularities, are wavelets [Daubechies 1992; Mallat 2009]. Since we are interested in signals defined over finite intervals $[a, b] \subset \mathbb{R}$, we will employ the Cohen-Daubechies-Vial (CDV) wavelets [Cohen et al. 1993] which are the classical Daubechies wavelets [Daubechies 1992] modified to correctly handle finite intervals $[a, b]$. In its standard form, a CDV wavelet basis $\{\varphi_{0,0}, \psi_{i,j}\}$, $j = 0 \cdots$, $i = 1 \cdots 2^j$, is globally supported over $[a, b]$ since the scaling function $\varphi_{0,0}(x)$ on level zero is nonzero over all of $[a, b]$. However, by the recursive structure of the wavelet spaces V_j that are spanned by the scaling and wavelet functions, an equivalent basis is given by the scaling functions $\varphi_{n,i}(x)$ on a finer level $n > 0$ and the wavelets functions $\psi_{j,i}$ on the following levels $j > n$. We hence have the following basis consisting of localized basis functions

$$\Psi = \left\{ \{\varphi_{n,i}\}, \{\psi_{j,i}\} \mid j = n, \cdots, N, i = 1, \cdots, 2^j \right\} \qquad (11)$$

where we also truncated the sequence of wavelets on some finite level N to obtain a representation amenable to numerical computations. Given the basis Ψ, we can employ Eq. 2.3 in the paper to construct reproducing kernels $k_\lambda(x)$ for the space spanned by Ψ. Since the

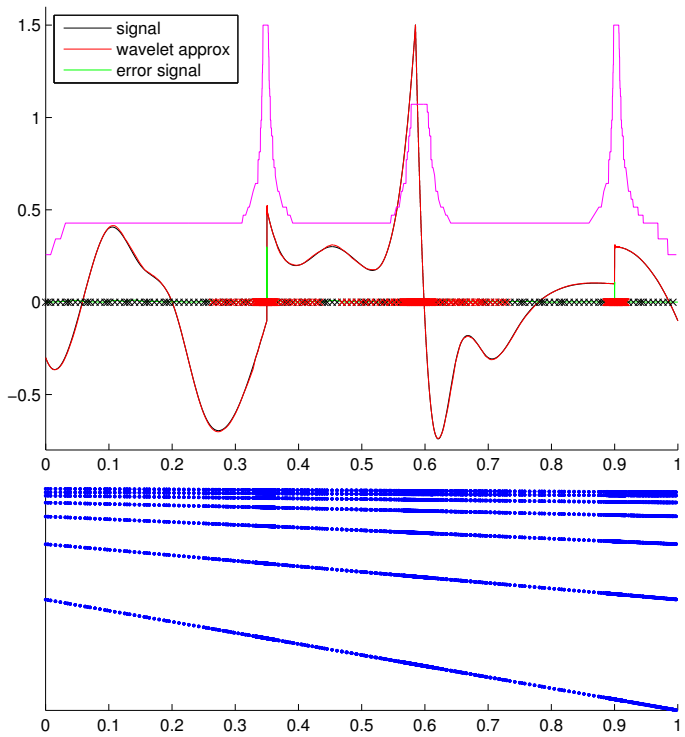


Figure 4: *Top: Adaptive wavelet reconstruction. Shown in magenta is the relative local dimension of the wavelet approximation, or, equivalently, the local depth of the wavelet tree. Samples are marked by crosses on the x-axis, black are samples from a base set and red those used for refinement around singularities. Bottom: Sparsity pattern of transposed kernel matrix K^T for wavelet reconstruction in the top plot (abscissa values and columns, corresponding to samples, are aligned).*

wavelet basis functions are localized, the reproducing kernels will also only have local support in the vicinity of the sampling location $\lambda \in \Lambda \subset [a,b]$. Choosing suitable locations $\Lambda = \{\lambda_i\}_{i=1}^m$, we can construct a reproducing kernel basis $\{k_i(x)\}_{i=1}^m$ that is spanned by locally supported kernel functions $k_i(x)$. The degree of localization is thereby determined by the support of the basis functions at the coarsest level n that is employed. As in the general case, numerically the reproducing kernel basis is represented using the kernel matrix:

$$K = \begin{pmatrix} \varphi_{n,1}(\lambda_1) & \cdots & \varphi_{n,2^n}(\lambda_1) & \varphi_{n,1}(\lambda_1) & \cdots & \psi_{N,2^N}(\lambda_1) \\ \vdots & & & & & \vdots \\ \varphi_{n,1}(\lambda_m) & \cdots & \varphi_{n,2^n}(\lambda_m) & \varphi_{n,1}(\lambda_m) & \cdots & \psi_{N,2^N}(\lambda_m) \end{pmatrix}$$

Since the reproducing kernels are locally supported the above kernel matrix is sparse. The sparsity pattern for CDV wavelets with three vanishing moments and $n = 3$ and $N = 8$ is shown in Fig. 4, bottom. Clearly visible is how each reproducing kernel basis function, corresponding to one column, depends only on a small subset of the localized wavelet basis functions. Apparent are also the different levels of the wavelet basis functions, with localization increasing as the level increases from top to bottom.

Through the sparsity of the kernel matrix K it is possible to invert it by considering only a small subset of the matrix to determine a row of the inverse. In particular, the dual kernel function $\tilde{k}_j(x)$, whose basis function coefficients are the elements of the inverse of K, are defined by the biorthogonality condition $\langle k_i(x), \tilde{k}_j(x)\rangle = \delta_{ij}$. Hence, the duals can only depend on those basis functions $\psi_{a,b}(x)$ of the wavelet basis that are nonzero over the support of $k_j(x)$. Conversely, $\tilde{k}_j(x)$ can only be nonzero over the support of these basis functions $\psi_{a,b}(x)$. Therefore, each dual kernel function $\tilde{k}_j(x)$ can be determined from only a small subset of the rows and columns of K and, moreover, the rows correspond to a localized neighborhood whose size depends on the localization of the basis functions. In Matlab pseudo-code, the local inversion is:

```
for( piv = 1 : size(K,1))
  % wavelets contributing to kernel at lambda(piv)
  idx = find( abs(K(piv,:)) > 0);

  % find all abscissa values that are relevant
  jdx = [];
  for( j = 1 : numel(idx))
    jdx = union( jdx, find(abs(K(:,idx(j))) > 0));
  end

  % subset of kernel matrix
  Kl = K(jdx,idx);
  % Invert *local* matrix
  Sl = pinv(Kl);

  % Copy relevant part of Sl to inverse
  ...

end
```

For example, when we again consider the CDV wavelets with three vanishing moments and $n = 3$, $N = 8$, and employ 1024 uniform sampling points in $[a,b]$ then the dual kernel functions $\tilde{k}_j(x)$ in the interior of the interval can be determined from a sub-matrix of size 182×35 instead of the full kernel matrix of size 1024×4096. The element-wise L_2 error that results from using the local approximation of the pseudo-inverse is 1.94×10^{-10}. A local inversion is also beneficial when adaptive sampling is used and the local sampling density is for example determined by the local regularity of the signal. We can then decide on the number of wavelet functions that is used in a neighborhood, or equivalently on the local depth of the wavelet tree, based on the number of samples that is adaptively determined in a neighborhood. A simple example of adaptive sampling and localized reconstruction is shown in Fig. 4 where we use $N = 8$ only around singularities and otherwise $N = 6$. There, we also used adaptive sampling to ensure that we are locally in an oversampling regime without wasting samples where the signal is very regular.

For ease of implementation, we employed in our proof-of-concept application wavelets. However, when the ideas are used for sampling on the image plane then curvelets [Candès and Donoho 1999] or bandlets [Mallat and Peyré 2007] should be employed, which are better adapted for the representation of natural images. An interesting avenue for future work is also to exploit that the scaling basis functions are closely related to reproducing kernels [Daubechies 1992].

5 Density Estimation and Reproducing Kernels

As was already known to Aronszajn [1950] when he developed the theory of reproducing kernels, Green's functions for bounded partial differential equations provide an important example of reproducing kernels. In the following, we will show that Green's functions also enable the interpretation of density estimation, typically given by

$$f(x) \approx \frac{1}{n\,h} \sum_{i=1} k\left(\frac{|x - x_i|}{h}\right) \tag{12}$$

for some translation invariant kernel $k(x,y) = k(\|x - y\|)$, from the point of view of reproducing kernel bases.

Let us begin with an example. The heat equation

$$\frac{\partial \phi_t(x)}{\partial t} - \Delta \phi_t(x) = 0 \tag{13}$$

describes the diffusion of a time-dependent function $\phi_t : \mathbb{R}^n \to \mathbb{R}$. As is by now well known in computer graphics, the Green's function for the heat equation is the heat kernel $H_t(x, y)$ given by

$$H_t(x, y) = \frac{1}{\sqrt{4\,\pi\,t}^n} \exp\left(\frac{-|x-y|^2}{4t}\right). \tag{14}$$

By definition, we hence have for the image of the function $\phi_0(x)$ at time $t = 0$ under the flow of the heat equation

$$\phi_t(y) = \frac{1}{\sqrt{4\,\pi\,t}^n} \int_{\mathbb{R}^n} \phi_0(x) \exp\left(\frac{-|x-y|^2}{4t}\right) dx. \tag{15}$$

Following [Koenker et al. 2012], let us assume that the function $\phi_0(x)$ corresponds to a normalized "train of Dirac-δ's" at a set of locations x_i,

$$\phi_0(x) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}, \tag{16}$$

we will try to interpret this from the point of view of our theory shortly. By Eq. 15, we then formally have

$$\begin{aligned}
\phi_t(y) &= \frac{1}{\sqrt{4\,\pi\,t}^n} \int_{\mathbb{R}^n} \phi_0(x) \exp\left(\frac{-|x-y|^2}{4t}\right) dx && \text{(17a)}\\
&= \frac{1}{\sqrt{4\,\pi\,t}^n\,N} \int_{\mathbb{R}^n} \sum_i \delta_{x_i} \exp\left(\frac{-|x-y|^2}{4t}\right) dx && \text{(17b)}\\
&= \frac{1}{\sqrt{4\,\pi\,t}^n\,N} \sum_i \int_{\mathbb{R}^n} \delta_{x_i} \exp\left(\frac{-|x-y|^2}{4t}\right) dx && \text{(17c)}\\
&= \frac{1}{\sqrt{4\,\pi\,t}^n\,N} \sum_i \exp\left(\frac{-|x_i-y|^2}{4t}\right). && \text{(17d)}
\end{aligned}$$

The last equation is the density estimate for a Gaussian kernel, cf. Eq. 12. Hence, density estimation can in this case be interpreted as smoothing by the heat flow. In the following, we will take an approach based on reproducing kernels to generalize the derivation to other kernels.

Let us assume a differential equation has orthonormal eigenfunctions $\phi_i(x)$ associated with eigenvalues λ_i. Its Green's function then has the representation

$$G_t(x, y) = \sum_{i=1}^{n} \lambda^{-t}\,\phi_i(x)\,\phi_i(y) \tag{18}$$

and the reproducing kernel for the domain of the Green's function can be written with the eigenfunctions using Eq. 2.3. For a tight reproducing kernel basis, a function $f(x)$ in the domain has then the representation

$$f(x) = \frac{n}{m} \sum_{i=1}^{m} f(\lambda_i)\,k_i(x). \tag{19}$$

Using $f(x) \equiv f_0(x)$ as the function at time $t = 0$ we obtain for a unit time flow $f(y) \equiv f_1(y)$ that

$$\begin{aligned}
f(y) &= \int_X f_0(x)\,G(x, y)\,dx && \text{(20a)}\\
&= \int_X \left(\sum_{i=1}^{m} f(\lambda_i)\,k_i(x)\right) G(x, y)\,dx && \text{(20b)}\\
&= \sum_{i=1}^{m} f(\lambda_i) \int_X k_i(x)\,G(x, y)\,dx. && \text{(20c)}
\end{aligned}$$

Expanding both the reproducing kernel and the Green's function in the eigenbasis and using the orthogonality of the $\phi_a(x)$ leads to

$$\int_X k_i(x)\,G(x, y)\,dx = \sum_a \lambda_a\,\phi_a(\lambda_i)\,\phi_a(y) = G(\lambda_i, y) \tag{20d}$$

Hence, we have

$$f(y) = \frac{n}{m} \sum_{i=1}^{m} f(\lambda_i)\,G(\lambda_i, y). \tag{20e}$$

When we choose $f(x)$ to be the indicator function so that $f(\lambda_i) = 1$, the last equation has again the form of Eq. 12. The additional terms compared to Eq. 12 result, as before, when we make a concrete choice for a Green's function, for example the heat kernel, and when $m \gg n$.

In the literature [Pharr and Humphreys 2010], a special type of density estimation are kernel weighted averages of the form

$$f(x) \approx \frac{\sum_{i=1}^{m} f(x_i)\,k(x - x_i)}{\sum_{i=1}^{m} k(x - x_i)} \tag{21}$$

with an additional normalization term in the denominator. In the statistics literature, Eq. 21 is known as Nadaraya-Watson estimator. In the above discussion of density estimation we have ignored the question of what happens to the function under the flow of the partial differential equation. For a well behaved flow, a smoothing will take place but unless the partial differential equation is unitary the flow will alter the norm of $f_t(x)$. When density estimation is uses as generalized interpolation then this is undesirable. It is easy to see that Eq. 21 enforces norm preservation for $f(x)$ being the characteristic function. In wavelet parlance, the denominator enforces a zero-th order vanishing moment.

Ideas similar to those discussed in the present section can be found in the work by Fasshauer [2010; 2011].

6 An Error Bound for Pointwise Techniques

In the following, we will provide a more formal and rigorous discussion of the material that was presented in Sec. 2.4 in the main paper. Let us being by restating the error formula in Eq. 2.11.

Proposition 1. *Let $\mathcal{G}(X)$ be a reproducing kernel Hilbert space with orthonormal basis $\{\phi_i\}_{i=1}^{p}$, with p possibly being infinity, and $\mathcal{H}(X) \subset \mathcal{G}(X)$ a finite, closed subspace such that $\{\phi_i\}_{i=1}^{n}$ forms an orthonormal basis for $\mathcal{H}(X)$. Furthermore, let $\{k_i(x)\}_{i=1}^{m}$ be a reproducing kernel basis for $\mathcal{G}(X)$ defined over locations $\Lambda = \{\lambda_i\}$. For a function $\hat{f} + \check{f} = f \in \mathcal{G}(X)$ with $\hat{f} \in \mathcal{H}(X)$ and $\check{f} \in \mathcal{G}(X) \setminus \mathcal{H}(X)$, the error $err_k(f)$ in the k^{th} basis function coefficient $f_k(\phi) = \langle f, \phi_k \rangle$ when obtained from only pointwise samples $f(\lambda_i)$ by $f(\Lambda) = K\,f(\phi)$ is bounded by*

$$|err_k(f)| \leq \|\check{f}\| \|\gamma_k\|$$

where the vector $\gamma_k = (\gamma_{n+1}^k, \ldots, \gamma_p^k)$ has elements

$$\gamma_i^k = \sum_{j=1}^{m} \phi_i(\lambda_j)\,r_{kj}.$$

Proof. Let $f(\lambda_i)$ be given values of $f \in \mathcal{G}(X)$ at the m locations $\lambda_i \in X$. We obtain for the k^{th} basis function coefficient $f_k \equiv f_k(\phi)$ with respect to ϕ_k that

$$\begin{aligned}
f_k &= \langle f\,,\,\phi_k \rangle && \text{(22a)}\\
&= \left\langle \sum_{j=1}^{m} f(\lambda_j)\,\tilde{k}_j\,,\,\phi_k \right\rangle. && \text{(22b)}
\end{aligned}$$

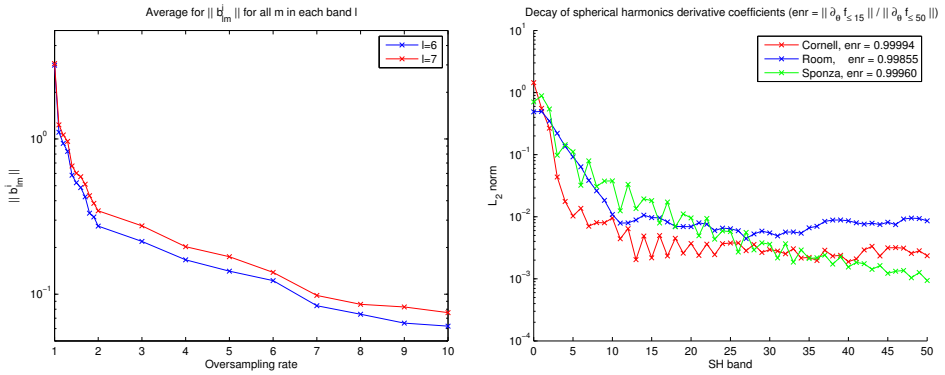


Figure 5: ***Left:*** *The effect of oversampling on $\|\gamma_k\|$ for approximate basis projection for the spherical harmonics space $\mathcal{H}_{\leq 5}$. Shown is $\|\gamma_k\|$ assuming the residual signal is contained in one band $l = 6$.* ***Right:*** *Decay of the spherical harmonics coefficients for typical scenes in light transport. As can be seen, the main signal components lie in low dimensional subspaces, making our theory applicable.*

Since $f = \hat{f} + \check{f}$, we can write this as

$$f_k = \left\langle \sum_{j=1}^{m} \left(\hat{f}(\lambda_j) + \check{f}(\lambda_j) \right) \tilde{k}_j \, , \, \phi_k \right\rangle . \tag{22c}$$

Exploiting linearity then yields

$$f_k = \underbrace{\left\langle \sum_{j=1}^{m} \hat{f}(\lambda_j) \tilde{k}_j \, , \, \phi_k \right\rangle}_{\hat{f}_k(\phi)} + \underbrace{\left\langle \sum_{j=1}^{m} \check{f}(\lambda_j) \tilde{k}_j \, , \, \phi_k \right\rangle}_{\mathrm{err}_k(f)} \tag{22d}$$

where $\hat{f}_k(\phi)$ is the exact basis function coefficient, corresponding to the orthogonal projection of the input signal, and the second term $\mathrm{err}_k(f)$ represents the error which is caused by an input not contained in $\mathcal{H}(X)$. Again exploiting linearity we obtain for the error term

$$\mathrm{err}_k(f) = \left\langle \sum_{j=1}^{m} \check{f}(\lambda_j) \, \tilde{k}_j \, , \, \phi_k \right\rangle \tag{23a}$$

$$= \sum_{j=1}^{m} \check{f}(\lambda_j) \left\langle \tilde{k}_j \, , \, \phi_k \right\rangle \tag{23b}$$

$$= \sum_{j=1}^{m} \check{f}(\lambda_j) \, s_{kj} \tag{23c}$$

where the s_{kj} are the elements of the sampling matrix $S_\phi(\Lambda)$. With the expansion of the residual signal $\check{f}$ in the basis $\{\phi_i\}_{i=1}^{p}$ for $\mathcal{G}(X)$ one obtains

$$\mathrm{err}_k(f) = \sum_{j=1}^{m} \left(\sum_{i=n+1}^{p} \check{f}_i(\phi) \, \phi_i(\lambda_j) \right) s_{kj} \tag{23d}$$

and by interchanging the order of the summations this becomes

$$\mathrm{err}_k(f) = \sum_{i=n+1}^{p} \check{f}_i(\phi) \underbrace{\sum_{j=1}^{m} \phi_i(\lambda_j) \, s_{kj}}_{\gamma_i^k} . \tag{23e}$$

The error term can hence be written concisely as

$$\mathrm{err}_k(f) = \sum_{i=n+1}^{p} \check{f}_i(\phi) \, \gamma_i^k . \tag{23f}$$

The elements s_{kj} of the sampling matrix can be interpreted as the integration weights for the projection of $\phi_i(\lambda_j)$ for $i > n$ onto the k^{th} basis function $\phi_k(x)$ for $\mathcal{H}(X)$. The γ_i^k thus represents the error of the "projection" using the point samples λ_j as integration nodes, which would vanish were it exact since $\phi_i \notin \mathcal{H}(X)$ for $i > n$.

The effect of the signal and the coefficient γ_i^k on the error $\mathrm{err}_k(f)$ can be separated using the Cauchy-Schwarz inequality. Eq. 23f, with the sum interpreted as a dot product, then becomes

$$|\mathrm{err}_k(f)| \leq \|\check{f}\| \|\gamma_k\| \tag{24}$$

where $\check{f}$ is the vector of basis function coefficients beyond the band-limit and γ_k the corresponding vector of the γ_i^k. □

A similar error analysis can be found, for example, in [Mallat 2009], see in particular Chapter 5.1.4.

In Fig. 5 we demonstrate the practical relevance of the error formula in Proposition 1. Shown on the left is how the signal independent term $\|\gamma_k\|$ decreases as the oversampling rate increases. On the right we see that for typical light transport scenes the main signal component lies in a low dimensional function space so that the residual component $\|\check{f}\|$ is sufficiently small for our theory to be applicable.

7 Representation of Color Information

In Fig. 6 we provide additional experimental results for the representation of color information using reproducing kernel bases discussed in Sec. 4.1 in the paper. The figure is an extension of Fig. 3 in the main paper that shows the dependence of the quality of the color reproduction on the number of samples that are used. The figure was inspired by Fig. 6 in Peercy's original paper [1993].

References

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

BABUŠKA, I., BANERJEE, U., AND OSBORN, J. E. 2003. Survey of Meshless and Generalized Finite Element Methods: A Unified Approach. *Acta Numerica 12* (May), 1–125.

BENEDETTO, J. J., AND FERREIRA, P. J. S. G. 2001. *Modern Sampling Theory: Mathematics and Applications*. Applied and Numerical Harmonic Analysis. Birkhäuser.

BERLINET, A., AND THOMAS-AGNAN, C. 2004. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers.

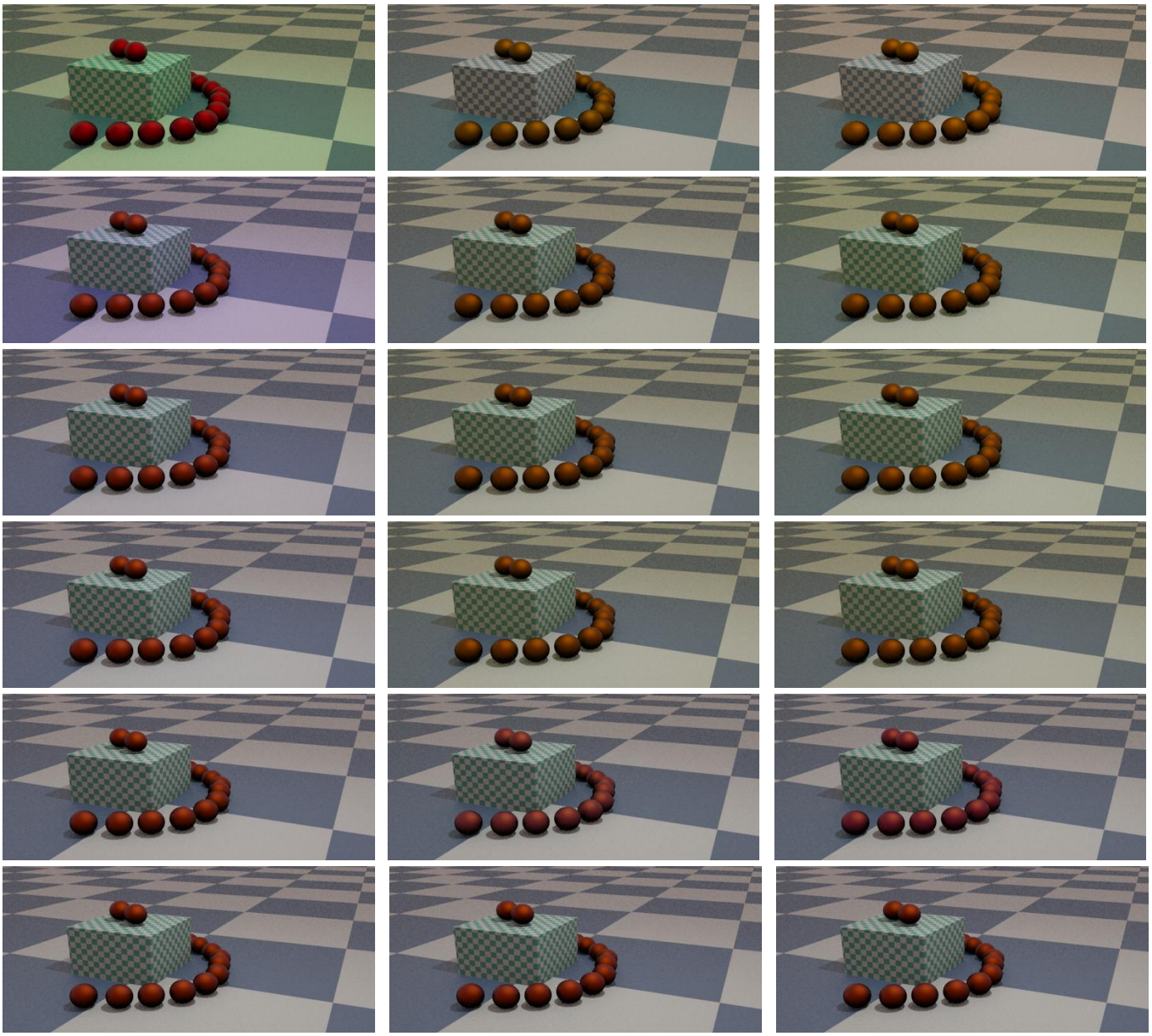


Figure 6: *Comparison of Peercy's point sampling, Peercy's linear model, linear model with sampling and final-step RKHS reprojection. The first column is Peercy's point sampling with 4, 9, 16, 25, and 36 samples from top to bottom; the second column is Peercy's linear model with 2, 3, 4, 5, and 6 basis functions projected from complete spectral density functions; the third column is linear model projected from 4, 9, 16, 25, and 36 samples of spectral density functions. The last row is in all columns a reference image computed with samples spaced at* 5 *nm intervals.*

CANDÈS, E. J., AND DONOHO, D. L. 1999. *Curvelets: A Surprisingly Effective Nonadaptive Representation of Objects with Edges*. Vanderbilt University Press, Nashville, TN.

COHEN, A., DAUBECHIES, I., AND VIAL, P. 1993. Wavelets on the Interval and Fast Wavelet Transforms. *Applied and Computational Harmonic Analysis 1*, 1 (Dec.), 54–81.

CUCKER, F., AND SMALE, S. 2002. On the Mathematical Foundations of Learning. *Bull. Amer. Math. Soc. 39*, 1–49.

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

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

FASSHAUER, G. E. 2010. Green's Functions: Taking Another Look at Kernel Approximation, Radial Basis Functions, and Splines. In *Approximation Theory XIII: San Antonio*, M. Neamtu and L. Schumaker, Eds., Springer Proceedings in Mathematics. Springer, New York.

FASSHAUER, G. E. 2011. Positive Definite Kernels: Past, Present and Future. *Dolomites Research Notes on Approximation 4*, 21–63.

GRÄF, M., AND POTTS, D. 2011. On the computation of spherical designs by a new optimization approach based on fast spherical Fourier transforms. *Numerische Mathematik 119*, 4 (July), 699–724.

GRÄF, M., POTTS, D., AND STEIDL, G. 2011. Quadrature rules, discrepancies and their relations to halftoning on the torus and the sphere. Tech. rep., Fakultät für Mathematik, TU Chemnitz.

GRIEBEL, M., AND SCHWEITZER, M. A., Eds. 2007. *Meshfree Methods for Partial Differential Equations III*. Lecture Notes in Computational Science and Engineering. Springer, Berlin.

KOENKER, R., MIZERA, I., AND YOON, J. 2012. What Do Kernel Density Estimators Optimize? *Journal of Econometric Methods 1*, 1, 15–22.

LI, S., AND LIU, W. K. 2004. *Meshfree Particle Methods*. Springer, Berlin / Heidelberg.

MALLAT, S. G., AND PEYRÉ, G. 2007. A Review of Bandlet Methods for Geometrical Image Representation. *Numerical Algorithms 44*, 205–234.

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

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.

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.

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

POGGIO, T., AND SMALE, S. 2003. The Mathematics of Learning: Dealing with Data. *Notices of the American Mathematical Society 50*, 5, 537–544.

RUDIN, W. 1987. *Real and Complex Analysis*. McGraw-Hill International Editions: Mathematics Series. McGraw-Hill Education - Europe.

SCHABACK, R., AND WENDLAND, H. 2006. Kernel Techniques: From Machine Learning to Meshless Methods. *Acta Numerica 15* (May), 543–639.

SCHÖLKOPF, B., AND SMOLA, A. J. 2002. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA.

SLOAN, I. H., AND WOMERSLEY, R. 2004. Extremal Systems of Points and Numerical Integration on the Sphere. *Advances in Computational Mathematics 21*, 1, 107–125.

SLOAN, I. H., AND WOMERSLEY, R. S. 2009. A variational characterisation of spherical designs. *Journal of Approximation Theory 159*, 2 (Aug.), 308–318.

SMALE, S., AND ZHOU, D.-X. 2004. Shannon Sampling and Function Reconstruction from Point Values. *Bull. Amer. Math. Soc. 41*, 279–305.

SMALE, S., AND ZHOU, D.-X. 2005. Shannon Sampling II: Connections to Learning Theory. *Applied and Computational Harmonic Analysis 19*, 3, 285–302.

SMALE, S., AND ZHOU, D.-X. 2007. Learning Theory Estimates via Integral Operators and Their Approximations. *Constructive Approximation 26*, 2, 20.

UNSER, M. 2000. Sampling—50 Years After Shannon. *Proceedings of the IEEE 88* (Apr.), 569–587.

ZAYED, A. I. 1993. *Advances in Shannon's Sampling Theory*. CRC Press, Boca Raton.