

# A Model for Geometric Reflections from Gaussian Surfaces extended to Grazing Angles

Peter E. Creasey  
ORCID:0000-0002-4049-4928

February 2025

## Abstract

A simple analytical model of geometrical reflections from Gaussian surfaces is presented that extends linear theory to high angles. By construction it is positive, reciprocal and conserves reflected rays. An interpretation is given as a random walk governed by a Fokker–Planck equation with an operator closely related to that of Zernike. For surfaces with isotropic deviations we compare the analytical BRDF to the ray-tracing of several realisations, finding very good agreement for rough surfaces with slope of the mean gradient  $\lesssim 15^\circ$  at all incident angles. Rougher surfaces deviate over a range of angles of incidence, where they exhibit non-Lambertian effects. An open-source implementation of the distribution computations is provided at <https://github.com/pec27/urdf>.

*Keywords:* Reflectance modeling, BRDF, Bidirectional Reflectance Distribution Function, Glint, Roughness, Physically-Based Rendering

## 1 Introduction

What direction does a ray travel in after geometrical reflection by a rough surface? The answer to this question, in combination with models of absorption, subsurface and diffraction effects, allows the construction of bidirectional reflectance distribution functions (BRDFs). Knowledge of the BRDFs of surfaces, along with radiative transfer, allows both the simulation of images, and the inference of the phys-

ical environment from real images. As such BRDFs are used across disciplines from computer graphics [8, 3], to astronomy [9, 7], oceanography [11], and remote sensing more broadly [13, 14, 15].

For surfaces with a Gaussian distribution of gradients, the linear theory for the geometric distribution of reflections at small incident and viewing angles for small slopes [2, 11] exhibits the physical properties of positivity and reciprocity, but is not unitary. Unitarity of geometrical reflections has the physical interpretation that every ray which intersects the surface exits somewhere; non-unitary BRDFs correspond to energy absorption. The linear theory deviates from unitarity due to the truncation of the distribution to prevent predicted reflections outside of the exterior hemisphere, a truncation whose magnitude grows with angle and slope.

At large slopes and/or angles, effects such as foreshortening, occlusion and multiple reflections become important, the latter two being somewhat dependent on the joint distribution of slope and displacement. These can be numerically sampled for a given surface distribution via ray tracing (e.g. [11]), though this brute-force approach is computationally expensive at high resolution due to the four-dimensional domain of a BRDF. An analytical approach is to apply a multiplicative correction to the density, such as a geometrical attenuation factor [18, 4] or visibility function [16, 7, 17, 12]. In general these corrections reduce deviations from unitarity, which is suggestive that the converse may apply, i.e. a model constructed to be unitary might be valid over a larger domain

than the linear theory.

Few models are explicitly unitary, exceptions include perfect specular reflection, Lambertian scattering or perfect retro-reflection. Constructing a unitary model from a non-unitary one is difficult due to the restriction of reciprocity, which prevents a naive re-normalisation of the densities. On the other hand, models that are already unitary remain so under convolution.

In this paper, we take a continuous-time version of the convolution approach, perturbing the specular model via the integrated action of a 2nd order linear differential operator corresponding to diffusion via a random walk. With suitably chosen coefficients we maintain reciprocity, unitarity and match linear perturbations in the low-angle, small slope regime. In the large-slope limit we recover the Lambertian model.

The utility of this model is studied for the case of isotropic surface deviations, in terms of its computational cost and the accuracy with respect to ray tracing. For computation we find a link with Zernike’s equation and provide a series expansion and recursive algorithm, an open-source implementation of which is made available at <https://github.com/pec27/urdf>. For comparison with ray tracing we simulate surfaces of squared-exponential autocovariance, demonstrating a very good agreement at all angles of incidence for mean slopes  $\lesssim 15^\circ$ .

The structure of this paper is as follows. In Section 2 we describe a random-walk model for the geometric reflections for surfaces with a Gaussian distribution of slopes. In Section 2.2 we consider the special case of isotropic deviations and provide a series expansion. In Section 3 we perform a comparison of the model to simulated BRDFs produced by the ray-tracing of random surfaces and to the Smith model. In Section 4 we give some concluding remarks and suggestions for future work.

## 2 Occlusion through Diffusion

In this section we describe the construction and evaluation of a model for the unitary distribution of geometric reflections. Section 2.1 describes the choice

of linear differential operator with coefficients chosen such that the integrated action forms a unitary reciprocal model that matches linear perturbations in the low-angle, low roughness limit, and show that it asymptotes to Lambertian scattering in the high-roughness limit. Section 2.2 considers the model in the case of isotropic surface deviations, and shows evaluation of the model is equivalent to a series in Zernike polynomials, and describes a summation with faster convergence.

Let us begin by making a choice of notation that simplifies later calculation. For an incident ray from direction  $\omega_i = (\sin \theta_i \cos \phi_i, \sin \theta_i \sin \phi_i, \cos \theta_i)$ , where the angle of incidence  $\theta_i \in [0, 90^\circ]$  denotes the exterior hemisphere to the surface, we form the projection onto the unit disk  $\{\mathbf{x} \in \mathbb{R}^2 : |\mathbf{x}| < 1\}$  of the direction of specular reflection, i.e.  $\mathbf{s} = (-\sin \theta_i \cos \phi_i, -\sin \theta_i \sin \phi_i)$ , with non-unit projected magnitude  $s = \sin \theta_i$ . Since we will only consider directions in this hemisphere, this is a one-to-one mapping. Similarly we transform the reflected direction  $\omega_r = (\sin \theta_r \cos \phi_r, \sin \theta_r \sin \phi_r, \cos \theta_r)$  to its projection onto the disk,  $\mathbf{r} = (\sin \theta_r \cos \phi_r, \sin \theta_r \sin \phi_r)$ , with magnitude  $r$ . The BRDF  $f_r(\omega_i, \omega_r)$ , defined as the density of rays reflected from  $\omega_i$  to  $\omega_r$ , per unit projected direction  $\cos \theta_r d\omega_r$ , can then be re-written

$$f_r(\omega_i, \omega_r) = f(\mathbf{r}, \mathbf{s}) . \quad (1)$$

An illustration of this projection is shown in Fig. 1.

### 2.1 Linear operator construction

The distribution of continuous-time random walks initialised to the specular direction has time-evolution governed by a Fokker–Planck equation. Introducing a continuous ‘time’ parameter  $t \in [0, 1]$ , and setting initial conditions to be the Dirac function  $\rho(\mathbf{r}; \mathbf{s}, t = 0) = \delta(\mathbf{r} - \mathbf{s})$ , the time-evolution is given by

$$\partial_t \rho(\mathbf{r}) = \mathcal{D}[\rho(\mathbf{r})] , \quad (2)$$

s.t.  $f(\mathbf{r}, \mathbf{s}) = \rho(\mathbf{r}; \mathbf{s}, 1)$ ,  $\mathcal{D}$  being a second-order linear differential operator which we take by ansatz to be

$$\mathcal{D}[f(\mathbf{r})] = \nabla \cdot [2(1 - r^2) \boldsymbol{\Sigma} \cdot \nabla f] , \quad (3)$$

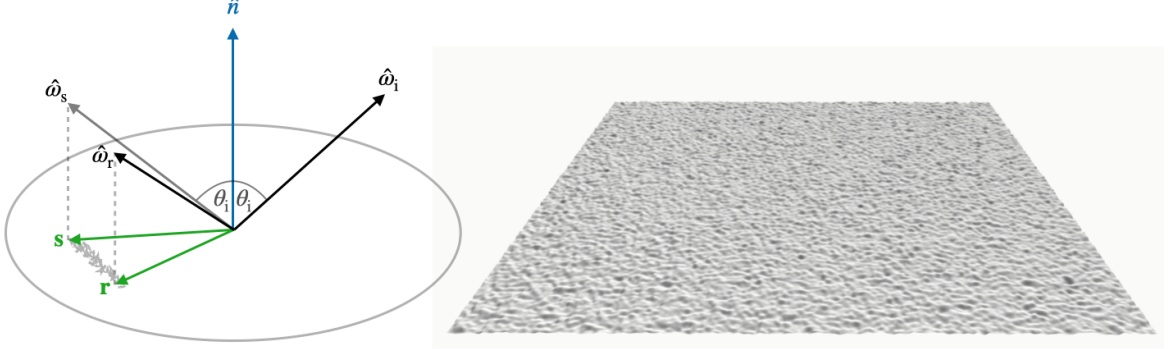


Figure 1: *Left* Illustration of the procedure in Section 2.1. An incoming ray from direction  $\omega_i$  is reflected from a surface with mean normal  $\hat{n}$ . The implied specular reflection due to this mean normal has outgoing direction  $\omega_s$ , which in projection is  $\mathbf{s}$ . The density of geometric reflections with projected direction  $\mathbf{r}$  (corresponding to  $\omega_r$ ) is the density of random-walks which travel from  $\mathbf{s}$  to  $\mathbf{r}$ , according to Fokker–Planck equation (2) or Itô equation (5). *Right* An example surface with Gaussian statistics (taken from the process in Sec. 3 with slope of the mean gradient  $7.1^\circ$ ).

where  $\Sigma_{ij}$  is the symmetric covariance matrix of the heightfield gradients, and all  $\nabla$ 's are taken w.r.t.  $\mathbf{r}$ . We then justify this choice of  $\mathcal{D}$  by showing it results in a distribution that satisfies the following three attributes:

1. Reciprocity,
2. Unitarity,
3. Matches perturbations for single-scattering at low order in incidence angle and  $|\Sigma|$ .

(1) Reciprocity, i.e.  $f(\mathbf{r}, \mathbf{s}) = f(\mathbf{s}, \mathbf{r})$ , corresponds to the time-reversibility of geometric optics. Interpreted in terms of random walks from the specular direction, we require walks starting at  $\mathbf{s}$  and traveling to  $\mathbf{r}$  at  $t = T$  (with density w.r.t.  $d\mathbf{r}$ ) to equal the density of walks starting  $\mathbf{r}$  and traveling to  $\mathbf{s}$  (w.r.t.  $d\mathbf{s}$ ). It is sufficient if the density  $f(\mathbf{r}, \mathbf{s})$  is given by the inner product of a symmetric linear operator. For symmetry, it is sufficient to be self adjoint. The integral of a self-adjoint linear differential operator is necessarily a linear self-adjoint operator.  $\mathcal{D}$  is evidently linear, and we note Eqn. (2) implicitly contains within it the diffusion tensor  $D_{ij} = 2(1 - r^2)\Sigma_{ij}$

which vanishes as  $r \rightarrow 1$ . As such it can be verified to be self-adjoint over the unit disk via integration by parts.

(2) Unitarity requires  $\int_{r < 1} d\mathbf{r} f = 1$ . This is satisfied for  $\rho(t = 0)$  and is automatically satisfied  $\forall t > 0$  if  $\partial_t \int_{r < 1} d\mathbf{r} \rho = 0$ , i.e. it is sufficient to show  $\mathcal{D}$  is conservative over the unit disk. This is immediate since Eqn. (3) is a divergence over  $r < 1$ , and as  $r \rightarrow 1$  the argument of the divergence meets the Neumann (fluxless) condition due to the disappearance of the diffusion tensor.

(3) First-order perturbations for Gaussian facets are discussed at length in works such as [2, 11]. In projection these correspond to a binormal distribution of reflections

$$\mathbf{r}_{\text{linear}} \sim \mathcal{N}(\mathbf{s}, 4(1 - s^2)\Sigma_{ij}). \quad (4)$$

Notably this is not normalised when restricted to  $r \leq 1$ , however here we are attempting only a first-order analysis for  $r, s, |\Sigma|$  small. In Itô notation, Eqn. (2) corresponds to a continuous-time random walk with instantaneous drift  $\partial_i D_{ij}$  and variance  $2D_{ij}$ , i.e.

$$d\mathbf{r}_t = -4(\Sigma \cdot \mathbf{r}) dt + 2\sqrt{1 - r^2} \mathbf{Q} d\mathbf{W}_t, \quad (5)$$

(see e.g. [6, 4.3]) where  $\mathbf{W}_t \in \mathbb{R}^2$  a Wiener process and  $\mathbf{Q}$  satisfies  $\mathbf{Q}\mathbf{Q}^T = \Sigma$ , e.g. a Cholesky decomposition. For both  $s$  and  $|\Sigma|$  small the mean corresponds to Eqn. (4) since the drift term (coefficient of  $dt$ ) is non-linear, and so to leading order the mean remains the initial value  $\mathbf{s}$ . The variance of Eqn. (4) ( $4(1-s^2)\Sigma_{ij}$ ) corresponds to the diffusion rate in Eqn. (5) ( $2D_{ij} = 4(1-r^2)\Sigma_{ij}$ ) to leading order, since  $\mathbf{r} \approx \mathbf{s} + O(\Sigma \cdot \mathbf{s})$ .

This completes the satisfaction of properties (1)-(3). A reasonable question is whether there are any other 2nd order operators which satisfy these. Inspection reveals that the term  $(1-r^2)$  in Eqn. (3) could be replaced by functions of  $g(r) = 1-r^2 + O(r^3)$  which satisfy  $\lim_{r \rightarrow 1} g \rightarrow 0$ , and as such  $1-r^2$  is the lowest order function which matches the Gaussian 1st order theory.

The BRDF described above can be evaluated by the usual three methods for diffusion equations. That is, we may perform Monte-Carlo sampling (via discretising Eqn. 5), we may use the Fokker-Planck equation (2) via finite-differences (or other numerical method) the to evaluate over  $\mathbf{r}$  at a fixed  $\mathbf{s}$  (or vice-versa), or we may evaluate a single  $f(\mathbf{r}, \mathbf{s})$  directly via construction and summation of the orthogonal eigenfunctions of  $\mathcal{D}$ .

For this latter method, we remark that these eigenfunctions are polynomials. This can be seen by substitution into  $\mathcal{D}[x^n y^m]$ , with  $n, m \in \mathbb{N}$  and  $x, y$  the cartesian components, forms a polynomial of degree  $\leq n+m$  (in  $x$  and  $y$ ), i.e. we form invariant subspaces. The polynomial coefficients of these eigenfunctions may be found for small  $n+m$  via matrix methods (e.g. Gram-Schmidt). For large  $n+m$  this presents the usual difficulties with numerical stability. One case special enough to merit further investigation of evaluation, however, is that of isotropic deviations, which we consider in the next section.

Finally, let us remark that the behaviour in the high roughness limit converges to Lambertian scattering, i.e.  $f \sim \frac{1}{\pi}$ . Since we have a self-adjoint operator on a bounded domain we are guaranteed a spectrum of eigenfunctions, and inspection reveals  $\mathcal{D}$  to be positive (all eigenvalues  $\geq 0$ ), and that for  $\Sigma$  positive definite the only eigenfunction with eigenvalue 0 must be constant, i.e. the stationary state must be

the Lambertian model of uniform radiance  $\frac{1}{\pi}$ .

One might be suspicious that this last attribute was almost ‘free’, in particular if you chose an operator to be conservative and self-adjoint w.r.t. the radiance measure ( $d\mathbf{r} = \cos\theta_r d\omega_r$ ), then you are guaranteed that a constant density function w.r.t. the same measure has eigenvalue zero (albeit you can choose operators with multiple orthogonal functions of zero eigenvalue, e.g.  $\mathcal{D}$  if  $\Sigma$  degenerate). In this sense one might view the Lambertian model as more fundamental than just a zero-th order approximation, being the only function to have zero eigenvalue with respect to all conservative self-adjoint operators w.r.t. the radiance measure.

## 2.2 Isotropic evaluation and relationship to Zernike’s Operator

In the limit of isotropic surface gradients,  $\Sigma_{ij}^{\text{iso}} = \sigma^2 \delta_{ij}$ , we may write  $\mathcal{D}^{\text{iso}}[\rho] = 2\sigma^2 \mathcal{L}[\rho]$ , where

$$\mathcal{L}[f(\mathbf{r})] = \nabla \cdot ((1-r^2) \nabla f). \quad (6)$$

This is very similar to the operator of Zernike (in modern formulation  $\mathcal{L}^{\text{Zernike}} = \nabla^2 - (\mathbf{r} \cdot \nabla)^2 - 2\mathbf{r} \cdot \nabla$ , see e.g. [10]), from which Eqn. (6) differs only by a factor of  $-\partial_{\phi_r \phi_r}$ . The eigenfunctions of Zernike’s operator are the well-known Zernike polynomials,  $Z_n^m(\mathbf{r})$ , and since these are separable in  $r$  and  $\phi_r$ , these are also eigenfunctions of  $\mathcal{L}$ . By inspection we may write  $-\mathcal{L}[Z_n^m] = \lambda_n^m Z_n^m$ , with eigenvalue

$$\lambda_n^m = n(n+2) - m^2. \quad (7)$$

This is a sum of  $-m^2$  with the eigenvalue of  $-\mathcal{L}^{\text{Zernike}}$ ,  $\lambda_{mn}^Z = n(n+2)$ , the former due to the aforementioned  $-\partial_{\phi_r \phi_r}$ . Since by construction the Zernike polynomials have  $n \geq |m|$  it follows that the  $\lambda_n^m$  in (7) are positive for every Zernike polynomial.

We can thus write the BRDF model corresponding to isotropic surface gradients with components of variance  $\sigma^2$  as the sum over normalised eigenfunctions

$$f_{\sigma}^{\text{iso}}(\mathbf{r}, \mathbf{s}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n+2}{\pi \epsilon_m} Z_n^m(\mathbf{r}) Z_n^m(\mathbf{s}) e^{-2\sigma^2 \lambda_n^m}, \quad (8)$$

with Neumann factor  $\epsilon_0 = 2$ ,  $\epsilon_{m \neq 0} = 1$ . As with many direct substitutions of Zernike polynomials, however, caution must be exercised when naively evaluating them to high order, due to accumulation of numerical error in sums involving many relevant  $Z_n^m$  before the exponential decay of  $\exp(-2\sigma^2 \lambda_n^m)$  provides a sufficient bound on the remaining series. In particular, the summation of terms of eigenvalue up to  $\lambda$  takes  $O(\lambda^2)$  terms<sup>1</sup>.

A re-summation of Eqn. (8) with more rapid convergence, detailed in Appendix A, is given by

$$f_\sigma^{\text{iso}}(\mathbf{r}, \mathbf{s}) = \frac{1}{\pi} \sum_{n=0}^{\infty} e^{-8\sigma^2 n(n+1)} \text{Re} \left[ \Omega_n(1-2r^2, 1-2s^2, rse^{i(\phi_s - \phi_r) - 4\sigma^2(2n+1)}) \right], \quad (9)$$

with  $\Omega_n(\mu, \nu, z)$  a function of  $\mu, \nu \in \mathbb{R}$ ,  $z \in \mathbb{C}$ , symmetric under exchange<sup>2</sup> of  $\mu, \nu$ , defined as the double sum of Jacobi polynomials

$$\Omega_n(\mu, \nu, z) = \sum_{\alpha, \beta=0}^n \binom{n}{\alpha} \binom{n}{\beta} (1 + \alpha + \beta) \times \frac{z^{\alpha+\beta}}{(1-z)^{2+\alpha+\beta}} \left( \frac{2n+1}{\binom{\alpha+\beta+n}{n}} P_n^{(\alpha, -\alpha)}(\mu) P_n^{(\alpha+\beta, -\beta)}(\nu) + \frac{2z}{\binom{1+\alpha+\beta+n}{n+1}} P_n^{(1+\alpha, -\alpha)}(\mu) P_n^{(1+\alpha+\beta, -\beta)}(\nu) - \frac{(2n+1)z^2}{\binom{2+\alpha+\beta+n}{n+2}} P_n^{(2+\alpha, -\alpha)}(\mu) P_n^{(2+\alpha+\beta, -\beta)}(\nu) \right) \quad (10)$$

where  $P_n^{(\alpha, \beta)}$  are the standard Jacobi polynomials and  $\binom{n}{a}$  refers to the binomial coefficient. For ref-

erence, the first three terms of this function are

$$\Omega_0(\mu, \nu, z) = \frac{1+2z-z^2}{(1-z)^2} \quad (11)$$

$$\Omega_1(\mu, \nu, z) = \frac{3+2z-3z^2}{(1-z)^2} \mu\nu + \frac{z(2-z)}{(1-z)^3} (3\mu\nu + \mu + \nu - 1) + \frac{z(4-z)}{(1-z)^4} (\mu+1)(\nu+1) \quad (12)$$

$$\Omega_2(\mu, \nu, z) = \frac{5+2z-5z^2}{4(1-z)^2} (3\mu^2-1)(3\nu^2-1) + \frac{z(3-2z)}{12(1-z)^3} [(9\mu^2+2\mu-3)(9\nu^2+2\nu-3) - 16\mu\nu] + \frac{3z(2-z)}{8(1-z)^4} [4(\mu-\nu)^2 + (19\mu\nu-9\mu-9\nu+3)(\mu+1)(\nu+1)] + \frac{3z(3-z)}{(1-z)^5} (\mu+1)(\nu+1)(\mu\nu-1) + \frac{3z(6-z)}{4(1-z)^6} (\mu+1)^2(\nu+1)^2 \quad (13)$$

The summation of terms of Eqn. (9) up to missing eigenvalue  $\lambda$  is reduced to  $O(\lambda^{\frac{3}{2}})$  terms (including the double sum in Eqn. (10)). This is both numerically more efficient and more accurate, since including all terms up to a given eigenvalue requires fewer iterations of recurrence relations for orthogonal polynomials. Details of the recurrences used are described in Appendix B. An open-source implementation of the calculation of these is provided at <https://github.com/pec27/urdf>.

In Figure 2 we plot the relative error due to the truncation of the series expansion in Eqn. (9). Each symbol corresponds to a simulation for a randomly chosen  $\sigma$ ,  $n_{\text{max}}$  with  $\sigma^2(n_{\text{max}}+1)(n_{\text{max}}+2)$  in the range  $[0.5, 2.5]$ , and at four different incidence angles. For each simulation we give the maximum error over all  $\mathbf{r}$ , taken relative to the peak  $\max_{\mathbf{r}} f_\sigma^{\text{iso}}(\mathbf{r}, \mathbf{s})$ . As one might expect, the scaling of the relative error is

<sup>1</sup>The partial sum up to  $n = n_{\text{max}} - 1$ , for all  $m$  takes  $O(n_{\text{max}}^2)$  terms and the smallest missing eigenvalue is  $\lambda_{n_{\text{max}}}^{n_{\text{max}}} = 2n_{\text{max}}$ .

<sup>2</sup>The symmetry under permutations is somewhat opaque in this formation, see Appendix A for details.

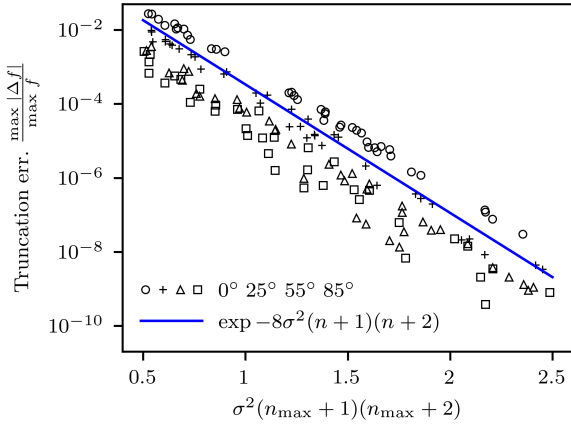


Figure 2: Truncation error of the series in Eqn. (9) as a function of  $-8\sigma^2(n_{\max}+1)(n_{\max}+2)$ . Truncation error is given as the maximum error over all  $\mathbf{r}$  relative to the peak ( $\max_{\mathbf{r}} f$ ), at different combinations of fixed  $\sigma$ ,  $n_{\max}$ ,  $\theta_i$ . Circles, plusses, triangles and squares correspond to incident angles  $\theta_i = 0^\circ$ ,  $25^\circ$ ,  $55^\circ$  and  $85^\circ$  respectively. The solid line indicates the exponential of  $-8\sigma^2(n+1)(n+2)$ , the negative of the smallest missing eigenvalue in Eqn. (8).

strongly related to the largest missing exponent in Eqn. (8), i.e.  $-2\sigma^2\lambda = -8\sigma^2(n_{\max}+1)(n_{\max}+2)$ .

### 3 Comparison with Simulated Surfaces

In the previous section we considered a model of geometric reflections from a surface with Gaussian deviations constructed to match several properties that distribution is known to obey. To be of more than theoretical novelty, however, we are interested in whether the model is a good model of those reflections. In this section we construct BRDFs by ray-tracing large numbers of rays across surfaces sampled to have isotropic Gaussian deviations, repeated for several roughness values and incident angles. We compare the model with both the simulation, linear theory, and the popular shadowing and masking term of Smith [16]. We begin in section 3.1 by describing the methodology used, before showing results in section 3.2.

#### 3.1 Microfacet distributions and ray-tracing

One type of surface that has a binormal distribution of surface gradients with variance  $\Sigma$  is one whose height function is itself a Gaussian random field. In our case we choose the mean-zero Gaussian random field  $v(\mathbf{u})$  with the squared-exponential covariance function

$$\mathbb{E}[v(\mathbf{u}_1)v(\mathbf{u}_2)] = \exp\left[-\frac{1}{2}(\mathbf{u}_1 - \mathbf{u}_2) \cdot \Sigma \cdot (\mathbf{u}_1 - \mathbf{u}_2)\right]. \quad (14)$$

The function is realised by convolving the discretised transfer function with white noise using a Fourier method on a regular grid to create a periodic surface. The facets are then given by a triangulation of this heightfield, with surface elements varying from  $2 \times 128^2$  to  $2 \times 4096^2$ . An illustration is shown in the right panel of Fig. 1, of a surface with  $\sigma = 0.1$  with  $2 \times 1024^2$  elements.

In order to efficiently ray-trace reflections for a large number of rays ( $> 10^7$  for the largest simu-

lations), a quad-tree data structure is constructed where each node stores the maximum height of its leaves. Ray-intersections can then be efficiently calculated by marching across the quadtree, refining when necessary. Upon intersection with a triangle the reflection is calculated using the analytic normal, and the ray-tracing repeats for further reflections until the ray leaves the upper bound for the entire simulated surface. A code for such simulations is available on request from the author.

In Table 1 we provide some statistics for the simulations used in this paper, which are every combination of the four surface roughness values  $\sigma$  of 0.05, 0.1, 0.2 and 0.3 with three different incident angles  $\theta_i$  of  $0^\circ$ ,  $30^\circ$  and  $75^\circ$ . For intuition we also quote, for each  $\sigma$ , the slope of the mean gradient  $\alpha_\sigma = \arctan(\sqrt{\frac{\pi}{2}}\sigma)$ .

For each of these simulations we provide the number of elements in the periodic surface, and the number of rays in the simulation. The start point of these rays is chosen at a fixed height above the surface, with horizontal positions sampled from a regular lattice, and angle  $\phi_i$  regularly sampled from  $0$ - $360^\circ$  in order to reduce the variance of the simulated distributions.

For a final comparison we consider the implied BRDF from single-reflections using a shadowing-masking term

$$f^{\text{smith}}(\omega_i, \omega_r) = \frac{D(\omega_h) G_1(\omega_i, \omega_h) G_1(\omega_r, \omega_h)}{4 \cos \theta_i \cos \theta_r}, \quad (15)$$

where  $\omega_h$  is the normalised half-vector that would reflect from  $\omega_i$  to  $\omega_r$ ,  $D$  is the density of surface elements with the given normal, and  $G_1$  is the Smith [16] model for the Gaussian surface, and the surface is assumed fully reflective (for a more detailed discussion of these terms see e.g. Walter [20]).

### 3.2 Results

In the right-hand column of Table 1 we quote a statistical comparison of the distribution of simulated rays with the numerical integration of the model, where  $Z_n$  is the Fasano and Franceschini extension of the Kolmogorov-Smirnov test to 2-dimensions [5]. We use  $n = 5000$  rays randomly sampled from the distribution and compute the statistic in the Carte-

sian projection of  $\mathbf{s}, \mathbf{r}$  (unlike the K-S test, the two-dimensional extension is dependent on choice of coordinates). We cannot use the entire distribution, due to the dependence of sampling only a single surface (per  $\sigma$ ,  $\theta_i$  pair), i.e. rays which reflect from distinct facets are correlated due to the spatial autocorrelation. The 95% confidence level of  $Z_{5000}$  is around 1.71 for uncorrelated data, and we have identified those simulations in excess of this. At high incidence angle this projection results in somewhat correlated coordinates, however the confidence intervals on  $Z_n$  exhibit only a weak dependence on correlation [5], and as such we use the same threshold.

In Figure 3 the results from the simulations in Table 1 have been plotted as BRDFs in the frame where  $\phi_s = 45^\circ$ , where we show the iso-radiance contours that contain 39.3% and 86.4% of the distribution respectively (these percentages correspond to 1- and 2-sigma for an unbound bivariate normal distribution).

Let us first note that the model BRDFs over the top two rows ( $\sigma = 0.05$  and  $0.1$ , blue contours) show insignificant differences compared to the noise floor of the simulated surfaces (magenta contours). This noise floor is visible by eye as non-smooth contours, and is a result of the finite numbers of rays and elements in the surface (both  $\geq 10^6$ ). The numerical evaluation of the model in Eqn. (9) also has a precision limit (described in the previous section), but this is not discernible by eye. In the left-most column is incidence  $\theta_i = 0^\circ$ , and it should be no surprise that the simulated surfaces are well described by the linear theory for low incidence and slope, and that the model (which has been constructed to match linear theory in this regime), matches both.

At larger incidence angles or greater slopes we expect deviations from the linear theory, which is apparent at  $\sigma \geq 0.1$ ,  $\theta_i = 30^\circ$  (middle panel, second row). Here the iso-radiance contour for the linear model starts to exhibit an offset from the simulated surfaces. What is remarkable, however, is the very close match of the model to the simulated surface (it may even be that deviations here are due to the finite sampling of the surface), which continues to  $\theta_i = 75^\circ$  (2nd row, right-most panel).

At the largest two roughness values  $\sigma = 0.2$  and  $0.3$  (the lowest two rows of panels) the simulated surfaces

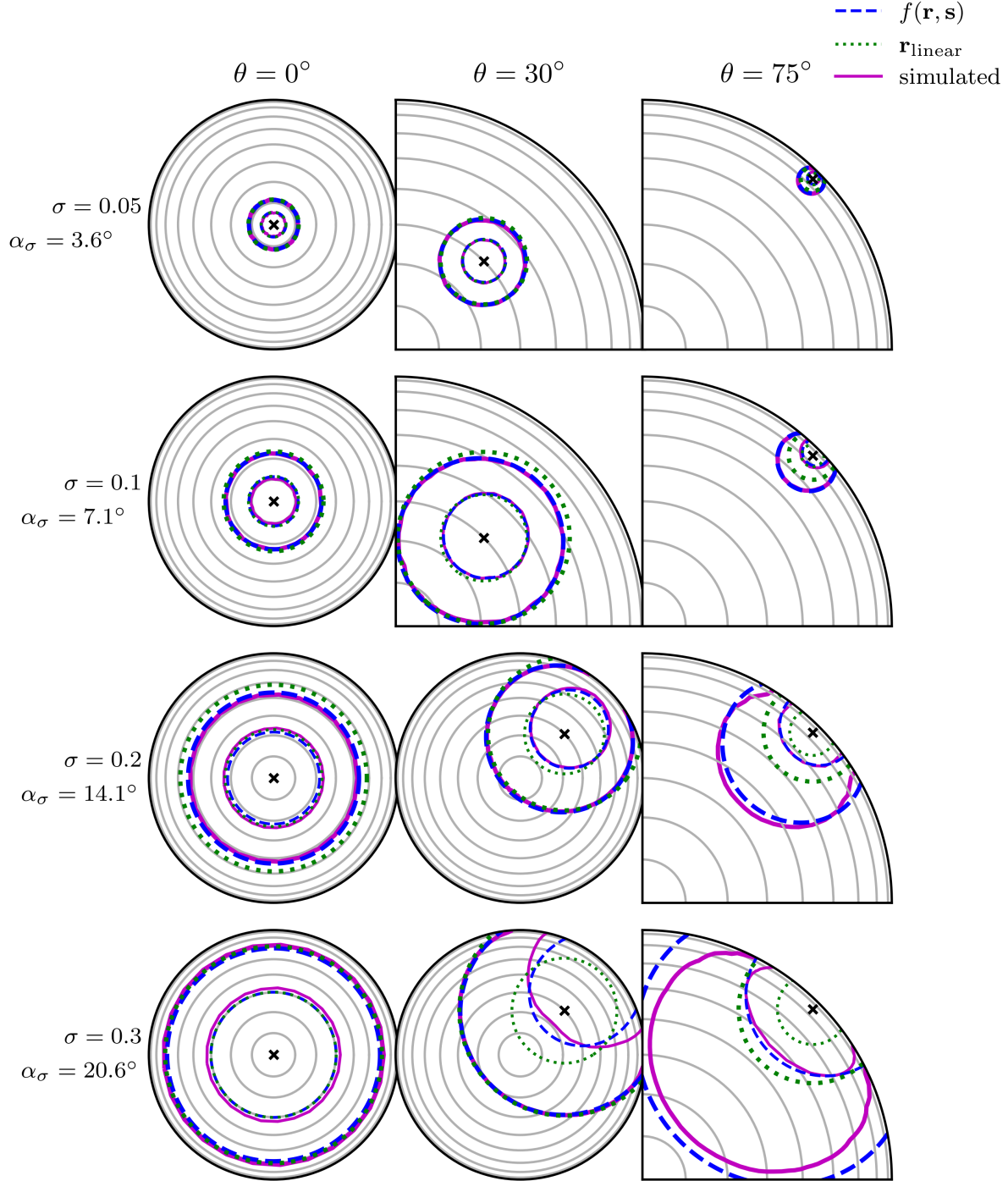


Figure 3: BRDFs for three different incidence angles (*columns*, from left to right  $0^\circ$ ,  $30^\circ$  and  $75^\circ$ ) and for surfaces with four isotropic roughness values  $\sigma$  (*rows*, from top to bottom 0.05, 0.1, 0.2 and 0.3) and their corresponding slope of the mean gradient  $\alpha_\sigma$ . *Light* and *heavy lines* indicate iso-radiance contours that contain 39.3% and 86.4% of the distribution respectively. *Blue dashed lines* indicate the BRDF given by Eqn. (9), whilst *solid magenta lines* are numerically realised via ray-tracing geometric reflections from a simulated surface. *Green dotted lines* indicate the linear theory. *Black crosses* indicate the specular reflection point for the given incident direction.



Roughness $\sigma$	Slope $\alpha_\sigma$	Incidence $\theta_i$	Surface elements	No. of rays	Significance $Z_{5000}$
0.05	3.6°	0°	$2 \times 128^2$	163,840	0.94
		30°	$2 \times 256^2$	655,360	1.54
		75°	$2 \times 256^2$	1,966,080	1.48
0.1	7.1°	0°	$2 \times 128^2$	163,840	0.95
		30°	$2 \times 256^2$	655,360	1.43
		75°	$2 \times 256^2$	1,966,080	1.68
0.2	14.1°	0°	$2 \times 256^2$	4,194,304	1.52
		30°	$2 \times 512^2$	33,554,432	1.62
		75°	$2 \times 4096^{2\dagger}$	50,331,648	2.56
0.3	20.6°	0°	$2 \times 256^2$	4,194,304	2.60
		30°	$2 \times 512^2$	33,554,432	2.99
		75°	$2 \times 512^2$	33,554,432	3.46

Table 1: The simulations used in this paper. From left to right,  $\sigma$ , the standard deviation of the isotropic surface gradient in a single direction,  $\alpha_\sigma$ , the corresponding slope of the mean gradient,  $\theta_i$ , the incident angle, the number of surface elements in the realisation of a periodic heightfield, the number of randomly sampled rays, the  $Z_n$  measure of deviation from the model in Section 2, for  $n = 5000$ , where values in bold are above the 95% confidence level of 1.71 Fasano [5].

† The number of surface elements was increased beyond the others to resolve the 86.4% contour in this simulation, which appears to contain a rapid transition likely due to  $\theta_i + \alpha_\sigma \approx 90^\circ$ .

exhibit discrepancies both from the linear model and the model in this paper. At incidence  $\theta = 0^\circ$  the light is scattered slightly towards grazing angles, visible in both 1 and 2-sigma contours. At  $\theta = 30^\circ$  the 2-sigma contour is very close (except for grazing angles), yet the 1-sigma contour shows significant deviation in the centroid. At  $\theta = 75^\circ$  the model predicts the 2-sigma contour as too broad, and the 1-sigma contour exhibits deviations at grazing angles. For even higher roughness values (not shown), the numerically simulated surfaces start to exhibit opposition effects.

In Figure 4 we make a comparison of the highest (75°) incidence angle simulation at the three smaller surface roughness values with the Smith BRDF in Eqn. (15). For additional comparison we include a third iso-radiance contour that contains 11.8% of the distribution (corresponding to the 0.5-sigma contour of the bivariate normal distribution) to describe the central peak of the distributions.

At  $\sigma = 0.05$  (left-most panel) the models are almost indistinguishable from each other and the numerical results, showing that at this incidence both

grazing terms are very accurate. At  $\sigma = 0.1$  (middle panel) we see some discrepancy, with the 1 and 0.5 sigma contours being closer to those of  $f(\mathbf{r}, \mathbf{s})$ , with the Smith values suggesting too much concentration at grazing angles. Interestingly for the 2-sigma contour the result is marginal, with  $f(\mathbf{r}, \mathbf{s})$  closer to the simulation at lower incidence angles, whilst Smith does better as we approach grazing (high incidence) angles.

For the  $\sigma = 0.2$  (right panel) result we see that again for the 1 and 0.5 sigma contours  $f(\mathbf{r}, \mathbf{s})$  is a significantly better predictor of the simulation, with the Smith distribution concentrated too much at high angles. Note however that for the 2-sigma contour the Smith result is better at both low and high incidence angles. Taken together with the previous results, it seems that  $f(\mathbf{r}, \mathbf{s})$  is a better model for the central peak, whilst the Smith result is more accurate in the tail of the distribution.

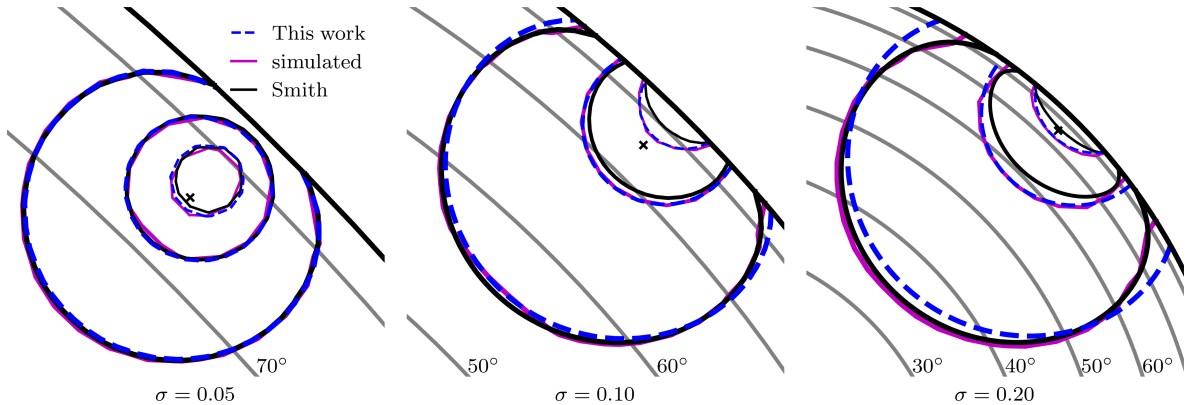


Figure 4: BRDFs with incidence angle  $75^\circ$  for three different isotropic surface roughness values  $\sigma$ , from left to right of 0.05, 0.1 and 0.2. *Heavy, normal and narrow lines* indicate iso-radiance contours that contain 86.4, 39.3 and 11.8% of the distribution respectively (corresponding to 2, 1 and 0.5-sigma contours for a bivariate normal distribution). *Blue dashed lines* indicate the BRDF given by Eqn. (9), whilst *solid magenta lines* are numerically realised via ray-tracing geometric reflections from a simulated surface. *Black lines* indicate results from the Smith shadowing-masking function, given in Eqn. (15). *Black crosses* indicate the specular reflection point for the given incident direction.

## 4 Concluding remarks

In this work we have described a model for the distribution of geometric reflections for surfaces with Gaussian deviations constructed via the continuous action of a linear differential operator. This model is by construction unitary and obeys Helmholtz reciprocity. It asymptotically matches linear theory at small roughness and small incidence angles, and at high roughness values asymptotes to Lambertian scattering. This model is equivalent to a distribution of reflections found via random-walk starting from the specular direction.

For the case of isotropic deviations we provide an analytic expansion for the evaluation of the BRDF in terms of Zernike polynomials. We also provide a novel re-summation with more rapid convergence, and an open-source implementation is provided at <https://github.com/pec27/urdf>.

The main result of this paper is that this model provides a very good match to the geometric reflection distribution for isotropic surfaces at all incidence

angles for low roughness. This is found via the ray-tracing of simulated surfaces, with deviations only starting to become significant at mean slope angles above  $15^\circ$  ( $\sigma \geq 0.2$ ). This is superior to linear theory, which matches only in the low-roughness, small-angle regime. In comparison with the Smith masking model this work gives more accurate results around the peak of the distribution near grazing, and comparable results elsewhere, except for away from the peak at higher roughness.

Directions for future work could include investigation of anisotropic deviations, either by finding a new spectral method or by brute-force sampling the random walks or finite-differencing the Fokker-Planck equation. As an explicitly unitary model, The random-walk interpretation may find use in Monte-Carlo rendering. Other directions could include modifying the operator coefficients to better match the distribution at higher roughness values. It may also be interesting to consider absorption and/or transmission via breaking the unitary assumption. Finally, one may be interested to model non-Gaussian devia-

tions, though non-smooth distributions may be difficult to match via a spectral method.

## Acknowledgements

PEC would like to thank Annika Lang for thoughtful comments, Samuel C. Barden for useful conversations and Peter Shirley for useful feedback on an early version of this article.

## A Re-summation of Zernike polynomials

Let us note that the radial component of the Zernike polynomial  $R_n^m(r)$ , with  $n = m + 2k$ , can be written

$$R_{m+2k}^m(r) = \sum_{p=0}^k \binom{k}{p} \frac{(-1)^{k-p}}{k!} r^p \frac{d^k}{dr^k} r^{m+k+p}. \quad (16)$$

As such  $m$  only appears in the right-most term in the geometric form  $r^m$  (albeit with combinatorial terms once differentiation has been applied).

Using the above we may write geometric sums of radial Zernike terms with another term  $z^m$  as

$$\sum_{m=0}^{\infty} \frac{2m+4k+2}{\epsilon_m} R_{m+2k}^m(x) R_{m+2k}^m(y) z^m = \sum_{l,p=0}^k \binom{k}{l} \binom{k}{p} \frac{(-1)^{l+p}}{(k!)^2} x^l y^p \frac{\partial^k}{\partial y^k} \frac{\partial^k}{\partial x^k} x^{k+l} y^{k+p} \left[ \frac{(2k+1)(1-\tau^2) + 2\tau}{(1-\tau)^2} \right]_{\tau=xyz}, \quad (17)$$

where we have used a geometric sum

$$\sum_{m=0}^{\infty} \frac{2(m+2k+1)}{\epsilon_m} z^m = \frac{(1-z^2)(2k+1) + 2z}{(z-1)^2}. \quad (18)$$

The RHS of Eqn. (17) is explicitly symmetric under permutation of  $x$  and  $y$ , however, by binomial expansion of the derivatives and using a re-summation, one

can find the double-sum identity

$$\begin{aligned} & \sum_{l,p=0}^k \binom{k}{l} \binom{k}{p} \frac{(-1)^{l+p}}{(k!)^2} x^l y^p \frac{\partial^k}{\partial y^k} \frac{\partial^k}{\partial x^k} \\ & x^{k+l} y^{k+p} \left[ \frac{\tau^n}{(1-\tau)^2} \right]_{\tau=xyz} \\ &= \sum_{\alpha,\beta=0}^k (1+\alpha+\beta) \frac{\binom{k}{\alpha} \binom{k}{\beta}}{\binom{k+\alpha+\beta+n}{k+n}} \times \\ & \left[ \frac{\tau^{n+\alpha+\beta}}{(1-\tau)^{2+\alpha+\beta}} \right]_{\tau=xyz} \times \\ & P_k^{(n+\alpha,-\alpha)}(1-2x^2) P_k^{(n+\alpha+\beta,-\beta)}(1-2y^2), \quad (19) \end{aligned}$$

(where the RHS lacks explicit symmetry), where we have used an expression for the standard Jacobi polynomial

$$P_k^{(\alpha+\beta,-\beta)}(1-2y^2) = \frac{(\alpha+\beta+k)!}{k!(\alpha+k)!} \sum_{p=0}^k \binom{k}{p} \times \frac{(\alpha+k+p)!}{(\alpha+\beta+p)!} (-y^2)^p. \quad (20)$$

We may thus combine (17) and (19) to write

$$\sum_{m=0}^{\infty} \frac{2m+4k+2}{\epsilon_m} R_{m+2k}^m(r) R_{m+2k}^m(s) z^m = \Omega_k(1-2r^2, 1-2s^2, rsz) \quad (21)$$

with  $\Omega_n(\mu, \nu, z)$  defined as in Eqn. (10).

Finally we note that re-writing the eigenvalue of  $-\mathcal{L}$  in Eqn. (7) as

$$\lambda_{m+2k}^m = 4k(k+1) + 2m(2k+1), \quad (22)$$

is linear in  $m$  (at fixed  $k$ ), we may write sums over the normalised Zernike polynomials with an exponential in the eigenvalue of  $\mathcal{L}$  as

$$\begin{aligned} & \sum_{n=0}^{\infty} \sum_{m=-n}^m \frac{2n+2}{\pi \epsilon_m} Z_n^m(\mathbf{r}) Z_n^m(\mathbf{s}) e^{-\lambda_n^m t} = \\ & \frac{1}{\pi} \sum_{n=0}^{\infty} e^{-4n(n+1)t} \operatorname{Re} [ \\ & \Omega_n(1-2r^2, 1-2s^2, rse^{i(\phi_s-\phi_r)-2(2n+1)t}) ], \quad (23) \end{aligned}$$

leading directly to Eqn. (9).

## B Jacobi polynomial recurrences for evaluating $\Omega_n$

Evaluation of  $\Omega_n(\mu, \nu, z)$  defined in Eqn. (10) requires a double summation over Jacobi polynomials. Whilst this may be performed using standard open-source packages (e.g. `scipy.special.eval_jacobi` [19]), for numerical purposes it is attractive to efficiently evaluate subsequent terms from earlier ones using the some of the many recurrence relations that exist for orthogonal polynomials of integer order. An open-source implementation of these algorithms is provided at <https://github.com/pec27/urdf>. As with any numerical iteration of recurrence relations we expect accumulation of errors due to finite precision, and these have only been tested up to  $n = 20$ , i.e. around 1,300 summation terms.

Let us begin by noting that all the  $P_n^{(\alpha, \beta)}$  in Eqn. (10) have  $\alpha, \beta, n \in \mathbb{Z}$ , with  $n$  non-negative,  $\beta \in \{-n, 1-n, \dots, 0\}$  and  $\alpha \geq -\beta \geq 0$ . As such we may start with the ‘initial’ value for  $\beta = -n$ ,

$$P_n^{(\alpha, -n)}(z) = \binom{n+\alpha}{n} \left( \frac{z+1}{2} \right)^n. \quad (24)$$

All the other terms we need can be found by some combination of two operations, (i) lowering  $\alpha$  but keeping  $\beta$  fixed, (ii) lowering  $\alpha$  and raising  $\beta$ .

These relationships for Jacobi polynomials are found via the contiguous relations for the Gauss hypergeometric function (see e.g. [1, 15.2]). Using the six contiguous relations for  $F(a, b, c; z)$  with  $\{a, b, c\}$  raised/lowered,

$$z \frac{dF}{dz} = z \frac{ab}{c} F(a+, b+, c+) \quad (25)$$

$$= a(F(a+) - F) \quad (26)$$

$$= b(F(b+) - F) \quad (27)$$

$$= (c-1)(F(c-) - F) \quad (28)$$

$$= \frac{(c-a)F(a-) + (a-c+bz)F}{1-z} \quad (29)$$

$$= \frac{(c-b)F(b-) + (b-c+az)F}{1-z} \quad (30)$$

$$= z \frac{(c-a)(c-b)F(c+) + c(a+b-c)F}{c(1-z)} \quad (30)$$

which transforms directly into Jacobi polynomials as

$$(z-1) \frac{d}{dz} P_n^{(\alpha, \beta)}(z) = \frac{1}{2}(z-1)(1+\alpha+\beta+n)P_{n-}^{(\alpha+, \beta+)} \quad (31)$$

$$= nP - (\alpha+n)P_{n-}^{(\beta+)} \quad (32)$$

$$= (1+\alpha+\beta+n)(P^{(\beta+)} - P) \quad (33)$$

$$= (\alpha+n)P^{(\alpha-, \beta+)} - \alpha P \quad (34)$$

$$= \frac{2(n+1)P_{n+}^{(\beta-)}}{1+z} - \left( 1+\alpha+\beta+n - \frac{2\beta}{1+z} \right) P \quad (35)$$

$$= \frac{(2\beta+n+nz)P - 2(\beta+n)P^{(\beta-)}}{1+z} \quad (36)$$

$$= \frac{1-z}{1+z} \left( \beta P - (\beta+n)P^{(\alpha+, \beta-)} \right). \quad (37)$$

where e.g.  $\alpha+$  denotes  $\alpha+1$ , and omitted indices denote their original values.

In particular we may combine Eqns. (36) & (37) to find

$$(1+z)P_n^{(\alpha, \beta)} = 2P_n^{(\alpha, \beta-1)} - (1-z)P_n^{(\alpha+1, \beta-1)} \quad (38)$$

and combine Eqns. (33) & (34) to find

$$P_n^{(\alpha-1, \beta)} = \frac{\alpha+\beta+n}{\alpha+n} P_n^{(\alpha, \beta)} - \frac{\beta+n}{\alpha+n} P_n^{(\alpha, \beta-1)}. \quad (39)$$

In combination these allow the inference of  $\{P_n^{(\alpha-1, \beta+1)}, P_n^{(\alpha, \beta+1)}, \dots, P_n^{(n+\alpha-1, \beta+1)}\}$  from  $\{P_n^{(\alpha, \beta)}, P_n^{(\alpha+1, \beta)}, \dots, P_n^{(n+\alpha, \beta)}\}$ , i.e. operation (ii).

For operation (i) we may combine Eqns. (33), (34), (36) & (37) to find the following relation to decrement  $\alpha$  at fixed  $\beta$ ,

$$P_n^{(\alpha-1, \beta)} = \left( \frac{\alpha}{\alpha+n} + \frac{1-z}{2} \frac{2n+1+\alpha+\beta}{\alpha+n} \right) P_n^{(\alpha, \beta)}(z) - \frac{1-z}{2} \frac{1+\alpha+\beta+n}{\alpha+n} P_n^{(\alpha+1, \beta)}(z). \quad (40)$$

## References

- [1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover Publications, 1965.
- [2] Petr Beckmann and Andre Spizzichino. The scattering of electromagnetic waves from rough surfaces. *Norwood, MA, Artech House, Inc., 1987, 511 p.*, 1, 1987.
- [3] Brent Burley. Physically based shading at disney. In *ACM SIGGRAPH*, volume 2012, pages 1–7. vol. 2012, 2012.
- [4] R. L. Cook and K. E. Torrance. A reflectance model for computer graphics. *ACM Trans. Graph.*, 1(1):7–24, January 1982.
- [5] G. Fasano and A. Franceschini. A multidimensional version of the Kolmogorov-Smirnov test. *MNRAS*, 225:155–170, March 1987.
- [6] Crispin Gardiner. *Stochastic methods*, volume 4. Springer Berlin, 2009.
- [7] B. Hapke. Bidirectional reflectance spectroscopy 3. Correction for macroscopic roughness. *Icarus*, 59(1):41–59, July 1984.
- [8] Xiao D. He, Kenneth E. Torrance, François X. Sillion, and Donald P. Greenberg. A comprehensive physical model for light reflection. *SIGGRAPH Comput. Graph.*, 25(4):175–186, July 1991.
- [9] M. Minnaert. The reciprocity principle in lunar photometry. *The Astrophysical Journal*, 93:403–410, May 1941.
- [10] George S. Pogosyan, Cristina Salto-Alegre, Kurt Bernardo Wolf, and Alexander Yakhno. Quantum superintegrable Zernike system. *Journal of Mathematical Physics*, 58(7):072101, July 2017.
- [11] Rudolph W Preisendorfer and Curtis D Mobley. Unpolarized irradiance reflectances and glitter patterns of random capillary waves on lakes and seas, by Monte Carlo simulation. NOAA Technical Memorandum ERL PMEL-63, 1985.
- [12] Ingmar G. E. Renhorn and Glenn D. Boreman. Analytical fitting model for rough-surface BRDF. *Opt. Express*, 16(17):12892–12898, Aug 2008.
- [13] Jean-Louis Roujean, Marc Leroy, and Pierre-Yves Deschamps. A bidirectional reflectance model of the earth’s surface for the correction of remote sensing data. *Journal of Geophysical Research: Atmospheres*, 97(D18):20455–20468, 1992.
- [14] Crystal B Schaaf, Feng Gao, Alan H Strahler, Wolfgang Lucht, Xiaowen Li, Trevor Tsang, Nicholas C Strugnell, Xiaoyang Zhang, Yufang Jin, Jan-Peter Muller, et al. First operational BRDF, albedo nadir reflectance products from MODIS. *Remote sensing of Environment*, 83(1-2):135–148, 2002.
- [15] G. Schaepman-Strub, M.E. Schaepman, T.H. Painter, S. Dangel, and J.V. Martonchik. Reflectance quantities in optical remote sensing—definitions and case studies. *Remote Sensing of Environment*, 103(1):27–42, 2006.
- [16] B. Smith. Geometrical shadowing of a random rough surface. *IEEE Transactions on Antennas and Propagation*, 15(5):668–671, 1967.
- [17] Yinlong Sun. Statistical ray method for deriving reflection models of rough surfaces. *J. Opt. Soc. Am. A*, 24(3):724–744, Mar 2007.
- [18] Kenneth E Torrance and Ephraim M Sparrow. Theory for off-specular reflection from roughened surfaces. *JOSA*, 57(9):1105–1112, 1967.
- [19] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng,

Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

- [20] Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. Microfacet Models for Refraction through Rough Surfaces. In Jan Kautz and Sumanta Pattanaik, editors, *Rendering Techniques*. vol. 2007, The Eurographics Association, 01 2007.