Supplemental Material for Gradient-Domain Path Tracing

Markus Kettunen[1] Marco Manzi[2] Miika Aittala[1] Jaakko Lehtinen[1,3] Frédo Durand[4] Matthias Zwicker[2]
[1]Aalto University [2]University of Bern [3]NVIDIA [4] MIT CSAIL

Abstract

This document supplements the SIGGRAPH 2015 technical paper "Gradient-domain Path Tracing". We provide additional details and steps in the derivation of a frequency analysis of gradient sampling and reconstruction for Monte Carlo integration, including a numerical experiment that we performed to validate the theory, under its simplifications. We also present full derivations of the Jacobian determinants for our the shift mapping.

1 Frequency Analysis Derivations

Here we provide some additional detail and derivation steps for the frequency analysis of gradient-domain Monte Carlo rendering presented in the paper. We follow the same outline as in the paper.

1.1 Frequency Analysis of Gradient Estimation

We start by expressing the path difference function as a 1D convolution. For our analysis we use the simple shift mapping $T_{ij}(\bar{x}) = (x + j - i, \bar{p})$ that moves the path from pixel i to j without changing the other path parameters. We only compute differences between neighboring pixels in 1D, so we only have one mapping $T(\bar{x}) = (x-1, \bar{p})$. Clearly, the Jacobian of this mapping is identity and it has a unit determinant. The difference between pixel i and the one next to it is given by integrating the path difference function $g(x, \bar{p}) = f(x, \bar{p}) - f(x - 1, \bar{p})$,

$$\Delta_i(x_i) = \left(h(x) * \int_\Omega f(x, \bar{p}) - f(x - 1, \bar{p}) \mathrm{d}\mu(\bar{p}) \mathrm{d}x \right) (x_i),$$

where Ω denotes the space of non-image path coordinates. Using a path space difference operator $d(\bar{x}) = \delta(x) - \delta(x - 1)$, where $\delta(x)$ is the Dirac delta distribution, we write this as a convolution

$$f(x, \bar{p}) - f(x - 1, \bar{p}) = (d * f)(x, \bar{p}) = g(x, \bar{p}). \quad (1)$$

In addition, instead of evaluating gradients only at discrete pixels i, we can write a continuous gradient function $\Delta(x)$ as

$$\Delta(x) = \left(h * \int_{\Omega_{\bar{p}}} g(x, \bar{p}) \, \mathrm{d}\mu(\bar{p}) \right) (x). \quad (2)$$

We now turn to spectral analysis. Denote the Fourier transform $\mathcal{F}\{f\}$ of the image contribution function by $F(\omega_x, \omega_{\bar{p}})$, where ω_x are frequencies (ordinary, not angular) along the image axis, and $\omega_{\bar{p}}$ is a vector of frequencies along the remaining path coordinates. Similarly, $D(\omega_x, \omega_{\bar{p}})$ is the Fourier transform of the difference operator d,

$$D(\omega_x, \omega_{\bar{p}}) = 1 - e^{i2\pi\omega_x},$$

and $|D|^2$ is its power spectrum, that is, the square magnitude of the Fourier transform,

$$|D(\omega_x, \omega_{\bar{p}})|^2 = 2 - 2\cos(2\pi\omega_x).$$

According to the standard convolution-multiplication relationship, the frequency representation of $g = d * f$ (Equation 1) is

$$G(\omega_x, \omega_{\bar{p}}) = DF = (1 - e^{i2\pi\omega_x}) F(\omega_x, \omega_{\bar{p}}), \quad (3)$$

and its power spectrum is

$$|G(\omega_x, \omega_{\bar{p}})|^2 = |DF|^2 = (2 - 2\cos(2\pi\omega_x)) |F(\omega_x, \omega_{\bar{p}})|^2.$$

showing that the difference operator scales down amplitudes of low frequencies, and it multiplies the square amplitudes of waves at the Nyquist frequency $\omega_x = 1/2$ of the pixel grid (unit spacing between pixels) by a factor of four.

Sampling Gradients and Pixels Since we are sampling stochastically we are interested in the expected values of sampled signals, as opposed to the result of sampling with one concrete instance of the stochastic process. Our analysis of stochastic sampling follows the work of Dippé and Wold [1985], who state the assumption that both the sampling process and the image contribution function are wide-sense stationary (WSS) processes. We show below, however, that it is only necessary to make this assumption for the sampling process, and not for the image contribution function. Wide-sense stationarity means that the first moment and the autocovariance of the process are constant over space, and this assumption simplifies the analysis considerably. Finally, we restrict our analysis to uniform random sampling, although the analysis can be extended to other stochastic sampling techniques [Dippé and Wold 1985].

As a key consequence, assuming stationarity of stochastic sampling implies that our sampling process extends over an infinite domain, and we cannot model the restriction of sampling to a finite region. Note that the means of our integrands $E[f(X)]$, where X is a random sample over an infinite domain, will be zero, and we will always overestimate sample variance compared to practical approaches. Considering the sample variance defined as

$$\mathrm{Var}[f(X)] = E[f(X)^2] - E[f(X)]^2 < E[f(X)]^2, \quad (4)$$

we see that if X is drawn from a finite sampling domain, the square of the mean $E[f(X)]^2$ is strictly positive. In contrast, the variance estimate under our assumptions is $E[f(X)^2]$, which we also call the energy of f.

The frequency domain representation of a wide-sense stationary stochastic signal is given by its power spectral density (PSD), which is defined as the Fourier transform of its autocorrelation function [Miller and Childers 2012]. Intuitively, this can be understood as the expected power spectrum of the stochastic process. We denote a (multi-dimensional) stationary, uniform random sample process s, and assume the sampling density is n samples per unit area.

Modeling uniform random sampling as a Poisson process, Leneman [1966] derives the PSD $|S(\omega_x, \omega_{\bar{p}})|^2$ of s as

$$|S(\omega_x, \omega_{\bar{p}})|^2 = n + n^2\delta(\omega_x, \omega_{\bar{p}}), \tag{5}$$

that is, a constant n plus a scaled impulse at the origin. We now sample both the image contribution function and the path difference function with s. We start by analyzing the sampled image contribution function. In the frequency domain, sampling is the convolution of the Fourier transform of the sampling grid with the Fourier transform of the signal. Since we are sampling stochastically, we are interested in the *expected* square magnitude of the Fourier transform of the sampled signal, which is $E[|F * S|^2]$ where $*$ denotes convolution.

The crucial consequence of assuming a wide sense-stationary sampling process is that we can express $E[|F * S|^2]$ as the convolution $|F|^2 * |S|^2$ [Dippé and Wold 1985]. This can be derived in a few simple steps. We first reformulate to change the order of taking the expected value and the Fourier transform, which we can do because both are linear operators,

$$\begin{aligned} E[|F * S|^2] &= E\left[\mathcal{F}\{sf\} \cdot \mathcal{F}\{sf\}^*\right] \\ &= E[\mathcal{F}\{(sf) \star (sf)\}] \\ &= \mathcal{F}\left\{E[(sf) \star (sf)]\right\}, \end{aligned}$$

where $\mathcal{F}\{\cdot\}^*$ is the complex conjugate of the Fourier transform, $\star$ denotes correlation, and we used the convolution theorem multiple times. Now let us denote paths by vectors $\bar{x} = (x, \bar{p})$ and τ and explicitly write the auto-correlation as

$$\begin{aligned} E[(sf) \star (sf)] &= E\left[\int s(\bar{x})s(\bar{x}+\tau)f(\bar{x})f(\bar{x}+\tau)d\bar{x}\right] \\ &= \int E[s(\bar{x})s(\bar{x}+\tau)]f(\bar{x})f(\bar{x}+\tau)d\bar{x} \\ &= R_{ss}(\tau)\int f(\bar{x})f(\bar{x}+\tau)d\bar{x} \\ &= R_{ss}(\tau)(f \star f)(\tau), \end{aligned}$$

where the first step is the definition of correlation. The second step exploits linearity of integration and expectation, and the fact that terms involving f are deterministic values that can be pulled out of the expectation. The crucial step is the third step, where we exploit the key property that the auto-correlation function R_{ss} of the wide-sense stationary sampling process depends only on the offset τ and can be pulled out of the integral over $\bar{x}$, and finally we also abbreviate the auto-correlation of f using the symbol $\star$. Now we plug this back into the Fourier transform and again leverage the convolution theorem. Here we use the key fact that, by definition, the Fourier transform of the auto-correlation function of the WSS process yields its power spectral density, that is $\mathcal{F}\{R_{ss}\} = |S|^2$,

$$\begin{aligned} E[|F * S|^2] &= \mathcal{F}\{R_{ss} \cdot (f \star f)\} \\ &= |S|^2 * \mathcal{F}\{f \star f\} \\ &= |S|^2 * |F|^2. \end{aligned}$$

Using the PSD for the Poisson process from Equation Equation 5 we now obtain

$$\begin{aligned} &\left(|F|^2 * |S|^2\right)(\omega_x, \omega_{\bar{p}}) \\ &= n^2|F|^2(\omega_x, \omega_{\bar{p}}) + n\int |F|^2(\omega_x{}', \omega_{\bar{p}}{}')\mathrm{d}\omega_x{}'\mathrm{d}\omega_{\bar{p}}{}' \\ &= n^2|F|^2(\omega_x, \omega_{\bar{p}}) + n\|F\|^2. \end{aligned}$$

Observe that this is the power spectrum of the sampled signal in the combined $(x, \bar{p})$ path space, before integration. It consists of a scaled version of the power spectrum of the original signal, plus an error term proportional to the total energy $\|F\|^2$ of the signal. This is analogous to the analysis by Dippé and Wold [1985], with the difference that they investigated sampling in the image plane only, whereas we sample in the higher dimensional space. In contrast to their assumption about f being wide-sense stationary, we also showed here that this is not necessary.

Integration over Paths and Pixel Filter Integration over $\bar{p}$ corresponds to slicing at $\omega_{\bar{p}} = 0$ (a vector of all zeros) in the frequency domain. In addition, to normalize the integral we need to divide each sample by one over the sample density, that is $1/n$. This means the PSD of the sampling grid is divided by $1/n^2$. Finally, to reconstruct the output image we convolve with the pixel filter $h(x)$ in the spatial domain, which is a multiplication with the Fourier transform $H(\omega_x)$ in the frequency domain. Hence the power spectrum of the image obtained using Monte Carlo integration is

$$\begin{aligned} &\frac{1}{n^2}\left(|F|^2 * |S|^2\right)(\omega_x, 0)|H(\omega_x)|^2 \\ &\qquad = |F(\omega_x, 0)|^2|H(\omega_x)|^2 + \frac{1}{n}\|F\|^2|H(\omega_x)|^2. \end{aligned}$$

The first term here is the power spectrum of the ground truth image. This reveals the mean squared error (MSE) $|\epsilon_F(\omega_x)|^2$ (slight abuse of notation, expected value is implied) of sampling, integration, and pixel filtering as the second term

$$|\epsilon_F(\omega_x)|^2 = \frac{1}{n}\|F\|^2|H(\omega_x)|^2, \tag{6}$$

which is a function of spatial frequency ω_x, and equivalent to the variance of the whole process. Assuming an ideal pixel filter, the MSE is simply a constant $\|F\|^2/n$ for frequencies $|\omega_x| < 1/2$ below the Nyquist limit of the output pixel grid. That is, the MSE is white noise with square magnitude proportional to the signal energy $\|F\|^2$.

Analysis of Gradients The power spectrum of the path difference function (Equation 3) $|G(\omega_x, \omega_{\bar{p}})|^2$ is

$$|G(\omega_x, \omega_{\bar{p}})|^2 = (2 - 2\cos(2\pi\omega_x))|F(\omega_x, \omega_{\bar{p}})|^2,$$

and we denote its total energy analogous to above as

$$\|G\|^2 = \int (2 - 2\cos(2\pi\omega_x{}'))|F(\omega_x{}', \omega_{\bar{p}}{}')|^2\mathrm{d}\omega_x{}'\mathrm{d}\omega_{\bar{p}}{}'. \tag{7}$$

Finally, the MSE of the Monte Carlo estimate of the gradients $|\epsilon_G(\omega_x)|^2$ is similar as above,

$$|\epsilon_G(\omega_x)|^2 = \frac{1}{n}\|G\|^2|H(\omega_x)|^2. \tag{8}$$

This leads to a first simple insight, which is that the ratio of the MSEs of conventional versus gradient sampling is simply the ratio of the total energies,

$$\boxed{\frac{|\epsilon_G(\omega_x)|^2}{|\epsilon_F(\omega_x)|^2} = \frac{\|G\|^2}{\|F\|^2},} \tag{9}$$

for $|\omega_x| < 1/2$. This ratio is typically significantly smaller than one, because the factor $(2 - 2\cos(2\pi\omega_x{}'))$ in Equation 7 strongly attenuates low frequency energy.

1.2 Poisson Reconstruction

We next analyze the reconstruction of a final image based on sampled gradients and a "coarse" image sampled conventionally. As Lehtinen et al. [2013] mention, this Poisson reconstruction problem can be expressed in the frequency domain. Here, we extend their analysis by explicitly deriving the MSE of the Poisson reconstruction using the MSEs of the sampled image and its gradients derived above.

Let us express the result of sampling, integration, and pixel filtering as the ground truth image plus an error $\epsilon_F(\omega_x)$,

$$(F * S)(\omega_x, 0)H(\omega_x) = F(\omega_x, 0)H(\omega_x) + \epsilon_F(\omega_x),$$

and similarly for the gradients with an error $\epsilon_G(\omega_x)$,

$$(DF * S)(\omega_x, 0)H(\omega_x) = D(\omega_x, 0)F(\omega_x, 0)H(\omega_x) + \epsilon_G(\omega_x).$$

In the following, all frequency domain functions are interpreted as their restriction to the slice $\omega_{\bar{p}} = 0$. Hence F denotes the Fourier transform of the ground truth image, and DF the Fourier transform of the ground truth image gradients. Poisson reconstruction R_α from the sampled data is then [Lehtinen et al. 2013]

$$\begin{aligned} R_\alpha &= \frac{\alpha^2(FH + \epsilon_F) + D^*(DFH + \epsilon_G)}{\alpha^2 + |D|^2} \\ &= FH + \frac{\alpha^2\epsilon_F + D^*\epsilon_G}{\alpha^2 + |D|^2} \\ &= FH + \epsilon_{R_\alpha}, \end{aligned} \tag{10}$$

where we omitted the frequency parameter ω_x for simplicity, and we introduced the term ϵ_{R_α} to denote the error of the reconstruction. The variance (that is, the MSE) of the reconstruction (at a given frequency ω_x, omitted again) is

$$\begin{aligned} |\epsilon_{R_\alpha}|^2 = \mathrm{Var}[\epsilon_{R_\alpha}] &= \mathrm{Var}\left[\frac{\alpha^2\epsilon_F + D^*\epsilon_G}{\alpha^2 + |D|^2}\right] \\ &= \frac{\alpha^4|\epsilon_F|^2 + |D|^2|\epsilon_G|^2}{(\alpha^2 + |D|^2)^2}, \end{aligned} \tag{11}$$

where we assumed the errors ϵ_F and ϵ_G are uncorrelated and their expected values are zero, and the notation $|\epsilon_F|^2$ and $|\epsilon_G|^2$ represents their variance. In our scenario, the variance of the image $|\epsilon_F|^2$ and its gradients $|\epsilon_G|^2$ are now given by Equation 6 and Equation 8.

Minimizing the MSE $|\epsilon_{R_\alpha}|^2$ over α^2 yields the optimal value α^2_*, which is

$$\boxed{\alpha^2_*(\omega_x) = \frac{|\epsilon_G(\omega_x)|^2}{|\epsilon_F(\omega_x)|^2} = \frac{\|G\|^2}{\|F\|^2}.} \tag{12}$$

We include the frequency parameter ω_x to emphasize that in general the optimal $\alpha^2_*(\omega_x)$ depends on the frequency, but in our scenario this happens to be a constant. Using the optimal α^2_*, the MSE reduces to

$$|\epsilon_{R_{\alpha_*}}|^2 = \frac{|\epsilon_G|^2|\epsilon_F|^2}{|\epsilon_F|^2|D|^2 + |\epsilon_G|^2}. \tag{13}$$

With our definition of the variances $|\epsilon_F|^2$ and $|\epsilon_G|^2$ from Equation 6 and Equation 8 we write this as

$$\boxed{\frac{|\epsilon_F(\omega_x)|^2}{|\epsilon_{R_{\alpha_*}}(\omega_x)|^2} = \frac{\|G\|^2/\|F\|^2 + 2 - 2\cos(2\pi\omega_x)}{\|G\|^2/\|F\|^2},} \tag{14}$$

for $|\omega_x| < 1/2$. This expresses the factor, at each frequency, by which gradient sampling and reconstruction reduces the MSE compared to conventional sampling.

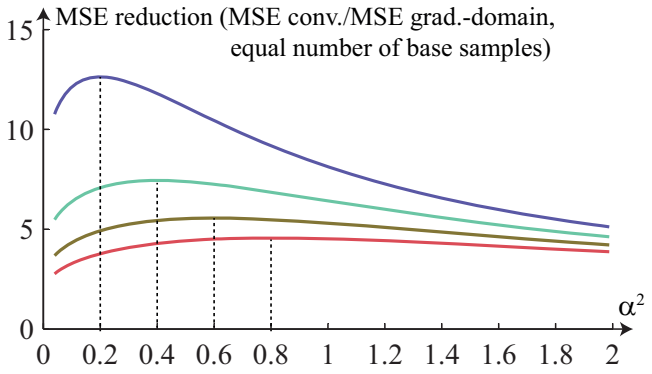

Figure 1: *Error reduction of gradient sampling and reconstruction for 2D images: We plot the factor by which gradient sampling and reconstruction reduces the total MSE compared to conventional sampling and integration at equal sample count. We show different energy ratios $\|G\|^2/\|F\|^2 \in \{0.05, 0.1, \ldots 0.5\}$, corresponding to the curves from top to bottom (purple to red). Note that each curve achieves its maximum at $\alpha^2 = \|G\|^2/\|F\|^2$.*

Poisson Reconstruction for 2D Images It is straightforward to generalize the error analysis of Poisson reconstruction to 2D images, where we have two distinct gradient operators along the horizontal and vertical image axes. Similar to Equation 11, for the MSE of the reconstruction we obtain

$$|\epsilon_{R_\alpha}|^2 = \frac{\alpha^4|\epsilon_F|^2 + |D^x|^2|\epsilon^x_G|^2 + |D^y|^2|\epsilon^y_G|^2}{(\alpha^2 + |D^x|^2 + |D^y|^2)^2}, \tag{15}$$

where $|D^x|^2$ and $|D^y|^2$ are the power spectra of the horizontal and vertical gradient operators, and $|\epsilon^x_G|^2$ and $|\epsilon^y_G|^2$ are the MSEs of the sampled horizontal and vertical gradients. Let us assume that the total energy in both gradients is the same, that is $\|G^x\|^2 = \|G^y\|^2$ and both are defined analogous to Equation 7. We denote this again $\|G\|^2$. Then we still have $\alpha^2_* = \|G\|^2/\|F\|^2$ as before.

Finally, let us analyze the total (as opposed to per frequency) MSE of the reconstructed image by integrating the reconstruction error in Equation 15 over all frequencies $|\omega_{x0}|, |\omega_{x1}| < 1/2$ below the Nyquist limit of the output pixel grid. For this purpose, we slightly rewrite the equation as

$$\frac{|\epsilon_F|^2}{|\epsilon_{R_\alpha}|^2} = \frac{(\alpha^2 + |D^x|^2 + |D^y|^2)^2}{\alpha^4 + (|D^x|^2 + |D^y|^2)\|G\|^2/\|F\|^2}, \tag{16}$$

where we exploited that $|\epsilon^x_G|^2/|\epsilon_F|^2 = |\epsilon^y_G|^2/|\epsilon_F|^2 = \|G\|^2/\|F\|^2$ (Equation 9). Assuming an ideal pixel filter the MSE of conventional sampling $|\epsilon_F|^2$ is a constant for $|\omega_x|, |\omega_y| < 1/2$ (Equation 6). Hence the integral of the term on the right yields the factor by which gradient sampling and reconstruction reduces the total MSE of conventional sampling. In Figure 1 we plot this factor for various ratios $\|G\|^2/\|F\|^2 \in \{0.05, 0.1, \ldots 0.5\}$ over values $\alpha \in (0, 1)$. The plots show an equal sample count comparison, meaning that we devide the integral of the term above by three to account for the fact that gradient-domain rendering in 2D uses N conventional samples and N samples for horizontal and vertical gradients each.

Bottom Line In summary, we conclude that under our simplifications we can quantify the MSE reduction of gradient sampling and reconstruction based on the ratio of the total energy in the path difference versus the image contribution function, $\|G\|^2/\|F\|^2$. Since the gradient operator attenuates low frequencies, and the signal

power is typically dominated by low frequencies, this ratio tends to be small. This is the fundamental reason behind the effectiveness of gradient-domain rendering.

1.3 Empirical Validation

In this section we provide an empirical validation of the theory presented so far in a 2D scenario. Here, "paths" are given as $\bar{x} = (x, p)$, where x is the 1D image coordinate, and p is an additional 1D integration domain. The image contribution function $f(x, p)$ is a 2D function that we specify explicitly. Hence we can easily compute its total energy $\|F\|^2$, and the energy of its gradients $\|G\|^2$, which is all we need to predict the optimal value α_*^2 (Equation 12), the error of the Poisson reconstruction at each frequency (Equation 14), and the reduction of the total MSE compared to conventional sampling as a function of α^2.

In Figure 2 we show an example image contribution function that approximates ambient occlusion on a planar receiver parallel to the image plane. Each black line represents an occluder at a certain distance to the receiver, where occluders at infinity are horizontal lines, and occluders close to the receiver are nearly vertical lines. In this particular scenario, $\|F\|^2 = 3.4 \times 10^2$ and $\|G\|^2 = 0.5 \times 10^2$. We uniformly distribute samples over a large 2D region to approximate a WSS sampling process over an infinite domain. In the middle we plot the predicted MSE $|\epsilon_F(\omega_x)|^2$ (Equation 6) of conventional Monte Carlo integration and the predicted error from gradient sampling and reconstruction $|\epsilon_{R_{\alpha_*}}(\omega_x)|^2$ (Equation 14) in black. The frequency range we plot is restricted to the Nyquist limit $|\omega_x| < 0.5$ of the output pixel grid. We also plot the corresponding empirically measured MSEs (blue and red) over 1000 runs of the experiment. At the bottom we plot the reduction of the total MSE compared to conventional sampling as a function of α^2, both from our empirical data (red) and the prediction by the theory (black). The theoretical prediction was obtained by integrating the 1D version of Equation 16. We observe that the theoretical predictions match the empirical data very well, where the mismatch can be explained by the variance in the empirical statistics that we gathered.

We evaluate the benefits and limitations of gradient sampling and reconstruction in a more realistic scenario in Figure 3. Here we render direct illumination from a planar, square area light source onto a planar receiver. Between we place an occluder plane with an infinite grid of circular holes, with a fixed ratio $1/4$ of hole size over hole spacing. Light, occluder, and receiver planes are all parallel. We sample points on the light source randomly with a uniform distribution. By scaling the grid of holes we generate occluders of various complexity. We present results of three different occluders of increasing complexity from top to bottom.

The left column in Figure 3 shows the result of conventional rendering with 64 samples per pixel (spp). The next column contains the result of gradient sampling and reconstruction at 64 spp for the coarse image, and additionally 64 spp for horizontal and vertical gradients. Here we reuse the samples to compute the coarse image and the base paths of the gradients. That is we compute three times as many samples as in conventional rendering.

We further show a reference image for comparison at $6400spp$. Then we illustrate the light source visibility from the central pixel in the images, where red dots indicate visible samples, and black dots occluded samples on the light. The figures highlight how the occlusion pattern becomes more complex from top to bottom. The rightmost column shows the MSE reduction factor of gradient sampling and reconstruction compared to conventional rendering as a function of α^2. Note how from top to bottom the ratio $\|G\|^2/\|F\|^2$ increases, and the amount of error reduction we obtain with optimal α decreases from approximately 15 to 5 and 3. Hence, in the last row the benefits and additional cost of gradient sampling and reconstruction approximately balance each other (three times MSE reduction for three times as many samples).

The MSE reduction plots here are very similar to the theoretical predictions in Figure 1, but they do not exactly match because we reuse samples to compute the coarse image and the base paths of the gradients. This reduces the effectiveness of gradient sampling and reconstruction, but the reduced sampling costs outweigh the slight quality degradation in practice. This also explains why the empirically best value α^2 in the plots does not exactly match $\|G\|^2/\|F\|^2$, although the discrepancy is insignificant in practice.

2 Derivation of Jacobians for G-PT

The Jacobian for the shift mapping used in G-PT can be computed by conceptually treating it as a concatenation of three steps: reparameterization of the path to coordinates where the shift map is trivial, a simple shift (like in Section 4 of the main paper), and a reverse parameterization change. Let us denote the sequence of path parameters of the base path in the original parameterization by $\bar{x}$ and in the new reparameterization by $\hat{x}$, and similarly for the shifted offset path by $\bar{y} = T(\bar{x})$ and $\hat{y}$, respectively. The Jacobian determinant of the shift is the product of the Jacobian determinants of these steps,

$$|T'| = \left|\frac{\partial \bar{y}}{\partial \bar{x}}\right| = \left|\frac{\partial \bar{y}}{\partial \hat{y}}\right| \left|\frac{\partial \hat{y}}{\partial \hat{x}}\right| \left|\frac{\partial \hat{x}}{\partial \bar{x}}\right|$$

The simple shift in the reparameterized space is just the addition of a constant to the intersection of the path with the image plane, hence its Jacobian is $|\partial \hat{y}/\partial \hat{x}| = 1$. For the other Jacobians, we interpret the reparameterizations as a concatenation of separate steps, where we start at the first path vertex (from the eye, excluding the eye itself), and apply the change of parameters according to the three distinct cases occurring in our shift (Section 5.2) step by step. The overall Jacobian determinant then is simply the product of the determinants for each step. After a case (iii) reconnects to the base path, all Jacobian determinants are one. We next describe the Jacobian determinants for the three cases.

The Jacobian for case (i) is unity, because we keep the image position as the path parameter under the reparameterization. The Jacobian for case (ii), that is, the change of parameterization from incident direction to projected half-vector, is more involved. We first derive the Jacobian $|\partial \omega_i / \partial \mathbf{h}|$ for reflection. We start by writing the incident direction ω_i as a function of the projected half-vector $\mathbf{h}$,

$$\omega_i(\mathbf{h}) = -\omega_o + 2\mathbf{h}(\omega_o \cdot \mathbf{h}), \tag{17}$$

where ω_o is the outgoing direction, considered constant.

Assume we want to compute the Jacobian determinant at a given half-vector $\mathbf{h}_0$. For our derivation we introduce an intermediate step, which is the projection of a vector $\mathbf{v}$ onto the plane perpendicular to $\mathbf{h}_0$, denoted $\mathbf{v}^{\perp \mathbf{h}_0}$, where $\mathbf{h}_0$ is considered a constant unit vector. We now have

$$\begin{aligned}\left|\frac{\partial \omega_i}{\partial \mathbf{h}}\right| &= \left|\frac{\partial \omega_i}{\partial \omega_i^{\perp \mathbf{h}_0}}\right| \left|\frac{\partial \omega_i^{\perp \mathbf{h}_0}}{\partial \mathbf{h}^{\perp \mathbf{h}_0}}\right| \left|\frac{\partial \mathbf{h}^{\perp \mathbf{h}_0}}{\partial \mathbf{h}}\right| \\ &= \frac{1}{\omega_i \cdot \mathbf{h}_0} \left|\frac{\partial \omega_i^{\perp \mathbf{h}_0}}{\partial \mathbf{h}^{\perp \mathbf{h}_0}}\right| (\mathbf{h}_0 \cdot \mathbf{h}),\end{aligned} \tag{18}$$

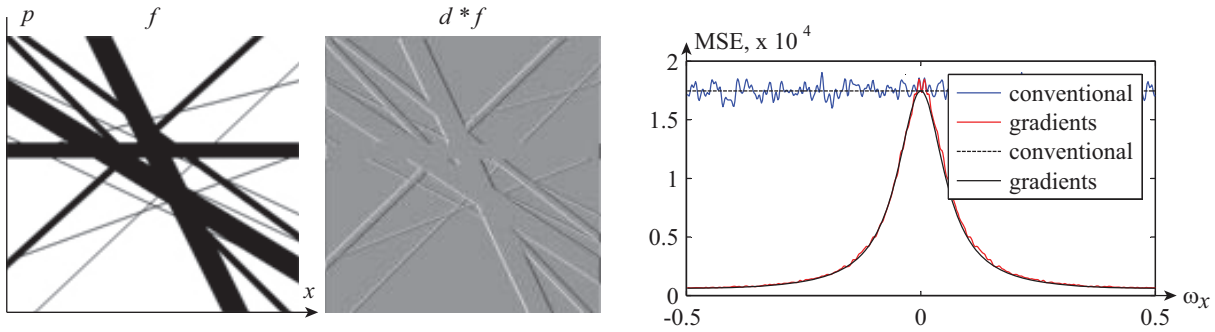

Figure 2: *An example "image contribution function" f and its finite difference $d * f$ in a 2D "path space" $\bar{x} = (x, p)$. Middle: we compare the predicted MSE in the frequency domain of conventional Monte Carlo integration (black dotted, Equation 6) and gradient domain Monte Carlo sampling and reconstruction (black, Equation 14). We compare with empirically acquired MSEs (conventional in blue, gradients in red) over* 1000 *runs of random sampling.*

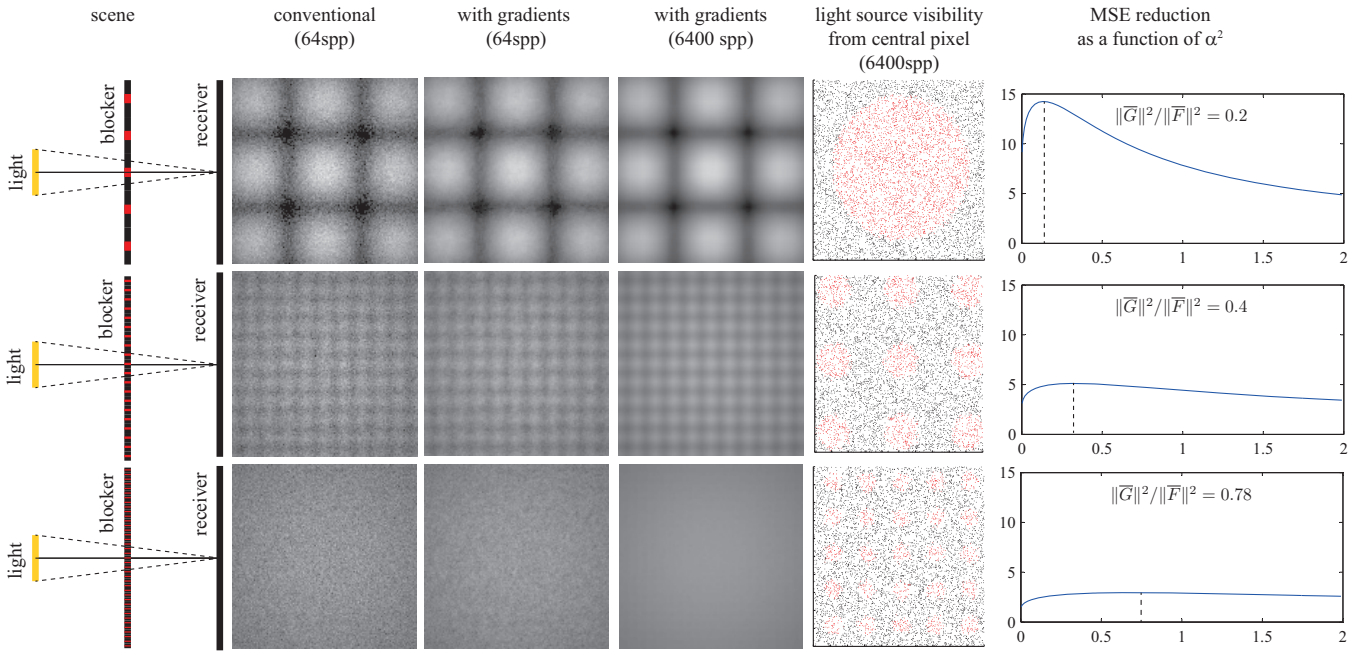

Figure 3: *We empirically investigate the effectiveness of gradient sampling and reconstruction on a simple scene with an area light source and a planar occluder with a rectangular grid of circular holes. As the grid becomes denser and the holes smaller, the energy in the gradients increases. We plot the empirically observed MSE reduction compared to conventional sampling on the right. Notice the close correspondence to the predicted error to the theoretical prediction in 1.*

where we used the fact that the Jacobian determinant of the projection of a vector onto a plane is the dot product with the unit normal. We derive the missing Jacobian by expressing $\mathbf{h}$ and ω_i locally in coordinates relative to $\mathbf{h}_0$, that is

$$\mathbf{h}(\epsilon_1, \epsilon_2) = \mathbf{h}_0 + \epsilon_1 \mathbf{e}_1 + \epsilon_2 \mathbf{e}_2,$$
$$\omega_i(\omega_{i1}, \omega_{i2}) = \mathbf{h}_0 + \omega_{i1} \mathbf{e}_1 + \omega_{i2} \mathbf{e}_2,$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are basis vectors on the plane perpendicular to $\mathbf{h}_0$. This means that by definition $\mathbf{h}^{\perp \mathbf{h}_0} = (\epsilon_1, \epsilon_2)$, and $\omega_i^{\perp \mathbf{h}_0} = (\omega_{i1}, \omega_{i2})$. Hence the missing Jacobian consists of the partial derivatives

$$\left| \frac{\partial \omega_i^{\perp \mathbf{h}_0}}{\partial \mathbf{h}^{\perp \mathbf{h}_0}} \right| = \begin{vmatrix} \frac{\partial \omega_{i1}}{\partial \epsilon_1} & \frac{\partial \omega_{i1}}{\partial \epsilon_2} \\ \frac{\partial \omega_{i2}}{\partial \epsilon_1} & \frac{\partial \omega_{i2}}{\partial \epsilon_2} \end{vmatrix}.$$

By plugging $\mathbf{h}(\epsilon_1, \epsilon_2)$ into Equation 17 and differentiating with respect to ϵ_1 and ϵ_2 we obtain

$$\frac{\partial \omega_{i1}}{\partial \epsilon_1} = \frac{\partial \omega_{i2}}{\partial \epsilon_2} = 2(\omega_o \cdot \mathbf{h}_0),$$
$$\frac{\partial \omega_{i1}}{\partial \epsilon_2} = \frac{\partial \omega_{i1}}{\partial \epsilon_2} = 0.$$

We can now remove the distinction between $\mathbf{h}$ and $\mathbf{h}_0$ and substitute this intermediate result back into Equation 18. We obtain

$$\left| \frac{\partial \omega_i}{\partial \mathbf{h}} \right| = 4\omega_o \cdot \mathbf{h}, \tag{19}$$

and this leads to the result used in Equation 11 of the main paper:

$$\left|\frac{\partial \omega_i^y}{\partial \omega_i^x}\right| = \left|\frac{\partial \omega_i^y}{\partial \mathbf{h}^y}\right| \left|\frac{\partial \mathbf{h}^x}{\partial \omega_i^x}\right| = \frac{\omega_o^y \cdot \mathbf{h}^y}{\omega_o^x \cdot \mathbf{h}^x}. \tag{20}$$

In the case of refraction the incident direction is given by Snell's law,

$$\omega_i(\mathbf{h}) = -\frac{n_2}{n_1}\omega_o + \mathbf{h}\left(\omega_o \cdot \mathbf{h} - \sqrt{1 - \frac{n_2^2}{n_1^2}(1 - (\omega_o \cdot \mathbf{h})^2)}\right), \tag{21}$$

where n_1 and n_2 are the indices of refraction. We follow the same steps as in the case of reflection, where the only difference is the Jacobian $\left|\partial \omega_i^{\perp \mathbf{h}_0} / \partial \mathbf{h}^{\perp \mathbf{h}_0}\right|$ in Equation 18. For refraction we obtain this by pluging $\mathbf{h}(\epsilon_1, \epsilon_2)$ into Snell's law in Equation 21 and differentiating with respect to ϵ_1 and ϵ_2. We obtain the elements of the Jacobian

$$\frac{\partial \omega_{i1}}{\partial \epsilon_1} = \frac{\partial \omega_{i2}}{\partial \epsilon_2} = \mathbf{h}_0 \cdot \omega_o - \sqrt{1 - \frac{n_2^2}{n_1^2}(1 - \omega_o \cdot \mathbf{h}_0)},$$
$$\frac{\partial \omega_{i1}}{\partial \epsilon_2} = \frac{\partial \omega_{i1}}{\partial \epsilon_2} = 0.$$

We can now remove the distinction between $\mathbf{h}$ and $\mathbf{h}_0$ and substitute this intermediate result back into Equation 18. We obtain

$$\left|\frac{\partial \omega_i}{\partial \mathbf{h}}\right| = \frac{\left[\mathbf{h} \cdot \omega_o - \sqrt{1 - \frac{n_2^2}{n_1^2}(1 - \omega_o \cdot \mathbf{h})}\right]^2}{\omega_i \cdot \mathbf{h}.}$$

By comparing this with Equation 21 we can write this simply as

$$\left|\frac{\partial \omega_i}{\partial \mathbf{h}}\right| = \frac{\left|\omega_i + \frac{n_2}{n_1}\omega_o\right|^2}{\omega_i \cdot \mathbf{h}}, \tag{22}$$

and this leads to the result used in Equation 12 of the main paper:

$$\left|\frac{\partial \omega_i^y}{\partial \omega_i^x}\right| = \left|\frac{\partial \omega_i^y}{\partial \mathbf{h}^y}\right| \left|\frac{\partial \mathbf{h}^x}{\partial \omega_i^x}\right| = \frac{\left|\omega_i^y + \frac{n_2^y}{n_1^y}\omega_o^y\right|^2}{\left|\omega_i^x + \frac{n_2^x}{n_1^x}\omega_o^x\right|^2} \frac{\omega_i^x \cdot \mathbf{h}^x}{\omega_i^y \cdot \mathbf{h}^y}. \tag{23}$$

References

DIPPÉ, M. A. Z., AND WOLD, E. H. 1985. Antialiasing through stochastic sampling. *SIGGRAPH Comput. Graph. 19*, 3 (July), 69–78.

LEHTINEN, J., KARRAS, T., LAINE, S., AITTALA, M., DURAND, F., AND AILA, T. 2013. Gradient-Domain Metropolis Light Transport. *ACM Trans. Graph. 32*, 4.

LENEMAN, O. A. 1966. Random sampling of random processes: Impulse processes. *Information and Control 9*, 4, 347 – 363.

MILLER, S. L., AND CHILDERS, D. G. 2012. *Probability and Random Processes*. Academic Press.