A linear heuristic for multiple importance sampling

Mateu Sbert[1*] and László Szirmay-Kalos[2]

*Correspondence:
mateu@ima.udg.edu

[1] Department of Informatics,
Applied Mathematics
and Statistics, University
of Girona, Campus de Montilivi,
Girona, Spain
[2] Department of Control
Engineering and Information
Technology, Budapest
University of Technology
and Economics, Muegyetem rkp.
3, 1111 Budapest, Hungary

Abstract

Multiple importance sampling combines the probability density functions of several sampling techniques into an importance function. The combination weights are the proportion of samples used for the particular techniques. This paper addresses the determination of the optimal combination weights from a few initial samples. Instead of the numerically unstable optimization of the variance, in our solution the quasi-optimal weights are obtained by solving a linear equation, which leads to simpler computations and more robust estimations. The proposed method is validated with 1D numerical examples and with the direct lighting problem of computer graphics.

Keywords: Multiple importance sampling, Monte Carlo, Kullback–Leibler divergence, BRDF and light source sampling

1 Introduction and previous work

Multiple Importance Sampling (MIS) [1, 2] combining several sampling techniques has been proven efficient in Monte Carlo integration. The weighting scheme of individual sampling techniques depends on the combined pdfs and also on the sample budgets, i.e., the number of samples generated with each of them.

This paper proposes a simple and efficient approach to automatically determine the sampling budgets of the combined methods based on the statistics of a few initial samples by solving a linear equation.

The organization of the paper is the following. In Sect. 2 the most relevant previous work is surveyed. Section 3 reviews the balance heuristic MIS. Section 4 presents the research problem of finding the optimum of weights and formulates it as the minimization of the divergence of the integrand and the mixture density. Section 5 presents our new heuristics resulting in a linear equation and the results are shown in Sect. 6. The paper is closed with discussion and conclusions.

2 Previous work

MIS has been originally proposed with allocating equal sample budgets for the combined techniques [1, 2]. In [3, 4] different equal sample number strategies were analysed. Adaptive budget allocation strategies have been examined in [5–10]. Sbert et al. [11] considered also the cost associated with the sampling strategies. Recently, a theoretical formula has been

elaborated for the weighting functions [12]. In [13] the balance heuristic estimator was generalized by decoupling the weights from the sampling rates, and implicit solutions for the optimal case were given.

These techniques offer lower variance and therefore theoretically outperform MIS with equal number of samples. However, equations determining the optimal weighting and the sample budget require the knowledge of the integrand. In computer graphics, for example, this integrand is not available in analytical form, so previous initial samples should be used for the approximation, which introduces errors in the computation.

It was pointed out in [14] that instead of the variance, the Kullback–Leibler divergence should be optimized since its estimator is more robust when just a few samples are available. Several authors have shown that the variance of an importance sampling estimator is equal to a Chi-square divergence [13, 15–17], which in turn can be approximated by the Kullback–Leibler divergence up to the second order [18]. In [19] this approach has been generalized to arbitrary divergences and the relevance of Tsallis-divergence was emphasized. The optimal sample budget has also been obtained by neural networks [17, 20].

In all these methods, the optimality criterion is expressed by nonlinear equations. The solution of these equations prohibit the lumping of the samples into a few variables or alternatively, several iterations using different samples need to be executed. If the sample number is not too high, either the number of iterations is limited or a single iteration can utilize very few samples.

In this paper we focus on simple equations and robust estimations, and show that quasi-optimal weights can be obtained directly from a linear equation where the few parameters to be stored are sums of integrand and pdf values.

3 Balance heuristic MIS

Monte Carlo methods estimate integral $\mu = \int f(x)\mathrm{d}x$ with random samples. Assume that we have m proposal pdfs $p_i(x)$ to generate the jth sample X_{ij} of method i in the domain of the integral. With method i, N_i samples are drawn, so the total number of samples is $\sum_{i=1}^{m} N_i = N$.

The mixture density of all sampling techniques is

$$p(\alpha, x) = \sum_{k=1}^{m} \alpha_k p_k(x). \tag{1}$$

where weight $\alpha_k = N_k/N$ is the fraction of the sample budgets allocated to method k and α is the vector of the weights of the combined methods.

The balance heuristics estimator [1, 2] of the integral is based on this mixture density:

$$\mu \approx F = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{N_i} \frac{f(X_{i,j})}{p(\alpha, X_{i,j})}. \tag{2}$$

Its variance, normalized to $N = 1$, is [13]

$$V[F] = \sum_{i=1}^{m} \alpha_i \sigma_i^2, \tag{3}$$

where

$$\sigma_i^2 = \int \frac{f^2(x)}{p^2(\alpha,x)} p_i(x)\mathrm{d}x - \mu_i^2, \tag{4}$$

and

$$\mu_i = \int \frac{f(x)}{p(\alpha,x)} p_i(x)\mathrm{d}x. \tag{5}$$

4 Optimization by minimizing the Kullback–Leibler divergence

The task is to find fractions α_i that provide the minimal integration error with the constraint that the sum of sample numbers must be equal to the total sample budget, i.e. $\sum_{k=1}^m \alpha_k = 1$.

As we are just estimating the integral, the integration error is not available. The variance of the estimator can be used instead, but it must be approximated from the available samples since neither can the variance be computed analytically. If the number of samples is small, then the variance estimator is numerically unstable and the variance estimates used in the optimization are unreliable. This uncertainty may significantly affect the goodness of the final results.

Importance sampling looks for a probability density $p(\alpha,x)$ that mimics $f(x)$ as much as possible. If we have non-negative integrand $f(x)$ with integral μ, then the integrand scaled down by the integral $g(x) = f(x)/\mu$ is also a pdf. The importance sampling problem can be stated as finding the mixture pdf $p(\alpha,x)$ that minimizes its divergence from $g(x)$ with the constraint that the sum of weights must be equal to 1.

Choosing the Kullback–Leibler divergence [18, 19]

$$KL(g,p) = \int g(x) \log\left(\frac{g(x)}{p(x)}\right)\mathrm{d}x \tag{6}$$

and using Lagrange multipliers for the constraint, the following function must be minimized:

$$KL\left(\frac{f(x)}{\mu}, p(\alpha,x)\right) + \lambda\left(\sum_k \alpha_k - 1\right). \tag{7}$$

Taking the partial derivatives with respect to α_i, we obtain that

$$\int \left(\frac{f(x)}{p(\alpha,x)}\right) p_i(x)\mathrm{d}x = \lambda\mu \tag{8}$$

are equal for each technique i when the solution is in the interior of the $(m-1)$-simplex, i.e., for positive weights. As the Kullback-Leibler divergence is convex in α, Eq. 8 guarantees a global minimum [19].

Forming the mixture of these equations, the value of Lagrange multiplier λ can be determined:

$$\lambda\mu = \sum_{i=1}^{m} \alpha_i \lambda \mu$$

$$= \sum_{i=1}^{m} \alpha_i \int \left(\frac{f(x)}{p(\alpha,x)} \right) p_i(x) dx$$

$$= \int \left(\frac{f(x)}{p(\alpha,x)} \right) \sum_{i=1}^{m} \alpha_i p_i(x) dx \qquad (9)$$

$$= \int \frac{f(x)}{p(\alpha,x)} p(\alpha,x) dx = \mu,$$

and thus $\lambda = 1$.

While solving this equation, we only have n_1 initial discrete samples $\{X_{1,1}, \ldots, X_{1,n_1}\}$ generated according to pdf $p_1(x)$, n_2 samples $\{X_{2,1}, \ldots, X_{2,n_2}\}$ generated with pdf $p_2(x)$, etc. Integrand values $f(X_{i,j})$ and pdf values $p_k(X_{i,j})$ are also evaluated for these discrete samples.

The integral of Eq. 8 can be given two different interpretations. On the one hand, it is the expected value of $f(x)/p(\alpha,x)$ when x takes random values distributed with $p_i(x)$:

$$\int \left(\frac{f(x)}{p(\alpha,x)} \right) p_i(x) dx = E_{p_i} \left[\frac{f(x)}{p(\alpha,x)} \right]$$

$$\approx \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{f(X_{i,j})}{p(\alpha,X_{i,j})}. \qquad (10)$$

On the other hand, the same integral expresses the expectation of $f(x)p_i(x)/p^2(\alpha,x)$ when x takes random values distributed with $p(\alpha,x)$:

$$\int \left(\frac{f(x)}{p(\alpha,x)} \right) p_i(x) dx = E_{p(\alpha)} \left[\frac{f(x)p_i(x)}{p^2(\alpha,x)} \right]$$

$$\approx \frac{1}{n} \sum_{k=1}^{m} \sum_{j=1}^{n_k} \frac{f(X_{k,j})p_i(X_{k,j})}{p^2(\alpha,X_{k,j})} \qquad (11)$$

where $n = \sum_{k=1}^{m} n_k$ is the total number of initial samples.

Both formulas are nonlinear functions of weights α_k, thus the optimization requires the solution of nonlinear equations.

One possibility is the application of the Newton–Raphson method [19] since the derivatives of Eqs. 10 and 11 with respect to α_k can be easily expressed. Such scheme decomposes the estimation of α_k into Newton–Raphson iterations each using a disjoint set of samples.

5 A linear heuristic

In order to find a simple solution for the quasi-optimal weights, we use an approximation in Eq. 10 to determine the optimum. As we already approximated the integral by a discrete sum, the additional simplification is acceptable. We also show that this additional simplification has small effect if the combined density is a relatively good importance sampling function.

The simplification is based on the relation of arithmetic and geometric means. The weighted sum of Eq. 10 is the arithmetic mean of ratios $f(X_{i,j})/p(\alpha, X_{i,j})$, which is approximated by the geometric mean:

$$\mu \approx \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{f(X_{i,j})}{p(\alpha, X_{i,j})} \approx \left(\prod_{j=1}^{n_i} \frac{f(X_{i,j})}{p(\alpha, X_{i,j})} \right)^{1/n_i}. \tag{12}$$

The error of this approximation is small if the terms of the arithmetic mean are close to each other. In our case, the terms are the ratios of the integrand and the combined density, which must be close to the integral if the combined density is a good importance sampling function. In Fig. 1 we compare the arithmetic and geometric means for the three 1D examples of Sect. 6.

The product of ratios in the geometric mean is replaced by the ratios of the products, then for the numerator and the denominator, we separately apply the approximation of the geometric mean by the arithmetic one:
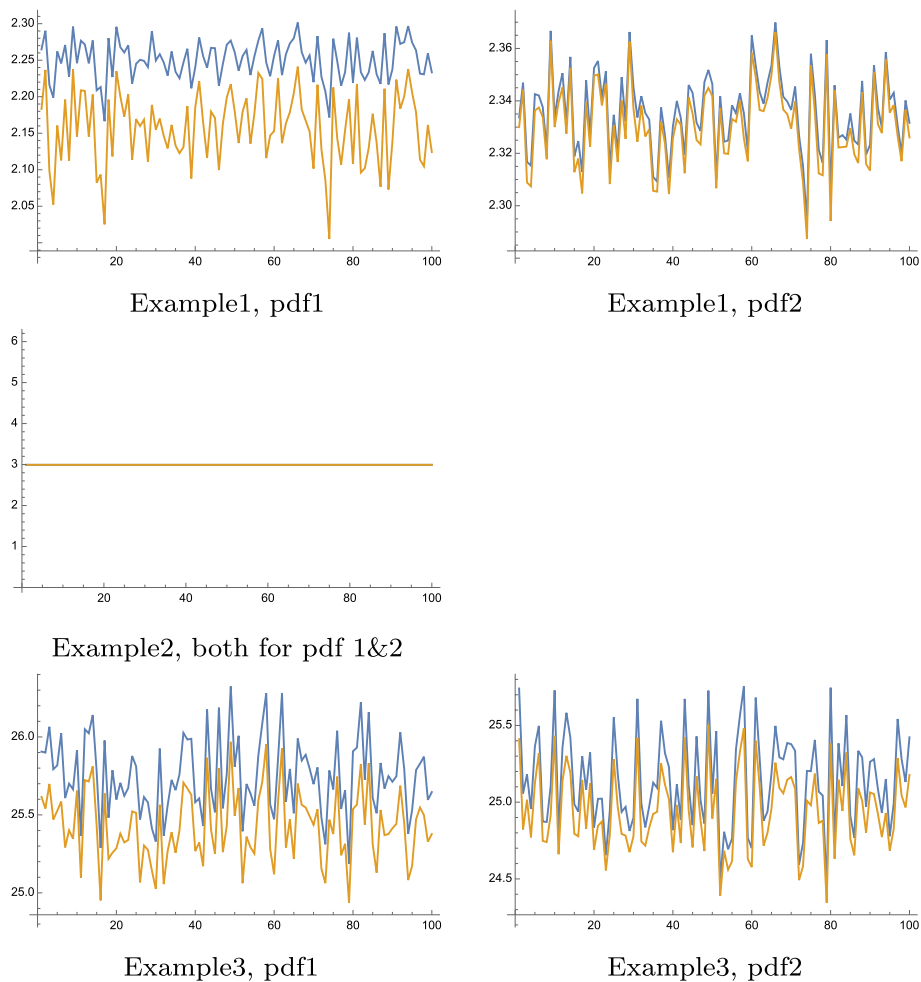
Fig. 1 Comparison of arithmetic (in blue) and geometric (in orange) means for Example1 (top), Example2 (middle), and Example3 (bottom) for 100 independent executions with 100 samples of pdf1 (left) and pdf2 (right). For Example2, pdf2 (not shown) is identical to pdf1

$$\left(\prod_{j=1}^{n_i} \frac{f(X_{i,j})}{p(\alpha, X_{i,j})}\right)^{1/n_i} = \frac{\left(\prod_{j=1}^{n_i} f(X_{i,j})\right)^{1/n_i}}{\left(\prod_{j=1}^{n_i} p(\alpha, X_{i,j})\right)^{1/n_i}}$$

$$\approx \frac{\frac{1}{n_i}\sum_{j=1}^{n_i} f(X_{i,j})}{\frac{1}{n_i}\sum_{j=1}^{n_i} p(\alpha, X_{i,j})}$$

$$= \frac{\sum_{j=1}^{n_i} f(X_{i,j})}{\sum_{j=1}^{n_i} p(\alpha, X_{i,j})}.$$

We cannot claim that in this approximation the factors are similar. The replacements of the geometric mean by the arithmetic one introduce an error. However, if the combined density is an already good importance sampling function, the terms in the numerator and denominator are similar with the exception of a constant scaling factor, i.e. $f(x) \approx \mu p(\alpha, x)$. Therefore, the errors of the numerator and the denominator are strongly correlated, thus the division makes the estimation more accurate.

Having applied this approximation, the optimality condition is the solution of the following equation:

$$\frac{\sum_{j=1}^{n_i} f(X_{i,j})}{\sum_{j=1}^{n_i} p(\alpha, X_{i,j})} = \mu. \tag{13}$$

Taking the reciprocals of both sides, we obtain a linear equation for the weights α_k since combined density $p(\alpha, x)$ is a linear function:

$$\frac{\sum_{j=1}^{n_i} p(\alpha, X_{i,j})}{\sum_{j=1}^{n_i} f(X_{i,j})} = \frac{1}{\mu}. \tag{14}$$

For example, if we combine technique 1 and technique 2, then we have to solve the following equation

$$\frac{\sum_{j=1}^{n_1} p(\alpha, X_{1,j})}{\sum_{j=1}^{n_1} f(X_{1,j})} = \frac{\sum_{j=1}^{n_2} p(\alpha, X_{2,j})}{\sum_{j=1}^{n_2} f(X_{2,j})}. \tag{15}$$

Let $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$ in order to satisfy the constraint. With this, the combined density is

$$p(\alpha, x) = \alpha p_1(x) + (1 - \alpha)p_2(x). \tag{16}$$

The condition of optimality can be expressed as

$$\alpha = \frac{P_{22}F_1 - P_{21}F_2}{P_{11}F_2 - P_{21}F_2 - P_{12}F_1 + P_{22}F_1} \tag{17}$$

where

$$P_{ik} = \sum_{j=1}^{n_k} p_i(X_{k,j}), \quad F_k = \sum_{j=1}^{n_k} f(X_{k,j}). \tag{18}$$

The structure of the parameters in the equations for the linear heuristic shows another advantage. As new samples are arriving, the parameters can be easily updated by adding the pdf and integrand values to the respective variables. Thus, previous and new samples can be exploited simultaneously unlike in the Newton–Raphson method which either individually stores all acquired samples or uses only those samples which have been generated since the last update of the weights.

For m pdfs, we have to solve the linear system of Eq. 14 with $m - 1$ independent equations setting $\alpha_m = 1 - \sum_{k=1}^{m-1} \alpha_k$.

5.1 The case of weight equal to zero

Equation 14 will fail to deliver a convex solution for the case where the optimal weight corresponding to a given technique is equal to zero. This reflects the fact that Eq. 8 will fail for this case. To account for this, we propose two techniques.

First, by construction, either $\{\alpha_i\}_{i=1}^{m}$ is a convex solution (i.e. for all i, $\alpha_i \geq 0$ and $\sum_{i=1}^{m} \alpha_i = 1$) or there is at least one negative weight. In the latter case, we take the minimum of the negative weights and substitute it with zero in Eq. 14, where we are left with a system of $m - 2$ equations and $m - 2$ unknowns. We repeat the procedure till we either obtain a convex solution or reach $m = 1$ when the weight of the remaining technique is set to 1.

The second technique is not as straightforward as the first one but gives more accurate solutions (see Example 6 in Sect. 6.1). It consists in considering each α_i equal to zero in turn and solving the remaining $m - 2$ linear system. From all the feasible solutions, we choose the one that has the smallest sample variance. Equation 3 is used as an estimator for the variance $V[F]$:

$$
V[F] \approx
$$
$$
\sum_{i=1}^{m} \alpha_i \left[\frac{1}{n_i} \sum_{j=1}^{n_i} \frac{f^2(X_{i,j})}{p^2(\alpha, X_{i,j})} - \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \frac{f(X_{i,j})}{p(\alpha, X_{i,j})} \right)^2 \right]. \tag{19}
$$

6 Results

6.1 Numerical 1D examples

In the first three examples two pdfs are combined. We obtain the optimal α with the Newton–Raphson minimization of the Kullback-Leibler divergence [19] and with our linear heuristic, respectively, and then compute the variance $V[F]$ for the obtained α by numerical integration with Mathematica. In all cases we execute 100 independent runs.

Additional examples combine more than two pdfs, for which we only compute the optimal α with the linear heuristic and then the variance $V[F]$.

1 *Example 1*

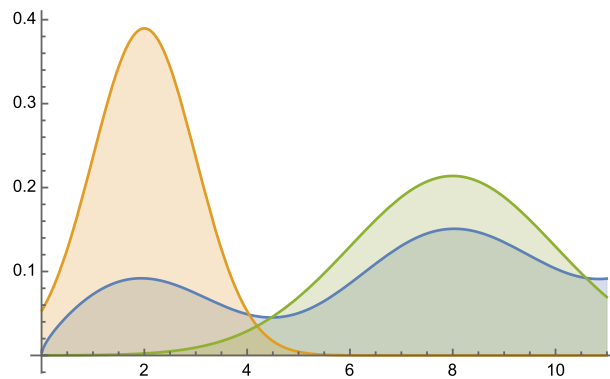*Suppose we want to evaluate the integral (see Fig. 2)*

Fig. 2 Example 1: $f(x)/\mu$ (in blue) superimposed on the two pdfs used for MIS integration
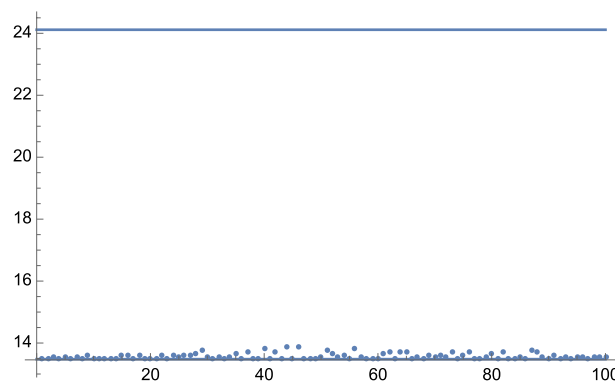

Fig. 3 The variances $V[F]$ of Example 1 computed for 100 independent executions with 4 Newton–Raphson iterations for the Kullback–Leibler divergence exploiting 50 samples in each iteration. The horizontal lines correspond to the optimal and the equal sampling budget variances, respectively

$$\int\limits_{0.01}^{3.5\pi} \left(\sqrt{x} + \sin x\right) \mathrm{d}x \approx 25.3065 \tag{20}$$

by MIS using pdfs $\phi_{2,1}(x)$ and $\phi_{8,2}(x)$, where $\phi_{m,\sigma}(x)$ stands for the pdf of Gaussian normal distribution of mean m and standard deviation σ.

For this example, equal sample budget MIS has variance $V[F] = 24.1152$. In Fig. 3 we show the values of $V[F]$ for the optimal α fractions minimizing the Kullback–Leibler divergence with four Newton–Raphson iterations relying on 50 samples in each iteration. Thus, 200 samples are taken in total for each run. Figure 4 shows the result of our new heuristic assigning 100 samples for each of the two techniques. Thus, the total number of samples is the same as in the Newton-Raphson solution.

1 *Example 2*

Let us consider integral (see Fig. 5)
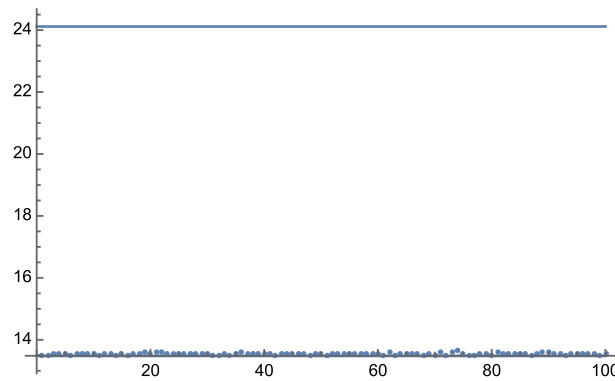
Fig. 4 The variances of Example 1 computed for 100 independent executions with the linear heuristic, taking 100 samples from each technique. The horizontal lines correspond to the optimal and equal sampling budget variances, respectively
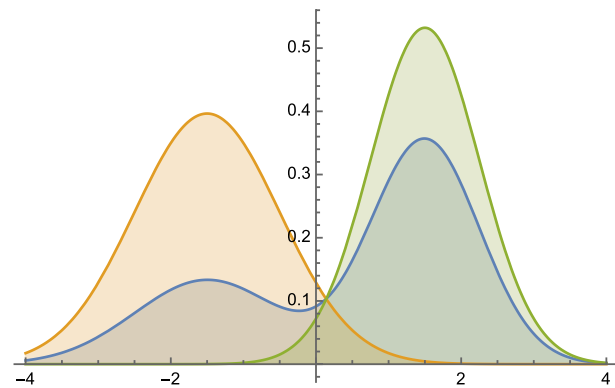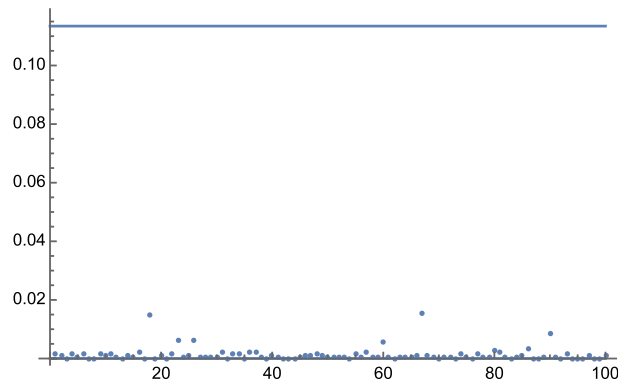

Fig. 5 Example 2: $f(x)/\mu$ (in blue) superimposed on the two pdfs used for MIS integration

$$\int_{-4}^{4} \phi_{-\frac{3}{2},1}(x) + 2\phi_{\frac{3}{2},\frac{3}{4}}(x)\mathrm{d}x \approx 2.9929 \qquad (21)$$

by MIS using pdfs $\phi_{-\frac{3}{2},1}(x)$ and $\phi_{\frac{3}{2},\frac{3}{4}}(x)$.

For this example, equal sample budget MIS has variance $V[F] = 0.1134$. Figure 6 presents the values of $V[F]$ for the optimal α fractions minimizing the Kullback–Leibler divergence by four Newton–Raphson iterations with 50 samples in each iteration, thus with 200 samples in total in a single run. Figure 7 shows the result of the linear heuristic with 100 runs of 100 samples for each technique, thus 200 samples in total in a single run.

1 *Example 3*

Consider the approximation of the following integral (see Fig. 8)

Fig. 6 The variances *V*[*F*] of Example 2 computed for 100 independent executions with 4 Newton–Raphson iterations for the Kullback–Leibler divergence exploiting 50 samples in each iteration. The horizontal lines correspond to the optimal and the equal sampling budget variances, respectively
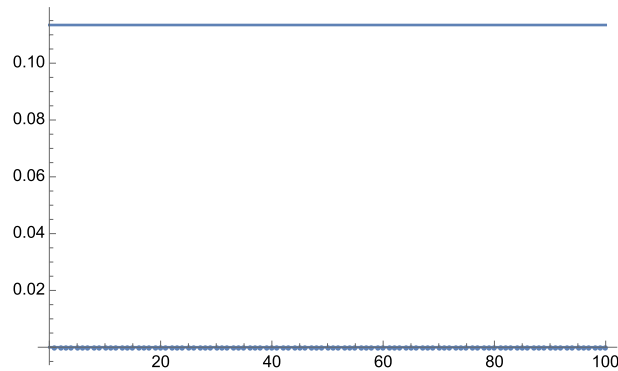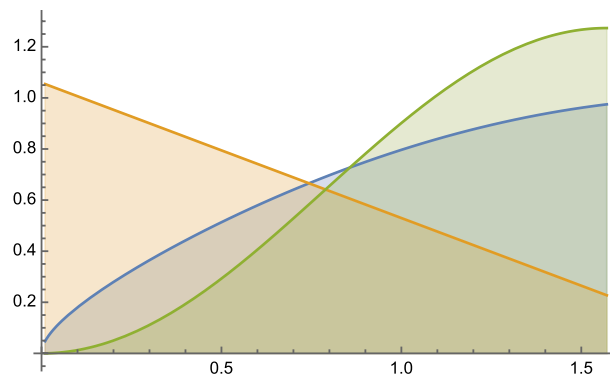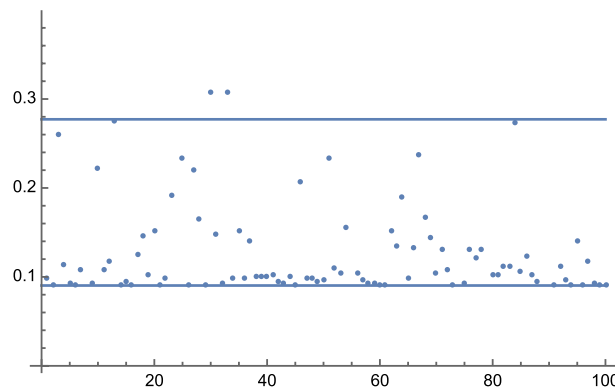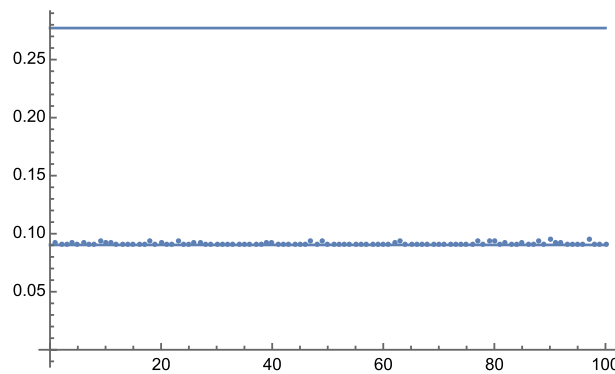

Fig. 7 The variances *V*[*F*] of Example 2 computed for 100 independent executions with the linear heuristic, using a total of 100 samples for each technique. The horizontal lines correspond to the optimal and equal sampling budget variances, respectively

Fig. 8 Example 3: $f(x)/\mu$ (in blue) superimposed on the two pdfs used for MIS integration

Fig. 9 The variances $V[F]$ of Example 3 computed for 100 independent executions with 4 Newton–Raphson iterations for the Kullback–Leibler divergence exploiting 50 samples in each iteration. The horizontal lines correspond to the optimal and the equal sampling budget variances, respectively

Fig. 10 The variances $V[F]$ of Example 3 computed for 100 independent executions with the linear heuristic, using a total of 100 samples for each technique. The horizontal lines correspond to the optimal and the equal sampling budget variances, respectively

$$\int\limits_{0.01}^{\pi/2} \left(\sqrt{x} + \sin x \right) \mathrm{d}x \approx 2.3118 \tag{22}$$

by MIS using importance functions $2 - x$ and $\sin^2(x)$. For this example, equal sample number MIS has variance $V[F] = 0.2772$. Figure 9 depicts the values of $V[F]$ for the optimal α fractions minimizing the Kullback-Leibler divergence using 100 runs each consisting of 4 Newton-Raphson iterations with 50 samples in each iteration, thus 200 samples in total for each run. Figure 10 shows the result of the linear heuristic with 100 runs of 100 samples for each technique.

1 *Example 4*

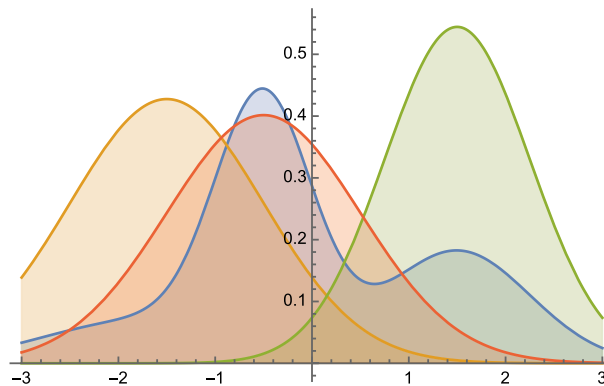*Consider the example with three pdfs, evaluating the following integral (see Fig. 11)*

Fig. 11 Example 4: $f(x)/\mu$ (in blue) superimposed with the three pdfs used for MIS integration
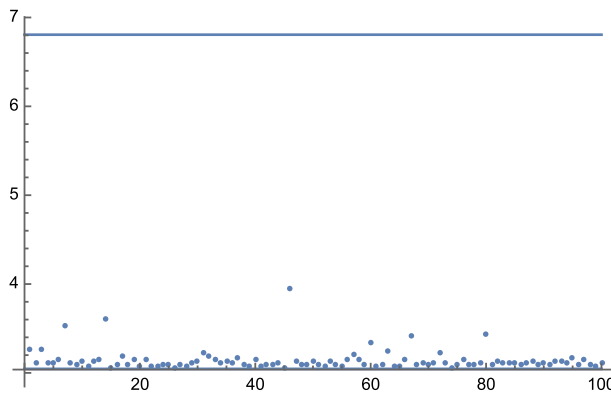

Fig. 12 The variances $V[F]$ of Example 4 computed for 100 independent executions with the linear heuristic, using 100 samples for each technique. The horizontal lines correspond to the optimal and the equal sampling budget variances, respectively

$$\int_{-3}^{3} \phi_{-1.8,1}(x) + 2\phi_{\frac{3}{2},\frac{3}{4}}(x) + 3\phi_{-\frac{1}{2},\frac{1}{2}}(x)\mathrm{d}x \approx 5.8394 \tag{23}$$

using pdfs $\phi_{-\frac{3}{2},1}(x)$, $\phi_{\frac{3}{2},\frac{3}{4}}(x)$, and $\phi_{-\frac{1}{2},1}(x)$.

For this example, equal sample number MIS has variance $V[F] = 6.8063$ while the minimum variance is $V[F] = 3.0454$. In Fig. 12 we show the result of our linear heuristic with 100 runs of 100 samples for each technique.

1 *Example 5*

Example 5 solves the following integral (see Fig. 13)

$$\int_{-3}^{3} \phi_{-1.8,1}(x) + 6\phi_{\frac{3}{2},\frac{3}{4}}(x) + 3\phi_{-\frac{1}{2},\frac{1}{2}}(x) + 3\phi_{\frac{1}{2},\frac{1}{2}}(x)\mathrm{d}x$$
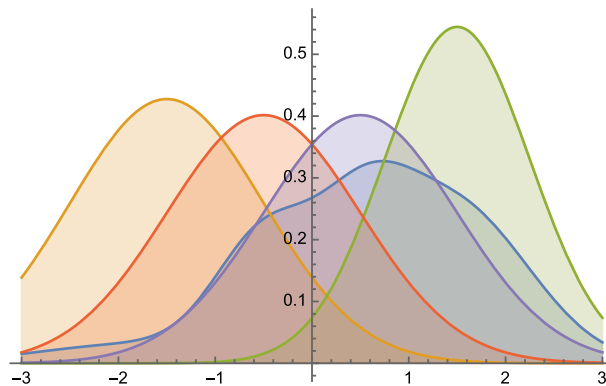$$\approx 12.7484 \tag{24}$$

Fig. 13 Example 5, $f(x)/\mu$ (in blue) superimposed with the four pdfs used for MIS integration
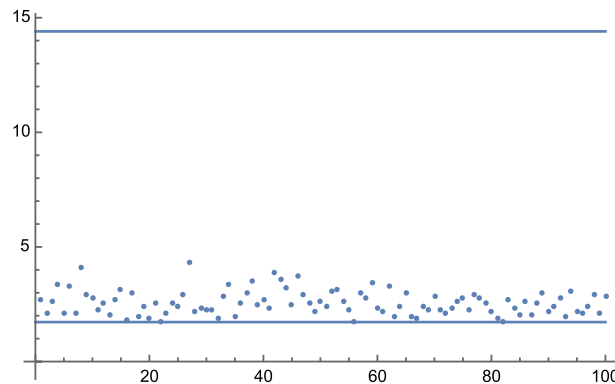

Fig. 14 The variances $V[F]$ of Example 5 computed for 100 independent executions with the linear heuristic, using a total of 100 samples for each technique. The horizontal lines correspond to the optimal and the equal sampling budget variances, respectively
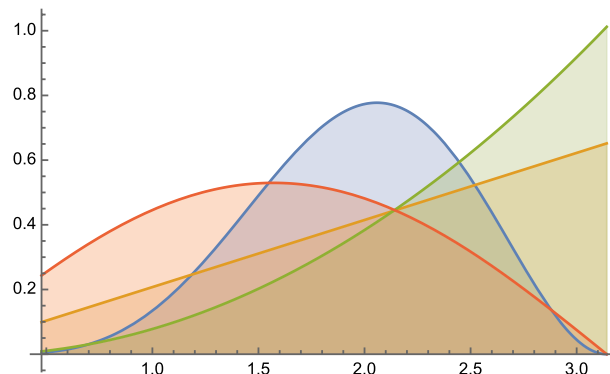

Fig. 15 Example 6, $f(x)/\mu$ (in blue) superimposed with the three pdfs used for MIS integration

using pdfs $\phi_{-\frac{3}{2},1}(x)$, $\phi_{\frac{3}{2},\frac{3}{4}}$, $\phi_{-\frac{1}{2},1}(x)$, and $\phi_{\frac{1}{2},1}(x)$. Equal sample number MIS has variance $V[F] = 14.4033$ and the minimum variance is $V[F] = 1.7217$. In Fig. 14 we show the result of our linear heuristic with 100 runs of 100 samples for each technique.
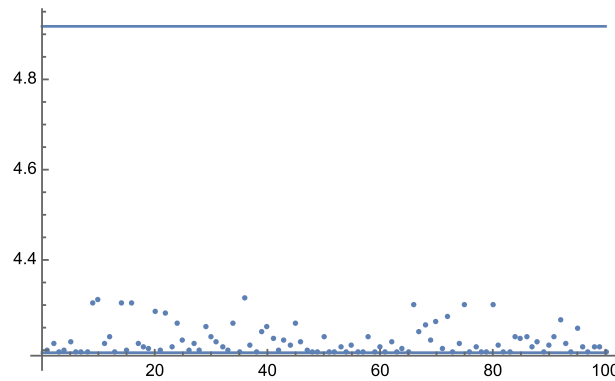
Fig. 16 The variances $V[F]$ of Example 6 computed for 100 independent executions with the linear heuristic, where we set $\alpha_1 = 0$, using 100 samples for each of the other two techniques. The horizontal lines correspond to the optimal and the equal sampling budget variances, respectively
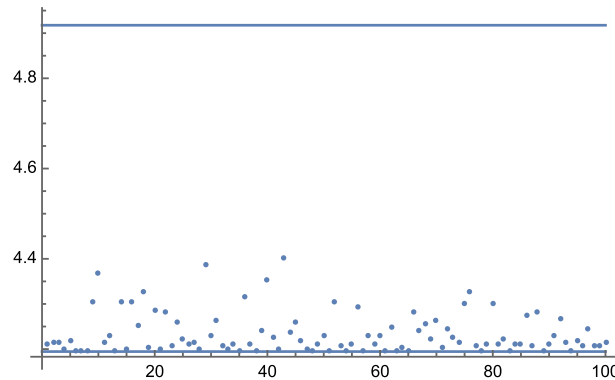

Fig. 17 The variances $V[F]$ of Example 6 computed for 100 independent executions with the linear heuristic and using the first technique of Sect. 5.1 to sort out negative solutions. We use the same 100 samples from each technique as in Fig. 16. The horizontal lines correspond to the optimal and the equal sampling budget variances, respectively. Compare it with Figs. 16 and 18

1 *Example 6*

This example with three pdfs shows what happens when one of the optimal weights is 0. Consider solving the following integral (see Fig. 15)

$$\int_{3/2\pi}^{\pi} \left(x^2 - \frac{x}{\pi}\right) \sin^2(x)\mathrm{d}x \approx 3.5962 \tag{25}$$

using importance functions x, $x^2 - x/\pi$ and $\sin(x)$. Equal sample budget MIS has variance $V[F] = 4.9175$ while the minimum variance is $V[F] = 4.1945$. The minimum $V[F]$ value corresponds to $\alpha_1 = 0$, $\alpha_2 = 0.1986$, and $\alpha_3 = 0.8014$. In Fig. 16 we set $\alpha_1 = 0$ and apply the linear heuristic to a two-pdf problem of 100 samples for each technique. In Fig. 17 we apply the first technique of Sect. 5.1 to sort out negative solutions, and use 100 samples from each technique (same ones as in Fig. 16). In Fig. 18 we apply the second technique of Sect. 5.1 with the same samples to get rid of the negative solution. Observe the better results of this second technique.
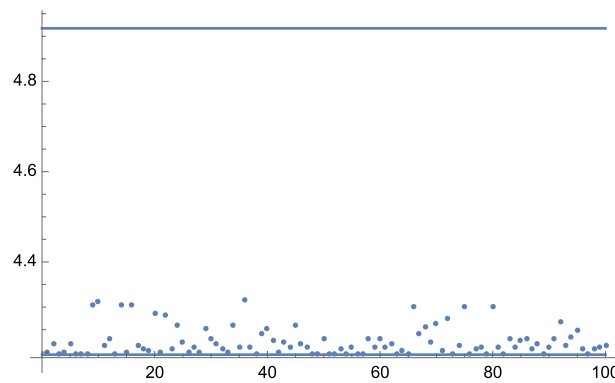
Fig. 18 The variances $V[F]$ of Example 6 computed for 100 independent executions with the linear heuristic and using the second technique of Sect. 5.1 to sort out negative solutions. We use the same 100 samples from each technique as in Fig. 16. The horizontal lines correspond to the optimal and the equal sampling budget variances, respectively. Compare it with Figs. 16 and 17

6.2 Combination of light source sampling and BRDF sampling in computer graphics

In order to demonstrate the efficiency of the proposed linear heuristic, we address the direct lighting problem and combine light source sampling and BRDF sampling. The reflected radiance $L^r(\mathbf{p}, \omega)$ of a surface point $\mathbf{p}$ at direction ω is expressed as an integral in the hemispherical domain of incident directions Ω:

$$L^r(\mathbf{p}, \omega) = \int_\Omega L(\mathbf{p}', \omega') f_r(\omega', \mathbf{p}, \omega) \cos \theta' \mathrm{d}\omega' \tag{26}$$

where $L(\mathbf{p}', \omega')$ is the radiance of point $\mathbf{p}'$ visible from point $\mathbf{p}$ in incident direction $-\omega'$, $f_r(\omega', \mathbf{p}, \omega)$ is the BRDF expressing the portion of the light beam that is reflected from direction ω' to ω at point $\mathbf{p}$, and θ' is the angle between the surface normal at point $\mathbf{p}$ and incident direction $-\omega'$. We have two sampling methods $p_1(\omega')$ approximately mimicking the $f_r(\omega, \mathbf{p}, \omega') \cos \theta'$ factor and $p_2(\omega')$ mimicking the incident radiance $L(\mathbf{p}', \omega')$.

MIS would use a combined pdf:

$$p(\alpha, \omega') = \alpha p_1(\omega') + (1 - \alpha) p_2(\omega'). \tag{27}$$

where we need to find optimal weight α.

We render the classic scene of Veach with combined light source and BRDF sampling. The illuminated rectangles have max-Phong BRDF [21] with shininess parameters 200, 500, 2000, and 5000, respectively. The four spherical light sources emit the same power.

For each pixel, we use 100 samples in total organized in 10 iterations of 10 samples each. The process starts with 5 BRDF and 5 light source samples per pixel, and the per-pixel α weights are updated at the end of each iteration. Figure 19 shows the rendered images together with the α maps, and we compare the original sampling techniques, equal count MIS, the optimization of the Kullback–Leibler divergence with the Newton–Raphson method, and the proposed linear heuristic.

Concerning the complexity and the overhead of the method, having identified visible point $\mathbf{p}$, BRDF sampling finds a random direction ω', obtains the first intersection

BRDF sampling (RMSE = 466)   Light source sampling (RMSE = 461)

Equal budget MIS (RMSE = 277)   Kullback-Leibler (RMSE = 248)
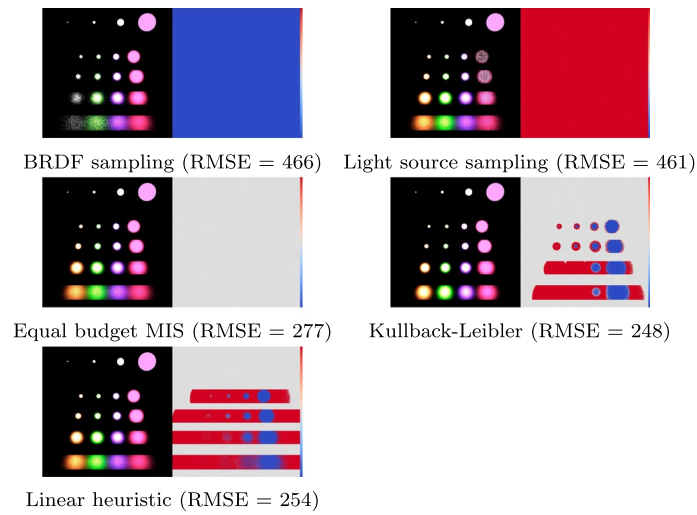
Linear heuristic (RMSE = 254)

Fig. 19 Comparison of MIS weighting schemes for the direct lighting problem of computer graphics. The left part is the image rendered with 100 rays per pixel, the right part is weight α of the light source sampling. The RMSE is computed as the average of 10 independent executions. The [0,1] interval of possible α values is visualized by the color bar

$\mathbf{p}'$ of the ray of start $\mathbf{p}$ and direction ω', evaluates the emitted radiance of the intersection point $\mathbf{p}'$, and finally divides the product of the $f_r(\omega, \mathbf{p}, \omega') \cos \theta'$ factor and the emitted radiance by the pdf of the random direction. For the scene of Fig. 19, the calculation of a light path sample took 1.786 μs on average on an Intel Core 7 CPU. When BRDF sampling is a part of MIS, the algorithm should also find the pdf of the light source sampling if a light source is hit by the ray, which increases the computation time to 1.789 μs. The proposed linear heuristic requires six additions to update the parameters in Eq. 18. After every 10 samples, weight α is updated according to Eq. 17. With these additional computations, a single sample needs 1.791 μs.

In case of light source sampling, first the emission point on the light source is sampled, and it is checked whether the line segment between visible point $\mathbf{p}$ and the light source point $\mathbf{p}'$ intersects any object. If there is no intersection, i.e., the light source is not occluded, then the emitted radiance is multiplied by the $f_r(\omega, \mathbf{p}, \omega') \cos \theta'$ factor and divided by the pdf of light source sampling. The calculation of a single pixel sample by light source sampling took 2.462 μs on average, which is increased to 2.687 μs with the MIS overhead, and to 2.703 μs with the overhead of the linear heuristic.

Enabling equal sample count MIS, half of the samples will be generated basically with BRDF sampling and half with light source sampling, which means that the average time needed by a light path sample becomes 2.246 μs. Linear heuristic sets the combination weights independently in every pixel. Its average computation time is 2.259 μs, which is around a 0.5% larger than that of equal count MIS partly because of its added overhead, and partly because it prefers more expensive light source sampling to BRDF sampling for this scene. This increase in computation cost is compensated by the reduced error, resulting in a gain in efficiency (defined as inverse of cost times mean square error) of about a 15%.

7 Discussion

The linear heuristic compares advantageously to the Newton–Raphson solution in the three 1D two-pdf examples shown. For the zero variance case of Example 2, the solution is exact. Observe also the higher robustness of our heuristic in the other three examples, especially in Example 3. For the direct lighting problem, where the function cannot be integrated analytically, our heuristic is better than equal sampling but it does not improve on the Kullback–Leibler minimization. For the 1D examples shown with three and four pdfs, the heuristic also works well. Assigning zero value to the most negative solution is less accurate than the comparison of the sample variance for the different feasible solutions.

8 Conclusions and future work

Inspired by the solution to the Kullback–Leibler representation of the MIS problem, we have presented a linear heuristic to obtain the optimal weights in MIS. Our linear heuristic compares advantageously to the Newton–Raphson solution in the shown 1D examples, both in accuracy ad robustness. It is better than equal sampling in the presented direct light source problem solution. It scales to any number of techniques used by solving a linear system of equations. Negative solutions appear when the optimal solution is on the border of the simplex domain, and are dealt with by assigning zero to the most negative weight and solving the linear system again without the corresponding technique, or by comparing the sample variances of the different feasible solutions, which gives a more accurate result although at a higher cost. In the future, possible additional linear heuristics will be investigated, potentially based on other divergence representation of MIS, and the cost of sampling will be taken into account.

References
1. E. Veach, L.J. Guibas, Optimally combining sampling techniques for Monte Carlo rendering, in *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '95*, pp. 419–428. ACM, New York, NY, USA (1995). https://doi.org/10.1145/218380.218498
2. E. Veach, Robust Monte Carlo Methods for Light Transport Simulation. Ph.D. thesis, Stanford University (1997)
3. V. Elvira, L. Martino, D. Luengo, M.F. Bugallo, Efficient multiple importance sampling estimators. IEEE Signal Process. Lett. **22**(10), 1757–1761 (2015). https://doi.org/10.1109/LSP.2015.2432078
4. V. Elvira, L. Martino, D. Luengo, M.F. Bugallo, Generalized Multiple Importance Sampling. ArXiv e-prints (2015). arXiv:1511.03095

5.   H. Lu, R. Pacanowski, X. Granier, Second-order approximation for variance reduction in multiple importance sampling. Comput. Graph. Forum **32**(7), 131–136 (2013). https://doi.org/10.1111/cgf.12220
6.   M. Sbert, V. Havran, L. Szirmay-Kalos, Variance analysis of multi-sample and one-sample multiple importance sampling. Comput. Graph. Forum **35**(7), 451–460 (2016). https://doi.org/10.1111/cgf.13042
7.   V. Havran, M. Sbert, Optimal combination of techniques in multiple importance sampling, in *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry. VRCAI '14*, pp. 141–150. ACM, New York, NY, USA (2014). https://doi.org/10.1145/2670473.2670496
8.   M. Sbert, V. Havran, Adaptive multiple importance sampling for general functions. Vis. Comput. **33**, 845–855 (2017). https://doi.org/10.1007/s00371-017-1398-1
9.   M. Sbert, V. Havran, L. Szirmay-Kalos, Multiple importance sampling revisited: breaking the bounds. EURASIP J. Adv. Signal Process. **2018**(1), 15 (2018). https://doi.org/10.1186/s13634-018-0531-2
10.   M. Sbert, V. Havran, L. Szirmay-Kalos, Optimal deterministic mixture sampling, in *Eurographics 2019-Short Papers* (2019). https://doi.org/10.2312/egs.20191018. Eurographics
11.   M. Sbert, V. Havran, L. Szirmay-Kalos, V. Elvira, Multiple importance sampling characterization by weighted mean invariance. Vis. Comput. **34**(6–8), 843–852 (2018)
12.   I. Kondapaneni, P. Vevoda, P. Grittmann, T. Skřivan, P. Slusallek, J. Křivánek, Optimal multiple importance sampling. ACM Trans. Graph. (2019). https://doi.org/10.1145/3306346.3323009
13.   M. Sbert, V. E, Generalizing the balance heuristic estimator in multiple importance sampling. Entropy **24**(2), 191 (2022). https://doi.org/10.3390/e24020191
14.   J. Vorba, J. Hanika, S. Herholz, T. Mueller, J. Krivanek, A. Keller, Path guiding in production, in *ACM SIGGRAPH 2019 Courses* (2019). ACM
15.   J. Cornebise, E. Moulines, J. Olsson, Adaptive methods for sequential importance sampling with application to state space models (2008). arXiv:0803.0054
16.   J. Míguez, On the performance of nonlinear importance samplers and population Monte Carlo schemes, in *2017 22nd International Conference on Digital Signal Processing (DSP)*, pp. 1–5 (2017). IEEE
17.   T. Müller, B. Mcwilliams, F. Rousselle, M. Gross, J. Novák, Neural importance sampling. ACM Trans. Graph. **2019**(38), 5 (2019)
18.   F. Nielsen, R. Nock, On the chi square and higher-order chi distances for approximating f-divergences. IEEE Signal Process. Lett. **21**(1), 10–13 (2014). https://doi.org/10.1109/LSP.2013.2288355
19.   M. Sbert, L. Szirmay-Kalos, Robust multiple importance sampling with Tsallis φ-divergences. Entropy **24**, 1240 (2022). https://doi.org/10.3390/e24091240
20.   D. Murray, S. Benzait, R. Pacanowski, X. Granier, On learning the best local balancing strategy, in *Eurographics 2020-Short Papers* (2020). https://doi.org/10.2312/egs.20201009. Eurographics
21.   L. Neumann, A. Neumann, L. Szirmay-Kalos, Compact metallic reflectance models. Comput. Graph. Forum **1999**(18), 161–172 (1999)

Publisher's Note