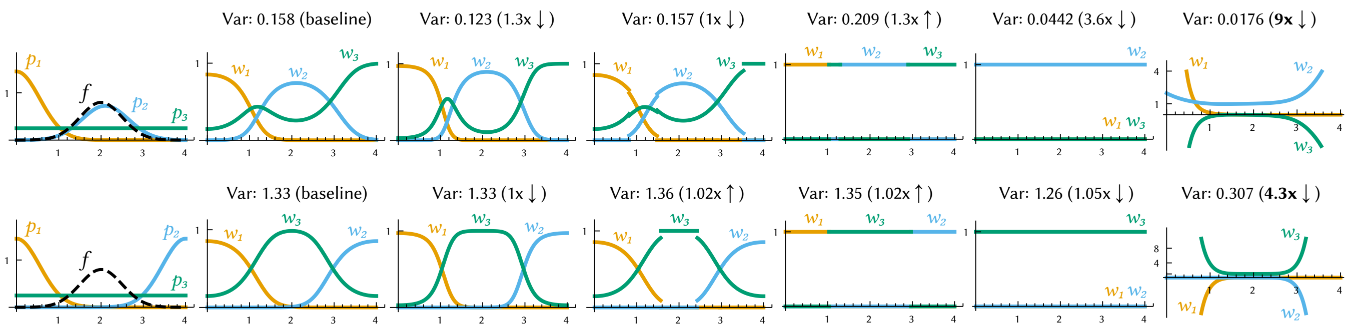
Optimal Multiple Importance Sampling: Supplemental Document

IVO KONDAPANENI, Charles University, Prague
PETR VÉVODA, Charles University, Prague and Render Legion, a. s.
PASCAL GRITTMANN, Saarland University
TOMÁŠ SKŘIVAN, IST Austria
PHILIPP SLUSALLEK, Saarland University and DFKI
JAROSLAV KŘIVÁNEK, Charles University, Prague and Render Legion, a. s.



a) integrand and sampling techniques | b) balance heuristic weights | c) power heuristic weights | d) cutoff heuristic weights ($\alpha = 0.1$) | e) maximum heuristic weights | f) best-technique heuristic weights | g) optimal weights (unconstrained sign)

Fig. 1. Extended version of Fig. 2 from the main paper with cutoff (d) and maximum (e) heuristic weights included. a) depicts an integration problem where the integral of a function f is estimated via MIS. Three sampling techniques, p_1, p_2, and p_3, are used, and one sample is taken from each. The two rows differ solely in the sampling technique p_2: while p_2 closely matches f in the first row, in the second row it is fairly different. b) - e) plot, respectively, the balance, power, cutoff, and maximum heuristic weights as defined by Veach [1997]. f) and g) depict, respectively, the best-technique heuristic and the optimal weights as defined in the paper.

This supplemental document provides several additional results to complete the main article.

1 WEIGHTS COMPARISON FOR TWO SIMPLE 1D INTEGRATION PROBLEMS

In Fig. 1 we provide an extended version of Fig. 2 from the paper. It compares different MIS combination strategies on two simple 1D integration problems.

We also provide the Mathematica notebook used for its production (`figure2/Figure2.nb`). Running the notebook requires the Mathematica software (version 11.0 or newer)[1] with the MaTex package (version 1.7.4 or newer)[2]. Label positions in the produced image were tweaked manually.

2 LIGHT SAMPLING TECHNIQUES FORMULAS

In Sec. 8.3 and Fig. 7b in the paper we discuss different light sampling techniques. Here we provide a derivation of the quantities illustrated in Fig. 7b. If expressed in the solid angle measure, the integrand and probability density functions of the techniques read:

$$\begin{aligned}
f(\theta) &= L_e \cos\theta \propto \cos\theta \\
p_{\mathrm{Spherical}}(\theta) &= \frac{1}{|A_{\mathrm{Spherical}}|} \propto 1 \\
p_{\mathrm{UniformArea}}(\theta) &= \frac{1}{|A_{\mathrm{UniformArea}}|}\frac{d(\theta)^2}{\cos l(\theta)} \\
&= \frac{1}{|A_{\mathrm{UniformArea}}|}\frac{d_\perp^2}{\cos^3 l(\theta)} \propto \frac{1}{\cos^3 l(\theta)} \\
p_{\mathrm{Parallel}}(\theta) &= \frac{1}{|A_{\mathrm{Parallel}}|}\frac{d(\theta)^2}{\cos l(\theta)} = \frac{1}{|A_{\mathrm{Parallel}}|}\frac{d_\perp^2}{\cos^3 l(\theta)} \\
&\propto \frac{1}{\cos^3 l(\theta)} = \frac{1}{\cos^3\theta}
\end{aligned} \tag{S1}$$

The quantities used in the formulas are shown in Fig. 2. As discussed in Sec. 8.3 in the main paper, linear combination of the *Uniform area* and *Spherical* techniques is a good approximation for the integrand as long as the lit surface is parallel to the light source. For that case it holds $\cos^{-3}\theta = \cos^{-3} l(\theta)$, but that relation breaks for points on differently oriented surfaces, and the linear combination of the *Uniform area* and *Spherical* techniques on such surfaces can no longer approximate the integrand well.

[1] http://www.wolfram.com/mathematica/
[2] http://library.wolfram.com/infocenter/MathSource/9355/

The above problem does not occur with the *Parallel* technique, which first projects the light onto a plane parallel to the shaded surface, and then samples that projection. Therefore, a linear combination of its sampling density ($\propto \cos^{-3}\theta$) with the *Spherical* technique ($\propto 1$) better approximates f irrespective of the light orientation.

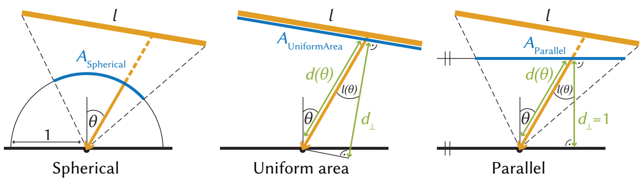


Fig. 2. Illustration of the quantities used in formulas in Sec. 2. A denotes surface area of the sampled light/projection, θ angle at the surface, $l(\theta)$ angle at the light/projection, $d(\theta)$ distance between the point on the surface and on the light/projection, $d_\perp$ perpendicular distance of the light/projection.

3 RELATIONSHIP TO OWEN AND ZHOU

Approximating $\boldsymbol{\alpha}$ in Eq. (16) in the paper can be viewed as a regression problem, as Owen and Zhou [2000] did. To explain their approach, we denote parts of (16) using the following notation

$$f_{ij} = f(X_{ij})/p_{\mathbf{c}}(X_{ij}), \qquad d_{ijk} = p_k(X_{ij})/p_{\mathbf{c}}(X_{ij}), \tag{S2}$$

where $p_{\mathbf{c}}(x) = \sum_{k=1}^{N} n_k p_k(x)/M$ and $M = \sum_{k=1}^{N} n_k$. Let us uniquely map an index pair $(i,j), i = 1,\dots,N,\ j = 1,\dots,n_i$ to an index $l = 1,\dots,M$ and denote quantities from (S2) as f_l and d_{lk} in the following text. Owen and Zhou approximate the optimal coefficients $\boldsymbol{\alpha}$ by multiple linear regression of observations f_l on regressors d_{lk} along with an intercept term $\langle\alpha_0\rangle$, i.e.,

$$\langle\alpha_0\rangle + \textstyle\sum_{k=1}^{N}\langle\alpha_k\rangle d_{lk} \approx f_l, \quad l = 1,\dots,M. \tag{S3}$$

In matrix form,

$$\mathbf{Dh} \approx \mathbf{f}, \tag{S4}$$

where each row corresponds to (S3) for a particular index l. Therefore $\mathbf{D}$ is a matrix $M \times (N+1)$ with the first column composed of ones and the $(k+1)$-th column being $(d_{1k},\dots,d_{Mk})^\intercal$, $\mathbf{f}$ is a column vector $(f_1,\dots,f_M)^\intercal$, and $\mathbf{h}$ is a column vector of length $N+1$ representing the terms $\langle\alpha_0\rangle$ and $\langle\alpha_k\rangle, k = 1,\dots,N$. Note that the above regression problem can be composed from several MIS sample batches by concatenating the corresponding matrices and vectors.

To solve the regression problem (S4), Owen and Zhou minimize $\|\mathbf{Dh} - \mathbf{f}\|_2^2$ in terms of $\mathbf{h}$, which leads to the *normal equation* for $\mathbf{h}$

$$\mathbf{D}^\intercal\mathbf{Dh} = \mathbf{D}^\intercal\mathbf{f}. \tag{S5}$$

The above equation is singular, because the first column of ones in $\mathbf{D}$ is a linear combination of the others, i.e., $\sum_{k=1}^{N} n_k d_{lk}/M = 1, l = 1,\dots,M$. Let $\mathbf{h}_0$ be a solution of (S5), and $\mathbf{v} = (-M, n_1,\dots,n_N)^\intercal \in \mathrm{Null}(\mathbf{D}^\intercal\mathbf{D})$. Then each $\mathbf{h} \in \{\mathbf{h}_0 + s\mathbf{v} | s \in \mathbb{R}\}$ solves (S5). Because the sum of elements of $\mathbf{v}$ equals 0, it holds for all $\mathbf{h}$ that the sum of their elements equals *the same* number, and we show in the next paragraph that it must be an estimate of the integral F. We also show that an alpha estimator extracted from any $\mathbf{h}$ estimates some $\widetilde{\boldsymbol{\alpha}}$, which belongs to the full solution for alphas (see Appendix C of the paper).

We can find a solution to (S5) by SVD applied directly (preferred by Owen and Zhou), but we can also solve a *truncated* system $\hat{\mathbf{D}}^\intercal\hat{\mathbf{D}}\hat{\mathbf{h}} = \hat{\mathbf{D}}^\intercal\mathbf{f}$, where $\hat{\mathbf{D}}$ is obtained by dropping one column from $\mathbf{D}$. That yields a truncated solution vector $\hat{\mathbf{h}}$, and it is equivalent to finding a solution $\mathbf{h}$ which has the element corresponding to the skipped column equal to zero. Therefore, summing up the elements of such a truncated vector gives the same estimate of F. Dropping the first column from $\mathbf{D}$ related to $\langle\alpha_0\rangle$ *makes the truncated system even the same (up to a scaling factor) as our system estimated by the balance heuristic* (21) described in Sec. 7.1, because then

$$\hat{\mathbf{D}}^\intercal\hat{\mathbf{D}} = M^2\langle\mathbf{A}\rangle, \quad \text{and} \quad \hat{\mathbf{D}}^\intercal\mathbf{f} = M^2\langle\mathbf{b}\rangle. \tag{S6}$$

The truncated vector $\hat{\mathbf{h}}$ solving such a system is then equal to the $\langle\boldsymbol{\alpha}\rangle$ estimate described in Sec. 7.2, and there exists an $\mathbf{h}_0$, with the first component equal to zero, corresponding to such a truncated vector. Therefore, using $\mathbf{n} = (n_1,\dots,n_N)^\intercal$, an alpha estimate represented by $\mathbf{h} = \mathbf{h}_0 + s\mathbf{v}, s \in \mathbb{R}$ equals to $\langle\boldsymbol{\alpha}\rangle + s\mathbf{n}$, which is an estimate of $\widetilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + s\mathbf{n}$ from the full solution for alphas. It follows that the sum of elements of any such $\mathbf{h}$ must be equal to the sum of elements of $\langle\boldsymbol{\alpha}\rangle$ and therefore it is an estimator of F. In other words, the solutions given by Owen and Zhou's approach are equivalent to the solution of the system from Theorem 5.2 as long as the system parts $\mathbf{A}$ and $\mathbf{b}$ are estimated by the balance heuristic. Our result is more general, and it suggests the existence of some alternative strategies how to approximate $\mathbf{A}$, $\mathbf{b}$, and $\boldsymbol{\alpha}$.

4 PSEUDOCODE OF FAN ET AL.

Here we present pseudocode of our adaptation of the method by Fan et al. [2006] who modified Owen and Zhou's approach by using regularization and applied it in rendering (see the previous section for details of Owen and Zhou's method). The computation is performed in batches. In each batch, n_i samples are drawn from each of the N sampling techniques $p_i, i = 1,\dots,N$ (we use $N = 2, n_1 = n_2 = 4$,

ALGORITHM 3: Fan et al.

$M \leftarrow \sum_{i=1}^{N} n_i$; // batch size
$result \leftarrow 0$;
for $batch \leftarrow 1$ **to** $maxBatches$ **do**
    $\mathrm{D} \leftarrow 0^{M\times(N+1)}$; $\mathrm{f} \leftarrow 0^{M\times 1}$; $l \leftarrow 0$;
    for $i \leftarrow 1$ **to** N **do**
        for $j \leftarrow 1$ **to** n_i **do**
            $X_{ij} \leftarrow$ draw j-th sample from technique p_i;
            $l \leftarrow l + 1$;
            $\mathrm{D}_{l0} \leftarrow 1$; // intercept term
            for $k \leftarrow 1$ **to** N **do**
                $\mathrm{D}_{lk} \leftarrow d_{ijk}$; // (S2)
            end
            $\mathrm{f}_l \leftarrow f_{ij}$; // (S2)
        end
    end
    $\hat{\mathrm{D}} \leftarrow$ drop one column of D;
    $\hat{\mathrm{h}} \leftarrow$ solve regularized linear system $(\hat{\mathrm{D}}^\intercal\hat{\mathrm{D}} + \lambda\mathrm{I})\hat{\mathrm{h}} = \hat{\mathrm{D}}^\intercal\mathrm{f}$;
    $result \leftarrow result + \sum_{i=1}^{N}\hat{\mathrm{h}}_i$
end
return $result/maxBatches$

each batch therefore consists of $M = 8$ samples, the same total number of samples as 4 iterations of our Direct estimator). For each sample one row of the data matrix $\mathbf{D}$ and vector $\mathbf{f}$ is computed according to (S2). After all samples in one batch are processed, one column of $\mathbf{D}$, corresponding to $\langle \alpha_k \rangle, k = 1, \ldots, N$, is dropped. Then the regularized truncated system $(\hat{\mathbf{D}}^\mathsf{T}\hat{\mathbf{D}} + \lambda\mathbf{I})\hat{\mathbf{h}} = \hat{\mathbf{D}}^\mathsf{T}\mathbf{f}$ is solved, where $\mathbf{I}$ is the identity matrix and λ is the weight of the regularization (we use $\lambda = 1$ as suggested by Fan et al.). Finally, the sum of the elements of the solution $\hat{\mathbf{h}}$ is added to the final result and the algorithm proceeds to the next batch.

Note that in practice this algorithm can be implemented to directly compute $\hat{\mathbf{D}}^\mathsf{T}\hat{\mathbf{D}}$ and $\hat{\mathbf{D}}^\mathsf{T}\mathbf{f}$ instead of first computing $\hat{\mathbf{D}}$ and $\mathbf{f}$ and then multiplying by $\hat{\mathbf{D}}^\mathsf{T}$. Such an implementation has the same computational complexity but smaller memory requirements. Fan et al. do not mention this optimization but our implementation of this algorithm applies it. This optimized implementation is included in the provided source code.

5 IMPLEMENTATION SOURCE CODE

We implemented the optimal MIS weights as a new integrator in pbrt-v3. Its source code including our changes can be found in the folder `implementation/src/integrators`. Our integrator is called *optmis*. It computes direct illumination only, and it has several parameters (described in `implementation/Params.html`) for specifying light selection/light sampling techniques, combination strategies, switching to the method of Fan et al., etc.

6 ADDITIONAL RESULTS

We provide additional results for scenes in the paper accessible through the html file `Supplemental_Results.html`.

REFERENCES

Shaohua Fan, Stephen Chenney, Bo Hu, Kam Wah Tsui, and Yu Chi Lai. 2006. Optimizing control variate estimators for rendering. *Comput. Graph. Forum (EUROGRAPHICS 2006)* 25, 3 (2006), 351–357.

Art Owen and Yi Zhou. 2000. Safe and Effective Importance Sampling. *J. Amer. Statist. Assoc.* 95, 449 (2000), 135–143.