Supplementary Document: Generalized Resampled Importance Sampling: Foundations of ReSTIR

DAQI LIN*, University of Utah, USA
MARKUS KETTUNEN*, NVIDIA, Finland
BENEDIKT BITTERLI, NVIDIA, USA
JACOPO PANTALEONI, NVIDIA, Germany
CEM YUKSEL, University of Utah, USA
CHRIS WYMAN, NVIDIA, USA

Contents

S.1 DERIVATION OF RESAMPLING MIS WEIGHTS

In this section we first derive the generalized Talbot MIS (Equation 36) and Pairwise MIS (Equation 37, Equation 38) weights from the requirement that the resampling weights w_i given by Equation 19 must have a finite upper bound.

We then derive the upper bounds for the resampling weights w_i with the above MIS weights schemes, and for variants of the MIS weights that use tractable PDFs p_i instead of $\hat{p}_i$.

We assume that of the M input samples X_i, indices in the set R are canonical (Definition 5.2), i.e., their domain is Ω, the shift mapping is identity, and the target density $\hat{p}_i = \hat{p}$. The number of canonical samples is denoted $|R|$.

*Joint first authors; equal contribution.

Authors' addresses: Daqi Lin, University of Utah, USA; Markus Kettunen, NVIDIA, Finland; Benedikt Bitterli, NVIDIA, USA; Jacopo Pantaleoni, NVIDIA, Germany; Cem Yuksel, University of Utah, USA; Chris Wyman, NVIDIA, USA.

S.1.1 Generalizing Talbot MIS Weights

We first require that the resampling weights stay bounded, and derive MIS weights m_i that fulfill this condition. Denote $Y_i = T_i(X_i)$ and assume that $Y_i \in T_i(\operatorname{supp} X_i)$. The resampling weight of X_i is then, by Equation 19,

$$w_i = m_i(Y_i) \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial Y_i}{\partial X_i}\right|. \tag{S.1}$$

Assuming that $\hat{p}_i(X_i) W_i \le C_i$, we can bound the above as

$$w_i = m_i(Y_i) \cdot \frac{\hat{p}(Y_i)\hat{p}_i(X_i)W_i}{\hat{p}_i(X_i)} \cdot \left|\frac{\partial Y_i}{\partial X_i}\right| \tag{S.2}$$

$$\le m_i(Y_i) \cdot \frac{\hat{p}(Y_i)\, C_i}{\hat{p}_i(X_i)} \cdot \left|\frac{\partial Y_i}{\partial X_i}\right|. \tag{S.3}$$

We require this to be at most some $\tilde{C}_i$, which we may choose freely as long as we still find suitable functions m_i. Then, $w_i \le \tilde{C}_i$ will also hold:

$$w_i \le m_i(Y_i) \cdot \frac{\hat{p}(Y_i)C_i}{\hat{p}_i(X_i)} \cdot \left|\frac{\partial Y_i}{\partial X_i}\right| \le \tilde{C}_i. \tag{S.4}$$

The latter inequality is equivalent to

$$m_i(Y_i) \le \frac{\tilde{C}_i}{C_i} \frac{\hat{p}_i(X_i)\left|\frac{\partial Y_i}{\partial X_i}\right|^{-1}}{\hat{p}(Y_i)}. \tag{S.5}$$

We observe that if j is any canonical index, then we have $\hat{p}(Y_i) = \hat{p}_j(T_j^{-1}(Y_i)) \left|\frac{\partial T_j^{-1}}{\partial Y_i}\right|$, and the numerator and denominator begin to look similar:

$$m_i(Y_i) \le \frac{\tilde{C}_i}{C_i} \frac{\hat{p}_i\left(T_i^{-1}(Y_i)\right)\left|\frac{\partial T_i^{-1}}{\partial Y_i}\right|}{\hat{p}_j(T_j^{-1}(Y_i))\left|\frac{\partial T_j^{-1}}{\partial Y_i}\right|} = \frac{\tilde{C}_i}{C_i} \frac{\hat{p}_{\leftarrow i}(Y_i)}{\hat{p}_{\leftarrow j}(Y_i)}. \tag{S.6}$$

Writing the expressions in terms of $\hat{p}_{\leftarrow i}$ and $\hat{p}_{\leftarrow j}$ is justified since $Y_i = T_i(X_i)$ was sampled (hence $Y_i \in T_i(\operatorname{supp} X_i)$), which implies $Y_i \in \operatorname{supp} \hat{p}$. Since j is a canonical index, $\operatorname{supp} \hat{p} \subset \operatorname{supp} X_j = \operatorname{supp} T_j(\operatorname{supp} X_j)$, and hence $Y_i \in \operatorname{supp} T_j(X_j)$.

If m_i is such that it fulfills the above inequality but with a larger denominator, it will also fulfill the above inequality. We make the denominator of m_i symmetric by summing over *all* indices $j \in \{1, \ldots, M\}$. We additionally choose $\tilde{C}_i = C_i$, leading to

$$m_i(y) = \frac{\hat{p}_{\leftarrow i}(y)}{\sum_{j=1}^{M} \hat{p}_{\leftarrow j}(y)}. \tag{S.7}$$

These m_i fulfill Equation 20 (see the definition of $\hat{p}_{\leftarrow i}$, Equation 35) and are valid, non-negative resampling MIS weights.

S.1.2 Generalizing Pairwise MIS Weights

In order to derive the generalized Pairwise MIS weights, we proceed as before until Equation S.5,

$$m_i(Y_i) \leq \frac{\tilde{C}_i}{C_i} \frac{\hat{p}_{\leftarrow i}(Y_i)}{\hat{p}(Y_i)}, \tag{S.8}$$

but then treat canonical samples $i \in R$ and non-canonical indices $i \notin R$ differently. Instead of including all the indices in the denominator like before, we only increase the denominator by the term corresponding to index i, with a positive multiplier α_i. We set for non-canonical samples

$$m_i(y) = \frac{\tilde{C}_i}{C_i} \frac{\hat{p}_{\leftarrow i}(y)}{\hat{p}(y) + \alpha_i\, \hat{p}_{\leftarrow i}(y)} \quad \text{for } i \notin R, \tag{S.9}$$

and observe that this choice fulfills Equation S.8. We still need m_i to sum to 1 over the i that can generate y, in order to fulfill Equation 20, so we simply divide the remainder uniformly to the canonical samples $i \in R$,

$$m_i(y) = \frac{1}{|R|} \left(1 - \sum_{j \notin R} m_j(y) \right) \quad \text{if i} \in \text{R}. \tag{S.10}$$

Different choices for $\tilde{C}_i$ and α_i yield a family of potential MIS weights: denoting $\beta_i = \tilde{C}_i / C_i$, we reach

$$\begin{aligned} m_i(y) &= \frac{1}{|R|} \left(1 - \sum_{j \notin R} \beta_j \frac{\hat{p}_{\leftarrow j}(y)}{\hat{p}(y) + \alpha_j \hat{p}_{\leftarrow j}(y)} \right) \qquad\qquad (\text{S.11}) \\ &= \frac{1}{|R|} \left(1 - \sum_{j \notin R} \frac{\beta_j}{\alpha_j} + \sum_{j \notin R} \frac{\beta_j}{\alpha_j} \frac{\hat{p}(y)}{\hat{p}(y) + \alpha_j \hat{p}_{\leftarrow j}(y)} \right) \quad \text{if} \in R \end{aligned}$$

$$m_i(y) = \beta_i \frac{\hat{p}_{\leftarrow i}(y)}{\hat{p}(y) + \alpha_i \hat{p}_{\leftarrow i}(y)} \quad \text{if } i \notin R. \tag{S.12}$$

Restricting this family by requiring that the parameters do not depend on i, and denoting $\alpha_i = \alpha$ and $\kappa = \sum_{i \notin R} \beta_i / \alpha_i = (M - |R|)\beta/\alpha$, we reach the family

$$m_i(y) = \frac{1}{|R|} \left(1 - \kappa + \frac{\kappa}{M - |R|} \sum_{j \notin R} \frac{\hat{p}(y)}{\hat{p}(y) + \alpha \hat{p}_{\leftarrow j}(y)} \right) \quad \text{if } i \in R \tag{S.13}$$

$$m_i(y) = \alpha \frac{\kappa}{M - |R|} \frac{\hat{p}_{\leftarrow i}(y)}{\hat{p}(y) + \alpha \hat{p}_{\leftarrow i}(y)} \quad \text{if } i \notin R. \tag{S.14}$$

Since MIS weights must be non-negative and sum to one, we must have $0 \leq m_i \leq 1$ for all i and y. We must generally have $\kappa \leq 1$ since otherwise $m_i(y)$ could be negative in Equation S.13. Since $\alpha > 0$ and we must have $m_i(y) \geq 0$ in Equation S.14, we must also have $0 \leq \kappa$. We have $0 \leq \kappa \leq 1$, and we interpret the above MIS weights as linear interpolation between uniformly choosing one of the canonical samples ($\kappa = 0$, $m_i = 1/|R|$ for $i \in R$, $m_i = 0$ for $i \notin R$), and a fundamental MIS scheme ($\kappa = 1$) parametrized by α:

$$m_i(y) = \frac{1}{|R|} \left(\frac{1}{M - |R|} \sum_{j \notin R} \frac{\hat{p}(y)}{\hat{p}(y) + \alpha \hat{p}_{\leftarrow j}(y)} \right) \quad \text{if } i \in R \tag{S.15}$$

$$m_i(y) = \alpha \frac{1}{M - |R|} \frac{\hat{p}_{\leftarrow i}(y)}{\hat{p}(y) + \alpha \hat{p}_{\leftarrow i}(y)} \quad \text{if } i \notin R. \tag{S.16}$$

The uniform case. To find a sensible value for α for the fundamental MIS scheme, we consider the simple case when all X_i are i.i.d. with $\hat{p}_i = \hat{p}$ for all i. In this case, we have no reason to favor any of the samples and we should have $m_1 = \cdots = m_M = 1/M$, yielding

$$m_i(y) = \frac{1}{|R|} \left(\frac{1}{M - |R|} \sum_{j \notin R} \frac{1}{1 + \alpha} \right) = \frac{1}{M} \quad \text{if } i \in R \tag{S.17}$$

$$m_i(y) = \alpha \frac{1}{M - |R|} \frac{1}{1 + \alpha} = \frac{1}{M} \quad \text{if } i \notin R, \tag{S.18}$$

from which we solve

$$\alpha = \frac{M}{|R|} - 1. \tag{S.19}$$

In the fundamental case $\kappa = 1$, we substitute $\alpha = M/|R| - 1$ and reach

$$m_i(y) = \frac{1}{M - |R|} \sum_{j \notin R} \frac{\hat{p}(y)}{|R|\, \hat{p}(y) + (M - |R|) \hat{p}_{\leftarrow j}(y)} \quad \text{if } i \in R \tag{S.20}$$

$$m_i(y) = \frac{\hat{p}_{\leftarrow i}(y)}{|R|\, \hat{p}(y) + (M - |R|)\, \hat{p}_{\leftarrow i}(y)} \quad \text{if } i \notin R, \tag{S.21}$$

which we call *uniform* Pairwise MIS.

The defensive case. If we instead treat the canonical samples as more reliable than the other samples, we may interpolate the previous solution towards always choosing one of the canonical samples by keeping $\alpha = M/|R| - 1$ and choosing $0 \leq \kappa < 1$. One such a heuristic could be to ensure that the MIS weights of the canonical samples are always at least as large as those of the other samples. With $\alpha = M/|R| - 1$, the canonical $m_i(y)$ cannot be less than $(1 - \kappa)/|R|$ (set $\hat{p}_j(y) \to \infty$ in Equation S.13), and the non-canonical m_i cannot exceed $\kappa/(M - |R|)$ (set $\hat{p}_{\leftarrow i}(y) \to \infty$ in Equation S.14). These bounds can be made equal by choosing $(1 - \kappa)/|R| = \kappa/(M - |R|)$, i.e., $\kappa = (M - |R|)/M$, which gives us the *defensive* generalized Pairwise MIS weights,

$$m_i(y) = \frac{1}{M} + \frac{1}{M} \sum_{j \notin R} \frac{\hat{p}(y)}{|R|\, \hat{p}(y) + (M - |R|)\, \hat{p}_{\leftarrow j}(y)} \quad \text{if } i \in R \tag{S.22}$$

$$m_i(y) = \frac{M - |R|}{M} \frac{\hat{p}_{\leftarrow i}(y)}{|R|\, \hat{p}(y) + (M - |R|)\, \hat{p}_{\leftarrow i}(y)} \quad \text{if } i \notin R. \tag{S.23}$$

S.1.3 Resampling Weight Bounds

We next derive more accurate bounds for the resampling weights for our generalized Talbot and Pairwise MIS weights. In all cases we achieve the same bound, $C_i/|R|$, where $\hat{p}_i(X_i) W_i \leq C_i$.

Generalized Talbot MIS. A direct substitution of the Generalized Talbot MIS weights (Equation S.7) into the formula of w_i (Equation S.1) yields, assuming that Y_i exists (otherwise $w_i = 0$),

$$w_i = m_i(Y_i) \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial T_i}{\partial X_i}\right| \tag{S.24}$$

$$= \left(\frac{\hat{p}_{\leftarrow i}(Y_i)}{\sum_{j=1}^{M} \hat{p}_{\leftarrow j}(Y_i)}\right) \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial T_i}{\partial X_i}\right| \tag{S.25}$$

$$= \frac{\hat{p}_i(X_i) \left|\frac{\partial T_i^{-1}}{\partial Y_i}\right|}{\sum_{j \in R} \hat{p}_{\leftarrow j}(Y_i) + \sum_{j \notin R} \hat{p}_{\leftarrow j}(Y_i)} \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial T_i}{\partial X_i}\right| \tag{S.26}$$

$$= \frac{\hat{p}(Y_i)}{|R|\,\hat{p}(Y_i) + \sum_{j \notin R} \hat{p}_{\leftarrow j}(Y_i)} \cdot \hat{p}_i(X_i) W_i \tag{S.27}$$

$$\leq \frac{1}{|R|} \cdot C_i = \frac{C_i}{|R|}. \tag{S.28}$$

Generalized Pairwise MIS. We first cover the feasible parameter combinations $\alpha > 0$ and $0 \leq \kappa \leq 1$ in one go, assuming that $\hat{p}_i(X_i) W_i \leq C_i$. For canonical samples $i \in R$, we have $\hat{p}_i = \hat{p}$ and $Y_i = T_i(X_i) = X_i$, we substitute Equation S.13 and get the bound

$$w_i = m_i(Y_i) \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial Y_i}{\partial X_i}\right| \tag{S.29}$$

$$= \frac{1}{|R|}\left(1 - \kappa + \frac{\kappa}{M - |R|} \sum_{j \notin R} \frac{\hat{p}(Y_i)}{\hat{p}(Y_i) + \alpha \hat{p}_{\leftarrow j}(Y_i)}\right) \cdot \hat{p}(Y_i) W_i \cdot 1 \tag{S.30}$$

$$\leq \frac{1}{|R|}\left(1 - \kappa + \frac{\kappa}{M - |R|} \sum_{j \notin R} \frac{\hat{p}(Y_i)}{\hat{p}(Y_i)}\right) \cdot \hat{p}(Y_i) W_i \tag{S.31}$$

$$\leq \frac{C_i}{|R|}. \tag{S.32}$$

For non-canonical samples $i \notin R$, assuming again $\hat{p}_i(X_i) W_i \leq C_i$, substituting Equation S.14, we get the bound

$$w_i = m_i(Y_i) \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial Y_i}{\partial X_i}\right| \tag{S.33}$$

$$= \left(\alpha \frac{\kappa}{M - |R|} \frac{\hat{p}_{\leftarrow i}(Y_i)}{\hat{p}(Y_i) + \alpha \hat{p}_{\leftarrow i}(Y_i)}\right) \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial Y_i}{\partial X_i}\right| \tag{S.34}$$

$$= \alpha \frac{\kappa}{M - |R|} \frac{\hat{p}_i(X_i)}{\hat{p}(Y_i) + \alpha \hat{p}_{\leftarrow i}(Y_i)} \cdot \hat{p}(Y_i) W_i \tag{S.35}$$

$$\leq \alpha \frac{\kappa}{M - |R|} C_i. \tag{S.36}$$

$$\tag{S.37}$$

In the case that $\alpha \leq M/|R| - 1$, we simplify

$$w_i \leq \kappa \frac{C_i}{|R|} \leq \frac{C_i}{|R|}. \tag{S.38}$$

S.1.4 Tractable Marginal PDFs

Sometimes the PDFs of the input samples X_i are tractable functions p_i. In that case, the PDFs p_i may be used in place of the $\hat{p}_i$ in the MIS weight formulas, effectively replacing $\hat{p}_{\leftarrow i}$ with the following "p from i":

$$p_{\leftarrow i}(y) = \begin{cases} p_i\left(T_i^{-1}(y)\right) \left|T_i^{-1\prime}\right|(y), & \text{if } y \in \mathcal{D}(T_i^{-1}) \\ 0 & \text{otherwise} \end{cases}, \tag{S.39}$$

resulting in the following expression for the generalized Talbot MIS:

$$m_i(y) = \frac{p_{\leftarrow i}(Y_i)}{\sum_{j=1}^{M} p_{\leftarrow j}(Y_i)}. \tag{S.40}$$

The Pairwise MIS expression additionally contains terms $\hat{p}(y)$ whose normalization may differ significantly from that of the PDFs p_i. As such, we suggest replacing the terms $\hat{p}(y)$ in the MIS with a fixed canonical importance sampler $c \in C$ that is reasonable for integrating $\hat{p}$ ($\hat{p}(y) \leq C_c p_c(y)$). We show the uniform case as an example of this translation to known PDFs:

$$m_i(y) = \frac{1}{M - |R|} \sum_{j \notin R} \frac{p_c(y)}{|R|\, p_c(y) + (M - |R|) p_{\leftarrow j}(y)} \quad \text{if } i \in R \tag{S.41}$$

$$m_i(y) = \frac{\hat{p}_{\leftarrow i}(y)}{|R|\, p_c(y) + (M - |R|)\, p_{\leftarrow i}(y)} \quad \text{if } i \notin R. \tag{S.42}$$

We then derive the resampling weight bounds for these updated formulas. Since the p_i are tractable, we assume unbiased contribution weights $W_i = 1/p_i(X_i)$. We also assume that the canonical samples are reasonably importance sampled for $\hat{p}$, i.e., $\hat{p}(x) \leq C_i p_i(x)$ for all $i \in R$.

Talbot MIS. Substituting Equation S.40 into Equation S.1, yields, remembering that $p_{\leftarrow j}(y) = p_j(y)$ for canonical j,

$$w_i = m_i(Y_i) \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial T_i}{\partial X_i}\right| \tag{S.43}$$

$$= \left(\frac{p_{\leftarrow i}(Y_i)}{\sum_{j=1}^{M} p_{\leftarrow j}(Y_i)}\right) \cdot \frac{\hat{p}(Y_i)}{p_i(X_i)} \cdot \left|\frac{\partial T_i}{\partial X_i}\right| \tag{S.44}$$

$$= \frac{p_i(X_i)}{\sum_{j \in R} p_{\leftarrow j}(Y_i) + \sum_{j \notin R} p_{\leftarrow j}(Y_i)} \cdot \frac{\hat{p}(Y_i)}{p_i(X_i)} \tag{S.45}$$

$$= \frac{\hat{p}(Y_i)}{\sum_{j \in R} p_j(Y_i) + \sum_{j \notin R} p_{\leftarrow j}(Y_i)} \tag{S.46}$$

$$\leq \frac{\hat{p}(Y_i)}{\sum_{j \in R} p_j(Y_i)} \leq \frac{\hat{p}(Y_i)}{|R| \min_{j \in R} p_j(Y_i)} \tag{S.47}$$

$$= \frac{1}{|R|} \max_{j \in R} \frac{\hat{p}(Y_i)}{p_j(Y_i)} \leq \frac{1}{|R|} \max_{j \in R} C_j. \tag{S.48}$$

Pairwise MIS. We now derive bounds for the resampling weights in the case of Generalized Pairwise MIS weights with $0 \leq \alpha \leq M/|R| - 1$ and $0 \leq \kappa \leq 1$, using p_i instead of $\hat{p}_i$.

If i is a canonical index, we use Equation S.13 with $\hat{p}_{\leftarrow j}$ replaced with $p_{\leftarrow j}$ and $\hat{p}$ replaced with p_c. Noting that for canonical indices

$Y_i = T_i(X_i) = X_i$, we reach

$$w_i = m_i(Y_i) \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial T_i}{\partial X_i}\right| \tag{S.49}$$

$$= \frac{1}{|R|}\left(1 - \kappa + \frac{\kappa}{M - |R|} \sum_{j \notin R} \frac{p_c(Y_i)}{p_c(Y_i) + \alpha p_{\leftarrow j}(Y_i)}\right) \cdot \frac{\hat{p}(Y_i)}{p_i(X_i)} \cdot 1 \tag{S.50}$$

$$\leq \frac{1}{|R|}\left(1 - \kappa + \frac{\kappa}{M - |R|} \sum_{j \notin R} 1\right) \cdot \frac{\hat{p}(X_i)}{p_i(X_i)} \tag{S.51}$$

$$= \frac{1}{|R|} \cdot \frac{\hat{p}(X_i)}{p_i(X_i)} \leq \frac{C_i}{|R|}. \tag{S.52}$$

Similarly, for non-canonical i we use Equation S.14 with the same substitutions and reach

$$w_i = m_i(Y_i) \cdot \hat{p}(Y_i) W_i \cdot \left|\frac{\partial T_i}{\partial X_i}\right| \tag{S.53}$$

$$= \left(\alpha \frac{\kappa}{M - |R|} \frac{p_{\leftarrow i}(Y_i)}{p_c(Y_i) + \alpha p_{\leftarrow i}(Y_i)}\right) \cdot \frac{\hat{p}(Y_i)}{p_i(X_i)} \cdot \left|\frac{\partial T_i}{\partial X_i}\right| \tag{S.54}$$

$$= \left(\alpha \frac{\kappa}{M - |R|} \frac{p_i(X_i) \left|\frac{\partial T_i^{-1}}{\partial Y_i}\right|}{p_c(Y_i) + \alpha p_{\leftarrow i}(Y_i)}\right) \cdot \frac{\hat{p}(Y_i)}{p_i(X_i)} \cdot \left|\frac{\partial T_i}{\partial X_i}\right| \tag{S.55}$$

$$= \alpha \frac{\kappa}{M - |R|} \frac{\hat{p}(Y_i)}{p_c(Y_i) + \alpha p_{\leftarrow i}(Y_i)} \leq \alpha \frac{\kappa}{M - |R|} \frac{\hat{p}(Y_i)}{p_c(Y_i)} \tag{S.56}$$

$$\leq \alpha \frac{\kappa}{M - |R|} C_c \leq \left(\frac{M}{|R|} - 1\right) \frac{1}{M - |R|} C_c = \frac{C_c}{|R|}. \tag{S.57}$$

We can combine the above results into a single, slightly looser bound that works for any i and both MIS weight families (Talbot and Pairwise) when used with tractable PDFs:

$$w_i \leq \frac{1}{|R|} \max_{j \in R} C_j. \tag{S.58}$$

S.2 CONVERGENCE WITH DEPENDENT SAMPLES

In Section 5.7, we assume dependent input samples such that

(1) the ratio of canonical samples, $|R|/M$, never falls below a positive constant γ for large enough M,
(2) there exists a $C > 0$ such that $w_i \leq C/|R|$ for all i,
(3) there exists a non-negative sequence b_k such that the correlation $\rho_{i,i+k} \leq b_k$ for all i, and $b_k \to 0$.

Then,

$$\mathrm{Var}\left[\sum_{i=1}^{M} w_i\right] = \sum_{i=1}^{M} \mathrm{Var}\left[w_i\right] + 2 \sum_{i=1}^{M} \sum_{k=1}^{M-i} \mathrm{Cov}(w_i, w_{i+k}) \tag{S.59}$$

converges to zero:

The convergence of the first term is proved in Section 5.7, and for the second term we have

$$\sum_{i=1}^{M} \sum_{k=1}^{M-i} \mathrm{Cov}(w_i, w_{i+k}) = \sum_{i=1}^{M} \sum_{k=1}^{M-i} \rho_{i,i+k} \sqrt{\mathrm{Var}\, w_i \ \mathrm{Var}\, w_{i+k}} \tag{S.60}$$

$$\leq \sum_{i=1}^{M} \sum_{k=1}^{M-i} \max(0, \rho_{i,i+k}) \frac{C^2}{4|R|^2} \leq \sum_{i=1}^{M} \sum_{k=1}^{M-i} b_k \frac{C^2}{4M^2\gamma^2}, \tag{S.61}$$

$$= \frac{C^2}{4\gamma^2} \frac{1}{M^2} \sum_{k=1}^{M} \sum_{i=1}^{M-k} b_k = \frac{C^2}{4\gamma^2} \frac{1}{M^2} \sum_{k=1}^{M} (M-k) b_k \tag{S.62}$$

$$\leq \frac{C^2}{4\gamma^2} \left(\frac{1}{M} \sum_{k=1}^{M} b_k\right) \xrightarrow{M \to \infty} 0. \tag{S.63}$$

To reach Equation S.61, we used Popoviciu's inequality: since $0 \leq w_i \leq C/|R|$, we know $\mathrm{Var}\, w_i \leq \frac{C^2}{4|R|^2}$. The next step used $|R|/M \geq \gamma$, and for Equation S.62 we reversed the summation order: $\sum_{i=1}^{M} \sum_{k=1}^{M-i} = \sum_{k=1}^{M} \sum_{i=1}^{M-k}$. The mean of b_k converges to zero since b_k converges to zero, and Equation S.63 implies

$$\mathrm{Var}\left[\sum_{i=1}^{M} w_i\right] \xrightarrow{M \to \infty} 0.$$

We can slightly generalize the result: assume that

$$|R| \geq c_M M^{0.5} \sqrt{\sum_{i=1}^{M} b_i}\,, \tag{S.64}$$

where (c_M) is any non-negative sequence that approaches infinity. Then, like above,

$$\sum_{i=1}^{M} \sum_{k=1}^{M-i} \mathrm{Cov}(w_i, w_{i+k}) \leq \sum_{i=1}^{M} \sum_{k=1}^{M-i} \max(0, \rho_{i,i+k}) \frac{C^2}{4|R|^2} \tag{S.65}$$

$$= \sum_{k=1}^{M} \sum_{i=1}^{M-k} \max(0, \rho_{i,i+k}) \frac{C^2}{4|R|^2} \leq \sum_{k=1}^{M} (M-k) b_k \frac{C^2}{4|R|^2} \tag{S.66}$$

$$\leq \left(\sum_{k=1}^{M} b_k\right) \frac{MC^2}{4|R|^2} \leq \left(\sum_{k=1}^{M} b_k\right) \frac{MC^2}{4c_M^2 M \left(\sum_{i=1}^{M} b_i\right)} \tag{S.67}$$

$$= \frac{C^2}{4c_M^2} \xrightarrow{M \to \infty} 0. \tag{S.68}$$

S.3 PRIMARY SAMPLE SPACE

Performing integration in a Monte Carlo setting typically starts from primary sample sequences, i.e., streams of random numbers $\bar{\mathbf{U}} = (U_1, U_2, \ldots)$, where $U_i \in [0, 1)$ are uniformly distributed. Each $\bar{\mathbf{U}}$ is used to estimate a contribution $F(\bar{\mathbf{U}})$, and the Monte Carlo integration result is

$$I = \mathbb{E}\left[F(\bar{\mathbf{U}})\right] = \int_{\mathcal{U}} F(\bar{\mathbf{u}})\, \mathrm{d}\bar{\mathbf{u}}. \tag{S.69}$$

A unidirectional path tracer builds a sequence of paths of different lengths from the random sequences $\bar{\mathbf{U}}$. Often, the path tracer produces, for each length d, N paths $X_{d,n} \in \Omega_d$ by different strategies. Here, Ω_d is the space of all paths of length d. N is often 2, and the paths $X_{d,n}$ with different n correspond to a next-event-estimation path connected to a random light, and path continued to a direction importance sampled according to the BSDF. The paths $X_{d,n}$ are functions of $\bar{\mathbf{U}}$, i.e., $X_{d,n} = x_{d,n}(\bar{\mathbf{U}})$.

The total path contribution is the sum of the integrals of the fixed-length path contributions,

$$I = \sum_{d=1}^{\infty} \int_{\Omega_d} f(x)\, \mathrm{d}x. \tag{S.70}$$

We factor in the MIS weights $\omega_{d,n}$ of the n sampling strategies and reach

$$I = \sum_{d=1}^{\infty} \int_{\Omega_d} \left[\sum_{n=1}^{N} \omega_{d,n}(x)\right] f(x)\, \mathrm{d}x \tag{S.71}$$

$$= \sum_{d=1}^{\infty} \sum_{n=1}^{N} \int_{\Omega_d} \omega_{d,n}(x) f(x)\, \mathrm{d}x. \tag{S.72}$$

We assume for each term a proper importance sampler that produces paths $X_{d,n} \in \Omega_d$ with density $p_{d,n}$. We assume that $\omega_{d,n}(x) = 0$ whenever $p_{d,n}(x) = 0$ to retain the partition of unity. This yields us

$$I = \sum_{d=1}^{\infty} \sum_{n=1}^{N} \mathbb{E}\left[\omega_{d,n}(X_{d,n}) \frac{f(X_{d,n})}{p_{d,n}(X_{d,n})}\right]. \tag{S.73}$$

Since the $X_{d,n}$ are generated from the random variables $\bar{\mathbf{U}}$ by $X_{d,n} = x_{d,n}(\bar{\mathbf{U}})$, we may write

$$I = \sum_{d=1}^{\infty} \sum_{n=1}^{N} \mathbb{E}\left[\omega_{d,n}\left(x_{d,n}(\bar{\mathbf{U}})\right) \frac{f\left(x_{d,n}(\bar{\mathbf{U}})\right)}{p_{d,n}\left(x_{d,n}(\bar{\mathbf{U}})\right)}\right]. \tag{S.74}$$

The fact that many $\bar{\mathbf{U}}$ may lead to the same path $X_{d,n}$ does not complicate this fact. We then write the expectations as integrals and reach

$$I = \sum_{d=1}^{\infty} \sum_{n=1}^{N} \int_{\mathcal{U}} \omega_{d,n}\left(x_{d,n}(\bar{\mathbf{u}})\right) \frac{f\left(x_{d,n}(\bar{\mathbf{u}})\right)}{p_{d,n}\left(x_{d,n}(\bar{\mathbf{u}})\right)}\, \mathrm{d}\bar{\mathbf{u}} \tag{S.75}$$

$$= \int_{\mathcal{U}} \sum_{d=1}^{\infty} \sum_{n=1}^{N} \omega_{d,n}\left(x_{d,n}(\bar{\mathbf{u}})\right) \frac{f\left(x_{d,n}(\bar{\mathbf{u}})\right)}{p_{d,n}\left(x_{d,n}(\bar{\mathbf{u}})\right)}\, \mathrm{d}\bar{\mathbf{u}}. \tag{S.76}$$

This yields us the F for Equation S.69,

$$F(\bar{\mathbf{u}}) = \sum_{d=1}^{\infty} \sum_{n=1}^{N} \omega_{d,n}\left(x_{d,n}(\bar{\mathbf{u}})\right) \frac{f\left(x_{d,n}(\bar{\mathbf{u}})\right)}{p_{d,n}\left(x_{d,n}(\bar{\mathbf{u}})\right)}. \tag{S.77}$$

Algorithm 1: Content of the reservoir struct (88 bytes).

```
struct Reservoir
    float M; // Confidence weight (for e.g., M-capping).
    float W; // Unbiased contribution weight.
    float3 F; // Cached integrand value of the sample.
    uint pathFlags; // Path length, technique type, reconn. vertex id, etc.
    uint initRandomSeed; // Random state at primary hit x̄1.
    // Information about the reconnection vertex (rc):
    uint rcVertexRandomSeed; // Random state at reconn. vertex.
    uint rcVertexInstanceID; // Hit point information:
    uint rcVertexPrimitiveIndex;
    float2 rcVertexBarycentrics;
    float3 rcVertexWi; // Direction to next vertex of base path.
    float3 rcVertexRadiance; // Incident radiance from next vertex.
    float4 rcVertexCachedValues; // Various partial terms for
        evaluating the Jacobian at reconnection. Light sampling PDF for the
        MIS weight is also stored here.
```

S.4 RESERVOIR STORAGE

We provide an overview of our reservoir data structure in Algorithm 1. A key takeaway is that most of the storage is used for enabling a reconnection to the base path's vertex: reconnection requires evaluating offset path's visibility to the reconnection vertex and the BSDF towards base path's next vertex. Note that our reservoir data structure is unoptimized and highly compressible–real-time use would allow lossy compression for increased performance, but our prototype implementation does not do it.

S.5 PROOFS OF THEOREMS

S.5.1 Unbiased Contribution Weights (Theorem A.1)

PROOF OF THEOREM A.1. Assume Item 1: W and X are such that $\mathbb{E}[f(X)W] = \int_{\operatorname{supp} X} f(x)\,\mathrm{d}x$ for any integrable f. We prove Item 2: Let $A \subset \operatorname{supp}(X)$ be measurable. Then,

$$\begin{aligned}\int_A p_X(x)\,\mathrm{d}x &= \int_{\operatorname{supp}(X)} \mathbb{1}_A(x) p_X(x)\,\mathrm{d}x \\ &= \mathbb{E}\left[\mathbb{1}_A(X) p_X(X) W\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_A(X) p_X(X) W \mid X\right]\right] \\ &= \mathbb{E}\left[\mathbb{1}_A(X) p_X(X)\,\mathbb{E}\left[W \mid X\right]\right] \\ &= \int_A p_X(x)^2\,\mathbb{E}\left[W \mid X = x\right]\mathrm{d}x.\end{aligned} \tag{S.78}$$

Since this holds for all measurable $A \subset \operatorname{supp}(X)$, we must have, almost everywhere in $\operatorname{supp}(X)$,

$$p_X(x) = p_X(x)^2\,\mathbb{E}\left[W \mid X = x\right]. \tag{S.79}$$

Since $p_X(x) > 0$ in $\operatorname{supp}(X)$, we deduce $\mathbb{E}\left[W \mid X = x\right] = 1/p_X(x)$ a.e. in $\operatorname{supp}(X)$.

Next, assume Item 2: W and X are such that $\mathbb{E}\left[W \mid X = x\right] = 1/p_X(x)$ a.e. in $\operatorname{supp}(X)$. We prove Item 1: Let $f : \Omega \to \mathbb{R}$ be integrable. We have

$$\begin{aligned}\mathbb{E}\left[f(X)W\right] &= \mathbb{E}\left[\mathbb{E}\left[f(X)W \mid X\right]\right] = \mathbb{E}\left[f(X)\,\mathbb{E}\left[W \mid X\right]\right] \\ &= \mathbb{E}\left[\frac{f(X)}{p_X(X)}\right] = \int_{\operatorname{supp}(X)} f(x)\,\mathrm{d}x.\end{aligned} \tag{S.80}$$

□

S.5.2 Asymptotic Sample Distribution (Theorem A.2)

We start from Equation 22 and derive the equality

$$\sum_{i=1}^{M} w_i = \hat{p}(Y) W_Y. \tag{S.81}$$

We then prove the following, slightly more general result: If $(Y_M)_{M=M_0}^{\infty}$ is a sequence of random variables that fulfill $\operatorname{supp}\hat{p} \subset \operatorname{supp} Y_M$ (Equation 57) and have non-negative unbiased contribution weights W_{Y_M} such that

$$\operatorname{Var}\left[\hat{p}(Y_M) W_{Y_M}\right] \xrightarrow{M\to\infty} 0 \tag{S.82}$$

(a generalization of Equation 58), then the conclusions of Theorem A.2 hold:

PROOF OF THEOREM A.2 (ITEM 1), EQUATION 59. Let $\varepsilon > 0$ be given. We prove that

$$\Pr\left[|p_Y(Y) - \bar{p}(Y)| > \varepsilon\right] \xrightarrow{M\to\infty} 0:$$

For any $0 < \varepsilon_2 < 1$ we have

$$\Pr\left[|\bar{p}(Y) - p_Y(Y)| > \varepsilon\right] = A + B + C, \quad \text{where} \tag{S.83}$$

$$A = \Pr\left[\left(|\bar{p}(Y) - p_Y(Y)| > \varepsilon\right) \wedge \left(p_Y(Y) \geq \frac{1}{\varepsilon_2}\right)\right],$$

$$B = \Pr\left[\left(|\bar{p}(Y) - p_Y(Y)| > \varepsilon\right) \wedge \left(1 < p_Y(Y) < \frac{1}{\varepsilon_2}\right)\right],$$

$$C = \Pr\left[\left(|\bar{p}(Y) - p_Y(Y)| > \varepsilon\right) \wedge \left(p_Y(Y) \leq 1\right)\right].$$

Since $\Pr\left[X \wedge Y\right] \leq \Pr\left[Y\right]$ and p_Y integrates to 1, we have

$$A \leq \Pr\left[p_Y(Y) \geq \frac{1}{\varepsilon_2}\right] \leq \varepsilon_2$$

(otherwise p_Y would integrate to more than $\varepsilon_2 \cdot 1/\varepsilon_2 = 1$). Since in case B we have $1/p_Y(Y) > \varepsilon_2$, we get

$$\begin{aligned}B &= \Pr\left[\left(|\bar{p}(Y) - p_Y(Y)|\,\varepsilon_2 > \varepsilon \cdot \varepsilon_2\right) \wedge \left(1 < p_Y(Y) < \frac{1}{\varepsilon_2}\right)\right] \\ &\leq \Pr\left[\left(\frac{|\bar{p}(Y) - p_Y(Y)|}{p_Y(Y)} > \varepsilon \cdot \varepsilon_2\right) \wedge \left(1 < p_Y(Y) < \frac{1}{\varepsilon_2}\right)\right] \\ &\leq \Pr\left[\left|\frac{\bar{p}(Y)}{p_Y(Y)} - 1\right| > \varepsilon \cdot \varepsilon_2\right].\end{aligned}$$

Similarly, in case C we have $1/p_Y(Y) \geq 1$, and thus

$$\begin{aligned}C &= \Pr\left[\left(|\bar{p}(Y) - p_Y(Y)| > \varepsilon\right) \wedge \left(p_Y(Y) \leq 1\right)\right] \\ &\leq \Pr\left[\left|\frac{\bar{p}(Y)}{p_Y(Y)} - 1\right| > \varepsilon\right].\end{aligned}$$

By Chebyshev's inequality, we have for any $s > 0$ (such as $s = \varepsilon \cdot \varepsilon_2$ for B and $s = \epsilon$ for C), that

$$\Pr\left[\left|\frac{\bar{p}(Y)}{p_Y(Y)} - 1\right| > s\right] < \frac{1}{s^2}\,\mathbb{E}\left[\left|\frac{\bar{p}(Y)}{p_Y(Y)} - 1\right|^2\right] \xrightarrow{M\to\infty} 0, \tag{S.84}$$

and thus

$$0 \leq \lim_{M\to\infty} A + B + C \leq \varepsilon_2 + 0 + 0 \xrightarrow{\varepsilon_2 \to 0} 0, \tag{S.85}$$

i.e.,

$$\Pr\left[|\bar{p}(Y) - p_Y(Y)| > \varepsilon\right] \xrightarrow{M\to\infty} 0. \tag{S.86}$$

■

PROOF OF THEOREM A.2 (ITEM 2), EQUATION 60. By assumption, we have $\operatorname{supp}\hat{p} \subset \operatorname{supp} Y_M$ for each M. Dropping the index M for brevity, we thus have

$$\mathbb{E}\left[\frac{\hat{p}(Y)}{p_Y(Y)}\right] = \int_{\operatorname{supp} Y} \hat{p}(y)\,\mathrm{d}y = \|\hat{p}\|. \tag{S.87}$$

Thus, we deduce from the law of total variance that

$$\begin{aligned}\operatorname{Var}\left[\hat{p}(Y) W_Y\right] &= \mathbb{E}\left[\operatorname{Var}\left[\hat{p}(Y) W_Y \mid Y\right]\right] + \operatorname{Var}\left[\mathbb{E}\left[\hat{p}(Y) W_Y \mid Y\right]\right] \\ &\geq \operatorname{Var}\left[\mathbb{E}\left[\hat{p}(Y) W_Y \mid Y\right]\right] = \operatorname{Var}\left[\frac{\hat{p}(Y)}{p_Y(Y)}\right] \\ &= \mathbb{E}\left[\left|\frac{\hat{p}(Y)}{p_Y(Y)} - \|\hat{p}\|\right|^2\right].\end{aligned} \tag{S.88}$$

Since $\operatorname{Var}\left[\hat{p}(Y) W_Y\right]$ by assumption tends to 0, we reach convergence of p_Y to $\bar{p}$ in mean square:

$$\begin{aligned}\mathbb{E}\left[\left|\frac{\bar{p}(Y)}{p_Y(Y)} - 1\right|^2\right] &= \frac{1}{\|\hat{p}\|^2}\,\mathbb{E}\left[\left|\frac{\hat{p}(Y)}{p_Y(Y)} - \|\hat{p}\|\right|^2\right] \\ &\leq \frac{1}{\|\hat{p}\|^2}\operatorname{Var}\left[\hat{p}(Y) W_Y\right] \xrightarrow{M\to\infty} 0.\end{aligned} \tag{S.89}$$

■

Proof of Theorem A.2 (Item 3), Equation 61. By the Cauchy-Schwarz inequality, convergence of a random variable in mean-square implies convergence in mean:

$$\mathbb{E}\left[|Z_i - Z_\infty|\right] \leq \sqrt{\mathbb{E}\left[1^2\right]}\sqrt{\mathbb{E}\left[|Z_i - Z_\infty|^2\right]} \xrightarrow{i\to\infty} 0. \tag{S.90}$$

By Equation 60, $\mathbb{E}\left[\left|\frac{\bar{p}(Y)}{p_Y(Y)} - 1\right|\right] \xrightarrow{M\to\infty} 0$. Since we have $\operatorname{supp}\hat{p} \subset \operatorname{supp} Y$, we have $p_Y(y) - \hat{p}(y) = 0 - 0$ outside of $\operatorname{supp} Y$ for all M, and thus

$$\begin{aligned}\int_\Omega |p_Y(y) - \bar{p}(y)|\,\mathrm{d}y &= \int_{\operatorname{supp} Y} |p_Y(y) - \bar{p}(y)|\,\mathrm{d}y \\ &= \mathbb{E}\left[\left|\frac{\bar{p}(Y)}{p_Y(Y)} - 1\right|\right] \xrightarrow{M\to\infty} 0. \end{aligned}\tag{S.91}$$

■

Proof of Theorem A.2 (Item 4). Let G be the set of $y \in \Omega$ for which $p_{Y_M}(y)$ converges, and let $g(y)$ be the limit, i.e.,

$$g(y) = \lim_{M\to\infty} p_{Y_M}(y) \quad \text{for all } y \in G. \tag{S.92}$$

All subsequences $p_{Y_{a_k}}(y)$ also converge pointwise to $g(y)$ in G. By Equation 61, $p_Y(y)$ converges to $\bar{p}$ in the Lebesgue L^1 sense:

$$\|\bar{p}(y) - p_{Y_M}(y)\|_{L^1} = \int_\Omega \left|\bar{p}(y) - p_{Y_M}(y)\right|\mathrm{d}y \xrightarrow{M\to\infty} 0. \tag{S.93}$$

Since p_{Y_M} converges to $\bar{p}$ in the L^1-norm, it converges also in the L^1-measure [Bartle 2014, p. 69]. Thus, it has a subsequence $p_{Y_{a_k}}$ that converges to $\bar{p}$ almost everywhere [Bartle 2014, p. 69]. But if $y \in G$, then $p_{Y_{a_k}}(y)$ also converges to $g(y)$, so we must have $g(y) = \bar{p}(y)$ almost everywhere in G. ■

Proof of Theorem A.2 (Item 5). Let X be distributed with PDF $\bar{p}$ and A be an arbitrary measurable subset of Ω. Then, by Theorem A.2 (Item 3),

$$\begin{aligned}|\Pr[Y \in A] - \Pr[X \in A]| &= \left|\int_A p_Y(y)\,\mathrm{d}y - \int_A \bar{p}(y)\,\mathrm{d}y\right| \\ &\leq \int_\Omega |p_Y(y) - \bar{p}(y)|\,\mathrm{d}y \xrightarrow{M\to\infty} 0.\end{aligned}$$

■

S.5.3 Asymptotic Variance (Theorem A.3)

We first prove the support and then Item 1 – Item 3:

Proof of Theorem A.3, support. By assumption (Equation 57), $\operatorname{supp}\hat{p} \subset \operatorname{supp} Y_M$ for all M. We also assume $f \leq C_f \hat{p}$ for some $C_f > 0$. Thus, $f(x) > 0$ implies $\hat{p}(x) > 0$, and we have $\operatorname{supp} f \subset \operatorname{supp}\hat{p} \subset \operatorname{supp} Y_M$. ■

Proof of Theorem A.3 (Item 1), Equation 63 and Equation 64. We first prove convergence in mean square:

$$\begin{aligned}\mathbb{E}\left[\left|f(Y)W_Y - \frac{f(Y)}{\bar{p}(Y)}\right|^2\right] &= \mathbb{E}\left[\left|\frac{f(Y)}{\hat{p}(Y)}\sum_{i=1}^M w_{M,i} - \frac{f(Y)}{\bar{p}(Y)}\right|^2\right] \\ &= \mathbb{E}\left[\frac{f(Y)^2}{\hat{p}(Y)^2}\left|\sum_{i=1}^M w_{M,i} - \frac{\hat{p}(Y)}{\bar{p}(Y)}\right|^2\right] = \mathbb{E}\left[\frac{f(Y)^2}{\hat{p}(Y)^2}\left|\sum_{i=1}^M w_{M,i} - \|\hat{p}\|\right|^2\right] \\ &\leq C_f^2\,\mathbb{E}\left[\left|\sum_{i=1}^M w_{M,i} - \|\hat{p}\|\right|^2\right] = C_f^2 \operatorname{Var}\left[\sum_{i=1}^M w_{M,i}\right] \xrightarrow{M\to\infty} 0.\end{aligned}$$

Convergence in mean square implies convergence in mean and in probability. E.g., by Chebyshev's inequality, given $\varepsilon > 0$, we have

$$\Pr\left[\left|f(Y)W_Y - \frac{f(Y)}{\bar{p}(Y)}\right| > \varepsilon\right] \leq \frac{1}{\varepsilon^2}\,\mathbb{E}\left[\left|f(Y)W_Y - \frac{f(Y)}{\bar{p}(Y)}\right|^2\right] \xrightarrow{M\to\infty} 0.$$

■

Proof of Theorem A.3 (Item 2), Equation 65. We proceed in three steps.

Step 1. We show that

$$\operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] \xrightarrow{M\to\infty} 0. \tag{S.94}$$

From the law of total variance we get

$$\operatorname{Var}\big[f(Y)W_Y\big] = \operatorname{Var}\big[\mathbb{E}\left[f(Y)W_Y|Y\right]\big] + \mathbb{E}\big[\operatorname{Var}\left[f(Y)W_Y \mid Y\right]\big],$$

which we rewrite as

$$\begin{aligned}\mathbb{E}\big[\operatorname{Var}\left[f(Y)W_Y \mid Y\right]\big] &= \operatorname{Var}\big[f(Y)W_Y\big] - \operatorname{Var}\big[\mathbb{E}\left[f(Y)W_Y|Y\right]\big] \\ &= \operatorname{Var}\big[f(Y)W_Y\big] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right].\end{aligned}$$

Since a conditional variance is non-negative, and we have

$$\begin{aligned}0 &\leq \operatorname{Var}\big[f(Y)W_Y\big] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] = \mathbb{E}\big[\operatorname{Var}\left[f(Y)W_Y \mid Y\right]\big] \\ &= \mathbb{E}\left[\operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\sum_{i=1}^M w_{M,i} \mid Y\right]\right] = \mathbb{E}\left[\frac{f(Y)^2}{\hat{p}(Y)^2}\operatorname{Var}\left[\sum_{i=1}^M w_{M,i} \mid Y\right]\right] \\ &\leq C_f^2\,\mathbb{E}\left[\operatorname{Var}\left[\sum_{i=1}^M w_{M,i} \mid Y\right]\right] \\ &\leq C_f^2\left(\mathbb{E}\left[\operatorname{Var}\left[\sum_{i=1}^M w_{M,i} \mid Y\right]\right] + \operatorname{Var}\left[\mathbb{E}\left[\sum_{i=1}^M w_{M,i} \mid Y\right]\right]\right) \\ &= C_f^2 \operatorname{Var}\left[\sum_{i=1}^M w_{M,i}\right] \xrightarrow{M\to\infty} 0.\end{aligned}\tag{S.95}$$

Step 2. We show that

$$\operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] \xrightarrow{M\to\infty} 0 \tag{S.96}$$

where X has density $\bar{p}(x)$. We start by writing

$$\begin{aligned}
&\operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] \\
&= \mathbb{E}\left[\frac{f(X)^2}{\bar{p}(X)^2}\right] - \mathbb{E}\left[\frac{f(X)}{\bar{p}(X)}\right]^2 - \mathbb{E}\left[\frac{f(Y)^2}{p_Y(Y)^2}\right] + \mathbb{E}\left[\frac{f(Y)}{p_Y(Y)}\right]^2 \qquad \text{(S.97)} \\
&= \mathbb{E}\left[\frac{f(X)^2}{\bar{p}(X)^2}\right] - \mathbb{E}\left[\frac{f(Y)^2}{p_Y(Y)^2}\right]. \qquad \text{(S.98)}
\end{aligned}$$

Since $\operatorname{supp} Y \subset \operatorname{supp}\hat{p}$ (Equation 15) and we assume $\operatorname{supp}\hat{p} \subset \operatorname{supp} Y$, we have $\operatorname{supp}\hat{p} = \operatorname{supp} Y$. With $\operatorname{supp}\bar{p} = \operatorname{supp}\hat{p}$, we continue the above as

$$\begin{aligned}
&= \int_{\operatorname{supp} Y} \bar{p}(x)\frac{f(x)^2}{\bar{p}(x)^2}\,\mathrm{d}x - \int_{\operatorname{supp} Y} p_Y(y)\frac{f(y)^2}{p_Y(y)^2}\,\mathrm{d}y \\
&= \int_{\operatorname{supp} Y} \bar{p}(y)\frac{f(y)^2}{\bar{p}(y)^2} - p_Y(y)\frac{f(y)^2}{p_Y(y)^2}\,\mathrm{d}y \\
&= \int_{\operatorname{supp} Y} p_Y(y)\left(\frac{f(y)^2}{\bar{p}(y)^2}\right)\left(\frac{\bar{p}(y)}{p_Y(y)} - \frac{\bar{p}(y)^2}{p_Y(y)^2}\right)\mathrm{d}y \\
&= \mathbb{E}\left[\frac{f(Y)^2}{\bar{p}(Y)^2}\left(\frac{\bar{p}(Y)}{p_Y(Y)} - \frac{\bar{p}(Y)^2}{p_Y(Y)^2}\right)\right],
\end{aligned}$$

and thus,

$$\begin{aligned}
&\left|\operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right]\right| \\
&\le \mathbb{E}\left[\frac{f(Y)^2}{\bar{p}(Y)^2}\left|\frac{\bar{p}(Y)}{p_Y(Y)} - \frac{\bar{p}(Y)^2}{p_Y(Y)^2}\right|\right] \\
&\le \|\hat{p}\|^2 C_f^2\,\mathbb{E}\left[\left|\frac{\bar{p}(Y)}{p_Y(Y)} - \frac{\bar{p}(Y)^2}{p_Y(Y)^2}\right|\right] \\
&= \|\hat{p}\|^2 C_f^2\,\mathbb{E}\left[\left|1 - \left(1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right)\right|\left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|\right] \\
&\le \|\hat{p}\|^2 C_f^2\,\mathbb{E}\left[\left(1 + \left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|\right)\left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|\right] \\
&= \|\hat{p}\|^2 C_f^2\left(\mathbb{E}\left[\left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|\right] + \mathbb{E}\left[\left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|^2\right]\right) \\
&\le \|\hat{p}\|^2 C_f^2\left(\sqrt{\mathbb{E}\left[\left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|^2\right]} + \mathbb{E}\left[\left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|^2\right]\right) \qquad \text{(S.99)} \\
&\xrightarrow{M\to\infty} 0
\end{aligned}$$

by Theorem A.2 (Item 2).

Step 3. Combining steps 1 and 2, we reach

$$\begin{aligned}
&\operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right] \qquad \text{(S.100)} \\
&= \left(\operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right]\right) - \left(\operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right]\right) \\
&\xrightarrow{M\to\infty} 0.
\end{aligned}$$

■

Proof of Theorem A.3 (Item 3). Substituting $\hat{p}(x) = Cf(x)$, i.e., $\bar{p}(x) = f(x)/\|f\|$, to the previous result, yields

$$\operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(X)\|f\|}{f(X)}\right] \xrightarrow{M\to\infty} 0,$$

i.e.,

$$\operatorname{Var}\left[f(Y)W_Y\right] \xrightarrow{M\to\infty} 0.$$

■

S.5.4 Variance in the Finite Case (Theorem 1)

Proof of Theorem 1. We continue the proof of Theorem A.3 (Item 2) in Section S.5.3, and deduce

$$\begin{aligned}
&\operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right] \\
&= \operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] + \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] - \operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right] \\
&\le \operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] + \left|\operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] - \operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right]\right|.
\end{aligned}$$

We denote $V = \operatorname{Var}\left[\sum_{i=1}^M w_i\right]$ and derive from Equation S.88 in the proof of Theorem A.2 that

$$\mathbb{E}\left[\left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|^2\right] \le \frac{\operatorname{Var}\left[\hat{p}(Y)W_Y\right]}{\|\hat{p}\|^2} = \frac{V}{\|\hat{p}\|^2}. \tag{S.101}$$

We combine this with Equation S.99 to get

$$\begin{aligned}
&\left|\operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right]\right| \qquad \text{(S.102)} \\
&\le \|\hat{p}\|^2 C_f^2\left(\sqrt{\mathbb{E}\left[\left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|^2\right]} + \mathbb{E}\left[\left|1 - \frac{\bar{p}(Y)}{p_Y(Y)}\right|^2\right]\right) \qquad \text{(S.103)} \\
&\le \|\hat{p}\|^2 C_f^2\left(\sqrt{\frac{V}{\|\hat{p}\|^2}} + \frac{V}{\|\hat{p}\|^2}\right) = C_f^2\left(\|\hat{p}\|\sqrt{V} + V\right). \qquad \text{(S.104)}
\end{aligned}$$

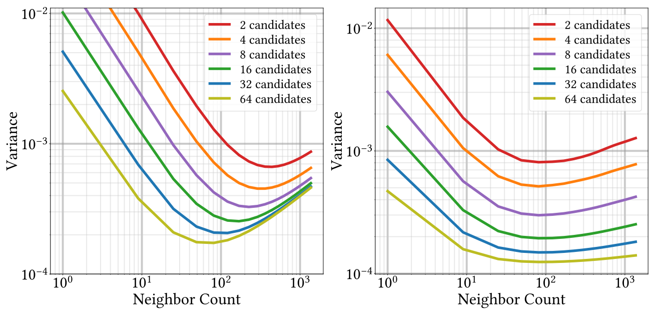
We then borrow from Equation S.95 that

$$0 \le \operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] \le C_f^2 \operatorname{Var}\left[\sum_{i=1}^M w_i\right] = C_f^2 V,$$

and reach

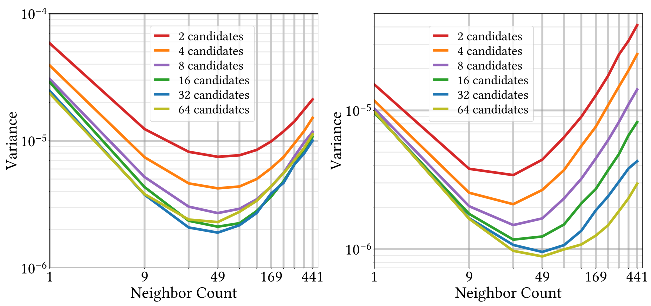
$$\begin{aligned}
&\operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right] \\
&\le \operatorname{Var}\left[f(Y)W_Y\right] - \operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] + \left|\operatorname{Var}\left[\frac{f(Y)}{p_Y(Y)}\right] - \operatorname{Var}\left[\frac{f(X)}{\bar{p}(X)}\right]\right|. \\
&\le C_f^2 V + C_f^2\left(\|\hat{p}\|\sqrt{V} + V\right) = C_f^2\sqrt{V}\left(\|\hat{p}\| + 2\sqrt{V}\right). \qquad \text{(S.105)}
\end{aligned}$$

□



(a) Cornell Box (b) Kitchen

Fig. 1. A parameter exploration of the number of initial samples per pixel (S, colored lines) and the number of neighbors used for resampling (X-axis). Each pixel generates S candidate samples and uses RIS to select one. Then GRIS resamples for each pixel a path from one from the neighboring 3x3, 5x5, 7x7, etc. pixels. Pairwise MIS is used.



(a) Cornell Box (b) Kitchen

Fig. 2. Sampling efficiency. This experiment repeats the setting of Figure 1, but outputs the MSE after ten seconds of rendering, revealing the most efficient parameters for multisample rendering.

S.6 PARAMETER EXPLORATION

In this section, we analyze the effect of parameters such as neighbor count, reuse window size and the number of spatial reuse passes for ReSTIR PT for offline rendering, justifying our choice of default parameters. We perform the experiments leading to our conclusions in two scenes, the simple Cornell Box and the more complex Kitchen scene.

We first demonstrate how sampling cost can be amortized by increasing the number of pixels we reuse from and the number of candidate path samples per pixel into optimal numbers. For the first experiment, we densely reuse paths from each pixel in a square around the current pixel, and after analyzing this case, we generalize the results to randomly sampled sparse neighbors from a larger neighborhood.

S.6.1 Parameters for Dense Block of Pixels

Path reuse without GRIS (e.g., the path reuse algorithm by Bekaert et al. [2002]) can already achieve higher sampling efficiency than pure path tracing because resampling is cheaper than generating samples from scratch. By defining unbiased contribution weights,

(a) Dense Reuse
MSE: 2.45e-7
MAPE: 0.0571

(b) Random Reuse
MSE: 2.37e-7
MAPE: 0.0502

(c) Reference

Fig. 3. A comparison of 60-second equal-time rendering of the Kitchen scene. Dense reuse and random reuse have similar MSE/MAPE with their respective optimal parameters, but random reuse produces results visually closer to the reference.

our GRIS supports more aggressive amortization of the sampling cost–with a reservoir, we can increase the number of input samples for the initial RIS to further amortize the sample generation cost. For S candidate samples, we approximately "gain" S samples when reusing one sample. To analyze the sampling efficiency, we form a simplified model that measures the ratio between the number of rays gained and the number of rays computed. Assuming a simplified ideal case where all pixels generate paths with a fixed length L and reusing a neighbor effectively gains all samples it generates, we write the following equation for reusing K pixels (including self),

$$\frac{\text{\# rays gained}}{\text{\# rays computed}} = \frac{KSL}{SL + \eta(K-1)}, \tag{S.106}$$

where η is the ray cost we pay for resampling (usually $\eta < L$).

If $S = 1$, the equation evaluates to $KL/(L + \eta(K-1))$. Even when $K \to \infty$, the efficiency is still bounded by L/η. If we have $S \to \infty$, the efficiency is bounded by K, which means that the efficiency improvement becomes theoretically unlimited in this simplified ideal case. In practice, we can measure sampling efficiency by comparing the variance at equal render time.

This suggests that we should use a relatively large S and a relatively large K to achieve higher sampling efficiency. In practice, we observe in Figure 1 that increasing S and K eventually becomes harmful: increasing the neighbor count K increases MSE instead of reducing it when the neighborhood becomes large enough, and a larger S leads to problems even sooner. This is because enlarging the neighborhood size adds samples that are farther away, and path space similarity generally decreases by distance. This eventually offsets the benefit from more samples. Enlarging S (the candidate count for initial resampling) can also eventually lead to diminishing returns, as we show in Figure 2. In both scenes, we see that using a combination of $S = 32$ and $K = 49$ (i.e., a 7×7 neighborhood) results in near-optimal sampling efficiency.

S.6.2 Parameters for Sparse Neighbors

The cost of using a large number of input samples for GRIS can be amortized also by chaining multiple spatial reuse passes. A caveat is that chaining spatial reuse passes with a small, fixed neighborhood can lead to excessive correlation between the samples, which can lead to reduced sampling efficiency. This motivates using sparse neighbors randomized from a larger neighborhood to minimize correlation.

From our parameter exploration, we found that near-optimal sampling efficiency for random reuse can be achieved with $S = 32$ (similar to the dense reuse case), 2-3 rounds of spatial reuse with 6-10 random neighbors in 5-10 pixel radius (a diameter of 10-20 pixels). After comparing visual quality in this parameter range for both scenes, we select a parameter set of 10 pixel radius, 3 rounds of spatial reuse, and 6 random neighbors. We see a slight variance reduction in equal render time compared to dense reuse with its near-optimal parameters. The improvement of sampling efficiency is small, because sparse reuse requires a larger neighborhood to reduce correlation, which also lowers the similarity between the pixels. This partially cancels the benefits from amortizing the sampling cost. The visual improvement is, however, much larger (Figure 3), since random reuse reduces visual correlation between nearby pixels. We find the 10-pixel radius can still be enlarged for real-time rendering, as chaining many reuse passes over multiple frames builds up more correlation. For real-time rendering, we use a radius of 20 pixels, only one initial candidate sample per pixel, and one spatial reuse pass between the current pixel and three random others, to keep the rendering time low.

S.7 MATHEMATICAL NOTES

S.7.1 Constraints on w_i for Zero Bias in Section 4.3

We assume, according to the section, that

$$g_i(x) = [x \in \mathcal{D}(T_i)]\; c_i(y_i) \cdot f(y_i) \left| \frac{\partial T_i}{\partial x} \right|,$$

where y_i is a shorthand for $T_i(x)$, and $[\cdot]$ is 1 if $\cdot$ is true and 0 otherwise, and that either

- (easy case) $w_i > 0$ exactly when $X_i \in \mathcal{D}(T_i)$ and $Y_i = T_i(X_i) \in \operatorname{supp} \hat{p}$, otherwise $w_i = 0$, or
- (general case) w_i are also allowed to be 0 also when $c_i(Y_i) = 0$ or $W_i = 0$, and Equation 17 holds for all $y \in \operatorname{supp} \hat{p} \cap \bigcup_{i=1}^{M} T_i(\operatorname{supp} X_i)$.

Then, we show that the derived estimator is unbiased,

$$E\left[g_s(X_s) \frac{\sum_{j=1}^{M} w_j}{w_s} f(X_s) W_s \right] = \int_{\operatorname{supp} Y} f(y)\, \mathrm{d}y, \tag{S.107}$$

and that Equation 15 holds, i.e.,

$$\operatorname{supp} Y = \operatorname{supp} \hat{p} \cap \bigcup_{i=1}^{M} T_i(\operatorname{supp} X_i). \tag{S.108}$$

Let us first prove Equation S.108 in both cases.

Proof. Easy case. If $y \in \operatorname{supp} \hat{p} \cap \bigcup_{i=1}^{M} T_i(\operatorname{supp} X_i)$, then $y = T_i(x_i)$ for some $x_i \in \operatorname{supp} X_i$ where x_i can be sampled with a positive PDF. Since also $y \in \operatorname{supp} \hat{p}$, we have $\hat{p}(y) = \hat{p}(T_i(x_i)) > 0$, and therefore by assumption $w_i > 0$. Thus, we have an x_i with positive PDF, $w_i > 0$, and $y = T_i(x_i)$, and we have a way of sampling y with a positive PDF, i.e., $y \in \operatorname{supp} Y$. We showed that $\operatorname{supp} \hat{p} \cap \bigcup_{i=1}^{M} T_i(\operatorname{supp} X_i) \subset \operatorname{supp} Y$.

If, on the other hand, $y \in \operatorname{supp} Y$, the PDF of sampling y is positive. Thus, there exists an x_i with a positive PDF and a positive selection probability for $y = T_i(x_i)$. Hence $w_i > 0$, which implies $\hat{p}(y) > 0$, and hence $y \in \operatorname{supp} \hat{p} \cap \bigcup_{i=1}^{M} T_i(\operatorname{supp} X_i)$. Combined with the previous, we have $\operatorname{supp} Y = \operatorname{supp} \hat{p} \cap \bigcup_{i=1}^{M} T_i(\operatorname{supp} X_i)$.

General case. Let $y \in \operatorname{supp} \hat{p} \cap \bigcup_{i=1}^{M} T_i(\operatorname{supp} X_i)$. Let I be the set of indices for which $y \in T_i(\operatorname{supp} X_i)$. Since $\sum_{i \in I} c_i(y) = 1$ by assumption, there exists at least one i such that $c_i(y) > 0$, and therefore an x_i such that $y = T_i(x_i)$ with $p_{X_i}(x_i) > 0$. Since $0 < p_{X_i}(x_i) = 1/\mathbb{E}[W_i \mid X_i = x_i]$, the conditional expectation of W_i is positive and hence $\Pr[W_i > 0 \mid X_i = x_i] > 0$ and $p_Y(y) > 0$, i.e., $y \in \operatorname{supp} Y$.

Assume then that $y \in \operatorname{supp} Y$. The PDF of generating y is positive, so there have to exist i and $x_i \in \Omega_i$ such that $p_{X_i}(x_i) > 0$, $y = T_i(x_i)$, and with a positive conditional probability $w_i > 0$. If w_i may be positive, it follows that $y \in \operatorname{supp} \hat{p}$, and hence $y \in \operatorname{supp} \hat{p} \cap \bigcup_{i=1}^{M} T_i(\operatorname{supp} X_i)$. Combined with the previous, we have $\operatorname{supp} Y = \operatorname{supp} \hat{p} \cap \bigcup_{i=1}^{M} T_i(\operatorname{supp} X_i)$. □

Next, we will prove Equation S.107.

Proof. First,

$$\mathbb{E}\left[g_s(X_s)\frac{\sum_{j=1}^M w_j}{w_s}W_s\right] \tag{S.109}$$

$$= \mathbb{E}\left[\sum_{s=1}^M [w_s > 0]\frac{w_s}{\sum_{j=1}^M w_j} g_s(X_s)\frac{\sum_{j=1}^M w_j}{w_s}W_s\right] \tag{S.110}$$

$$= \mathbb{E}\left[\sum_{s=1}^M [w_s > 0] g_s(X_s) W_s\right]. \tag{S.111}$$

Next, we substitute the definition of g_s and reach

$$= \mathbb{E}\left[\sum_{s=1}^M [w_s > 0][X_s \in \mathcal{D}(T_s)]\ c_s(Y_s)f(Y_s)\left|\frac{\partial T_s}{\partial X_s}\right| W_s\right]. \tag{S.112}$$

We add in the assumption which holds in both cases, $w_s = 0$ if $\hat{p}(Y_s) = 0$, and reach

$$= \mathbb{E}\left[\sum_{s=1}^M [Y_s \in \operatorname{supp}\hat{p}][w_s > 0][X_s \in \mathcal{D}(T_s)]\ c_s(Y_s)f(Y_s)\left|\frac{\partial T_s}{\partial X_s}\right| W_s\right]. \tag{S.113}$$

Next we substitute $[w_s > 0] = 1 - [w_s = 0]$, and reach

$$\begin{aligned} &= \mathbb{E}\left[\sum_{s=1}^M [Y_s \in \operatorname{supp}\hat{p}][X_s \in \mathcal{D}(T_s)]\ c_s(Y_s)f(Y_s)\left|\frac{\partial T_s}{\partial X_s}\right| W_s\right] \\ &\quad - \mathbb{E}\left[\sum_{s=1}^M [Y_s \in \operatorname{supp}\hat{p}][w_s = 0][X_s \in \mathcal{D}(T_s)]\ c_s(Y_s)f(Y_s)\left|\frac{\partial T_s}{\partial X_s}\right| W_s\right]. \end{aligned} \tag{S.114}$$

Using the definition of unbiased contribution weights (everything except W_s is a function of X_s), we get for the first term,

$$\mathbb{E}\left[\sum_{s=1}^M [Y_s \in \operatorname{supp}\hat{p}][X_s \in \mathcal{D}(T_s)]\ c_s(Y_s)f(Y_s)\left|\frac{\partial T_s}{\partial X_s}\right| W_s\right] \tag{S.115}$$

$$= \sum_{s=1}^M \int_{\operatorname{supp} X_s} [y_s \in \operatorname{supp}\hat{p}][x_s \in \mathcal{D}(T_s)]\ c_s(y_s)f(y_s)\left|\frac{\partial T_s}{\partial x_s}\right| \mathrm{d}x_s \tag{S.116}$$

$$= \sum_{s=1}^M \int_{\mathcal{D}(T_s)} [y_s \in \operatorname{supp}\hat{p}][x_s \in \operatorname{supp} X_s]\ c_s(y_s)f(y_s)\left|\frac{\partial T_s}{\partial x_s}\right| \mathrm{d}x_s, \tag{S.117}$$

which, with a change of variables $y = T_s(x_s)$ for each of the terms, and then denoting $x_s = T_s^{-1}(y)$, simplifies into

$$= \sum_{s=1}^M \int_{\mathcal{I}(T_s)} [y \in \operatorname{supp}\hat{p}][x_s \in \operatorname{supp} X_s]\ c_s(y)f(y)\,\mathrm{d}y \tag{S.118}$$

$$= \sum_{s=1}^M \int_{\operatorname{supp}\hat{p}} [y \in \mathcal{I}(T_s)][x_s \in \operatorname{supp} X_s]\ c_s(y)f(y)\,\mathrm{d}y \tag{S.119}$$

$$= \int_{\operatorname{supp}\hat{p}} \left(\sum_{s=1}^M [y \in \mathcal{I}(T_s)][x \in \operatorname{supp} X_s]\ c_s(y)\right) f(y)\,\mathrm{d}y. \tag{S.120}$$

We then write the product of the brackets as a summation condition and reach

$$= \int_{\operatorname{supp}\hat{p}} \left(\sum_{\substack{s=1 \\ y \in T_s(\operatorname{supp} X_s)}}^M c_s(y_s)\right) f(y)\,\mathrm{d}y \tag{S.121}$$

$$= \int_{\operatorname{supp}\hat{p} \cap \bigcup_i T_i(\operatorname{supp} X_i)} \left(\sum_{\substack{s=1 \\ y \in T_s(\operatorname{supp} X_s)}}^M c_s(y_s)\right) f(y)\,\mathrm{d}y \tag{S.122}$$

$$= \int_{\operatorname{supp} Y} \left(\sum_{\substack{s=1 \\ y \in T_s(\operatorname{supp} X_s)}}^M c_s(y)\right) f(y)\,\mathrm{d}y \tag{S.123}$$

$$= \int_{\operatorname{supp} Y} f(y)\,\mathrm{d}y. \tag{S.124}$$

$$\tag{S.125}$$

For the method be unbiased for integrating f over $\operatorname{supp} Y$, the second term,

$$\mathbb{E}\left[\sum_{s=1}^M [Y_s \in \operatorname{supp}\hat{p}][w_s = 0][X_s \in \mathcal{D}(T_s)]\ c_s(Y_s)f(Y_s)\left|\frac{\partial T_s}{\partial X_s}\right| W_s\right],$$

must be zero.

For the easier case, $w_i > 0$ if and only if $X_i \in \mathcal{D}(T_i)$ and $Y_i = T_i(X_i) \in \operatorname{supp}\hat{p}$. The above second term contains for each s the factor

$$[Y_s \in \operatorname{supp}\hat{p}][X_s \in \mathcal{D}(T_s)][w_s = 0] \tag{S.126}$$

$$= [w_s > 0][w_s = 0] = 0, \tag{S.127}$$

and thus the second term is zero and the estimator is unbiased.

For the general case, w_i is additionally allowed to be 0 when either $W_i = 0$ or $c_i(Y_i) = 0$. The second term is also then zero: for it to be non-zero, we need to have $w_s = 0$ for some s. However, if $w_s = 0$, then by assumption, either $X_s \notin \mathcal{D}(T_s)$, $Y_s \notin \operatorname{supp}\hat{p}$, $c_s(Y_s) = 0$, or $W_s = 0$. It is easy to check all these cases: in all cases the second term is zero. The estimator is unbiased.

□

S.7.2 Non-Negativity of m_i and W_i in Section 4.4

Here we prove that in Equation 19, we must require $m_i \geq 0$ and $W_i \geq 0$ to guarantee non-negative probabilities.

Proof. The selection probabilities $\Pr[s = i] = w_i / \sum_{j=1}^M w_j$ all need to be non-negative. They are divided by the same denominator, which flips the sign of either none or all the $w_i / \sum_{j=1}^M w_j$ expressions. Therefore, all w_i must have the same sign, so that the probabilities can all be non-negative. If we had $w_i \leq 0$ for all i, we could simply flip the signs of the w_i to reach $w_i \geq 0$; without loss of generality, we restrict ourselves to the case $w_i \geq 0$.

If it is the case that $w_i \geq 0$, then it is also the case that

$$\mathbb{E}[w_i \mid X_i] \geq 0 \tag{S.128}$$

almost surely.

Next, we substitute Equation 19 for w_i. If $x \notin \mathcal{D}(T_i)$, we have $w_i = 0$. Otherwise, denoting $y = T_i(x)$ and $Y_i = T_i(X_i)$, we must have

$$0 \le \mathbb{E}\left[w_i \mid X_i = x\right] = \mathbb{E}\left[m_i(Y_i)\hat{p}(Y_i)W_i\left|\frac{\partial T_i}{\partial X_i}\right| \mid X_i = x\right] \tag{S.129}$$

$$= m_i(y)\hat{p}(y)\left|\frac{\partial T_i}{\partial x}\right|\mathbb{E}\left[W_i \mid X_i = x\right] \tag{S.130}$$

$$= m_i(y)\hat{p}(y)\left|\frac{\partial T_i}{\partial x}\right|\frac{1}{p_{X_i}(x)}. \tag{S.131}$$

Since $\hat{p}(y)$, $\left|\frac{\partial T_i}{\partial x}\right|$ and $p_{X_i}(x)$ are all non-negative and the full product is non-negative, $m_i(y)$ must also be non-negative.

Looking at Equation 19, since we have $m_i \ge 0$, $\hat{p}(y) \ge 0$ and $\left|\frac{\partial T_i}{\partial x}\right| \ge 0$, and their product with W_i is $w_i \ge 0$, we must also have $W_i \ge 0$. □

S.7.3 Resampling MIS Must Be Positive in Support of c_i in Section 4.4

To guarantee unbiased integration, the resampling MIS weights m_i must fulfill $m_i(y) > 0$ whenever $c_i(y) \neq 0$ in addition to non-negativity and Equation 20:

Proof. Substituting g_i into the left-hand-side of Equation 10 yields

$$\mathbb{E}\left[g_s(X_s)\frac{\sum_{j=1}^M w_j}{w_s}W_s\right] = \mathbb{E}\left[[w_s > 0]g_s(X_s)\frac{\sum_{j=1}^M w_j}{w_s}W_s\right] \tag{S.132}$$

$$= \mathbb{E}\left[\sum_{s=1}^M [w_s > 0]\frac{w_s}{\sum_{j=1}^M w_j}g_s(X_s)\frac{\sum_{j=1}^M w_j}{w_s}W_s\right] \tag{S.133}$$

$$= \sum_{s=1}^M \mathbb{E}\left[[w_s > 0]g_s(X_s)W_s\right]. \tag{S.134}$$

We know that w_s (Equation 19) is positive if and only if $X_s \in \mathcal{D}(T_s)$, $m_s(Y_s) > 0$, $\hat{p}(Y_s) > 0$, $W_s > 0$ and $\left|\frac{\partial T_s}{\partial X_s}\right| > 0$. We assume $W_s \ge 0$, and the case $W_s = 0$ is already handled by the factor W_s. The Jacobian determinant is nonzero with probability 1 in $\mathcal{D}(T_i)$ since T_i is bijective. Hence, substituting g_s, we reach

$$= \sum_{s=1}^M \mathbb{E}\left[[X_s \in \mathcal{D}(T_s)][m_s(Y_s), \hat{p}(Y_s) > 0]\ c_s(Y_s)f(Y_s)\left|\frac{\partial T_s}{\partial X_s}\right|W_s\right].$$

Everything left from the unbiased contribution weight W_s is a function of X_s. Hence, by the definition of unbiased contribution weights, we reach

$$= \sum_{s=1}^M \int_{\operatorname{supp} X_s} [x_s \in \mathcal{D}(T_s)][m_s(y_s), \hat{p}(y_s) > 0]\ c_s(y_s)f(y_s)\left|\frac{\partial T_s}{\partial x_s}\right| \mathrm{d}x_s.$$

Swapping the integration domain and $[x_s \in \mathcal{D}(T_s)]$, we reach

$$= \sum_{s=1}^M \int_{\mathcal{D}(T_s)} [x_s \in \operatorname{supp} X_s][m_s(y_s), \hat{p}(y_s) > 0]\ c_s(y_s)f(y_s)\left|\frac{\partial T_s}{\partial x_s}\right| \mathrm{d}x_s,$$

and the change of variables $y = T_s(x_s)$ yields

$$= \sum_{s=1}^M \int_{\mathcal{I}(T_s)} [x_s \in \operatorname{supp} X_s][m_s(y), \hat{p}(y) > 0]\ c_s(y)f(y)\,\mathrm{d}y.$$

Swapping the integration domain and $[\hat{p}(y) > 0]$ yields

$$= \sum_{s=1}^M \int_{\operatorname{supp} \hat{p}} [x_s \in \operatorname{supp} X_s][y \in \mathcal{I}(T_s)][m_s(y) > 0]\ c_s(y)f(y)\,\mathrm{d}y,$$

which allows moving the sum inside the integral:

$$= \int_{\operatorname{supp} \hat{p}} \left(\sum_{s=1}^M [x_s \in \operatorname{supp} X_s][y \in \mathcal{I}(T_s)][m_s(y) > 0]\ c_s(y)\right) f(y)\,\mathrm{d}y.$$

Next, we simplify the first two indicators into the summation condition:

$$= \int_{\operatorname{supp} \hat{p}} \left(\sum_{\substack{s=1 \\ y \in T_s(\operatorname{supp} X_s)}}^M [m_s(y) > 0]\ c_s(y)\right) f(y)\,\mathrm{d}y. \tag{S.135}$$

The integrand is zero unless the sum contains at least one index and we can shrink the integration domain accordingly:

$$= \int_{\operatorname{supp} \hat{p} \cap \bigcup_{i=1}^M T_i(\operatorname{supp} X_i)} \left(\sum_{\substack{s=1 \\ y \in T_s(\operatorname{supp} X_s)}}^M [m_s(y) > 0]\ c_s(y)\right) f(y)\,\mathrm{d}y.$$

By Equation 15, this domain is exactly $\operatorname{supp}(Y)$:

$$= \int_{\operatorname{supp} Y} \left(\sum_{\substack{s=1 \\ y \in T_s(\operatorname{supp} X_s)}}^M [m_s(y) > 0]\ c_s(y)\right) f(y)\,\mathrm{d}y. \tag{S.136}$$

Finally, if we assume $m_s(y) > 0$ whenever $c_s(y) \neq 0$, by the constraint of the contribution MIS weights (Equation 17), we reach

$$= \int_{\operatorname{supp} Y} \left(\sum_{\substack{s=1 \\ y \in T_s(\operatorname{supp} X_s)}}^M c_s(y)\right) f(y)\,\mathrm{d}y \tag{S.137}$$

$$= \int_{\operatorname{supp} Y} f(y)\,\mathrm{d}y. \tag{S.138}$$

If we assumed a non-zero probability that $m_s(y) = 0$ while $c_s(y) \neq 0$, the multiplier in front of $f(y)$ in Equation S.136 would not simplify to 1 with the constraints of c_i, and the result would be wrong. □

REFERENCES

Robert G Bartle. 2014. *The elements of integration and Lebesgue measure.* John Wiley & Sons.

Philippe Bekaert, Mateu Sbert, and John H Halton. 2002. Accelerating Path Tracing by Re-Using Paths.. In *Rendering Techniques*. 125–134.