Multiple Importance Sampling Optimization with Linear Equations

Mateu Sbert
Informatics and Applications Institute
University of Girona
Girona, Spain
0000-0003-2164-6858

László Szirmay-Kalos
Dept. of Control Eng. and Information Tech.
Budapest Uni. of Technology and Economics
Budapest, Hungary
0000-0002-8523-2315

Víctor Elvira
School of Mathematics
University of Edinburgh
Edinburgh, UK
0000-0002-8967-4866

***Abstract*—Multiple importance sampling is an efficient Monte Carlo method to approximate integrals. It is able to simulate samples from several proposal distributions and then weight these samples accordingly. Different weighting strategies have been proposed in the literature, some of which largely reducing the variance of the Monte Carlo estimators. However, obtaining good weighting mechanisms is often challenging due to intractable integrals and non-convex optimization problems. In this work, we propose two new weighting mechanisms that are both efficient to design and obtain provable high performance. We build upon our recent work, where quasi optimal weights are obtained by solving a linear equation. Our new methods lead to more efficient computations and more robust estimations through a fast convex minimization of the variance of the estimator. The contribution of this paper is also to provide a better understanding of previous works. We validate the new methods in three experiments, showing their excellent performance.**

***Index Terms*—Multiple importance sampling, Monte Carlo**

I. Introduction

Multiple importance sampling (MIS) [1], [2] is an efficient method in Monte Carlo integration that has been widely used in statistical signal processing [3], [4]. MIS consists on combining samples from several sampling proposals in order to produce more efficient estimators. In the well-known deterministic mixture estimator (also known as balance heuristic), the weights used are proportional to the count of samples from each proposal [2], [5], [6]. Several estimators have been proposed that provide better performance than the balance heuristics with equal sample budget [7]–[10]. In [6], [11], different equal sample number strategies were analysed. The cost associated with the sampling strategies was considered in [12], and an adaptive solution by optimizing the variance using the Newton-Raphson method was presented in [13].

The optimal weights can be obtained with convex optimization, which require knowledge of the integrand function and its gradient. However, in some cases as in global illumination, this integrand is not analytically available, so previous samples should be used for the approximation, which introduces errors in the computation.

Using the fact that the MIS optimization can be represented as minimizing a Tsallis f-divergence [14], [15], from which the variance or χ^2 divergence [16]–[19] and Kullback-Leibler divergence [20] are particular cases, in previous work [21] we presented a linear heuristic to obtain the optimal weights, based on approximating arithmetic by geometric mean of primary estimators.

In this paper, we present two new methods that optimize MIS estimators through efficient and effective linear solutions. One method is based on minimizing the χ^2-divergence between the target and the proposals, and the other method minimizes the L_2 distance between the proposal and the normalized function to integrate. A new theoretical result is also found for the method in [21], showing that it can be interpreted as a Bayesian update of individual proposals with the mixed proposal. We compare the three methods in numerical examples where we have access to the ground truth, and we show the good performance of our novel approaches.

The rest of the paper is organized as follows. In Section II, we provide background material. We present our methodological contribution in Section III. We close the paper with numerical results in Section IV and concluding remarks in Section V.

II. Background

In this section, we present the problem statement and review some basics of multiple importance sampling (MIS).

A. Problem statement

The goal is to approximate the intractable integral

$$\mu = \int f(x)\mathrm{d}x,$$

where $x \in \mathbb{R}^{d_x}$ and $f(x)$ is an integrable function. In the Bayesian signal processing literature, a common problem is $f(x) = h(x)\pi(x|y)$, where $h(x)$ is a test function and $\pi(x|y)$ is the posterior distribution of the parameter x given the data $y \in \mathbb{R}^{d_y}$ [22]. Here, we work with $f(x)$ so we tackle a more generic problem.

B. Multiple importance sampling

In importance sampling (IS), the integral μ is approximated by N samples $\{X_n\}_{n=1}^N$ that are independently simulated from a so-called *proposal* probability density function (PDF), $p_1(x)$. Each sample receives an importance weight $w_n = \frac{f(X_n)}{p_1(X_n)}$,

This work was supported by PID2023-149976OB-C22 from Spanish Ministry of Science and Innovation. L.S-K. is supported by NKFIH K-145970 grant. V. E. is supported by ARL/ARO under grant W911NF-22-1-0235.

and one can approximate the integral as $\widehat{\mu} = \frac{1}{N}\sum_{n=1}^{N} w_n X_n$ [23]. In MIS, the samples are instead simulated from a set of m proposal PDFs, $\{p_i(x)\}_{i=1}^{m}$. In the following, we review several sampling and weigthing mechanisms, with associated estimators, that are often used to approximate μ. We refer the interested reader to [6].

Deterministic mixture sampling or balance heuristic. In this methods, a number n_i of samples are drawn from proposal $p_i(x)$, so the total number of samples is $\sum_{i=1}^{m} n_i = N$. When the sample numbers n_i are determined deterministically, the approach is called *multi-sample* and its estimator is given by

$$F = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{f(X_{i,j})}{\psi_\alpha(X_{i,j})}, \tag{1}$$

where $\alpha_i = n_i/N$ is the fraction of the samples allocated to the i-th proposal, and

$$\psi_\alpha(x) = \sum_{k=1}^{m}\alpha_k p_k(x) \tag{2}$$

is the mixture PDF, where $\alpha = (\alpha_1, ..., \alpha_m)$ is the vector with the mixture weights. The variance of this estimator is

$$V[F] = \frac{1}{N}\left(\int \frac{f^2(x)}{\psi_\alpha(x)}\mathrm{d}x - \sum_{i=1}^{m}\alpha_i\mu_i^2\right), \tag{3}$$

where $\mu_i = \int \frac{f(x)}{\psi_\alpha(x)} p_i(x)\mathrm{d}x$ and it can be shown that $\sum_{i=1}^{m}\alpha_i\mu_i = \mu$ [6].

The task is to optimize the vector α that minimizes $V[F]$ with the budget constraint $\sum_{k=1}^{m}\alpha_k = 1$. However, there are two additional challenges. First, the variance integrals cannot be computed analytically and must be estimated from samples, introducing uncertainty that affects the final results. Second, the optimization process should be efficient and not introduce excessive overhead. Since direct variance optimization fails to meet these two requirements, standard approaches in the approach rely on the so-called *heuristic* rules that provide simple and robust estimations, although do not necessarily guarantee to find the minimum variance.

Standard mixture sampling. In standard mixture sampling, the N samples $\{X_j\}_{j=1}^{N}$ are i.i.d. according to the mixture of (2), and the estimator is

$$\mathcal{F} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(X_j)}{\psi_\alpha(X_j)} \tag{4}$$

with variance

$$V[\mathcal{F}] = \frac{1}{N}\left(\int \frac{f^2(x)}{\psi_\alpha(x)}\mathrm{d}x - \mu^2\right). \tag{5}$$

Observe that $V[F] \leq V[\mathcal{F}]$, since this approach introduces additional randomization w.r.t. deterministic mixture sampling [1], [19], [24]. However, standard mixture sampling can also be used when the numbers of samples is less than the number of proposals.

C. MIS as divergence between distributions

In MIS, a common approach is to optimize $\psi_\alpha(x)$ in such a way that *mimics* the integrand $f(x)$. If $f(x) \geq 0$ in the whole integration space, then the integrand scaled down by the integral $g(x) = f(x)/\mu$ is also a PDF. The MIS problem can be stated as finding mixture PDF $\psi_\alpha(x)$ that minimizes a particular divergence between two distributions. For instance, it is easy to see that the variance of the estimator is minimized when the proposal minimizes the χ^2 divergence from $f(x)/\mu$ to the mixture proposal $\psi_\alpha(x)$ [17], [14].

D. Optimal MIS solution with Kullback-Leibler divergence

An optimal solution for Kullback-Leibler divergence was given in [14], the optimum is such that for all i, the first moments with respect to $p_i(x)$

$$M_i^1 = \int \left(\frac{f(x)}{\psi_\alpha(x)}\right) p_i(x)\mathrm{d}x = E_{p_i}[\mathcal{F}], \tag{6}$$

have to be equal. Eq. (6) guarantees a global minimum. It retrieves the solution in [25]. As

$$\begin{aligned}\sum_{i=1}^{m}\alpha_i M_i^1 &= \int \left(\frac{f(x)}{\psi_\alpha(x)}\right)\sum_{i=1}^{m}\alpha_i p_i(x)\mathrm{d}x \\ &= \int \left(\frac{f(x)}{\psi_\alpha(x)}\right)\psi_\alpha(x)\mathrm{d}x = M^1,\end{aligned} \tag{7}$$

where M^1 is the first moment respect to $\psi_\alpha(x)$, the condition for minimum in Eq. (6) is equivalent to for all i,

$$M_i^1 = M^1. \tag{8}$$

III. Optimizing MIS estimators via linear equations

We now propose our three methods to optimize the MIS estimator through fast and efficient novel mechanisms. In particular, we propose three approaches. The first method, approximates the non-linear equations for optimal α in Kullback-Leibler divergence through linear equations. The second method minimizes a χ^2 divergence between the proposal PDF and the normalized function to integrate, approximating the problem such that the equations for optimal α are linear. The third method follows a similar approach as the second one, using instead the L^2-norm of the difference of both PDFs.

A. First linear method

Given N_1 samples $\{X_1^1, \ldots, X_{N_1}^1\}$ from $p_1(x)$ and N_2 samples $\{X_1^2, \ldots, X_{N_2}^2\}$ from $p_2(x)$, the Kullback-Leibler optimal happens when the first moments of $\mathcal{F}$ are equal [14], thus we equal the estimators for each moment,

$$\widehat{E_{p_1}[\mathcal{F}]} \approx \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{f(X_i^1)}{\psi_\alpha(X_i^1)} = \frac{1}{N_2}\sum_{i=1}^{N_2}\frac{f(X_i^2)}{\psi_\alpha(X_i^2)} \approx \widehat{E_{p_2}[\mathcal{F}]}. \tag{9}$$

However, solving Eq. (9) would mean to solve a polynomial of grade max(N_1, N_2) equation for α. Let us consider instead solving the equation,

$$\frac{\sum_{i=1}^{N_1} f(X_i^1)}{\sum_{i=1}^{N_1}\psi_\alpha(X_i^1)} = \frac{\sum_{i=1}^{N_2} f(X_i^2)}{\sum_{i=1}^{N_2}\psi_\alpha(X_i^2)}. \tag{10}$$

Observe that Eq. (10) is a linear equation in α:

$$\frac{\sum_{i=1}^{N_1} f(X_i^1)}{\alpha \sum_{i=1}^{N_1} p_1(X_i^1) + (1-\alpha) \sum_{i=1}^{N_1} p_2(X_i^1)} \tag{11}$$
$$= \frac{\sum_{i=1}^{N_2} f(X_i^2)}{\alpha \sum_{i=1}^{N_2} p_1(X_i^2) + (1-\alpha) \sum_{i=1}^{N_2} p_2(X_i^2)}.$$

It is easily checked that Eq. (10) estimates the solution of

$$\begin{aligned} E_{\psi_\alpha p_1}[\mathcal{F}] &= \int \frac{f(x)}{\psi_\alpha(x)} \frac{\psi_\alpha(x) p_1(x)}{\int \psi_\alpha(x) p_1(x) \mathrm{d}x} \mathrm{d}x \quad (12) \\ &= \frac{\int f(x) p_1(x) dx}{\int \psi_\alpha(x) p_1(x) \mathrm{d}x} \\ &\approx \frac{\sum_{i=1}^{N_1} f(X_i^1)}{\sum_{i=1}^{N_1} \psi_\alpha(X_i^1)} = \frac{\sum_{i=1}^{N_2} f(X_i^2)}{\sum_{i=1}^{N_2} \psi_\alpha(X_i^2)} \\ &\approx \frac{\int f(x) p_2(x) dx}{\int \psi_\alpha(x) p_2(x)} \mathrm{d}x \\ &= \int \frac{f(x)}{\psi_\alpha(x)} \frac{\psi_\alpha(x) p_2(x)}{\int \psi_\alpha(x) p_2(x) \mathrm{d}x} \mathrm{d}x \\ &= E_{\psi_\alpha p_2}[\mathcal{F}]. \end{aligned}$$

Thus we have substituted the optimality condition of $E_{p_1}[\mathcal{F}] = E_{p_2}[\mathcal{F}]$ by the condition $E_{\psi_\alpha p_1}[\mathcal{F}] = E_{\psi_\alpha p_2}[\mathcal{F}]$ obtaining a linear equation in α. Observe that $\psi_\alpha p_1, \psi_\alpha p_2$ can be interpreted as the Bayesian update of p_1, p_2 with ψ_α.

Eq. (10) was introduced in [21] by considering the arithmetic mean of primary estimators of the first moments to be approximately equal to the geometric mean of these estimators. For m pdfs, Eq. (10) generalizes to $m-1$ independent equations

$$\frac{\sum_{l=1}^{N_i} f(X_l^i)}{\sum_{k=1}^{N_i} \psi_\alpha(X_k^1)} = \frac{\sum_{l=1}^{N_j} f(X_l^2)}{\sum_{k=1}^{N_j} \psi_\alpha(X_k^j)} \tag{13}$$

for all i, j, that taken together with $\sum_{k=1}^{m} \alpha_k = 1$ gives us m independent equations for the m unknowns $\{\alpha_k\}_{k=1}^{m}$. **The case of weight equal to zero** Eq. (13) will fail to deliver a convex solution for weights for the case where the optimal weight corresponding to a given proposal is equal to zero. This reflects the fact that Eq. (8) will fail for this case. To account for this, in [21] it was proposed considering each α_i equal to zero in turn and solve the remaining $m-2$ linear system. From all the feasible solutions, we choose the one that has less sample variance, although this can potentially have a combinatorial cost.

B. Second linear method: χ^2 divergence minimization

If we optimize $\chi^2(g, \psi_\alpha)$, equal to the variance normalized by μ^2 of the estimator $\mathcal{F}$, no linear approach would result. Let us instead use the other direction of the divergence

$$\chi^2(\psi_\alpha, g) = \int \frac{(\psi_\alpha(x) - g(x))^2}{g(x)} \mathrm{d}x. \tag{14}$$

It is convex in α. Using Lagrange multipliers, we obtain that the optimality condition of Eq. (14) is when $\forall i, j$

$$\int \frac{\psi_\alpha(x)}{g(x)} p_i(x) \mathrm{d}x = \int \frac{\psi_\alpha(x)}{g(x)} p_j(x) \mathrm{d}x. \tag{15}$$

We consider two proposals $p_1(x), p_2(x)$ with N_1, N_2 samples, respectively. Then, the optimality condition leads into the linear equation

$$\frac{1}{N_1} \sum_{i=1}^{N_1} \frac{\psi_\alpha(X_i^1)}{g(X_i^1)} = \frac{1}{N_2} \sum_{i=1}^{N_2} \frac{\psi_\alpha(X_i^2)}{g(X_i^2)},$$

or

$$\frac{1}{N_1} \sum_{i=1}^{N_1} \frac{\psi_\alpha(X_i^1)}{f(X_i^1)} = \frac{1}{N_2} \sum_{i=1}^{N_2} \frac{\psi_\alpha(X_i^2)}{f(X_i^2)}. \tag{16}$$

Observe that Eq. (16) corresponds to the equality of the harmonic means of the primary estimators of the first moments in Eq. (6). Thus, as the first linear method was derived from the equality of the arithmetic means of the first moment estimators, this second linear method derives from the equality of the harmonic means. Finally, observe that Eq. (16) solves for the optimum of

$$\int \left(\frac{\psi_\alpha(x)}{g(x)} - 1 \right)^2 dx. \tag{17}$$

C. Third linear method: L^2-norm minimization

Let us consider the minimum of the squared L^2-norm

$$\int \big(g(x) - \psi_\alpha(x) \big)^2 \mathrm{d}x, \tag{18}$$

which is convex in α. Using Lagrange multipliers, we obtain that the optimality condition is when $\forall i, j$

$$\int (g(x) - \psi_\alpha(x)) p_i(x) \mathrm{d}x = \int (g(x) - \psi_\alpha(x)) p_j(x) \mathrm{d}x. \tag{19}$$

Suppose two proposals $p_1(x), p_2(x)$ with N_1, N_2 samples, respectively, the optimality condition translates into the linear equation

$$\frac{1}{N_1} \Big(\sum_{i=1}^{N_1} g(X_i^1) - \sum_{i=1}^{N_1} \psi_\alpha(X_i^1) \Big) = \frac{1}{N_2} \Big(\sum_{i=1}^{N_2} g(X_i^2) - \sum_{i=1}^{N_2} \psi_\alpha(X_i^2) \Big),$$

which can be re-written as

$$\frac{N_2 \sum_{i=1}^{N_1} f(X_i^1) - N_1 \sum_{i=1}^{N_2} f(X_i^2)}{\mu} = N_2 \sum_{i=1}^{N_1} \psi_\alpha(X_i^1) - N_1 \sum_{i=1}^{N_2} \psi_\alpha(X_i^2),$$

where μ is estimated with the deterministic mixture estimator F with $\alpha_0 = \{\frac{N_1}{N_1+N_2}, \frac{N_2}{N_1+N_2}\}$ as

$$\mu \approx \frac{1}{N} \Big(\sum_{i=1}^{N_1} \frac{f(X_i^1)}{\psi_{\alpha_0}(X_i^1)} + \sum_{i=1}^{N_2} \frac{f(X_i^2)}{\psi_{\alpha_0}(X_i^2)} \Big). \tag{20}$$

IV. Numerical Examples

We present here three examples. In all examples, we were able to compute the ground truth of the values of variances $V[F]$ by using a solver for Eq. (3).

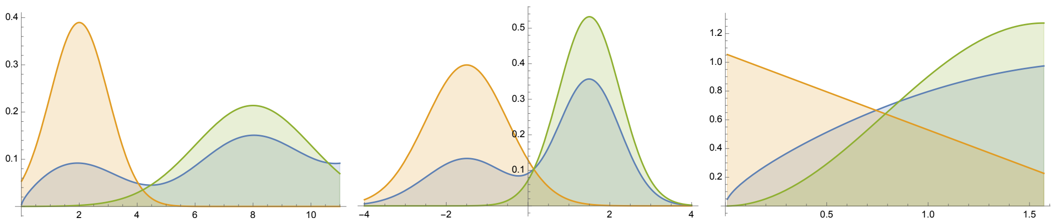

Fig. 1. $f(x)/\mu$ (in blue) superimposed on the two PDFs used for MIS integration in Example 1 (left), Example 2 (middle), and Example 3 (right).

Example 1

Suppose we want to evaluate the integral (see Fig. 1(left))

$$\mu = \int_{0.01}^{3.5\pi} \left(\sqrt{x} + \sin x\right) \mathrm{d}x \approx 25.3065 \tag{21}$$

by MIS using PDFs $\mathcal{N}(2,1)$ and $\mathcal{N}(8,2)$, where $\mathcal{N}(m,\sigma)$ stands for the normal distribution of mean m and standard deviation σ. For this example, equal sample number MIS has variance $V[F] = 24.1152$. In Fig. 2 we show the values of $V[F]$ for the optimal α fractions for the K-L divergence using 100 runs each consisting of 4 Newton-Raphson iterations with 50 total samples in each iteration, thus 200 samples in total for each run. In Fig. 3 we show the values of $V[F]$ for the three linear methods for 100 runs, each with 100 samples from each proposal, thus 200 samples in total.

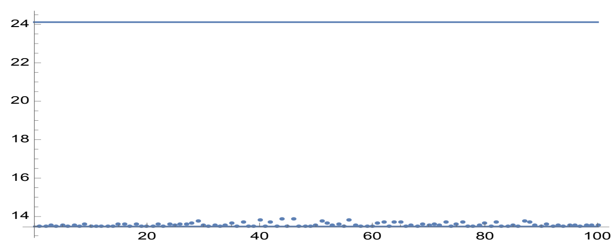

Fig. 2. The variances, in vertical axis, obtained for 100 runs with Newton-Raphson for K-L divergence method with 4 iterations of 50 samples each applied to Example 1. The horizontal lines correspond to the optimal and the equal sampling variance $V[F]$, respectively.

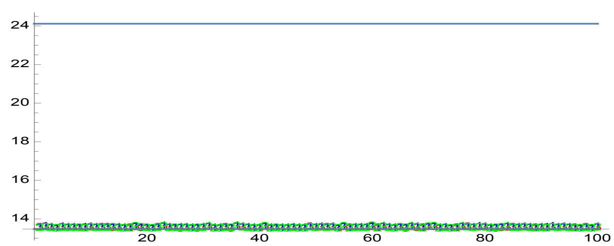

Fig. 3. The variances, in vertical axis, obtained for 100 runs with the three methods (first in blue, second in green and third in red) applied to Example 1, with a total of 100 samples for each proposal. The horizontal lines correspond to the optimal and the equal sampling variance $V[F]$, respectively.

Example 2

Let us consider integral (see Fig. 1(center))

$$\mu = \int_{-4}^{4} \left(\mathcal{N}(-1.5,1) + 2\mathcal{N}(1.5,0.75)\right) \mathrm{d}x \approx 2.9929 \tag{22}$$

by MIS using functions $\mathcal{N}(-1.5,1)$ and $\mathcal{N}(1.5,0.75)$. For this example, equal sample number MIS has a variance of $V[F] = 0.1134$. In Fig. 4 we show the values of $V[F]$ for the optimal α fractions for the K-L divergence using 100 runs of 4 Newton-Raphson iterations with same number of samples as in Example 1. In Fig. 5 we show the values of $V[F]$ for the three linear methods for 100 runs, each run with 100 samples from each proposal, thus 200 samples in total.

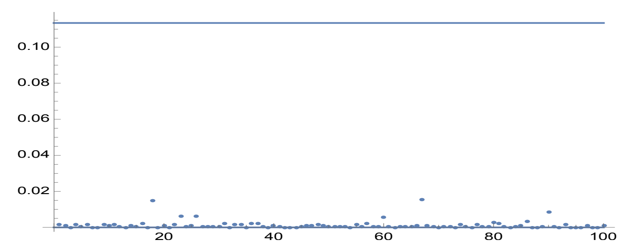

Fig. 4. The variances, in vertical axis, obtained for 100 runs for K-L divergence with Newton-Raphson method with 4 iterations of 50 samples each applied to Example 2. The horizontal lines correspond to the optimal and the equal sampling variance $V[F]$, respectively.

Fig. 5. The variances, in vertical axis, obtained for 100 runs with the three methods (first in blue, second in green and third in red) applied to Example 2, with a total of 100 samples for each proposal. The horizontal lines correspond to the optimal and the equal sampling variance $V[F]$, respectively.

Example 3

Consider the approximation of the following integral (see Fig. 1(right))

$$\mu = \int_{0.01}^{\pi/2} \left(\sqrt{x} + \sin x\right) \mathrm{d}x \approx 2.3118 \tag{23}$$

by MIS using functions $2-x$, and $\sin^2(x)$. For this example, equal sample number MIS has a variance of $V[F]=0.2772$. In Fig. 6 we show the values of $V[F]$ for the optimal α fractions for the K-L divergence using 100 runs of 4 Newton-Raphson iterations with same number of samples as in Example 1. In Fig. 7 we show the values of $V[F]$ for the three linear methods for 100 runs, each run with 100 samples from each proposal, thus 200 samples in total.

Finally, in Table I we compare, for the examples described above, the minimum $V[F]$ value, Eq. (7), the value obtained with the solution of equal first moments, Eq. (8), and the value with the solution of linearizing equal first moments, Eq. (12).

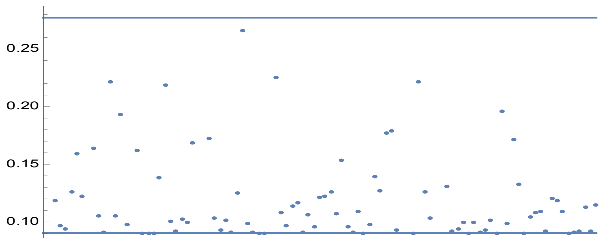

Fig. 6. The variances, in vertical axis, obtained for 100 runs with Newton-Raphson method for K-L divergence with 4 iterations of 50 samples each applied to Example 3. The horizontal lines correspond to the optimal and the equal sampling variance $V[F]$, respectively.

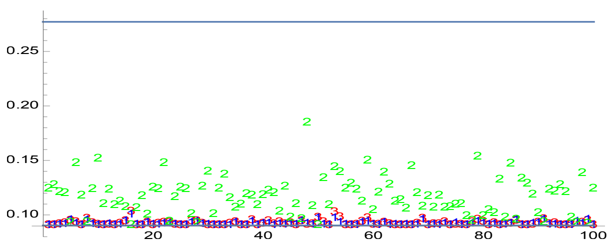

Fig. 7. The variances, in vertical axis, obtained for 100 runs with the three methods (first in blue, second in green and third in red) applied to Example 3, with a total of 100 samples for each proposal. The horizontal lines correspond to the optimal and the equal sampling variance $V[F]$, respectively.

TABLE I

VALUES OF $V[F]$: OPTIMUM , EQUAL FIRST MOMENTS, AND FIRST LINEAR METHOD SOLUTION FOR EXAMPLES 1–3.

Example	Optimum	Equal first moments	Linear method
1	13.4788	13.5021	13.5398
2	0	0	0
3	0.09032	0.09268	0.09042

V. CONCLUSION

In this paper we have revisited a linear method, based on K-L divergence optimization, to approximate the optimal weights in Multiple importance sampling by solving linear equations, added further motivation to the method and showed the relationship to the exact optimal values. We have also introduced two new linear method to solve for the quasi-optimal weights. Our preliminary results show that the new methods work also better than the Newton-Raphson based solution, and the one based on optimizing the L^2 squared norm appears to be as good as our original method. This method was showed in [21] to be the state of the art in global illumination, thus in future work we will evaluate the new methods in the global illumination problem too. We will also consider the cost associated with the sampling strategies.

REFERENCES

[1] E. Veach and L. J. Guibas, "Optimally combining sampling proposals for monte carlo rendering," pp. 419–428, 1995.

[2] E. Veach, *Robust Monte Carlo Methods for Light Transport Simulation*, Ph.D. thesis, Stanford University, 1997.

[3] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Miguez, and P. M. Djuric, "Adaptive importance sampling: The past, the present, and the future," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 60–79, 2017.

[4] V. Elvira, L. Martino, M. F. Bugallo, and P. M. Djuric, "Elucidating the auxiliary particle filter via multiple importance sampling [lecture notes]," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 145–152, 2019.

[5] A. Owen and Y. Zhou, "Safe and effective importance sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 135–143, 2000.

[6] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized Multiple Importance Sampling," *Statistical Science*, vol. 34, no. 1, pp. 129 – 155, 2019.

[7] H. Lu, R. Pacanowski, and X. Granier, "Second-order approximation for variance reduction in multiple importance sampling," *Computer Graphics Forum*, vol. 32, pp. 131–136, 2013.

[8] M. Sbert, V. Havran, and L. Szirmay-Kalos, "Variance analysis of multi-sample and one-sample multiple importance sampling," *Computer Graphics Forum*, vol. 35, pp. 451–460, 2016.

[9] V. Havran and M. Sbert, "Optimal combination of proposals in multiple importance sampling," in *VRCAI '14*, 2014, pp. 141–150.

[10] M. Sbert, V. Havran, and L. Szirmay-Kalos, "Multiple importance sampling revisited: Breaking the bounds," *EURASIP J. Adv. Signal Process*, vol. 15, 2018.

[11] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Efficient multiple importance sampling estimators," *IEEE Signal Process. Lett.*, vol. 10, pp. 1757–1761, 2015.

[12] M. Sbert, V. Havran, L. Szirmay-Kalos, and V. Elvira, "Multiple importance sampling characterization by weighted mean invariance," *The Visual Computer*, 2018.

[13] M. Sbert, V. Havran, and L. Szirmay-Kalos, "Optimal deterministic mixture sampling," in *Eurographics 2019 - Short Papers*, 2019.

[14] M. Sbert and L. Szirmay-Kalos, "Robust multiple importance sampling with tsallis ϕ-divergences," *Entropy*, vol. 24, pp. 1240, 2022.

[15] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observation," *Studia Scientiarum Mathematicarum Hungarica*, pp. 229–318, 1967.

[16] J. Cornebise, E. Moulines, and J. Olsson, "Adaptive methods for sequential importance sampling with application to state space models," *Stat. Comput.*, vol. 18, pp. 461–480, 2008.

[17] J. Miguez, "On the performance of nonlinear importance samplers and population monte carlo schemes," in *22nd International Conference on Digital Signal Processing (DSP)*, 2017.

[18] T. Müller, B. Mcwilliams, F. Rousselle, M. Gross, and J. Novák, "Neural importance sampling," *ACM Trans. Graph.*, vol. 38, no. 5, 2019.

[19] M. Sbert and V. Elvira, "Generalizing the balance heuristic estimator in multiple importance sampling," *Entropy*, vol. 24, 2022.

[20] J. Vorba, J. Hanika, S. Herholz, T. Mueller, P. Krivanek, and A. Keller, "Path guiding in production," in *ACM SIGGRAPH 2019 Courses*, 2019.

[21] M. Sbert and L. Szirmay-Kalos, "A linear heuristic for multiple importance sampling," *EURASIP J. Adv. Signal Process*, vol. 31, 2023.

[22] J. V. Candy, *Bayesian signal processing: classical, modern, and particle filtering methods*, John Wiley & Sons, 2016.

[23] V. Elvira and L. Martino, "Advances in importance sampling," *arXiv preprint arXiv:2102.05407*, 2021.

[24] R. Mukerjee and V. Elvira, "Variance analysis of multiple importance sampling schemes," *Statistics*, vol. 57, no. 6, pp. 1392–1401, 2023.

[25] L. Szirmay-Kalos and M. Sbert, "Robust sample budget allocation for mis," in *Eurographics 2022 - Short Papers*, 2022.