Multiple Importance Sampling for Efficient Symbol Error Rate Estimation

Víctor Elvira, *Member, IEEE,* and Ignacio Santamaría, *Senior Member, IEEE,*

Abstract—Digital constellations formed by hexagonal or other non-square two-dimensional lattices are often used in advanced digital communication systems. The integrals required to evaluate the symbol error rate (SER) of these constellations in the presence of Gaussian noise are in general difficult to compute in closed-form, and therefore Monte Carlo simulation is typically used to estimate the SER. However, naive Monte Carlo simulation can be very inefficient and requires very long simulation runs, especially at high signal-to-noise ratios. In this letter, we adapt a recently proposed multiple importance sampling (MIS) technique, called ALOE (for "At Least One rare Event"), to this problem. Conditioned to a transmitted symbol, an error (or rare event) occurs when the observation falls in a union of half-spaces or, equivalently, outside a given polytope. The proposal distribution for ALOE samples the system conditionally on an error taking place, which makes it more efficient than other importance sampling techniques. ALOE provides unbiased SER estimates with simulation times orders of magnitude shorter than conventional Monte Carlo.

Index Terms—Improper constellations, lattices, Monte Carlo, multiple importance sampling, symbol error rate.

I. INTRODUCTION

In several communication systems the transmitted signal belongs to a non-square 2D lattice. For instance, hexagonal constellations are the densest packing of regularly spaced points in 2D and, as the number of constellation points grows to infinity, they also minimize the probability of error in Gaussian noise under an average power constraint being therefore optimal [1]. Other examples of constellations that provide non-rectangular decision boundaries are the θ quadrature amplitude modulation (QAM) family [2], and the recently proposed family of improper constellations in [3].

When these constellations are used, maximum likelihood decoding amounts to finding the closest lattice point to a given noisy observation, x, which is known to be NP-hard for generic channels [4], [5] meaning that the problem complexity is exponential on the dimension of the lattice [6]. Further, to evaluate the symbol error rate probability conditioned to a given constellation point we need to compute the probability that the observation x lies outside a polytope defining the corresponding Voronoi region. The resulting integrals are in

general difficult to compute in closed form for arbitrary decision regions, and typically one resorts to standard Monte Carlo simulations for performance evaluation, which can be very time consuming especially when the number of symbols of the constellation is large and the targeted symbol error rate (SER) is below 10^{-6}.

As an alternative to naive Monte Carlo, importance sampling (IS) has been used as a method for variance reduction in SER or BER simulations in a wide range of scenarios since the late seventies [7]–[10] (for a review we refer the reader to [11]). Despite this, many digital communication researchers are still unaware of the potential benefits of IS techniques to characterize the statistical performance of digital communication systems.

The basic IS methodology samples from a proposal distribution that increases the number of errors during simulation, and then weights the samples by the ratios of the target to the proposal densities [12]. However, designing a good proposal with high density in regions where samples should be drawn might not be easy, and typically is problem dependent. For instance, a biased channel distribution is proposed in [13] for the simulation of orthogonal space-time block codes (OSTBCs) on Nakagami channels. Other recent work is [14], where IS is used for the estimation of the outage probability of multi-antenna receivers with generalized selection combining. A general approach is using multiple importance sampling (MIS), where several proposals are used for simulating the samples [15]–[18].

In this letter, we apply a recently proposed multiple importance sampling technique called ALOE ("At Least One rare Event") [19] to estimate the SER in additive white Gaussian noise (AWGN) channels.

The conditional symbol error probability is the integral of a Gaussian outside a polytope formed by the intersection of K hyperplanes; or, in other words, the integral over the union of the half-spaces formed by these hyperplanes. ALOE estimates this integral by taking samples from a mixture of K truncated Gaussians where errors take place. The method can be easily extended to arbitrary 2D constellations and provides unbiased SER estimated with bounded variance.

V. Elvira is with IMT Lille Douai & CRISTAL (UMR 9189), Lille, France (e-mail: victor.elvira@imt-lille-douai.fr).

I. Santamaria is with the Department of Communications Engineering, University of Cantabria, Santander, Spain (e-mail: i.santamaria@unican.es).

II. EFFICIENT SER ESTIMATION

A. Problem statement

Our motivating problem is the efficient SER estimation of digital constellations like the one depicted in Fig. 1a. Assume that symbol $\mathbf{s}_m$, marked with a star in Fig. 1b, is transmitted. The observation $\mathbf{x} = \mathbf{s}_m + \mathbf{n}$ is the symbol perturbed by additive
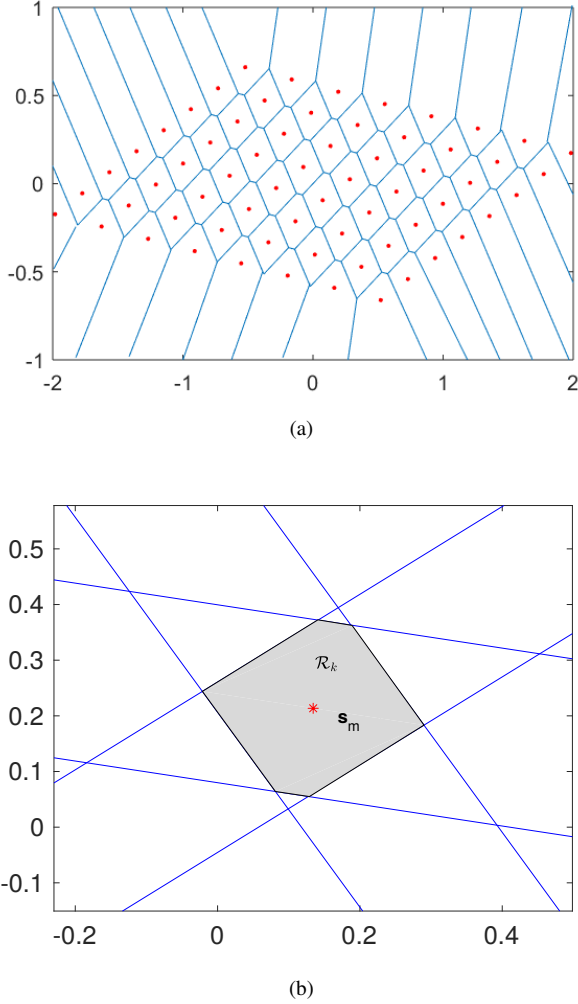
(a)

(b)

Fig. 1: a) 2D lattice for an improper constellation with $M = 64$ and $\kappa = 0.8$. The decision regions are also depicted, and b) Zoom of the decision region for a given transmitted symbol $\mathbf{s}_m = 0.1343 + 0.2136j$.

Gaussian noise. The conditional probability of error is the integral of a Gaussian distribution centered at $\mathbf{s}_m$ outside the shaded decision region in Fig. 1b: $p_m \triangleq P(e|\mathbf{s}_m) = \mathbb{P}(\mathbf{x} \notin \mathcal{R}_m|\mathbf{s}_m)$, and, assuming that all symbols are transmitted with equal probability, the symbol error probability is

$$P_e = \frac{1}{M} \sum_{m=1}^{M} p_m. \tag{1}$$

Our goal is to develop an efficient multiple important sampling (MIS) scheme to estimate p_m, $m = 1, \ldots, M$ in (1).

For notational simplicity let us denote the transmitted symbols as $\mathbf{s}$, and let $\mathcal{R}$ be its associated Voronoi region, which is a polytope defined by the intersection of finitely many hyperplanes in $\mathbb{R}^2$. The probability of interest can be then expressed as

$$p = \int_{\mathbb{R}^2} \mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x})\tilde{\pi}(\mathbf{x})d\mathbf{x} \tag{2}$$

where $\tilde{\pi}(\mathbf{x}) \triangleq \mathcal{N}(\mathbf{s}, \sigma^2 \mathbf{I})$ is the Gaussian distribution of the observation, and $\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x})$ is the indicator function taking value

1 for all $\mathbf{x} \notin \mathcal{R}$, i.e., out of the shaded decision region in Fig. 1b. The integral in Eq. (2) is intractable in the general case where $\mathcal{R}$ is an arbitrary polytope. However, for some constellations such as QAM modulations, the integral can of course be obtained in a closed-form, and there is no need to resort to any simulation procedure.

A naive Monte Carlo estimator of p simulates N samples from $\tilde{\pi}(\mathbf{x})$ and approximates (2) as

$$\widehat{p}^{\,(\mathrm{MC})} = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}_n), \tag{3}$$

i.e., it simulates from the observation distribution and counts the rate of observations that are out of $\mathcal{R}$. This estimator is unbiased, but its efficiency decays dramatically when p is small.

B. Importance Sampling

Importance sampling (IS) is a more advanced Monte Carlo methodology, used when sampling from $\tilde{\pi}(\mathbf{x})$ is either not possible or not efficient. The N samples are simulated instead from a so-called *proposal* distribution, $q(\mathbf{x})$, and the estimator of p is built as

$$\widehat{p}^{\,(\mathrm{IS})} = \frac{1}{N} \sum_{n=1}^{N} w_n \mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}_n), \qquad \mathbf{x}_n \sim q(x), \qquad n = 1, ..., N, \tag{4}$$

where $w_n = \frac{\tilde{\pi}(\mathbf{x}_n)}{q(\mathbf{x}_n)}$ is the importance weight. Note that $\widehat{p}^{\,(\mathrm{IS})}$ is an unbiased and consistent estimator of p.

The variance of the estimator $\widehat{p}^{\,(\mathrm{IS})}$ in Eq. (4) is given by

$$\mathrm{Var}_q(\widehat{p}^{\,(\mathrm{IS})}) = \frac{1}{N} \int \frac{\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x})\tilde{\pi}^2(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} - \frac{p^2}{N}. \tag{5}$$

It can be shown that the optimal proposal, q^*, that minimizes the variance is $q^*(\mathbf{x}) \propto \mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}) \cdot \tilde{\pi}(\mathbf{x})$. Intuitively, the performance is poor when the targeted integrand $\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}) \cdot \tilde{\pi}(\mathbf{x})$ and the distribution that generates the samples have a large mismatch. In our problem, the performance is very bad when we find few or no errors for the simulated observation. This explains the very poor performance of the naive Monte Carlo $\widehat{p}^{\,(\mathrm{MC})}$, whose variance can be obtained as a particular case of (5) as

$$\mathrm{Var}_{\tilde{\pi}}(\widehat{p}^{\,(\mathrm{MC})}) = \frac{1}{N}\left(p - p^2\right), \tag{6}$$

with a relative root mean square error (RRMSE) of

$$\mathrm{RRMSE}(\widehat{p}^{\,(\mathrm{MC})}) = \frac{\sqrt{\mathrm{Var}(\widehat{p}^{\,(\mathrm{MC})})}}{p} \tag{7}$$

$$= \frac{1}{\sqrt{N}} \frac{\sqrt{p - p^2}}{p} \tag{8}$$

$$= \frac{1}{\sqrt{N}} \sqrt{\frac{1}{p} - 1}. \tag{9}$$

Note that N needs to be inversely proportional to the true value p in order to provide a reliable estimate, which explains why N must be huge for high signal-to-noise ratio (SNR).

C. Multiple Importance Sampling

In the case of IS, it is difficult to find a good unique proposal with a low mismatch w.r.t. the multimodal product $\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}) \cdot \tilde{\pi}(\mathbf{x})$. Multiple importance sampling (MIS) is a natural extension of IS in this setup, allowing for the simulation from a set of K proposals, $\{q_k(\mathbf{x})\}_{k=1}^{K}$, instead of just one [17]. However, the extension from one to several proposals is not straightforward and many sampling and weighting schemes can be devised (see [20] for a review). A conventional way to proceed is to simulate from the mixture proposal as

$$\mathbf{x}_n \sim q_{\boldsymbol{\alpha}} = \sum_{k=1}^{K} \alpha_k q_k(\mathbf{x}), \qquad n = 1, ..., N, \qquad (10)$$

where $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_K]$ is a simplex vector with all non-negative weights in the mixture such that $\sum_{k=1}^{K} \alpha_k = 1$. The MIS extension of Eq. (4) is

$$\widehat{p}^{\text{ (MIS)}} = \frac{1}{N} \sum_{n=1}^{N} \frac{\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}_n)\tilde{\pi}(\mathbf{x}_n)}{q_{\boldsymbol{\alpha}}(\mathbf{x}_n)}. \qquad (11)$$

We recall that we aim at integrating $\tilde{\pi}(\mathbf{x})$ in region $\overline{\mathcal{R}}$. This region can be described as the union of all half-spaces in $\mathbb{R}^2$ generated by the K hyperplanes that define the border of $\mathcal{R}$ (e.g, $K = 6$ for the symbol in Fig. 1b). More precisely, $\overline{\mathcal{R}} = \bigcup_{k=1}^{K} \mathcal{S}_k$, where $\mathcal{S}_k = \{\mathbf{x} \in \mathbb{R}^2 \,|\, \mathbf{x}^T \boldsymbol{\gamma}_k \geq \beta_k\}$ is the half-space defined by the k-th hyperplane, which is parametrized by $\boldsymbol{\gamma}_k$ and β_k.

We follow the choice of proposals in [19]. First, the number of proposals, K, is equal to the number of hyperplanes, being each proposal a truncated version of the target distribution (a Gaussian centered at the received symbol) beyond each hyperplane, i.e., $q_k(\mathbf{x}) = \frac{\mathbb{I}_{\mathcal{S}_k(\mathbf{x})}\tilde{\pi}(\mathbf{x})}{P_k}$, where $P_k = \int \mathbb{I}_{\mathcal{S}_k(\mathbf{x})}\tilde{\pi}(\mathbf{x})d\mathbf{x}$ is the integral of the target distribution beyond the hyperplane (the procedure for the efficient simulation from a generic truncated Gaussian distribution is described in Appendices A and B). For reasons that will become apparent shortly, it is useful to define $\overline{p} = \sum_{k=1}^{K} P_k$, which is an upper union bound of p.[1] Then, Eq. (11) yields

$$\widehat{p}^{\text{ (MIS)}} = \frac{1}{N} \sum_{n=1}^{N} \frac{\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}_n)\tilde{\pi}(\mathbf{x}_n)}{q_{\boldsymbol{\alpha}}(\mathbf{x}_n)} \qquad (12)$$

$$= \frac{1}{N} \sum_{n=1}^{N} \frac{\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}_n)}{\sum_{k=1}^{K} \alpha_k \mathbb{I}_{\mathcal{S}_k}(\mathbf{x}_n)P_k^{-1}}. \qquad (13)$$

The weight of each proposal in the mixture defined in Eq. (10) is chosen as $\alpha_k = P_k/\overline{p}$, for $k = 1, ..., K$. Then, Eq. (13) yields

$$= \frac{\overline{p}}{N} \sum_{n=1}^{N} \frac{\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}_n)}{\sum_{k=1}^{K} \mathbb{I}_{\mathcal{S}_k}(\mathbf{x}_n)} \qquad (14)$$

$$= \frac{\overline{p}}{N} \sum_{n=1}^{N} \frac{1}{C(\mathbf{x}_n)}, \qquad (15)$$

where $C(\mathbf{x}_n) = \sum_{k=1}^{K} \mathbb{I}_{\mathcal{S}_k}(\mathbf{x}_n)$ is the number of half-spaces $\mathcal{S}_k$ where $\mathbf{x}_n$ is present. Note that we have used $\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x}_n) = 1$, since all samples are generated in $\overline{\mathcal{R}}$. Fig. 2 shows samples generated from the proposal corresponding to the symbol $\mathbf{s}_m =$

---

[1] Note that the bound becomes an equality when $S_i \cap S_j = \varnothing$, for all $i \neq j$.

0.1343+0.2136j in the constellation of Fig. 1a. Note that there are $K = 6$ proposals in the mixture, and the samples of each proposal are represented by a different color.

D. Theoretical guarantees of the MIS estimator

The variance of the estimator (15) can be bounded by [19]

$$\text{Var}(\widehat{p}^{\text{ (MIS)}}) \leq \frac{p(\overline{p} - p)}{N}. \qquad (16)$$

First, note that when the upper union bound $\overline{p}$ gets closer to the true value p, the variance of the MIS estimator goes to zero. This is of special interest in our application since the performance of the raw Monte Carlo estimate deteriorates for high SNRs (i.e., when p is small). Conversely, the MIS estimator gets better for high SNR since the upper bound $\overline{p}$ becomes tighter. The reason is that most errors fall very close to the separating hyperplane, and hence in just only one region $\mathcal{S}_k$ (i.e., $C(\mathbf{x}_n) = 1$ for all $\mathbf{x}_n$). Note that in the limit, when all $C(\mathbf{x}_n) = 1$, the estimator (15) is always $\overline{p}$.

Let us compare the variance bound of the MIS estimator and the closed-form variance of raw Monte Carlo. We look for values of $\overline{p}$ where

$$\frac{p(\overline{p} - p)}{N} \leq \frac{1}{N}\left(p - p^2\right), \qquad (17)$$

which turns to be $\overline{p} \leq 1$. Note that for moderate SNRs, the upper bound $\overline{p}$ is smaller than 1. Only for very low SNRs, $\overline{p} \geq 1$, which is not a problem since (a) we are comparing the MIS upper bound (not the variance itself), and (b) the low SNR scenario is not challenging and both MIS and raw Monte Carlo estimators are very accurate with a few samples.

Eq. (16) gives us an upper bound for the variance of the conditional probability of error of an arbitrary symbol of the constellation that we denoted for simplicity as $\mathbf{s}$. To estimate the symbol error probability in Eq. (1) we run M independent ALOEs, then, the variance of the P_e estimator can be bounded as

$$\text{Var}(\widehat{P}_e^{\text{ (MIS)}}) \leq \frac{1}{M^2 N} \sum_{m=1}^{M} p_m(\overline{p}_m - p_m). \qquad (18)$$

Note that several symbols in Fig. 1a have identical Voronoi regions, which can be exploited to avoid approximating all corresponding error probabilities. Moreover, one can allocate different number of samples, N_m, for the approximation of the error probability associated to each of the M symbols. However, this fine tuning goes beyond the scope of this paper.

III. SIMULATION RESULTS

In this section we estimate the SER of arbitrary 2D digital constellations with the proposed MIS technique. As an example, we consider the family of improper constellations proposed in [3], whose two-dimensional signal points belong to a non-square lattice. Improper signals, which are correlated with their complex conjugate, have proven beneficial in several interference-limited wireless networks such as the interference channel [21]–[23], Z-interference channel [24], underlay cognitive radio networks [25], [26], and relay channels [27].
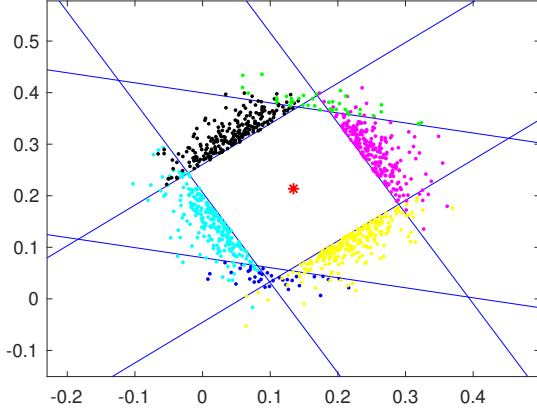
Fig. 2: The red asterisk represents the symbol $\mathbf{s}_m = 0.1343 + 0.2136j$. The blue solid lines represent the $K = 6$ hyperplanes that define the half-spaces, whose union describes $\overline{\mathcal{R}}$. The points represent the samples simulated from the mixture proposal $q_{\boldsymbol{\alpha}}$ (each component in a different color).

In all these scenarios, we wish to transmit a digital communications signal with a given circularity coefficient, which is the parameter that determines the degree of impropriety of the signal. The circularity coefficient of a zero-mean complex random variable X, is defined as [28]

$$\kappa = \frac{\left|\mathrm{E}[X^2]\right|}{\mathrm{E}[|X|^2]}, \qquad 0 \leq \kappa \leq 1 \tag{19}$$

We consider a constellation with $M = 64$ symbols and circularity coefficient $\kappa = 0.8$, whose signal points are shown in Fig. 1a. We estimate the SER curve using ALOE and conventional Monte Carlo transmitting only $N = 1280$ symbols at each E_b/N_0 value, i.e., each estimator uses $N = 1280$ samples (we transmit 20 times each of the $M = 64$ symbols). Moreover, we also compare with a standard IS algorithm with a unique Gaussian proposal centered at each symbol $\mathbf{s}$, and variance $\alpha^2 \sigma^2 \mathbf{I}$, with $\alpha > 1$, i.e., the proposal is overdispersed w.r.t. the target $\tilde{\pi}(\mathbf{x})$, since $\mathbb{I}_{\overline{\mathcal{R}}}(\mathbf{x})\tilde{\pi}(\mathbf{x})$ is clearly more dispersed than $\tilde{\pi}(\mathbf{x})$ (see the discussion about the optimal proposal in [29], [30] and the use of overdispersed IS proposals in [31]). We try a grid of values of $\alpha \in \{1, 1.5, 2, ..., 5\}$, and we select the case with the smallest RRMSE. The experiment is repeated 200 times and the RRMSEs defined in (9) are depicted in Fig. 3. For high E_b/N_0 values, ALOE provides four to five orders of magnitude improvement over the usual naive Monte Carlo.

IV. CONCLUSIONS

A multiple importance sampling technique called ALOE is used in this paper to estimate the symbol error rate probability of 2D constellations designed over arbitrary lattices. ALOE is extremely efficient to estimate the integral of a Gaussian in a region defined by a union of half-spaces, which is precisely the error event in a digital communications system. At high signal-to-noise ratios, ALOE provides orders of magnitude speedup with respect to conventional Monte Carlo simulation. As
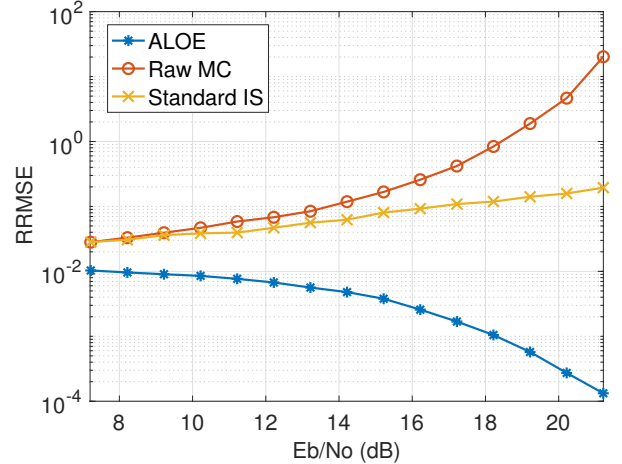

Fig. 3: Relative error variance of SER curves estimated with ALOE, Raw-MC, and a standard IS algorithm.

future work, we plan to extend the proposed SER estimation technique to other noise distributions, to higher dimensional lattices originated by multi-antenna systems, as well as to fading channels. constellations with enhanced features.

APPENDIX

A. Simulation from a truncated Gaussian $\mathcal{N}(\boldsymbol{0}, \mathbf{I})$

The proposals in the MIS implementation are truncated Gaussian distributions. Let us first describe the simulation of a truncated Gaussian $\mathcal{N}(\boldsymbol{0}, \mathbf{I})$ in the half-space described by $\mathbf{x}^T \boldsymbol{\omega} \geq \tau$, which first proceeds by simulating the sample $\mathbf{x}$ from the complementary half-space (i.e., $\mathbf{x}^T \boldsymbol{\omega} < \tau$), and then delivering the $-\mathbf{x}$ for numerical stability. The algorithm described in [19] proceeds as follows:

1) Simulate $\mathbf{z} \sim \mathcal{N}(\boldsymbol{0}, \mathbf{I})$
2) Simulate $u \sim \mathcal{U}(0,1)$
3) Let $y = \Phi^{-1}(u\Phi(-\tau))$, where $\Phi(\tau)$ denotes the cumulative distribution function for the standard Gaussian
4) Let $\mathbf{x} = \boldsymbol{\omega}y + \left(\mathbf{I} - \boldsymbol{\omega}\boldsymbol{\omega}^T\right)\mathbf{z}$
5) Output $\mathbf{x} = -\mathbf{x}$

B. Extension to a generic truncated Gaussian $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

When the target distribution is a generic truncated Gaussian $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ that must be integrated over the union of the half-spaces described by the hyperplanes $\{\boldsymbol{\gamma}_k, \beta_k\}_{k=1}^K$, one can transform the problem as follows. First, we describe K half-spaces with the hyperplanes $\boldsymbol{\omega}_k^T \boldsymbol{x} \geq \tau_k$, where

$$\boldsymbol{\omega}_k = \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\Sigma}^{1/2}}{\sqrt{\boldsymbol{\gamma}_k^T \boldsymbol{\Sigma} \boldsymbol{\gamma}_k}}, \quad \text{and} \quad \tau_k = \frac{\beta_k - \boldsymbol{\gamma}_k^T \boldsymbol{\mu}}{\sqrt{\boldsymbol{\gamma}_k^T \boldsymbol{\Sigma} \boldsymbol{\gamma}_k}}, \tag{20}$$

for $k = 1, ..., K$. Then, a Gaussian $\mathcal{N}(\boldsymbol{0}, \mathbf{I})$ is integrated over the union of those half-spaces as described in previous section.

REFERENCES

[1] G. Forney and G. Ungerboek, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2384–2415, 1998.

[2] K. N. Pappi, A. S. Lioumpas, and G. K. Karagiannidis, "θ-QAM: A parametric quadrature qmplitude modulation family and its performance in AWGN and fading channels," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 1014–1019, 2010.

[3] I. Santamaria, P. M. Crespo, C. Lameiro, and P. J. Schreier, "Information-theoretic analysis of a family of improper discrete constellations," *Entropy*, vol. 20, no. 1, 2018.

[4] M. Grotschel, L. Lovasz, and A. Schriver, *Geometric Algorithms and Combinatorial Optimization*. Springer Verlag, 2nd ed., 1993.

[5] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, 2003.

[6] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge University Press, 2014.

[7] P. Balaban, "Statistical evaluation of the error rate of the fiber-guide repeater using importance sampling," *Bell Syst. Tech. J.*, vol. 55, no. 6, pp. 745–766, 1976.

[8] K. S. Shanmugam and P. Balaban, "A modified Monte-Carlo simulation technique for the evaluation of error rate in digital communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1916–1924, 1980.

[9] D. Lu and K. Yao, "Improved importance sampling technique for efficient simulation of digital communication systems," *IEEE Journal on Sel. Area Comm.*, vol. 6, no. 1, pp. 67–75, 1988.

[10] W. Al-Qaq and J. Townsend, "A stochastic importance sampling methodology for the efficient simulation of adaptive systems in frequency nonselective Rayleigh fading channels,," *IEEE Journal on Sel. Area Comm.*, vol. 15, no. 4, pp. 614–625, 1997.

[11] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: A review of importance sampling techniques in communications systems," *IEEE Journal on Sel. Area Comm.*, vol. 15, no. 4, pp. 597–613, 1997.

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

[13] L. Jacobs and M. Moeneclaey, "Efficient BER simulation of Orthogonal Space-Time Block codes in Nakagami-m fading," in *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications*, Washington, DC, USA, 2014, pp. 676–680.

[14] N. Ben Rached, Z. Botev, A. Kammoun, M.-S. Alouini, and R. Tempone, "Importance sampling estimator of outage probability under generalized selection combining model," in *Int. Conf Acoust., Speech, Signal Process. (ICASSP)*, Calgary, Alberta, Canada, 2018, pp. 3909–3913.

[15] E. Veach and L. J. Guibas, "Optimally combining sampling techniques for Monte Carlo rendering," in *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, 1995, pp. 419–428.

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

[17] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Efficient multiple importance sampling estimators," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1757–1761, 2015.

[18] ——, "Heretical multiple importance sampling," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1474–1478, 2016.

[19] A. B. Owen, Y. Maximov, and M. Chertkov, "Importance sampling the union of rare events with an application to power systems analysis," *arXiv preprint arXiv:1710.06965*, 2017.

[20] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized multiple importance sampling," *Statistical Science (to appear), preprint: arXiv:1511.03095*, 2018.

[21] V. Cadambe, S. Jafar, and C. Wang, "Interference alignment with asymmetric complex signaling—settling the Høst-Madsen-Nosratinia conjecture," *IEEE Trans. Inf. Theory*, vol. 56, pp. 4552–4565, 2010.

[22] Z. Ho and E. Jorswieck, "Improper Gaussian signaling on the two-user SISO interference channel," *IEEE Trans. Wireless Communications*, vol. 11, pp. 3194–3203, 2012.

[23] S. Lagen, A. Agustin, and J. Vidal, "Coexisting linear and widely linear transceivers in the MIMO interference channel," *IEEE Trans. Signal Proc.*, vol. 64, pp. 652–664, 2016.

[24] C. Lameiro, I. Santamaria, and P. Schreier, "Rate region boundary of the SISO Z-interference channel with improper signaling," *IEEE Trans. Commun.*, vol. 65, pp. 1022–1034, 2017.

[25] C. Lameiro, I. Santamaria, and P. J. Schreier, "Benefits of improper signaling for underlay cognitive radio," *IEEE Wireless Comm. Letters*, vol. 4, pp. 22–25, 2015.

[26] O. Amin, W. Abediseid, and M. S. Alouini, "Underlay cognitive radio systems with improper Gaussian signaling: Outage performance analysis," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 4875–4887, 2016.

[27] M. Gaafar, O. Amin, A. Ikhlef, A. Chaaban, and M. Alouini, "On alternate relaying with improper Gaussian signaling," *IEEE Comm. Letters*, vol. 20, pp. 1683–1686, 2015.

[28] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge Univ. Press, 2010.

[29] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer, 2004.

[30] A. B. Owen, *Monte Carlo theory, methods and examples*, 2013.

[31] F. J. R. Ruiz, M. K. Titsias, and D. M. Blei, "Overdispersed black-box variational inference," in *Uncertainty in Artificial Intelligence*, 2016.