Deterministic Sampling of Arbitrary Densities Using Equal Sphere Packing of Volume under the Density (PoVuD)

Daniel Frisch, Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
daniel.frisch@ieee.org, uwe.hanebeck@ieee.org

***Abstract*—We present a new deterministic sampling method for arbitrary densities, unnormalized densities, and likelihoods. Our rejection-free and kernel-free method uses dense equal sphere packing of the volume under the density function (PoVuD). In order to obtain an ensemble that is better than independent random particles, we enforce some local homogeneity.**

I. Introduction

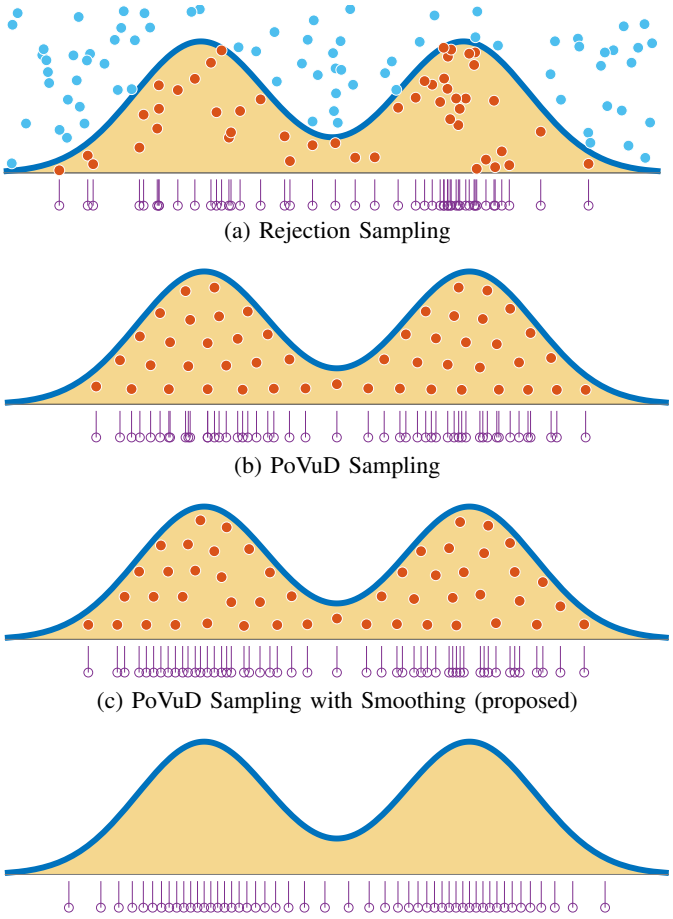
A. Context

Deterministic sampling methods provide sets of samples or particles, ensembles, nodes, abscissas, or "point clouds" that approximate a given density function. These discrete density approximations can be used, e.g., to approximate expected values of nonlinear functions of random variables that could not be computed in closed form. Common application scenarios are nonlinear filtering and control, where deterministic samples produce better results (for the same number of samples) than the often used independent identically distributed (iid) random samples. Deterministic sampling methods for, e.g., Gaussian densities include the unscented transform [1], [2], Gauss-Hermite quadrature [3], samples based on the Localized Cumulative Distribution (LCD) [4], [5], [6], and the projected cumulative distribution (PCD) [7], [8]. If the density function is known to be e.g. Gaussian and also the nonlinear weighting function (such as the likelihood in a Bayesian state estimation setting [9], [10]) is known to be polynomial of some degree, then a tabulated cubature rule with weighted samples can give guaranteed results [11]. Various methods of unweighted deterministic sampling are known for the uniform distribution, most notably quasi-random or low-discrepancy sequences [12], [13]. They can also be transformed to Gaussian samples if certain constraints are met [14], [15].

All these methods generally work well only for certain densities such as the normal density, the uniform density, or for sample reduction with the Dirac mixture density. In the following, we will focus on methods for arbitrary density functions.

B. State-of-Art

One of the basic methods for sampling from arbitrary densities is rejection sampling [16], [17]. It involves sampling



(a) Rejection Sampling

(b) PoVuD Sampling

(c) PoVuD Sampling with Smoothing (proposed)

(d) Inverse Transform Sampling of Equidistant Samples

Fig. 1: Four methods to draw 50 univariate samples (purple) from an arbitrary density function (blue line): classical rejection sampling (a), our proposed methods (b,c), and inverse transform sampling (d) of equidistant samples as "gold standard". Note that random samples as in (a) can also be obtained via Markov Chain Monte Carlo sampling methods for arbitrary density functions. Note also that inverse transform sampling (d) for arbitrary densities is only available in the univariate case. See Fig. 7 for the corresponding cumulatives.

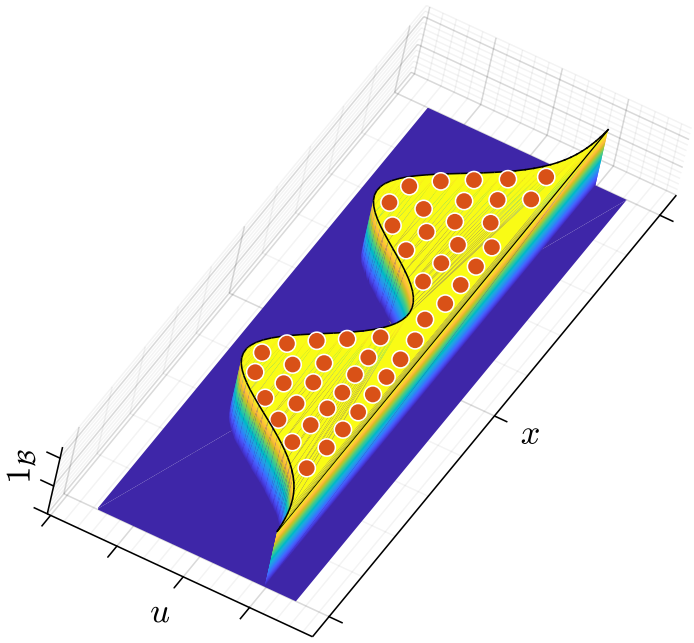
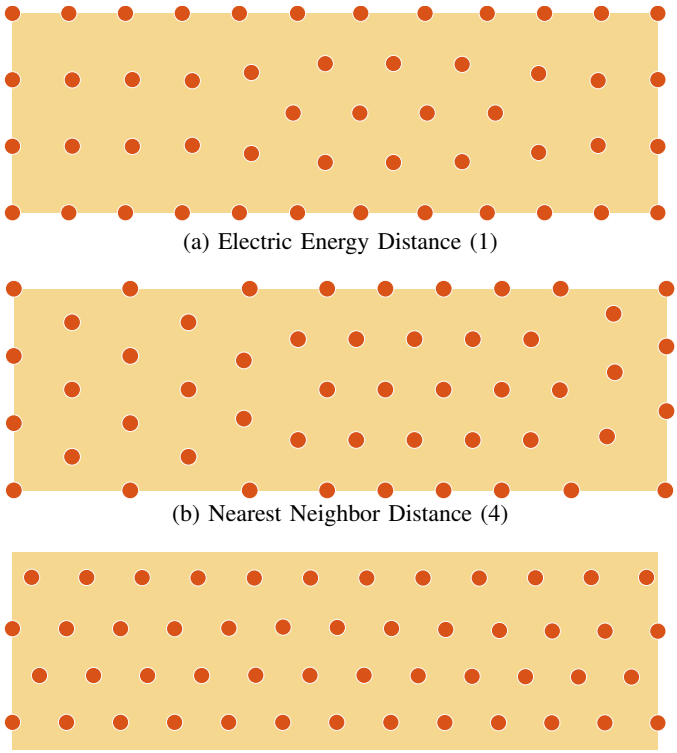


Fig. 2: Given the state space x, some density function $u = f(x)$, and the joint space $\underline{\xi} = [x, u]^\top$, we obtain the area $\mathcal{B}$ under the density function, i.e., the area between the x axis and $f(x)$. Now we draw samples from $1_{\mathcal{B}}$, i.e., the indicator function of that area interpreted as density. After projection to the x axis, we obtain the desired samples from $f(x)$ (not shown here).

from a proposal density and accepting them with a probability proportional to the local density function value. Equivalently, the proposal samples can be augmented with a uniform random value in an auxiliary dimension and accepted if that value is smaller than the density function value, see Fig. 1a. The samples can also be stratified in this setting, i.e., made more locally homogeneous via quasi-random or low-discrepancy sequences [18]. However, the number of samples that have to be rejected for each accepted sample increases exponentially with the number of dimensions. Thus, the method is not applicable to arbitrarily high dimensions. Markov Chain Monte Carlo (MCMC) methods also obtain samples from arbitrary density functions [19]. This does work in higher dimensions, but the samples are again rather random and not locally homogeneously distributed. Finally, if the given density function is expensive to evaluate, it can also be approximated or interpolated via a suitable surrogate model that can be used for both sampling and computation of the desired results [20], [21]. Our method, on the other hand, does without kernel approximation.

C. Key Idea

Our scope is obtaining deterministic samples $\underline{x}_i \in \mathbb{R}^D$ from an arbitrary, not necessarily normalized density function $f(\underline{x})$, $\underline{x} \in \mathbb{R}^D$. In a first step, we compute augmented uniform samples $\underline{\xi}_i \in \mathbb{R}^{D+1}$ from under the density function as shown in Fig. 2. Therefore we solve the corresponding equal hypersphere packing problem, hence the name "Packing of Volume under Density" (PoVuD). After a simple projection,



(a) Electric Energy Distance (1)

(b) Nearest Neighbor Distance (4)

(c) Packing with distance to vertical boundaries (17)

Fig. 3: Simple example: uniform packing of 50 samples in a rectangular region. Space can either be filled completely (a,b) or with some distance to the vertical boundaries (c). Note that minimizing the electric energy distance (a) results in a somewhat higher point density on the boundary.

the resulting samples will have a higher density in regions with higher density function values and vice versa, see Fig. 1c. In total, the samples represent a discrete and deterministic approximation of $f(\underline{x})$.

This principle can potentially be applied to arbitrarily high dimensions and arbitrary density functions as well as likelihoods with finite integral. Computational burden is the computation of an approximate close packing of equal hyperspheres under the density function, i.e., a constrained optimization problem. Note that rejection sampling and Markov Chain Monte Carlo sampling can draw from arbitrary densities as well, but the former is limited to lower dimensions, and the latter to random, unsmoothed samples.

II. Uniform Packing

Close-packing or dense ball-packing of equally sized circles ($D = 1$), spheres ($D = 2$), or hyperspheres ($D > 2$) in the infinite space or in certain bounded regions has been researched mainly in very specific configurations such as circles in squares [22], while we require general nonlinear boundaries for our purposes. We define an optimality measure for uniformity, so we can obtain a sufficiently optimal solution with constrained nonlinear optimization.

A. Optimality Measures

The first idea that may come to mind is minimizing the electric or Coulomb energy of equal point charges

$$\underline{x}_{1:L}^{\text{opt,el}} = \arg\min_{\underline{x}_{1:L}} \left\{ \sum_{i=1}^{L} \sum_{j=1}^{L} \frac{1}{\|\underline{x}_i - \underline{x}_j\|} \right\} \quad , \tag{1}$$
$$\underline{x}_{1:L} = \{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_L\} \quad ,$$
$$\underline{x}_i \in \mathbb{R}^D \quad .$$

However, it turns out that the resulting points are located somewhat more densely on the boundaries than in the interior, see Fig. 3a. This would lead to systematic errors in density approximation, so we continue with other optimality measures. The gold standard is to maximize the minimum distance between any two samples [23], [24]

$$\underline{x}_{1:L}^{\text{opt,gs}} = \arg\max_{\underline{x}_{1:L}} \left\{ \min_{\substack{i,j \\ i \neq j}} \{\|\underline{x}_i - \underline{x}_j\|\} \right\} \quad , \tag{2}$$

which is equivalent to optimizing the dispersion [25, p. 524]. This measure is however not very well suited for gradient-based optimization, as it focuses only on one single pair of points at a time. It can be relaxed by using the electric energy distance with an additional exponent p

$$\underline{x}_{1:L}^{\text{opt,rel}} = \arg\min_{\underline{x}_{1:L}} \left\{ \sum_{i=1}^{L} \sum_{\substack{j=1 \\ j \neq i}}^{L} \left(\frac{1}{\|\underline{x}_i - \underline{x}_j\|} \right)^p \right\} \quad , \tag{3}$$

where for $p \to 1$, we obtain (1), and for $p \to \infty$ it is equivalent to (2). A different way to avoid the electric energy measure populating the boundaries more than the interior is by taking into account the nearest neighbors only

$$\underline{x}_{1:L}^{\text{opt,nn}} = \arg\min_{\underline{x}_{1:L}} \left\{ \sum_{i=1}^{L} \frac{1}{\min\limits_{\substack{j \\ j \neq i}} \{\|\underline{x}_i - \underline{x}_j\|\}} \right\} \quad . \tag{4}$$

This is the optimality measure we will be using in the following. It is good for numerical optimization and yields uniform distributions in the entire region, see Fig. 3b.

B. Uniform Sampling in Arbitrary Domain

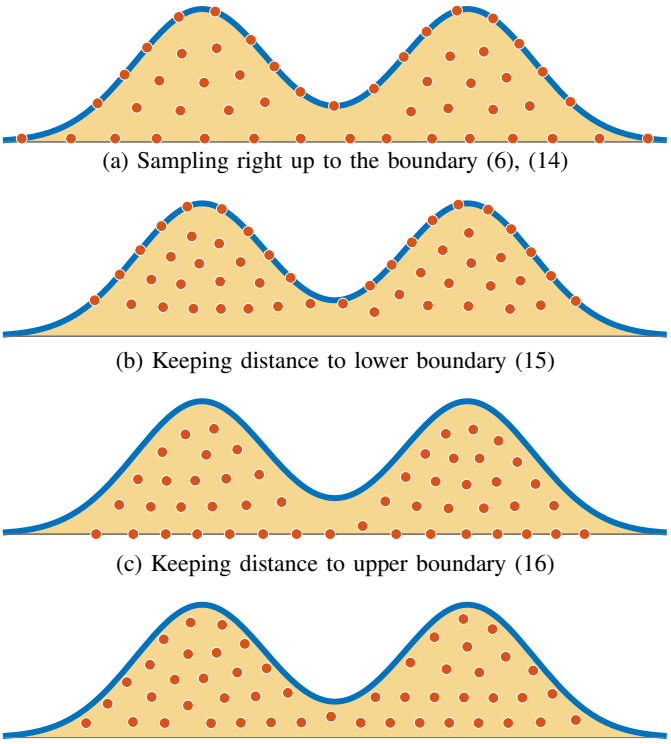
We define the distance measure

$$\Theta(\underline{x}_{1:L}) = \sum_{i=1}^{L} \frac{1}{\min\limits_{\substack{j \\ j \neq i}} \{\|\underline{x}_i - \underline{x}_j\|\}} \quad . \tag{5}$$

By minimizing this distance measure according to

$$\underline{x}_{1:L}^{\text{opt}} = \arg\min_{\underline{x}_{1:L}} \{\Theta(\underline{x}_{1:L})\} \tag{6}$$
$$\text{w.r.t} \quad \underline{x}_i \in \mathcal{B} \quad \forall\, i \quad ,$$

we obtain a distribution of samples $\underline{x}_{1:L}^{\text{opt}}$ uniformly populating the domain $\mathcal{B} \subset \mathbb{R}^D$. See Fig. 3b for an example with $L = 50$ samples where $\mathcal{B}$ is a simple rectangle. For solving the nonlinear constrained minimization problem (6), we used the sequential quadratic programming (SQP) algorithm from the Matlab optimization toolbox [26, Chapter 18].



(a) Sampling right up to the boundary (6), (14)

(b) Keeping distance to lower boundary (15)

(c) Keeping distance to upper boundary (16)

(d) Half distance to upper and lower boundary, respectively (17)

Fig. 4: Various ways to keep some distance to the boundary while packing samples under a density function, see section IV.

III. Deterministic Sampling

A. Univariate Densities ($D = 1$)

In the univariate case, $D = 1$, we augment the scalar sample space $x \in \mathbb{R}$ with a second auxiliary dimension u, yielding a vector $\underline{\xi}$,

$$\underline{\xi} = (u, x) \in \mathbb{R}^2 \quad , \tag{7}$$
$$u = \gamma \cdot f(x) \quad . \tag{8}$$

The auxiliary variable u accounts for the sample's density function value. Note that scaling factor γ can be chosen arbitrarily for now. This means also that $f(x)$ does not need to be normalized, i.e., it can be either a likelihood (with finite integral) or a density (with unit integral by definition).

Now define the domain $\mathcal{B}$ as the region under the (scaled) density function or likelihood

$$\mathcal{B} = \left\{ \underline{\xi} \mid u \leq \gamma \cdot f(x) \right\} \tag{9}$$

and compute uniform samples $\underline{\xi}_{1:L}^{\text{opt}}$ for this area using the distance function (5)

$$\underline{\xi}_{1:L}^{\text{opt}} = \arg\min_{\underline{\xi}_{1:L}} \left\{ \Theta(\underline{\xi}_{1:L}) \right\} \tag{10}$$

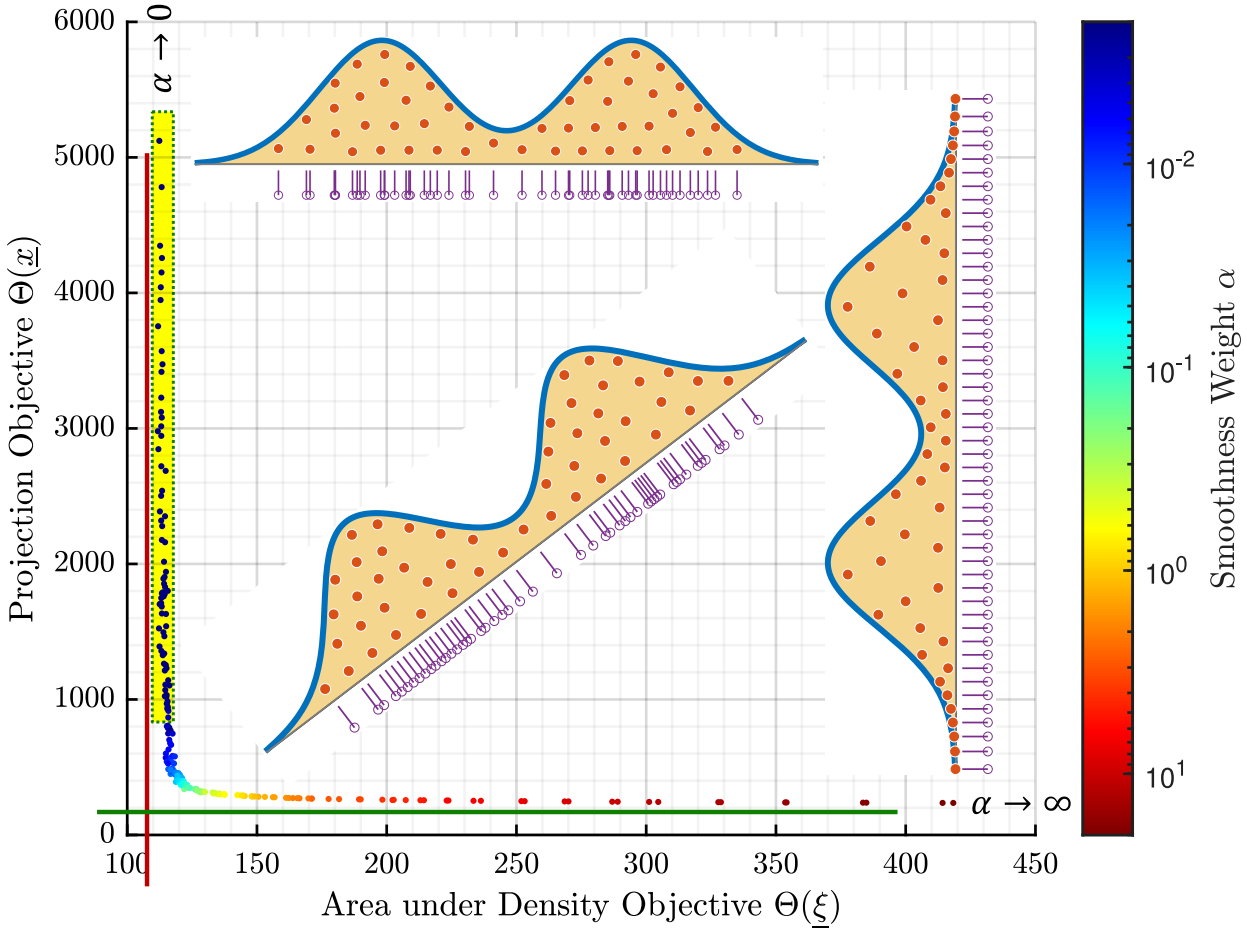


Fig. 5: Pareto front of two objectives $\Theta(\underline{\xi}_{1:L})$ and $\Theta(\underline{x}_{1:L})$. Approximate null space of uniform $\underline{\xi}_{1:L}$ (vertical left area, yellow marking) and corresponding single-objective optimization result for $\alpha \to 0$ in (18) (top middle). Null space of uniform "over-smoothed" $\underline{x}_{1:L}$ (horizontal lower area), with single-objective optimization result for $\alpha \to \infty$ (right, rotated 90°). Best choices in the lower left corner, corresponding optimization result shown rotated by 45°.

$$\text{w.r.t} \quad \underline{\xi}_i \in \mathcal{B} \quad \forall\, i \ .$$

See Fig. 4a for a visual example. After projection, which is equivalent to removing the auxiliary variable u in $\underline{\xi}$ by marginalization, we obtain the resulting samples $x_{1:L}$ that represent $f(x)$.

B. Multivariate Densities

The same principle can be applied to higher dimensions $D > 1$. For $\underline{x} \in \mathbb{R}^D$, we define

$$\underline{\xi} = (u, \underline{x}) \in \mathbb{R}^{D+1} \ , \tag{11}$$

$$\mathcal{B} = \left\{ \underline{\xi} \mid u \leq \gamma \cdot f(\underline{x}) \right\} \ , \tag{12}$$

then compute a packing of samples $\underline{\xi}_{1:L}$ covering the volume $\mathcal{B}$ under the (scaled) density function $\gamma \cdot f(\underline{x})$ uniformly (PoVuD), and obtain $\underline{x}_{1:L}$ being a discrete approximation of $f(\underline{x})$.

IV. Distance to Boundary

Consider the classical rejection sampling, see Fig. 1a. The proposal samples are placed randomly without knowing where the boundary between rejected and accepted samples runs. Therefore, not many samples are placed right on that boundary. Now, packing samples equally into the area under the density function as in Fig. 4a does however cause a significant number of samples to be placed right at the boundary, i.e., the density function. This would introduce a systematic error.

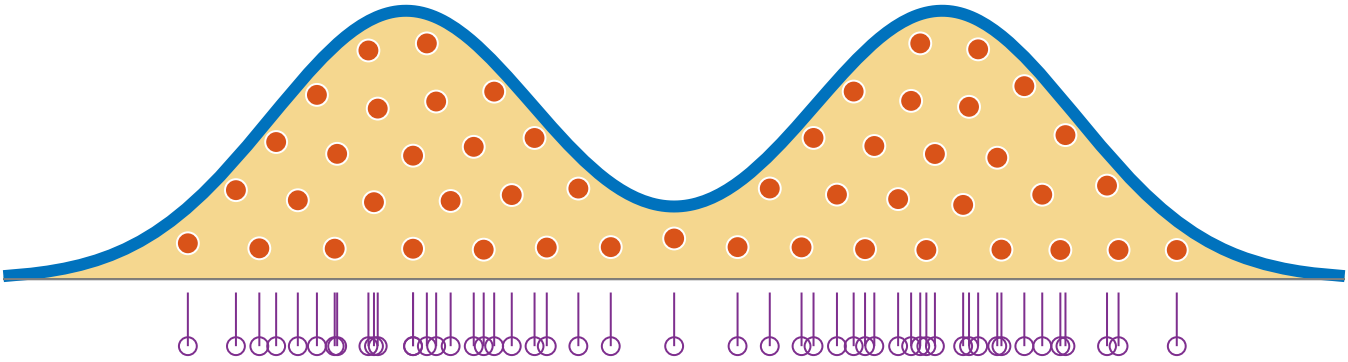
To mitigate this, we augment the Euclidean distance term (14) inside the objective function (5), (13) to include a distance to the boundary, yielding variants (15), (16), and (17)

$$\Theta(\underline{\xi}_{1:L}) = \sum_{i=1}^{L} \frac{1}{\Theta_i} \ , \tag{13}$$

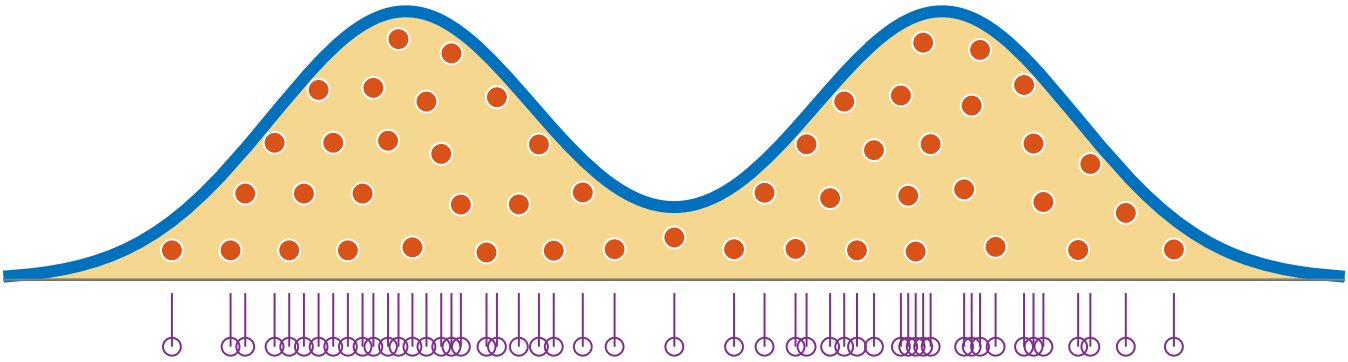
$$\Theta_i^{\text{none}} = \min_{\substack{j \\ j \neq i}} \left\{ \left\| \underline{x}_i - \underline{x}_j \right\| \right\} \ , \tag{14}$$

$$\Theta_i^{\text{lower}} = \min \left\{ \min_{\substack{j \\ j \neq i}} \left\{ \left\| \underline{x}_i - \underline{x}_j \right\| \right\}, u_i \right\} \ , \tag{15}$$

$$\Theta_i^{\text{upper}} = \min \left\{ \min_{\substack{j \\ j \neq i}} \left\{ \left\| \underline{x}_i - \underline{x}_j \right\| \right\}, f(\underline{x}_i) - u_i \right\} \ , \tag{16}$$



(a) PoVuD Sampling



(b) PoVuD Sampling with Smoothing

Fig. 6: Uniformly packed points (red dots) in area (yellow) under density function (blue line). Projections, i.e., the desired samples of the blue density function, are shown in purple.

$$\Theta_i^{\text{both}} = \min\left\{ \min_{\substack{j \\ j\neq i}} \left\{ \left\| \underline{x}_i - \underline{x}_j \right\| \right\}, \frac{f(\underline{x}_i) - u_i}{2}, \frac{u_i}{2} \right\} . \tag{17}$$

This causes the samples to keep the same distance that they have between each other also to the lower boundary (15), the upper boundary (16), or half the distance to both (17), respectively. For visual examples of all the described variants, see Fig. 4. For a visual example including the resulting projected samples, see Fig. 6a and its cumulative in Fig. 7a.

V. Smoothing

We can see in Fig. 6a that the projected points are not distributed locally homogeneously. Often, several samples come to lie very close to each other. This is to be expected, as the point set $\underline{\xi}_{1:L}$ is solely optimized to fill the area under the density as smoothly and uniformly as possible, with no requirements on any projections. A similar effect has also been observed with optimal deterministic Gaussian samples [27, Fig. 4a+5a].

A. Two Objectives

It is evident that many configurations of $\underline{\xi}_{1:L}$ are nearly equally uniformly distributed under the density function, and out of that approximate null space, we want to choose a solution where the projection $\underline{x}_{1:L}$ is locally homogeneous as well. Therefore we introduce a second objective $\Theta(\underline{x}_{1:L})$ that accounts for the smoothness of the projection $\underline{x}_{1:L}$ only. We can solve this multi-objective problem via a new objective function $\widetilde{\Theta}(\cdot)$ that includes both objectives

$$\widetilde{\Theta}(\underline{\xi}_{1:L}) = \Theta(\underline{\xi}_{1:L}) + \alpha \cdot \Theta(\underline{x}_{1:L}) \ . \tag{18}$$

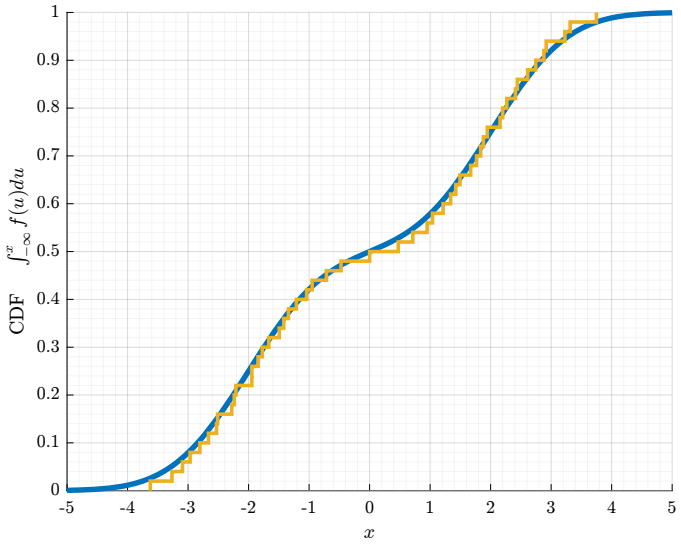
The parameter α must now be chosen small enough that the uniformity of $\underline{\xi}_{1:L}$ is not significantly affected, and large enough that $\underline{x}_{1:L}$ is adequately smoothed.

B. Pareto Front

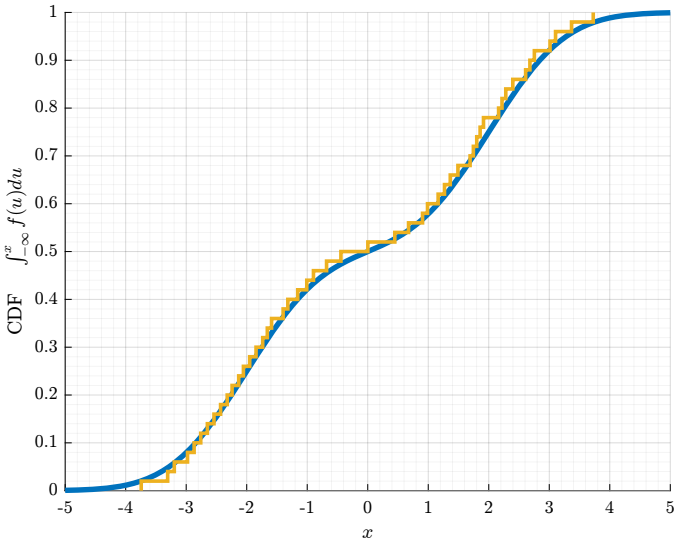
By trying different choices of α, we can inspect the resulting Pareto front, see Fig. 5. It shows how the two objectives interact. In the top left corner, the augmented samples $\underline{\xi}_{1:L}$ are most equidistant but their projection $\underline{x}_{1:L}$ is very inhomogeneous. The lower right corner holds samples where $\underline{x}_{1:L}$ is entirely uniform and not distributed according to $f(\underline{x})$ anymore. The desired solutions where $\underline{x}_{1:L}$ is relatively smooth *and* distributed according to $f(\underline{x})$ are in the lower left corner.

C. Normalization

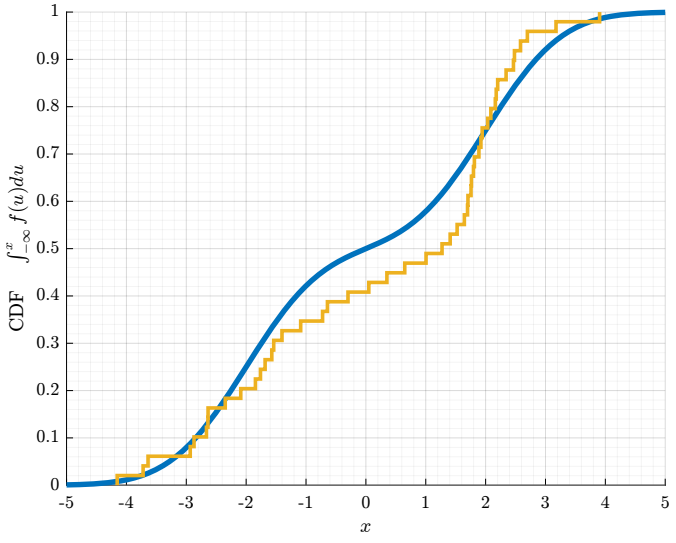
In general, $\Theta(\underline{x}_{1:L})$ attains larger values than $\Theta(\underline{\xi}_{1:L})$ due to the different average distance between points. Therefore, we want to normalize (18) such that we do not need to search for an optimal smoothing factor for every new density, dimension, or number of samples.



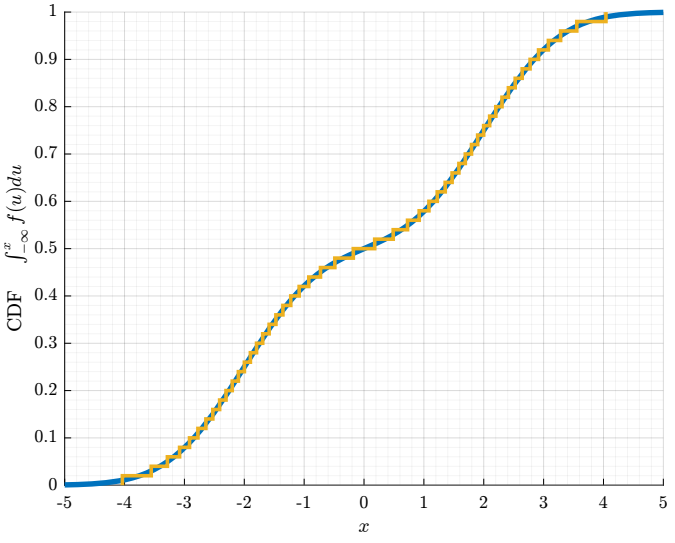
(a) PoVuD Sampling



(b) PoVuD Sampling with Smoothing (proposed)



(c) Random Rejection



(d) Inverse Transform

Fig. 7: Cumulatives of reference density (blue) and Dirac mixture approximation (yellow). This figure serves as a simple evaluation of our proposed sampling method (b). Note that rejection sampling (c) is available only for lower dimensions, and inverse transform sampling (d), at least for arbitrary density functions, only for univariate ones. The corresponding samples and density functions are shown in Fig. 1 and, for the PoVuD samples, also bigger in Fig. 6.

We can always compute, without having to specify a smoothing factor, the two single-objective endpoints representing $\alpha \to 0$ and $\alpha \to \infty$, respectively. They are visualized in Fig. 5 as red vertical and green horizontal line, respectively. With the rule of three, we can now level out the different scaling of both objectives in (18) according to

$$\widetilde{\Theta}(\underline{\xi}_{1:L}) = \Theta(\underline{\xi}_{1:L}) + \beta \cdot \frac{\min_{\underline{\xi}_{1:L}} \left\{\Theta(\underline{\xi}_{1:L})\right\}}{\min_{\underline{x}_{1:L}} \{\Theta(\underline{x}_{1:L})\}} \cdot \Theta(\underline{x}_{1:L}) \ . \tag{19}$$

In other words, we divide α into a universal smoothing factor β and a normalization factor η

$$\begin{aligned} \alpha &= \beta \cdot \eta \ , \\ \eta &= \frac{\min_{\underline{\xi}_{1:L}} \left\{\Theta(\underline{\xi}_{1:L})\right\}}{\min_{\underline{x}_{1:L}} \{\Theta(\underline{x}_{1:L})\}} \ , \end{aligned} \tag{20}$$

where β represents the actual weighting of the smoothing part. Choosing $\beta = 1$ would give the same weight to the smoothing as to the accuracy of the density. However, we put more emphasis on accurate density representation and consider smoothing as a rather minor objective to make optimal use of the remaining degrees of freedom from the null space. In our experiments, $\beta = 0.2$ was, by inspecting the cumulatives,

a good choice to apply some smoothing without significantly distorting the cumulative distribution of the samples. Refer to Fig. 6b for a visual example of the resulting samples, and, as evaluation, Fig. 7b for the cumulatives.

VI. Conclusion

The beauty of our method is its simple, intuitive rationale and that it uses two widely available tools, of which there is a wide selection, namely constrained optimization, where many software implementations exist, and a uniformity measure, that can be taken, e.g., from works in close sphere packing. The method allows for drawing locally homogeneous deterministic samples from arbitrary density functions. Other than rejection-based methods, the computational complexity does not increase exponentially with the dimension, and other than iid random samples or MCMC samples, our ensembles are locally homogeneous and smooth.

However, the constrained optimization problem to be solved is not quite simple. One should try different solvers and parameters to get good results. Furthermore, the probability density shape has an influence. Multiple modes are a problem if the density goes near zero somewhere between them. Narrow peaks only provide space for a "tower" of samples, while wide, flat areas provide space for only one "layer", which may lead to inaccuracies if the number of samples is rather small. To summarize, computation is rather complex and slow, but not exponentially increasing with higher dimensions, and does not require closed-form solutions of any integrals as with kernel-based methods.

VII. Future Work

For optimization, we so far employed sequential quadratic programming (SQP) [26, Chapter 18] as ships with Matlab. However, it is tedious to manually implement the gradients of the objective functions. In order to easily try different configurations of optimality measures and solvers, we will implement the algorithm in Julia [28] and use its automatic differentiation features [29], [30] in the optimization. We also plan to try probabilistic solvers like Simulated Annealing [31]. We will replace the boundary constraints of the volume under the density with a soft repulsion to simplify the constrained optimization problem to an unconstrained one. Poisson Disc Sampling also provides an interesting alternative to obtain uniform samples that are locally homogeneous in given domains [32], even in higher dimensions [33]. We are going to implement our method in higher dimensions, as described in section III-B, and perform extensive simulations on the accuracy. Additional constraints ensuring exact moment matching of mean and covariance could improve the results further. Finally, in a progressive optimization setup, we can start with, e.g., the 10^{th} root of the wanted density function, which flattens narrow peaks [34, Figure 3]. Over the course of the optimization, the density function is then progressively adapted until it reaches the wanted density function in the end [35].

The authors acknowledge support by the state of Baden-Württemberg through bwHPC.

References

[1] S. J. Julier, "The Scaled Unscented Transformation," in *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, vol. 6, May 2002, pp. 4555–4559.

[2] S. J. Julier and J. K. Uhlmann, "New Extension of the Kalman Filter to Nonlinear Systems," in *Signal Processing, Sensor Fusion, and Target Recognition VI*, vol. 3068. International Society for Optics and Photonics, Jul. 1997, pp. 182–193.

[3] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-Time Nonlinear Filtering Algorithms Using Gauss–Hermite Quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.

[4] U. D. Hanebeck and V. Klumpp, "Localized Cumulative Distributions and a Multivariate Generalization of the Cramér-von Mises Distance," in *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)*, Seoul, Republic of Korea, Aug. 2008, pp. 33–39.

[5] U. D. Hanebeck, M. F. Huber, and V. Klumpp, "Dirac Mixture Approximation of Multivariate Gaussian Densities," in *Proceedings of the 2009 IEEE Conference on Decision and Control (CDC 2009)*, Shanghai, China, Dec. 2009.

[6] J. Steinbring, M. Pander, and U. D. Hanebeck, "The Smart Sampling Kalman Filter with Symmetric Samples," *Journal of Advances in Information Fusion*, vol. 11, no. 1, pp. 71–90, Jun. 2016.

[7] U. D. Hanebeck, "Deterministic Sampling of Multivariate Densities based on Projected Cumulative Distributions," in *Proceedings of the 54th Annual Conference on Information Sciences and Systems (CISS 2020)*, Princeton, New Jersey, USA, Mar. 2020.

[8] D. Prossel and U. D. Hanebeck, "Dirac Mixture Reduction Using Wasserstein Distances on Projected Cumulative Distributions," in *Proceedings of the 25th International Conference on Information Fusion (Fusion 2022)*, Linköping, Sweden, Jul. 2022.

[9] J. H. Kotecha and P. M. Djuric, "Gaussian Sum Particle Filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2602–2612, Oct. 2003.

[10] V. Elvira, L. Martino, and P. Closas, "Importance Gaussian Quadrature," *IEEE Transactions on Signal Processing*, vol. 69, pp. 474–488, 2021.

[11] R. Cools and P. Rabinowitz, "Monomial Cubature Rules Since "Stroud": A Compilation," *Journal of Computational and Applied Mathematics*, vol. 48, no. 3, pp. 309–326, 1993.

[12] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, 1992.

[13] J. Dick, F. Y. Kuo, and I. H. Sloan, "High-Dimensional Integration: The Quasi-Monte Carlo Way," *Acta Numerica*, vol. 22, pp. 133–288, 2013.

[14] D. Frisch and U. D. Hanebeck, "Deterministic Gaussian Sampling With Generalized Fibonacci Grids," in *Proceedings of the 24th International Conference on Information Fusion (Fusion 2021)*, Sun City, South Africa, Nov. 2021.

[15] ——, "The Generalized Fibonacci Grid as Low-Discrepancy Point Set for Optimal Deterministic Gaussian Sampling," *Journal of Advances in Information Fusion*, 2023.

[16] J. Von Neumann, "Various Techniques Used in Connection With Random Digits," *John von Neumann, Collected Works*, vol. 5, pp. 768–770, 1963.

[17] B. D. Flury, "Acceptance–Rejection Sampling Made Easy," *SIAM Review*, vol. 32, no. 3, pp. 474–476, 1990.

[18] D. Frisch and U. D. Hanebeck, "Rejection Sampling from Arbitrary Multivariate Distributions Using Generalized Fibonacci Lattices," in *Proceedings of the 25th International Conference on Information Fusion (Fusion 2022)*, Linköping, Sweden, Jul. 2022.

[19] C. J. Geyer, "Practical Markov Chain Monte Carlo," *Statistical science*, pp. 473–483, 1992.

[20] F. L. Fernández, L. Martino, V. Elvira, D. Delgado, and J. López-Santiago, "Adaptive Quadrature Schemes for Bayesian Inference via Active Learning," *IEEE Access*, vol. 8, pp. 208 462–208 483, 2020.

[21] J. Westermann, A. Zea, and U. D. Hanebeck, "Adaptive Sampling for Global Meta Modeling Using a Gaussian Process Variance Measure," in *Proceedings of the 2021 European Control Conference (ECC 2021)*, Virtual, Jun. 2021.

[22] P. G. Szabó, M. C. Markót, and T. Csendes, *Global Optimization in Geometry — Circle Packing Into the Square*. Boston, MA: Springer US, 2005, pp. 233–265.

[23] R. L. Graham, B. D. Lubachevsky, K. J. Nurmela, and P. R. J. Östergård, "Dense Packings of Congruent Circles in a Circle," *Discrete Mathematics*, vol. 181, no. 1, pp. 139–154, 1998.

[24] M. Locatelli and U. Raber, "Packing Equal Circles in a Square: A Deterministic Global Optimization Approach," *Discrete Applied Mathematics*, vol. 122, no. 1, pp. 139–166, 2002.

[25] H. Niederreiter, *A quasi-Monte Carlo method for the approximate computation of the extreme values of a function*. Basel: Birkhäuser Basel, 1983, pp. 523–529. [Online]. Available: https://doi.org/10.1007/978-3-0348-5438-2_45

[26] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

[27] D. Frisch and U. D. Hanebeck, "Efficient Deterministic Conditional Sampling of Multivariate Gaussian Densities," in *Proceedings of the 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2020)*, Virtual, Sep. 2020.

[28] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM review*, vol. 59, no. 1, pp. 65–98, 2017.

[29] J. Revels, M. Lubin, and T. Papamarkou, "Forward-Mode Automatic Differentiation in Julia," *arXiv:1607.07892 [cs.MS]*, 2016.

[30] P. K. Mogensen and A. N. Riseth, "Optim: A Mathematical Optimization Package for Julia," *Journal of Open Source Software*, vol. 3, no. 24, p. 615, 2018.

[31] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[32] M. S. Ebeida, A. A. Davidson, A. Patney, P. M. Knupp, S. A. Mitchell, and J. D. Owens, "Efficient Maximal Poisson-Disk Sampling," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4, pp. 1–12, 2011.

[33] R. Bridson, "Fast Poisson Disk Sampling in Arbitrary Dimensions," *SIGGRAPH sketches*, vol. 10, no. 1, p. 1, 2007.

[34] J. Steinbring and U. D. Hanebeck, "Progressive Gaussian Filtering Using Explicit Likelihoods," in *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*, Salamanca, Spain, Jul. 2014.

[35] D. Frisch and U. D. Hanebeck, "Progressive Bayesian Filtering with Coupled Gaussian and Dirac Mixtures," in *Proceedings of the 23rd International Conference on Information Fusion (Fusion 2020)*, Virtual, Jul. 2020.