# Visualization of rare event and importance sampling proposal in high-dimensional space

**Tianyu ZHANG**
Department of Civil and Environmental Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
ty.zhang@connect.ust.hk

**Hanzhe CUI**
Department of Civil and Environmental Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
hcuiah@connect.ust.hk

**Wenxi QIU**
Department of Industrial Engineering and Decision Analytics
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
wqiuae@connect.ust.hk

**Erjia FU**
Department of Industrial Engineering and Decision Analytics
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
efuaa@connect.ust.hk

## Abstract

Visualizing rare events and importance sampling distributions in high-dimensional spaces is a challenging task. While two-dimensional visualizations provide valuable insights, they become infeasible for problems with three or more dimensions. Researchers in the field of adaptive importance sampling (AIS) have primarily relied on quantitative metrics to assess the quality of importance distributions. In this report, we explore a new perspective by employing dimensionality reduction algorithms to visualize rare events and importance distributions in high-dimensional spaces. By compressing the data into a two-dimensional space, we can intuitively evaluate the quality of importance distributions generated by AIS methods. We investigate three specific methods: Importance Conditional Expectation (ICE) with a Single Gaussian (ICE-SG), ICE with a Gaussian Mixture (ICE-GM), and ICE with the von Mises–Fisher–Nakagami mixture (ICE-vMFNM). We apply these methods to two datasets: a series system dataset and a heat diffusion dataset. The visualizations obtained through dimensionality reduction provide a qualitative assessment of the importance distributions, complementing the quantitative metrics used in the literature. This new perspective enhances our understanding of the quality of importance distributions and verifies the results of existing quantitative approaches.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

# 1 Introduction

Reliability analysis is a critical component in the stochastic evaluation of engineering systems, which aims to quantify the probability of a system reaching predefined failure states. Formally, consider a system with $n$-dimensional uncertain model parameters $\mathbf{x} \in \mathbf{X}$, and let $f$ denotes the failure event for which the limit state function $g(\mathbf{x})$ takes non-positive values. The probability of failure $P_f$ can be defined as:

$$P_f = \int_{\mathbb{R}^n} \mathbb{I}\left(g(\mathbf{x}) \leq 0\right) p(\mathbf{x}) d\mathbf{x} \tag{1}$$

where $p(\mathbf{x})$ is the joint probability distribution function (PDF) for the uncertain variables, $\mathbb{I}(\cdot)$ is the indicator function (1 if the condition is met and 0 otherwise).

The crude Monte Carlo estimation of $P_f$ requires prohibitively many samples to achieve an acceptable accuracy. For example, for $P_f$ around $1 \cdot 10^{-5}$, at least $10^7$ samples are required to get an reliable estimate. Therefore, importance sampling is often used to estimate $P_f$. The importance sampling estimator of $P_f$ is given by:

$$P_f \approx \frac{1}{N} \sum_{i=1}^{N} \frac{p(\mathbf{x}_i)}{q(\mathbf{x}_i)} \mathbb{I}(g(\mathbf{x}_i) \leq 0) \tag{2}$$

where $q(\mathbf{x})$ is the importance distribution, from which $\mathbf{x}_i$ is sampled. According to the importance sampling theory, the optimal proposal distribution is given as follows:

$$q^*(\mathbf{x}) = \frac{\mathbb{I}(g(\mathbf{x}) \leq 0) p(\mathbf{x})}{\int_{\mathbb{R}^n} \mathbb{I}(g(\mathbf{x}) \leq 0) p(\mathbf{x}) d\mathbf{x}} \tag{3}$$

which will lead to 0 estimation variance. However, the optimal proposal distribution is unknown in practice since the denominator is the quantity of interest. Therefore, adaptive importance sampling (AIS) is used to approximate the optimal proposal distribution. In this report, we will consider the improved cross entropy method (ICE) proposed by Papaioannou et al. [2019], combined with different density models, which is currently one of the most advanced AIS algorithm.

For $n = 2$ problems, both failure domain and $q(\mathbf{x})$ can be directly visualized, providing valuable insights into AIS behavior. However, visualization becomes challenging for problems where $n \geq 3$. In the field of AIS, researchers have traditionally relied solely on quantitative metrics such as relative error and coefficient of variation to evaluate importance distribution quality. This report will show how dimension reduction algorithms can help to assess the quality of importance distributions generated by ICE.

# 2 Problem formulation and Methodology

Given a limit state function $g(\mathbf{x})$ and an input density $p(\mathbf{x})$, a dataset $\{X, y\}$ will be generated using crude Monte Carlo sampling, where $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ is a set of $N$ samples drawn from $p(\mathbf{x})$, and $y = \{g(\mathbf{x}_1), g(\mathbf{x}_2), \ldots, g(\mathbf{x}_N)\}$ is the corresponding set of limit state function evaluations. A dimension reduction algorithm $h : \mathbb{R}^n \to \mathbb{R}^2$ will be fitted on the dataset $X, y$, mapping the high-dimensional input space to a two-dimensional space.

Let $q(\mathbf{x})$ denote the importance sampling density obtained from the ICE algorithm. A new set of samples $X_1 = \{\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_M\}$ will be generated from $q(\mathbf{x})$, and the fitted dimension reduction algorithm $h$ will be applied to $X_1$, projecting the samples into the low-dimensional space:

$$h(X_1) = \{h(\mathbf{x}'_1), h(\mathbf{x}'_2), \ldots, h(\mathbf{x}'_M)\} \tag{4}$$

By visualizing the positions of the $h(X_1)$ in the two-dimensional space, we can gain insights into the quality of the importance sampling density $q(\mathbf{x})$. This approach allows for a more intuitive understanding of the high-dimensional importance sampling distribution, complementing the traditional quantitative metrics such as relative error and coefficient of variation used in AIS algorithms, which especially lack information about the geometrical structure of $q(\mathbf{x})$.

# 3 Datasets

In this section, we present the dimension reduction results on two datasets. We investigate three AIS methods: ICE with a Single Gaussian (ICE-SG), ICE with a Gaussian Mixture (ICE-GM), and ICE

with the von Mises–Fisher–Nakagami mixture (ICE-vMFNM). The details of these methods can be found in Papaioannou et al. [2019].

For dimensionality reduction, we primarily employ Uniform Manifold Approximation and Projection (UMAP)McInnes et al. [2020], a nonlinear technique for visualizing high-dimensional data. UMAP consists of two steps. Firstly, it computes a graph representing the input data. Secondly, it learns the embedding for the graph in the first step. Specifically, we will evaluate three UMAP algorithms: 1. Unsupervised UMAP, which constructs the graph without using label information. 2. Supervised UMAP, which uses the label information during graph construction by increasing the affinity between points sharing the same label. 3. Parametric UMAP, which replaces the second step by minimizing the same objective function as UMAP, but learning the relationship between the data and embedding using a neural network (here we use a 3-layer 100-neuron fully connected neural network). Here, for label information during the step of fitting UMAP, we only use the sign of $g(\mathbf{x})$, and accordingly use terminologies positive sample (related to the failure event) and negative sample (related to the non-failure event).

### 3.1 Series system dataset

In this example, we consider an analytical limit state function given as follows,

$$g_1(\mathbf{x}) = \min \left\{ \begin{array}{l} \beta - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \\ \beta + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \end{array} \right\} \tag{5}$$

where we set $n = 100$, leading to a 100-dimensional problem. The input density is standard Gaussian. We choose $\beta = 3.5$ which results in a failure probability of $P_f = 4.65 \cdot 10^{-4}$.

We generate the original dataset of size $10^6$ using crude Monte Carlo. To address the high class imbalance, we downsample the dataset by retaining only 5000 negative class instances while keeping all positive class instances. This improves the class balance, resulting in a final dataset of approximately 5500 samples, which serves as the training dataset for our analysis.

The results of dimension reduction to the 2D space on the training set are displayed in Fig. 1. In the absence of label information, unsupervised UMAP fails to effectively distinguish between negative and positive samples, resulting in a lack of clear separation between the two classes. However, when label information is provided, both supervised and parametric UMAP demonstrate the ability to effectively separate the dataset into the two desired clusters.

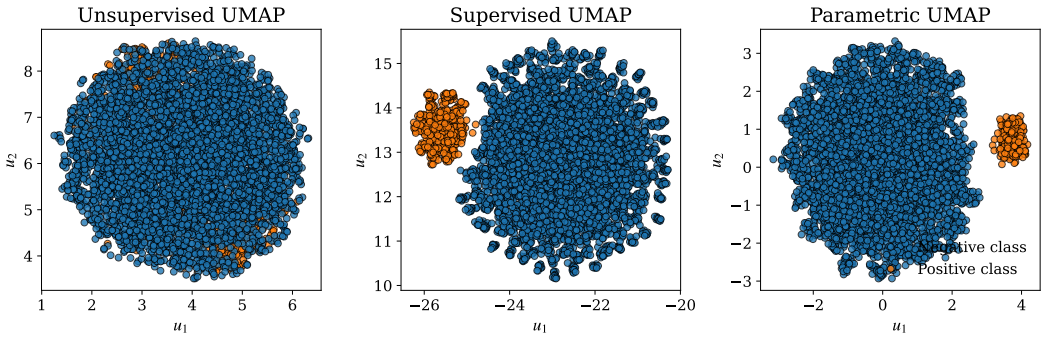


Figure 1: UMAP dimensionality reduction comparison for training set of series system dataset: unsupervised, supervised, and parametric approaches.

Next, we apply the fitted UMAP algorithms to the original dataset generated by crude Monte Carlo. The results are presented in Fig. 2. Both algorithms perserve the high-dimensional structure of the original dataset. Different to the results in Fig. 6, both algorithms display a continuous change in the value of $g(\mathbf{x})$, which is because we include sufficiently many negative samples in the training step. This shows a trade-off between the geometrical structure and effective discrimination, which are respectively continuous change and clearly separated clusters in our examples.

The embedding of samples from ICE-SG, ICE-GM, and ICE-vMFNM to the 2D space using parametric UMAP is illustrated in Fig. 3. In this example, the learning processes of both ICE-SG

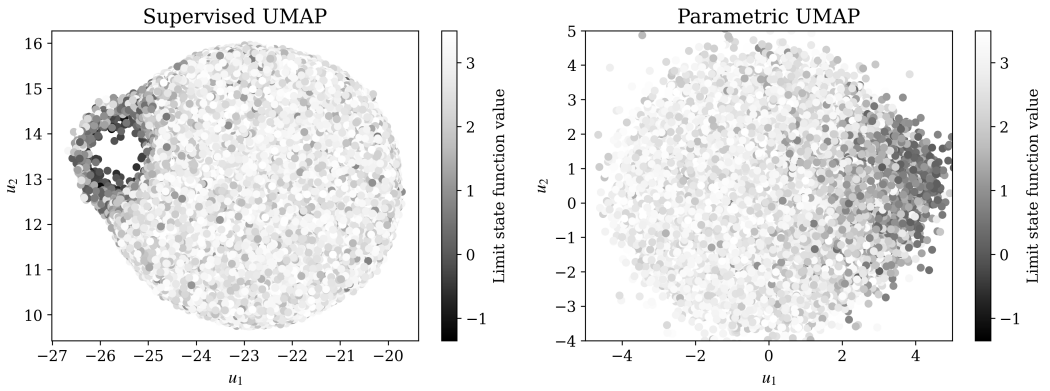


Figure 2: UMAP dimensionality reduction comparison for original series system dataset: supervised, and parametric approaches.

and ICE-GM generate samples that converge to a single point, indicating a significant failure in capturing the data distribution. In contrast, the visualization corresponding to ICE-vMFNM exhibits a substantial portion of failure samples, suggesting a strong capability to model the failure domain in high-dimensional space.

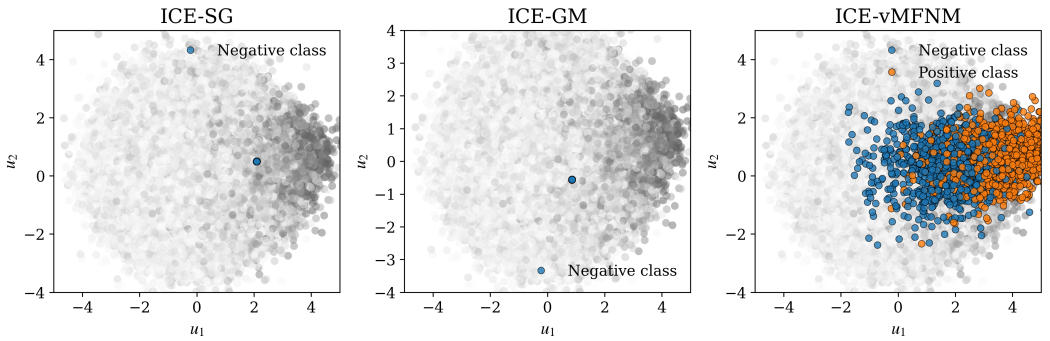


Figure 3: Parametric UMAP dimensionality reduction comparison for ICE dataset: ICE-SG, ICE-GM, and ICE-vMFNM. The result of original series system dataset are added for comparison.

### 3.2 Heat diffusion dataset

In this example, we consider a planar FEM example, as indicated in Fig. 4. The target variable is the average temperature within a small square domain $B = (-0.3, -0.2)$ m $\times$ $(-0.3, -0.2)$ m. We define the failure event as $f = \{T \geq 7.5°\text{C}\}$ with a reference failure probability of $1.11 \cdot 10^{-3}$. For more details of this example, see Konakli and Sudret [2016]. The input variables follow the standard Gaussian density in 53-dimensional space.

We generate the original dataset of size $10^5$ using crude Monte Carlo. To address the high class imbalance, we downsample the dataset by retaining only 1000 negative class instances while keeping all positive class instances. This improves the class balance, resulting in a final dataset of approximately 1400 samples, which serves as the training dataset for our analysis.

The results of dimension reduction for the 2D space on the training set are displayed in Fig. 5. The unsupervised UMAP results show that the two classes form distinct clusters, indicating inherent separability in the high-dimensional space, despite some overlap. Supervised UMAP and parametric UMAP significantly improve class separation by using label information. Supervised UMAP brings same-class samples closer and pushes different-class samples apart. Parametric UMAP further optimizes the manifold structure, minimizing within-class distances and maximizing between-class distances, resulting in more compact and distinct low-dimensional representations.

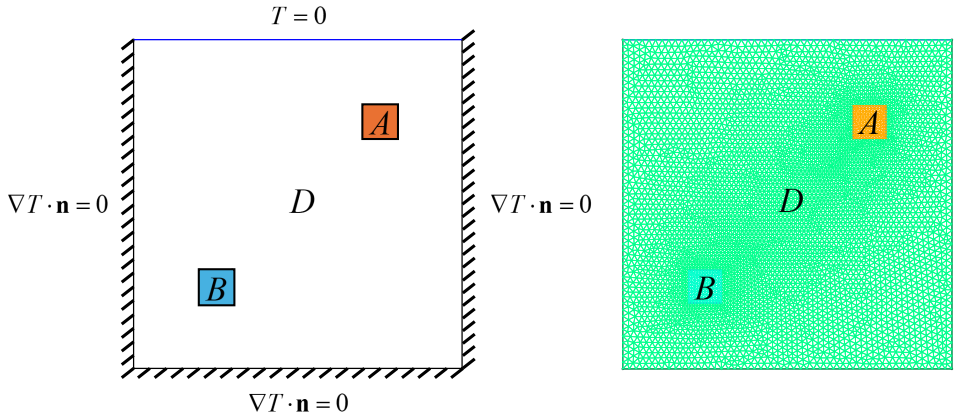


Figure 4: Two-dimensional heat conduction example. Left: domain geometry and boundary conditions; Right: finite element mesh.

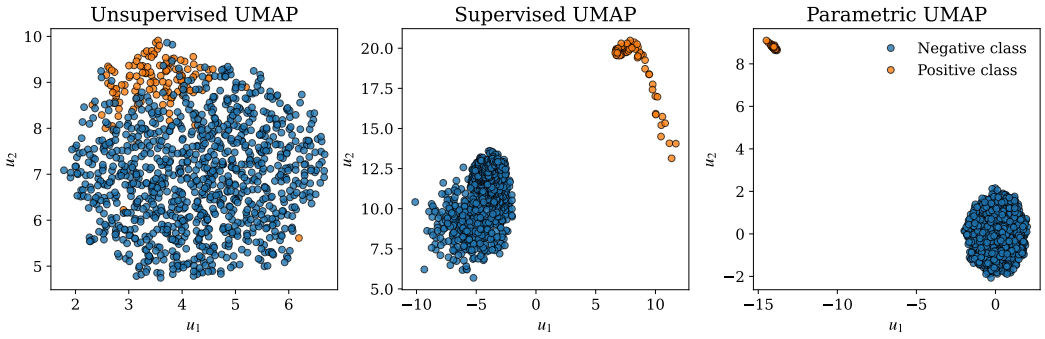


Figure 5: UMAP dimensionality reduction comparison for training set of heat diffusion dataset: unsupervised, supervised, and parametric approaches.

Next, we apply the fitted UMAP algorithms to the original dataset generated by crude Monte Carlo. The results are presented in Fig. 6. Note that for each new data point, unsupervised and regular supervised UMAP finds its nearest neighbors within the original training set in the high-dimensional space, followed by approximately embedding the new data point into the target low-dimensional space. In many cases we have observed, the performance of these two methods is poorer than parametric UMAP does on this task. Hence, we focus on supervised UMAP and parametric UMAP. The results demonstrate that both algorithms effectively preserve the original high-dimensional structure of the original dataset. However, parametric UMAP exhibits superior performance, as its low-dimensional representation of the original dataset displays a continuous change in the value of $g(\mathbf{x})$, which is expected. This continuity in the low-dimensional space reflects the inherent continuous nature of $g(\mathbf{x})$ in the high-dimensional space, highlighting the advantage of parametric UMAP in capturing and preserving the underlying data structure.

The embedding of samples from ICE-SG, ICE-GM, and ICE-vMFNM to the 2D space using parametric UMAP are illustrated in Fig. 7. The results indicate that for this particular example, none of the sampling methods has generated failure samples, suggesting that they all exhibit suboptimal performance in this case. The absence of failure samples in the low-dimensional representation highlights the limitations of these sampling techniques in capturing the critical regions of the high-dimensional space that contribute to system failures. This emphasizes the need for more effective sampling strategies that can better explore and identify the failure regions, enabling more comprehensive reliability analysis and design optimization.

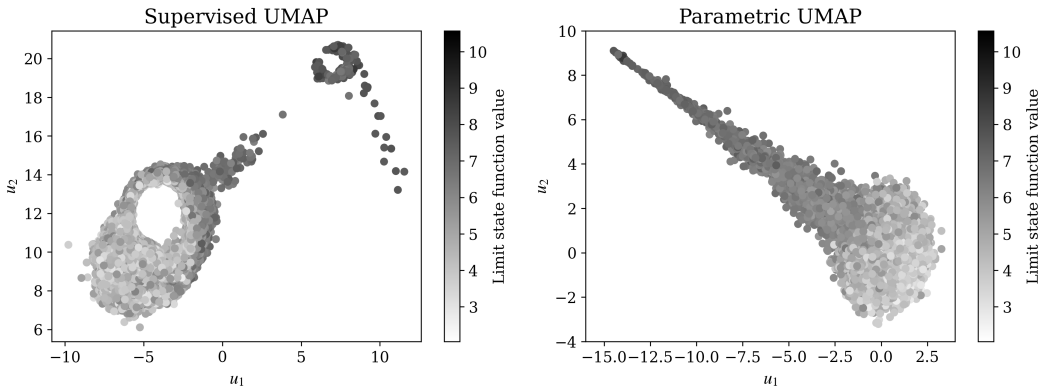


Figure 6: UMAP dimensionality reduction comparison for original heat diffusion dataset: supervised, and parametric approaches.

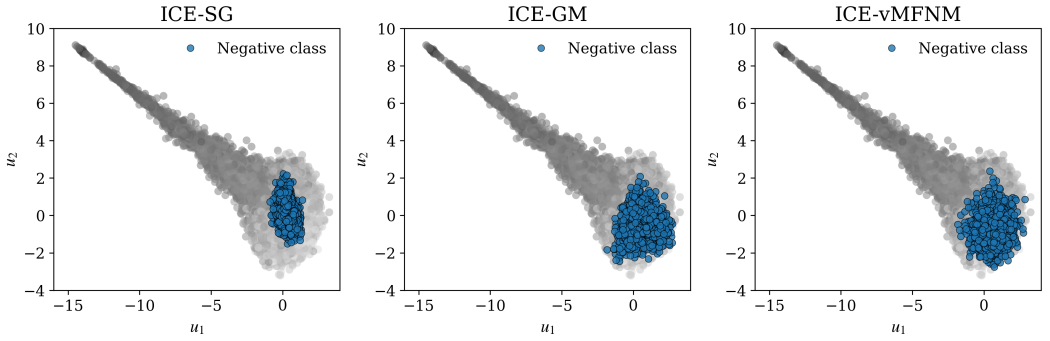


Figure 7: Parametric UMAP dimensionality reduction comparison for ICE datasets: ICE-SG, ICE-GM, and ICE-vMFNM. The result of the original heat diffusion dataset is displayed for comparison.

## 4 Conclusions

This report explored the application of dimensionality reduction algorithms as a means to visualize and qualitatively assess importance sampling distributions in high-dimensional spaces, a task traditionally dominated by quantitative metrics. We specifically employed unsupervised, supervised, and parametric UMAP to visualize samples related to rare events defined by a series system (100-dimensional) and a heat diffusion model (53-dimensional), comparing the importance distributions generated by ICE-SG, ICE-GM, and ICE-vMFNM methods.

Our results indicate that dimensionality reduction, especially using supervised or parametric UMAP which leverage label information (sign of $g(\mathbf{x})$), can effectively create low-dimensional representations that distinguish between failure and non-failure regions. Parametric UMAP often yields embeddings that better reflected the underlying structure. Critically, these visualizations provided direct qualitative feedback on the AIS methods' performance. For the series system problem, the visualization illustrated the failure of ICE-SG and ICE-GM, contrasting with the more successful performance of ICE-vMFNM in generating samples appropriately. For the heat diffusion problem, none of the tested ICE methods appeared to adequately cover the failure region in the embedding.

Therefore, we conclude that dimensionality reduction serves as a valuable diagnostic tool for AIS. It allows for an intuitive grasp of the geometric properties of the generated importance distributions, complementing quantitative metrics and offering insights into structures of different AIS strategies when tackling high-dimensional rare event probability estimation.

## 5 Contribution

Tianyu ZHANG: Provided the idea for this report, wrote the report and slides, wrote code, and coordinated the work among team members.
Hanzhe CUI: Wrote the code for Example 1 and delivered part of the presentation.
Wenxi QIU: Wrote the code for Example 2, delivered part of the presentation, and contributed to writing example 2 report.
Erjia FU: Reviewed and verified the code.

## 6 Code

Link: `https://github.com/jonas-tzhang/CSIC5011_project`

## References


Katerina Konakli and Bruno Sudret. Global sensitivity analysis using low-rank tensor approximations. *Reliability Engineering & System Safety*, 156:64–83, 2016.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020. URL `https://arxiv.org/abs/1802.03426`.

Iason Papaioannou, Sebastian Geyer, and Daniel Straub. Improved cross entropy-based importance sampling with a flexible mixture model. *Reliability Engineering & System Safety*, 191:106564, 2019.