# POF-Darts: Geometric Adaptive Sampling for Probability of Failure

Mohamed S. Ebeida[a,*], Scott A. Mitchell[a], Laura P. Swiler[a], Vicente J. Romero[a], Ahmad A. Rushdi[a,b]

*[a]Sandia National Laboratories, Albuquerque, NM, USA*
*[b]Institute for Computational Engineering and Sciences, University of Texas, Austin, TX, USA*

## Abstract

We introduce a novel technique, POF-Darts, to estimate the Probability Of Failure based on random disk-packing in the uncertain parameter space. POF-Darts uses hyperplane sampling to explore the unexplored part of the uncertain space. We use the function evaluation at a sample point to determine whether it belongs to failure or non-failure regions, and surround it with a protection sphere region to avoid clustering. We decompose the domain into Voronoi cells around the function evaluations as seeds and choose the radius of the protection sphere depending on the local Lipschitz continuity. As sampling proceeds, regions uncovered with spheres will shrink, improving the estimation accuracy. After exhausting the function evaluation budget, we build a surrogate model using the function evaluations associated with the sample points and estimate the probability of failure by exhaustive sampling of that surrogate. In comparison to other similar methods, our algorithm has the advantages of decoupling the sampling step from the surrogate construction one, the ability to reach target POF values with fewer samples, and the capability of estimating the number and locations of disconnected failure regions, not just the POF value. We present various examples to demonstrate the efficiency of our novel approach.

*Keywords:* Probability of Failure, Percentile Estimation, Reliability, Computational Geometry, Surrogate Models

## 1. Introduction

### 1.1. Problem Statement

Estimating the probability of failure based on computational simulation models is a challenging problem in several engineering applications including device and system design [3, 4, 5, 6], structural and reliability analysis [7, 8, 9], fault-tree analysis [10, 11, 12], and financial systems [13, 14, 15].

The probability of failure, $P$, quantifies the probability that a "failure" condition occurs, where failure is defined by the value of some scalar function, $f(x)$, falling above/below a threshold $T$, e.g., the probability that a device exceeds a certain temperature when subject to varying environmental conditions: $P(f(x) > T)$, or the voltage of a circuit node falls below a certain point when power supply noise fluctuates: $P(f(x) < T)$. A few design constraints contribute to the difficulty of estimating $P$, including:

• Dimensionality: Probability of failure problems are often more difficult when the domain $\mathcal{D}$ (the parameter space of $f(x)$) is of high dimension, $d$. In real applications involving hundreds or thousands of parameters, often the value of $f(x)$ is dominated by only a few of them. Therefore, $f(x)$ can be replaced by a lower dimensional function involving just the significant parameters, but this is not always easy to achieve.

• Noise: If $f(x)$ is noisy or has discontinuities, evaluating it at a point returns little about its behavior within a neighborhood of the point, at which characterizing the function becomes difficult and expensive.

• Cost: The number of required samples, which is proportional to the number of points $x$ where $f(x)$ is evaluated, is a key metric for evaluating the running time of a probability of failure method. Cost is a critical challenge, especially when a single evaluation may require a finite-element simulation to solve an implicit limit state function. Cost also is a challenge for the estimation small failure probabilities, e.g. Monte Carlo methods require more samples to provide predictions with small
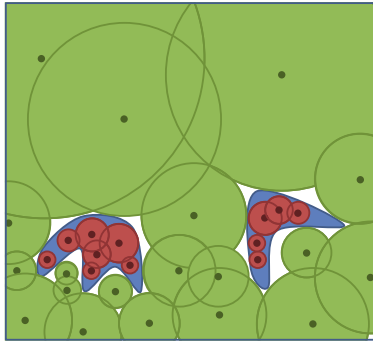
(a) POF-Darts.

*"Puff and Darts" game, circa 1902, courtesy FCIT.*

*"POF Darts" researchers hard at work, circa 2013*

(b) "Puff and Darts" courtesy of FCIT [1].

Figure 1: The name of our algorithm, POF-Darts, is inspired by the old game of "puff the dart," [2] a.k.a. "puff and darts,". POF-Darts (left) throws random disks and determines the radius of each one based on the function evaluation at the disk center relative to the failure threshold and an estimate of the Lipschitz continuity associated with the underlying function.

statistical uncertainty as the magnitude of the failure probability decreases.

• Topology: Quantifying $P$ is often associated with concluding the parameter values leading to failure, the location of the failure region, and its shape and topology. As the failure probability decreases, locating the failure region becomes a bigger challenge.

### 1.2. Probability Of Failure Estimation Approaches

Approaches to estimate the probability of failure of a computational simulation model can be, in general, categorized into a few groups, such as:

† **Direct Sampling Methods.** Monte Carlo (MC) sampling is probably the most widely used approach for calculating uncertainty and failure probability [16, 17, 18, 19]. MC sampling methods have many advantages: they are simple, reproducible, easy to implement, and work well with a variety of sensitivity analysis procedures. Their main drawback is expense. A large number of samples are needed to get good estimates of small probabilities. Many approaches have been designed to overcome MC sampling's expense. Latin Hypercube Sampling (LHS) is a stratified MC sampling method. It ensures that samples are placed in equiprobable strata for each input parameter and "mixed" across dimensions. Compared to standard Monte Carlo sampling, LHS tends to give better coverage of the input space, especially for small sample sizes. Also, LHS output statistics have lower variance, especially for separable functions [20, 21]. However, LHS and sampling-based methods in general are very expensive as they require a high sample budget to cover the entire domain. Importance sampling [22] involves sampling from a density which concentrates the samples in the important (failure) region and weights those samples relative to the original density of the input parameters. However, importance sampling requires a priori knowledge of the failure region, which is often not known.

† **Surrogate-Based Approaches.** Surrogate models (also called meta-models or emulators) are commonly used to understand output responses of systems in several UQ problems [23]. Surrogates, in general, are very useful in understanding trends and sensitivities [24]. They are called surrogates because they serve as a substitute for evaluating the original function using a few function evaluations. Among the popular surrogates, Gaussian process (GP) models [25] (also called Kriging models) are governed by a covariance function, provide a spatial interpolation and an estimate of uncertainty at new prediction points, and are guaranteed to go through the points to which they are fitted. Other surrogate models include radial basis functions, smoothing splines, neural networks, polynomial regression, etc. Using surrogates to solve probability of failure problems can be tricky, however, for their sensitivity to inaccuracies around the failure boundary.

Another surrogate approach is to construct stochastic expansions which represent stochastic responses. For example, the Polynomial Chaos Expansion (PCE) is based on a multidimensional orthogonal polynomial approximation. In non-intrusive PCE for black-box functions, the calculation of chaos expansion coefficients for response metrics of interest is based on a set of simulation response evaluations. The calculation of these coefficients is usually performed using regression methods or spectral projection methods. The regression approach finds the set of PCE coefficients which

best match the simulation model responses. The spectral projection approach projects the response against each basis function using inner products and employs the polynomial orthogonality properties to extract each coefficient. Each inner product involves a multidimensional integral which can be evaluated numerically using sampling, tensor-product quadrature, or sparse grid approaches. Much work in the past decade has focused on efficient calculation of these coefficients using sparse grids and adaptive grid methods [26, 27, 28, 29].

† **Approximate Methods.** There are a variety of approximate methods such as FORM and SORM which approximate the limit state function with first or second order expansions, respectively. [30, 22]. These methods originated in civil and structural engineering, e.g. to determine when a probabilistic load would exceed a probabilistic capacity for a structure. These methods are often more efficient at computing low probability events, the tail statistics, than sampling. To make probability calculations more tractable, the user-specified uncertain variables are transformed to standard normal variables, i.e. independent Gaussian random variables with mean zero and variance one. The original variables may be non-normal and correlated. In the transformed space, probability contours are circular. Also, the multidimensional integrals which define the POF calculation can be approximated by simple functions of a single parameter, $\beta$, called the reliability index. $\beta$ is the minimum Euclidean distance from the origin in the transformed space to the failure boundary. This point is also known as the Most Probable Point (MPP) of failure. There are global and local reliability methods. Global methods can find multiple MPPs if they exist [31, 32, 33]. Local methods use local optimization to locate one MPP. Subset simulation [34, 35] is a reliability based method for estimating small failure probabilities, converting the small region problem into a series of larger conditional probabilities.

† **Hybrid Methods.** Several estimation approaches combine existing methods. For example, LHS-GP is a global Gaussian process surrogate built on LHS sample points instead of Monte Carlo sample points. (Note: there are some methods which adaptively select sample points based on the prediction variance of the GP. In this paper, when we refer to LHS-GP we mean a non-adaptive version, where a GP is built over a fixed LHS sample). EGRA [32] is a global reliability method designed to overcome some of the limitations of local reliability methods. The EGRA method begins with a GP emulator using a very small number of LHS samples, and then adaptively chooses where to generate subsequent samples in an attempt to increase the emulator accuracy in the vicinity of the failure boundary. The resulting GP model is then sampled using multimodal adaptive importance sampling to calculate the probability of failure. By locating multiple points on or near the failure boundary, complex and nonlinear boundaries can be modeled, allowing a more accurate POF estimate. Because EGRA concentrates samples in the area where accuracy is important (i.e. in the vicinity of the failure boundary), it is relatively efficient in number of samples required.

### 1.3. Paper Contribution

The known hybrid sampling-surrogate methods do not have all of our desired features (e.g., efficiency and accuracy). For example, LHS sampling is accurate but inefficient. Accuracy comes from covering the whole domain to find the failure regions. But this is also inefficient, requiring many samples, and these are not placed adaptively. At the other extreme, EGRA is very efficient because it uses the information gained from previous samples to guide future placement, and hence requires few samples. Its drawback is generality: it is tied to a specific GP surrogate that does not accurately approximate noisy or discontinuous functions.

In this paper, we present a new approach to failure probability estimation, called POF-Darts; see Figure 1. Our key target is to quickly cover the entire domain with sample points surrounded by protective spheres of variable radii. We focus on an adaptive sampling phase which uses basic concepts from computational geometry and random sampling to speed up the convergence towards the failure regions. Each time we throw a new sample point, we surround it with a sphere based on information we estimate about the Lipschitz condition which bounds the function value change within the sphere. Then, based on this bounding value, we categorize that sphere as "green" when the sphere is in the no-failure region and "red" when the sphere is in the failure region. We use spheres to cut off wide swaths of space requiring no further exploration. Additionally, line-guided sampling improves the efficiency of exploring the remainder of the domain. Every time we throw a new point, the radii of the spheres around the existing points are re-sized based on this new information added.

Once the sample budget in exhausted, we build a surrogate on top of the sample points to approximate the function everywhere and sample the surrogate exhaustively to spot the failure regions and estimate the failure probability. In this way, we decouple the sampling step from the surrogate construction and evaluation steps. This decoupling enables the freedom of selecting a proper surrogate model.

3

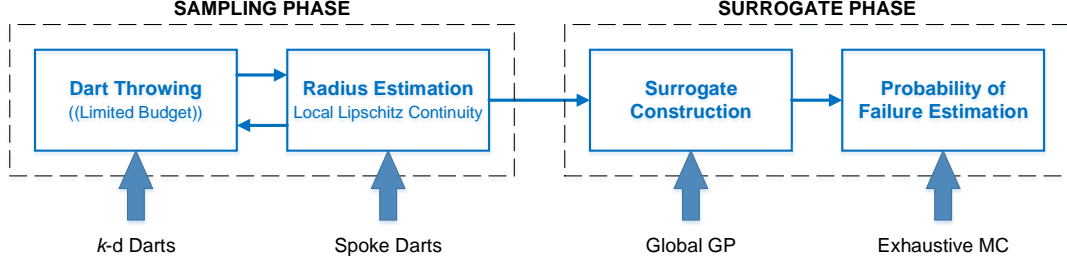| SAMPLING PHASE | | SURROGATE PHASE | |
| --- | --- | --- | --- |
| **Dart Throwing** ((Limited Budget)) | **Radius Estimation** Local Lipschitz Continuity | **Surrogate Construction** | **Probability of Failure Estimation** |
| *k*-d Darts | Spoke Darts | Global GP | Exhaustive MC |

Figure 2: An overview of the steps of POF-Darts.

To further motivate the POF-Darts approach, we highlight a preliminary comparison with MC in Table 1. In this example, we estimate a small failure probability of $10^{-4}$ for the function given in Equation (7). The table shows that Monte-Carlo is not accurate until $10^6$ sample points are evaluated. In fact, for a 100/1000 MC samples, no failures were found as no samples hit the failure region(s), resulting in a failure probability of 0. In contrast, POF-Darts accurately obtained the correct failure probability with 1000 samples. Further details of the algorithm and complete experimental results are to follow in sections 2 and 4, respectively.

| # samples: | 1E2 | 1E3 | 1E4 | 1E5 | 1E6 |
| --- | --- | --- | --- | --- | --- |
| MC: | 0 | 0 | 3E-4 | 5E-5 | 9E-5 |
| POF-Darts: | 8E-5 | 1E-4 | 1E-4 | 1E-4 | 1E-4 |

Table 1: Convergence comparison between POF-Darts and Monte Carlo's probability of failure estimates, when used to sample the function defined in Equation (7), for a true failure probability = $10^{-4}$.

### 1.4. Algorithm Overview

We start with a *Sampling Phase*. We "throw" darts sequentially until we reach the sampling budget. With each sample, the domain is re-classified into failure, not-failure, and unexplored subregions. We use k-d Darts in this step. When a sample is placed, a protective sphere surrounds it to avoid future sampling in its close vicinity. New samples are only inserted in the unexplored regions forcing it to shrink. By decomposing the domain into the corresponding Voronoi diagram, we estimate the radius of the protection sphere around a new sample based on the size of its Voronoi cell and its local Lipschitz continuity. We approximate the local Lipschitz continuity as the maximum gradient between the function value at the sample and at its significant Voronoi neighbors found using spoke-darts.

After exhausting the sample budget, the *Surrogate Phase* follows. We build a surrogate using the functions evaluations at the sample points inserted in the sampling phase. Then, we measure the volume of the failure subregion by exhaustively sampling the surrogate using Monte Carlo sampling.

These steps are summarized in Figure 2. Further details about each step and supporting algorithms will follow in Section 2.

### 1.5. POF-Darts Features

POF-Darts is generally a global method; it does not rely on the underlying function being convex as several methods do to guarantee convergence. By decoupling sampling and surrogate construction, the method enables tunable efficient designs. The sampling phase of POF-Darts is especially efficient because POF-Darts tend to generate larger disks first. This forces areas far from failure boundaries to be quickly covered and ignored. A new disk may not only overlap, but actually contain the center of an old disk of the same *color*, where red disks are around failure points, and green disks around not-failure. The disk radius approaches zero at the failure boundary, and red and green disks are disjoint. The uncovered space tends to be concentrated near the failure boundary, where we are most interested in gaining more information about the function's behavior by another sample. The entire domain will never be covered, i.e. there will always be room for another sample, unless all the disks are the same color. The user may specify an arbitrary budget of function evaluations. However, we can enrich the sample set with any number of new samples if the budget increases. Our method can always make use of data from a sample point, regardless of its location, even if it was generated by another method. A more conservative estimate of the Lipschitz continuity of the underlying function would force

all disks to shrink leaving more room to add new samples efficiently.

## 1.6. Organization of Paper

Section 2 gives the algorithmic details of POF-Darts. We provide analytic bounds on the convergence of POF-Darts in Section 3. Section 4 presents experimental results that uses a number of analytic functions and real engineering case studies in different dimensions, with comparisons to LHS-GP, EGRA, and PCE. Finally, Section 5 summarizes the major conclusions of this work and outlines future directions.

## 2. POF-Darts Algorithm Details

In this section, we describe the details of the POF-Darts algorithm, focusing on the novel adaptive sampling phase. Algorithm 1 demonstrates the pseudocode listing of the algorithm.

---

**Algorithm 1:** Pseudocode of POF-Darts.

**Input**: input parameters $x^d$, with bounds
**Input**: $T$ failure threshold
**Input**: Simulation to produce response $f(x)$
**Input**: $N$ number of samples of simulation
**Output**: Probability of Failure Estimate: $P =$ Prob(Response $\leq T$)
**begin**
  **for** $i = 1$ **to** $N$ **do**
    Find a point i outside prior spheres, using $k$-d Darts ;
    **while** *no points found to add* **do**
      reduce radii of existing spheres ;
      re-run kd-darts placement to find i ;
    **end**
    evaluate $f(x)$ at point i;
    find the significant neighbors of i using Spoke-Darts ;
    compute the local Lipschitz constant at i ;
    estimate the radius of the sphere at i ;
    update the local Lipschitz constant at i's neighbors ;
    update the radii of the spheres centered around i's neighbors ;
  **end**
  Calculate $P$ ;
    Construct Surrogate over the $N$ points ;
    Evaluate Surrogate $N_s$ times (e.g., 1E+6) ;
    $P = (\sum \# \text{ surrogate responses} \leq T)/N_s$ ;
**end**

---

### 2.1. Sampling Phase

Our method is based on sampling, yet is different than MC and LHS because the decision of where to place a new sample depends on the locations of previous samples. We iteratively place a sample, protect its neighborhood with a disk/sphere associated with a classification and a radius. This sphere both prevents the introduction of new sample points within it, and covers part of the domain. Adaptivity makes it possible to approach the failure region quickly, accurately, and often with a lower sample budget.

A popular adaptive sampling approach is Maximal Poisson-disk Sampling (MPS) [36], where samples are added and surrounded by disks/spheres until the whole domain is fully covered. New sample points are sequentially and uniformly selected from the remainder of the domain not already covered, the *void*. In a *maximal* sampling, this proceeds until all of the domain is covered. Poisson-disk sampling with a constant disk radius produces a half-radius sphere packing, while variable-radii Poisson-disk Sampling [37] is a non-uniform version of MPS where the radius of a disk depends on some function of the position of its center. Poisson-disk sampling, however, poses serious algorithmic and implementation challenges in high dimensions. Moreover, in a probability of failure problem, coverage maximality is not necessarily required, especially with a limited sample budget. To efficiently address high-dimensional problems, an adaptive sampling solution is $k$-d Darts, which is a sampling method that uses $k$-dimensional dart searches [38] to sample high-dimensional domains, with no background data structures. Most similar algorithms use some sort of background grid to represent high-dimensional domains, while $k$-d Darts does not. Therefore, it is most suitable for exploring and tracking the failure regions in a domain as part of POF-Darts. Basically, $k$-d Darts evaluates a function along a $k$-dimensional dart. In general, a $k$-d dart is a generalization of a point sample to higher dimensions, and is defined as a set of flats of dimension $k$, where $k = 0$ is points, $k = 1$ is lines, $k = 2$ is planes, etc. The flats are embedded in $d$ dimensions, and aligned with the coordinate directions for simplicity and efficiency. A $k$-flat has $(d-k)$ fixed coordinates, and $k$ free coordinates. One dart has ($d$ choose $k$) flats, one flat for each combination of fixed coordinates.

Figure 3(a–c) explains that the higher the dimension of a dart, the more adept it is at finding (intersecting) a small region of interest. Figure 3(d) provides an illustrative example of $k$-d darts in 2d where lines are used to sample planes and then points are used to sample the uncovered regions of that line. At each iteration, the

new dart is subdivided into segments that are not covered by any prior disks, and one new point is uniformly selected from these segments surrounded by a new disk centered there, declaring a *hit*. A *miss* is when there are no such segments, at which we proceed to the next dart throw. Note that in general, there are $O(n)$ subsegments, which can be generated in $O(n \log n)$ time. In low dimensions there are good ways of speeding up proximity checks, but these become less effective as the dimension increases. Upon applying the $k$-d darts component within POF-Darts (in contrast to basic MPS), it is important to note that the radii of the disks are not constant, which enables fast convergence by placing large disks in areas far from the failure regions. Also, maximality of domain coverage is not the target. The termination criteria is typically exhausting the sample budget.

### 2.1.1. Disk Category and Radius Estimation

As a new sample is added at $x_i$, we implicitly construct the corresponding Voronoi diagram around the sample points as "seeds". The Voronoi diagram is a space partitioning technique where each partition or cell combines all points that are closer to its seed than all other seeds. We then compute two scalar quantities:

- The function evaluation $f(x_i)$. The sign of $f(x_i) - T$ determines whether it is a red failure disk or a green not-failure disk.

- The Lipschitz continuity, approximated as the *local* maximum gradient

$$g_i = \max_j \frac{|f(x_i) - f(x_j)|}{\|x_i - x_j\|} \qquad (1)$$

for every pair of Voronoi neighbor samples $j$. We use the Spoke Darts algorithm [39] to approximately find the significant neighbors of any Voronoi cell/seed.

The radius of the newly added sphere is given by

$$r(x_i) = \frac{|f(x_i) - T|}{\alpha g_i} \qquad (2)$$

where $T$ is the failure threshold, and $\alpha$ is a scale factor typically set to 2 to account for the error resulting from calculating the Lipschitz continuity at a discrete number of neighbors (rather than a continuous range).

Note that the computation of the Lipschitz continuity is not computationally demanding. The Voronoi neighbors of each point are found with the Spoke Darts algorithm which employs random dart throwing and efficiently finds the neighbor cells. We then use complete enumeration over all neighbor points in Eq. 1 to find the maximum gradient: we do not use an optimization algorithm. We use this maximum gradient to update the radius around point $x_i$ as shown in Eq. 2. In a local operation, the neighbors of $x_i$ also update their radii using the same rule in Eq. 2. If all the neighbors had the same function value (up to a small tolerance), $\alpha g_i$ could be very close to 0, which results in spheres with very large radii. To avoid this problem, we lower bound $\alpha g_i$ at a value of 0.1. Finally, note that the failure threshold $T$ can be zero or any real number less than infinity. The case of $T=0$ can represent an implicit limit state.

### 2.1.2. Improving Disk Radius Estimation

To further improve the estimate of the disk radius $r(x)$, we apply a few more rules to adjust sphere radii:
**Mitigating Estimation Errors.** There are a few possible types of radius estimation errors, mainly: overestimation and underestimation; see Figure 4 for a pictorial overview. Overestimation causes inaccuracy, and underestimation causes inefficiency. This could be controlled by changing the value of $\alpha$ in (2). If the $k$-d darts algorithm reaches a maximum number of misses, that means we have over-estimated radii. We, then, multiply $\alpha$ by 3/2, forcing radii to decrease, and apply (2) to all existing disks. Shrinking all disks opens up more space for throwing more $k$-d darts successfully.
**Avoiding Overlapping Disks.** If red and green disks overlap, this shows that the estimation of the radius was too large; see Figure 4. We use a simple rule that rectifies the overlap locally and concentrates future samples where our models and assumptions have proven to be inaccurate. We revise each pair of opposite-color disks sequentially as they are discovered. We use a linear approximation of the function between the two disk centers $a$ and $b$. We define slope $L = (f(a) - f(b))/\|a - b\|_2$, and new radii $r(a) = |f(a)|/L$ and $r(b) = |f(b)|/L$. This results in the two disks being just tangent.
**Correcting Covered Failure Regions.** If a non-failure disk is too large, it may completely cover a failure region. A large failure disk overlapping the non-failure region also produces inaccuracy. We occasionally (1 in 25 samples) add the farthest Voronoi vertex of an approximate cell, which may lie inside a disk. This sample may correct an overestimated radius.

### 2.2. Surrogate Phase

### 2.2.1. Surrogate Construction

After reaching the user-specified function evaluations budget, we build a surrogate based on the disk centers to approximate the function everywhere in the domain.
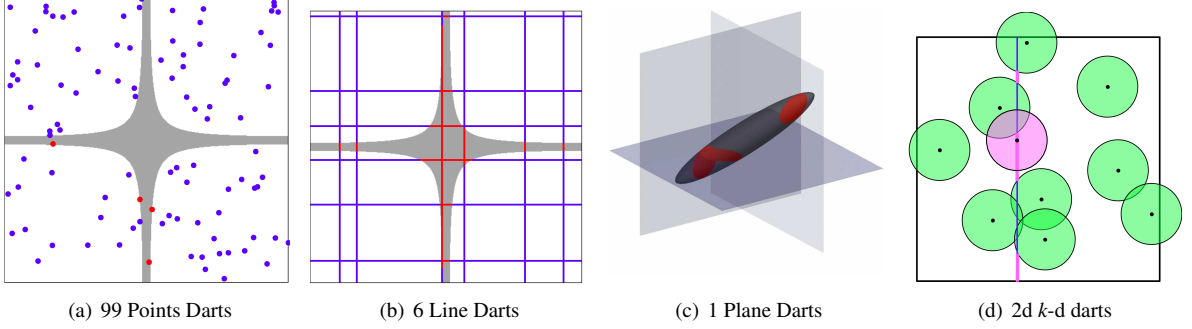
(a) 99 Points Darts (b) 6 Line Darts (c) 1 Plane Darts (d) 2d $k$-d darts

Figure 3: $k$-d darts sampling using points (a), lines (b), and planes (c). $k$-d darts of higher dimension are better at intersecting the region of interest (red), especially narrow regions. To place a new 2d disk using $k$-d darts (d): we generate a random line, subtract the disks of previous samples to find the uncovered part(s) of the line, and uniformly choose the center of the new disk from the uncovered part(s).



(a) Guaranteed non-failure (b) Underestimated radius (c) Optimal radius

(d) Guaranteed-failure radius (e) Overestimated radius (f) Resolving conflicts
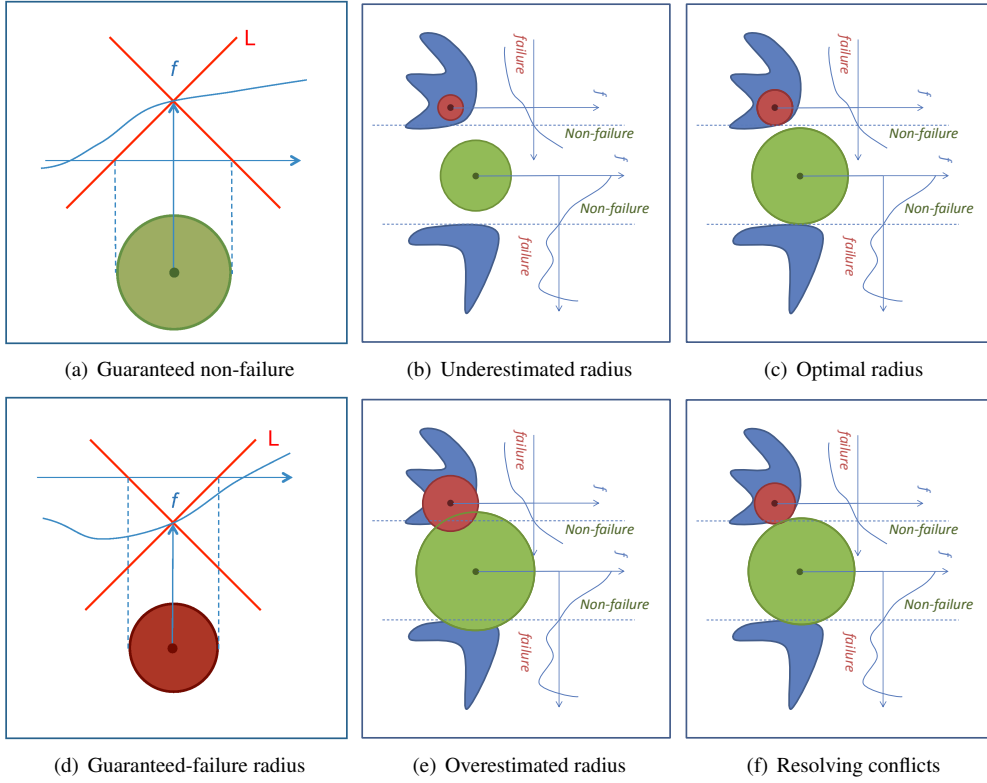
Figure 4: If the global Lipschitz constant $L$ is known, then a neighborhood disk of radius $|f(x) - T|/L$ is conservative, and guaranteed to be consistently non-failure (a) or failure (d). The optimal radius (c) is the maximum that still prevents crossing the failure threshold within the disk. We estimate the radius with an estimate of the local Lipschitz continuity. Underestimating (b) results in less coverage and hence more function evaluations. Overestimating (e) could result in crossing the failure threshold, and conflicts between failure and non-failure disks. Conflicts (f) are detected and mitigated by geometrically shrinking the disks. If the overestimate is too large, a non-failure disk can cover a failure region completely, which decreases the accuracy of the POF. We reduce the chance of that happening by conservative estimates, adjusting radii, and occasionally inserting the farthest approximate Voronoi vertex.

7

Since our sampling phase is decoupled from the surrogate construction, several surrogate models can be easily employed at this point. In general, we use *Global Surrogate models*, where the function value approximation anywhere depends on all available samples. Gaussian Processes (GP) models are fairly standard, with a squared exponential correlation function. The correlation lengths are calculated by maximum likelihood estimation.

Different surrogate models have different features and capabilities as we will point out in Section 4, but a complete study of the surrogate selection for the estimation of the probability of failure is outside the scope of this paper.

### 2.2.2. Probability of Failure Estimation

Once a surrogate model is available, we use exhaustive Monte Carlo sampling to estimate the probability of failure from the surrogate directly. Figure 5 demonstrates that using a GP surrogate is an efficient way to classify the unexplored regions in the coverage phase even with few function evaluations.



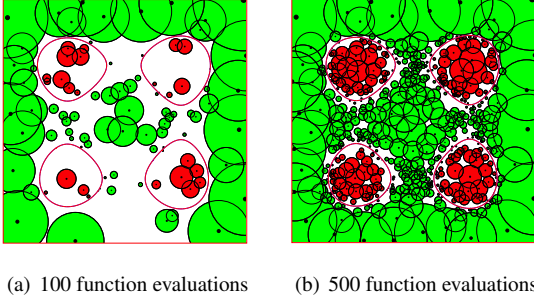(a) 100 function evaluations     (b) 500 function evaluations

Figure 5: Applying POF-Darts to a test problem with 4 failure regions. The estimated failure isocontours (red) and the exact ones (blue) are almost on top of each other here indicating an accurate estimation.

## 3. Convergence Analysis

Here we study the bounds on the expected accuracy of our $P$ estimates. We will provide arguments leading to the following two conclusions. First, the algorithm is efficient except perhaps when the failure function is inherently difficult to explore, because the function value is near the failure threshold over much of the domain. (An example is a function with a large noise component). Second, the algorithm converges very quickly (exponentially) once a sufficient number of samples has filled up most of the domain where the function value is far from failure. The analysis is based on the intuition

that disks are only small near the failure threshold, so the area far from the failure threshold is quickly filled up with big disks. The analysis quantifies that statement. Yet, as with many worst-case theory bounds, the exact constants here are not tight. The main value of the analysis is the rates in the bounds. We parameterize the user-desired *accuracy* by two parameters: $t$ and $k$. Here $t$ is a threshold of near-failureness, and $k$ is multiplicative factor of accuracy. We parameterize the function *difficulty* by $A_t$, $L$, and the dimension $d$ (indirectly). Let the domain area be 1. Without loss of generality, let the failure threshold be 0. Let $A(R)$ denote the area $A$ of a $d$-dimensional region $R$. Let $A_t$ be the area of the subdomain where $|f(t)| < t$, that is, where the function is close to the failure threshold. The larger this area is, the more difficult it is for our method to estimate the POF accurately. (Note the lack of dependence on the area of the failure region itself, just a dilation of its boundary. Figure 3 shows an example where $A_t$ is large compared to the actual failure region because the function is nearly flat and close to failure.) Let $\overline{A_t}$ be the complement of $A_t$. Let $j$ be the number of samples we have generated, and let $i$ be the number in $\overline{A_t}$, i.e. where $|f(x_i)| > t$ and we are far from the failure boundary but could be inside the failure region. Let $D$ be the domain area covered by sample disks, and $\overline{D}$ its complement.
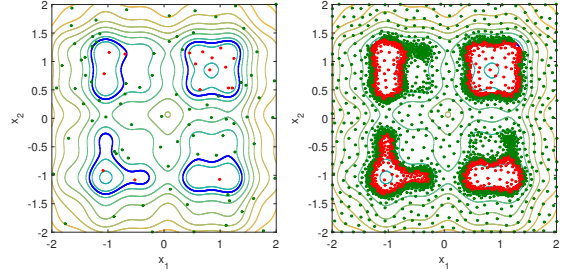


Figure 6: POF-Darts places more points where the function value is near the failure threshold, that is, within $A_t$. Note that flat regions often appear near the actual failure boundary, but we also see extended lobes where the function is nearly constant and near failure.

To reach our conclusions, we explore the following two questions: "After $j$ samples, what is an upper bound on the probability that $A(\overline{D}) > kA_t$?" and "How many samples $j$ do we need so that the area $A(\overline{D}) < kA_t$ with high probability?" The outline of the answers are that disks outside $A_t$ have large radius, hence there can not be very many of them, and each one reduces the uncovered area outside $A_t$ significantly. Samples inside $A_t$ could have small radius, approaching zero, and hence these do not help us. If we have taken many samples then we most likely have filled up a large fraction of the

complement of $A_t$, relative to the size of $A_t$, and hence the uncertainty in our probability-of-failure estimate is probably within some constant times $A_t$. We make three simplifying assumptions for analysis: sampling is uniform by area in the uncovered region, $L$ is known, and the surrogate is not used. Generating samples by lines introduces additional dependencies on where the region lies relative to prior sample disks. Since Lipschitz constant $L$ exists and is known, we can size the disk for sample $x$ correctly with radius $r_x = |f(x)|/L$. [1]

## 3.1. Accuracy on the Fly

We reach our first conclusion indirectly, by considering an estimate of the accuracy of a run in progress. Run the algorithm, and select $t$ and $k$ and observe quantities $j$ and $i$, and also the uncovered area $\overline{D}$. Consider the remaining area $\overline{D}_j = G_j$ of the domain outside all disks after $j$ samples. Suppose that the uncovered area is large compared to $A_t$, that is $G_j > kA_t$; then we show this happens with low probability. For $m \leq j$ let $\overline{F}_m$ be the uncovered area inside $\overline{A}_t$, and $F_m$ be the uncovered area inside $A_t$. Note $\overline{F}_0 = \overline{A}_t = 1 - A_t$, and $F_j$ is decreasing. Note $\overline{D} = G_0 = 1$ and $G_m = (\overline{F}_m + F_m)$ is strictly decreasing. The probability of the next sample landing in $\overline{F}_m$ is $\overline{F}_m/G_m$ and in $F_m$ is $F_m/G_m$. There are many permutations of which of the $i$ samples lie in $\overline{A}_t$ and which $(j - i)$ of them lie in $A_t$. For example, for the first permutation $P$ of all $i$ samples in $\overline{A}_t$ occurring first,

$$p_{P\bar{i}(j-i)} = \frac{\overline{F}_0}{G_0}\frac{\overline{F}_1}{G_1} \cdots \frac{\overline{F}_i}{G_i}\frac{F_{i+1}}{G_{i+1}}\frac{F_{i+2}}{G_{i+2}} \cdots \frac{F_j}{G_j}$$

To find an upper bound on this value, note $\overline{F}_m/G_m \leq 1$ so we may discard the over-lined terms. Moreover $F_m \leq F_0 = A_t$, so we may replace all the non-over-lined $F$ by $A_t$. By assumption $G_m > kA_t$, so we replace the non-discarded denominators by $kA_t$. Hence the probability of this, or any permutation, occurring is less than $k^{-j+i}$. There are $\binom{j}{i}$ permutations, so the probability that $G_j > kA_t$ is bounded by

$$p(G_j > kA_t) < \binom{j}{i}k^{-j+i}. \tag{3}$$

Observe that this is a rapidly decreasing function of $j$ for $j > i$. The right hand side is implicitly dependent on $t$,

through the definition of $i$. In principle, one could compute this value during a run of the algorithm for a concrete problem, answer the question "At this stage of my sampling process, what is the probability that the ambiguous uncovered region is larger than $kA_t$?" and use it as a stopping criteria. This formula is easier to compute than the union-volume POF upper bound, given by the area of the observed uncovered region plus the failure disks. However Equation 3 only gives a smaller upper bound than union-volume in the case that the uncovered region is large compared to the observed failure region. But that is precisely when the POF estimate is inaccurate, so the smaller bound is not useful. The main thing that this formula does provide is the confidence that if the estimate is inaccurate, it is because there is a large area that is near failure. That is, this shows that the algorithm is only inefficient when the failure function is inherently difficult to explore, our first conclusion.

## 3.2. A Priori Sample Budget

We reach our second conclusion, on the expected convergence rate, also indirectly. We consider the question of "Before I've begun sampling, how many samples $j$ should I budget to ensure $p(G_j > kA_t)$ is small?"

We first consider the question of how many large sample disks, with $|f(x)| > t$, can exist. We answer this with an area argument. Note that half-radius disks do not overlap. Since $r_x = |f(x)|/L \geq t/L$, each large sample disk covers a unique area of at least $V(t/(2L))$ where $V(r)$ is the volume of a sphere with radius $r$, and is exponential in dimension. For unit-area domains without boundary, e.g. a periodic unit cube, an upper bound on the number of samples $I$ that can lie outside of $A_t$ is

$$I < \frac{1}{V(t/(2L))}. \tag{4}$$

Let us select $j > I$, and recall $I$ is dependent on $t$. Note $p(G_j > kA_t)$ for all $i \leq I$ is bounded by (3). Hence

$$p(G_j > kA_t) < \sum_{i \leq I} \binom{j}{i}k^{-j+i}. \tag{5}$$

This is a very rapidly decreasing function, which leads to our second conclusion: after sufficient samples are added, the accuracy of the probability of failure estimate increases exponentially. Here "sufficient" depends on the desired accuracy $t$ and $k$, and the difficulty of the domain based on its $A_t$, $L$ and dimension. Comprehending this abstract statement is challenging due to the large number of parameters and caveats. For illustration let us consider a concrete example. Let $t = 0.1$,

---

[1] Occasionally sampling at the farthest Voronoi vertex (e.g. 1 in 25 samples) shows that the algorithm still converges in the absence of a known or even extant Lipschitz constant, because in the limit the sampling is dense. Given a bad $L$ it will still converge, but we cannot guarantee our good convergence rate. This is similar to the convergence proof of DIRECT [40, 41].

$k = 2$, and $L = 1$. Thus the unique coverage radius of far-from-failure disks is $r_x = 0.05$. Let the domain be two-dimensional, so $V(r_x) \approx 0.0079$. Let the domain be a non-periodic unit square; let us dilate it by $t$ in order to fully contain the unique-coverage disks. Then the area of the dilated domain is 1.44 and $I = 183$. We plot this function in Figure 7 for $j$ near $4I$.
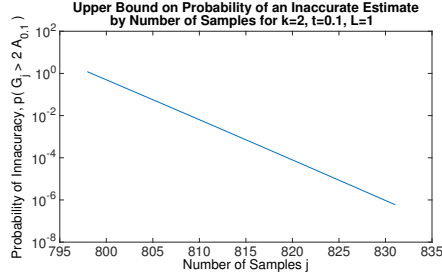


Figure 7: The upper bound on the probability that the POF estimate is inaccurate becomes vanishingly small as enough samples are added in the ambiguous near-failure regions.

Note that in the span of thirty samples the probability of inaccuracy becomes vanishingly small. The reader may be concerned that this concrete example is 2D, has a high POF, and requires hundreds of samples. We reiterate that the analysis is not tight, e.g. we assumed all far-from-failure disks were the smallest possible and made many loose simplifications in the combinatoric ratios. That is, the value of the analysis is the convergence rate rather than the particular constants. The exponential convergence is not directly to the actual failure value, but to a small factor of the size of the near-failure region. The "small-factor" and "near-failure" can be chosen arbitrarily close to the actual failure value and region, which pushes out the threshold number of iterations, but not the convergence rate once that threshold is reached. We believe this rate and approximation compare very favorably with the $j^{-1/2}$ of Monte Carlo sampling, and even with the polynomial convergence of methods such as sparse grids.

## 4. Experimental Results

This section evaluates the performance of POF-Darts compared to EGRA, LHS-GP and PCE, when applied to various low to moderate dimensional (2D to 9D) test problems. The particular implementations of the methods evaluated in this study are those in the DAKOTA software package [42]. The DAKOTA implementation of the POF-Darts method is presently restricted to problems with uniform PDFs for the input uncertainties. Ac-

cordingly, the test problems involve only uniform input uncertainties.

With Polynomial Chaos Expansions, we use a full tensor product quadrature grid of points in the 2-D problems below, with a quadrature order of five for a total of 25 sample points. We use sparse grids for the higher-dimensional problems. These grid points are used to build the expansion using a spectral projection approach which projects the response against each basis function using inner products and employs the polynomial orthogonality properties to extract each coefficient. After the coefficients of the expansion are determined, we sample the expansion with one million samples to obtain the probability of failure estimates.

With LHS-GP, a Gaussian process is constructed over a specified number of LHS sample points. Then, the GP is evaluated at a large number of sample points ($10^6$), and the ratio of the sampled failed points to the total points is used as the estimate of the probability of failure. Note that POF-Darts and EGRA involve adaptive sampling, whereas PCE and the sampling of LHS-GP are non-adaptive, without feedback from the response samples to guide further sampling in an adaptive manner. Non-adaptive sample values can be used to estimate POF values for multiple threshold levels for various response quantities from a physics model such as pressure, temperature, etc. and at multiple points in time and/or space simultaneously (e.g. with one set of samples). This generality, however, makes non-adaptive methods often less accurate than adaptive methods for a single POF quantity of interest. Non-adaptive methods are included in the current study to assess any accuracy-cost penalty that these methods incur relative to the adaptive methods (POF-Darts and EGRA) when one is interested in only one specific POF quantity.

We first apply POF-Darts and the other methods to four analytic problems that have been chosen to highlight various features. We study their adaptive sampling behavior, error performance, and sensitivity to failure threshold. Furthermore, we extend our experimental tests to a 5D circuit design engineering problem, and a 9D steel column engineering problem. For each problem, we ran each algorithm twenty times (POF-Darts, EGRA, and the LHS-GP). Note that PCE uses deterministic sampling. We then calculated the mean of the probability of failure over all twenty estimates, the standard deviation, and the relative average absolute error of the twenty estimates. These statistics are reported in tables. Additionally, we repeated each experiment varying the number of samples to understand the convergence behavior against the number of samples. EGRA has some internal controls which specify the termination criteria

in its current implementation, so we cannot specify the sample size; it is determined by the EGRA algorithm.

### 4.1. 2-d Analytic Functions

To highlight the capabilities of POF-Darts, we first apply it to a few analytic 2D test problems that have different features, making them challenging for failure estimation and also optimization and uncertainty quantification problems in general. These test functions are (1) the smooth Herbie function [43] which is a relatively smooth function, (2) the Herbie function which has a high frequency sine component (smooth noise) that creates a large number of local minima and maxima, (3) the circular cone which has a single singularity at the origin, and its local (and global) Lipschitz is unity everywhere, and (4) the planar cross [38] which expands the cone's single singularity along the main axes. The D-dimensional versions of these functions are given by the following equations, respectively:

$$f_1(\underline{x}) = -\prod_{d=1}^{M} e^{-(x_d-1)^2} + e^{-0.8(x_d+1)^2} \tag{6}$$

$$f_2(\underline{x}) = f_1(\underline{x}) - 0.05 \sin\left(8\left(x_d + 0.1\right)\right) \tag{7}$$

$$f_3(\underline{x}) = \left(\sum_{d=1}^{M} x_d^2\right)^{1/M} \tag{8}$$

$$f_4(\underline{x}) = \left(\prod_{d=1}^{M} \frac{1 + \cos(2\pi x_d)}{2}\right)^{1/M} \tag{9}$$

For illustration, Figure 8 shows these functions in 2D along with their iso-contours at different constant threshold values. The topology of failure regions vary from multiple closed shapes to open regions.

**Comparing Sampling Approaches.** We first look at the sampling phase of the POF-Darts, EGRA, LHS-GP and PCE. In Figure 9, we study the Herbie function when the failure regions are associated with a threshold $T = -0.875$. Figure 9 shows the actual sample points placed with each method; there are no surrogate evaluations shown in this figure. The true probability of failure associated with this threshold is $P = 0.2885$. The two non-adaptive methods: PCE (structured), and LHS (the sampling part of LHS-GP, random) continuously place points all over the domain, without attention to information gained from previous samples about the failure regions. On the adaptive side, EGRA guides points towards the failure regions, however, it converges before enough samples are placed around the boundaries of the failure regions. Our proposed

method, POF-Darts, quickly guides points towards the boundaries of the failure regions to quickly identify its volume and topology. Figure 10 illustrates the power of adaptivity in placing 5000 samples around the failure regions for the smooth Herbie, Herbie, cone, and cross functions, respectively. All 5000 samples are plotted in this figure, and no surrogate evaluations are shown.

**Comparing POF Approaches.**

To quantify the performance of different methods, we run each experiment $N_t$ times and compute the following metrics: $\mu$: the average of the $N_t$ probability of failure estimates, $\sigma$: the standard deviation over the $N_t$ trials, and $e$: the relative average absolute error given by the following formula:

$$e = \frac{\mu_{|e_t|}}{P_{true}} = \frac{\frac{1}{N_t}\sum_{t=1}^{N_t} |P_t - P_{true}|}{P_{true}} \tag{10}$$

where $P_{true}$ is the true value of the probability of failure, $|e_t|$ is the absolute trial error, and $\mu_{|e_t|}$ is the average of trials' absolute errors. Table 2 shows these metrics for the four functions when POF, EGRA, LHS-GP, and PCE are used. Three different threshold levels are employed to result in true POF values of $10^{-4}$, $10^{-3}$, and $10^{-2}$. Figure 11 shows the error variation of the four methods against the sample budget.

**Extension to High-Dimensions.** To illustrate performance in higher dimensions, we study the smooth Herbie function in 2, 3, 4, and 5-dimensions, when the probability of failure is estimated using POF-Darts and LHS-GP.

Figure 12 shows the error performance in these four cases. Comparing the performance of LHS-GP and POF, POF curves clearly continued to converge even at higher dimensions while LHS-GP started to saturate in 4D and 5D. Furthermore, varying the failure threshold in high dimensions pushes the error performance of LHS-GP to saturate quickly at lower failure probabilities which are harder to estimate, while that of POF continues to converge.

### 4.2. 5-d Circuit problem

A numerical model of a proprietary circuit with uniform PDFs for 5 model inputs is assessed next. Details of the circuit model cannot be presented, but the performance of the POF methods on representative failure probability problems can be presented. A threshold value for one of the output responses which gives a circuit probability of failure value around 0.0001 was determined. Table 3 lists failure probability estimates, while Figure 13(a) shows the error variation of the four
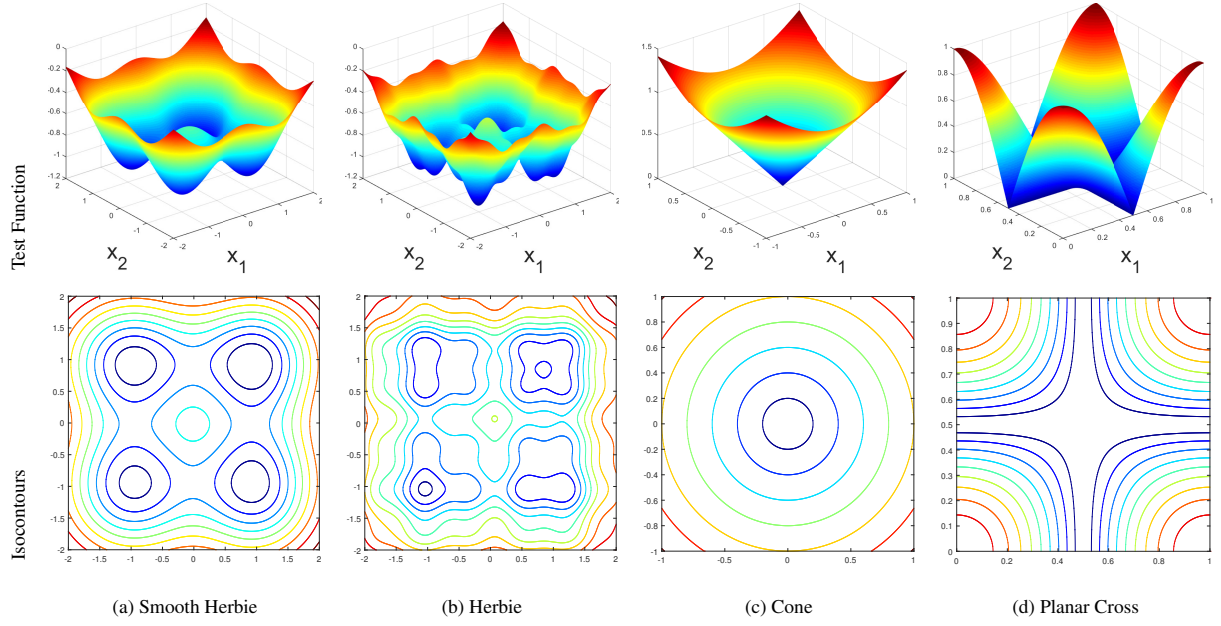
(a) Smooth Herbie  (b) Herbie  (c) Cone  (d) Planar Cross

Figure 8: 2D test functions and their iso-contours for different threshold values.

| f($\underline{x}$) | T | P | POF (N=1000) | | | EGRA (N=55) | | | LHS-GP (N=1000) | | | PCE (N=529) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\mu$ | $\sigma$ | e | $\mu$ | $\sigma$ | e | $\mu$ | $\sigma$ | e | $\mu$ | $\sigma$ | e |
| Sm.Herbie | -1.0937 | $10^{-4}$ | 1.0E-4 | 9.4E-6 | 0.07 | 4.9E-5 | 6.1E-5 | 0.69 | 1.0E-4 | 9.9E-6 | 0.07 | 4.5E-3 | 0.0 | 0.51 |
| | -1.0901 | $10^{-3}$ | 1.0E-3 | 2.7E-5 | 0.02 | 5.4E-4 | 4.1E-4 | 0.46 | 1.0E-3 | 2.1E-5 | 0.02 | 8.8E-2 | 0.0 | 0.43 |
| | -1.0631 | $10^{-2}$ | 1.0E-2 | 9.3E-5 | 0.008 | 8.5E-3 | 2.9E-3 | 0.15 | 9.9E-3 | 1.2E-4 | 0.008 | 4.7E-2 | 0.0 | 0.18 |
| Herbie | -1.1257 | $10^{-4}$ | 9.7E-5 | 9.3E-6 | 0.09 | 5.6E-3 | 6.7E-5 | 55.2 | 9.5E-5 | 8.3E-6 | 0.07 | 5.6E-3 | 0.0 | 55.2 |
| | -1.1152 | $10^{-3}$ | 9.9E-4 | 3.6E-5 | 0.03 | 3.0E-4 | 4.4E-4 | 0.72 | 9.9E-4 | 3.1E-5 | 0.03 | 8.2E-3 | 0.0 | 7.17 |
| | -1.0731 | $10^{-2}$ | 1.0E-2 | 1.1E-4 | 0.01 | 5.9E-3 | 4.3E-3 | 0.47 | 1.0E-2 | 7.6E-5 | 0.006 | 8.2E-3 | 0.0 | 0.18 |
| Cone | 0.0113 | $10^{-4}$ | 1.0E-4 | 8.7E-6 | 0.007 | 8.3E-5 | 4.9E-5 | 0.41 | 5.9E-5 | 1.9E-4 | 1.34 | 0.0 | 0.0 | 1 |
| | 0.0357 | $10^{-3}$ | 9.9E-4 | 3.1E-5 | 0.002 | 6.6E-4 | 2.6E-4 | 0.34 | 9.8E-4 | 4.9E-4 | 0.35 | 0.0 | 0.0 | 1 |
| | 0.1128 | $10^{-2}$ | 9.8E-3 | 9.2E-5 | 0.008 | 9.7E-3 | 1.3E-3 | 0.07 | 9.5E-3 | 1.4E-3 | 0.12 | 1.0E-2 | 0.0 | 1.4E-2 |
| Cross | 1.87E-5 | $10^{-4}$ | 1.2E-4 | 2.2E-4 | 1.26 | 9.0E-3 | 2.5E-4 | 0.07 | 6.0E-5 | 2.7E-4 | 1.50 | 0.0 | 0.0 | 1 |
| | 0.0002 | $10^{-3}$ | 9.4E-4 | 3.4E-4 | 0.28 | 7.1E-3 | 9.6E-3 | 0.34 | 1.7E-3 | 1.2E-3 | 1.14 | 0.0 | 0.0 | 1 |
| | 0.0031 | $10^{-2}$ | 1.2E-2 | 1.4E-3 | 0.17 | 1.2E-2 | 4.7E-4 | 0.08 | 9.9E-3 | 2.4E-3 | 0.19 | 9.1E-4 | 0.0 | 0.9 |

Table 2: Performance metrics: $\mu$, the mean of the probability of failure over 20 trials, $\sigma$, the standard deviation, and e, the relative average absolute error, for different 2d analytic functions using POF, EGRA, LHS-GP, and PCE. Note that an average of 0 means no failure regions were found indicating that $e = 1$. Also, the standard deviation of PCE experiments is set to 0.0 as it involves deterministic sampling with no trial variation.

methods against the sample budget. The *reference value* mean estimate of the exact failure probability based on four million samples is $1.6475E-4 \pm 2.81E-5$ and the 95% confidence interval is [1.370E-4, 1.930E-4].[2]

EGRA converged with 31 function evaluations and the PCE sparse grid for this problem required 71 evaluations. Both EGRA and PCE gave estimates of the

---

[2]Note: we used the following to obtain these confidence intervals: If the random-variable uncertainty space is randomly sampled via Monte Carlo sampling, then the number of model response values (sample values) that exceed the threshold T, divided by the total number of samples, provides an estimate $P*$ of the true failure probability $P$ of the system. If enough samples are taken, then the estimate

$P*$ can be said with some percent likelihood or "confidence" to lie within a corresponding confidence interval of the true result $P$. When the number N of total samples meets the condition [44]

$$N * P \geq 5 \qquad (11)$$

then the following formula for the 95% confidence intervals (CI) applies:

$$|P - P^*| \leq 1.96[P(1 - P)/N]^{1/2}. \qquad (12)$$

(a) EGRA.

(b) PCE with 25, 225, and 625 samples.

(c) LHS with 100, 500, 1000, and 5000 samples.

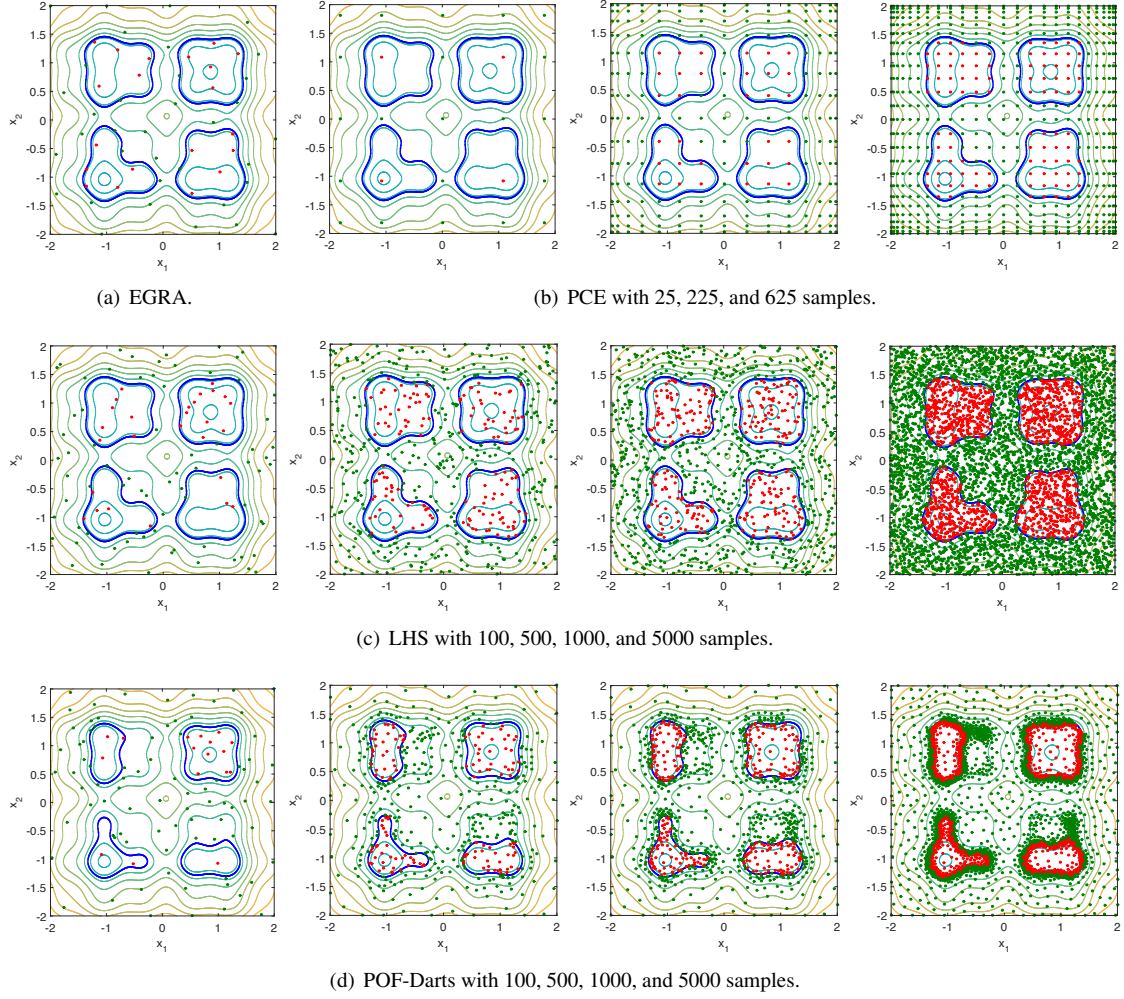(d) POF-Darts with 100, 500, 1000, and 5000 samples.

Figure 9: Comparing the sampling phase of different methods applied to the Herbie function. Plots show the samples when using EGRA, PCE (with quadrature order of 5, 15, and 25), LHS (with 100, 500, 1000, 5000 samples), and POF-Darts (with 100, 500, 1000, 5000 samples). In this illustrative example, the target failure threshold is $T = -0.875$ which results in a probability of failure $P = 0.2885$. Although EGRA is adaptive, it stops before the failure boundaries are sampled enough. On the other hand, PCE is structured and LHS is random; however they are both non-adaptive, hence the waste a lot of samples away from the failure regions. Our method, POF-Darts, adaptively places new samples around the boundaries of the failure regions, quickly capturing its volume and topology.
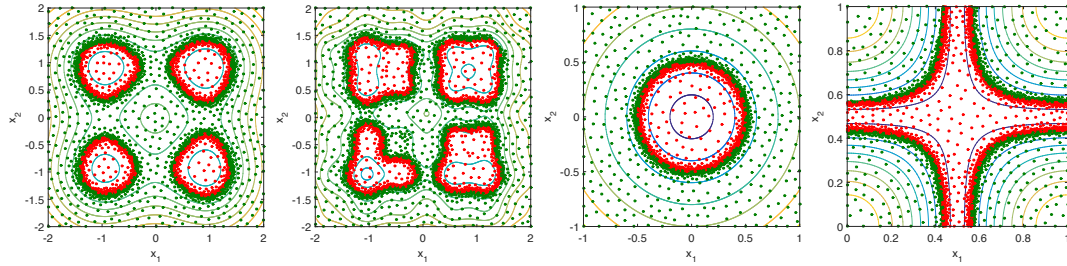


Figure 10: The sample distribution of POF-Darts when 5000 samples are used to adaptively sample smooth Herbie ($T = -0.9$), Herbie ($T = -0.875$), circular cone ($T = 0.5$), and planar cross ($T = 0.2$), listed left to right, respectively.
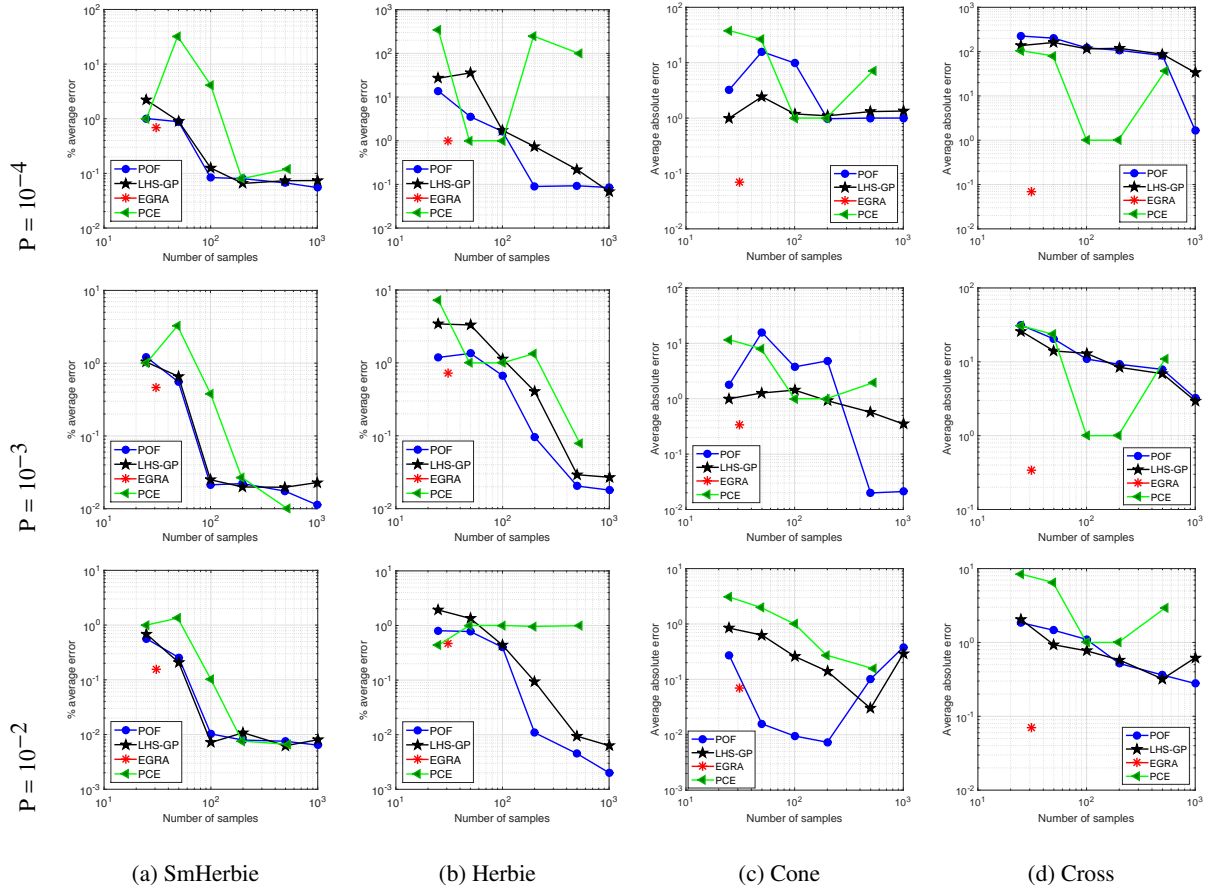
13

Figure 11: Error variation against the sample budget for different 2d analytic functions using POF, EGRA, LHS-GP, and PCE.

| T | P | POF (N=100) | | | EGRA (N=31) | | | LHS-GP (N=100) | | | PCE (N=71) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mu$ | $\sigma$ | $e$ | $\mu$ | $\sigma$ | $e$ | $\mu$ | $\sigma$ | $e$ | $\mu$ | $\sigma$ | $e$ |
| 0.84 | $1.6475E-4$ | 1.43E-5 | 1.24E-5 | 0.13 | 1.57E-4 | 1.09E-5 | 0.07 | 1.056E-5 | 3.58E-5 | 0.36 | 1.612E-4 | 1.147E-5 | 0.05 |

Table 3: Failure probability estimates for 5D Circuit problem using POF, EGRA, LHS-GP, and PCE.

mean POF that are reasonably accurate. LHS-GP did not perform well on this problem, even at large number of samples. The relative average error was quite high for the GP surrogates based on LHS points: even with 1000 samples, it was 40%. POF-Darts performs well on this problem: at 100 samples, it has 13% average error and the average error decreases to 8% at 200 samples. The standard deviations of the POF estimates for EGRA, PCE, and POF-Darts are similar.

### 4.3. 9-d Steel column problem

This problem involves determining the probability that the stress on a steel column will not meet a specified margin of safety relative to its yield stress $F_s$. The problem is based on the reliability problem in [45]

and [46]. Nine uncorrelated random variables are used in the problem as defined in Table 4.

| Variable | | Lower Bound | Upper Bound |
|---|---|---|---|
| $F_s$: | yield stress, MPa | 260 | 575 |
| $P_1$: | dead weight load, kN | 250 | 650 |
| $P_2$: | variable load, kN | 150 | 870 |
| $P_3$: | variable load, kN | 150 | 870 |
| $B$: | flange breadth, mm | 185 | 215 |
| $T$: | flange thickness, mm | 11.5 | 27.5 |
| $H$: | profile height, mm | 75 | 125 |
| $D_0$: | initial deflection, mm | -20 | 80 |
| $E$: | elastic modulus, MPa | 1 | 41 |

Table 4: Uniform PDF Inputs for Steel Column Problem

The failure boundary function defining the margin $g$
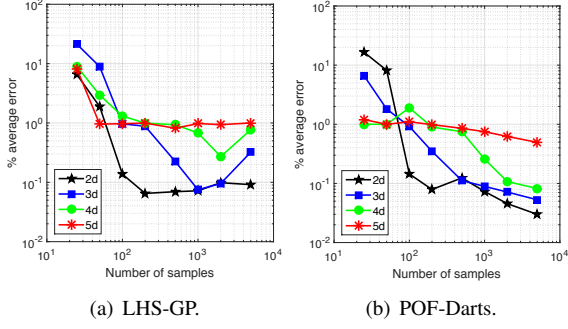
14

(a) LHS-GP.  (b) POF-Darts.

Figure 12: Comparing error performance of POF and LHS when GP global surrogates are employed for 2, 3, 4, and 5-dimensional smooth Herbie functions targeting $P = 10^{-4}$.
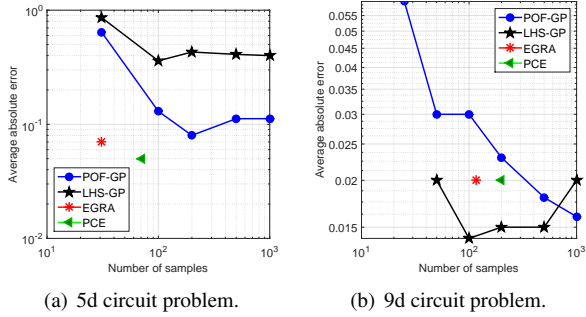


(a) 5d circuit problem.  (b) 9d circuit problem.

Figure 13: Error variation against the sample budget for the 5d circuit problem and 9d steel problem using POF, EGRA, LHS-GP, and PCE.

is calculated from the following equations:

$$g = F_s - P\left(\frac{1}{2BT} + \frac{F_0}{BTH}\frac{E_b}{E_b - P}\right) \qquad (13)$$

where

$$P = P_1 + P_2 + P_3 \qquad (14)$$

$$E_b = \frac{\pi^2 EBTH^2}{2L^2} \qquad (15)$$

and the column length $L$ is 7500mm.

A positive safety margin exists when the value of $g$ is positive. We specify a desired safety margin of g=260 MPa, and determine the probability that this margin is not met: Prob($g \leq 260$ MPa). The failure probability for this problem with the specified inputs is approximately 0.001. Table 5 lists failure probability estimates from the various methods, while Figure 13(b) shows the error variation of the four methods against the sample budget.

The *reference value* mean estimate of the exact failure probability based on three million samples is 1.035E-3 ± 2.4843E-5 and the 95% confidence interval on this estimate is [1.0102E-3, 1.0598E-3].

In this problem, all methods performed comparably. EGRA and PCE required 116 and 199 samples, respectively, and both had 2% average error. The GP surrogates performed very well on this problem, achieving excellent accuracy and precision with as little as 50 samples. Reasonable performance with such few samples indicates that the function is probably only mildly nonlinear over the UQ space. For a very small sample size of 25 samples, POF-Darts was not as accurate, but became more accurate based on more samples. GPs were not tried with 25 samples because the current implementation in DAKOTA does not allow GPs to be used with such few samples for a 9D problem.

## 5. Concluding Remarks

We introduced a new technique to estimate the Probability Of Failure $P$ based on a computational geometry approach: POF-Darts. POF-Darts involves two separate steps: sampling and surrogate construction. For the sampling phase, POF-Darts uses $k$-d Darts to sample high-dimensional domains. Using spheres with variable radii, we adaptively cover the domain, using local Lipschitz information to decide on the radius of an added disk and update the radii of previously placed ones. Once the sampling budget is completely used, we construct a global surrogate model and sample it exhaustively with Monte Carlo to evaluate the function everywhere, quantify the probability of failure, and identify the topology of the failure regions.

We evaluated POF-Darts for a variety of analytic and realistic problems ranging in dimensionality from 2-d to 9-d. POF-Darts performs favorably compared to similar methods such as LHS-GP, EGRA, and PCE. The adaptive placement of points is a very useful feature of the algorithm: POF-Darts can efficiently cut away large swaths of the space which are far from the failure region with large hyperspheres in regions far away from failure. In this way, POF-Darts can concentrate its samples around the failure boundaries as shown in Figures 9 and 10. Monte Carlo methods would require many more samples to get such coverage around a failure boundary. Because of this, we expect our method to perform well on problems even higher than 9-dimensions. Our method has the advantage of being able to detect the failure topology (number of simply connected failure regions) and discontinuities.

We present a convergence analysis of the POF-Darts method in Section 3. The convergence is primarily governed by the number of sample points it takes to fill the non-failure regions and not the accuracy of the surrogate. As stated in Section 3, the POF-Darts algorithm

15

| T | P | POF (N=100) | | | EGRA (N=116) | | | LHS-GP (N=100) | | | PCE (N=199) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mu$ | $\sigma$ | $e$ | $\mu$ | $\sigma$ | $e$ | $\mu$ | $\sigma$ | $e$ | $\mu$ | $\sigma$ | $e$ |
| 260.0 | 1.035E-3 | 1.030E-5 | 3.26E-5 | 0.03 | 1.023E-5 | 2.126E-5 | 0.02 | 1.029E-3 | 1.865E-5 | 0.01 | 1.027E-3 | 1.793E-5 | 0.02 |

Table 5: Failure probability estimates for 9-D Steel Column Problem using POF, EGRA, GP, and PCE.

converges very quickly (exponentially) once a sufficient number of samples has filled up most of the domain where the function value is far from failure. Our convergence analysis focuses on bounds on the probability that the remaining volume in parameter space is not covered by disks. While we have established the convergence of POF-Darts in the limit, we have not yet addressed the accuracy of the surrogate and its effect on the resulting Probability of Failure estimate. This is the subject of future work. POF-Darts decouples the sampling and surrogate construction, which allows for the possibility of investigating different surrogates and developing more elegant designs. When evaluating smooth functions, this decoupling could lead to less efficiency than coupled sampling/surrogate methods like EGRA, where the probability of failure estimation based on the GP is used to augment the sample set. However, decoupling is advantageous for the case of discontinuous functions. Even without a tight coupling, the empirical evidence that we present demonstrates competitive results for failure estimates from POF-Darts. We plan to investigate methods to incorporate coupling between the sample points and the surrogate type/accuracy. This will involve changing the POF-Darts algorithm to allow for the placement of additional points based on the Probability of Failure accuracy from surrogate information. We are also developing a version of POF-Darts which uses local surrogates (e.g. constructing a separate surrogate within each disk and not one global GP surrogate.) The POF-Darts approach is unique due to its basis in computational geometry, and can be extended to address various use cases.

# References

[1] T. S. Meek, The Home Library of Entertainment Instruction and Amusement, N.D. Thompson Publishing Company, 1902, clipart courtesy FCIT, http://etc.usf.edu/clipart/, under Free Classroom License.

[2] P. Chaplin, Darts in England, 1900–1939. A Social History, Studies in Popular Culture, Manchester University Press, 2009.

[3] F. Silva, T. W. Lambe, W. A. Marr, Probability and risk of slope failure, Journal of geotechnical and geoenvironmental engineering 134 (12) (2008) 1691–1699.

[4] J. D. Kalbfleisch, R. L. Prentice, The statistical analysis of failure time data, Vol. 360, John Wiley & Sons, 2011.

[5] K. Lamár, J. Neszveda, Average probability of failure of aperiodically operated devices, Acta Polytechnica Hungarica 10 (8) (2013) 153–167.

[6] K. W. Miller, L. J. Morell, R. E. Noonan, S. K. Park, D. M. Nicol, B. W. Murrill, J. M. Voas, Estimating the probability of failure when testing reveals no failures, Software Engineering, IEEE Transactions on 18 (1) (1992) 33–43.

[7] A. M. Freudenthal, Safety and the probability of structural failure, American Society of Civil Engineers Transactions.

[8] W. Duckett, Risk analysis and the acceptable probability of failure, Structural Engineer 83 (15).

[9] S. Lacaze, L. Brevault, S. Missoum, M. Balesdent, Probability of failure sensitivity with respect to decision variables, Structural and Multidisciplinary Optimization 52 (2) (2015) 375–381.

[10] A. M. Rushdi, On reliability evaluation by network decomposition, Reliability, IEEE Trans. on 33 (5) (1984) 379–384.

[11] A. M. Rushdi, Uncertainty analysis of fault-tree outputs, Reliability, IEEE Trans. on 34 (5) (1985) 458–462.

[12] A. M. Rushdi, K. F. Kafrawy, Uncertainty propagation in fault-tree analyses using an exact method of moments, Microelectronics Reliability 28 (6) (1988) 945–965.

[13] J. H. Boyd, G. De Nicolo, The theory of bank risk taking and competition revisited, The Journal of finance 60 (3) (2005) 1329–1343.

[14] A. M. Santomero, J. D. Vinso, Estimating the probability of failure for commercial banks and the banking system, Journal of Banking & Finance 1 (2) (1977) 185–205.

[15] M. Koehn, A. M. Santomero, Regulation of bank capital and portfolio risk, Journal of finance (1980) 1235–1244.

[16] J. C. Helton, Conceptual and computational basis for the quantification of margins and uncertainty, Tech. Rep. SAND2009-3055, Sandia National Labs, Albuquerque, NM (Jul. 2009).

[17] J. C. Helton, F. J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, Reliability Engineering and System Safety 81 (1) (2003) 23–69.

[18] A. M. Law, W. K. Kelton, Simulation Modeling and Analysis, McGraw-Hill, 1999.

[19] M. G. Morgan, M. Henrion, Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press, 1993.

[20] R. L. Iman, M. J. Shortencarier, A Fortran 77 program and user's guide for the generation of latin hypercube samples for use with computer models, Tech. Rep. NUREG/CR-3624, SAND83-2365, Sandia National Laboratories, Albuquerque, NM (1984).

[21] M. Stein, Large sample properties of simulations using Latin Hypercube Sampling, Technometrics 29 (2) (1987) 143–151.

[22] G. I. Schueller, H. J. Pradlwarter, P. S. Koutsourelakis, A critical appraisal of reliability estimation procedures for high dimensions, Probabilistic Engineering Mechanics 19 (2004) 463–474.

[23] A. A. Giunta, J. M. McFarland, L. P. Swiler, M. S. Eldred, The promise and peril of uncertainty quantification using respone surface approximations, Structure and Infrastructure Engineering 2 (3-4) (2006) 175–189.

[24] C. Storlie, L. Swiler, J. Helton, C. Sallaberry, Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models, Reliability Eng. and System Safety 94 (11) (2009) 1735–1763.

[25] T. Santner, B. Williams, W. Notz, The Design and Analysis of Computer Experiments, New York, NY: Springer, 2003.

[26] R. G. Ghanem, P. D. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer-Verlag, New York, 1991.

[27] D. Xiu, G. M. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.

[28] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Element Approach, Princeton University Press, 2010.

[29] M. S. Eldred, C. G. Webster, P. Constantine, Evaluation of non-intrusive approaches for Wiener-Askey generalized polynomial chaos, in: Proceedings of the 10th AIAA Non-Deterministic Approaches Conference, no. AIAA-2008-1892, Schaumburg, IL, 2008.

[30] R. Rackwitz, Reliability analysis. a review and some perspectives, Structural Safety 23 (2000) 365–395.

[31] A. Haldar, S. Mahadevan, Probability, Reliability, and Statistical Methods in Engineering Design, Wiley, New York, 2000.

[32] B. J. Bichon, M. S. Eldred, L. P. Swiler, S. Mahadevan, J. M. McFarland, Efficient global reliability analysis for nonlinear implicit performance functions, AIAA Journal 46 (10) (2008) 2459–2468.

[33] M. S. Eldred, H. Agarwal, V. M. Perez, J. Wojtkiewicz, S. F., J. E. Renaud, Investigation of reliability method formulations in DAKOTA/UQ, Structure & Infrastructure Engineering: Maintenance, Management, Life-Cycle Design & Performance 3 (3) (2007) 199–213.

[34] S.-K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics 16 (4) (2001) 263–277.

[35] J.-M. Bourinet, F. Deheeger, M. Lemaire, Assessing small failure probabilities by combined subset simulation and support vector machines, Structural Safety 33 (6) (2011) 343–353.

[36] M. S. Ebeida, A. Patney, S. A. Mitchell, A. Davidson, P. M. Knupp, J. D. Owens, Efficient maximal Poisson-disk sampling, ACM Transactions on Graphics 30 (4) (2011) 49:1–49:12. doi:10.1145/1964921.1964944.

[37] S. A. Mitchell, A. Rand, M. S. Ebeida, C. Bajaj, Variable radii Poisson-disk sampling, extended version, in: Proceedings of the 24th Canadian Conf. on Comp. Geometry, 2012, pp. 1–9.

[38] M. S. Ebeida, A. Patney, S. A. Mitchell, K. R. Dalbey, A. A. Davidson, J. D. Owens, K-d darts: Sampling by k-dimensional flat searches, ACM Trans. Graph. 33 (1) (2014) 3:1–3:16. doi:10.1145/2522528.
URL http://doi.acm.org/10.1145/2522528

[39] M. S. Ebeida, S. A. Mitchell, M. A. Awad, C. Park, L. P. Swiler, D. Manocha, L. Wei, Spoke darts for efficient high dimensional blue noise sampling.
URL http://arxiv.org/abs/1408.1118

[40] B. O. Shubert, A sequential method seeking the global maximum of a function, SIAM Journal on Numerical Analysis 9 (3) (1972) 379–388.

[41] D. R. Jones, C. D. Perttunen, B. E. Stuckman, Lipschitzian optimization without the lipschitz constant, Journal of Optimization Theory and Applications 79 (1) (1993) 157–181.

[42] B. Adams, L. Bauman, W. Bohnhoff, K. Dalbey, M. Ebeida, J. Eddy, P. Hough, K. Hu, J. Jakeman, L. Swiler, , J. Stephens, D. Vigil, T. Wildey, Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantication, and sensitivity analysis. Version 6.2 user's manual, Sand Report SAND2014-4633, Sandia National Laboratories, https://dakota.sandia.gov/content/sand-reports (May 2014, updated 2015).

[43] H. Lee, R. Gramacy, C. Linkletter, G. Gray, Optimization Sub-

ject to Hidden Constraints via Statistical Emulation, Pacific Journal of Optimization 7 (2011) 467–478.

[44] J. Devore, Probability and Statistics for Engineering and the Sciences, Belmont, CA: Brooks/Cole Publishing, 1982.

[45] N. Kuschel, R. Rackwitz, Two basic problems in reliability-based structural optimization, Math. Method Oper. Res. 46 (1997) 309–333.

[46] B. J. Bichon, Efficient surrogate modeling for reliability analysis and design, Ph.D. Thesis, Vanderbilt University, Nashville, TN.