Acta Polytechnica CTU Proceedings **36**:269–279, 2022
https://doi.org/10.14311/APP.2022.36.0269

Published by the Czech Technical University in Prague

ADAPTIVE SEQUENTIAL SAMPLING FOR RELIABILITY ESTIMATION OF BINARY FUNCTIONS

Miroslav Vořechovský

Brno University of Technology, Institute of Structural Mechanics, Faculty of Civil Engineering, Veveří 331/95, 602 00 Brno, Czech Republic

correspondence: vorechovsky.m@vut.cz

Abstract. A novel method for estimation of rare event probability is proposed, which works also for computational models returning categorical information only: success or failure. It combines the robustness of *simulation* methods (counting failure events) with the strength of *approximation* methods which refine the boundary between the failure and safe sets. Two basic tasks are identified: (i) *extension* of the experimental design (ED) and (ii) *estimation* of probabilities. The new *extension* algorithm adds points for limit state evaluation to the ED by balancing the global *exploration* and local *exploitation*, and the *estimation* uses the pointwise information to build a simple surrogate and perform a novel optimized importance sampling. No connection is presumed between the limit function value at point and its proximity to the failure surface. A new global sensitivity measure of the failure probability to individual variables is proposed and obtained as a by-product of the proposed methods.

Keywords: Exploitation, exploration, reliability.

1. Introduction

This article promotes a recent work [1] by the author. We revisit the problem of failure probability estimation of an engineering product or a process, which is represented by a computationally expensive mathematical model $g(\boldsymbol{x})$. The model features uncertain or random variables $\boldsymbol{x}$ and we assume the joint probability density $f_{\boldsymbol{X}}(\boldsymbol{x})$ is known and can be uniquely transformed to the space of independent Gaussian random variables of dimension N_{var}. The model $g(\boldsymbol{x})$ may return categorical output only, that is, failure or success. The problem is to accurately estimate the probability of failure with the smallest possible number of $g(\boldsymbol{x})$ calls.

The spectrum of existing methods for failure probability estimation is very rich. Their categorization can be done depending on how they treat three groups of input information: (i) the geometry and topology of the input space of input variables indexed by a random vector $\boldsymbol{X}$, (ii) probability density function of $\boldsymbol{X}$, be it the true density or some modified "sampling" density, and (iii) the limit state function $g(\boldsymbol{x})$. From the point of view of the way that the methods for the estimation of failure probability handle the outputs of the limit state function, two basic groups can be identified. *Simulation methods* such as Monte Carlo integration, Importance Sampling [2–4] or Asymptotic Sampling [5] use the binary information only: they count the failure events out of all N_{sim} function calls and associate these events with some weights to be used in estimates performed via arithmetic or weighted averages. These methods are robust with respect to complicated functions $g(\boldsymbol{x})$, which may be noisy, can have disjoint failure sets, the landscapes may non-smooth or there can be even regions for which $g(\boldsymbol{x})$ return no answer at all. However, excessively high number function calls needed in simulation methods when dealing with practical problems, in which the true failure probability is very low. A disadvantage of the simulation methods is that they use no or only little information from previous function calls.

The other group of methods can be termed *approximation methods*. They make various assumptions about the shape of $g(\boldsymbol{x})$, and in most cases, the methods work for smooth functions only. They can be seen as methods making use of the gradient optimization in search for the most probable failure regions (FORM/SORM [6–14], Subset Simulation [15]) or they adaptively build approximations (surrogate models) of $g(\boldsymbol{x})$ [16]. The surrogate model can be a smooth approximation based of polynomials, using Kriging (Gaussian process regression) [17, 18], Polynomial Chaos Expansions [19, 20], radial basis functions [21, 22], or classifiers such as artificial neural networks [23–25] or Support vector regression surrogates [26–30]. The evaluation of the surrogate must be sufficiently fast to be useful for the robust sampling strategies making the probability estimations. There is also a class of methods which construct approximations of the true distribution function of numerical values of $g(\boldsymbol{x})$ based on the empirical histograms to estimate the cumulative probability distribution at zero value, that is, direct estimation of failure probability. It is clear that the shape of the cumulative distribution function is strongly dependent of the particular formulation of $g(\boldsymbol{x})$. In other words, various reformulations and reparametrizations of $g(\boldsymbol{x})$ which do *not* alter the failure surface lead to quite different distribution functions of the outcome which complicates this kind of approximation.

In practical applications of reliability estimation methods, we face a variety of challenges and there are difficulties which can not be overcome by the existing algorithms. Suppose, for example, that the state of an existing structure is modelled by an algorithm (e.g., a nonlinear FEM simulation). This model can provide highly nonlinear output in the space of input variables. The model can capture the existence of multiple failure modes whose probabilities are not many orders of magnitudes different (a situation with multiple design points in FORM/SORM), and there can be combinations of inputs for which the algorithm provides *no* meaningful response. The existence of these scenarios motivates the development of more robust algorithms such as the one presented in this paper. Another challenge is connected with the spurious dependence of answers obtained by many existing algorithms on the particular formulation of the limit state function. The same system condition can be expressed via different formulations of the function of input variables leading to dramatically different $g(\boldsymbol{x})$ function "landscapes" in the input space. The landscape is, however, important for gradient-based algorithms, be it the deterministic ones (design point search) or stochastic gradient descent methods (Subset Simulation). Moreover, the some models can also be formulated in a way that the $g(\boldsymbol{x})$ function *value* is not reasonably reflecting the condition of the system (differentiating between categories "safe" or "very safe"). Our argument is that in these cases, it can be better to consider the categorical information only: safe operation or failure. The failure category can actually be represented by a finite number of potentially different failure codes, and there can also be an additional category "no response". The algorithm for reliability estimation must be able to work with such a crude information and keep going upon receiving any of these state codes by (i) processing the information and (ii) proposing a new point for evaluation by the model.

In this paper, we briefly present a technique consisting of two algorithms. One for the adaptive sequential extension of the experimental design (ED), that is, the table of N_{sim} points each with N_{var} coordinates, and the other for importance sampling analysis performed in the standard Gaussian space. The method is able to work with binary functions, however, if the $g(\boldsymbol{x})$ function returns values usable for building smooth approximation, the estimation task can be performed with an advanced surrogate function with the chance of obtaining more accurate results for small EDs.

2. Gaussian space with independent marginals

The methods presented in this paper use the fact that the rotationally symmetrical Gaussian space of dimension N_{var} can be indexed either by N_{var} orthogonal coordinates using the Cartesian coordinate system, or by (i) $N_{\text{var}}-1$ independent and identically distributed directions and (ii) one radial distance from the origin

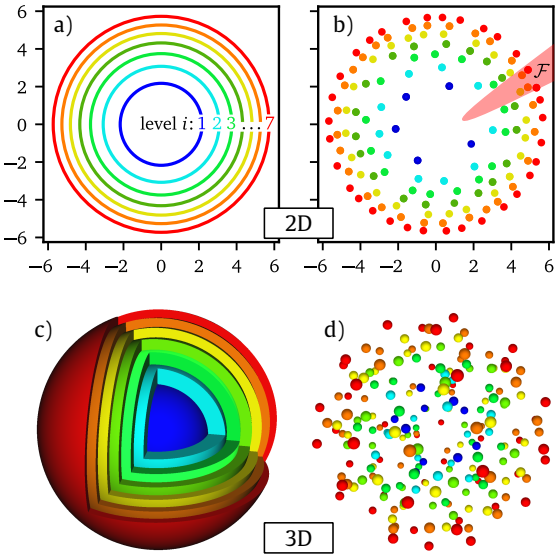

Figure 1. The first seven levels of exploration sets (layers) in $N_{\text{var}}=2$ and $N_{\text{var}}=3$ dimensional standard Gaussian space. Panels a,c: nested N_{var}-balls. Panels b,d: realizations of random exploration sets.

which is independent from the directions. The radial distance from is χ-distributed with N_{var} degrees of freedom. This allows for direct computations of probabilities associated with N_{var}-balls of a given radius r, see [1] for mathematical details and the available numerical libraries which facilitate the work in the (hyper)spherical coordinate system.

We now define a sampling density for a ring (annulus), which is a rotationally symmetrical region between two N_{var}-balls defined by two different radii, $r<R$. The probability density for a point $\boldsymbol{x}$ is a scaled standard Gaussian density

$$h_{\text{ann}}(\boldsymbol{x};r,R)=\frac{\Pi_{v=1}^{N_{\text{var}}}\varphi(\boldsymbol{x}_v)}{p_{\text{ann}}}=\frac{\Pi_{v=1}^{N_{\text{var}}}\varphi(\boldsymbol{x}_v)}{p_R-p_r} \quad (1)$$

and zero otherwise, where $\varphi(\boldsymbol{x}_v)$ is the univariate standard Gaussian density. In this equation, the scaling denominator $p_{\text{ann}}=p_R-p_r$ is formed by the difference between two probabilities corresponding to probability content in the two N_{var}-balls. This density can be rewritten in terms if the radial distance $\rho\equiv\|x\|$.

3. Extension of the experimental design

3.1. Exploration set

Assume now that the input space is divided into rotationally symmetric layers with predefined probability content. Each such layer is obtained as the difference between two N_{var}-dimensional co-centric balls. The first ball contains, say, 90% of the total probability, the next ball contains 99% so that the ring (an annulus) in between adds 1% to the smaller ball. The next onion-like shell encloses 99.9% probability etc. In this way, the space is divided into layers such that the probability contents are covered in a controlled

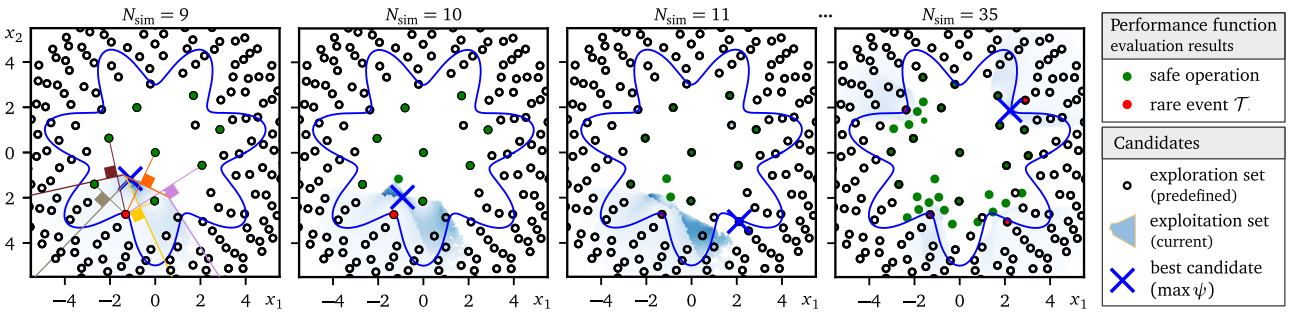

FIGURE 2. Two-dimensional illustration of the combined set candidates and the selection process (standard Gaussian space) for a binary performance function dividing the space into a "flower-shaped" safe domain and the surrounding failure domain. At $N_{\text{sim}} = 9$ model evaluations, the rare event is hit for the first time. At $N_{\text{sim}} = 10$ an *exploitation* candidate is selected, while at $N_{\text{sim}} = 11$ the *exploration* candidate wins using the highest value of the ϕ criterion expressed by the blue color saturation. A complete evolution of the extension of the ED using the ϕ-based selection process for another run of the same problem is available as a video.

fashion and the radii of the N_{var}-balls forming the boundaries are easily obtained using the χ distributed variable discussed above. Assume now that the surface of each such ball is uniformly covered by a predefined of points. The complete set of point corresponding to all considered layers is called the *exploration set*. Figure 1 illustrates the nested rings and their coverage by points for $N_{\text{var}} = 2$ dimensions and 3D.

3.2. Exploitation set

Suppose that the experimental design contains at least one failure point and at least one safe point. From that moment it makes sense to expect that the true boundary between the safe and failure sets passes somewhere between the two points with a different event type. We propose to generate a cloud of "candidate dots" centered at each failure point and with a Gaussian density of standard deviation equal to $\sqrt{N_{\text{var}} - 1}$. From all these candidate dots, we only consider those whose two nearest neighbors have two different classifications from $g(\boldsymbol{x})$. The set of all dots fulfilling this requirement collectively forms the *exploitation set* and they are considered in competition for the best candidate alongside the exploration set.

3.3. ψ criterion for candidate selection

Suppose the combined exploration-exploitation set of candidates is available. We propose to select the most informative candidate using the criterion which balances the local exploitation with global exploration. Each candidate is associated with the criterion value which roughly estimates the probability being resolved by evaluating the $g(\boldsymbol{x})$ function in that candidate

$$\psi_{\text{c}} = \underbrace{\sqrt{f_{\text{c}} f_{\text{s}}}}_{\text{ave probability}} \quad \underbrace{(l_{\text{c,s}})^{N_{\text{var}}}}_{\propto \text{vol.}} \tag{2}$$

The square root term represents the geometric mean of probability density f_{c} in the candidate "c", and the density f_{s} in its nearest neighbor "s". This average is multiplied by a term proportional to the volume occupied by the region between the candidate and its nearest neighbor. This volume is estimated as their Euclidean distance $l_{\text{c,s}}$ raised to the domain dimension. The candidate delivering the highest value of the ψ criterion is selected to extend the current ED. In this way the algorithm automatically balances between the refinement of the failure surface approximation and the expansion of the ED in the input space towards infinity. Simply, the candidate yielding the largest (approximate) probability amount is selected. The ψ criterion can be monitored during the extension and its value can support the stopping criterion. Figure 2 illustrates the combined exploration and exploitation sets and shows the ranking of candidates using the ψ criterion via the darkness of the blue infill.

4. Estimation of probabilities

At any stage of the ED extension process, the desired probabilities can be estimated based on point-wise information, i.e., the current ED with known $g(\boldsymbol{x})$ outcomes. We propose to substitute a true computationally expensive model (a simulator) by a computationally cheaper model constructed in a solely non-intrusive way with respect to the original simulator, i.e., in a purely data-driven manner. When the performance function $g(\boldsymbol{x})$ returns continuously distributed, trustworthy, and well-behaved output, traditional smooth surrogate models such as PCE, Kriging, radial basis functions, etc., can be employed for fast sampling analysis. Such a surrogate model has the potential to improve the accuracy of the importance sampling estimation presented in this section because the failure surface may be approximated more accurately. It is guaranteed that the above-described extension of the ED was performed in such a way that the surrogate model was well supported, especially when close to the failure surface. In this paper, however, we focus on cases in which the original simulator is trusted to provide categorical information only, i.e., one of a finite set of classifications. Similarly, the surrogate will simply be a finite-state classifier. We propose to use simply a classifier based on the nearest neighbor: at any location in the input space, the event

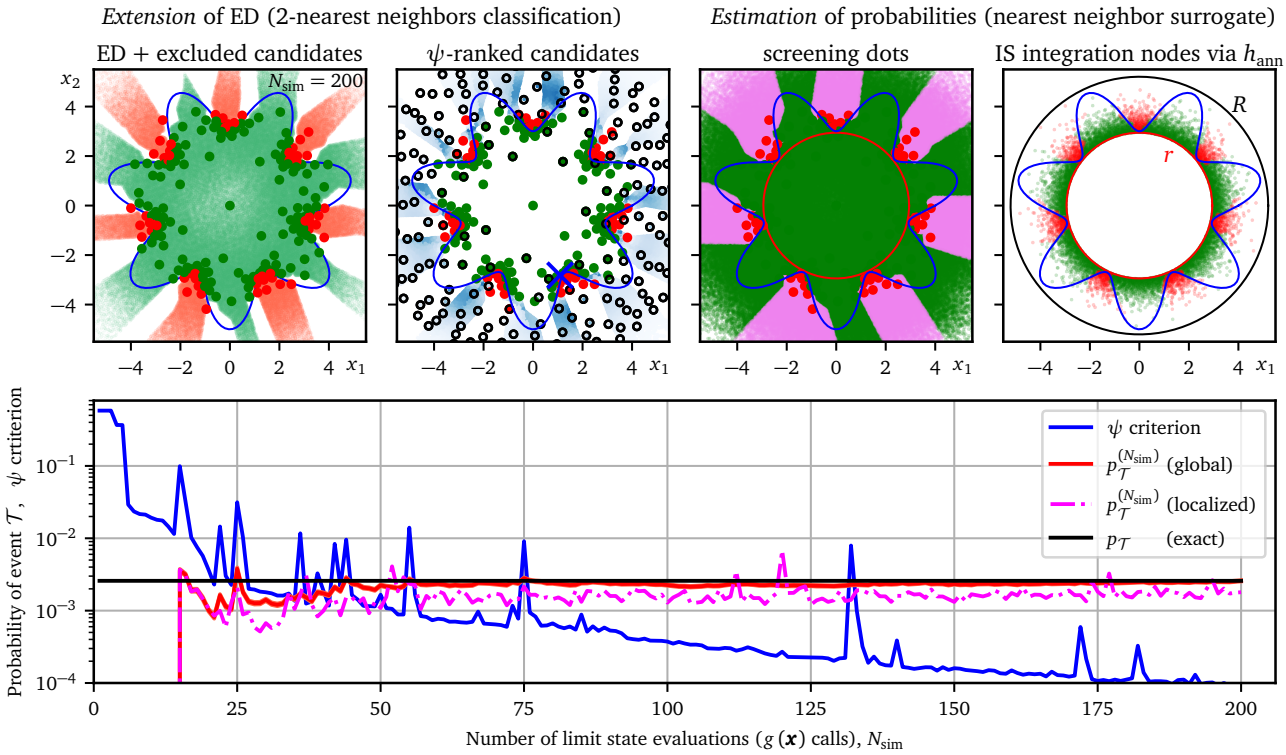

FIGURE 3. Top row: on-the-fly classification of territories. Two top left panels: two nearest neighbors for the classification of boundary regions – red and white candidates are censored out, and the retained (blue) candidates are considered for the *extension* of ED. Two top right panels: classification of dots and nodes via the nearest neighbor. Bottom: history of the failure probability *estimation* and the ψ criterion quantifying the approximate "probability bites" by selected candidates. The complete evolution of all panels is demonstrated in the Wavy circle video.

type is approximated by the event previously detected in the nearest ED point.

Suppose we have the classifier at hand in the form of an indicator function $\mathbf{1}_{\mathcal{F}}^{(N_{\text{sim}})}(\boldsymbol{x})$ based on N_{sim} evaluated points in which the $g(\boldsymbol{x})$ was evaluated previously. $\mathbf{1}_{\mathcal{F}}^{(N_{\text{sim}})}(\boldsymbol{x})$ approximates the true indicator function $\mathbf{1}_{\mathcal{F}}(\boldsymbol{x})$.

Using the importance sampling (IS) numerical integration, the probability of an event type $\mathcal{T}$ occurring when the indicator function $\mathbf{1}_{\mathcal{T}}(\boldsymbol{X})$ signals it is defined as the expectation: $p_{\mathcal{T}} = \mathbb{E}[\mathbf{1}_{\mathcal{T}}(\boldsymbol{X})] = \int \cdots \int_{\mathcal{D}} \mathbf{1}_{\mathcal{T}}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$. Let $h(\boldsymbol{x})$ be the IS density which is positive wherever event $\mathcal{T}$ occurs. The probability of event $\mathcal{T}$ can be rewritten as

$$p_{\mathcal{T}} = \int_{\mathcal{D}} \frac{\mathbf{1}_{\mathcal{T}}(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})}{h(\boldsymbol{x})} h(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \tag{3}$$

$$= \mathbb{E}_h \left[\frac{\mathbf{1}_{\mathcal{T}}(\boldsymbol{X}) f_{\boldsymbol{X}}(\boldsymbol{X})}{h(\boldsymbol{X})} \right], \tag{4}$$

where $\mathbb{E}_h[\cdot]$ denotes the expectation for $\boldsymbol{X}$ being distributed according to h: $\boldsymbol{X} \sim h$. The IS estimation of $p_{\mathcal{T}}$ based on the current approximation of $\mathbf{1}_{\mathcal{T}}^{(N_{\text{sim}})}(\boldsymbol{x})$ of the true indicator function $\mathbf{1}_{\mathcal{T}}(\boldsymbol{x})$ is made with n_{IS} integration nodes via the arithmetic average

$$p_{\mathcal{T}}^{(N_{\text{sim}})} \approx \frac{1}{n_{\text{IS}}} \sum_{i=1}^{n_{\text{IS}}} \mathbf{1}_{\mathcal{T}}^{(N_{\text{sim}})}(\boldsymbol{x}_i) \frac{f_{\boldsymbol{X}}(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)}, \quad \boldsymbol{X}_i \sim h. \tag{5}$$

The IS estimator is unbiased by construction.

We propose to use the following cubature via a *global* IS using the rotationally invariant density $h_{\text{ann}}(\boldsymbol{x}; r, R)$ in an important ring, which was introduced in Eq. (1). Suppose the screening dots were not localized in a very small region for which the localized importance sampling would be efficient. We propose the use of use a large pool of "integration nodes" selected from sampling density $h_{\text{ann}}(\boldsymbol{x})$ which excludes the useless N_{var}-ball of the radius r (= the distance of the most central "screening dot" from the origin; see the magenta points in Figure 3). Additionally, we propose that the density also excludes the exterior of the N_{var}-ball with the radius $R > r$. The outer radius R is obtained using the previous estimation of the rare event $p_{\mathcal{T}}^{(N_{\text{sim}})}$

$$R(p_{\mathcal{T}}^{(N_{\text{sim}})}; N_{\text{var}}) = F_{\rho}^{-1}\left(1 - \frac{p_{\mathcal{T}}^{(N_{\text{sim}})}}{10^4}; N_{\text{var}}\right). \tag{6}$$

Based on the information from the screening dots, we presume that the interior of the N_{var}-ball of radius r does not contain any failure event. The number 10^4 is selected to guarantee that the ignored space outside the ball of radius R is associated with a probability, which is four orders of magnitude less than the current estimate of the rare event probability. Different choices of this threshold are possible.

Let us now consider the sampling density $h_{\text{ann}}(\boldsymbol{x}; r, R)$ introduced in Eq. (1) and substitute it

into Eq. (5)

$$p_{\mathcal{T}}^{(N_{\text{sim}})} \approx \frac{1}{n_{\mathsf{IS}}} \sum_{i=1}^{n_{\mathsf{IS}}} \mathbf{1}_{\mathcal{T}}^{(N_{\text{sim}})}(\boldsymbol{x}_i) \frac{f_{\boldsymbol{X}}(\boldsymbol{x}_i)}{h_{\text{ann}}(\boldsymbol{x}_i)}$$
$$= \frac{p_{\text{ann}}}{n_{\mathsf{IS}}} \underbrace{\sum_{i=1}^{n_{\mathsf{IS}}} \mathbf{1}_{\mathcal{T}}^{(N_{\text{sim}})}(\boldsymbol{x}_i)}_{n_{\mathsf{IS},\mathcal{T}}} = p_{\text{ann}} \frac{n_{\mathsf{IS},\mathcal{T}}}{n_{\mathsf{IS}}}. \quad (7)$$

In other words, computation of the likelihood ratio ($f_{\boldsymbol{X}}/h$) at each node is *not* needed as the Gaussian densities cancel out. This is because the samples have their density proportional to $f_{\boldsymbol{X}}$. It suffices to simply compute the proportion of IS nodes that signaled event $\mathcal{T}$ (e.g., failure) and multiply it with p_{ann}. The closer the true event domain is to the annulus, the closer h_{ann} is to the optimal IS density, and so the estimation variance vanishes. With this strategy, the annulus between the two radii r and R is effectively examined. Whatever event occurs outside the outer radius R is associated with a negligible probability.

The variance of such an IS estimator is also simple to obtain

$$\text{Var}_h[p_{\mathcal{T}}^{(N_{\text{sim}})}] \approx \frac{1}{n_{\mathsf{IS}}} \left\{ \left[\frac{n_{\mathsf{IS},\mathcal{T}}}{n_{\mathsf{IS}}} p_{\text{ann}}^2 \right] - \left(p_{\mathcal{T}}^{(N_{\text{sim}})} \right)^2 \right\}$$
$$= \frac{p_{\mathcal{T}}^{(N_{\text{sim}})}}{n_{\mathsf{IS}}} \left(p_{\text{ann}} - p_{\mathcal{T}}^{(N_{\text{sim}})} \right), \quad (8)$$

and therefore, by using Eq. (7), the coefficient of variation of the estimator associated with the indicator function used is simply

$$\text{CoV}_h[p_{\mathcal{T}}^{(N_{\text{sim}})}] \approx \frac{1}{\sqrt{n_{\mathsf{IS}}}} \sqrt{\frac{p_{\text{ann}}}{p_{\mathcal{T}}^{(N_{\text{sim}})}} - 1} \quad (9)$$
$$= \frac{1}{\sqrt{n_{\mathsf{IS}}}} \sqrt{\frac{n_{\mathsf{IS}}}{n_{\mathsf{IS},\mathcal{T}}} - 1} \quad (10)$$

Figure 3 presents all the important information about the process of *extension* of ED (top left) and the *estimation* of probabilities (top right) which can be performed at any time during the process. The evolution of both is shown in the Wavy circle video covering the history from the very first limit state function evaluation up to $N_{\text{sim}} = 200$. The video frames, which can be displayed one by one, correspond to individual stages of the process, thus enabling a detailed inspection of the process. The accuracy of the estimation is excellent already at $N_{\text{sim}} \approx 70$ when all of the seven "failure regions" become discovered. The blue line plotted in the bottom diagram shows the amount of the "probability bite ψ" occupied by the neighborhood of the selected candidate. It can be seen that once about $N_{\text{sim}} = 30$ points have been evaluated, the ψ contributions become smaller than the rare event probability itself. The decrease in ψ with an increasing N_{sim} can be used, along with the stabilization of probability estimations, to decide the profitability of further $g(\boldsymbol{x})$ function evaluations (stopping criterion).

5. Global sensitivity measures

In [1] we also introduce a new generalized importance measure that makes no use of the performance function's gradient and considers all (known) points contributing to failure, weighted by the original density. The proposed measure comes as a by-product of the above-proposed technique for reliability estimations. Consider that the local contribution to failure probability of *any* point $\boldsymbol{x} = \{x_1, \ldots, x_v, \ldots, \boldsymbol{x}_{N_{\text{var}}}\}$ can be decomposed into individual coordinates in the spirit of the FORM α-sensitivities

$$\alpha_v^2(\boldsymbol{x}) = \left(\frac{x_v}{\rho(\boldsymbol{x})} \right)^2 = \frac{x_v^2}{\sum_{v=1}^{N_{\text{var}}} x_v^2}, \quad (11)$$

so that $\sum_{v=1}^{N_{\text{var}}} \alpha_v^2(\boldsymbol{x}) = 1$ for any $\boldsymbol{x}$. The denominator $\rho^2(\boldsymbol{x})$ is simply the squared Euclidean distance between a general point $\boldsymbol{x}$ and the origin. These $\alpha_v^2(\boldsymbol{x})$ can be viewed as importance measures of the individual dimensions and can be used as additive shares of probability density at any point $\boldsymbol{x}$. This choice is natural as α_vs weigh the individual increments to the distance ρ: $\alpha_v(x) = \frac{\partial \rho(\boldsymbol{x})}{\partial x_v}$. Using these shares, the total failure of probability can be written as

$$p_{\mathcal{F}} \equiv \int \cdots \int_{\mathcal{F}} \sum_{v=1}^{N_{\text{var}}} \alpha_v^2(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$
$$= \sum_{v=1}^{N_{\text{var}}} \int \cdots \int_{\mathcal{F}} \alpha_v^2(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \sum_{v=1}^{N_{\text{var}}} p_{\mathcal{F},v}. \quad (12)$$

In this way, the probability of failure (or analogously any event type) is obtained as a sum of contributions $p_{\mathcal{F},v}$ of individual variables. Each variable v contributes to $p_{\mathcal{F}}$ by

$$p_{\mathcal{F},v} = \int \cdots \int_{\mathcal{F}} \alpha_v^2(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \quad (13)$$
$$= \int \cdots \int_{\mathcal{F}} \underbrace{\left(\frac{x_v}{\rho(\boldsymbol{x})} \right)^2 f_{\boldsymbol{X}}(\boldsymbol{x})}_{f_v(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x} \quad (14)$$
$$= \int \cdots \int_{\mathcal{F}} f_v(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \quad (15)$$

where we define a part of the standard Gaussian density ascribed to a single variable v as (see also the illustrations in Figs. 4c and d)

$$f_v(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}) \left(\frac{x_v}{\rho(\boldsymbol{x})} \right)^2. \quad (16)$$

This local contribution to $p_{\mathcal{F},v}(\boldsymbol{x})$ is easy to evaluate as it is dependent only on the coordinates of point $\boldsymbol{x}$ and its standard Gaussian density. Finally, the shares $p_{\mathcal{F},v}$ can be standardized by $p_{\mathcal{F}}$ to form the proposed global importance measures s_v^2 associated with individual variables

$$s_v^2 = \frac{p_{\mathcal{F},v}}{p_{\mathcal{F}}}, \quad \text{i.e.} \sum_{v=1}^{N_{\text{var}}} s_v^2(\boldsymbol{x}) = 1. \quad (17)$$

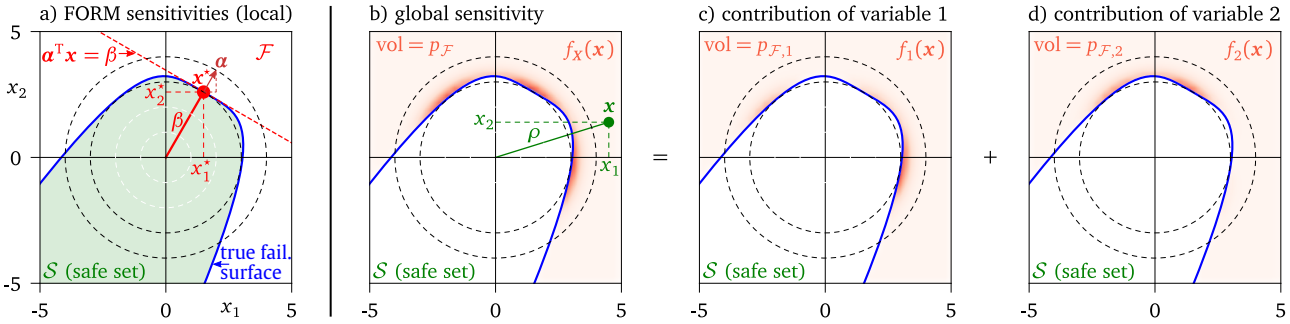

FIGURE 4. Measures of importance with respect to failure probability. a) illustration of FORM α-sensitivities constructed using linearization of a failure surface in the most central failure point $\boldsymbol{x}^\star$. b)-c) illustration of the proposed global sensitivity measures s_v^2 by decomposition of $p_\mathcal{F}$ into additive contributions by individual variables.

An important aspect is that the proposed s_v^2s are not based on values of the performance function, as only the binary information indicating an *event* is needed. This is an important property because the importance measures should *not* be dependent on the way a performance function is defined if it provides the same failure boundary. A robust importance measure for sensitivity to an event should be invariant under reformulations or reparametrizations of the underlying problem.

The evaluation of individual $p_{\mathcal{F},v}$ (and therefore also s_v^2) is very cheap as it can be seen as a by-product of the sampling analysis employed to deliver an estimation of $p_\mathcal{F}$. Suppose we have an existing set of points that were sampled proportionally to the standard Gaussian density $f_{\boldsymbol{X}}(\boldsymbol{x})$. It can be the global importance sample with N_{sim} points that was obtained outside the N_{var}-ball or in the annuloidal $h_{\text{ann}}(\boldsymbol{x}) \propto f_{\boldsymbol{X}}(\boldsymbol{x})$. From this sample, we only select a vector $\boldsymbol{x}$ containing $n_\mathcal{F}$ points marked as "failure". All these points are equally probable, and they each represent the same share of the estimated failure probability $p_\mathcal{F}(\boldsymbol{x}_i) = p_\mathcal{F}/n_\mathcal{F}$, $i = 1, \ldots, n_\mathcal{F}$. This share can be further split into individual directions in the spirit of Eq. (11) such that the contribution of the ith point in the vth direction reads

$$p_{\mathcal{F},v}(\boldsymbol{x}_i) = p_\mathcal{F}(\boldsymbol{x}_i)\frac{x_{i,v}^2}{\rho^2(\boldsymbol{x}_i)} = \frac{p_\mathcal{F}}{n_\mathcal{F}}\frac{x_{i,v}^2}{\rho^2(\boldsymbol{x}_i)}, \qquad (18)$$

i.e., each point $\boldsymbol{x}_i$ contributes

$$p_\mathcal{F}(\boldsymbol{x}_i) = \sum_{v=1}^{N_{\text{var}}} p_{\mathcal{F},v}(\boldsymbol{x}_i) = \frac{p_\mathcal{F}}{n_\mathcal{F}} \qquad (19)$$

where $x_{i,v}$ is the vth coordinate of point $\boldsymbol{x}_i$ and $\rho^2(\boldsymbol{x}_i) = \sum_{v=1}^{N_{\text{var}}} \boldsymbol{x}_v^2$ is the squared Euclidean distance of point $\boldsymbol{x}_i$ from the origin. The desired estimation of importance measures s_v^2 of individual directions can be obtained by summation over $n_\mathcal{F}$ failing nodes with a fixed direction index v and dividing by the failure probability

$$s_{\mathcal{F},v}^2 \approx \sum_{i=1}^{n_\mathcal{F}} \frac{p_{\mathcal{F},v}(\boldsymbol{x}_i)}{p_\mathcal{F}} = \frac{1}{n_\mathcal{F}} \sum_{i=1}^{n_\mathcal{F}} \frac{x_{i,v}^2}{\rho^2(\boldsymbol{x}_i)}. \qquad (20)$$

Therefore, the proposed sensitivity measure is just a cheap by-product of the proposed method. In cases when the sampling probability is not proportional to the standard Gaussian density, it is no longer true that all points have the same contribution of $1/n_\mathcal{F}$, and therefore straightforward re-scaling in a similar manner to importance sampling must be employed

$$s_{\mathcal{F},v}^2 \approx \frac{1}{p_\mathcal{F}} \frac{1}{n_{\text{IS}}} \sum_{i=1}^{n_{\text{IS}}} \mathbf{1}_\mathcal{F}^{(N_{\text{sim}})}(\boldsymbol{x}_i) \frac{f_{\boldsymbol{X}}(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)} \frac{x_{i,v}^2}{\rho^2(\boldsymbol{x}_i)}. \qquad (21)$$

6. Numerical examples

A variety of numerical examples which have been selected to explore different classes of problems posing unique challenges are presented in [1]. For the present short paper, we selected four two-dimensional problems defined in the space of independent standard Gaussian random variables; see Figure 5 for a quick overview of the selected functions. Then, problems in higher dimensions are analyzed.

In all definitions of the functions [1], we present the expressions that return smoothly varying output variables. However, the proposed extension algorithm receives categorical information only (such as binary "failure-success" codes). The same holds for the estimation, which uses only the indicator functions signaling an event. The only exceptions are "Four Branch" and "Metaballs" examples for which we also examine the degree of improvement in probability estimation when a smooth interpolation of the point-wise information in the ED via the Radial Basis Function is employed as the classifier.

We now briefly comment on numerical results obtained for the Four Branch example, which is a frequently used problem for demonstration of new techniques and for comparison of results. The following 2-dimensional "Four Branch function" [31] is a common benchmark problem in reliability analyses; see, e.g., studies with various parameter settings [4, 17, 32]. The function describes the failure of a *series system* with four distinct limit state components: two linear and two nonlinear branches of the failure surface. The

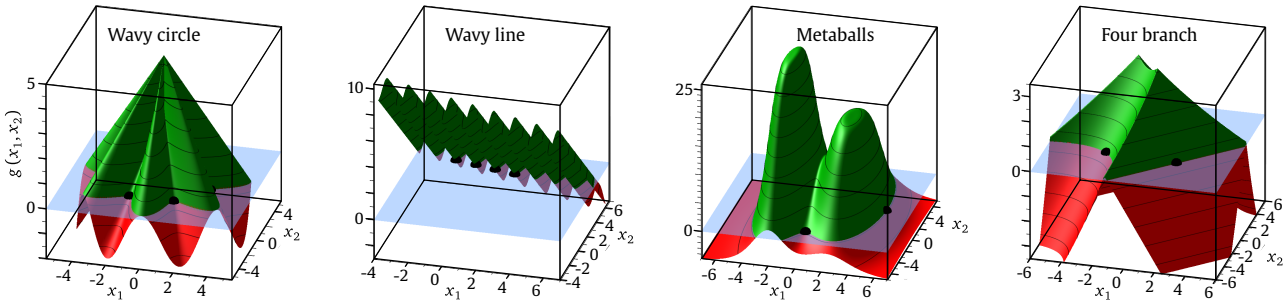

FIGURE 5. Overview of four two-dimensional examples used for demonstration of the proposed method. Design points are visualized using solid black balls.

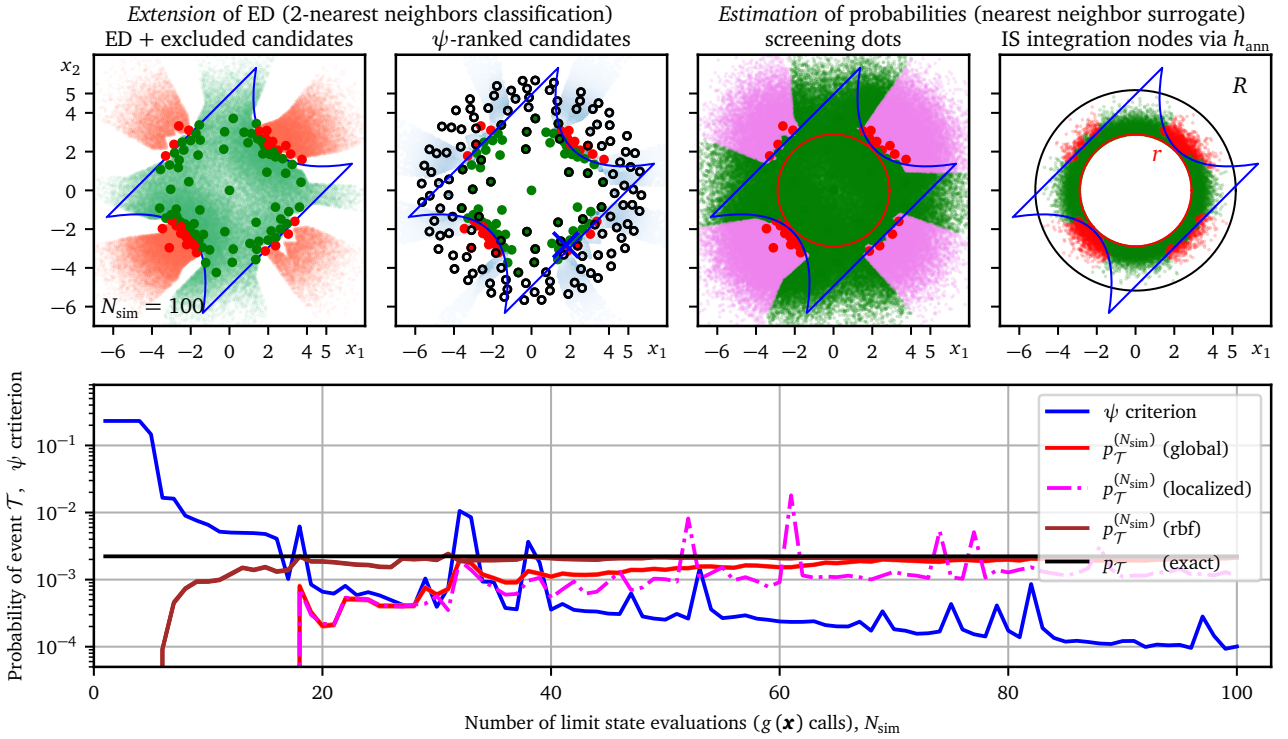

FIGURE 6. The "Four branch" problem. The complete evolution of all panels is available as Four Branch video.

limit state function reads

$$g(x_1,x_2)=\min\begin{cases}3+0.1\,(x_1-x_2)^2-(x_1+x_2)/\sqrt{2}\\ 3+0.1\,(x_1-x_2)^2+(x_1+x_2)/\sqrt{2}\\ x_1-x_2+7/\sqrt{2}\\ x_2-x_1+7/\sqrt{2}\end{cases} \tag{22}$$

This function can be also be viewed in the context of practical "assessment of existing structures": there are four distinct part of the failure set which are signalling failure. They can be viewed as different failure modes each of which can bring the structure down. The failure event is defined as $g(\boldsymbol{x})\le 0$. In the present definition, there are two pairs of design points: two points when $x_1=x_2=\pm 3\sqrt{2}/2$ at a distance $\beta_{1,2}=3$ and another two when $x_1=-x_2=\pm 7\sqrt{2}/4$ at a distance $\beta_{1,2}=3.5$. The exact result failure probability is $p_{\mathcal{F}}\approx 2.222\cdot 10^{-3}$, and the proposed global sensitivities read $s^2_{\mathcal{F},1}=s^2_{\mathcal{F},2}=0.5$.

Figure 6 shows the four blue lines forming the failure surface. The associated Four Branch video demonstrates the behavior of the proposed method by adding points one by one.

The extension algorithm refines the boundary proportionally to the probability density featured in the ψ criterion. The consequence of this is that the classification close to the four remote intersections of failure surfaces is not performed correctly for small designs. While this is no problem for the global IS probability estimation, which focuses on the high-density regions, it is a source of erroneous estimation for the local IS. As can be seen, the magenta dash-dot line in Fig. 6 represents wrong and unstable results, which are degraded due to the accentuation of the inaccurately classified regions (heavily covered by the local IS integration nodes). The global IS applied to the same binary surrogate classifier provides accurate and stable results for EDs as small as $N_{\text{sim}}\approx 80$. The top right panel in Fig. 6 shows that the corner regions are not important for the global IS.

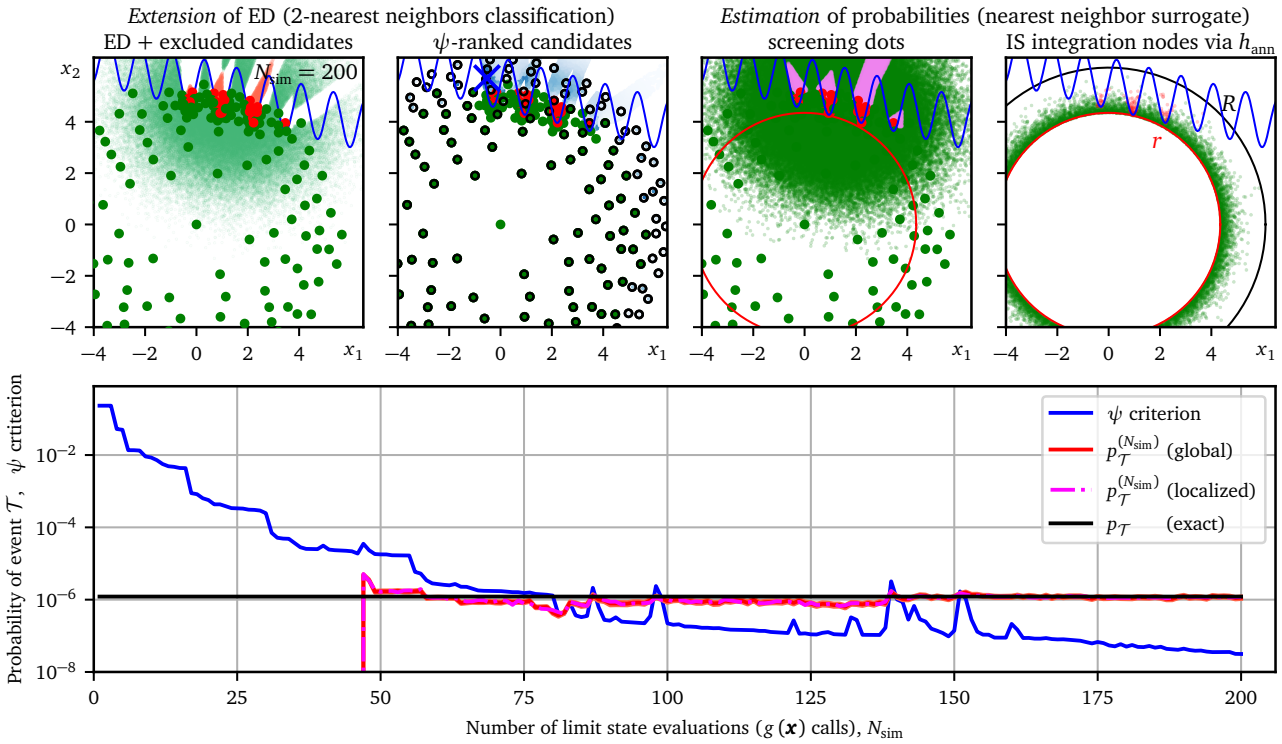

FIGURE 7. The "Wavy line" problem. The complete evolution of all panels is available as Wavy line video.

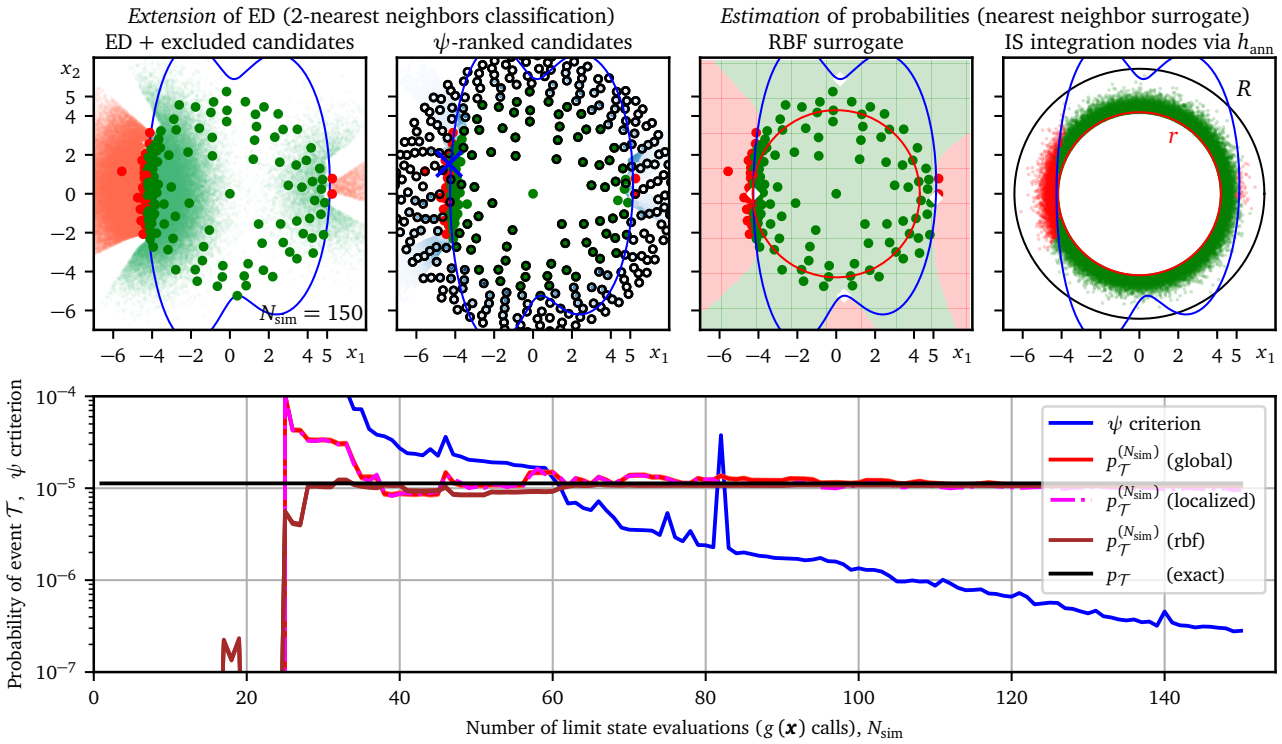

FIGURE 8. The "MetaBalls" problem. The complete evolution of all panels is available in the MetaBalls video.

We have examined the simplest Gaussian RBF-based classifier to show to what extent the estimation improves compared to the simple nearest neighbor classification. Using the global IS estimation leads to stable, almost exact results which are already in the range of $N_{\text{sim}} \in (20, 80)$, see the brown line in Fig. 6. The improvement is due to the more accurate classification of the failure surface in the vicinity of the four design points; the "corners" were not classified correctly neither by RBF nor by the binary surrogate. The efficiency of the RBF classifier is as good as the efficiency of the best methods known in the literature. Many existing sampling techniques, such as the sequential importance sampling employed in [4], use

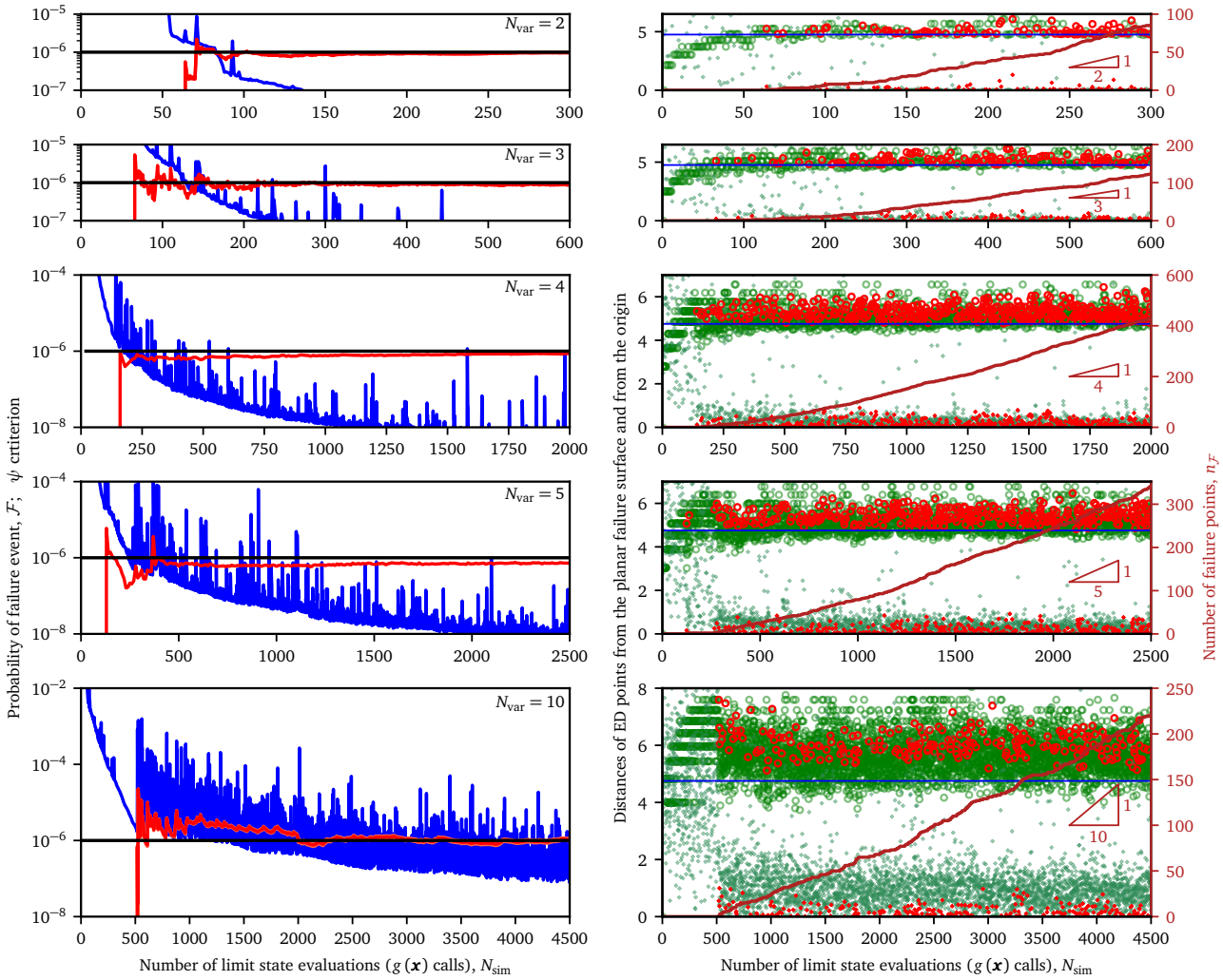

FIGURE 9. Recorded histories of exploration and estimation for the "Linear Failure Surface" problem in Eq. (23) studied for $N_{var} = 2$, 3, 4, 5 and 10 (rows). The left column reports histories of the ψ criterion and the estimated $p_{\mathcal{F}}^{(N_{sim})}$. The right column reports the radial distances of ED points from the origin (empty circles), as well as distances from the planar failure surface (small diamonds). The secondary vertical axis shows the number of failure points. The complete evolution of ED extension in $N_{var} = 2$ and 3 dimensions is shown in video 2D and video 3D.

too many limit state function calls, and good results are obtained only when building a surrogate model. We repeat again that the proposed algorithm achieves almost the same efficiency; however, it does so using only the categorical information about $g(\boldsymbol{x})$, which makes the proposed method very robust.

Figs. 7 and 8 show the behavior of the method for other functions. The detailed description can be found in [1], along with more examples posing other challenges to the existing methods, which are elegantly overcome by the proposed method.

6.1. LINEAR FAILURE SURFACE IN HIGHER DIMENSIONS

The previous 2D examples revealed the robustness of the algorithm regarding the complicated failure surface and $g(\boldsymbol{x})$ function values. What remains a question is how the algorithm efficiency scales with dimension. To present a reasonable higher-dimensional example relevant to many practical problems, we use a compromise: a simple linear failure boundary (a line, a plane, or generally a hyperplane). There is no reason to make the linear failure surface rotated in the space of input variables because the proposed framework is rotationally invariant in Gaussian space. Therefore, it suffices to make the limit state function simply depend on the first dimension only

$$g(x_1, x_2, \ldots) = \beta - x_1. \tag{23}$$

Such a problem has a trivial analytical solution: $p_{\mathcal{F}} = \Phi(-\beta)$. We set $\beta = 4.7534243$ to achieve the failure probability $p_{\mathcal{F}} = 10^{-6}$, i.e., a number relevant for "assessment of existing structures".

Fig. 9 presents the results for $N_{var} = 2, 3, 4, 5$ and 10 dimensions. The blue line in the left column is the estimated amount of probability resolved by evaluation of the corresponding $g(\boldsymbol{x})$ (ψ criterion), and the red line plots the estimates of the failure probability. It is clear that the purely "exploratory phase" with the expanding search until the first failure is hit

consumes increasingly more $g(\boldsymbol{x})$ function evaluations as the space dimension increases; compare the radial distances of points from the origin plotted as empty circles in the right column of Fig. 9, which are organized at individual distance levels. For $N_{\text{var}} = 10$, it takes about 500 $g(\boldsymbol{x})$ function calls to hit the failure event and begin refining the (very large) failure surface. Stabilization of the probability estimates necessitates very fine refinement of the failure surface, which also consumes many function calls. The need to spend higher numbers of points in the purely exploratory phase in higher dimensions is an inevitable consequence of the fact that the numerical value of $g(\boldsymbol{x})$ cannot be used to orient the search, e.g., in the direction of the negative gradient, as Subset Simulation or methods building a smooth surrogate do. The estimated $p_{\mathcal{F}}$ for $N_{\text{var}} = 10$ shows quite a serrated profile, although the coefficient of variation is very small due to the use of a high number n_{IS} of integration nodes. The reason is that the boundary approximated via the Voronoï cells in the nearest neighbor classification is also very serrated. Its extent is large, and the refinement would necessitate very many additional $g(\boldsymbol{X})$ calls. One can also notice that the decrease in ψ criterion is less rapid in high dimensions because the volume of the space simply increases with the space dimension. This is manifested via the increase of the extent of the failure surface part with a high Gaussian density. Apart from the radial distances plotted via empty circles in the right column of Fig. 9, we also plot the distance of points from the planar failure surface as small diamonds (green and red failure points). Once the first failure event is hit, the extension algorithm primarily selects the points to refine the failure surface. However, as can be seen, their distance from the origin is considerably greater than the shortest distance of the plane β, which is marked by the horizontal blue line. The complete evolution of the refinement process is captured point-by-point in individual frames of video 2D and video 3D.

The non-decreasing, maroon-colored line in the right column of Fig. 9 shows that the proposed refinement algorithm tends to the ratio $n_{\mathcal{F}}/N_{\text{sim}} = 1/N_{\text{var}}$, which is excellent in $N_{\text{var}} = 2$ dimensions where almost all limit state function evaluations refine the failure surface from both sides, but less efficient in higher dimensions in which increasingly more points are spent on the exploration of new territories and a smaller share is devoted to boundary refinement.

7. Conclusions

This paper briefly presents simple yet robust and efficient methods for the sequential *extension* of experimental design and *estimation* of failure probabilities for computational models, which can be non-smooth, or returns only a finite number of states or even have blind spots for which there is no result at all.

The *extension* of experimental design balances the *gradual exploration* of new territories by expanding the covered region and *refinement via the exploitation* of important regions. The balance is maintained by maximizing the proposed simple ψ criterion which expresses the approximate amount of probability being classified by any proposed candidate for extension. The *extension* of the experimental design makes no assumptions about the performance function and, therefore, is invariant with respect to its reparametrizations and reformulations, which do not alter the failure domain shape & location, and the method is resistant against noise and jumps.

The *estimation* task can be performed at any time during the extension process by quickly analyzing the point-wise information only. It uses standard importance sampling applied to a surrogate model.

The proposed method needs no fine-tuning of parameters depending on the analyst, because there are no such variables. The only freedom is in the density of the initial *exploration set* and with the option of refining it during the extension process. The method can help in solving hard practical reliability problems for which the existing methods fail due to their strong assumptions about the performance function being well-formulated and well-behaved.

Simple yet apt *global sensitivity measures* are proposed, which can be obtained for any rare event type as a by-product of the presented method.

The applicability of the methods has been demonstrated for small to medium dimensions; high dimensions (several tens to hundreds of independent input variables) remain a challenge.

Acknowledgements

The author acknowledges the financial support provided by the Ministry of Education, Youth and Sports of the Czech Republic under project No. LTAUSA19058, and additionally by the Czech Science Foundation under project No. 20-01781S.

References

[1] M. Vořechovský. Reliability analysis of discrete-state performance functions via adaptive sequential sampling with detection of failure surfaces. *Computer Methods in Applied Mechanics and Engineering* p. under review.

[2] A. Harbitz. Efficient and accurate probability of failure calculation by the use of the importance sampling technique. In G. Augusti (ed.), 4^{th} *Int. Conf. on Applications of Statistics and Probability in Soil and Structural Engineering*, vol. 3, pp. 825–836. Univ. di Firenze (Italy), Pitagora Editrice, Bologna, Italy, 1983.

[3] M. Shinozuka. Basic Analysis of Structural Safety. *Journal of Structural Engineering* **109**(3):721-40, 1983. `https://doi.org/10.1061/(asce)0733-9445(1983)109:3(721)`.

[4] I. Papaioannou, C. Papadimitriou, D. Straub. Sequential importance sampling for structural reliability analysis. *Structural Safety* **62**:66–75, 2016. `https://doi.org/10.1016/j.strusafe.2016.06.002`.

[5] C. Bucher. Asymptotic sampling for high-dimensional reliability analysis. *Probabilistic Engineering Mechanics* **24**(4):504–510, 2009. https://doi.org/10.1016/j.probengmech.2009.03.002.

[6] A. M. Hasofer, N. Lind. Exact and invariant second-moment code format. *Journal of Engineering Mechanics Division – ASCE* **100**(1):111–121, 1974. https://doi.org/10.1061/JMCEA3.0001848.

[7] R. Rackwitz, B. Fiessler. Structural reliability under combined random load sequences. *Computers & Structures* **9**(5):489–494, 1978. https://doi.org/10.1016/0045-7949(78)90046-9.

[8] B. Fiessler, H.-J. Neumann, R. Rackwitz. Quadratic limit states in structural reliability. *Journal of the Engineering Mechanics Division* **105**(4):661–676, 1979. https://doi.org/10.1061/jmcea3.0002512.

[9] M. Hohenbichler, S. Gollwitzer, W. Kruse, R. Rackwitz. New light on first- and second-order reliability methods. *Structural Safety* **4**(4):267–284, 1987. https://doi.org/10.1016/0167-4730(87)90002-6.

[10] K. W. Breitung, M. Hohenbichler. Asymptotic approximations for multivariate integrals with an application to multinormal probabilities. *Journal of Multivariate Analysis* **30**(1):80–97, 1989. https://doi.org/10.1016/0047-259x(89)90089-4.

[11] L. Tvedt. Distribution of Quadratic Forms in Normal Space-Application to Structural Reliability. *Journal of Engineering Mechanics* **116**(6):1183-97, 1990. https://doi.org/10.1061/(asce)0733-9399(1990)116:6(1183).

[12] O. Ditlevsen, H. O. Madsen. *Structural Reliability Methods*. John Wiley & Sons, 1996.

[13] H. Madsen, S. Krenk, N. Lind. *Methods of Structural Safety*. Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

[14] R. E. Melchers, A. T. Beck (eds.). *Structural Reliability Analysis and Prediction*. John Wiley & Sons Ltd, 2017. https://doi.org/10.1002/9781119266105.

[15] S.-K. Au, J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics* **16**(4):263–277, 2001. https://doi.org/10.1016/s0266-8920(01)00019-4.

[16] R. Teixeira, M. Nogal, A. O'Connor. Adaptive approaches in metamodel-based reliability analysis: A review. *Structural Safety* **89**:102019, 2021. https://doi.org/10.1016/j.strusafe.2020.102019.

[17] B. Echard, N. Gayton, M. Lemaire. AK-MCS: An active learning reliability method combining Kriging and Monte Carlo simulation. *Structural Safety* **33**(2):145–154, 2011. https://doi.org/10.1016/j.strusafe.2011.01.002.

[18] J. Wang, G. Xu, Y. Li, A. Kareem. AKSE: A novel adaptive Kriging method combining sampling region scheme and error-based stopping criterion for structural reliability analysis. *Reliability Engineering & System Safety* **219**:108214, 2022. https://doi.org/10.1016/j.ress.2021.108214.

[19] S. Marelli, B. Sudret. An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis. *Structural Safety* **75**:67–74, 2018. https://doi.org/10.1016/j.strusafe.2018.06.003.

[20] Y. Zhou, Z. Lu, W. Yun. Active sparse polynomial chaos expansion for system reliability analysis. *Reliability Engineering & System Safety* **202**:107025, 2020. https://doi.org/10.1016/j.ress.2020.107025.

[21] X. Li, C. Gong, L. Gu, et al. A sequential surrogate method for reliability analysis based on radial basis function. *Structural Safety* **73**:42–53, 2018. https://doi.org/10.1016/j.strusafe.2018.02.005.

[22] L. Shi, B. Sun, D. S. Ibrahim. An active learning reliability method with multiple kernel functions based on radial basis function. *Structural and Multidisciplinary Optimization* **60**(1):211–229, 2019. https://doi.org/10.1007/s00158-019-02210-0.

[23] W. J. de Santana Gomes. Structural reliability analysis using adaptive artificial neural networks. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* **5**(4), 2019. https://doi.org/10.1115/1.4044040.

[24] W. J. S. Gomes. Shallow and deep artificial neural networks for structural reliability analysis. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* 2020. https://doi.org/10.1115/1.4047636.

[25] S. S. Afshari, F. Enayatollahi, X. Xu, X. Liang. Machine learning-based methods in structural reliability analysis: A review. *Reliability Engineering & System Safety* **219**:108223, 2022. https://doi.org/10.1016/j.ress.2021.108223.

[26] H.-S. Li, Z.-Z. Lü, Z.-F. Yue. Support vector machine for structural reliability analysis. *Applied Mathematics and Mechanics* **27**(10):1295–1303, 2006. https://doi.org/10.1007/s10483-006-1001-z.

[27] J.-M. Bourinet, F. Deheeger, M. Lemaire. Assessing small failure probabilities by combined subset simulation and Support Vector Machines. *Structural Safety* **33**(6):343–353, 2011. https://doi.org/10.1016/j.strusafe.2011.06.001.

[28] J.-M. Bourinet. Rare-event probability estimation with adaptive support vector regression surrogates. *Reliability Engineering & System Safety* **150**:210–221, 2016. https://doi.org/10.1016/j.ress.2016.01.023.

[29] Q. Pan, D. Dias. An efficient reliability method combining adaptive Support Vector Machine and Monte Carlo simulation. *Structural Safety* **67**:85–95, 2017. https://doi.org/10.1016/j.strusafe.2017.04.006.

[30] A. Roy, S. Chakraborty. Reliability analysis of structures by a three-stage sequential sampling based adaptive support vector regression model. *Reliability Engineering & System Safety* **219**:108260, 2022. https://doi.org/10.1016/j.ress.2021.108260.

[31] A. Borri, E. Speranzini. Structural reliability analysis using a standard deterministic finite element code. *Structural Safety* **19**(4):361–382, 1997. https://doi.org/10.1016/s0167-4730(97)00017-9.

[32] L. Schueremans, D. V. Gemert. Benefit of splines and neural networks in simulation based structural reliability analysis. *Structural Safety* **27**(3):246–261, 2005. https://doi.org/10.1016/j.strusafe.2004.11.001.