Langevin Dynamics for Importance Sampling in Reliability Analysis

Armin Tabandeh
Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, IL, USA

Gaofeng Jia
Department of Civil and Environmental Engineering, Colorado State University, CO, USA

Paolo Gardoni
Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, IL, USA

ABSTRACT: Importance Sampling (IS) is a popular simulation method for the reliability analysis of general engineering systems. Its core idea is to build a biased density that can sample from the failure domain more frequently than in the Monte Carlo simulation. The optimal IS density, derived from the Euler-Lagrange equation, is inversely proportional to the unknown failure probability to be estimated; thus, directly sampling from it is impossible. However, samples can indirectly be drawn from the optimal IS density by simulating a Markov chain. In particular, the Hamiltonian Monte Carlo (HMC) sampling algorithm and its variants design a dynamical system that can explore the failure domain better than, for example, the Metropolis-Hastings algorithm. However, such sampling algorithms can be prohibitively slow if the problem is high-dimensional or involves expensive model evaluations. This paper presents a method to convert the inference problem in HMC into an optimization problem and discusses its connection with existing optimization-based algorithms for IS. The starting point is the Langevin equation, which offers a unified formulation for several variants of HMC. The optimal IS density is the steady-state solution of the Fokker-Planck equation associated with the Langevin equation. The proposed method approximates the steady-state solution of the Fokker-Planck equation via optimization. The process is illustrated through a benchmark reliability problem.

Monte Carlo simulation (MCS) is among the most versatile solution methods for reliability analysis. However, its slow convergence is a primary concern in evaluating small failure probabilities. Variance-reduction methods are popular strategies to improve the convergence rate of MCS while retaining its versatility. Importance Sampling (IS) is perhaps the most researched variance-reduction method in reliability analysis (Engelund and Rackwitz 1993, Tabandeh et al. 2022). Its core idea is to define a biased density (i.e., the IS density) that can sample from the failure domain more frequently than in MCS (Robert and Casella 2004). The optimal choice for the biased density is the one that minimizes the variance of the estimator for failure probability (hence, the name variance-reduction.) The closed-form solution to the optimization problem depends on the sought failure probability. However, we may devise simulation methods to sample from the optimal biased density or approximate it in some statistical sense.

Simulation methods that sample from the optimal biased density generate a Markov chain or process that must converge to the target density and explore its support (i.e., failure domain). The performance of simulation methods depends on their ability to rapidly reach and explore the failure domain. A typical class of such simulation methods is based on the Metropolis-Hastings algorithm (Metropolis et al. 1953, Hastings 1970). The random walk behavior of the Markov chain in this algorithm has been criticized for its slow convergence (Neal 1993, Roberts et al. 1997). The Hybrid or Hamiltonian Monte Carlo algorithm (Duane et al., 1987) and its several variants avoid the random walk behavior by designing artificial dynamics for the problem. Parameters for the

artificial dynamics are tuned to control the algorithm's performance. Besides simulation algorithms, the literature includes reliability methods that estimate the biased density by minimizing a measure of discrepancy between the optimal solution and a family of approximating densities. We distinguish different variants by their choice of discrepancy measure and approximating density (see Tabandeh et al. 2022 for a recent review.)

This paper develops a hybrid solution for IS by leveraging artificial dynamics for simulations and the parametric approximation of the biased density. The core idea is to create a parametric density that approximately satisfies the governing equation of the designed dynamics for simulation. The probability density of the Markov process in simulation algorithms satisfies the Fokker-Planck equation (Risken 1996). We present conditions for the equation parameters that ensure its steady-state solution matches the optimal biased density. Thus, densities that approximately satisfy the Fokker-Planck equation also approximate the optimal biased density. We show the mathematical connection between the residual of the Fokker-Planck equation and discrepancy measures to approximate a density. The tuning parameters of the designed dynamics (i.e., those that also appear in the Fokker-Planck equation) that result in improved simulation performance can also help the specification of the discrepancy measure for density approximation methods and vice versa. We briefly discuss the geometric interpretation of such tuning parameters in the formulated problem to estimate the parametric biased density.

The rest of the paper consists of five sections. Section 1 reviews the general formulation of sampling methods for reliability analysis, including Monte Carlo and Importance Sampling. Section 2 presents the equations for the artificial dynamics in Hamiltonian Monte Carlo and its stochastic extension. Section 3 discusses the proposed approach, including its formulation and solution algorithm. Section 5 illustrates some of the concepts for a benchmark reliability problem. Finally, the last section summarizes the paper and draws some conclusions.

1. SAMPLING FOR RELIABILITY ANALYSIS

Consider a system described by the vector of state variables $\mathbf{x} \subseteq \mathbb{R}^d$ with probability density $p(\mathbf{x})$. The limit-state function $g: \mathbb{R}^d \to \mathbb{R}$ defines its failure domain as $\Omega_F = \{\mathbf{x} \in \mathbb{R}^d : g(\mathbf{x}) \leq 0\}$. The system's failure probability P_F is then written as (Ditlevsen and Madsen 1996, Gardoni 2017)

$$\begin{aligned} P_F &= \int_{\Omega_F} p(\mathbf{x}) \mathrm{d}\mathbf{x} \\ &= \int_{\mathbb{R}^d} I_{\Omega_F}(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x} = \mathbb{E}_p\left[I_{\Omega_F}(\mathbf{x})\right], \end{aligned} \tag{1}$$

where $I_{\Omega_F}(\mathbf{x})$ is the indicator function, and $\mathbb{E}_p(\cdot)$ is the expected value with respect to $p(\mathbf{x})$.

Computing P_F usually involves a high-dimensional integral, not analytically tractable. Simulation methods are the general approach to evaluating high-dimensional integrals based on an unbiased and consistent estimator. For example, the Monte Carlo estimator of P_F is

$$\hat{P}_{F,p}(\mathcal{D}_N) = \frac{1}{N} \sum_{n=1}^{N} I_{\Omega_F}\left(\mathbf{x}^{(n)}\right), \tag{2}$$

where $\mathcal{D}_N = \left\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\right\}$ is a set of statistically independent samples from $p(\mathbf{x})$.

The mean and variance of $\hat{P}_{F,p}$ are

$$\begin{aligned} \mathbb{E}_p\left[\hat{P}_{F,p}(\mathcal{D}_N)\right] &= P_F, \\ \mathrm{Var}_p\left[\hat{P}_{F,p}(\mathcal{D}_N)\right] &= \frac{P_F(1-P_F)}{N}, \end{aligned} \tag{3}$$

which indicate that $\hat{P}_{F,p}$ is an unbiased and consistent estimator. From its Coefficient of Variation (CoV)

$$\delta_{MC} = \frac{\sqrt{\mathrm{Var}_p\left[\hat{P}_{F,p}(\mathcal{D}_N)\right]}}{\mathbb{E}_p\left[\hat{P}_{F,p}(\mathcal{D}_N)\right]} = \sqrt{\frac{1-P_F}{NP_F}}, \tag{4}$$

we find that almost $\lceil 1/(\delta_{MC}^2 P_F) \rceil$ samples are required to estimate P_F with the confidence level of $(1-\delta_{MC})$.

Importance sampling aims to reduce the estimator's variance by introducing an alternative sampling density $q_{IS}(\mathbf{x})$ to estimate P_F as

$$P_F = \int_{\mathbb{R}^d} I_{\Omega_F}(\mathbf{x}) \frac{p(\mathbf{x})}{q_{IS}(\mathbf{x})} q_{IS}(\mathbf{x}) \mathrm{d}\mathbf{x} = \mathbb{E}_q\left[I_{\Omega_F}(\mathbf{x}) \frac{p(\mathbf{x})}{q_{IS}(\mathbf{x})} \right], \quad (5)$$

where $q_{IS}(\mathbf{x}) > 0$ for all $\mathbf{x} \in \Omega_F$. As for the Monte Carlo simulation, we estimate P_F as

$$\hat{P}_{F,q}(\mathcal{D}_N) = \frac{1}{N} \sum_{n=1}^{N} I_{\Omega_F}\left(\mathbf{x}^{(n)}\right) \frac{p\left(\mathbf{x}^{(n)}\right)}{q_{IS}(\mathbf{x}^{(n)})}, \quad (6)$$

where $\mathcal{D}_N$ is now a set of N independent samples from $q_{IS}(\mathbf{x})$. It is straightforward to show that $\mathbb{E}_q\left(\hat{P}_{F,q}\right) = P_F$, implying an unbiased estimator. We can also write the estimator's variance as

$$\mathrm{Var}_q\left(\hat{P}_{F,q}\right) = \frac{1}{N}\left\{ \mathbb{E}_q\left[I_{\Omega_F}(\mathbf{x}) \frac{p(\mathbf{x})^2}{q_{IS}(\mathbf{x})^2} \right] - P_F^2 \right\}, \quad (7)$$

which is minimized by $q_{IS}^*(\mathbf{x}) = I_{\Omega_F}(\mathbf{x}) p(\mathbf{x}) / P_F$, with $\mathrm{Var}_{q^*}\left(\hat{P}_{F,q^*}\right) = 0$ for all $N \in \mathbb{N}$ (Srinivasan 2013, Tabandeh et al. 2022).

2. LANGEVIN DYNAMICS FOR IMPORTANCE SAMPLING

Direct sampling from $q_{IS}^*(\mathbf{x})$ is not possible since it depends on P_F. However, we can sample from it indirectly by simulating a Markov Chain with stationary density $q_{IS}^*(\mathbf{x})$. Specifically, we may design a dynamical system whose response is a Markov process with stationary density $q_{IS}^*(\mathbf{x})$. Let $\mathbf{y}_t^\intercal = (\mathbf{x}_t^\intercal, \mathbf{v}_t^\intercal)$ be the state variables of the dynamical system, consisting of displacement $\mathbf{x}_t \in \mathbb{R}^d$ and velocity $\mathbf{v}_t \in \mathbb{R}^d$ vectors. The vector $\mathbf{y}_t$ satisfies the Langevin equation

$$\mathrm{d}\mathbf{y}_t = \mathbf{a}(\mathbf{y}_t)\mathrm{d}t + \sqrt{2\mathbf{D}(\mathbf{y}_t)}\mathrm{d}\mathbf{w}_t, \quad (8)$$

with initial condition $\mathbf{y}_0$, where $\mathbf{a}: \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ is the drift vector, $\mathbf{D}$ is a $(2d \times 2d)$ symmetric positive-semidefinite diffusion matrix, and $\mathbf{w}_t \in \mathbb{R}^{2d}$ is the Wiener process.

The transition probability density $f(\mathbf{y}, t)$ of the Markov process $\mathbf{y}_t$ satisfies the Fokker-Planck equation

$$\frac{\partial f(\mathbf{y}, t)}{\partial t} = \nabla^\intercal \cdot \{ -\mathbf{a}(\mathbf{y}) f(\mathbf{y}, t) + \nabla \cdot [\mathbf{D}(\mathbf{y}) f(\mathbf{y}, t)] \}, \quad (9)$$

with initial condition $f(\mathbf{y}, t = 0) = \delta(\mathbf{y} - \mathbf{y}_0)$, where $\delta(\cdot)$ is the Dirac delta function, and $\nabla \cdot$ is the divergence operator with respect to $\mathbf{y}$.

The design of $\mathbf{a}(\mathbf{y})$ and $\mathbf{D}(\mathbf{y})$ should ensure that $q_{IS}^*(\mathbf{x}) = f_s^*(\mathbf{x}) = \int f_s^*(\mathbf{x}, \mathbf{v}) \mathrm{d}\mathbf{v}$, where $f_s^*(\mathbf{x}, \mathbf{v}) = \lim_{t \to \infty} f(\mathbf{x}, \mathbf{v}, t)$ is the stationary density of $\mathbf{y}_t$ (i.e., the steady-state solution of the Fokker-Planck equation.) Following Ma et al. (2015), we define $\mathbf{a}(\mathbf{y})$ as

$$\mathbf{a}(\mathbf{y}) = -[\mathbf{D}(\mathbf{y}) + \mathbf{Q}(\mathbf{y})] \nabla H(\mathbf{y}) + \nabla \cdot [\mathbf{D}(\mathbf{y}) + \mathbf{Q}(\mathbf{y})], \quad (10)$$

where $\mathbf{Q}(\mathbf{y})$ is a $(2d \times 2d)$ skew-symmetric matrix (i.e., $\mathbf{Q} = -\mathbf{Q}^\intercal$), and $H: \mathbb{R}^{2d} \to \mathbb{R}$ is the Hamiltonian of the system, representing its total energy. i.e.,

$$H(\mathbf{y}) = U(\mathbf{x}) + K(\mathbf{v}), \quad (11)$$

in which $U(\mathbf{x}) = -\ln[q_{IS}^*(\mathbf{x})]$ is the potential energy, $K(\mathbf{v}) = 1/2\, \mathbf{v}^\intercal \boldsymbol{\Sigma} \mathbf{v}$ is the kinetic energy, and $\boldsymbol{\Sigma}$ is a (constant) symmetric, positive semidefinite matrix. The steady-state solution of the Fokker-Planck equation is $f_s^*(\mathbf{y}) \propto \exp[-H(\mathbf{y})] = q_{IS}^*(\mathbf{x}) \exp(-1/2\, \mathbf{v}^\intercal \boldsymbol{\Sigma} \mathbf{v})$ (Ma et al. 2015). Thus, the necessary condition $q_{IS}^*(\mathbf{x}) = \int f_s^*(\mathbf{x}, \mathbf{v}) \mathrm{d}\mathbf{v}$ is satisfied.

The dynamical system can better explore the target probability distribution than the Metropolis-Hastings algorithm (Neal 2011). Depending on the choice of $\mathbf{D}$ and $\mathbf{Q}$, we can retrieve some variants of the Langevin and Hamiltonian Monte Carlo algorithms (Ma et al. 2015). For example, the Hamiltonian Monte Carlo (HMC) (Neal 2011) is a special case with $\mathbf{D} = \mathbf{0}$ and

$$\mathbf{Q} = \begin{pmatrix} \mathbf{0}_{d \times d} & -\mathbf{I}_{d \times d} \\ \mathbf{I}_{d \times d} & \mathbf{0}_{d \times d} \end{pmatrix}, \quad (12)$$

where $\mathbf{0}_{d \times d}$ is the $(d \times d)$-matrix of zeros, and $\mathbf{I}_{d \times d}$ is the $(d \times d)$ identity matrix. The Riemannian Manifold Langevin Monte Carlo (Xifara 2014) is another special case with $\boldsymbol{\Sigma} = \mathbf{0}$, $\mathbf{Q} = \mathbf{0}$, and

$$\mathbf{D} = \begin{pmatrix} \boldsymbol{\Gamma}(\mathbf{x})^{-1} & \mathbf{0}_{d \times d} \\ \mathbf{0}_{d \times d} & \mathbf{0}_{d \times d} \end{pmatrix}, \quad (13)$$

where $\boldsymbol{\Gamma}(\mathbf{x})$ is a metric tensor (e.g., the Fisher information metric.) Such algorithms rely on the designed continuous dynamics to accelerate

convergence. However, discretizing continuous dynamics for numerical implementations may not preserve the target convergence rate and stationary distribution (Mou et al. 2021, Barp et al. 2022). The literature includes physical and geometrical solutions to ensure that time-integration schemes leave the stationary distribution invariant (Barp et al. 2022). However, preserving the convergence rate remains an open challenge.

Instead of directly using dynamical systems to sample from $q_{IS}^*(\mathbf{x})$, we use the associated Fokker-Planck equation to approximate $q_{IS}^*(\mathbf{x})$ via optimization. In particular, using $\hat{f}_s(\mathbf{y})$ in writing $H(\mathbf{y})$ and, hence, in the drift vector, we can express the residual of the steady-state Fokker-Planck equation as

$$r(\mathbf{y};\hat{f}_s) = \nabla^{\intercal}\cdot\{-\mathbf{a}(\mathbf{y};\hat{f}_s)f_s^*(\mathbf{y}) + \nabla \cdot [\mathbf{D}(\mathbf{y})f_s^*(\mathbf{y})]\}, \tag{14}$$

which is uniformly zero at all $\mathbf{y}$ values for $\hat{f}_s = f_s^*$. In the next section, we discuss a learning algorithm to estimate a specific $\hat{f}_s(\mathbf{y})$ with $r(\mathbf{y};\hat{f}_s) \approx 0$.

3. LEARNING ALGORITHM TO ESTIMATE THE IMPORTANCE SAMPLING DENSITY

Let $\mathcal{F} = \{f_s(\mathbf{y}) \propto q_{IS}(\mathbf{x})\exp(-1/2\,\mathbf{v}^{\intercal}\boldsymbol{\Sigma}\mathbf{v}) : q_{IS}(\mathbf{x}) > 0 \text{ for all } \mathbf{x} \in \Omega_F\}$ be the space of admissible densities to approximate $f_s^*(\mathbf{y})$. The learning algorithm involves two main tasks, defining the subspace of candidate densities $\hat{\mathcal{F}} \subseteq \mathcal{F}$ to solve the steady-state Fokker-Planck equation, and finding the density $\hat{f}_s^* \in \hat{\mathcal{F}}$ that closely approximates $f_s^*(\mathbf{y})$ in the sense that $r(\mathbf{y};f_s^*) \approx 0$ at selected $\mathbf{y}$ values. Finding $\hat{f}_s^*(\mathbf{y})$ amounts to solving the optimization problem $\hat{f}_s^* = \arg\min_{\hat{f}_s\in\hat{\mathcal{F}}}\|r(\mathbf{y};\hat{f}_s)\|$, where $\|\cdot\|$ is a norm.

3.1. Formulation of the optimization problem

To define the optimization problem, we take a closer look at $r(\mathbf{y};\hat{f}_s)$. Substituting $\mathbf{a}(\mathbf{y};\hat{f}_s)$ into Eq. (14) yields $r(\mathbf{y};\hat{f}_s)$, in indicial notation, as

$$r(\mathbf{y};\hat{f}_s) = \frac{\partial}{\partial y_i}\left[(D_{ij}+Q_{ij})\frac{\partial \widehat{H}}{\partial y_j}f_s^* - \frac{\partial Q_{ij}}{\partial y_j}f_s^* + D_{ij}\frac{\partial f_s^*}{\partial y_j}\right], \tag{15}$$

where $i,j = 1,\dots,2d$, and a summation runs over repeated indices in each term. Also, $\widehat{H}$ indicates an approximation to the Hamiltonian, in which q_{IS}^* is replaced by $\hat{q}_{IS}$.

We rewrite $\partial Q_{ij}/\partial y_j\, f_s^*$ in Eq. (15), using the identity

$$\frac{\partial}{\partial y_i}\left(\frac{\partial Q_{ij}}{\partial y_j}f_s^*\right) = \frac{\partial^2 Q_{ij}}{\partial y_i \partial y_j}f_s^* + \frac{\partial Q_{ij}}{\partial y_j}\frac{\partial f_s^*}{\partial y_i}, \tag{16}$$

in which the first term is zero since Q_{ij} is skew-symmetric. We then rewrite Eq. (16) as

$$\begin{aligned}\frac{\partial}{\partial y_i}\left(\frac{\partial Q_{ij}}{\partial y_j}f_s^*\right) &= \frac{\partial Q_{ij}}{\partial y_j}\frac{\partial f_s^*}{\partial y_i} - Q_{ij}\frac{\partial^2 f_s^*}{\partial y_i \partial y_i} \\ &= -\frac{\partial Q_{ij}}{\partial y_i}\frac{\partial f_s^*}{\partial y_j} - Q_{ij}\frac{\partial^2 f_s^*}{\partial y_i \partial y_j} \\ &= -\frac{\partial}{\partial y_i}\left(Q_{ij}\frac{\partial f_s^*}{\partial y_j}\right),\end{aligned} \tag{17}$$

where $Q_{ij}\,\partial^2 f_s^*/\partial y_i \partial y_j$, added in the first line, is zero due to the skew symmetry of Q_{ij} and the symmetry of $\partial^2 f_s^*/\partial y_i \partial y_j$; and the second line follows by switching the indices i and j in the first term and using $Q_{ji} = -Q_{ij}$. Substituting back Eq. (17) into Eq. (15) results in

$$\begin{aligned}r(\mathbf{y};\hat{f}_s) &= \frac{\partial}{\partial y_i}\left[(D_{ij}+Q_{ij})\left(\frac{\partial \widehat{H}}{\partial y_j}f_s^* + \frac{\partial f_s^*}{\partial y_j}\right)\right] \\ &= \frac{\partial}{\partial y_i}\left[(D_{ij}+Q_{ij})e^{-\widehat{H}}\frac{\partial\left(f_s^* e^{\widehat{H}}\right)}{\partial y_j}\right].\end{aligned} \tag{18}$$

Using $f_s^*(\mathbf{y}) \propto q_{IS}^*(\mathbf{x})\exp(-1/2\,\mathbf{v}^{\intercal}\boldsymbol{\Sigma}\mathbf{v})$ and $\widehat{H} = -\ln[\hat{q}_{IS}(\mathbf{x})] + 1/2\,\mathbf{v}^{\intercal}\boldsymbol{\Sigma}\mathbf{v}$ in Eq. (18) yields

$$r(\mathbf{y};\hat{f}_s) = \frac{\partial}{\partial y_i}\left[(D_{ij}+Q_{ij})e^{-\widehat{H}}\frac{\partial}{\partial y_j}\frac{q_{IS}^*}{\hat{q}_{IS}}\right], \tag{19}$$

where $\mathbf{D}$ and $\mathbf{Q}$ in sampling algorithms are usually block matrices written as

$$\mathbf{D} = \begin{pmatrix}\mathbf{D}_{11}(\mathbf{x}) & \mathbf{0}_{d\times d}\\ \mathbf{0}_{d\times d} & \mathbf{D}_{22}(\mathbf{x})\end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix}\mathbf{0}_{d\times d} & \mathbf{Q}_{12}(\mathbf{x})\\ -\mathbf{Q}_{12}(\mathbf{x}) & \mathbf{0}_{d\times d}\end{pmatrix}. \tag{20}$$

Substituting the above $\mathbf{D}$ and $\mathbf{Q}$ into Eq. (19) and assuming a constant mass matrix $\boldsymbol{\Sigma}$ results in

$$r\left(\mathbf{y};\hat{f}_s\right) = e^{-1/2 v_k \Sigma_{kl} v_l}\left\{\frac{\partial}{\partial x_i}\left[q_{IS}^* S_{ij}\frac{\partial}{\partial x_j}\ln\frac{q_{IS}^*}{\hat{q}_{IS}}\right] + (\Sigma_{il}v_l + \Sigma_{ki}v_k)\left[q_{IS}^* C_{ij}\frac{\partial}{\partial x_j}\ln\frac{q_{IS}^*}{\hat{q}_{IS}}\right]\right\}, \quad (21)$$

where $i,j,k,l = 1,\dots,d$, and $\mathbf{S} = [S_{ij}]$ and $\mathbf{C} = [C_{ij}]$ are new notations for $\mathbf{D}_{11}$ and $\mathbf{Q}_{12}$.

The first term inside the curly bracket in Eq. (21) only depends on $\mathbf{x}$, whereas the second term depends on both $\mathbf{x}$ and $\mathbf{v}$. A sufficient condition to approximately satisfy $r\left(\mathbf{y};\hat{f}_s\right) = 0$ is then

$$\left|q_{IS}^* \Gamma_{ij}^{-1/2}\frac{\partial}{\partial x_j}\ln\frac{q_{IS}^*}{\hat{q}_{IS}}\right| \le \varepsilon, \quad (22)$$

for every $i = 1,\dots,d$, and $\mathbf{C} = \mathbf{S} = \boldsymbol{\Gamma}^{-1/2}$, where $\varepsilon > 0$ is a small constant. Accordingly, we write the following optimization problem to find $\hat{q}_{IS}^*$:

$$\hat{q}_{IS}^* = \arg\min_{\hat{q}_{IS}} \mathbb{E}_{q^*}\|\nabla \ln q_{IS}^* - \nabla \ln \hat{q}_{IS}\|_{\boldsymbol{\Gamma}^{-1}}^2. \quad (23)$$

where $\|\mathbf{x}\|_{\boldsymbol{\Gamma}^{-1}}^2 = \mathbf{x}^\intercal \boldsymbol{\Gamma}^{-1}\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^d$.

The above optimization problem is related to minimizing the Kullback-Leibler (KL) distance between q_{IS}^* and $\hat{q}_{IS}$ in the following way. Using the logarithmic Sobolev inequality, we show that

$$\mathcal{D}(q_{IS}^* \parallel \hat{q}_{IS}) \le \kappa \mathbb{E}_{q^*}\|\nabla \ln q_{IS}^* - \nabla \ln \hat{q}_{IS}\|_{\boldsymbol{\Gamma}^{-1}}^2, \quad (24)$$

where $\mathcal{D}(\cdot\|\cdot)$ is the KL distance, and $\kappa > 0$ is a constant. Past studies have used the KL distance to find q_{IS}^* (Cappé 2008, Tabandeh et al. 2022).

A probability measure μ on $\mathbb{R}^d$ satisfies the logarithmic Sobolev inequality with constant $\kappa' > 0$ if for any smooth function h, we have (Gross 1975, Zahm et al. 2022)

$$\int h^2 \ln\frac{h^2}{\int h^2 \mathrm{d}\mu}\mathrm{d}\mu \le \kappa' \int \|\nabla h\|_{\boldsymbol{\Gamma}^{-1}}^2 \mathrm{d}\mu. \quad (25)$$

Setting $h^2 = q_{IS}^*/\hat{q}_{IS}$, we can find that $\nabla h = 1/2\sqrt{q_{IS}^*/\hat{q}_{IS}}\,\nabla \ln(q_{IS}^*/\hat{q}_{IS})$. Using this definition and the Radon-Nikodym derivative $\mathrm{d}\mu/\mathrm{d}\mathbf{x} = \hat{q}_{IS}$ in Eq. (25) results in

$$\begin{aligned}\mathcal{D}(q_{IS}^* \parallel \hat{q}_{IS}) &= \int q_{IS}^* \ln(q_{IS}^*/\hat{q}_{IS})\,\mathrm{d}\mathbf{x} \\ &\le \frac{\kappa'}{4}\int q_{IS}^* \|\nabla \ln(q_{IS}^*/\hat{q}_{IS})\|_{\boldsymbol{\Gamma}^{-1}}^2 \mathrm{d}\mathbf{x} \\ &= \kappa \mathbb{E}_{q^*}\|\nabla \ln q_{IS}^* - \nabla \ln \hat{q}_{IS}\|_{\boldsymbol{\Gamma}^{-1}}^2.\end{aligned} \quad (26)$$

The field $\boldsymbol{\Gamma}^{-1}$ in the optimization acts as a filter that amplifies errors in specific directions. The metric tensor $\boldsymbol{\Gamma}$ identifies the principal curvatures of q_{IS}^* at every $\mathbf{x} \in \mathbb{R}^d$ and projects the error $\nabla \ln q_{IS}^* - \nabla \ln \hat{q}_{IS}$ onto principal directions. For smooth q_{IS}^* and comparable principal curvatures, we may use $\boldsymbol{\Gamma} = \mathbf{I}$ as in the HMC algorithm.

3.2. Solution algorithm

Using a simulation method to evaluate $\mathbb{E}_{q^*}\|\cdot\|_{\boldsymbol{\Gamma}^{-1}}^2$ results in the following optimization problem:

$$\hat{q}_{IS}^* = \arg\min_{\hat{q}_{IS}} \frac{1}{M}\sum_{m,i} I_{\Omega_F}\left|\Gamma_{ij}^{-1/2}\frac{\partial}{\partial x_j}\ln\frac{q_{IS}^*}{\hat{q}_{IS}}\right|_{\mathbf{x}=\mathbf{x}^{(m)}}^2, \quad (27)$$

where $\mathbf{x}^{(1)},\dots,\mathbf{x}^{(M)}$ are independent samples from $p(\mathbf{x})$. The solution we seek is based on approximating $g(\mathbf{x})$ in the definition of q_{IS}^*. Specifically, we write $\hat{q}_{IS}(\mathbf{x},\boldsymbol{\theta}) \propto I_{\widehat{\Omega}_F(\boldsymbol{\theta})}(\mathbf{x})p(\mathbf{x})$, where $\widehat{\Omega}_F(\boldsymbol{\theta}) = \{\mathbf{x} \in \mathbb{R}^d : \hat{g}(\mathbf{x},\boldsymbol{\theta}) \le 0\}$ is the failure domain based on a computationally efficient surrogate $\hat{g}(\mathbf{x},\boldsymbol{\theta})$ for $g(\mathbf{x})$ with the vector of unknown parameters $\boldsymbol{\theta}$. Thus, the optimization problem amounts to estimating the vector $\boldsymbol{\theta}^*$ of $\hat{q}_{IS}^*$. Also, capturing the mismatch between $\hat{g}(\mathbf{x},\boldsymbol{\theta})$ and $g(\mathbf{x})$, we rewrite the optimization problem as

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{M}\sum_{m,i} I_{\Omega_F}\left|\Gamma_{ij}^{-1/2}\frac{\partial \ln I_{\widehat{\Omega}_F}}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}^{(m)}}^2 + \lambda|g-\hat{g}|_{\mathbf{x}=\mathbf{x}^{(m)}}^2, \quad (28)$$

where $\lambda \ge 0$ is a prescribed regularization parameter controlling the contribution of $|g-\hat{g}|^2$ in estimating $\boldsymbol{\theta}^*$. Since $I_{\widehat{\Omega}_F}$ in the above equation is not differentiable, we use $I_{\widehat{\Omega}_F} \cong 1/\left(1+e^{\hat{g}/\eta}\right)$ as a smooth approximation, where $\eta > 0$ is a small constant. Hence, we arrive at

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{M}\sum_{m,i} I_{\Omega_F}\left|\frac{\Gamma_{ij}^{-1/2}\,\partial\hat{g}/\partial x_j}{1+e^{-\hat{g}/\eta}}\right|_{\mathbf{x}=\mathbf{x}^{(m)}}^2 + \lambda|g-\hat{g}|_{\mathbf{x}=\mathbf{x}^{(m)}}^2. \quad (29)$$

The first term in the above equation is non-zero only for $\mathbf{x}^{(m)} \in \Omega_F$. However, the majority of samples from $p(\mathbf{x})$ results in $I_{\Omega_F} = 0$ for rare

events. We use an adaptive approach to reduce the computational cost of generating samples in Ω_F. The approach involves an iterative process to generate more samples in Ω_F and improve the estimate of $\boldsymbol{\theta}^*$. Let $\boldsymbol{\theta}^\tau$ denote the most recent estimate of θ^*. The next iteration involves solving $\boldsymbol{\theta}^{\tau+1}$

$$= \arg\min_{\boldsymbol{\theta}} \frac{1}{\tau M} \sum_{m,i} I_{\Omega_F} \frac{p}{\hat{q}_{IS}^{\tau}} \left| \frac{\Gamma_{ij}^{-1/2} \, \partial \hat{g}/\partial x_j}{1 + e^{-\hat{g}/\eta}} \right|^2_{\mathbf{x}=\mathbf{x}^{(\tau} + \lambda |g - \hat{g}|^2_{\mathbf{x}=\mathbf{x}^{(m)}}, \tag{30}$$

where $\mathbf{x}^{(\tau M+1)}, \ldots, \mathbf{x}^{(\tau M+M)}$ are independent samples from $\hat{q}_{IS}^{\tau}(\mathbf{x}) \propto I_{\widehat{\Omega}_F(\boldsymbol{\theta}^\tau)}(\mathbf{x}) p(\mathbf{x})$, in which $\widehat{\Omega}_F(\boldsymbol{\theta}^\tau)$ is the failure domain defined based on the most recent estimate of the surrogate $\hat{g}(\mathbf{x}, \boldsymbol{\theta}^\tau)$.

We initiate the adaptive process with the sampling density $\hat{q}_{IS}^{0}(\mathbf{x}) = p(\mathbf{x})$; hence, the first few rounds of estimating $\boldsymbol{\theta}^*$ is mainly controlled by $|g - \hat{g}|^2$. However, after several iterations, we gradually generate more samples in Ω_F at a reduced cost and increase the contribution of the first term in Eq. (30)(28) to estimate $\boldsymbol{\theta}^*$.

Substituting the inferred surrogate $\hat{q}_{IS}^{*}(\mathbf{x}) = I_{\widehat{\Omega}_F(\boldsymbol{\theta}^*)}(\mathbf{x}) p(\mathbf{x}) / P_F^*$ into Eq. (5) results in

$$\begin{aligned} P_F &= P_F^* \int_{\mathbb{R}^d} \frac{I_{\Omega_F}(\mathbf{x})}{I_{\widehat{\Omega}_F(\boldsymbol{\theta}^*)}(\mathbf{x})} \hat{q}_{IS}^{*}(\mathbf{x}) \mathrm{d}\mathbf{x} \\ &= P_F^* \mathcal{D}_{IS}(g, \hat{g}), \end{aligned} \tag{31}$$

where P_F^* is the Monte Carlo estimate of P_F by replacing I_{Ω_F} with $I_{\widehat{\Omega}_F(\boldsymbol{\theta}^*)}$ in Eq. (2), and $\mathcal{D}_{IS}(g, \hat{g})$ is the correction factor capturing the error in approximating g by $\hat{g}$, also evaluated using the Monte Carlo estimator with samples from $\hat{q}_{IS}^{*}(\mathbf{x})$.

4. NUMERICAL EXAMPLE

The example is a component reliability problem with the following metaball limit-state function (Breitung 2019):

$$g(\mathbf{x}) = \frac{30}{\left[\frac{4}{9}(x_1 + 2)^2 + \frac{1}{25} x_2^2\right]^2 + 1} + \frac{20}{\left[\frac{1}{4}(x_1 - 2.5)^2 + \frac{1}{25}(x_2 - 0.5)^2\right] + 1} - 5, \tag{32}$$

where x_1 and x_2 are independent random variables with a standard Gaussian distribution. Figure 1 shows $g(\mathbf{x})$ and the points (blue asterisks) on the level-sets $\{\mathbf{x} \in \mathbb{R}^2 : g(\mathbf{x}) = c\}$ with the shortest distance to the origin as c varies from 20 to 0. The figure also shows the contour plot of $g(\mathbf{x})$ with failure boundary in bold red (i.e., the level-set with $c = 0$.)

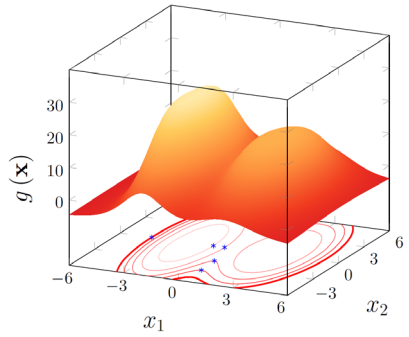
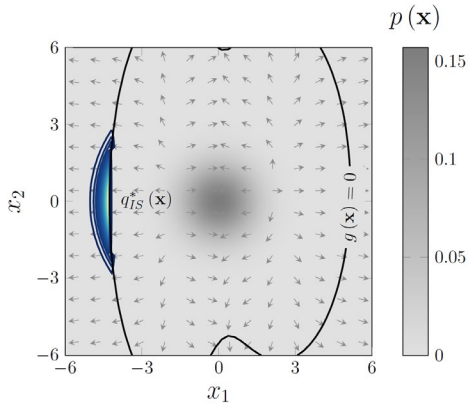


Figure 1: The limit-state function $g(\boldsymbol{x})$ and design points (blue asterisks) on its level-sets (red curves).

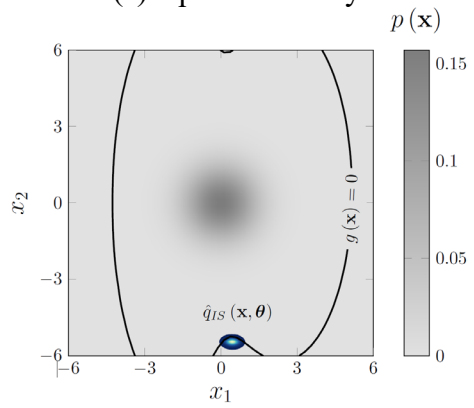
We evaluate the performance of the proposed method in terms of its computational cost and estimation accuracy, where the number of $g(\mathbf{x})$ calls is a proxy for computational cost. The convergence criterion is based on the CoV of the estimator, set to 0.05. Figure 2 shows the optimal IS density and the estimated ones based on two approximation methods, a Gaussian mixture with KL divergence (Geyer et al. 2019) and the proposed approach. Each subplot shows the contour of $q_{IS}(\mathbf{x})$, superimposed on $g(\mathbf{x}) = 0$ and $p(\mathbf{x})$. The first subplot also shows the vector field $-\nabla g / \|\nabla g\|$ to help us explain the results. The Monte Carlo estimate of P_F with $\delta_{MC} = 0.05$ is 1.02×10^{-5}. The estimate based on the Gaussian mixture model is 1.6×10^{-8} with a total of 7,359 calls to $g(\mathbf{x})$, and the proposed method yields 8.67×10^{-6} with a total of 221 calls to $g(\mathbf{x})$. The specific setting of the proposed method is as follows: we used a 2nd-order polynomial for $\hat{g}(\mathbf{x}, \boldsymbol{\theta})$ with $\boldsymbol{\Gamma} = \mathbf{I}$ and $M = 25$, and the adaptive algorithm converged after 8 iterations. We also had 21 (i.e., $N = 21$ in Eq. (6)) more calls to $g(\mathbf{x})$ to estimate $\mathcal{D}_{IS}(g, \hat{g})$ in Eq. (31).

We can observe that because of the changing topological structure of $g(\mathbf{x})$ (see Fig. 1),

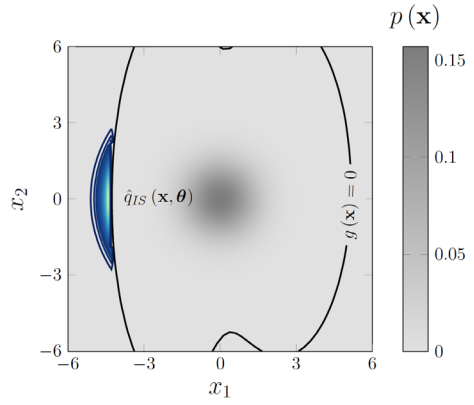
adaptive methods may get trapped within the band $x_1 \in [-0.5,1]$ (see the vector field in Fig. 2a) and converge to a locally optimal point, as in Fig. 2(b). However, random perturbations and information on ∇g can help escape the band $x_1 \in [-0.5,1]$ and find the design point, resulting in an improved $\hat{q}_{IS}^*$.



(a) Optimal density



(b) Gaussian mixture with KL divergence



(c) Proposed method

Figure 2: The contour plot of the importance sampling density for a metaball limit-state function.

5. CONCLUSIONS

The paper discussed state-of-the-art simulation methods for reliability analysis. These methods are based on artificial dynamics to improve the convergence rate of simulations. The (stochastic) Langevin equation governs the artificial dynamics. The solution of the Langevin equation is a Markov process that converges to the target density (e.g., the optimal importance sampling density in reliability analysis.) The probability density of the Markov process satisfies the Fokker-Planck equation. The current use of this equation is instrumental in the following way: any artificial dynamics is valid as long as its stationary density satisfies the Fokker-Planck equation. Instead, this paper presented a novel approach that uses the equation to estimate an importance sampling density via optimization. The paper also discussed the connection with existing Importance Sampling methods for reliability analysis. The insights from a benchmark example suggest that the designed artificial dynamics for reliability problems in the context of stochastic simulations or optimization is essentially a global search for design point(s). Thus, future works must investigate specific conditions (i.e., problem formulation and solution algorithms) that can help us solve such global optimization problems.

6. REFERENCES

Barp, A., Da Costa, L., França, G., Friston, K., Girolami, M., Jordan, M.I., and Pavliotis, G.A. (2022). "Geometric methods for sampling, optimization, inference and adaptive agents." in Srinivasa, R., Arni, S.R., Young, G.A., and Rao, C.R. (Eds.), *Handbook of Statistics*, Academic Press, 21-78.

Breitung, K. (2019). "The geometry of limit state function graphs and subset simulation: Counterexamples." *Reliability Engineering & System Safety*, 182, 98-106.

Cappé, O., Douc, R., Guillin, A., Marin, J.M., and Robert, C.P. (2008). "Adaptive importance sampling in general mixture classes." *Statistics and Computing*, 18, 447-459.

Ditlevsen, O., and Madsen, H.O. (1996). *Structural reliability methods*. Wiley, New York, NY.

Duane, S., Kennedy, A.D., Pendleton, B.J., and Roweth, D. (1987). "Hybrid Monte Carlo." *Physics Letters B*, 195 (2), 216-222.

Engelund, S., and Rackwitz, R. (1993). "A benchmark study on importance sampling techniques in structural reliability." *Structural Safety*, 12 (4), 255-276.

Gardoni, P., (Ed.), (2017). Risk and reliability analysis: Theory and applications, Springer.

Geyer, S., Papaioannou, I., and Straub, D. (2019). "Cross entropy-based importance sampling using Gaussian densities revisited." *Structural Safety*, 76, 15-27.

Gross, L. (1975). "Logarithmic Sobolev inequalities." *American Journal of Mathematics*, 97 (4), 1061-1083.

Hastings, W.K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57 (1), 97-109.

Ma, Y.A., Chen, T., and Fox, E.B. (2015). "A complete recipe for stochastic gradient MCMC." in *Advances in Neural Information Processing Systems* (*NIPS*).

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). "Equation of state calculations by fast computing machines." *The Journal of Chemical Physics*, 21 (6), 1087-1092.

Mou, W., Ma, Y.A., Wainwright, M.J., Bartlett, P.L., and Jordan, M.I. (2021). "High-order Langevin diffusion yields an accelerated MCMC algorithm." *Journal of Machine Learning Research*, 22 (42), 1-41.

Neal, R.M. (1993). *Probabilistic inference using Markov chain Monte Carlo methods* Technical Report CRG-TR-93-1, University of Toronto, ON

Neal, R.M. (2011). "MCMC using Hamiltonian dynamics." in Brooks, S., Gelman, A., Jones, G., and Meng, X.L. (Eds.), *Handbook of Markov Chain Monte Carlo*, CRC Press, 113-162.

Risken, H. (1996). The Fokker-Planck equation: Methods of solution and applications. Springer Berlin, Heidelberg.

Robert, C., and Casella, G. (2004). *Monte Carlo statistical methods*. Springer, New York, NY.

Roberts, G.O., Gelman, A., and Gilks, W.R. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithms." *The Annals of Applied Probability*, 7 (1), 110-120.

Srinivasan, R. (2013). Importance sampling: Applications in communications and detection. Springer Science & Business Media.

Tabandeh, A., Jia, G., and Gardoni, P. (2022). "A review and assessment of importance sampling methods for reliability analysis." *Structural Safety*, 97, 102216.

Xifara, T., Sherlock, C., Livingstone, S., Byrne, S., and Girolami, M. (2014). "Langevin diffusions and the Metropolis-adjusted Langevin algorithm." *Statistics & Probability Letters*, 91, 14-19.

Zahm, O., Cui, T., Law, K., Spantini, A., and Marzouk, Y. (2022). "Certified dimension reduction in nonlinear Bayesian inverse problems." *Mathematics of Computation*, 91 (336), 1789-1835.