SIAM J. SCI. COMPUT.
Vol. 40, No. 1, pp. A548–A572

ASYMPTOTICALLY EFFICIENT SIMULATION OF ELLIPTIC PROBLEMS WITH SMALL RANDOM FORCING*

XIAOLIANG WAN† AND XIANG ZHOU‡

Abstract. Recent rare-event simulations show that the large deviation principle (LDP) for stochastic problems plays an important role in both theory and simulation, for studying rare events induced by small noise. Practical challenges of applying this useful technique include minimizing the rate function numerically and incorporating the minimizer into the importance sampling scheme for the construction of efficient probability estimators. For a spatially extended system where the noise is modeled as a random field, even for simple steady state problems, many new issues are encountered in comparison to the finite dimensional models. We consider the Poisson equation subject to a Gaussian random forcing with vanishing amplitude. In contrast to the simplified rate functional given by space white noise, we consider the covariance operator of trace class such that the effects of small noise of moderate or large correlation length on rare events can be studied. We have constructed an LDP-based importance sampling estimator with a sufficient and necessary condition to guarantee the weak efficiency, where numerical approximation of the large deviation principle is also addressed. Numerical studies are presented.

Key words. small random perturbation, Gaussian random field, importance sampling, large deviation principle, rare events, uncertainty quantification

AMS subject classifications. 65C05, 65N25, 60F10, 35J25

DOI. 10.1137/17M111643X

1. Introduction.

During the past two decades, there has been a widespread interest in uncertainty quantification to develop stochastic models and approaches to quantitatively describe the propagation of uncertainty in complex systems. The following stochastic partial differential equation (SPDE) is a general model for spatially extended systems perturbed by the additive noise:

$$\partial_t u_\varepsilon(t, \mathbf{x}) + \mathcal{L}(u_\varepsilon) = \sqrt{\varepsilon} G(t, \mathbf{x}), \quad \mathbf{x} \in D \subset \mathbb{R}^d, \tag{1}$$

where $\mathcal{L}$ is a spatial differentiation operator, ε is a small positive number and G is a zero-mean Gaussian process which is white in time but is allowed to be colored in space. The distribution of $G(t, \mathbf{x})$ is specified by its spatio-temporal covariance. Assume G is white in time; then $\mathbb{E}(G(t, \mathbf{x})G(t', \mathbf{y})) = K(\mathbf{x}, \mathbf{y})\delta(t - t')$. The spatial covariance function K is the kernel of the covariance operator $\mathcal{Q}$ on a Hilbert space where $G(t)$ takes its values. The examples of (1) appear in many fields, such as the stochastic Navier–Stokes equations [26, 16, 31, 34], stochastic Cahn–Hillard equation [35], stochastic Kardar–Parisi–Zhang equation [12], etc.

We are concerned with the rare events in model (1). For this randomly perturbed system, the classic work is the Freidlin–Wentzell (F-W) large deviation principle

*Submitted to the journal's Methods and Algorithms for Scientific Computing section February 14, 2017; accepted for publication (in revised form) October 5, 2017; published electronically February 20, 2018.
http://www.siam.org/journals/sisc/40-1/M111643.html
Funding: The first author's work was partially supported by NSF grant DMS-1620026 and AFOSR grant FA9550-15-1-0051. The second author's work was supported by General Research Grants (11304314, 11304715, 11337216) from the Research Grants Council of the Hong Kong Special Administrative Region, China.
†Department of Mathematics, Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803 (xlwan@math.lsu.edu).
‡Department of Mathematics, City University of Hong Kong, Hong Kong, China (xiang.zhou@cityu.edu.hk).

(LDP) [14]. The key object of F-W LDP is the rate functional, or action functional. Assume that the time integration in model (1) is in the Itô sense. The rate functional is defined as [11]

$$I_{\mathcal{Q}}(u) = \frac{1}{2}\int_0^T \|\mathcal{Q}^{-1/2}(\partial_t u - \mathcal{L}(u))\|^2 dt, \tag{2}$$

where $\|\cdot\|$ indicates the L^2 norm on the physical domain D. Simply speaking, the main conclusion of F-W LDP is that for any Borel set B, the probability $\Pr(u_\varepsilon \in B)$ can be characterized by the minimizer of the rate functional, more specifically

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log \Pr(u_\varepsilon \in B) = -\inf_{u_\varepsilon\in B} I_{\mathcal{Q}}(u_\varepsilon). \tag{3}$$

From the application point of view, an immediate problem is to seek the minimizer of the rate functional. The associated Euler–Lagrange (E-L) equation takes the following abstract form:

$$\left(\partial_t + \mathcal{L}'\right)^* \mathcal{Q}^{-1}\left(\partial_t + \mathcal{L}\right)(u) = 0, \tag{4}$$

where $\mathcal{L}'$ is the first-order derivative of $\mathcal{L}$ so that for any infinitesimal perturbation function δu, $\mathcal{L}\left(u+\delta u\right) \approx \mathcal{L}(u) + \mathcal{L}'(u)\delta u$. Apparently, (4) is not a classical PDE since $\mathcal{Q}$ is defined globally. It was shown in [5, 2] that as $\mathcal{Q}$ converges to an identity operator $\mathcal{I}$, i.e., the kernel of $\mathcal{Q}$ goes to a delta function, the rate functional $I_{\mathcal{Q}}$ Γ-converges to $I_{\mathcal{I}}$. Replacing $\mathcal{Q}$ with $\mathcal{I}$, we see that the E-L equation associated with $I_{\mathcal{I}}$ is a classical PDE. This implies that when the correlation length is small in a proper sense, we can use the minimizer of $I_{\mathcal{I}}$ to approximate the minimizer of $I_{\mathcal{Q}}$ by solving a classical PDE. However, if the correlation length is moderate or large, such a strategy cannot be used, and we need to solve (4) directly.

Another practical issue is about an accurate approximation of $\Pr(u_\varepsilon \in B)$. From the LDP (3), we know that

$$\Pr(u_\varepsilon \in B) \sim C e^{-\frac{I_{\mathcal{Q}}(u^*)}{\varepsilon}}, \tag{5}$$

where C is a prefactor, and u^* is the minimizer of $I_{\mathcal{Q}}$. In general, the large deviation theory only provides an estimate of the probability $\Pr(u_\varepsilon \in B)$ in an asymptotic sense. For a finite ε, the prefactor cannot be ignored, which means that we need more than large deviation theory for the computation of $\Pr(u_\varepsilon \in B)$. An intuitive idea to incorporate the large deviation theory into the estimation of $\Pr(u_\varepsilon \in B)$ is to construct an importance sampling scheme around the neighborhood of u^*, since away from u^* the density decays exponentially. The simplest strategy to do this is to shift the mean of $\sqrt{\varepsilon}G$ from zero to the noise profile corresponding to u^*, which results in an exponential tilting estimator. Although such a straightforward strategy may not work well, as shown in [15], some remedy strategies have been developed in terms of weak efficiency [15, 4], mainly for finite-dimensional cases. For any unbiased estimator Z of $\Pr(u_\varepsilon \in B)$, we have

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{E}[Z^2] \geq \lim_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{E}[Z]^2 = 2\lim_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{E}[Z] = -2I_{\mathcal{Q}}(u^*).$$

When the lower bound provided by the large deviation theory is reached, we say that the estimator Z is asymptotically (or weakly) efficient. Although in general the control (sub-solution) based approach developed in [8, 7] for the dynamical case insightfully provides a connection between the large deviation theory and the weak efficiency of importance sampling, other forms of equivalent conditions for weak efficiency, which is easy to verify in practice, are always expected for specific problems arising from various applications.

In this work, we will explore the aforementioned two issues in an infinite dimensional system: the approximation of (4) for a moderate or large correlation length, and the accurate estimation of $\Pr(u_\varepsilon \in B)$. Before we describe our problem setting, we briefly discuss the related work to these two issues. First of all, the numerical algorithm of approximating the minimizer of the rate functional $I_\mathcal{Q}$ is usually under the name minimum action method (MAM) [9] when nongradient systems are considered. Although many variants of MAM have been developed so far [27, 18, 30, 32, 16, 33], most of these methods deal with the simplified case that $\mathcal{Q} = \mathcal{I}$ to focus on the difficulties from the phase transition in time direction. Second, rare or extreme events have been discussed in [22, 21, 20] for the elliptic model $-\nabla \cdot (a(\mathbf{x}, \omega)\nabla u) = f(\mathbf{x})$, where the force term is deterministic and the coefficient $a(\mathbf{x}, \omega)$ is log-normal. The minimization of the rate functional results in a PDE-constrained optimization problem, where the optimality condition has a different form than the E-L equation (4). Third, the sufficient condition for the weak efficiency of estimator Z is usually considered with respect to a stochastic ODE [8].

To focus on a PDE problem, we will simplify our problem by letting $\partial_t = 0$ and choosing $\mathcal{L} = -\Delta$. This way, the difficulties from small-noise-induced transitions are excluded and the choice of a simple Laplace operator will allow us to obtain more insights from the theoretical point of view. As for the random-event set B, we pick

$$B = \{\|u_\varepsilon\| \geq C_B\} \quad \text{for some constrant } C_B, \tag{6}$$

which consists of the random events that the L^2 norm of u_ε exceeds a certain threshold.

We now outline our work in this paper. First of all, we derive the E-L equation for the problem proposed, which is an eigenvalue problem of operator $\Delta\mathcal{Q}^{-1}\Delta$ subject to a Navier-type boundary conditions. Compared to the SPDE (1), the differentiation order is doubled. If $\mathcal{Q} = \mathcal{I}$, the operator $\Delta\mathcal{Q}^{-1}\Delta$ becomes the biharmonic operator. Second, we replace $\mathcal{Q}$ with $\mathcal{Q}_M$, where $\mathcal{Q}_M$ is a finite-rank approximation of $\mathcal{Q}$ given by the eigenfunctions associated with the largest M eigenvalues of $\mathcal{Q}$. The main question here is the convergence of eigenvalues of $\Delta(\mathcal{Q}_M|_{V_M})^{-1}\Delta$ to those of $\Delta\mathcal{Q}^{-1}\Delta$, where $\mathcal{Q}_M|_{V_M}$ indicates the restriction of $\mathcal{Q}_M$ onto its range V_M such that the inverse exists. We proved that the convergence rate of the approximated eigenvalues of $\Delta(\mathcal{Q}_M|_{V_M})^{-1}\Delta$ is consistent with the decay rate of the eigenvalues of $\mathcal{Q}$ with respect to M. Third, we use the minimizer of the rate functional to construct an exponential tilting estimator to approximate $\Pr(u_\varepsilon \in B)$, where we obtain a sufficient and necessary condition for the weak efficiency. More specifically, we proved that the importance sampling estimator is weakly efficient if and only if the ratio between the second smallest eigenvalue of $\Delta\mathcal{Q}^{-1}\Delta$ and the smallest one is larger than 3, under the assumption that the smallest eigenvalue of $\Delta\mathcal{Q}^{-1}\Delta$ is simple. Finally, we implement numerical experiments to verify our theoretical results and explore some interesting issues such as the relation between the smallest eigenvalue of $\Delta\mathcal{Q}^{-1}\Delta$ and correlation length. We demonstrate numerically that the relative error per sample given by our estimator has an algebraic increase, in contrast to the exponential increase given by a Monte Carlo estimator as $\varepsilon \to 0$.

The recent work [24] used importance sampling to estimate the escape probability from a ball within a certain time for a linear stochastic evolution equation in infinite dimensions, which generalized the the pre-asymptotic analysis in [7]. They also found that in their sub-solution method (after projection onto the leading mode of the covariance operator), a spectral gap condition is required for the efficiency of the importance sampling scheme. By considering a simple static problem, we avoid the difficulties from path sampling such that we are able to derive a sufficient and necessary condition for weak efficiency with respect to any prescribed covariance operator, while in [24] the covariance operator is defined on the eigensapce of the Laplacian for the convenience of analysis. Although we cannot directly compare the two spectral gap conditions yet, we believe that both discoveries show the intrinsic connection between certain spectral gaps and the rare-event conditional distribution. Our work and [24] both confirmed the challenges of efficient rare-event simulations in infinite dimensions, even for simple linear problems.

Our paper is organized as follows. The problem setting is described in section 2. In section 3, we construct an LDP-based importance sampling scheme along with a sufficient and necessary condition for the weak efficiency of the estimator. In section 4, we approximate the Euler–Lagrange equation associated with the rate functional by replacing the covariance operator with its finite-rank approximation. The fully discretized problem is given in section 5. Numerical results have been included in section 6, which is followed by a summary section.

2. Problem setting. We consider the following stochastic elliptic problem on a convex physical domain $D \subset \mathbb{R}^2$ with Lipschitzian boundary ∂D:

$$\begin{cases} -\Delta u_\varepsilon(\mathbf{x}) = \sqrt{\varepsilon} G(\mathbf{x}), & \mathbf{x} \in D, \\ u_\varepsilon(\mathbf{x}) = 0, & \mathbf{x} \in \partial D, \end{cases} \tag{7}$$

where ε is a small positive number, and $G(\mathbf{x})$ is a Gaussian field on the Hilbert space $L^2(D)$. The covariance operator $\mathcal{Q}$ of G is nondegenerate, i.e., the null space $\mathcal{N}(\mathcal{Q}) = \{0\}$. We also assume that $\mathcal{Q}$ is of trace class, positive, and self-adjoint. We consider the probability of the following random event:

$$B = \{u_\varepsilon(\mathbf{x}) | \, \|u_\varepsilon\| \geq C_B\}, \tag{8}$$

where $0 < C_B < \infty$ is constant, and $\|\cdot\|$ indicates the L^2 norm on D.

3. Estimation of $\mathbf{Pr}(u_\varepsilon \in B)$. In this section, we construct an importance sampling scheme based on the results from the theory of large deviations.

3.1. Large deviation principle. Let $G_\varepsilon = \sqrt{\varepsilon} G(\mathbf{x})$. For the scaled Gaussian random field G_ε, we have the following good rate functional [23]:

$$I_{\mathcal{Q}}(\phi) = \frac{1}{2} \left\| \mathcal{Q}^{-1/2} \phi \right\|^2 = \frac{1}{2} \left\langle \mathcal{Q}^{-1/2} \phi, \mathcal{Q}^{-1/2} \phi \right\rangle, \quad \phi \in L^2(D), \tag{9}$$

where we set $I_{\mathcal{Q}}(\phi) = +\infty$ if $\mathcal{Q}^{-1/2}\phi$ cannot be defined, and $\langle \cdot, \cdot \rangle$ indicates the inner product on $L^2(D)$. More precisely, $I_{\mathcal{Q}}(\phi)$ is defined on the Cameron–Martin space $\mathcal{H}_{\mathcal{Q}} = \mathcal{Q}^{1/2} L^2(D)$, i.e., the image of $\mathcal{Q}^{1/2}$ acting on $L^2(D)$. Let $H_0^1(D)$ consist of functions in $H^1(D)$ which vanish on the boundary ∂D. Let $\Delta^{-1} : L^2(D) \mapsto H_0^1(D)$ be the inverse of the Laplacian, which is continuous. According to the contraction principle [6], we know that u_ε has a good rate functional:

$$S(u) = \inf_{\phi \in L^2(D): u = -\Delta^{-1}\phi} \frac{1}{2} \left\| \mathcal{Q}^{-1/2} \phi \right\|^2.$$

When the domain boundary is good enough, e.g., Lipschitzian and convex [10], we have $u_\varepsilon \in H^2(D)$ almost surely. We can then rewrite the above rate functional as

$$S(u) = \begin{cases} \frac{1}{2}\left\|\mathcal{Q}^{-1/2}\Delta u\right\|^2 & \text{if } u \in H_0^1(D) \cap H^2(D) \text{ and } \Delta u \in \mathcal{H}_\mathcal{Q}, \\ \infty & \text{otherwise.} \end{cases} \tag{10}$$

Using the LDP for u_ε, we have an estimate of $\Pr(u_\varepsilon \in B)$ as

$$\lim_{\varepsilon\downarrow 0} \varepsilon \ln \Pr(u_\varepsilon \in B) = -\inf_{u\in B} S(u) = -S(u^*), \tag{11}$$

where u^* is the minimizer of $S(u)$ in B. In other words, when ε is small, we have

$$\Pr(u_\varepsilon \in B) \approx Ce^{-\frac{S(u^*)}{\varepsilon}}, \tag{12}$$

where C is a prefactor. To capture the effect of the prefactor C, we can consider importance sampling (IS) by using the asymptotic result (11) for the change of measure [4]. We then need to address two issues: (1) solving the optimization problem in equation (11), and (2) checking the effectiveness of the LDP-based IS estimator.

Remark 3.1. In general, the attainability of the infimum in (11) on a admissible set is of theoretical importance. For our problem, we expect that u^* is located in the admissible set $\{u | u \in H_0^1(D) \cap H^2(D), u \in B\}$ since $\mathcal{H}_\mathcal{Q}$ is dense in $L^2(D)$ and B is closed. Indeed, this fact can be established by the weakly lower-semicontinuity of $S(u)$ with respect to the norm $\|\mathcal{Q}^{-1/2}\Delta \cdot\|$ [10, 14].

3.2. The eigenvalue problem for minimizing the rate functional. We now address the optimization problem defined by LDP:

THEOREM 3.2. *The minimizers of $S(u)$ in (11) satisfy the following eigenvalue problem:*

$$\Delta\mathcal{Q}^{-1}\Delta u = \lambda u \tag{13}$$

subject to Navier-type boundary conditions, $u|_{\partial D} = (\mathcal{Q}^{-1}\Delta u)|_{\partial D} = 0$. The minimum of the rate functional is $S(u^*) = \frac{1}{2}\lambda_{\min} C_B^2$, where $\lambda_{\min}$ is the minimum eigenvalue of the problem (13), and u^* is the corresponding eigenfunction with $\|u^*\| = C_B$.

Proof. By the definition of $\mathcal{Q}$, we can rewrite the rate functional $S(u)$ as $S(u) = \frac{1}{2}\left\langle \Delta u, \mathcal{Q}^{-1}\Delta u\right\rangle$. It is easy to see that the minimum of $S(u)$ on B must be achieved at the boundary of B. Define the Lagrangian

$$\mathcal{L}(u,\lambda) = \frac{1}{2}\left\langle \mathcal{Q}^{-1}\Delta u, \Delta u\right\rangle - \frac{\lambda}{2}\left(\langle u, u\rangle - C_B^2\right), \tag{14}$$

where $\lambda/2$ is the Lagrange multiplier. The first-order variation of $\mathcal{L}$ in terms of u is

$$\delta\mathcal{L} = \left\langle \Delta\mathcal{Q}^{-1}\Delta u, \delta u\right\rangle + \left\langle \mathcal{Q}^{-1}\Delta u, \partial_\mathbf{n}\delta u\right\rangle_{\partial D} - \left\langle \partial_\mathbf{n}\left(\mathcal{Q}^{-1}\Delta u\right), \delta u\right\rangle_{\partial D} - \lambda\langle u, \delta u\rangle.$$

Letting $\delta u|_{\partial D} = 0$ and $(\mathcal{Q}^{-1}\Delta u)|_{\partial D} = 0$, we obtain (13). From the duality feasibility $\lambda \geq 0$ and the complementarity condition $\lambda\left(\langle u, u\rangle - C_B^2\right) = 0$, we have

$$S(u^*) = \frac{1}{2}\left\langle \mathcal{Q}^{-1}\Delta u^*, \Delta u^*\right\rangle = \frac{1}{2}\left\langle \Delta\mathcal{Q}^{-1}\Delta u^*, u^*\right\rangle = \frac{\lambda}{2}\langle u^*, u^*\rangle \geq \frac{1}{2}\lambda_{\min} C_B^2.$$

To verify that u^* is indeed a local minimizer, we can look at the second-order variation

$$2\delta^2\mathcal{L} = \left\langle \Delta\delta u, \mathcal{Q}^{-1}\Delta\delta u\right\rangle - \lambda\langle \delta u, \delta u\rangle.$$

The sign of $\delta^2\mathcal{L}$ depends on λ. If $\lambda > \lambda_{\min}$, $\delta^2\mathcal{L}$ can be either positive or negative. This implies that only $\lambda_{\min}$ corresponds to a local minimizer u^*. Due to the definition of B, we note that $-u^*$ is also a minimizer. ∎

To deal with the boundary conditions $(\mathcal{Q}^{-1}\Delta u)|_{\partial D} = 0$, we define a new variable $v = \mathcal{Q}^{-1}\Delta u$, and consider a mixed formulation of (13) [3]: Seek $(v, u) \in H_0^1(D) \times H_0^1(D)$, $(v, u) \neq 0$, such that

$$\Delta u = \mathcal{Q}v, \quad \Delta v = \lambda u, \tag{15}$$

subject to the Dirichlet boundary conditions $u|_{\partial D} = v|_{\partial D} = 0$. We include a simple estimate of the lower bound of $\lambda_{\min}$ and will address the numerical approximation of (15) in section 4.

PROPERTY 3.3. *The minimum eigenvalue $\lambda_{\min}$ of problem* (13) *can be bounded from below by*

$$\lambda_{\min} \geq \lambda_{\mathcal{Q},1}^{-1}\hat{\lambda}_{\min}, \tag{16}$$

where $\hat{\lambda}_{\min}$ is the minimum eigenvalue of a biharmonic eigenvalue problem: $\Delta^2 u = \hat{\lambda}u$, subject to Navier boundary conditions: $u|_{\partial D} = \Delta u|_{\partial D} = 0$. If D is a square $[0, L]^2$, $\lambda_{\min} \geq 4\pi^4\lambda_{\mathcal{Q},1}^{-1}/L^4$.

Proof. Note that

$$\lambda_{\min} = \min_{\substack{u \in H^2(D)\cap H_0^1(D),\\ \|u\|=1}} \left\langle \Delta u, \mathcal{Q}^{-1}\Delta u\right\rangle \geq \lambda_{\mathcal{Q},1}^{-1} \min_{\substack{u \in H^2(D)\cap H_0^1(D),\\ \|u\|=1}} \langle \Delta u, \Delta u\rangle.$$

The minimum on the right-hand side can be given by the eigenvalue problem $\Delta^2 u = \hat{\lambda}u$ subject to Navier boundary conditions $u|_{\partial D} = \Delta u|_{\partial D} = 0$. On convex domains, we know that $\hat{\lambda}$ will be just squares of the eigenvalues of the Laplace operator with the homogeneous Dirichlet boundary condition. On a unit square, $\hat{\lambda}_{\min} = 4\pi^4$. ∎

Since the behavior of $\lambda_{\min}$ is closely related to $\lambda_{\mathcal{Q},1}$, we also include a general property of $\lambda_{\mathcal{Q},1}$:

PROPERTY 3.4. *Assume that the covariance kernel of $\mathcal{Q}$ is in the form of*

$$K(\mathbf{x}, \mathbf{y}) = \frac{1}{l_c^d}\rho\left(\frac{\mathbf{x}-\mathbf{y}}{l_c}\right) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$$

for some nonnegative function ρ satisfying $\int_{\mathbb{R}^d}\rho(\mathbf{x})d\mathbf{x} = 1$. Then the maximal eigenvalue of $\mathcal{Q}$ associated with any bounded domain $D \subset \mathbb{R}^d$ for any $l_c > 0$ is strictly less than 1*, i.e.,*

$$\lambda_{\mathcal{Q},1} = \max_{\|v\|=1} \langle v, \mathcal{Q}v\rangle < 1. \tag{17}$$

Proof. For any $v(\mathbf{y}) \in L^2(D)$ with $\|v\| = 1$, we extend it to $\mathbb{R}^d$ by letting $v(\mathbf{y}) = 0$ if $\mathbf{y} \notin D$. Using the Cauchy–Schwarz inequality, we have

$$\begin{aligned}|\mathcal{Q}v| &\leq \int_{\mathbb{R}^d} K(\mathbf{x},\mathbf{y})|v|(\mathbf{y})d\mathbf{y} = \frac{1}{l_c^d}\int_{\mathbb{R}^d}\rho\left(\frac{\mathbf{x}-\mathbf{y}}{l_c}\right)|v|(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d}\rho(\mathbf{z})|v|(\mathbf{x}-l_c\mathbf{z})d\mathbf{z}\\ &< \left(\int_{\mathbb{R}^d}\rho(\mathbf{z})d\mathbf{z}\right)^{1/2}\left(\int_{\mathbb{R}^d}\rho(\mathbf{z})v^2(\mathbf{x}-l_c\mathbf{z})d\mathbf{z}\right)^{1/2} = \left(\int_{\mathbb{R}^d}\rho(\mathbf{z})v^2(\mathbf{x}-l_c\mathbf{z})d\mathbf{z}\right)^{1/2},\end{aligned}$$

where the strict inequality is due to the fact that the domain D is finite. Then

$$\|\mathcal{Q}v\|^2 < \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \rho(\mathbf{z}) v^2(\mathbf{x} - l_c\mathbf{z}) d\mathbf{z} d\mathbf{x} = \int_{\mathbb{R}^d} \rho(\mathbf{z}) \int_{\mathbb{R}^d} v^2(\mathbf{x} - l_c\mathbf{z}) d\mathbf{x} d\mathbf{z} = \|v\|^2.$$

We then have $\langle v, \mathcal{Q}v\rangle^2 < \|v\|\|\mathcal{Q}v\| \le \|v\|^2$, which yields the conclusion. ◻

Remark 3.5. As we mentioned earlier, an important idealized case is white noise, which corresponds to $\mathcal{Q}$ being an identity operator $\mathcal{I}$; then (13) becomes the eigenvalue problem for the biharmonic operator Δ^2 [17]. Note that $\mathcal{I}$ is not of trace class. When $\mathcal{Q}$ is of trace class with a finite correlation length, one can show that as $\mathcal{Q}$ goes to $\mathcal{I}$, i.e., the correlation length tends to zero, the rate functional $\|\mathcal{Q}^{-1/2}\Delta u\|^2$ Γ-converges to $\|\Delta u\|^2$ [5, 2], where the minimizers of the rate functional converge correspondingly. In the other extreme case, when the correlation length tends to infinity, the random function $G(\mathbf{x})$ tends to a random constant, and the kernel of $\mathcal{Q}$ is simply a constant. In this paper, we are mainly interested in a moderate correlation length in terms of the characteristic length of the domain D.

3.3. The importance sampling problem. Let

$$\gamma := \Pr(u_\varepsilon \in B) = \mathbb{E}\left[1_B(G_\varepsilon)\right] = \int 1_B(\psi)\mathbb{P}_0(d\psi), \tag{18}$$

where 1_B is the indicator function, the Gaussian measure $\mathbb{P}_0$ is the law of G_ε with zero mean and covariance operator εQ. The Monte Carlo estimator is defined as

$$\hat{P}_{\mathsf{MC}} = \frac{1}{N_{\mathsf{MC}}} \sum_{i=1}^{N_{\mathsf{MC}}} 1_B\left(G_\varepsilon^{(i)}\right), \tag{19}$$

where N_{MC} is the number of samples and the superscript of $G_\varepsilon^{(i)}$ indicates the index of the samples. The *relative error* of this estimator is

$$\frac{\mathrm{Var}^{1/2}\left(\hat{P}_{\mathsf{MC}}\right)}{\gamma} = \frac{1}{\sqrt{N_{\mathsf{MC}}}}\left(\frac{1-\gamma}{\gamma}\right)^{1/2} \sim (\gamma N_{\mathsf{MC}})^{-1/2}. \tag{20}$$

If the relative error is $O(1)$, we have $N_{\mathsf{MC}} \sim \gamma^{-1}$. Since γ decreases exponentially as $\varepsilon \to 0$, N_{MC} must increase exponentially, which makes the brute-force Monte Carlo method prohibitively expensive.

To reduce the variance, one typical method is the importance sampling (IS). We look for a new Gaussian measure $\mathbb{P}_\phi$, where the covariance operator εQ remains unchanged but the mean is shifted to a function ϕ. This results in an IS estimator of exponential tilting type

$$\hat{P}_{\mathsf{IS}} = \frac{1}{N_{\mathsf{MC}}} \sum_{i=1}^{N_{\mathsf{MC}}} Z_\phi\left(\tilde{G}_\varepsilon^{(i)}\right) \quad \text{with } Z_\phi(\cdot) = 1_B(\cdot)\frac{d\mathbb{P}_0}{d\mathbb{P}_\phi}(\cdot), \tag{21}$$

where $\tilde{G}_\varepsilon^{(i)}$ are generated according to $\mathbb{P}_\phi$ and $d\mathbb{P}_0/d\mathbb{P}_\phi$ is the Radon–Nikodym derivative:

$$\frac{d\mathbb{P}_0}{d\mathbb{P}_\phi}(\psi) = \exp\left\langle -\frac{1}{\varepsilon}\left\langle \phi, Q^{-1}\psi\right\rangle + \frac{1}{2\varepsilon}\left\langle \phi, Q^{-1}\phi\right\rangle\right\rangle. \tag{22}$$

To find a good alternative measure $\mathbb{P}_\phi$, we need to minimize the variance of the unbiased estimator Z_ϕ. Since $\mathbb{E}_\phi\left[Z_\phi^2\right] \geq \mathbb{E}_\phi\left[Z_\phi\right]^2 = \mathbb{E}\left[1_B\right]^2$, after taking logarithm on both sides, we have from the LDP that

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{E}_\phi\left[Z_\phi^2\right] \geq 2 \lim_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{E}[1_B] = -2S(u^*). \tag{23}$$

If the lower bound in (23) is reached, we say the estimator Z_ϕ is *asymptotically (or weakly) efficient*. The LDP suggests that one could use the "optimal" one ϕ^* given by the minimizer u^* to construct $\mathbb{P}_\phi$. However, many recent works show that this choice can be dangerous (especially when S has multiple local minimizers), since in some cases (22) may lead to an infinite variance as $\varepsilon \to 0$ (cf. [15]).

We subsequently present an IS estimator for $\Pr(u_\varepsilon \in B)$ and rigorously prove its weak efficiency under some appropriate conditions. Due to symmetry, we split B into two disjoint sets $B = B_+ \cup B_-$ and deal with the two subsets separately, where

$$B_+ = \{u_\varepsilon(\mathbf{x}) | u_\varepsilon \in B, \langle u_\varepsilon, u^*\rangle \geq 0\} \quad \text{and} \quad B_- = \{u_\varepsilon(\mathbf{x}) | u_\varepsilon \in B, \langle u_\varepsilon, u^*\rangle \leq 0\}. \tag{24}$$

We assume that the eigenvalues of the problem (13) are denoted by

$$\lambda_{\min} = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots,$$

with the smallest eigenvalue $\lambda_{\min}$ being simple. Then within B_+, u^* is the unique minimizer of $S(u)$, and within B_-, $-u^*$ is the unique minimizer. Using $\phi^* = \pm\Delta u$ and (21), we define the following two LDP-based IS estimators for $\Pr(u_\varepsilon \in B_+)$ and $\Pr(u_\varepsilon \in B_-)$, respectively:

$$Z_{\phi^*} = 1_{B_+} \frac{d\mathbb{P}_0}{d\mathbb{P}_{\phi^* = -\Delta u^*}} \quad \text{and} \quad Z_{-\phi^*} = 1_{B_-} \frac{d\mathbb{P}_0}{d\mathbb{P}_{\phi^* = \Delta u^*}}. \tag{25}$$

The estimator for $\Pr(u_\varepsilon \in B)$ is then defined as [15]

$$Z_{\phi^*} + Z_{-\phi^*}, \tag{26}$$

where Z_{ϕ^*} and $Z_{-\phi^*}$ are sampled independently.

Due to symmetry, we only need to focus on Z_{ϕ^*} for $\Pr(u_\varepsilon \in B_+)$. The analysis for $\Pr(B_-)$ is exactly the same. A sufficient and necessary condition for the weak efficiency of Z_{ϕ^*} is given in the following theorem:

THEOREM 3.6. *Let $\phi^* = -\Delta u^*$. Assume that the smallest eigenvalue of equation* (13) *is simple. Then Z_{ϕ^*} is asymptotically efficient for estimating* $\Pr(u_\varepsilon \in B_+)$ *if and only if $\lambda_2 \geq 3\lambda_{\min}$, where λ_2 is the second smallest eigenvalue of* (13).

The proof of this theorem is split into the following two lemmas.

LEMMA 3.7. *The following two conditions are equivalent:*

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{E}_{\phi^*}\left[Z_{\phi^*}^2\right] = -2S(u^*) \Leftrightarrow \lim_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{E}_{-\phi^*}\left[1_{B_+}\right] = -4S(u^*).$$

Proof. Using (22), we have

$$\begin{aligned}
\mathbb{E}_{\phi^*}\left[Z_{\phi^*}^2\right] &= \int_{B_+} Z_{\phi^*}^2(\psi)\, \mathbb{P}_{\phi^*}(d\psi) = \int_{B_+} \frac{d\mathbb{P}_0}{d\mathbb{P}_{\phi^*}}\, \mathbb{P}_0(d\psi) \\
&= \int_{B_+} \frac{d\mathbb{P}_0}{d\mathbb{P}_{\phi^*}} \frac{d\mathbb{P}_0}{d\mathbb{P}_{-\phi^*}}\, \mathbb{P}_{-\phi^*}(d\psi) = \int_{B_+} \exp\left(\frac{1}{\varepsilon}\left\langle \phi^*, \mathcal{Q}^{-1}\phi^*\right\rangle\right) \mathbb{P}_{-\phi^*}(d\psi) \\
&= \exp\left(\frac{1}{\varepsilon}\left\langle \Delta u^*, \mathcal{Q}^{-1}\Delta u^*\right\rangle\right) \mathbb{E}_{-\phi^*}\left[1_{B_+}\right],
\end{aligned}$$

which yields that

$$\lim_{\varepsilon\downarrow 0}\varepsilon\log\mathbb{E}_{\phi^*}\left[Z_{\phi^*}^2\right]=2S(u^*)+\lim_{\varepsilon\downarrow 0}\varepsilon\log\mathbb{E}_{-\phi^*}\left[1_{B_+}\right].$$

By the definition of weak efficiency, we reach the conclusion. ∎

Another lemma is needed for the asymptotic estimate of $\mathbb{E}_{-\phi^*}[1_{B_+}]$:

LEMMA 3.8. *We have that*

$$\min_{u\in B_+}\left[\hat{S}(u)=\frac{1}{2}\left\langle\Delta(u+u^*),\mathcal{Q}^{-1}\Delta(u+u^*)\right\rangle\right]=4S(u^*) \tag{27}$$

if and only if $\lambda_2\geq 3\lambda_{\min}$.

Proof. We introduce Lagrangian multipliers $\lambda_a/2$ and λ_b, respectively, for the two inequality constraints of B_+, i.e., $\|u\|\geq C_B$ and $\langle u,u^*\rangle\geq 0$, and consider the Lagrangian

$$\mathcal{L}(u)=\frac{1}{2}\left\langle\Delta(u+u^*),\mathcal{Q}^{-1}\Delta(u+u^*)\right\rangle-\frac{\lambda_a}{2}(\langle u,u\rangle-C_B^2)-\lambda_b\langle u,u^*\rangle,$$

whose Euler–Lagrange equation is

$$\Delta\mathcal{Q}^{-1}\Delta(u+u^*)-\lambda_a u-\lambda_b u^*=0, \tag{28}$$

with the same boundary conditions as those for (13). Then the Karush–Kuhn–Tucker (KKT) conditions for optimality include the first-order condition (28), dual feasibility $\lambda_a\geq 0$, $\lambda_b\geq 0$, and the complementary slackness $\lambda_a(\|u\|-C_B)=\lambda_b\langle u,u^*\rangle=0$.

$\hat{S}$ is quadratic in terms of u. The optimal point does not lie in the interior of B_+, but on its boundary. So either $\lambda_a\neq 0$ if the optimal solution satisfies $\|u\|=C_b$ or $\lambda_b\neq 0$ if the optimal solution satisfies $\langle u,u^*\rangle=0$. It is not possible that both are zeros. Before we look into the KKT conditions, we simplify $\hat{S}(u)$ first by writing

$$u=qu^*+\tilde{u},$$

where $\tilde{u}$ satisfies $\langle\tilde{u},u^*\rangle=0$. Since $u\in B_+$, then $q\geq 0$. The E-L equation (28) can be rewritten as

$$\Delta\mathcal{Q}^{-1}\Delta\tilde{u}-\lambda_a\tilde{u}+(\lambda_{\min}(q+1)-\lambda_a q-\lambda_b)u^*=0. \tag{29}$$

It is known that the orthogonal conditions holds for two eigenfunctions of (13) corresponding to distinct eigenvalues. Taking the inner product of (29) with u^* and using the assumption that $\lambda_{\min}$ is simple, after applying integration by parts for the first term, we have

$$(\lambda_{\min}(q+1)-\lambda_a q-\lambda_b)\langle u^*,u^*\rangle=0,$$

which implies that the E-L equation (28) is equivalent to the system

$$\begin{cases}\Delta\mathcal{Q}^{-1}\Delta\tilde{u}-\lambda_a\tilde{u}=0,\\ \lambda_{\min}(q+1)-\lambda_a q-\lambda_b=0,\end{cases} \tag{30}$$

where $\tilde{u}$ satisfies the Navier-type boundary conditions as u and u^*. So if $\tilde{u}$ is nonzero, it is also one of the eigenfunctions of $\Delta\mathcal{Q}^{-1}\Delta$ and its eigenvalue λ_a is strictly larger than the smallest one $\lambda_{\min}$ (i.e., $\lambda_a\geq\lambda_2$) since $\lambda_{\min}$ is simple. In addition, if $\lambda_a=0$, $\tilde{u}=0$ is the unique solution of $\Delta\mathcal{Q}^{-1}\Delta\tilde{u}=0$.

For the constraints, we have by definition

$$\begin{aligned}\langle u,u\rangle &= q^2\langle u^*,u^*\rangle + \langle\tilde u,\tilde u\rangle = q^2C_B^2 + \langle\tilde u,\tilde u\rangle\,,\\ \langle u,u^*\rangle &= q\langle u^*,u^*\rangle = qC_B^2.\end{aligned}\tag{31}$$

$u\in B_+$ if and only if $q\ge 0$ and $\langle\tilde u,\tilde u\rangle\ge(1-q^2)C_B^2$. The minimization over u is equivalent to the minimization over q and $\tilde u$. The complementary slackness condition becomes

$$(\langle\tilde u,\tilde u\rangle-(1-q^2)C_B^2)\lambda_a = q\lambda_b = 0.$$

The functional $\hat S$ can be rewritten as

$$\begin{aligned}\hat S(u) &= \frac12\left\langle\Delta(\tilde u+(q+1)u^*),\mathcal{Q}^{-1}\Delta(\tilde u+(q+1)u^*)\right\rangle\\ &= \frac12\left(\lambda_a\langle\tilde u,\tilde u\rangle+\lambda_{\min}(q+1)^2C_B^2\right).\end{aligned}\tag{32}$$

We now start to analyze two possible cases when the KKT conditions hold.

Case I: $\lambda_b=0$. This means $\langle u,u^*\rangle>0$ and the constraint $\|u\|=C_B$ has to be active ($\lambda_a>0$). By (31), we know q is strictly positive. From the second equation of (30) and $\lambda_b=0$ in this case, we obtain that

$$\lambda_a=\lambda_{\min}(1+1/q).$$

Then additionally with (32) and noting that $\langle\tilde u,\tilde u\rangle$ is equal to $(1-q^2)C_B^2$, we have

$$\begin{aligned}\hat S(u) &= \frac12\left(\lambda_{\min}(1+1/q)(1-q^2)C_B^2+\lambda_{\min}(q+1)^2C_B^2\right)\\ &= \frac{2+q+1/q}{2}\lambda_{\min}C_B^2\ge 2\lambda_{\min}C_B^2 = 4S(u^*).\end{aligned}$$

The equality holds at the optimal $q=1$, which implies $\lambda_a=2\lambda_{\min}$ and $\tilde u=0$. So, the minimizer of $\hat S$ is also u^*, the minimizer of S.

Case II: $\lambda_b\ne 0$. The constraint $\langle u,u^*\rangle=0$ is active now, and $q=0$. Then $u=\tilde u$. Notice that $\lambda_a\ne 0$ is also true in this case; otherwise, $\tilde u=0$ by (30) and it follows that u is zero too, which does not satisfy the constraint $\|u\|\ge C_B$. Nonzero λ_a then means $\|u\|=C_B$. Then

$$\begin{aligned}\hat S(u) &= \frac12(\lambda_a\langle\tilde u,\tilde u\rangle+\lambda_{\min}C_B^2) = \frac12(\lambda_a\langle u,u\rangle+\lambda_{\min}C_B^2)\\ &= \frac12 C_B^2(\lambda_{\min}+\lambda_a)\ge\frac12 C_B^2(\lambda_{\min}+\lambda_2),\end{aligned}$$

where we used the fact that the smallest possible value of λ_a is λ_2. In this case, the optimal u for $\hat S$ is the eigenfunction of $\Delta\mathcal{Q}^{-1}\Delta$ corresponding to the second smallest eigenvalue λ_2.

Now combining cases I and II, we know that the true minimum is eventually determined by the competition of the above two cases as follows:

$$\min_{u\in B_+}\hat S(u)=\min\left\{4S(u^*)=2\lambda_{\min}C_B^2,\ \frac12 C_B^2(\lambda_{\min}+\lambda_2)\right\}.$$

Therefore, the condition (27) holds if and only if

$$\frac12 C_B^2(\lambda_{\min}+\lambda_2)\ge 2\lambda_{\min}C_B^2,$$

which is equivalent to $\lambda_2\ge 3\lambda_{\min}$. ∎

We are now ready to prove Theorem 3.6:

Proof. First of all, Z_{ϕ^*} is unbiased:

$$\mathbb{E}_{\phi^*}\left[Z_{\phi^*}\right] = \int_{B_+} 1_{B_+} \frac{d\mathbb{P}_0}{d\mathbb{P}_{\phi^*}} \mathbb{P}_{\phi^*}(d\psi) = \int_{B_+} 1_{B_+} d\mathbb{P}_0(d\psi) = \mathbb{E}\left[1_{B_+}\right].$$

From Lemma 3.7, we need to estimate $\mathbb{E}_{-\phi^*}[1_{B_+}]$, which can be done by the large deviation principle. Shifting the mean of G_ε to $-\phi^*$, we have

$$-\Delta\hat{u}_\varepsilon = -\phi^* + G_\varepsilon \Rightarrow -\Delta(\hat{u}_\varepsilon + u^*) = G_\varepsilon.$$

From the contraction principle, we know that $\hat{u}_\varepsilon$ has a good rate functional $\hat{S}(u)$ defined in (27). The conclusion follows Lemma 3.7 and 3.8. ∎

Remark 3.9. In many cases where the simple exponential tilting scheme fails, the reason usually is the existence of multiple local minimizers of the rate function (see [8] for the first work to solve this issue based on the control theory). Here we point out one simple case in which, although the minimizer on B_+ is unique, one may still run into the risk of losing the weak efficiency, which is consistent with the observation in [15] for a finite-dimensional case. One sufficient condition given in Chapter 5.2 of [4] to obtain weak efficiency is the so called *dominating point*, which requires that the set is completely located on one side of a hyperplane tangent to the level set of rate function at the minimizer. Obviously, this sufficient condition is too strong for the set B_+ here. From the proof of Lemma 3.8, we have if $\lambda_2 < 3\lambda_{\min}$ that the relative error per sample of the IS estimator is

$$\frac{\mathrm{Var}^{1/2}\left(\hat{P}_{\mathsf{IS}}\right)}{\gamma}\sqrt{N_{\mathsf{MC}}} = \left(\frac{\mathbb{E}_{\phi^*}\left[Z_{\phi^*}^2\right] - \gamma^2}{\gamma^2}\right)^{1/2} \approx \frac{\mathbb{E}_{\phi^*}^{1/2}\left[Z_{\phi^*}^2\right]}{\gamma} \sim e^{\frac{3\lambda_{\min}-\lambda_2}{2\lambda_{\min}\varepsilon}S(u^*)},$$

which increases exponentially as $\varepsilon \to 0$.

4. Approximate $\mathcal{Q}$ by a finite-rank approximation $\mathcal{Q}_M$. When estimating $\Pr(u_\varepsilon \in B)$, the Gaussian field $G_\varepsilon(\mathbf{x})$ needs to be sampled, implying that $\mathcal{Q}$ must be approximated. In this work, we employ the Karhunen–Loevé (K-L) expansion to approximate G:

$$G(\mathbf{x}) \approx G_M(\mathbf{x}) = \sum_{i=1}^{M} \sqrt{\lambda_{\mathcal{Q},i}}\, e_{\mathcal{Q},i}(\mathbf{x})\xi_i, \tag{33}$$

where ξ_i are i.i.d. normal random variables, and $\{(\lambda_{\mathcal{Q},i}, e_{\mathcal{Q},i}(\mathbf{x}))\}_{i=1}^{\infty}$ the eigenpairs of $\mathcal{Q}$ with $\lambda_{\mathcal{Q},1} \geq \lambda_{\mathcal{Q},2} \geq \cdots > 0$. The K-L expansion is given by an eigenvalue problem associated with the kernel K of $\mathcal{Q}$:

$$\mathcal{Q}v = \int_D K(\mathbf{x},\mathbf{y})v(\mathbf{y})d\mathbf{y} = \lambda_{\mathcal{Q}} v(\mathbf{x}), \tag{34}$$

where $K(\mathbf{x},\mathbf{y}) = \mathbb{E}\left[G(\mathbf{x})G(\mathbf{y})\right]$ is the covariance matrix [19]. The finite-rank approximation of $\mathcal{Q}$ is

$$\mathcal{Q} \approx \mathcal{Q}_M = \sum_{i=1}^{M} \lambda_{\mathcal{Q},i} \left\langle \cdot, e_{\mathcal{Q},i}\right\rangle e_{\mathcal{Q},i}. \tag{35}$$

The variational formulation of (15) is to find $(v,u) \in H_0^1(D) \times H_0^1(D)$, $(v,u) \neq (0,0)$, such that

$$\begin{cases} \langle \mathcal{Q}v, w\rangle + \langle \nabla w, \nabla u\rangle = 0 & \forall w \in H_0^1(D), \\ \langle \nabla v, \nabla \hat{w}\rangle = -\lambda \langle u, \hat{w}\rangle & \forall \hat{w} \in H_0^1(D). \end{cases} \tag{36}$$

Replacing $\mathcal{Q}$ with $\mathcal{Q}_M$, we have the semidiscrete version of (36): Find $(\tilde{v},\tilde{u}) \in H_0^1(D)\times H_0^1(D)$, $(\tilde{v},\tilde{u}) \neq (0,0)$, such that

$$\begin{cases} \langle \mathcal{Q}_M\tilde{v}, w\rangle + \langle \nabla w, \nabla \tilde{u}\rangle = 0 & \forall w \in H_0^1(D), \\ \langle \nabla \tilde{v}, \nabla \hat{w}\rangle = -\tilde{\lambda}_M \langle \tilde{u}, \hat{w}\rangle & \forall \hat{w} \in H_0^1(D). \end{cases} \tag{37}$$

LEMMA 4.1. *Problem* (37) *has* M *eigenvalues, which are all positive.*

Proof. We look at the problem from the point of view of spectral approximation theory. We start from the source problem associated with (37): For $g \in L^2(D)$, find $(\tilde{v},\tilde{u}) \in H_0^1(D) \times H_0^1(D)$ such that

$$\begin{cases} A_M(\tilde{v}, w) + B(w,\tilde{u}) = 0 & \forall w \in H_0^1(D), \\ B(\tilde{v},\hat{w}) = -\langle g, \hat{w}\rangle & \forall \hat{w} \in H_0^1(D), \end{cases} \tag{38}$$

where we define two continuous symmetric bilinear forms, $A_M(v,w) = \langle \mathcal{Q}_M v, w\rangle$ and $B(v,w) = \langle \nabla v, \nabla w\rangle$. It is seen that (38) is uniquely solvable for any $g \in L^2(D)$ because $\tilde{v}$ and $\tilde{u}$ are solutions of two decoupled elliptic equations subject to homogeneous Dirichlet boundary conditions. We then introduce the component solution operator $\mathcal{S} : L^2(D) \to H_0^1(D)$, $\mathcal{S}g = \tilde{v}$ such that $B(\mathcal{S}g, \hat{w}) = -\langle g, \hat{w}\rangle$ for any $\hat{w} \in H_0^1(D)$, and the solution operator $\mathcal{T}_M : L^2(D) \to L^2(D)$, $\mathcal{T}_M g = \tilde{u}$, where

$$|\mathcal{S}g|_{H_0^1(D)} \leq C_D\|g\|,$$

with C_D being the Poincaré constant, i.e., $\mathcal{S}$ is bounded.

First of all, all eigenvalues of (37) are positive. Letting $w = \tilde{v}$ and $\hat{w} = \tilde{u}$ in (37), then we have

$$A_M(\tilde{v},\tilde{v}) = \langle \mathcal{Q}_M\tilde{v}, \tilde{v}\rangle = \tilde{\lambda}_M \langle \tilde{u}, \tilde{u}\rangle .$$

Since $A_M(\tilde{v},\tilde{v}) \geq 0$, we have $\tilde{\lambda}_M \geq 0$. We need to pay attention to the case that $A_M(\tilde{v},\tilde{v}) = 0$, since $\mathcal{Q}_M$ is of finite rank. Let $V_M = \operatorname{span}\{e_{\mathcal{Q},i}\}_{i=1}^M$ and write $\tilde{v} = \mathcal{P}_M\tilde{v} + (\mathcal{I} - \mathcal{P}_M)\tilde{v}$, where $\mathcal{P}_M$ indicates a projection onto V_M. We then have that

$$A_M(\tilde{v},\tilde{v}) = A_M(\mathcal{P}_M\tilde{v}, \mathcal{P}_M\tilde{v}) = 0,$$

implying that $\mathcal{P}_M\tilde{v} = 0$, i.e., $\tilde{v} \perp V_M$ and $\mathcal{Q}_M\tilde{v} = 0$. However, if $\mathcal{Q}_M\tilde{v} = 0$, then by the first equation of (37) we have $\tilde{u} = 0$ since $\Delta\tilde{u} = 0$ and $\tilde{u} \in H_0^1(D)$, which further implies that $\tilde{v} = 0$ by the second equation of (37). Thus, the condition $A_M(\tilde{v},\tilde{v}) = 0$ will lead to the contradiction that $(\tilde{u},\tilde{v}) = 0$. Consequently, we proved that $A_M(\tilde{v},\tilde{v}) > 0$ for any nonzero $\tilde{v}$, i.e., $\tilde{\lambda}_M > 0$.

Second, the eigenpairs of (37) can be characterized by $\mathcal{T}_M$. We can think of the eigenvalue problem (37) as a source problem with $g = \lambda_M\tilde{u}$. In other words, if $(\tilde{\lambda}_M, (\tilde{v},\tilde{u}))$ is an eigenpair of (37), then $\mathcal{T}_M(\tilde{\lambda}_M\tilde{u}) = \tilde{u}$. On the other hand, the source problem with $\mathcal{T}_M(\tilde{\lambda}_M\tilde{u}) = \tilde{u}$ and $\tilde{u} \neq 0$ corresponds to an eigenvalue problem, since there exists a $\mathcal{S}(\tilde{\lambda}_M\tilde{u}) = \tilde{v} \in H_0^1(D)$ such that $(\tilde{\lambda}_M, (\tilde{v},\tilde{u}))$ is an eigenpair of problem (37). Hence, $\tilde{\lambda}_M$ is an eigenvalue of (37) if and only if $\tilde{\lambda}_M^{-1}$ is an eigenvalue of $\mathcal{T}_M$. We can then focus on the property of $\mathcal{T}_M$.

Third, $\mathcal{T}_M$ is self-adjoint and nonnegative. Let $f \in L^2(D)$. We let $\hat{w} = \mathcal{T}_M f$ in the second equation of (38), and have

$$B(\mathcal{S}g, \mathcal{T}_M f) = -\langle g, \mathcal{T}_M f\rangle .$$

We let $\tilde{u} = \mathcal{T}_M f$, $\tilde{v} = \mathcal{S}f$, and $w = \mathcal{S}g$ in the first equation of (38), and have

$$A(\mathcal{S}f, \mathcal{S}g) + B(\mathcal{S}g, \mathcal{T}_M f) = 0.$$

The above two equations yield

$$\langle g, \mathcal{T}_M f\rangle = A_M(\mathcal{S}f, \mathcal{S}g) \quad \forall f, g \in L^2(D).$$

Since A_M is symmetric, we have

$$\langle g, \mathcal{T}_M f\rangle = A_M(\mathcal{S}f, \mathcal{S}g) = A_M(\mathcal{S}g, \mathcal{S}f) = \langle f, \mathcal{T}_M g\rangle .$$

When $f = g$, $\langle g, \mathcal{T}_M g\rangle = A_M(\mathcal{S}g, \mathcal{S}g) \geq 0$. Thus $\mathcal{T}_M$ is self-adjoint and nonnegative.

Fourth, $\mathcal{T}_M$ is of finite rank. For each $e_{\mathcal{Q},i} \in V_M$, $i = 1, \ldots, M$, there exists a unique $u_{Q,i}$ such that

$$A_M(e_{\mathcal{Q},i}, w) + B(w, u_{\mathcal{Q},i}) = 0 \quad \forall w \in H_0^1(D).$$

It is easy to see that $u_{\mathcal{Q},i}$ are linearly independent, where $i = 1, \ldots, M$. Then for any $g \in L^2(D)$, we have $\mathcal{T}_M g = \sum_{i=1}^{M} \langle \mathcal{Q}_M \mathcal{S}g, e_{\mathcal{Q},i}\rangle\, u_{\mathcal{Q},i}$. We only need to make sure that, for each $e_{\mathcal{Q},i}$, there exists g such that $\langle \mathcal{Q}_M \mathcal{S}g, e_{\mathcal{Q},i}\rangle \neq 0$. We pick $g = e_{\mathcal{Q},i}$, and let $\hat{w} = \mathcal{S}e_{\mathcal{Q},i}$. We have

$$\langle e_{\mathcal{Q},i}, \mathcal{Q}_M \mathcal{S}e_{\mathcal{Q},i}\rangle = \langle \mathcal{Q}_M e_{\mathcal{Q},i}, \mathcal{S}e_{\mathcal{Q},i}\rangle = \lambda_{\mathcal{Q},i} \langle e_{\mathcal{Q},i}, \mathcal{S}e_{\mathcal{Q},i}\rangle = -B(\mathcal{S}e_{\mathcal{Q},i}, \mathcal{S}e_{\mathcal{Q},i}) \leq 0.$$

If $B(\mathcal{S}e_{\mathcal{Q},i}, \mathcal{S}e_{\mathcal{Q},i}) = 0$, we have $\mathcal{S}e_{\mathcal{Q},i} = 0$ since $\mathcal{S}e_{\mathcal{Q},i} \in H_0^1(D)$. This is possible only when $e_{\mathcal{Q},i} = 0$.

Overall, $\mathcal{T}_M$ is self-adjoint, nonnegative, and of rank M. Thus, $\mathcal{T}_M$ has M positive eigenvalues $\tilde{\lambda}_M^{-1}$, implying that (37) has M positive eigenvalues $\tilde{\lambda}_M$. ∎

We now establish the convergence of $\tilde{\lambda}_M$ to λ.

Lemma 4.2. *There is a constant C such that the eigenvalues in* (36) *associated with $\mathcal{Q}$ and in* (37) *associated with $\mathcal{Q}_M$ satisfy*

$$|\lambda^{-1} - \tilde{\lambda}_M^{-1}| \leq C\lambda_{\mathcal{Q},M+1}^2, \tag{39}$$

where $\lambda_{\mathcal{Q},M+1}$ is the $(M+1)$th smallest eigenvalue of $\mathcal{Q}$.

Proof. This lemma is an application of Theorem 11.1 in [1]. We include the associated source problems of (36): For $g \in L^2(D)$, find $(v, u) \in H_0^1(D) \times H_0^1(D)$, such that

$$\begin{cases} A(v, w) + B(w, u) = 0 & \forall w \in H_0^1(D), \\ B(v, \hat{w}) = -\langle g, \hat{w}\rangle & \forall \hat{w} \in H_0^1(D), \end{cases} \tag{40}$$

where we define a new continuous bilinear forms $A(v, w) = \langle \mathcal{Q}v, w\rangle$, and B is the same as in the proof of previous lemma. We also need a new solution operator $\mathcal{T} : L^2(D) \to L^2(D)$ such that $\mathcal{T}g = u$, for any $g \in L^2(D)$. The component solution operator $\mathcal{S}g = v$ will be shared by (38) and (40).

To this end, we can consider the approximation of $\mathcal{T}$ given by $\mathcal{T}_M$. If $\|\mathcal{T}-\mathcal{T}_M\| \to 0$ as $M \to \infty$, we can use Theorem 11.1 in [1] to establish the convergence of $\tilde{\lambda}_M$ to λ. For any $g \in L^2(D)$, we have

$$A(\mathcal{S}g, w) + B(w, \mathcal{T}g) = 0 \quad \forall w \in H_0^1(D)$$

from the first equation of (40), and

$$A_M(\mathcal{S}g, w) + B(w, \mathcal{T}_M g) = 0 \quad \forall w \in H_0^1(D)$$

from the first equation of (38). The difference of the above two equations satisfies

$$\langle (\mathcal{Q} - \mathcal{Q}_M)\mathcal{S}g, w\rangle + \langle \nabla w, \nabla(\mathcal{T} - \mathcal{T}_M)g\rangle = 0 \quad \forall w \in H_0^1(D).$$

Letting $w = (\mathcal{T} - \mathcal{T}_M)g$, we can obtain

$$\|(\mathcal{T} - \mathcal{T}_M)g\| \le C_D|(\mathcal{T} - \mathcal{T}_M)g|_{H_0^1(D)} \le C_D\|(\mathcal{Q} - \mathcal{Q}_M)Sg\| \le C_D\|\mathcal{Q} - \mathcal{Q}_M\|\|\mathcal{S}\|\|g\|.$$

We then have

$$\|\mathcal{T} - \mathcal{T}_M\| \le C_D\lambda_{\mathcal{Q},M+1}\|\mathcal{S}\|,$$

where $\|\mathcal{S}\|$ is bounded. The conclusion is reached by applying Theorem 11.1 in [1]. ∎

5. Numerical discretization. For numerical experiments, we will consider the one-dimensional (1D) problem defined on $\Gamma = [-1, 1]$ and the two-dimensional problem defined on $D = \Gamma^2$. We choose the spectral method for spatial discretization in view of the simple geometry of D. In particular, we pick the following one-dimensional basis functions [25]:

$$\phi_i(x) = (L_i(x) - L_{i+2}(x))/\sqrt{4i + 6} \in H_0^1(\Gamma), \tag{41}$$

where L_k is the Legendre polynomial of degree k. Let $W_N = \mathrm{span}\{\phi_i(x)\}_{i=0}^{N-1}$. The two-dimensional approximation space is constructed by tensor product: $V_N = W_N \otimes W_N = \mathrm{span}\{\theta_{i(k,l)}(\mathbf{x}) = \phi_k(x)\phi_l(y)\}_{i=1}^{N^2} \subset H_0^1(D)$, where $i(k, l)$ indicates the global index corresponding to k and l. The fully discretized version of (37) takes the following matrix form:

$$\begin{pmatrix} \mathbf{K}^\top & \mathbf{M}_{\mathcal{Q}_M} \\ 0 & -\mathbf{K} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \tilde{\lambda}_M \begin{pmatrix} 0 & 0 \\ \mathbf{M} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}, \tag{42}$$

where the vectors $\mathbf{u}$ and $\mathbf{v}$ consist of unknown coefficients of the expansions of u and v in V_N, $\mathbf{K}$ is the matrix $\langle \nabla\theta_i, \nabla\theta_j\rangle$, $\mathbf{M}_{\mathcal{Q}_M}$ is $\langle \mathcal{Q}_M\theta_i, \theta_j\rangle$, and $\mathbf{M}$ is $\langle \theta_i, \theta_j\rangle$. The entries of $\mathbf{M}_{\mathcal{Q}_M}$ can be computed using the definition of Q_M:

$$\langle \mathcal{Q}_M\theta_i, \theta_j\rangle = \sum_{k=1}^{M} \lambda_{\mathcal{Q},i} \langle \theta_i, e_{\mathcal{Q},k}\rangle \langle \theta_j, e_{\mathcal{Q},k}\rangle . \tag{43}$$

The approximation of $e_{\mathcal{Q},k}(\mathbf{x})$ can be completely independent of problem (42), where any appropriate method for the eigenvalue problem (34) can be used. In this work, we use the Nyström method to compute $e_{\mathcal{Q},k}$, where (34) is enforced on some collocation points in D. Depending on the correlation length, we will choose the Legendre–Gauss–Lobatto (LGL) quadrature points either globally on D or locally after a finite element discretization of D. We finally move the eigenvalue $\tilde{\lambda}_M$ to the left-hand side to compute the largest value of $\tilde{\lambda}_M^{-1}$, because the matrix on the left-hand side is nonsingular.

For sampling we also need a numerical solver for the stochastic elliptic problem (7). The weak form is to find $u_{\varepsilon,N} \in V_N$ such that

$$\langle \nabla u_{\varepsilon,N}, \nabla v \rangle = \left\langle \phi + \sqrt{\varepsilon} G_M, v \right\rangle \quad \forall v \in V_N, \tag{44}$$

where $\phi = 0$ for the brute-force Monte Carlo method and $\phi = \phi^* = \pm\Delta u^* = \pm\mathcal{Q}_M v^*$ for the importance sampling. We use the truncated K-L expansion (33) to approximate G. Then the weak form (44) can be written in matrix form as

$$\mathbf{B}\mathbf{u} = \mathbf{f} + \sqrt{\varepsilon}\mathbf{C}\boldsymbol{\xi}, \tag{45}$$

where $\mathbf{B}$ is the stiffness matrix $\langle \theta_i, \theta_j \rangle$, $\mathbf{u}$ the vector consisting of all unknown coefficients, $\mathbf{f}$ the vector $\langle \phi, \theta_i \rangle$ induced by force term ϕ, $\mathbf{C}$ the matrix $\left\langle \theta_i, \sqrt{\lambda_{\mathcal{Q},j}} e_{\mathcal{Q},j} \right\rangle$, $j = 1, \dots, M$, and $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_M)^{\mathsf{T}}$. We can then sample equation (45). Taking $\phi = \phi^*$, one realization of the shifted Gaussian field can be represented as $\psi = \phi^* + \sqrt{\varepsilon} G_M^{(i)}$. Then the Radon–Nikodym derivative can be written in terms ϕ^* and $G_M^{(i)}$ as

$$\frac{d\mathbb{P}_0}{d\mathbb{P}_{\phi^*}}(\psi) = \exp\left(-\frac{\lambda_{M,\min} C_B^2}{2\varepsilon} - \frac{1}{\sqrt{\varepsilon}} \left\langle G_M^{(i)}, \left(\mathcal{Q}_M|_{V_M}\right)^{-1} \phi^* \right\rangle \right),$$

where $\lambda_{M,\min}$ is the minimum eigenvalue given by (37), and $\mathcal{Q}_M|_{V_M}$ is the restriction of $\mathcal{Q}_M$ onto V_M such that the inverse is well defined.

To this end, we summarize our algorithm as follows:

(1) The K-L expansion. Use any appropriate algorithm to solve the eigenvalue problem (34) to obtain $G_M(\mathbf{x})$ and $\mathcal{Q}_M$.

(2) The E-L equation. Use the mixed formulation (15) to solve the E-L equation (13) to obtain (v^*, u^*) and $\phi^* = \pm\Delta u^* = \pm\mathcal{Q}_M v^*$.

(3) The importance sampling. Sample the approximated Gaussian field $G_M(\mathbf{x})$ and use the IS estimator (26) to estimate $\Pr(u_\varepsilon \in B)$.

6. Numerical results.

6.1. The minimizer. We now study the eigenvalues of $\Delta\mathcal{Q}^{-1}\Delta$ numerically using $D = \Gamma^2$. We choose the Gaussian kernel:

$$K(\mathbf{x}, \mathbf{y}) = \frac{1}{\pi l_c^2} \exp\left(-\frac{|\mathbf{x} - \mathbf{y}|^2}{l_c^2} \right). \tag{46}$$

Note that $\lim_{l_c \to 0} K(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$, where $\delta(\cdot)$ indicates the Dirac delta function. As the correlation length l_c goes to 0, the colored noise becomes white, i.e., all eigenvalues of the kernel become 1. In this work, we are interested in the cases that l_c is moderate or large. The decay rate of eigenvalues is determined by the regularity of $K(\mathbf{x}, \mathbf{y})$, and the behavior of the leading eigenvalue is described in Property 3.4.

In the left plot of Figure 1, we plot the decay behavior of the eigenvalues of $K(\mathbf{x}, \mathbf{y})$ for various correlation length l_c. The eigenvalue problem (33) is computed by the Nyström method subject to $48 \times 48 = 2304$ Legendre–Gauss–Lobatto (LGL) quadrature points. We truncated the eigenvalues at the same level such that $\frac{\lambda_{\mathcal{Q},i}}{\lambda_{\mathcal{Q},1}} \leq 10^{-8}$. We see that the eigenvalues decay exponentially and will quickly reach the machine accuracy at a moderate or large correlation length [13]. Note that $\lambda_{\mathcal{Q},1} < 1$

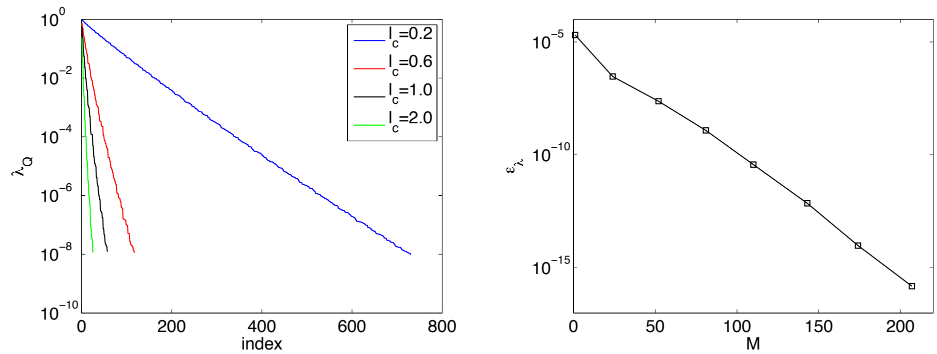


Fig. 1. *Left: The decay behavior of eigenvalues of kernel* (46) *for different correlation lengths.* $\lambda_{\mathcal{Q},1}|_{l_c=0.2} \approx 0.96$*,* $\lambda_{\mathcal{Q},1}|_{l_c=0.6} \approx 0.75$*,* $\lambda_{\mathcal{Q},1}|_{l_c=1.0} \approx 0.54$*,* $\lambda_{\mathcal{Q},1}|_{l_c=2.0} \approx 0.24$*. Note that* $\lambda_{\mathcal{Q},1} < 1$ *for* $l_c > 0$*. The eigenvalues are truncated such that* $\frac{\lambda_{\mathcal{Q},i}}{\lambda_{\mathcal{Q},1}} \leq 10^{-8}$*. Right: The convergence of* $\lambda_{M,\min}$ *with respect to* M*.*

for any $l_c > 0$, and $\lambda_{\mathcal{Q},1}$ decays as l_c increases. This is consistent with Property 3.4 with $\rho(\mathbf{x}) = \pi^{-d/2}\exp(-|\mathbf{x}|^2)$.

In practical applications, one may also concern with a covariance kernel with a constant variance regardless of the correlation length. So we also consider the Gaussian kernel in the following form:

$$\widetilde{K}(\mathbf{x},\mathbf{y}) := \exp\left(-\frac{|\mathbf{x}-\mathbf{y}|^2}{l_c^2}\right) = \pi l_c^2 K(\mathbf{x},\mathbf{y}), \tag{47}$$

$\widetilde{K}(\mathbf{x},\mathbf{x}) = 1$ regardless of the choice of l_c. The eigenvalues of $\tilde{K}$ and K are only different up to a factor πl_c^2, and the associated eigenfunctions are the same.

6.1.1. Convergence of $\boldsymbol{\lambda_{M,\min}}$. We look at the convergence of $\lambda_{M,\min}$ with respect to M, where $\lambda_{M,\min}$ is the minimum eigenvalue given by (37). We set up a rule to choose M by defining

$$M(\delta_{\mathcal{Q}}) = \sum_{i=1}^{\infty} 1_{\left\{\frac{\lambda_{\mathcal{Q},i}}{\lambda_{\mathcal{Q},1}} \leq \delta_{\mathcal{Q}}\right\}}, \tag{48}$$

where $\delta_{\mathcal{Q}}$ is a prescribed threshold. In other words, we keep all the terms such that $\frac{\lambda_{\mathcal{Q},i}}{\lambda_{\mathcal{Q},1}} \leq \delta_{\mathcal{Q}}$. We define a relative error

$$\epsilon_\lambda = \frac{|\lambda_{M,\min}^{-1} - \lambda_{\text{ref}}^{-1}|}{\lambda_{\text{ref}}^{-1}}, \tag{49}$$

where λ_{ref} is a reference solution. In the right plot of Figure 1, we plot the convergence behavior of $\lambda_{M,\min}$ with respect to M. $M(\delta_{\mathcal{Q}})$ is given by decreasing the threshold $\delta_{\mathcal{Q}}$. Kernel (46) is used for $\mathcal{Q}$. $N = 32$ is used to construct the approximation space V_N, i.e., $|V_N| = 32 \times 32$. The correlation length is set to be $l_c = 0.3$. We use the eigenvalue $\lambda_{M,\min}$ given by $\delta_{\mathcal{Q}} = 10^{-10}$ as the reference solution. From Lemma 4.2, we know the convergence rate is determined by the decay rate of the eigenvalues of $\mathcal{Q}$. Since kernel (46) is smooth, its eigenvalues decay exponentially. In the right plot of Figure 1, we observe that the error of $\lambda_{M,\min}$ also decays exponentially with respect to M, which is consistent with our theoretical results.

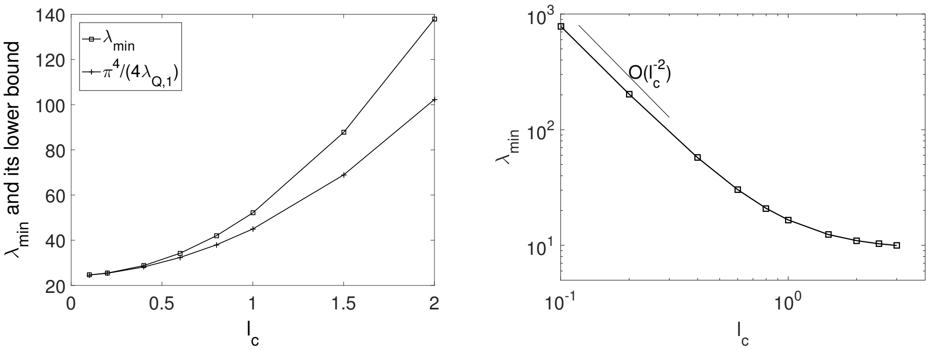


FIG. 2. *The dependence of* $\lambda_{\min}$ *on the correlation length* l_c. *Two-dimensional elliptic problems are considered. Left: Kernel* K. *Right: Kernel* $\widetilde{K}$.

6.1.2. Dependence of $\boldsymbol{\lambda_{\min}}$ on the correlation length. For the computation, we choose $M(\delta_{\mathcal{Q}} = 10^{-6})$ for $\mathcal{Q}_M$, which is good enough since the error in $\lambda_{\min}$ decays exponentially with respect to M. In Figure 2, we plot the dependence of $\lambda_{\min}$ on the correlation length for kernel K on the left and for kernel $\widetilde{K}$ on the right. Let $\lambda_{\min}^K$ and $\lambda_{\min}^{\widetilde{K}}$ be $\lambda_{\min}$ given by K and $\widetilde{K}$, respectively. In the left plot, we also include the curve given by $\pi^4/(4\lambda_{\mathcal{Q},1})$, which is the lower bound of $\lambda_{\min}$ given in Property 3.4 for $L = 2$. It is seen that as $l_c \to 0$, $\lambda_{\min}^K$ goes to $\pi^4/4$ since $\lambda_{\mathcal{Q},1}$ goes to 1 meanwhile. This is because that as $\mathcal{Q} \to \mathcal{I}$, $\langle \Delta u, \mathcal{Q}^{-1}\Delta u\rangle$ Γ-converges to $\langle \Delta u, \Delta u\rangle$ [2], implying the convergence of the corresponding minimizers of the two functionals. As $l_c \to \infty$, $K \sim l_c^{-2}$ and $\lambda_{\mathcal{Q},1} \to 0$, implying that $\lambda_{\min}$ must go to infinity. For kernel $\widetilde{K}$, which is a scaled version of K by πl_c^2, we simply have $\lambda_{\min}^{\widetilde{K}} = \lambda_{\min}^K/(\pi l_c^2)$. The figure for $\lambda_{\min}^{\widetilde{K}}$ is plotted in log-log scale. Since $\lambda_{\min}^K \to \pi^4/4$ as $l_c \to 0$, we should have $\lambda_{\min}^{\widetilde{K}} \sim l_c^{-2}$ when l_c is small. This is demonstrated by the straight line with a slope -2. When $l_c \to \infty$, it appears that $\lambda_{\min}^{\widetilde{K}}$ converges to a constant. When $l_c = \infty$, $\widetilde{K} = 1_D$, corresponding to the right-hand side of the SPDE model (7) being simply a Gaussian random variable. The kernel 1_D has one nonzero eigenvalue $|D|$ associated with a constant eigenfunction. For this case, $\lambda_{\min} = 9.1776$, which is exactly the constant that $\lambda_{\min}^{\widetilde{K}}$ converges to as $l_c \to \infty$.

In summary, the relation of $\lambda_{\min}$ to the correlation length l_c for K has an opposite trend compared to that for $\tilde{K}$. Since $S(u^*)$ is proportional to $\lambda_{\min}$, this implies that the choice of kernel is important for the estimation of $\Pr(B)$ even when the same correlation length and noise amplitude are used.

6.1.3. Profile of the minimizer $\boldsymbol{u^*}$. We now look at the profile of the minimizer (u^*, v^*) corresponding to the smallest eigenvalue of (37), where we normalize u^* such that $\|u^*\| = 1$. We use kernel $\widetilde{K}$ for the demonstration. In Figure 3, we plot u^* for $l_c = 0.1$, 1. The two profiles of u^* are quite similar visually, and are mainly determined by low order modes in V_N. More specifically, the difference between the two u^*'s is $\|u^*|_{l_c=0.1} - u^*|_{l_c=1}\| = 5.19\%$. However, such a slight difference can be amplified by Δ and $\mathcal{Q}^{-1}$. In Figure 4, we plot $(\mathcal{Q}_M|_{V_M})^{-1/2}\Delta u^* = \mathcal{Q}_M^{1/2} v^*$. It is seen that the difference between the two $\mathcal{Q}_M^{1/2} v^*$'s is significant. Indeed, the two minimum eigenvalues given by $\|(\mathcal{Q}_M|_{V_M})^{-1/2}\Delta u^*\|^2/\|u^*\|^2$ are 784.41 and 16.59 for $l_c = 0.1$ and 1, respectively. Note that we have $M(10^{-6}) = 1860$ for $l_c = 0.1$, where $\lambda_{\mathcal{Q},1} = 0.0311$, and $M(10^{-6}) = 41$ for $l_c = 1.0$, where $\lambda_{\mathcal{Q},1} = 1.70$.

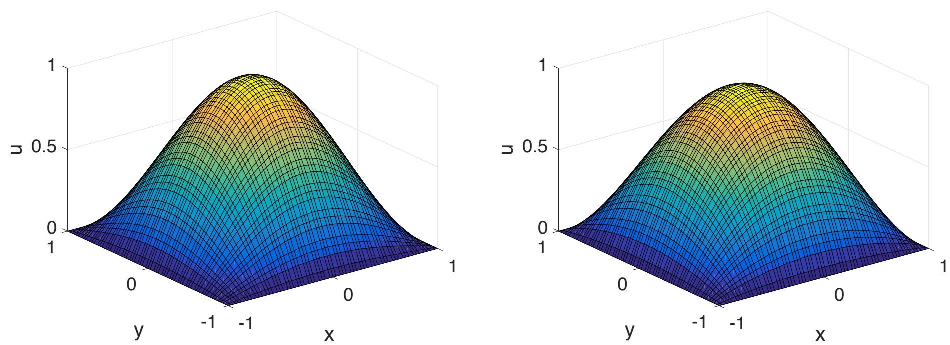


FIG. 3. *The profile of* u^*. *Left:* $l_c = 0.1$. *Right:* $l_c = 1.0$.

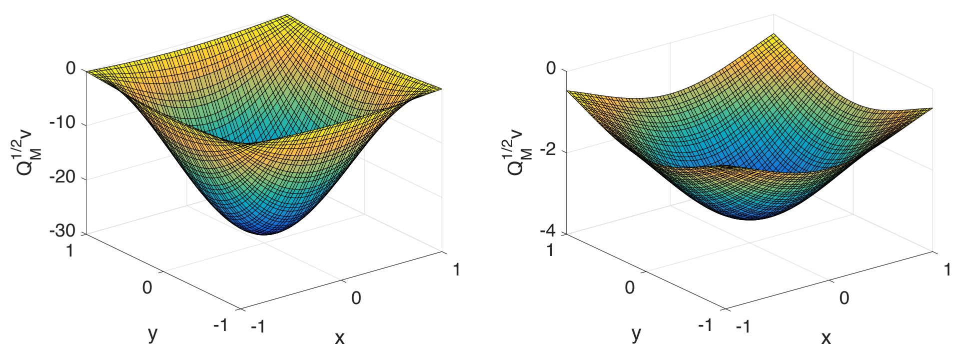
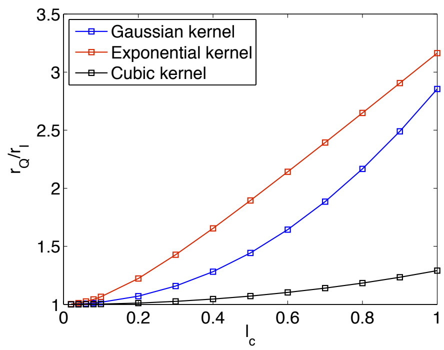


FIG. 4. *The profile of* $(\mathcal{Q}_M|_{V_M})^{-1/2}\Delta u^* = \mathcal{Q}_M^{1/2} v^*$. *Left:* $l_c = 0.1$. *Right:* $l_c = 1.0$.

6.2. Importance sampling for $\mathbf{Pr}(u_\varepsilon \in B_+)$. In this section we will test the efficiency and convergence of the LDP-based IS estimator for the rare event $\Pr(u_\varepsilon \in B_+)$ defined in (24). We consider both one- and two-dimensional elliptic problems.

First of all, we check if the condition for weak efficiency given in Theorem 3.6 is satisfied. As $l_c \to 0$, the kernel (46) approaches the delta function, i.e., $\mathcal{Q}$ goes to the identity operator $\mathcal{I}$. Then $\langle \Delta u, \mathcal{Q}^{-1}\Delta u\rangle$ Γ-converges to $\langle \Delta u, \Delta u\rangle$ [2]. Correspondingly the eigenvalue problem (13) becomes $\Delta^2 u = \lambda u$ subject to Navier boundary conditions, for which we have exact solutions of λ. For the biharmonic operator, $\lambda_2/\lambda_{\min} = 16$ for the 1D case, and $\lambda_2/\lambda_{\min} = 25/4 = 6.25$ for a square, and $\lambda_1/\lambda_{\min} = 4$ for a cube. For nonzero correlation lengths, we plot $\lambda_2/\lambda_{\min}$ of $\Delta\mathcal{Q}^{-1}\Delta$ versus l_c in Figure 5 using kernel K for $\mathcal{Q}$, where the left plot is for the 1D elliptic problem and the right plot is for the 2D elliptic problem. It is clearly shown that for both 1D and 2D cases, as l_c decreases to zero, $\lambda_2/\lambda_{\min}$ monotonically decreases to the limit ratio corresponding to the biharmonic operator. This implies that the IS estimator (26) is asymptotically efficient for any correlation length l_c for both 1D and 2D elliptic problems. Since kernel (47) is the same as kernel (46) up to a scaling factor, this observation holds for both kernels (46) and (47). Noticing that the limit ratio $\lambda_2/\lambda_{\min}$, given by the biharmonic operator, does not depend on the form of the kernel as long as the kernel approaches the delta function, an interesting question is whether the above observation holds for other kernels. In figure 5, we also include the convergence behavior of $\lambda_2/\lambda_{\min}$ given by the exponential kernel $e^{-|\mathbf{x}-\mathbf{y}|/l_c}$ and the



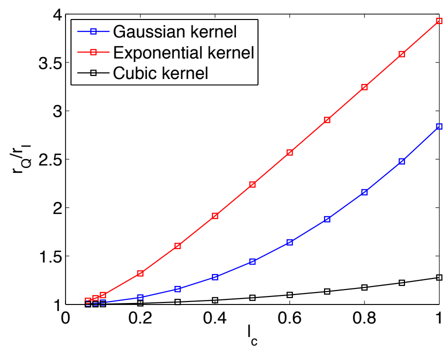


FIG. 5. *$r_{\mathcal{Q}}$ is the ratio $\lambda_2/\lambda_{\min}$ given by the smallest two eigenvalues of $\Delta\mathcal{Q}^{-1}\Delta u = \lambda u$ and $r_{\mathcal{I}}$ is the same ratio associated with the model $\Delta^2 u = \lambda u$. The y axis indicates $r_{\mathcal{Q}}/r_{\mathcal{I}}$. The x axis indicates the correlation length l_c. Left: One-dimensional elliptic problem. $r_{\mathcal{I}} = 16$. Right: Two-dimensional elliptic problem. $r_{\mathcal{I}} = 6.25$.*

following cubic kernel [29]:

$$\begin{cases} 1 - 7\left(\frac{\tau}{l_c}\right)^2 + \frac{35}{4}\left(\frac{\tau}{l_c}\right)^3 - \frac{7}{2}\left(\frac{\tau}{l_c}\right)^5 + \frac{3}{4}\left(\frac{\tau}{l_c}\right)^7 & \text{if } \tau = |\mathbf{x} - \mathbf{y}| < l_c, \\ 0 & \text{otherwise.} \end{cases}$$

The results qualitatively agree with those for the Gaussian kernel. We note that the limit ratio of $\lambda_2/\lambda_{\min}$ is also larger than 3 for 3D elliptic problems. Based on our 1D and 2D results, we expect that the IS estimator may be still asymptotic efficient for any correlation length for 3D elliptic problems.

We now look at the IS estimator $\hat{P}_{\mathsf{IS}}$ defined in (21). We choose the correlation length $l_c = 0.5$ around where the kernels K and $\widetilde{K}$ induce comparable $\lambda_{\min}$. In this experiment, we use kernel K. We pick $M(10^{-6}) = 12$ for the 1D case and $M(10^{-6}) = 110$ for the 2D case. The ratio of $\lambda_2/\lambda_{\min} \approx 23$ for the 1D case and ≈ 9 for the 2D case. Due to the symmetry, and for simplicity, we only consider the estimation of $\Pr(u_\varepsilon \in B_+)$ using Z_{ϕ^*} (see (25)). Let $\hat{\sigma}_{\mathsf{IS}}$ and $\hat{\sigma}_{\mathsf{MC}}$ be the unbiased estimates of the standard deviation of the IS estimator $\hat{P}_{\mathsf{IS}}$ and Monte Carlo estimator $\hat{P}_{\mathsf{MC}}$, respectively. We use

$$\epsilon_{\mathsf{IS}} = \hat{\sigma}_{\mathsf{IS}}/P_{\text{ref}} \quad \text{and} \quad \epsilon_{\mathsf{MC}} = \hat{\sigma}_{\mathsf{MC}}/P_{\text{ref}}$$

to indicate the relative errors, where P_{ref} is a reference estimate of $\Pr(B_+)$.

We compare the IS estimator and the Monte Carlo estimator by considering a sequence of increasing values of C_B so that $\Pr(u_\varepsilon \in B_+)$ decreases. The reference probability is given by the IS estimator subject to 10^7 realizations. The results are summarized in Table 1 for the 1D case and in Table 2 for the 2D case.

First, it is easy to check that both ϵ_{IS} and ϵ_{MC} have the convergence rate $O(N_{\mathsf{MC}}^{-1/2})$. Second, the IS estimator is always better than the MC estimator. For the same number of realizations, the ratio $\epsilon_{\mathsf{MC}}/\epsilon_{\mathsf{IS}}$ keeps increasing as C_B increases. When C_B is small enough, their performance is similar, as shown in Tables 1(a) and 2(a). When $\Pr(u_\varepsilon \in B_+)$ is of $O(10^{-3})$ and $O(10^{-5})$, the ratio $\epsilon_{\mathsf{MC}}/\epsilon_{\mathsf{IS}}$ is about of $O(10)$ and of $O(10^2)$, respectively, as shown in Tables 1(b)–1(c) and 2(b)–2(c). When C_B is large enough, i.e., $\Pr(u_\varepsilon \in B_+)$ is small enough, the Monte Carlo method starts to have difficulties in capturing the rare events, as shown in Tables 1(d) and 2(d). Interestingly,

TABLE 1
[1D case] Convergence behavior of MC and IS estimators for one-dimensional elliptic problems. The noise amplitude is chosen as $\varepsilon = 10^{-8}$.

(a) $C_B = 3 \times 10^{-5}$

N_{MC}	10^3	10^4	10^5	10^6	10^7
$\hat{P}_{\mathsf{IS}}$	2.154e-1	2.272e-1	2.249e-1	2.258e-1	2.254e-1
ϵ_{IS}	3.5e-2	1.1e-2	3.5e-3	1.1e-3	3.6e-4
$\hat{P}_{\mathsf{MC}}$	2.210e-1	2.241e-1	2.271e-1	2.254e-6	2.252e-6
ϵ_{MC}	5.8e-2	1.9e-2	5.9e-3	1.9e-3	5.9e-4

(b) $C_B = 1 \times 10^{-4}$

N_{MC}	10^3	10^4	10^5	10^6	10^7
$\hat{P}_{\mathsf{IS}}$	4.539e-3	4.399e-3	4.458e-3	4.446e-3	4.443e-3
ϵ_{IS}	5.6e-2	1.7e-2	5.4e-3	1.7e-3	5.5e-4
$\hat{P}_{\mathsf{MC}}$	8.000e-3	4.000e-3	4.150e-3	4.354e-3	4.420e-3
ϵ_{MC}	6.3e-1	1.4e-1	4.6e-2	1.5e-2	4.7e-3

(c) $C_B = 1.5 \times 10^{-4}$

N_{MC}	10^3	10^4	10^5	10^6	10^7
$\hat{P}_{\mathsf{IS}}$	4.533e-5	4.263e-5	4.190e-5	4.207e-5	4.200e-5
ϵ_{IS}	6.8e-1	2.1e-2	6.7e-3	2.1e-3	6.7e-4
$\hat{P}_{\mathsf{MC}}$	1.000e-3	1.000e-4	5.000e-5	3.700e-5	4.12e-5
ϵ_{MC}	23.8	2.4	5.3e-1	1.5e-1	4.8e-2

(d) $C_B = 2 \times 10^{-4}$

N_{MC}	10^3	10^4	10^5	10^6	10^7
$\hat{P}_{\mathsf{IS}}$	7.399e-8	7.806e-8	7.608e-8	7.688e-8	7.717e-8
ϵ_{IS}	7.4e-2	2.5e-2	7.6e-3	2.4e-3	7.7e-4
$\hat{P}_{\mathsf{MC}}$	-	-	-	-	1.000e-7
ϵ_{MC}	-	-	-	-	1.3

one out of the 10^7 realizations for the MC estimator of the 1D elliptic problem captured the rare event in our numerical experiment, where ϵ_{MC} is of $O(1)$, and $\epsilon_{\mathsf{MC}}/\epsilon_{\mathsf{IS}}$ is of $O(10^3)$. For the 2D case, the 10^7 realizations for the MC estimator did not capture the rare event. If we assume one realization captures the rare event for this case, it would induce a relative error 10.62 (bracketed number in the table), corresponding to a ratio $\epsilon_{\mathsf{MC}}/\epsilon_{\mathsf{IS}}$ of $O(10^4)$. Third, the IS estimator appears robust for both 1D and 2D elliptic problems in terms of all the four C_B's, since we do not observe a significant change in ϵ_{IS} for the same number of realizations as C_B increases. More specifically, we look at the change of the relative error per sample, i.e., $\epsilon_{\mathsf{IS}}\sqrt{N_{\mathsf{MC}}}$, with respect to ε. In Figure 6 we plot the relative error per sample versus ε^{-1} in log-log scale on the left, and the probability $\Pr(u_\varepsilon \in B_+)$ versus ε^{-1} with only the y axis in log scale on the right, which confirms that as ε decreases, $\Pr(u_\varepsilon \in B_+)$ decays exponentially. While the relative error per sample of the MC estimator increases exponentially, the scaling in figure 6 shows that the relative error per sample of the IS estimator appears to increase only algebraically, implying that our IS scheme is indeed weakly efficient.

6.3. The dependence of $\boldsymbol{\lambda_2/\lambda_{\min}}$ on random events. In section 3, we presented an LDP-based IS importance sampling scheme to estimate $\Pr(B)$, where a

TABLE 2

[2D case] Convergence behavior of MC and IS estimators for two-dimensional elliptic problems. The noise amplitude is chosen as $\varepsilon = 10^{-8}$.

(a) $C_B = 1 \times 10^{-5}$

N_{MC}	10^3	10^4	10^5	10^6	10^7
$\hat{P}_{\mathsf{IS}}$	3.369e-1	3.345e-1	3.366e-1	3.390e-1	3.391e-1
ϵ_{IS}	3.2e-2	1.0e-2	3.1e-3	1.0e-3	3.2e-4
$\hat{P}_{\mathsf{MC}}$	3.360e-1	3.341e-1	3.378e-1	3.387e-1	3.391e-1
ϵ_{MC}	4.4e-2	1.4e-2	4.4e-3	1.4e-3	4.4e-4

(b) $C_B = 5 \times 10^{-5}$

N_{MC}	10^3	10^4	10^5	10^6	10^7
$\hat{P}_{\mathsf{IS}}$	2.727e-3	2.620e-3	2.623e-3	2.614e-3	2.617e-3
ϵ_{IS}	5.8e-2	1.8e-2	5.7e-3	1.8e-3	5.7e-4
$\hat{P}_{\mathsf{MC}}$	4.000e-3	2.100e-3	2.750e-3	2.525e-3	2.576e-3
ϵ_{MC}	7.6e-1	1.8e-1	6.3e-2	1.9e-2	6.1e-3

(c) $C_B = 7 \times 10^{-5}$

N_{MC}	10^3	10^4	10^5	10^6	10^7
$\hat{P}_{\mathsf{IS}}$	4.380e-5	4.420e-5	4.270e-5	4.311e-5	4.308e-5
ϵ_{IS}	6.9e-2	2.2e-2	6.8e-3	2.2e-3	6.8e-4
$\hat{P}_{\mathsf{MC}}$	1.000e-3	3.000e-4	6.000e-5	4.000e-5	4.370e-5
ϵ_{MC}	23.2	4.0	5.7e-1	1.5e-1	4.9e-2

(d) $C_B = 1 \times 10^{-4}$

N_{MC}	10^3	10^4	10^5	10^6	10^7
$\hat{P}_{\mathsf{IS}}$	9.019e-9	9.213e-9	9.444e-9	9.423e-9	9.416e-9
ϵ_{IS}	8.2e-1	2.6e-2	8.4e-3	2.6e-3	8.3e-4
$\hat{P}_{\mathsf{MC}}$	-	-	-	-	(1.000e-7)
ϵ_{MC}	-	-	-	-	(10.6)

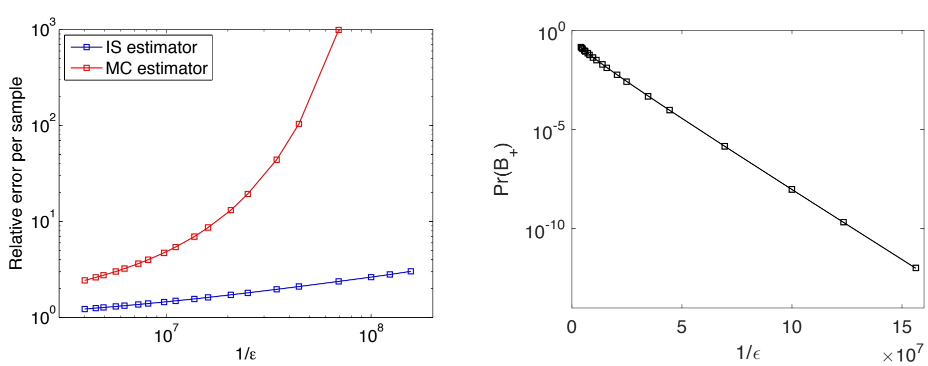


FIG. 6. *The IS estimator for two-dimensional elliptic problem with $C_B = 10^{-4}$ and $N_{MC} = 10^6$ realizations. Left: Relative error per sample versus ε^{-1}. Right:* $\Pr(u_\varepsilon \in B_+)$ *versus ε^{-1}.*

sufficient and necessary condition, i.e., $\lambda_2/\lambda_{\min} \geq 3$, for the weak efficiency is obtained. Numerical experiments in section 6.2 show that this condition holds for any correlation length in 1D and 2D cases. However, we note that the definition of the eigenvalue problem (13) depends on the definition of random-event set B, and so does the condition given in Theorem (3.6). To demonstrate this, we subsequently apply the procedure developed in this work to another rare-event set, where we pay particular condition to the condition for the weak efficiency of the IS estimator.

We consider

$$\hat{B} = \{\|\nabla u\| \geq C_{\hat{B}}\}. \tag{50}$$

It can be obtained that the minimizer of $S(u)$ restricted onto $\hat{B}$ corresponds to the following new biharmonic-type eigenvalue problem:

$$\Delta \mathcal{Q}^{-1} \Delta u = -\lambda \Delta u, \tag{51}$$

subject to boundary conditions $u|_{\partial D} = (\mathcal{Q}^{-1}\Delta u)|_{\partial D} = 0$, which has a mixed formulation

$$\Delta u = \mathcal{Q} v, \quad \Delta v = -\lambda \Delta u, \tag{52}$$

subject to homogeneous Dirichlet boundary conditions $u|_{\partial D} = v|_{\partial D} = 0$. Equation (52) can be decoupled into two subproblems:

$$\begin{cases} \Delta v = -\lambda \mathcal{Q} v, & \mathbf{x} \in D, \\ v|_{\partial D} = 0, & \mathbf{x} \in \partial D, \end{cases} \quad \text{and} \quad \begin{cases} \Delta u = \mathcal{Q} v, & \mathbf{x} \in D, \\ u|_{\partial D} = 0, & \mathbf{x} \in \partial D. \end{cases} \tag{53}$$

To construct an asymptotically efficient IS estimator, we consider the following splitting $\hat{B} = \hat{B}_+ \cap \hat{B}_-$:

$$\hat{B}_+ = \{u_\varepsilon(\mathbf{x}) | u_\varepsilon \in B, \langle u_\varepsilon, \phi^* \rangle \geq 0\} \quad \text{and} \quad \hat{B}_- = \{u_\varepsilon(\mathbf{x}) | u_\varepsilon \in B, \langle u_\varepsilon, \phi^* \rangle \leq 0\},$$

where $\phi^* = -\Delta u^* = -\mathcal{Q} v^*$, and we still use u^* to indicate the minimizer of $S(u)$. Note that we project u_ε onto ϕ^* instead of u^* due to the definition of $\hat{B}$. Using an argument similar to that in the proof of Theorem 3.6, we can obtain the same condition $\lambda_2 \geq 3\lambda_{\min}$ such that Z_{ϕ^*} is asymptotically efficient for $\Pr(\hat{B}_+)$, except that $\lambda_{\min}$ and λ_2 are the two smallest eigenvalues of (51) with $\lambda_{\min}$ being simple. We then obtain an asymptotically efficient IS estimator $Z_{\phi^*} + Z_{-\phi^*}$ for $\Pr(\hat{B})$.

For the random event $\hat{B}$ defined in (50), we see from (53) that the limit ratio of $\lambda_2/\lambda_{\min}$ as $l_c \to 0$ is defined by the eigenvalues of Laplace operator subject to homogeneous Dirichlet boundary conditions. When $\mathcal{Q} = \mathcal{I}$, $\lambda_2/\lambda_{\min} = 4$ for the 1D case, $\lambda_2/\lambda_{\min} = 2.5$ for a square and $\lambda_2/\lambda_{\min} = 2$ for a cube. This implies that the condition $\lambda_2/\lambda_{\min} > 3$ may break down for 2D and 3D cases when l_c is small enough. In Figure 7, we plot $\lambda_2/\lambda_{\min}$ for 1D and 2D elliptic problems. For 1D elliptic problems, $\lambda_2/\lambda_{\min} > 3$ for any correlation length l_c. For 2D elliptic problems, we plotted $r_{\mathcal{Q}}/3 = (\lambda_2/\lambda_{\min})/3$. It is seen that for all the three kernels studied, $\lambda_2/\lambda_{\min}$ is smaller than 3 if l_c is not large enough, indicated by the dotted horizontal line with $r_{\mathcal{Q}} = 3$. For l_c such that the corresponding $r_{\mathcal{Q}}/3$ is under the dotted horizontal line, the IS estimator Z_{ϕ^*} for $\hat{B}$ does not have weak efficiency.

We now compare the convergence behavior of the IS estimator at two correlation lengths $l_c = 0.1$ and $l_c = 0.5$ for the two-dimensional elliptic problems and the kernel K. When $l_c = 0.1$, $\lambda_2/\lambda_{\min} \approx 2.54$; when $l_c = 0.5$, $\lambda_2/\lambda_{\min} \approx 3.58$. In other

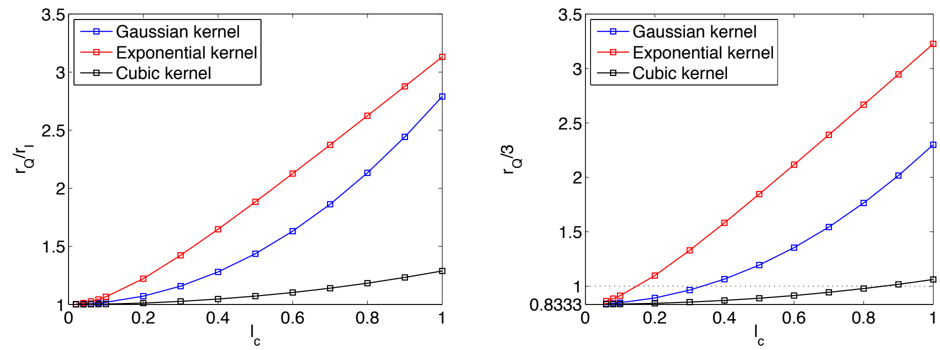


FIG. 7. *$r_{\mathcal{Q}}$ is the ratio $\lambda_2/\lambda_{\min}$ given by the two smallest eigenvalues of $\Delta\mathcal{Q}^{-1}\Delta u = -\lambda\Delta u$ and $r_{\mathcal{I}}$ is the same ratio given by $\Delta^2 u = -\lambda\Delta u$. The x axis indicates the correlation length l_c. Left: The y axis indicates $r_{\mathcal{Q}}/r_{\mathcal{I}}$. The one-dimensional elliptic problem is considered. $r_{\mathcal{I}} = 4$. Right: The y axis indicates $r_{\mathcal{Q}}/3$. The two-dimensional elliptic problem is considered. $r_{\mathcal{I}} = 2.5$. Note that $r_{\mathcal{I}}/3 \approx 0.8333$.*

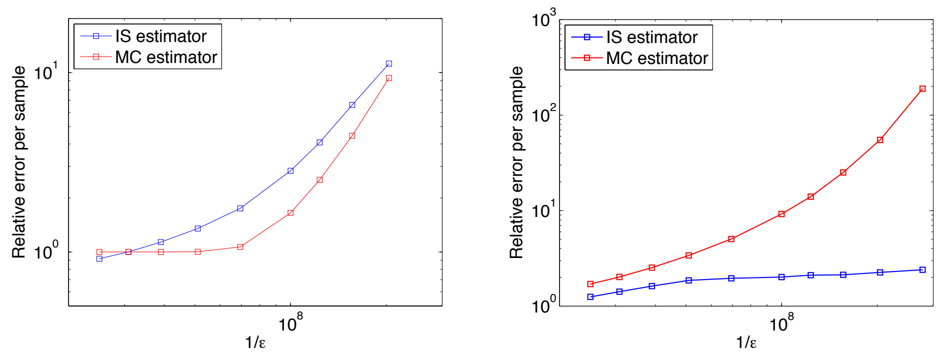


FIG. 8. *Relative error per sample for the two-dimensional elliptic problem with $C_{\hat{B}} = 10^{-4}$ and $N_{MC} = 10^6$ realizations. Left: $l_c = 0.1, \lambda_2/\lambda_{\min} \approx 2.54$. Right: $l_c = 0.5, \lambda_2/\lambda_{\min} \approx 3.58$.*

words, the condition for weak efficiency holds for $l_c = 0.5$, but fails for $l_c = 0.1$. We use $M(10^{-5}) = 1370$ and $M(10^{-5}) = 110$ to approximate $\mathcal{Q}_M$ for $l_c = 0.1$ and 0.5, respectively. We fix $N_{\mathsf{MC}} = 10^6$ and $C_{\hat{B}} = 10^{-4}$, and vary ε. We will focus on relatively large ε, such that the Monte Carlo estimator with $N_{\mathsf{MC}} = 10^6$ yields a reasonable approximation of $\mathbb{E}[1_{\hat{B}_+}]$. In Figure 8, we compare the relative error per sample given by both IS and MC estimators. It is seen that the relative error per sample given by the IS estimator increases exponentially for $l_c = 0.1$ and algebraically for $l_c = 0.5$, indicating that the LDP-based IS estimator for random-event set $\hat{B}$ is not weakly efficient when the condition $\lambda_2/\lambda_{\min} \geq 3$ fails.

7. Conclusion and discussion. In this work, we addressed the probability estimation of small-noise-induced rare events for an elliptic problem by employing the large deviation principle. The whole procedure consists of several steps. (1) Approximation of the Gaussian random field: We employed the Karhunen–Loéve expansion to do this, which results in a finite-rank approximation $\mathcal{Q}_M$ of $\mathcal{Q}$. (2) Minimizing the rate functional: We derived the associated Euler–Lagrange equation, which corresponds to an eigenvalue problem of a nonlocal operator $\Delta\mathcal{Q}^{-1}\Delta$. To deal with the Navier-type boundary conditions we considered a mixed formulation of the E-L equation. We have proved that $\Delta(\mathcal{Q}_M|_{V_M})^{-1}\Delta$ has M positive eigenvalues if $\mathcal{Q}$ is replaced by its finite-rank approximation $\mathcal{Q}_M$, and the convergence rate of eigenvalues as $M \to \infty$ is consistent with the decay rate of the eigenvalues of $\mathcal{Q}$. (3) Exponential tilting importance sampling estimator: We derived a sufficient and necessary condition that $\lambda_2/\lambda_{\min} \geq 3$ to guarantee the weak efficiency of the estimator.

Although our problem setting is relatively simple and specific, our work provides a fundamental understanding of the main difficulties, and a guidance to the probability estimation of other random events such as $\hat{B}$ given in section 6.3. We expect to generalize our work to a more general setting, where time dependence and nonlinearity are taken into account as in equation (4).

REFERENCES

[1] I. Babuška and J. Osborn, *Eigenvalue Problems*, in Handbook of Numerical Analysis II, P. G. Ciarlet and J. I. Lions, eds., North-Holland, Amsterdam, 1991, pp. 641–687.

[2] Z. Brzezniak, S. Cerrai, and M. Freidlin, *Quasipotential and exit time for 2D stochastic Navier-Stokes equations by space time white noise*, Probab. Theory Relat. Fields, 162 (2015), pp. 739–793.

[3] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, Berlin, 1991.

[4] J. A. Bucklew, *Introduction to Rare Event Simulation*, Springer-Verlag, Berlin, 2004.

[5] S. Cerrai and M. Freidlin, *Approximation of quasi-potential and exit problems for multidimensional RDE's with noise*, Trans. Amer. Math. Soc., 363 (2011), pp. 3853–3892.

[6] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer-Verlag, Berlin, 1998.

[7] P. Dupuis, K. Spiliopoulos, and X. Zhou, *Escaping from an attractor: Importance sampling and rest points I*, Ann. Appl. Probab., 25 (2015), pp. 2909–2958.

[8] P. Dupuis and H. Wang, *Importance sampling, large deviations, and differential games*, Stoch. Stoch. Reports, 76 (2004), pp. 481–508.

[9] W. E, W. Ren and E. Vanden-Eijnden, *Minimum action method for the study of rare events*, Commun. Pure Appl. Math., 57 (2004), pp. 637–565.

[10] L. C. Evans, *Partial Differential Equations*, American Mathematical Society, Providence, RI, 1998.

[11] W. Faris and G. Jona-Lasinio, *Large fluctuations for a nonlinear hear equation with noise*, J. Phys. A: Math. Gen., 15 (1982), pp. 3025–3055.

[12] H. C. Fogedby and W. Ren, *Minimum action method for the Kardar-Parisi-Zhang equation*, Phys. Rev. E, 80 (2009), p. 041116.

[13] P. Frauenfelder, C. Schwab, and R. A. Todor, *Finite elements for elliptic problems with stochastic coefficients*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 205–228.

[14] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems*, 2nd ed., Springer-Verlag, New York, 1998.

[15] P. Glassman and Y. Wang, *Counterexamples in importance sampling for large deviation probabilities*, Ann. Appl. Probab., 7 (1997), pp. 731–746.

[16] T. Grafke, R. Grauer, and T. Schäfer, *The instanton method and its numerical implementation in fluid mechanics*, J. Phys. A Math. Theor., 48 (2015), 333001.

[17] M. Hairer and H. Weber, *Large deviations for white-noise driven, nonlinear stochastic PDEs in two and three dimensions*, Ann. Fac. Sci. Toulouse, 24 (2015), pp. 55–92.

[18] M. Heymann and E. Vanden-Eijnden, *The geometric minimum action method: A least action principle on the space of curves*, Commun. Pure Appl. Math., 61 (2008), pp. 1052–1117.

[19] M. A. Lifshits, *Gaussian Random Functions*, Mathematics and Its Applications, Springer-Verlag, Dordrecht, 1995.

[20] X. Li, J. Liu, J. Lu, and X. Zhou, *Moderate deviation for random elliptic PDE with small noise*, Annu. Appl. Probab., to appear.

[21] J. Liu, J. Lu, and X. Zhou, *Efficient rare event simulation for failure problems in random media*, SIAM J. Sci. Comput., 37 (2015), pp. A609–A624.

[22] J. Liu, X. Zhou, R. Patra, and W. E., *Failure of random materials: A large deviation and computational study*, in Proceedings of the 2011 Winter Simulation Conference, IEEE, Piscataway, NJ, 2011, pp. 3784–3794.

[23] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Cambridge University Press, Cambridge, 1992.
[24] M. Salins and K. Spiliopoulos, *Rare event simulation via importance sampling for linear SPDE's*, arXiv:1609.04365, 2016.
[25] J. Shen, *Efficient spectral-Galerkin method I: Direct solvers for the second and fourth order equations using Legendre polynomials*, SIAM J. Sci. Comput., 15 (1994), pp. 1489–1504.
[26] S. Sritharan and P. Sundar, *Large deviations for the two-dimensional Navier-Stokes equations with multiplicative noise*, Stoch. Process. Appl., 116 (2006), pp. 1636–1659.
[27] X. Zhou, W. Ren, and W. E., *Adaptive minimum action method for the study of rare events*, J. Chem. Phys., 128 (2008), 104111.
[28] J. Weidmann, *Linear Operators in Hilbert Spaces*, Springer-Verlag, Berlin, 1980.
[29] E. Wackernagel, *Multivariate Geostatistics*, Springer-Verlag, Berlin, 1995.
[30] X. Wan, *An adaptive high-order minimum action method*, J. Comput. Phys., 230 (2011), pp. 8669–8682.
[31] X. Wan, H. Yu, and W. E., *Model the nonlinear instability of wall-bounded shear flows as a rare event: A study on two-dimensional Poiseuille flow*, Nonlinearity, 28 (2015), pp. 1409–1440.
[32] X. Wan, *A minimum action method with optimal linear time scaling*, Commun. Comput. Phys., 18 (2015), pp. 1352–1379.
[33] X. Wan, B. Zheng, and G. Lin, *An hp-adaptive minimum action method based on a posteriori error estimate*, Commun. Comput. Phys., 23 (2018), pp. 408–439.
[34] W. Yao and W. Ren, *Noise-induced transition in barotropic flow over topography and application to Kuroshio*, J. Comput. Phys., 300 (2015), pp. 352–364.
[35] W. Zhang, T. Li, and P. Zhang, *Numerical study for the nucleation of one-dimensional stochastic Cahn–Hilliard dynamics*, Commun. Math. Sci., 10 (2012), pp. 1105–1132.