Research Statement

Shanyin Tong

My research is in applied and computational mathematics, specially, uncertainty quantification, PDE-constrained optimization, optimization under uncertainty and rare events. As a PhD student at the Courant Institute, I have worked on several projects. The most representative one is quantifying extreme tsunami waves through large deviation theory and PDE-constrained optimization. This project is at the intersection of applied probability, optimization, sampling, partial differential equations and geophysics. With my advisors (Georg Stadler and Eric Vanden-Eijnden), I developed a new sampling-free method to estimate extreme event probabilities, obtaining experience both in theoretical derivation and numerical implementation. This method was further studied and applied to chance-constrained optimization problems during a summer internship at Argonne National Lab, in collaboration with Anirudh Subramanyam and Vishwas Rao. I additionally acquired a background in machine learning methods, and I plan to use these methods in scientific computing and uncertainty quantification. Currently, I am developing a reduced dimension importance sampling framework informed by large deviation theory and the related constrained optimization problem. In the following sections, I will talk about my experiences and projects in detail and discuss prospective directions for my future research.

1 Extreme event probability estimation and uncertainty quantification

Extreme events are infrequent but have severe consequences. Examples include hurricanes, energy grid blackouts, dam breaks, earthquakes, and pandemics [5]. Because estimating the probability of such events can inform strategies that mitigate their effects, we want to develop methods to study the distribution tail of these occurrences. The problem can be described by computing the probability measure of the extreme event set $\Omega(z) = \{\theta : F(\theta) \geq z\}$, where θ represents the uncertainty in the system, F is the parameter-to-observation map characterizing the dynamics, and z is a given threshold controlling the rareness of the event.

Explicitly calculating these small probabilities is generally infeasible, particularly when the probability density depends on complex dynamics and the random variables that enter these dynamical systems are

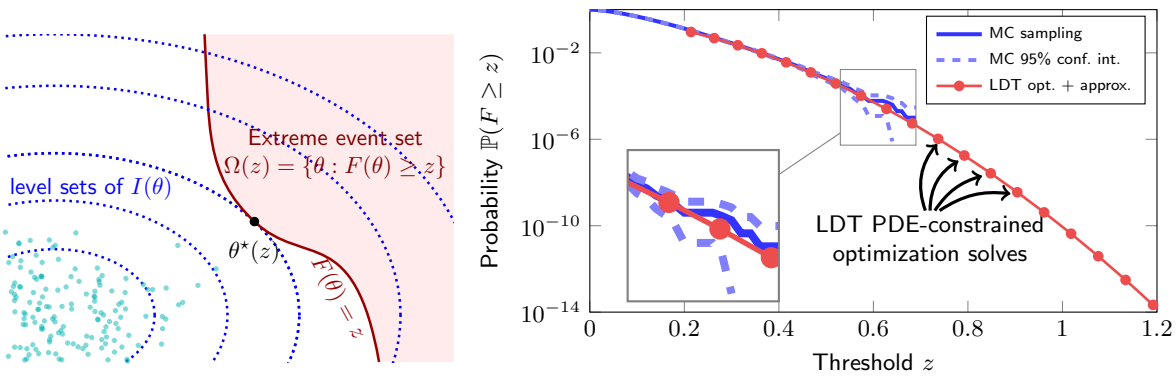

Figure 1: Left: Illustration of two-dimensional random parameter space with level sets (dotted blue lines) of rate functions $I(\theta)$. Monte Carlo samples (blue dots) typically fall outside of the set of parameters that lead to extreme events (in red). Large deviation theory (LDT) indicates that the extreme event probability is dominated by the point $\theta^\star(z)$: the least unlikely point that leads to an event of size z or larger. Right: Probability estimation for the tsunami problem. The event threshold z is plotted on the x-axis and the probability of events of size at least z is on the y-axis. The blue line depicts results that are obtained via standard Monte Carlo estimation with 10^5 samples and their 95 percent confidence intervals (blue dotted line). The red line depicts the estimation that results from large deviation theory (LDT) optimization combined with curvature information of $F(\theta) = z$. Figure taken from [13].

high dimensional. Standard Monte Carlo methods are inefficient for the exploration of probability tails because most samples are far from the extreme event set (see Figure 1). Sampling that is guided by tailored proposal distributions may improve this situation. Thus, how to choose the proposal distributions and how we can avoid sampling for the estimations are vital problems to study.

Main contributions. Following the work of large deviation theory (LDT) for extreme event quantification in dynamical systems with random components [3, 4], I derived an extreme event probability estimation framework that exploited connections between probability estimation and constrained optimization, proposed approaches to refine the asymptotic probability estimates from LDT, and applied our approaches for estimating probabilities of extreme tsunami waves. This work was featured at the homepage of SIAM News [13, 17].

The framework first requires the computation of the most important point in the extreme event set $\Omega(z)$; i.e., the minimizer of the rate function $I(\theta)$. This rate function is a convex function that is fully determined by the probability distribution of θ. Under reasonable assumptions, LDT implies that the minimizer $\theta^\star(z)$ (see Figure 1) holds crucial information to estimate the target probability. Thus, we transfer the probability estimation problem to the constrained optimization problem

$$\theta^\star(z) = \operatorname*{argmin}_{\theta:F(\theta)=z} I(\theta). \tag{1}$$

After solving this optimization problem, we show that importance sampling (IS) using the proposal distribution guided by this optimizer (mean shifted to $\theta^\star(z)$) can lead to an exponential reductions of relative errors in all parameter directions, which answers our first question about how to choose proposal densities for sampling to estimate the target small probability and reduce the required sample sizes.

Furthermore, we propose sampling-free approaches for estimating $\mathbb{P}(F(\theta) \geq z)$. We compute the probability measure bounded by the second-order Taylor expansion of F at $\theta^\star(z)$ as the approximation of the target probability, which we prove to be asymptotically accurate. The computation of this approximation only requires the eigenvalues of the covariance-preconditioned Hessian of F at $\theta^\star(z)$ when the random parameter is Gaussian. This approximation is similar to the Second Order Reliability Method (SORM) in engineering [12]. Our contribution is to combine it with randomized linear algebra methods [9], extending its applicability to high-dimensional problems. More importantly, the computational costs are independent of the extremeness of the events and the discretization dimension, which is a vital improvement for rare event study since the required efforts of sampling-based methods often scale exponentially with respect to the rareness.

We applied our approaches to tsunami hazard assessment [16, 17, 19], where we are interested in estimating the probability of tsunami wave higher than given thresholds while the tsunamis are generated by random slips under the earth plates. The LDT-based approaches allow for accurate estimation of tail probabilities by solving a sequence of optimization problems with different thresholds z and computing the eigenvalues of the preconditioned Hessian for each optimizer (results shown in Figure 1).

Current and future work. I am currently refining the proposal distribution for IS based on the LDT optimizer and the cross-entropy method [18] by finding a Gaussian distribution minimizing the Kullback-Leibler (KL) divergence with the optimal biasing distribution [14]. Here, I am using the dominant eigenspace of the covariance-preconditioned Hessian of F at $\theta^\star$ to explore the low-dimensional structure of the rare event simulation problem, and build connections with the likelihood-informed subspace (LIS) for dimension reduction techniques from Bayesian inverse problems [2].

Large deviation theory and asymptotic approximations can also be used in estimating expectation of quantity of interest (QoI) in stochastic differential equations (SDE). Together with Tobias Grafke (University of Warwick), we are currently studying the method of efficiently computing the expectation based on the second-order approximations of the QoI at the LDT minimizer, and its relation to the Laplace method and the approximations based on solving Riccati equations [8].

I want to extend the LDT-based extreme event estimation framework to other applications and combine it with methods from other areas including machine learning, optimal transport, optimal control and inverse problems. Normalizing flows and deep neural networks have recently been used for assisting Markov Chain Monte Carlo Methods (MCMC) sampling unknown distributions [6]. I am interested in studying these methods and their possible use in Bayesian inference and rare even study.

2 PDE-constrained optimization

The parameter-to-event map F in the tsunami application in section 1 involves the solution of the shallow water equations, computing minimizers results in a PDE-constrained optimization problem. The governing equation is a nonlinear hyperbolic conservation law, which makes the optimization problem challenging for multiple reasons, like the possible occurrence of shocks in the solution and difficulties that arise in the computation of adjoint-based gradients.

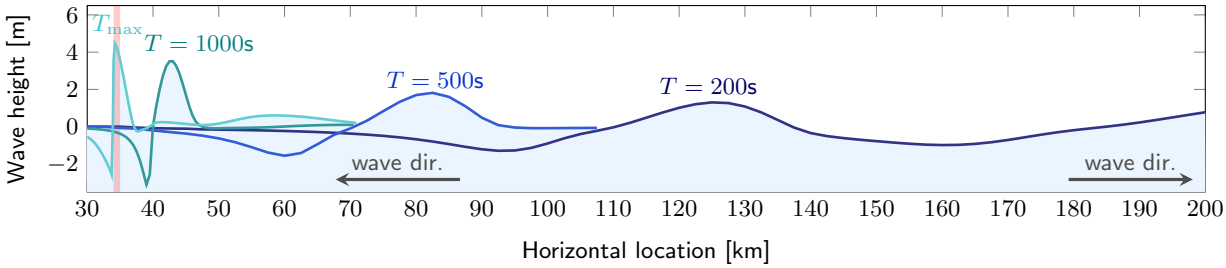

Figure 2: Snapshots at different times T of tsunami waves generated by the optimizer with maximal observed height $z = 4$ from the random log-normal source distribution whose mean correspond to the $M_w = 9$ magnitude earthquake. The ocean floor deformation generates waves in both directions, we focus on the waves traveling towards the shore on the left. The tsunami wave is compressed as it travels towards shore, its magnitude increases and its leading edge steepens.

Main contributions. The solutions of nonlinear hyperbolic equations may have shocks, and we also observed them in our simulation. To have well-defined discrete state and adjoint equations, we added artificial viscosity to prevent slopes that cannot be resolved by discretization, following the analysis in [7]. The key for solving PDE-constrained optimization problem is to compute the gradient, for which I used an adjoint method which only requires two PDE solves for one gradient computation. To solve the 1D shallow water equations, I used the discontinuous Galerkin finite element method (DG-FEM) with linear interpolating polynomials and a global Lax-Friedrichs flux to discretize the equations in space, and the strong stability-preserving second-order Runge-Kutta (SSP-RK2) method to discretize the equations in time [10]. The adjoint equation used a discretize-then-optimize method to maintain the consistency. To evaluate gradients, I integrated the state and adjoint solutions over time using the rule induced by the RK scheme. I derived the associated formula and equations, and implemented them using MATLAB. Snapshots of waves generated by the optimizer are shown in Figure 2.

Current and future work. One project I am involved in uses these tools on inversion of slips in earthquake. We target inference of the slip along one boundary of a 3D subduction zone based on point observations of the elastic displacement on another boundary. The problem is a linear steady-state inverse problem. I derived the system associated with the problem, and solved it by implementing the equations using FEniCS in Python. Currently, I collaborate with Alen Alexanderian (North Carolina State University) and his students to use my implementation for studying sensitivities of Bayesian inverse problem solutions, focusing on the K-L divergence between prior and posterior [1].

I am interested in studying PDE-constrained optimization problems under uncertainty, for example, the optimal control and mitigation of the tsunami waves by designing wave breakers on shore. For hyperbolic governing equations, I want to dive deeper on the analysis of adjoint method, especially their convergence when the solution has shocks. I also plan to extend our current derivations and codes to the 2D settings for the tsunami problem, since there are no adjoint-based tools for the 2D shallow water equation. Another interesting direction is to study the connections and differences between adjoint methods and machine learning methods for solving PDE-constrained optimization problems, to interpret the layer information in neural networks as time steps of an ODE, and to explore the reliable uses of machine learning methods for these problems.

3 Optimization with probabilistic constraints

Probabilistic constraints provide a principled framework to mitigate risks of high-impact extreme events. The rare occurrences of such events, however, impose severe sampling and computational requirements on classical solution methods that render them impractical. Inspired by the work discussed in section 1,

I collaborated with Anirudh Subramanyam and Vishwas Rao during my summer internship at Argonne National Lab, to study sampling-free approaches for rare chance constrained optimizations [15].

Main contributions. We extended the work of [17] and provided explicit formulas for the LDT approximations of rare event probabilities for Gaussian mixture random parameters. This can be used to tackle general continuous distributions by approximating them with a Gaussian mixture. By applying these approximations to optimization problems with constraint $\mathbb{P}(F(u,\theta) \geq z) \leq \alpha$ where u is the optimal control we want to solve for and $\alpha \ll 1$ is the given risk threshold, we provided bilevel optimization reformulations that can be solved by off-the-shelf solvers. To test the performance, we conducted computational experiments from applications in portfolio management, structural engineering and fluid dynamics, illustrated the broad applicability of our method and its advantages over classical sampling-based approaches in terms of both accuracy and efficiency. One example is the short column design problem, we compared the feasibility of probabilistic constraints for the optimal solution and total computation time for our LDT-based method and sampling based methods with 10^5 samples: sample average approximation (SAA) and conditional-value-at-risk (CVaR), the sampling based methods fail to satisfy the chance constraints and require much longer time compared with our methods (Figure 3). The formulations were implemented in Julia using JuMP, and optimization problems were solved using IPOPT, while the gradients were computed by automatic differentiation in Julia.

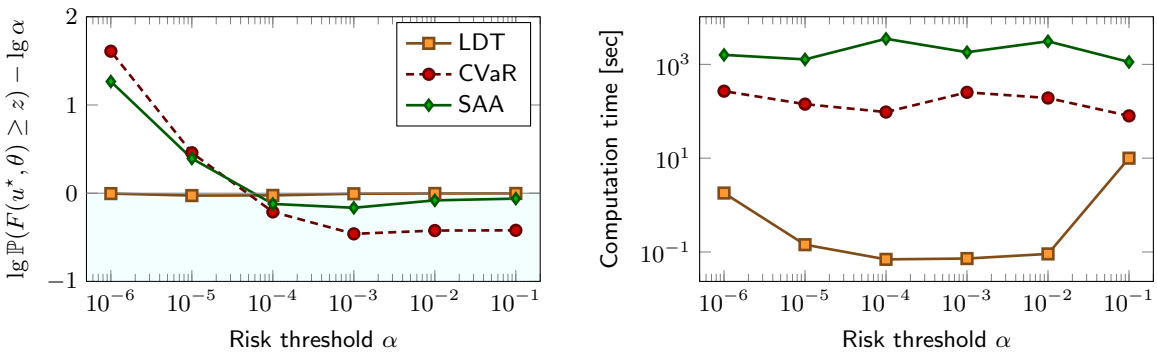

Figure 3: Comparison of LDT-based method and sampling-based methods: CVaR and SAA with 10^5 samples for solving the short column design problem under the probabilistic constraint, $\mathbb{P}(F(u,\theta) \geq z) \leq \alpha$ for different risk threshold α. Left: Feasibility of probabilistic constraint for each optimal solution $u^\star$ (falling in the cyan shaded part), the sampling-based method fails to satisfy the chance constraint for small $\alpha(\leq 10^{-5})$. Right: Computation efforts for different methods, LDT based method requires much shorter time. Figure taken from [15].

Current and future work. I plan to extend our methods by weakening some regularity assumptions, thus the method can be used for nonsmooth constrained functions, and also generalize our methods to problems with joint chance constraints. The automatic differentiation of the required derivatives of the probabilities and Hessian is still time-consuming for large-scale problems, I want to explore more efficient methods to compute these derivatives, for example, by involving randomized methods. We believe this may also spur the development of new techniques in automatic differentiation. Probabilistic constraints also provide a way to incorporate fairness constraints in optimization problems, which can be used to limit unfairness in algorithmic decision making caused by the bias in the collected real-world datasets. It provides a new direction for me to apply and refine our current methods and to build connection to machine learning and to a data training framework.

4 Computing and software

In my research, I enjoy using a mix of theoretical analysis and numerical implementation, from which I obtain a clear and concrete understanding of problems and methods. In this section, I summarize some of my computing projects and software libraries used, which provide me with a solid foundation for successful research in the future.

I took the advanced graduate-level class on *High Performance Computing* and implemented a 3D Biot Savart law on both CPUs (using OpenMP) and GPUs (using CUDA). The GPU involved further refinement by using shared memory between threads in the block and reciprocal square root directly.

The codes and implementations[1] may be ported to the modern open-source software SIMSOPT for solving stellarator optimization problems [11]. In the advanced graduate-level class: *Fast Solvers*, I implemented the fast multipole method (FMM) for computing electrostatic interactions in 2D using C++[2] and proved the linear computation cost of my implementation. In a *Deep Learning* class, I implemented the SimCLR algorithm for self-supervised learning for image classifications using PyTorch[3] and reached top 10 test accuracy among the class. Besides Python and C++, I used Julia for the probabilistic constraints problems in my summer internship at Argonne national lab and as the coding language for my teaching in the lab session of the undergrad optimization class.

My experience of using different coding languages and software libraries for numerical implementation of algorithms and methodologies brings many possibilities to my future research. The experiences of using PyTorch for deep learning training, FEniCS to solve PDEs in their weak form, MATLAB to discretize PDEs in space and time enable me to study connections between modern machine learning frameworks and the traditional numerical methods for solving PDEs and constrained optimization problems as well as sampling. I believe these are potential directions that may bring huge impact to the applied and computational fields. My experiences of using C++ and CUDA for high performance computing and Julia for optimization and automatic differentiation motivate me to think of implementing the PDE solves in a more efficient and interpretable way and utilizing these implementations in uncertainty quantification and PDE-constrained optimization.

5 Conclusion and contributions to diversity, equity and inclusion

I believe it is of vital importance to promote diversity, equity and inclusion. As a female student, I got encouraged to study maths and do research by several women-enhancing programs and kind people, and I also made my contributions to encourage classmates and students in underrepresented groups. I am a member of Association for Women in Mathematics at Courant. I mentored several incoming female PhD students, listened to their worries and problems, and tried to help and support them. I also encourage several female students to strive for further education.

In my future career, I would like to conduct research with students in underrepresented groups and mentor them, to provide them with more opportunities and to help them when they face inequality or other barriers. I also plan to be involved in the DEI group in the school and organize activities for students and professors to discuss difficulties underrepresented groups meet and what should be done to help and improve.

In conclusion, I have broad interests in computational and applied mathematics, am experienced in uncertainty quantification, optimization, inverse problems, machine learning and high performance computing. I am interested to explore more research directions and applications, and collaborate with creative researchers.

References

[1] Abhi Chowdhary, Shanyin Tong, Georg Stadler, and Alen Alexanderian. Sensitivity analysis of Bayesian inverse problem solutions. In preparation.

[2] Tiangang Cui, James Martin, Youssef M Marzouk, Antti Solonen, and Alessio Spantini. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems*, 30(11):114015, 2014.

[3] Giovanni Dematteis, Tobias Grafke, and Eric Vanden-Eijnden. Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences*, 115(5):855–860, 2018.

[4] Giovanni Dematteis, Tobias Grafke, and Eric Vanden-Eijnden. Extreme event quantification in dynamical systems with random components. *SIAM/ASA Journal on Uncertainty Quantification*, 7(3):1029–1059, 2019.

[5] Mohammad Farazmand and Themistoklis P Sapsis. Physics-based probing and prediction of extreme events. *SIAM News*, 2018.

[6] Marylou Gabrié, Grant M Rotskoff, and Eric Vanden-Eijnden. Efficient Bayesian sampling using normalizing flows to assist Markov chain Monte Carlo methods. In *ICML Workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models*, 2021.

[1] Link to the HPC project: `github.com/TongShanyin/HPC21_final`
[2] Link to the FMM project: `github.com/TongShanyin/Fastsolver-2DFMM`
[3] Link to the DL project: `github.com/TongShanyin/DL21SP_final`

[7] Mike Giles and Stefan Ulbrich. Convergence of linearized and adjoint approximations for discontinuous solutions of conservation laws. part 2: Adjoint approximations and extensions. *SIAM Journal on Numerical Analysis*, 48(3):905–921, 2010.

[8] Tobias Grafke, Tobias Schäfer, and Eric Vanden-Eijnden. Sharp asymptotic estimates for expectations, probabilities, and mean first passage times in stochastic systems with small noise. *arXiv preprint arXiv:2103.04837*, 2021.

[9] Nathan Halko, Per Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

[10] Jan S Hesthaven and Tim Warburton. *Nodal discontinuous Galerkin methods: algorithms, analysis, and applications*. Springer Science & Business Media, 2007.

[11] Matt Landreman, Bharat Medasani, Florian Wechsung, Andrew Giuliani, Rogerio Jorge, and Caoxiang Zhu. SIMSOPT: A flexible framework for stellarator optimization. *Journal of Open Source Software*, 6(65):3525, 2021.

[12] Rüdiger Rackwitz. Reliability analysis – a review and some perspectives. *Structural Safety*, 23(4):365–395, 2001.

[13] Shanyin Tong. Extreme event probability estimation with application to tsunamis. *SIAM News*, 2021.

[14] Shanyin Tong and Georg Stadler. LDT-based reduced dimension important sampling for rare events. *In preparation*.

[15] Shanyin Tong, Anirudh Subramanyam, and Vishwas Rao. Optimization under rare chance constraints. *arXiv preprint arXiv:2011.06052, submitted to SIAM Journal on Optimization*, 2020.

[16] Shanyin Tong, Eric Vanden-Eijnden, and Georg Stadler. Estimating earthquake-induced tsunami inundation probabilities without sampling. *arXiv preprint arXiv:2111.14325, submitted to Geophysics Journal International*, 2021.

[17] Shanyin Tong, Eric Vanden-Eijnden, and Georg Stadler. Extreme event probability estimation using PDE-constrained optimization and large deviation theory, with application to tsunamis. *Communications in Applied Mathematics and Computational Science*, 16(2):181–225, 2021.

[18] Felipe Uribe, Iason Papaioannou, Youssef M Marzouk, and Daniel Straub. Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):818–847, 2021.

[19] Amy L Williamson, Donsub Rim, Loyce M Adams, Randall J LeVeque, Diego Melgar, and Frank I González. A source clustering approach for efficient inundation modeling and regional scale probabilistic tsunami hazard assessment. *Frontiers in Earth Science*, page 442, 2020.