Large deviation theory-based adaptive importance sampling for rare events in high dimensions

Shanyin Tong[1]

jointly with: Georg Stadler[2]

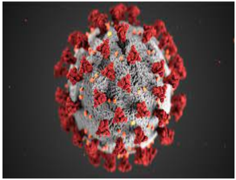
[1]Department of Applied Physics and Applied Mathematics, Columbia University, NY
[2]Courant Institute of Mathematical Sciences, New York University, NY

Feb 29, 2024
SIAM Conference on Uncertainty Quantification (UQ24), Trieste, Italy

Registration and travel support for this presentation was provided by National Science Foundation.

Outline

What are rare events?



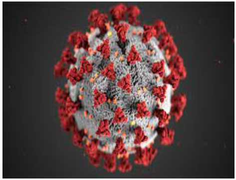
Examples:

- Material failure (bridge/tool/plane stress fractures)
- Extreme weather (tornadoes, hurricanes, heat waves)
- Rogue waves, tsunamis, earthquakes
- Financial sector/bank/company collapse
- Pandemics

Why study rare events?







Common to all these:

- Rare but high impact
- Prob. 10^{-3} or 10^{-7}: big difference
- Long simulation time
- Control and mitigation

$\Longrightarrow$

We need:

- accurate
- efficient

probability estimation

Rare event probability estimation

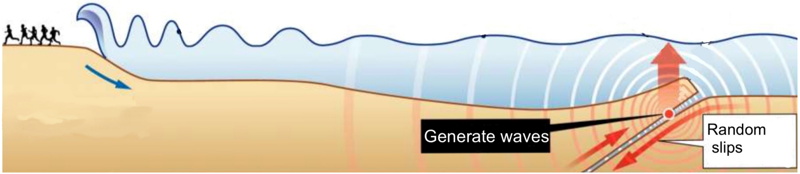
$$F : \boldsymbol{\theta} \in \mathbb{R}^n \longrightarrow \mathbb{R}$$

- $\boldsymbol{\theta}$: random parameter with PDF π_{pr} (high-dimensional)
- F: parameter-to-event map (involve PDE solves)

Rare event probability estimation

$$F : \boldsymbol{\theta} \in \mathbb{R}^n \longrightarrow \mathbb{R}$$

- $\boldsymbol{\theta}$: random parameter with PDF π_{pr} (high-dimensional) random slips
- F: parameter-to-event map (involve PDE solves) observed wave height



Shallow water equations

$$h_t + v_x = 0$$

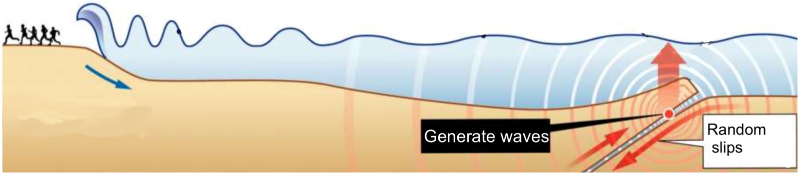
$$v_t + \left(\frac{v^2}{h} + \frac{1}{2} g h^2 \right)_x + g h B_x = 0$$

$$h(x,0) = -B_0(x), v(x,0) = 0$$

Rare event probability estimation

$$F : \boldsymbol{\theta} \in \mathbb{R}^n \longrightarrow \mathbb{R}$$

- $\boldsymbol{\theta}$: random parameter with PDF π_{pr} (high-dimensional) random slips
- F: parameter-to-event map (involve PDE solves) observed wave height



Shallow water equations

$$h_t + v_x = 0$$

$$v_t + \left(\frac{v^2}{h} + \frac{1}{2}gh^2\right)_x + ghB_x = 0$$

$$h(x,0) = -B_0(x), v(x,0) = 0$$

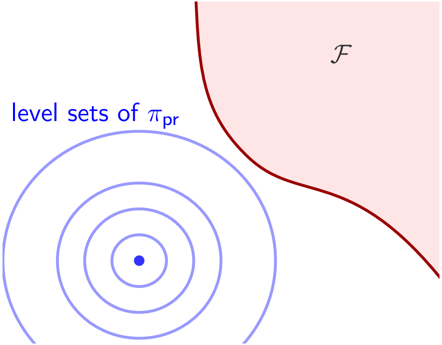
Target: Probability of wave height ≥ a given threshold z:

$$\mathbb{P}(F(\boldsymbol{\theta}) \geq z)$$

Why are rare events difficult to estimate?

$$p_{\mathcal{F}} := \mathbb{P}(F(\boldsymbol{\theta}) \geq z)$$

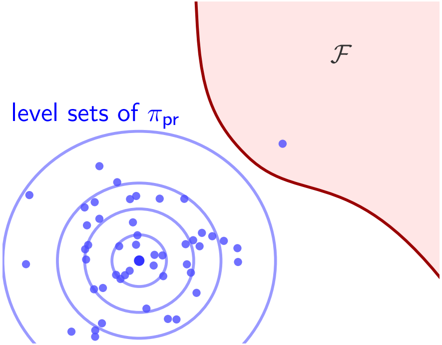
- Measure of rare event set $\mathcal{F} := \{\boldsymbol{\theta} : F(\boldsymbol{\theta}) \geq z\}$



Why are rare events difficult to estimate?

$$p_{\mathcal{F}} := \mathbb{P}(F(\boldsymbol{\theta}) \geq z) \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i), \text{ with samples } \{\boldsymbol{\theta}_i\}_{i=1}^{N} \overset{\text{i.i.d.}}{\sim} \pi_{\text{pr}}$$

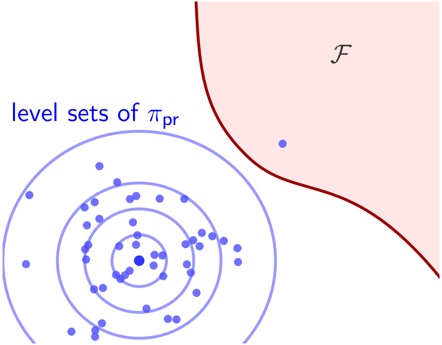
- Measure of rare event set $\mathcal{F} := \{\boldsymbol{\theta} : F(\boldsymbol{\theta}) \geq z\}$
- Standard Monte Carlo methods $\Longrightarrow$ fail when $p_{\mathcal{F}} \ll 1$



Why are rare events difficult to estimate?

$$p_{\mathcal{F}} := \mathbb{P}(F(\boldsymbol{\theta}) \geq z) \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i), \text{ with samples } \{\boldsymbol{\theta}_i\}_{i=1}^{N} \overset{\text{i.i.d.}}{\sim} \pi_{\text{pr}}$$

- Measure of rare event set $\mathcal{F} := \{\boldsymbol{\theta} : F(\boldsymbol{\theta}) \geq z\}$
- Standard Monte Carlo methods $\Longrightarrow$ fail when $p_{\mathcal{F}} \ll 1$



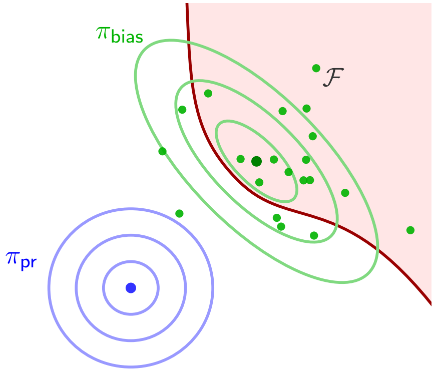
Main Challenges:

I. **Large sample size**

II. **High-dim random space**

III. **Expensive eval**

Methods for rare event estimation



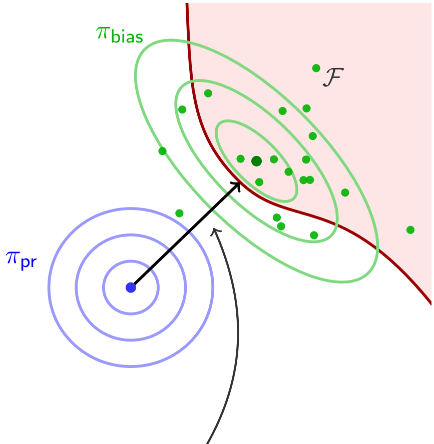
Sampling:

- Importance sampling, adaptive importance sampling, cross-entropy [Kahn, Marshall, Rubinstein, . . .]
- Sequential sampling, splitting [Au, Beck, Papaioannou, . . .]

Asymptotic approximation:

- Reliability methods [Breitung, Hasofer, Lind, . . .]
- Large deviation theory [Grafke, Vanden-Eijnden, . . .]

Methods for rare event estimation



Our contribution:
Adapt LDT to inform IS & improve efficiency

Sampling:

- Importance sampling, adaptive importance sampling, cross-entropy [Kahn, Marshall, Rubinstein, . . .]
- Sequential sampling, splitting [Au, Beck, Papaioannou, . . .]

Asymptotic approximation:

- Reliability methods [Breitung, Hasofer, Lind, . . .]
- Large deviation theory [Grafke, Vanden-Eijnden, . . .]

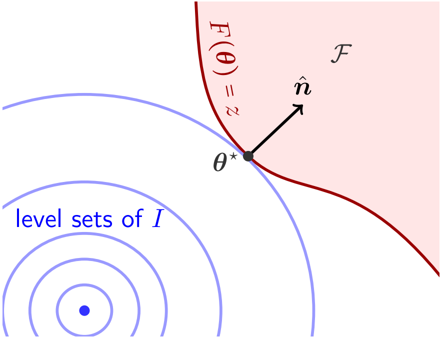
Outline

I. Large deviation theory and optimization

Large deviation theory [Tong, Vanden-Eijnden, Stadler (2021)]

Under regularity assumptions on F and rate function I (depends only on π_{pr}),

$$p_{\mathcal{F}} \approx C(z)\exp(-I(\boldsymbol{\theta}^\star)) \text{ as } z \to \infty, \quad \boldsymbol{\theta}^\star := \operatorname*{argmin}_{\boldsymbol{\theta}\in\mathcal{F}} I(\boldsymbol{\theta}).$$



- $\log$(decay rate of $p_{\mathcal{F}}$) is $I(\boldsymbol{\theta}^\star)$
- $C(z)$: sub-exp. prefactor
- $I(\boldsymbol{\theta})$: Legendre transform of cumulant generating fct of $\boldsymbol{\theta}$

$$I(\boldsymbol{\theta}) = \sup_{\eta} \eta^\top \boldsymbol{\theta} - \log \mathbb{E}_{\pi_{\text{pr}}}[\exp(\eta^\top \boldsymbol{\theta})]$$

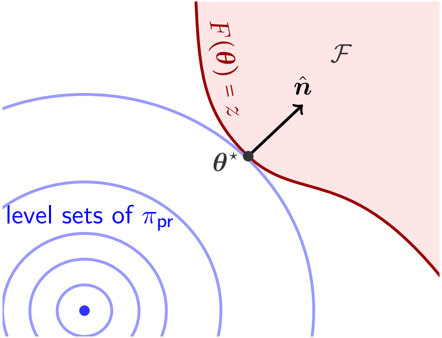
- $I(\boldsymbol{\theta}) = \frac{1}{2}\|\boldsymbol{\theta}\|^2$ for Gaussian $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$

I. Large deviation theory and optimization

Large deviation theory [Tong, Vanden-Eijnden, Stadler (2021)]

Under regularity assumptions on F and rate function I (depends only on π_{pr}),

$$p_{\mathcal{F}} \approx C(z)\exp(-I(\boldsymbol{\theta}^\star)) \text{ as } z \to \infty, \quad \boldsymbol{\theta}^\star := \operatorname*{argmin}_{\boldsymbol{\theta}\in\mathcal{F}} I(\boldsymbol{\theta}).$$



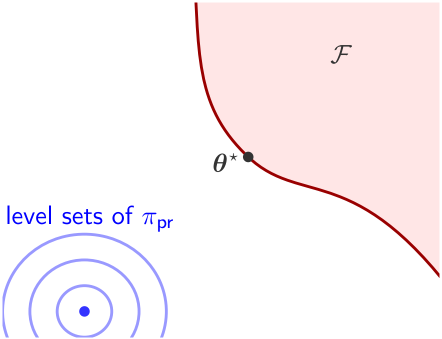
$\boldsymbol{\theta}^\star$ dominates $p_{\mathcal{F}}$ in $\mathcal{F}$

- LDT optimizer
- Reliability: most probable/design point
- Bayesian: MAP point
- SDE/SPDE: instanton/saddle point

I. Large deviation theory and optimization

LDT optimization for Gaussian $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$

$$\boldsymbol{\theta}^\star := \operatorname*{argmin}_{\boldsymbol{\theta} \in \mathcal{F}} \frac{1}{2}\|\boldsymbol{\theta}\|^2, \quad \mathcal{F} = \{\boldsymbol{\theta} : F(\boldsymbol{\theta}) \geq z\}$$

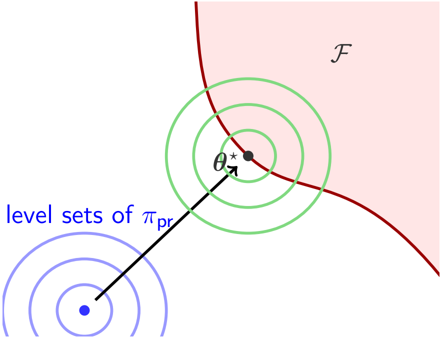


- (PDE-) constrained optimization problem
- Solved by adjoint method
- Mass concentrate near $\boldsymbol{\theta}^\star$

I. Large deviation theory and optimization

LDT optimization for Gaussian $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$

$$\boldsymbol{\theta}^\star := \operatorname*{argmin}_{\boldsymbol{\theta} \in \mathcal{F}} \frac{1}{2} \|\boldsymbol{\theta}\|^2, \quad \mathcal{F} = \{\boldsymbol{\theta} : F(\boldsymbol{\theta}) \geq z\}$$

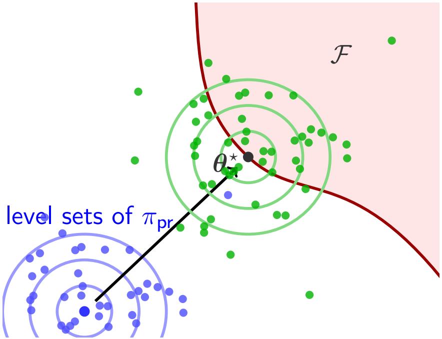


- (PDE-) constrained optimization problem
- Solved by adjoint method
- Mass concentrate near $\boldsymbol{\theta}^\star$
- Good initial biasing density $\mathcal{N}(\boldsymbol{\theta}^\star, \boldsymbol{I}_n)$

I. Large deviation theory and optimization

LDT optimization for Gaussian $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$

$$\boldsymbol{\theta}^\star := \underset{\boldsymbol{\theta} \in \mathcal{F}}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{\theta}\|^2, \quad \mathcal{F} = \{\boldsymbol{\theta} : F(\boldsymbol{\theta}) \geq z\}$$

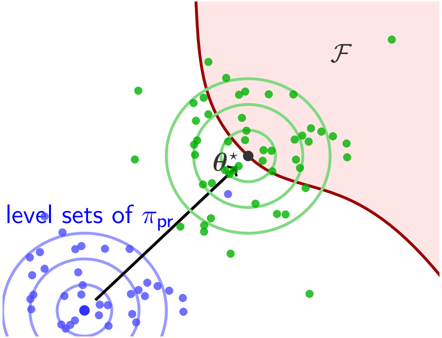


- (PDE-) constrained optimization problem
- Solved by adjoint method
- Mass concentrate near $\boldsymbol{\theta}^\star$
- Good initial biasing density $\mathcal{N}(\boldsymbol{\theta}^\star, \boldsymbol{I}_n)$
- Remove exponential term in variance

I. Large deviation theory and optimization

LDT optimization for Gaussian $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$

$$\boldsymbol{\theta}^\star := \underset{\boldsymbol{\theta} \in \mathcal{F}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{\theta}\|^2, \quad \mathcal{F} = \{\boldsymbol{\theta} : F(\boldsymbol{\theta}) \geq z\}$$



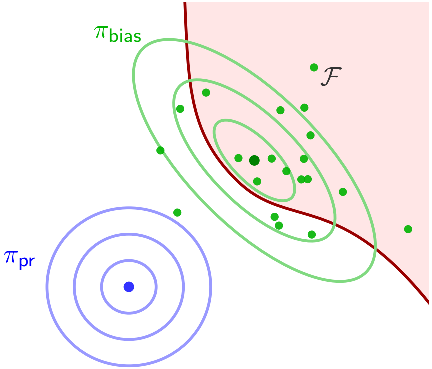
- (PDE-) constrained optimization problem
- Solved by adjoint method
- Mass concentrate near $\boldsymbol{\theta}^\star$
- Good initial biasing density $\mathcal{N}(\boldsymbol{\theta}^\star, \boldsymbol{I}_n)$
- Remove exponential term in variance

Challege I completed: Find a good initialization to reduce sample size!

What is the optimal biasing density?

Importance sampling (IS): sample from a biasing density $\{\boldsymbol{\theta}_i\}_{i=1}^N \overset{\text{i.i.d.}}{\sim} \pi_{\text{bias}}$

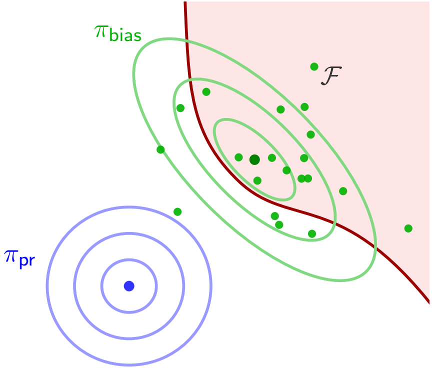
$$p_{\mathcal{F}} \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i) \text{ with } w(\boldsymbol{\theta}_i) = \frac{\pi_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi_{\text{bias}}(\boldsymbol{\theta}_i)}$$



What is the optimal biasing density?

Importance sampling (IS): sample from a biasing density $\{\boldsymbol{\theta}_i\}_{i=1}^N \overset{\text{i.i.d.}}{\sim} \pi_{\text{bias}}$

$$p_{\mathcal{F}} \approx \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i) \text{ with } w(\boldsymbol{\theta}_i) = \frac{\pi_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi_{\text{bias}}(\boldsymbol{\theta}_i)}$$



- Optimal biasing density $\pi_{\mathcal{F}}$

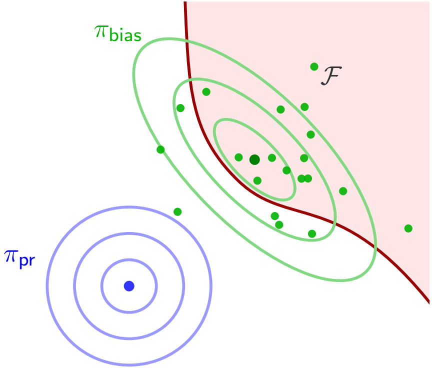
$$\pi_{\mathcal{F}}(\boldsymbol{\theta}) := \frac{1}{p_{\mathcal{F}}} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) \pi_{\text{pr}}(\boldsymbol{\theta})$$

"posterior" if view $\mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta})$ as likelihood

What is the optimal biasing density?

Importance sampling (IS): sample from a biasing density $\{\boldsymbol{\theta}_i\}_{i=1}^N \overset{\text{i.i.d.}}{\sim} \pi_{\text{bias}}$

$$p_{\mathcal{F}} \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i) \text{ with } w(\boldsymbol{\theta}_i) = \frac{\pi_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi_{\text{bias}}(\boldsymbol{\theta}_i)}$$



- Optimal biasing density $\pi_{\mathcal{F}}$

$$\pi_{\mathcal{F}}(\boldsymbol{\theta}) := \frac{1}{p_{\mathcal{F}}} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) \pi_{\text{pr}}(\boldsymbol{\theta})$$

  "posterior" if view $\mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta})$ as likelihood
- $p_{\mathcal{F}}$ is unknown (target) $\Longrightarrow$ impractical
- Approximate $\pi_{\mathcal{F}}$ with Gaussian

Cross-entropy method

Cross-entropy method (CE):

- min Kullback-Leibler divergence between $\pi_{\mathcal{F}}$ and Gaussian $\pi_{\text{bias}}(\cdot;\boldsymbol{v})$
- use evaluations from previous $\boldsymbol{\theta}_i \sim \pi_{\text{bias}}(\cdot;\boldsymbol{v}')$ (iterative)

Optimal mean $\boldsymbol{\mu}^\star$ and covariance $\boldsymbol{\Sigma}^\star$ are high-dimensional:

$$\boldsymbol{\mu}^\star = \frac{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i;\boldsymbol{v}')\boldsymbol{\theta}_i}{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i;\boldsymbol{v}')}, \quad \boldsymbol{\Sigma}^\star = \frac{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i;\boldsymbol{v}')(\boldsymbol{\theta}_i - \boldsymbol{\mu}^\star)(\boldsymbol{\theta}_i - \boldsymbol{\mu}^\star)^\top}{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i;\boldsymbol{v}')}$$

- Large sample size required for estimating optimal parameter

Cross-entropy method

Cross-entropy method (CE):

- min Kullback-Leibler divergence between $\pi_{\mathcal{F}}$ and Gaussian $\pi_{\text{bias}}(\cdot; \boldsymbol{v})$
- use evaluations from previous $\boldsymbol{\theta}_i \sim \pi_{\text{bias}}(\cdot; \boldsymbol{v}')$ (iterative)

Optimal mean $\boldsymbol{\mu}^\star$ and covariance $\boldsymbol{\Sigma}^\star$ are high-dimensional:

$$\boldsymbol{\mu}^\star = \frac{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i; \boldsymbol{v}') \boldsymbol{\theta}_i}{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i; \boldsymbol{v}')}, \quad \boldsymbol{\Sigma}^\star = \frac{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i; \boldsymbol{v}') (\boldsymbol{\theta}_i - \boldsymbol{\mu}^\star)(\boldsymbol{\theta}_i - \boldsymbol{\mu}^\star)^\top}{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i; \boldsymbol{v}')}$$

- Large sample size required for estimating optimal parameter
- Intrinsic low-dim $\Longrightarrow$ No need of changes in every direction

Cross-entropy method

Cross-entropy method (CE):

- min Kullback-Leibler divergence between $\pi_{\mathcal{F}}$ and Gaussian $\pi_{\text{bias}}(\cdot; \boldsymbol{v})$
- use evaluations from previous $\boldsymbol{\theta}_i \sim \pi_{\text{bias}}(\cdot; \boldsymbol{v}')$ (iterative)

Optimal mean $\boldsymbol{\mu}^\star$ and covariance $\boldsymbol{\Sigma}^\star$ are high-dimensional:

$$\boldsymbol{\mu}^\star = \frac{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i; \boldsymbol{v}')\boldsymbol{\theta}_i}{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i; \boldsymbol{v}')}, \; \boldsymbol{\Sigma}^\star = \frac{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i; \boldsymbol{v}')(\boldsymbol{\theta}_i - \boldsymbol{\mu}^\star)(\boldsymbol{\theta}_i - \boldsymbol{\mu}^\star)^\top}{\sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i; \boldsymbol{v}')}$$

- Large sample size required for estimating optimal parameter
- Intrinsic low-dim $\Longrightarrow$ No need of changes in every direction

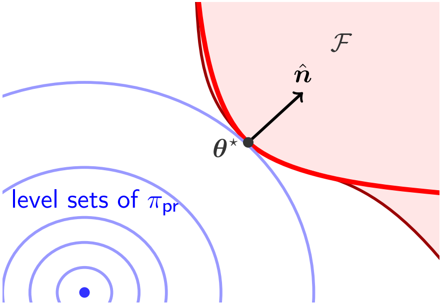
Our contributions:

- Build a low-dim subspace using local info of $\boldsymbol{\theta}^\star$
- Reweigh samples and evaluations to save costs

II. LDT-informed dimension reduction

Low-dimensional subspace includes

- **normal direction** $\hat{\boldsymbol{n}} = \nabla F(\boldsymbol{\theta}^\star)/\|\nabla F(\boldsymbol{\theta}^\star)\|$
- **dominating eigvec** of $\boldsymbol{H}_{\mathrm{LDT}} := (\boldsymbol{I}_n - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^\top)\nabla^2 F(\boldsymbol{\theta}^\star)(\boldsymbol{I}_n - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^\top)$ (large curvatures)



- Inspired by asymptotic approx/SORM:

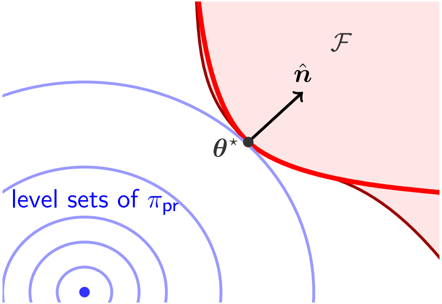
$$C(z) \approx \left[2\pi\|\boldsymbol{\theta}^\star\|^2 \prod_{i=1}^{n} \left(1 - \lambda\lambda_i\left(\boldsymbol{H}_{\mathrm{LDT}}\right)\right)\right]^{-\frac{1}{2}}$$

- $\lambda := \|\nabla I(\boldsymbol{\theta}^\star)\|/\|\nabla F(\boldsymbol{\theta}^\star)\|$
- $\lambda_i(\cdot)$: ith eigval
- Coincides with local likelihood-informed subspace from Bayesian inversion

II. LDT-informed dimension reduction

Low-dimensional subspace includes

- **normal direction** $\hat{\boldsymbol{n}} = \nabla F(\boldsymbol{\theta}^\star)/\|\nabla F(\boldsymbol{\theta}^\star)\|$
- **dominating eigvec** of $\boldsymbol{H}_{\text{LDT}} := (\boldsymbol{I}_n - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^\top)\nabla^2 F(\boldsymbol{\theta}^\star)(\boldsymbol{I}_n - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^\top)$ (large curvatures)

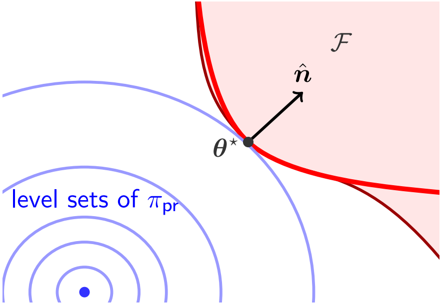


- Randomized SVD or Lanczos
- Only need Hessian-applies
- Apply CE only in this low-dim subspace
- Use π_{pr} in remaining orthogonal space

II. LDT-informed dimension reduction

Low-dimensional subspace includes

- **normal direction** $\hat{\boldsymbol{n}} = \nabla F(\boldsymbol{\theta}^\star)/\|\nabla F(\boldsymbol{\theta}^\star)\|$
- **dominating eigvec** of $\boldsymbol{H}_{\mathrm{LDT}} := (\boldsymbol{I}_n - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^\top)\nabla^2 F(\boldsymbol{\theta}^\star)(\boldsymbol{I}_n - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^\top)$ (large curvatures)



- Randomized SVD or Lanczos
- Only need Hessian-applies
- Apply CE only in this low-dim subspace
- Use π_{pr} in remaining orthogonal space

Challenge II completed: Build a low-dim subspace to tackle high dim!

III. Multiple importance sampling

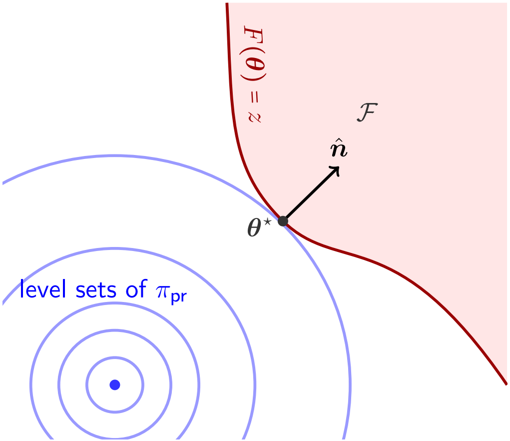
MIS: Reweigh all previous samples and evaluations (as from a mixture distribution)

$$p_{\mathcal{F}} \approx \frac{1}{JM} \sum_{j=1}^{J} \sum_{i=1}^{M} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i^{(j)}) w_{\text{MIS}}(\boldsymbol{\theta}_i^{(j)})$$

Recompute weights: $w_{\text{MIS}}(\boldsymbol{\theta}_i^{(j)}) := \dfrac{J}{\sum_{j'=1}^{J} 1/w_{j'}(\boldsymbol{\theta}_i^{(j)})}$

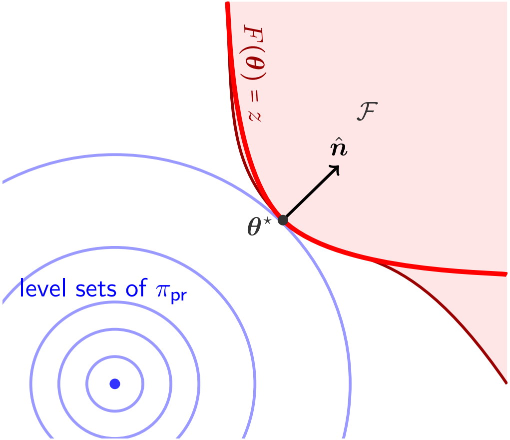
Challenge III completed: MIS to reuse evaluations and save costs!

Summary of LDT-based adaptive IS (LAIS)



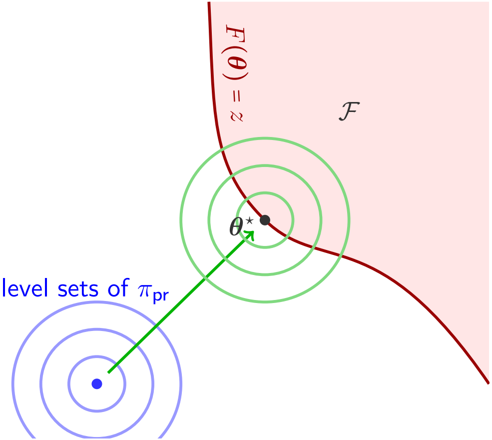
1. LDT optimization $\Rightarrow \boldsymbol{\theta}^\star$

Summary of LDT-based adaptive IS (LAIS)



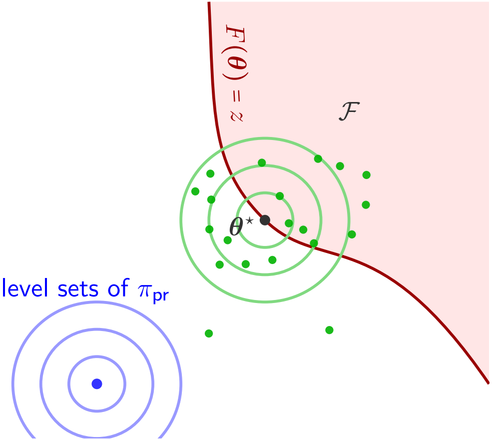
1. LDT optimization $\Rightarrow \boldsymbol{\theta}^\star$
2. Build low-dim subspace using local derivatives

Summary of LDT-based adaptive IS (LAIS)



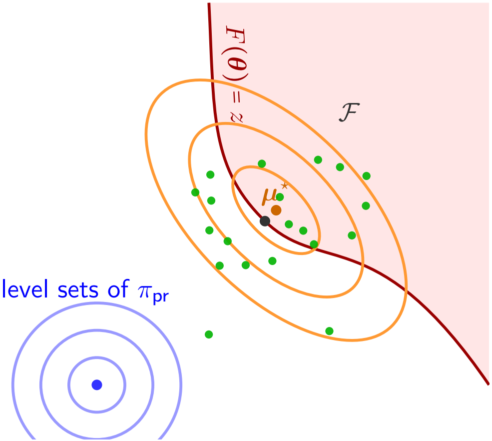
1. LDT optimization $\Rightarrow \boldsymbol{\theta}^\star$
2. Build low-dim subspace using local derivatives
3. Adaptive IS: initialization

Summary of LDT-based adaptive IS (LAIS)



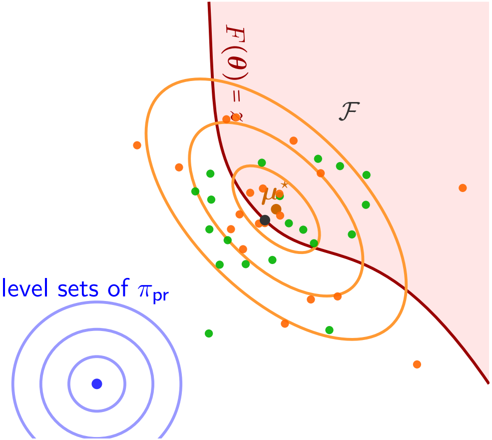
1. LDT optimization $\Rightarrow \boldsymbol{\theta}^\star$
2. Build low-dim subspace using local derivatives
3. Adaptive IS: initialization
   3.1 Sample from current biasing density

Summary of LDT-based adaptive IS (LAIS)



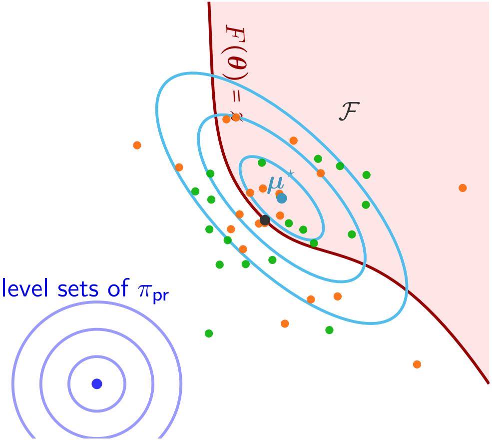
1. LDT optimization $\Rightarrow \boldsymbol{\theta}^\star$
2. Build low-dim subspace using local derivatives
3. Adaptive IS: initialization
   3.1 Sample from current biasing density
   3.2 CE to update $\boldsymbol{\mu}^\star$ and $\boldsymbol{\Sigma}^\star$

Summary of LDT-based adaptive IS (LAIS)



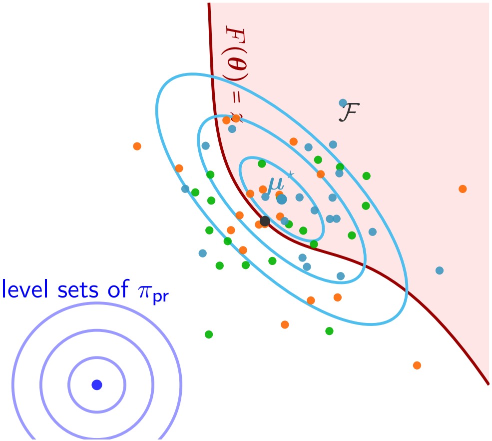
1. LDT optimization $\Rightarrow \boldsymbol{\theta}^\star$
2. Build low-dim subspace using local derivatives
3. Adaptive IS: initialization
   - 3.1 Sample from current biasing density
   - 3.2 CE to update $\boldsymbol{\mu}^\star$ and $\boldsymbol{\Sigma}^\star$
   - 3.3 Repeat

Summary of LDT-based adaptive IS (LAIS)



1. LDT optimization $\Rightarrow \boldsymbol{\theta}^\star$
2. Build low-dim subspace using local derivatives
3. Adaptive IS: initialization
   3.1 Sample from current biasing density
   3.2 CE to update $\boldsymbol{\mu}^\star$ and $\boldsymbol{\Sigma}^\star$, reuse samples
   3.3 Repeat

Summary of LDT-based adaptive IS (LAIS)



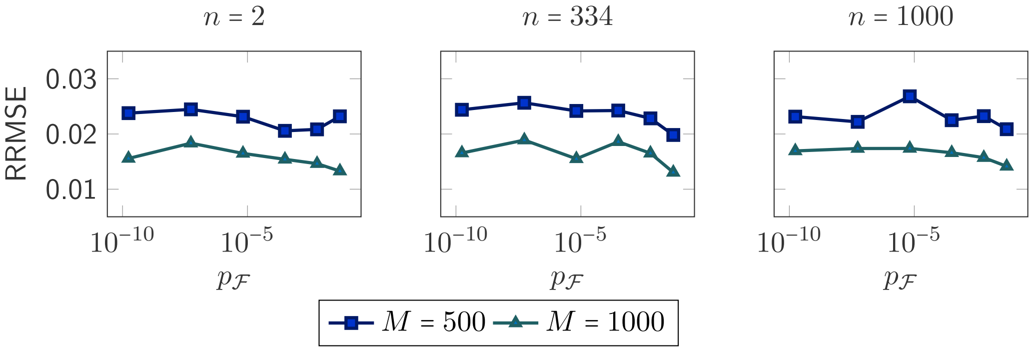
1. LDT optimization $\Rightarrow \boldsymbol{\theta}^\star$
2. Build low-dim subspace using local derivatives
3. Adaptive IS: initialization
   - 3.1 Sample from current biasing density
   - 3.2 CE to update $\boldsymbol{\mu}^\star$ and $\boldsymbol{\Sigma}^\star$, reuse samples
   - 3.3 Repeat
4. Estimate $p_{\mathcal{F}}$ using MIS and all previous eval

Outline

Quadratic map

$$F(\boldsymbol{\theta}) := -\frac{5}{4}(\theta_1 - \theta_2)^2 + \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\theta_k, \quad \boldsymbol{\theta} = [\theta_1, \ldots, \theta_n], \quad p_{\mathcal{F}} := \mathbb{P}(F(\boldsymbol{\theta}) \geq z)$$

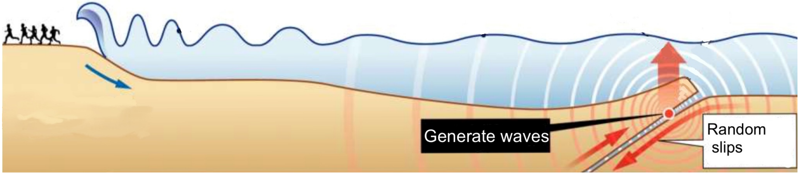
Relative root-mean-square error after 5 iterations, M samples per iteration



LAIS: insensitive to dimension & target probability $p_{\mathcal{F}}$

Extreme tsunami estimation

Target: Estimate $\mathbb{P}(\text{tsunami height} \geq z[\text{m}])$



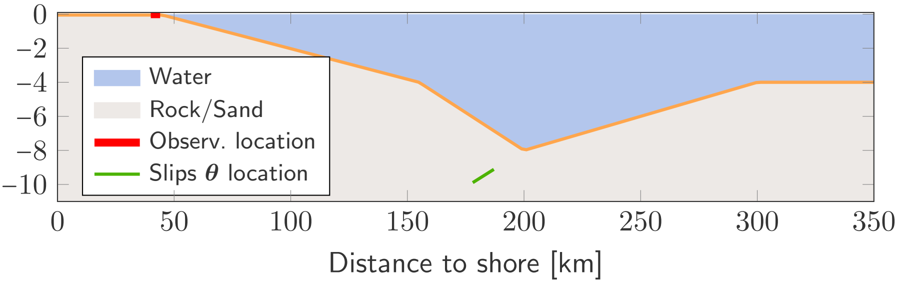
Shallow water equations

$$h_t + v_x = 0$$

$$v_t + \left(\frac{v^2}{h} + \frac{1}{2}gh^2\right)_x + ghB_x = 0$$

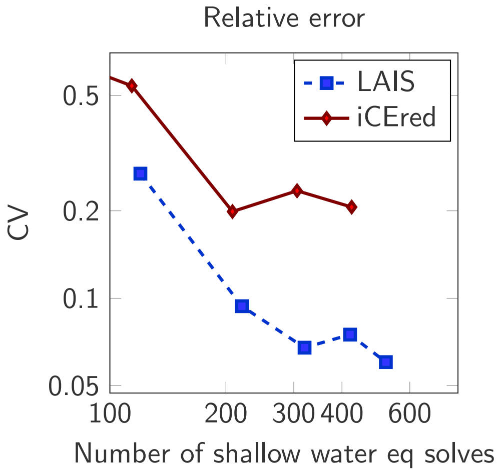
$$h(x,0) = -B_0(x), v(x,0) = 0$$

2011 Tohoku-Oki earthquake and tsunami



- Gaussian slips $\boldsymbol{\theta}$
- Bathymetry $B = O\boldsymbol{\theta} + B_0$
- Tsunami $F(\boldsymbol{\theta}) = \max_{[0,T]} \fint (h + B_0)\, dx$

Extreme tsunami estimation – Comparison with iCEred[1]



- $p_{\mathcal{F}} \approx 10^{-4}$
- each PDE solve: ≈ 10min
- MC (10% error): 10^6 SWE solves

Number of adjoint (extra) PDE solves:

- LAIS: $O(1) \approx 20$
- iCEred: $O(N)$

[1] Uribe, Felipe, Iason Papaioannou, Youssef M. Marzouk, and Daniel Straub, *Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation.* JUQ (2021)

Outline

Main takeaways

LAIS: Efficient for rare event probability estimation

- insensitive to extremeness and dimensions
- works for high-dim random space and expensive evaluations
- Connection to reliability methods and Bayesian inverse problems

Main takeaways

LAIS: Efficient for rare event probability estimation

- insensitive to extremeness and dimensions
- works for high-dim random space and expensive evaluations
- Connection to reliability methods and Bayesian inverse problems

Extensions:

- Infinite dim: SDE/SPDEs [Statistics and Computing (2023)]
  MS 223: Friday 5:30-5:55PM, Timo Schorlepp: Scalable Methods for Computing Sharp Extreme Event Probabilities in Infinite-Dimensional Stochastic Systems
- Non-Gaussian: triangular maps/normalizing flows/diffusion models (ongoing)
- Risk control: optimization under rare chance constraints [SIOPT (2022)]
- Digital twin: MS169: Thursday 5:30-5:55PM, Georg Stadler: Rare Event Estimation in Complex Physics Models

References

- S. Tong and G. Stadler, *Large deviation theory-based adaptive importance sampling for rare events in high dimensions*, SIAM/ASA Journal on Uncertainty Quantification 11.3 (2023)
- T. Schorlepp, S. Tong, T. Grafke and G. Stadler, *Scalable methods for computing sharp extreme event probabilities in infinite-dimensional stochastic systems*, Statistics and Computing (2023)
- S. Tong, E. Vanden-Eijnden and G. Stadler, *Estimating earthquake-induced tsunami height probabilities without sampling*, Pure and Applied Geophysics (2023)
- S. Tong, A. Subramanyam and V. Rao, *Optimization under rare chance constraints*, SIAM Journal on Optimization 32.2 (2022)
- S. Tong, E. Vanden-Eijnden and G. Stadler, *Extreme event probability estimation using PDE-constrained optimization and large deviation theory, with application to tsunamis*, CAMCoS (2021)