Estimation of extreme event probabilities in complex systems governed by PDEs

Georg Stadler

jointly with: Shanyin Tong Eric Vanden Eijnden

Courant Institute of Mathematical Sciences, New York University, NY

December 1, 2021

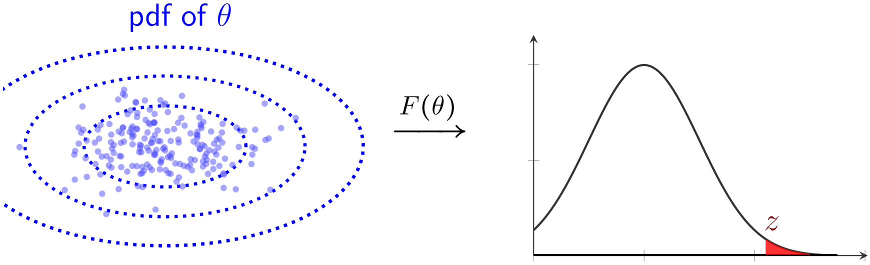
Supported by NSF under DMS-1723211, and DOE SciDAC.

Outline

Extreme event probability estimation

$$F : \theta \in (X, \mu) \to \mathbb{R}$$

- $(X, \mu) \subset \mathbb{R}^n \ldots$ random space of parameters θ with measure μ
- $F \ldots$ parameter-to-event map (involves PDE solve)



Target: Estimate the measure of extreme event set for $z \gg 0$:

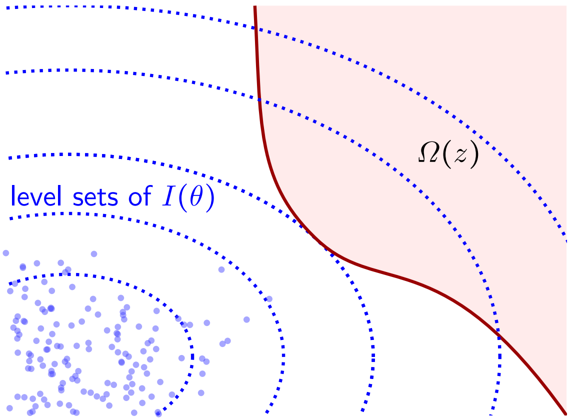
$$\Omega(z) := \{\theta : F(\theta) \geq z\}, \quad \text{i.e., compute} \quad \mathbb{P}(F(\theta) \geq z).$$

Why low probability events are difficult to estimate

We're interested in

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

where z is "large", i.e., the probability is small.



- The more extreme/rare the event, the more difficult it is to characterize its probability
- Even more so if computing $F(\theta)$ is expensive and θ is high-dimensional

Estimation of extreme/rare but important events

Examples:

- Material failure (e.g., bridge/tool/plane stress fractures)
- Extreme weather patterns (e.g., tornados, hurricanes, heat waves)
- Rogue waves, tsunamis, extreme turbulence behavior, earthquakes
- financial sector/bank/company collapse

Common to all these:

- Rare but high impact/cost/damage
- It makes a big difference if the probability of an event is $p = 10^{-3}$ or $p = 10^{-7}$!

Main application involves tsunamis, but approach is generic.

Probability Estimation and Optimization

- The maximum likelihood point of a density is usually statistically not important
- Optimization typically feasible in high dimensions
- High-dimensional statistics suffers from curse of dimensionality
- When can optimization ideas be used in probability estimation?

Main takeaways

- Methods for extreme/rare event estimation that are:
  insensitive to extremeness,
  work for high-dimensional parameters and
  expensive parameter-to-event maps F
- Theoretical foundation: large-deviation theory to argue connection between (PDE-constrained) optimization and extreme event estimation
- Nonlinear parameter-to-event map F where derivatives are available (using adjoints); parameters θ Gaussian or non-Gaussian
- Illustration using tsunami problem, where F involves time-dependent shallow water equation

Outline

Maths intro: probability & optimization

Physics intro: tsunamis

Extreme events and optimization

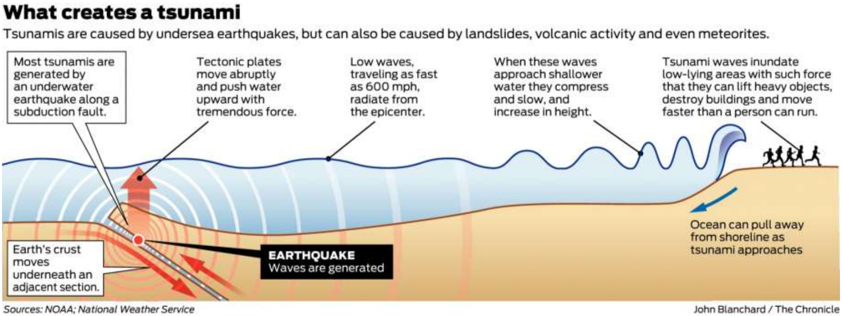
Probability approximations using sampling

Probability approximations using approximation of $\Omega(z)$

Aspects of PDE-constrained optimization

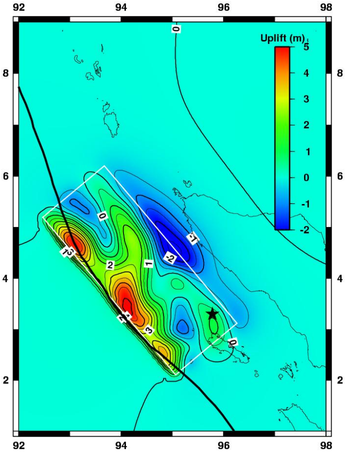
Physical setup for Tohoku earthquake

What causes tsunamis?



- tsunamis caused by sudden up/downlift of the ocean floor due to earthquake rupture (change of bathymetry)
- resulting waves can travel hundreds of km and lead to severe flooding

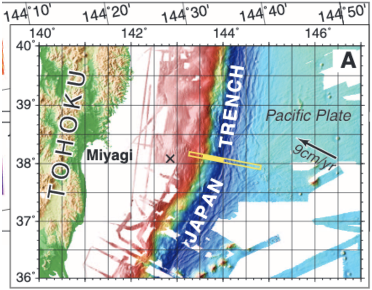
What causes tsunamis?

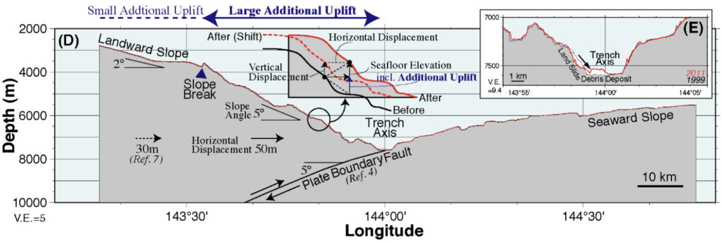


Sumatra 2004 earthquake up/downlift; Image courtesy by Chen Ji (Caltech)

- Bathymetry change is smooth but can be complicated
- . . . cannot be predicted, depends on slip at fault

Tohoku earthquake/tsunami problem



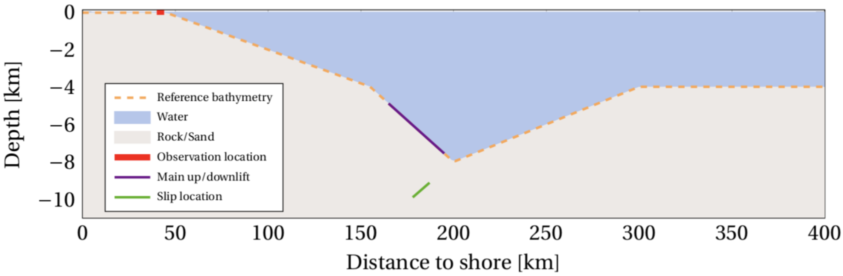


Modeling tsunamis: the shallow water equations

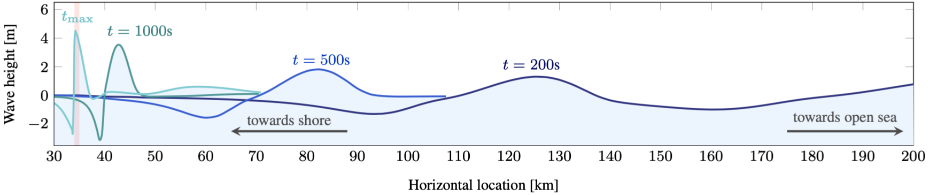
1-D, inviscit shallow water equations with bathymetry:

$$h_t + v_x = 0,$$
$$v_t + \left(\frac{v^2}{h} + \frac{1}{2}gh^2\right)_x + ghB_x = 0,$$
$$h(x,0) = -B_0(x), v(x,0) = 0,$$
$$+ \text{ bdry. cond.}$$

h :	water height
u :	velocity
v :	momentum hu
g :	gravity constant
B_0 :	reference bathymetry
B :	changed bathymetry



Modeling tsunamis: the shallow water equations



Measure added water volume in interval $[a, b]$ near shore, B is updated bathymetry:

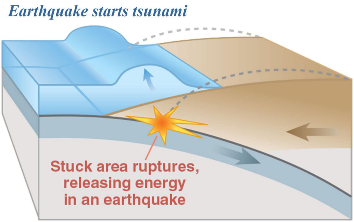
$$F(B) = F(B; h, v) = \max_{[0,T]} \fint_c^d (h + B_0)\, dx.$$

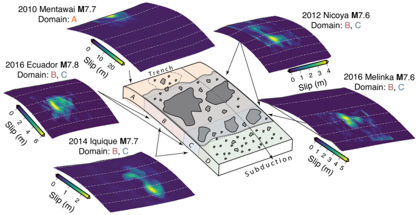
or smoothed version with $\gamma > 0$:

$$F_\gamma(B) = F_\gamma(B; h, v) = \gamma \log\left[\frac{1}{T}\int_0^T \exp\left(\frac{1}{\gamma}\fint_c^d (h + B_0)\, dx\right) dt\right].$$

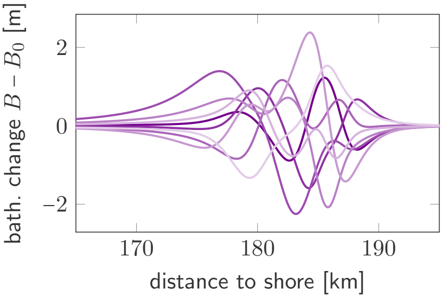
Distribution for random parameter θ (or B)

Target: Realistic model of earthquake-induced ocean floor up/downlift





- Slip patterns (under ocean floor) are unpredictable (modeled as random iid Gaussian variable)
- Slip leads to floor up/downlift (using Okada model); Gaussian random field



Outline

Maths intro: probability & optimization

Physics intro: tsunamis

Extreme events and optimization

Probability approximations using sampling

Probability approximations using approximation of $\Omega(z)$

Aspects of PDE-constrained optimization

Physical setup for Tohoku earthquake

Extreme tsunami events

Extreme event probability for $z \in \mathbb{R}$:

$$\boxed{P(z) := \mathbb{P}(F(B) \geq z),}\ B \sim \mathcal{N}(B_0, C),\ \ z \in \mathbb{R},$$

where h, v solve the shallow water equations and $F(B) = F(B; h, v)$.

Large Deviation Theory (LDT)?

- probability theory that quantifies tail behavior of distributions [Cramer, Varadhan, Dembo, Zeitouni, ...]
- Used in thermodynamics, statistical mechanics, recently in physical/dynamical systems with random perturbations [Dematteis, Farazmand, Grafke, Sapsis, ...]

Probability estimation in complex systems [Anitescu, Biros, Rao, Wahal, ...]

LDT and optimization

Set of extreme events:

$$\Omega(z) = \{B : F(B) \geq z\},$$

LDT connects the probability $P(z) = \mathbb{P}(F(B) \geq z)$ and the minimizer (≈instanton) of the rate function $I(\cdot)$ (which is a PDE-constrained optimization problem):

$$B^* = \operatorname*{argmin}_{B \in \Omega(z)} I(B)$$

$$\text{with} \quad I(B) = \frac{1}{2}\|(B - B_0)\|^2_{C^{-1}} = \frac{1}{2}(B - B_0)^T C^{-1}(B - B_0).$$

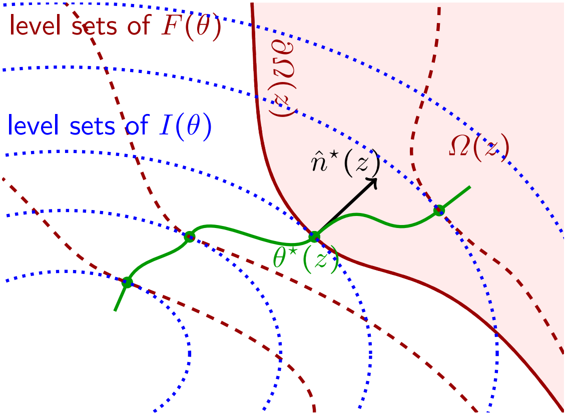
LDT ⇒ asymptotic identical behavior of logarithms

$$\frac{-I(B^*)}{\log(P(z))} \to 1 \quad \text{as } z \to \infty.$$

Thus, with a sub-exponential prefactor $C_0(z)$:

$$P(z) = C_0(z) \exp(-I(B^*(z)))$$

LDT and optimization



- LDT applies under assumptions on $F(\cdot)$ and the domain $\Omega(z)$
- LDT informally says that in the rare event limit, the probability of $\Omega(z)$ is dominated by one point $\theta^\star(z)$.
- Minimizer typically taken on boundary $\partial\Omega(z)$.

LDT and optimization

Assumption for rigorous proof:

1. F differentiable
2. Probability measure has differentiable cumulant generating function.
3. Rate function has unique global minimizer θ^* and $I(\theta^*)$ increases with z.
4. Regularity on the extreme event set

LDT and optimization

The minimizer of $I(B)$ in $\Omega(z)$ (most likely point for Gaussians):

$$B^*(z) = \underset{B\in\Omega(z)}{\operatorname{argmin}}\, I(B).$$

Relax this inequality-constrained optimization using penalty/Lagrangian multiplier $\lambda > 0$, resulting in family of optimization problems:

$$\min_B \; J(B; h, v) := I(B) - \lambda F(B; h, v),$$

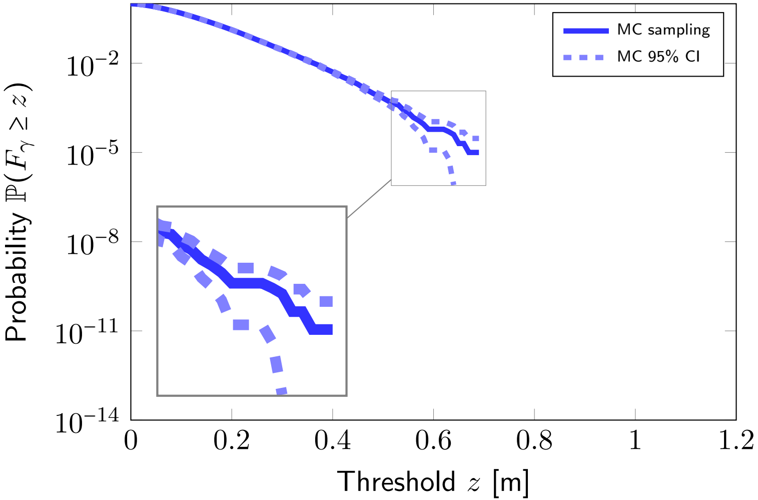
s.t. (B, h, v) satisfies shallow water eq.

Interpreting as PDE-constrained optim./optimal control:

- $I(B)$... regularization
- $F(B)$... objective

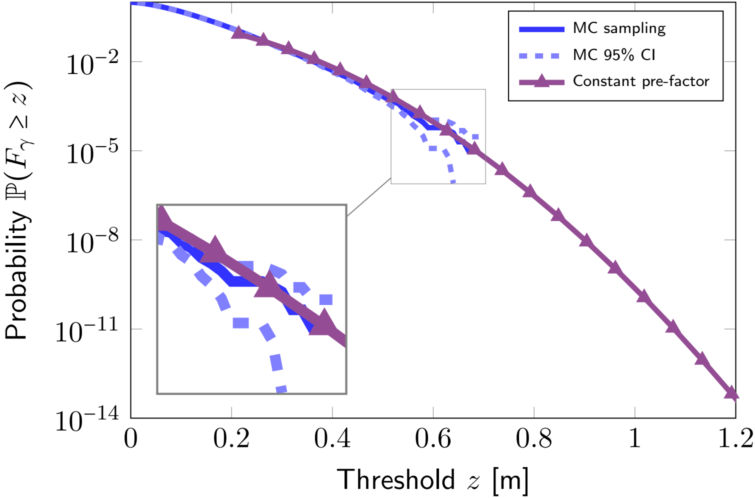
Outline

Standard MC and LDT



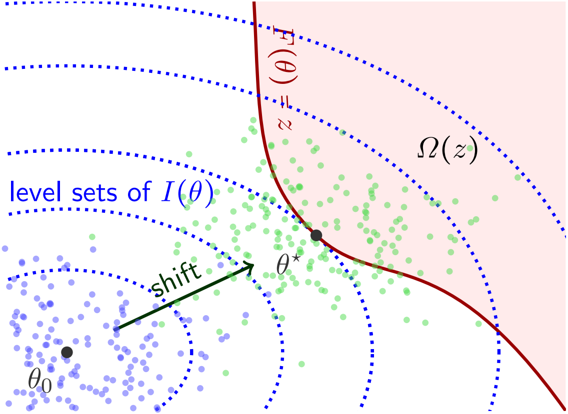
Comparison of MC with 10^5 samples and MC+LDT.

Standard MC and LDT



Comparison of MC with 10^5 samples and MC+LDT.

Importance sampling



- Improves significantly over standard MC using optimization solutions B^* as anchors
- Almost completely removes growth of variance as $z \to \infty$.
- Builds on underlying probability to be (multivariante) Gaussian.

Importance sampling

Relative root mean square error (RMSE) for N samples and estimator P_N:

$$e_N(z) := \frac{\sqrt{\mathbb{V}_\mu\left[P_N(z)\right]}}{\mathbb{E}_\mu\left[P_N(z)\right]}$$

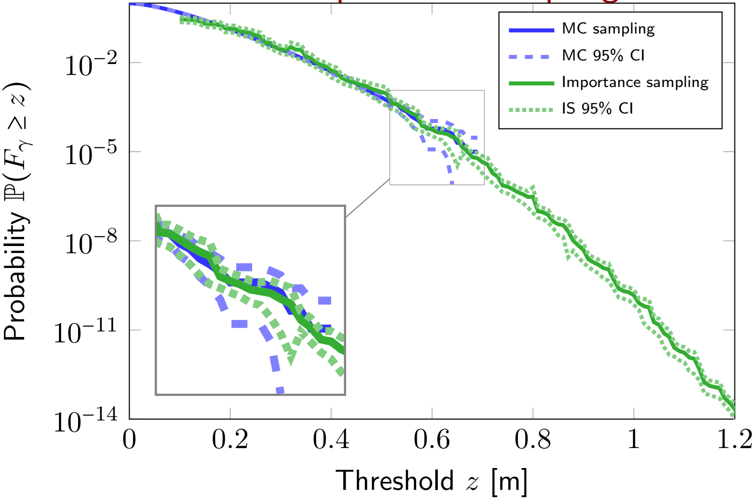
Standard MC:

$$e_N^{MC}(z) \approx \frac{1}{\sqrt{N}}\left[4\pi I(\theta^\star(z))\right]^{\frac{1}{4}} \exp\left(\frac{1}{2} I(\theta^\star(z))\right)$$

Importance sampling:

$$e_N^{IS}(z) \approx \frac{1}{\sqrt{N}}\left[\pi I(\theta^\star(z))\right]^{\frac{1}{4}}$$

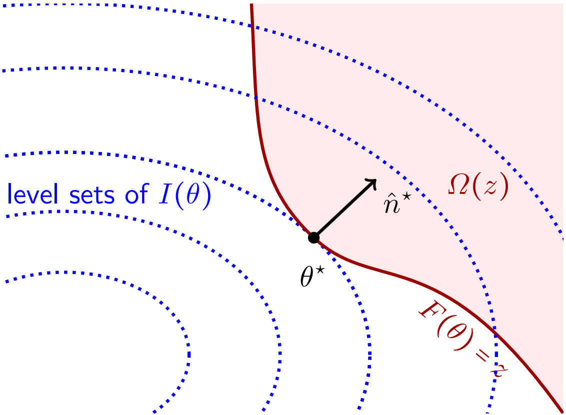
Importance sampling



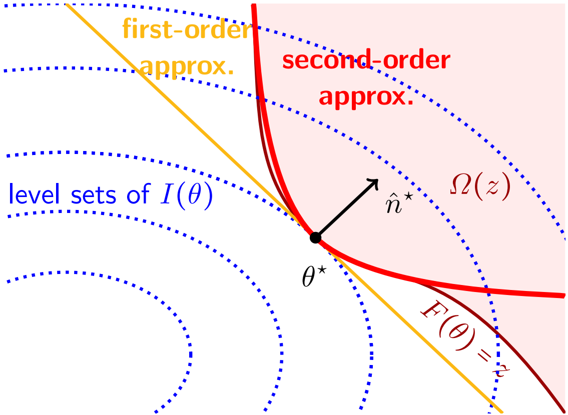
Comparison with importance sampling (100 samples for each optimizer).

Outline

First/Second-order approximation of $\Omega(z)$



First/Second-order approximation of $\Omega(z)$



Second-order approximation, Gaussian parameters

Improve rate estimates by quadratic approximation of F at B^*:

$$F(B) = F(B^*) + \nabla_B F(B^*)(B - B^*) + \frac{1}{2}(B - B^*)\nabla_B^2 F(B^*)(B - B^*) + \dots$$

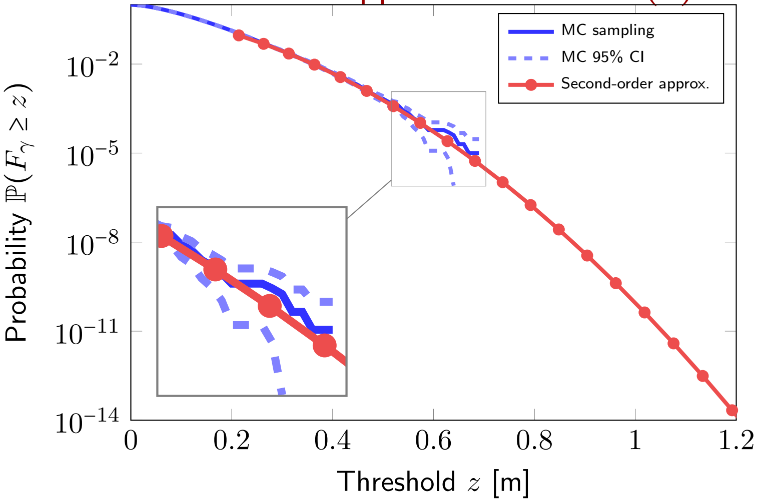
The second-order approximation is just the first-order approximation multiplied by correction term:

$$P(z) \approx \Phi(-\|\theta^*\|)\prod_{i=1}^{n-1}(1 + \|\theta^*\| k_i)^{-1/2},$$

where k_i are principle curvatures of the quadratic approximating F at B^*.

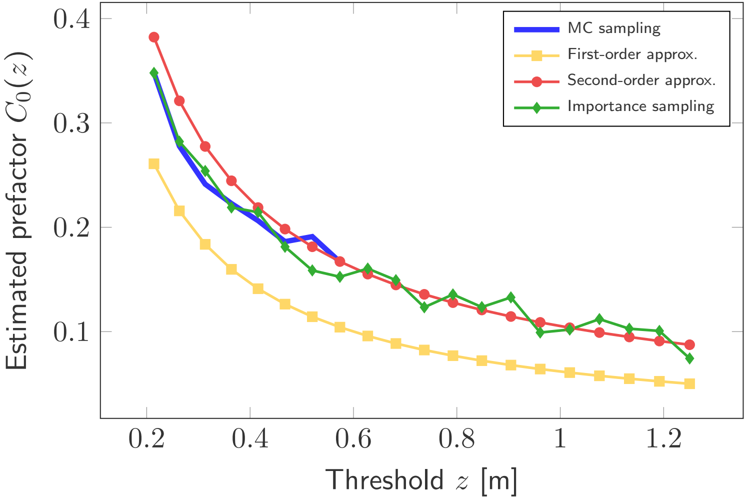
- Approximates extreme event set $\Omega(z)$ by paraboloid.
- Only need largest curvatues of $\Omega(z)$ important for distribution of B.
- Even better prefactor (but same log-asymptotic)
- In engineering, a variant of this is called *second-order reliability method (SORM)*.

SORM approximation of $\Omega(z)$



Comparison of probability estimation using approximations of $\Omega(z)$.

Prefactor comparison



Outline

Shallow water equation-constrained optimization

$$\min_{B \in \mathbb{B}} J(B; h, v) := I(B) - \lambda F(B; h, v),$$

s.t. (B, h, v) satisfies shallow water eq.

- The optimization is over the Cameron-Martin space $\mathbb{B}$.
- Gradients of the objective with respect to B, the bathymetry change are needed to solve this optimization problem.

Deriving adjoint shallow water equations

Shallow water equations (neglecting BCs):

$$
\begin{aligned}
h_t + v_x &= 0, \\
v_t + \left(\frac{v^2}{h} + \frac{1}{2}gh^2\right)_x + ghB_x &= 0, \\
h(x,0) = -B_0(x),\ \ v(x,0) &= 0.
\end{aligned}
$$

h :	water height
u :	velocity
v :	momentum hu
g :	gravity constant
B_0 :	reference bathymetry
B :	bathymetry change

Adjoint shallow water equation (neglecting BCs):

$$
\begin{aligned}
p_t + \left(gh - \frac{v^2}{h^2}\right)\mu_x - (gB_x)\mu + \lambda F_{,h} &= 0, \\
\mu_t + p_x + \frac{2v}{h}\mu_x &= 0, \\
p(x,T) = 0,\ \ \mu(x,T) &= 0
\end{aligned}
$$

μ :	adj. water height
p :	adj. velocity
$F_{,h}$:	deriv. of F w.r. to h

Deriving gradients

1. Solve state equation initial value problem for (v, h).
2. Solve adjoint equations final value problem for (p, μ).
3. Compute the gradient in direction $\hat{B}$:

$$\mathcal{G}(B)(\hat{B}) = (B - B_0, \hat{B})_{C^{-1}} + \int_0^T \int_a^b g\mu h \hat{B}_x \, dx \, dt.$$

4. Use the covariance-preconditioned gradient $C\mathcal{G}(B)$ as descent direction in minimization.

Precond. gradient descent iterations

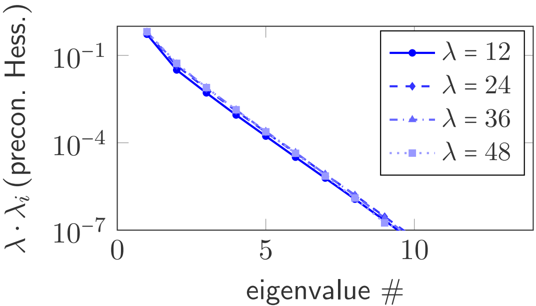
λ	$z := F_\gamma(B^\star(\lambda))$	$P^{SO}(z)$	# iter
12	0.263	4.80e-02	23
16	0.364	9.55e-03	31
20	0.468	1.24e-03	24
24	0.574	1.04e-04	31
28	0.682	5.45e-06	27
32	0.792	1.77e-07	33
36	0.905	3.54e-09	29
40	1.018	4.27e-11	32
44	1.134	3.09e-13	30
48	1.250	1.36e-15	37

Number of optimization iterations for different λ's to achieve 5 orders of magnitude gradient norm reduction.

Hessian-applies

Since B is a function/long vector, we use adjoints for that.

- SORM, which is based on the second-order approximation of the boundary of $\Omega(z)$ requires Hessian information.
- Only need dominating curvature directions of $\Omega(z)$ that are important for the probability of θ.
- Use finite differences of gradients; that's sufficient as we only need Hessian-applies
- Use randomized SVD



Shown on the left are the eigenvalues needed for second-order approximation for differently extreme points.

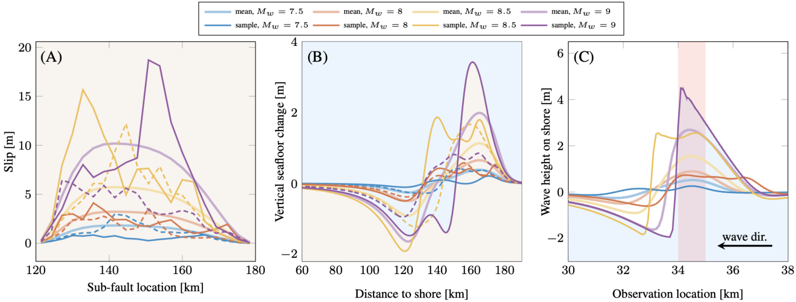
Numerical analysis challenges

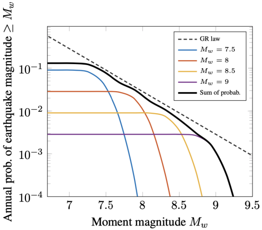
- The shallow water equations are hyperbolic can can have shocks, in which case adjoint-based gradients can be challenging: We use artificial viscosity by introducing an extra variable in the formulation.
- Spatial/temporal discretization, and optimized-then-discretize (OTD) vs. discretize-then-optimize (DTO): We DG in space, and SSP RK in time. Artificial viscosity results in convergence of adjoint-based gradients (theory for FD by Giles, S. Ulbrich . . .)
- Challenges for PDE-constrained optimization: Time-optimal control; existence/uniquiness of solutions; second-order conditions; Hessians; low-rank approximations for SORM using randSVD

Outline

Distribution modeling: Log-normal sum for slips

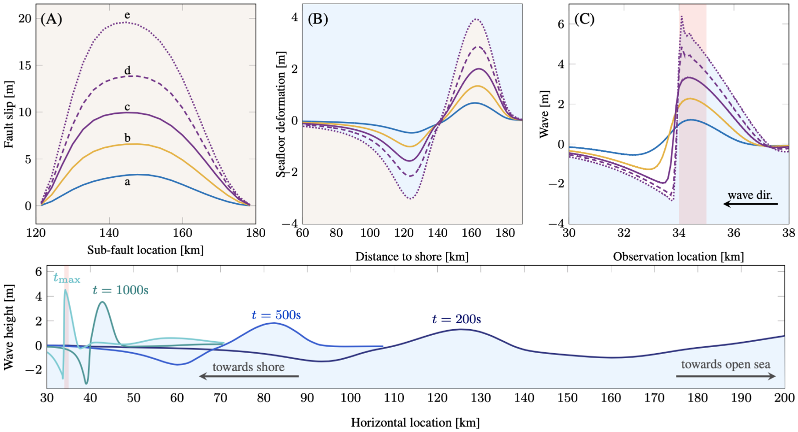
Use a sum of log-normal distributions for slips:





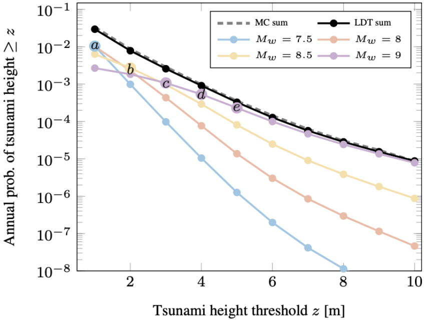
- Distribution is a sum of log-normals with different means
- distribution fits Gutenberg-Richter law for moment magnitudes 7–9

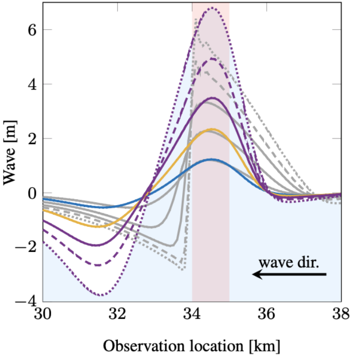
LDT-optimization results



- optimizers develop shocks close to observation region
- seems to be the mechanism to avoid energy dissipation

Probabilities and role of nonlinearity





- probability has concave trend
- nonlinearity in SWE plays significant role

Main takeaways

- Methods for extreme/rare event estimation that are:
  insensitive to extremeness,
  work for high-dimensional parameters and
  expensive parameter-to-event maps F
- Theoretical foundation: large-deviation theory to argue connection between (PDE-constrained) optimization and extreme event estimation
- Nonlinear parameter-to-event map F where derivatives are available (using adjoints); parameters θ Gaussian or non-Gaussian
- Illustration using tsunami problem, where F involves time-dependent shallow water equation; new class of PDE-constrained optimization
- Current work: Subspace important sampling based on second-order set approx.
- Extensions: 2D SWE, control/mitigation of extreme events

https://arxiv.org/abs/2007.13930
https://arxiv.org/abs/2111.14325

Summary

- Connection between PDE-constrained optimization and extreme event probability estimation, applied to 1D shallow water equations

Some literature:

- Wahal, Biros: *BIMC: The Bayesian Inverse Monte Carlo method for goal-oriented uncertainty quantification. Part I*, arXiv:1911.00619, (2019).
- Dematteis, Grafke, Vanden-Eijnden: *Extreme event quantification in dynamical systems with random components*, SIUQ, (2019).
- Farazmand, Sapsis, *Physics-based probing and prediction of extreme events*, SIAM News, (2018).
- LeVeque, George: *High-resolution finite volume methods for the shallow water equations with bathymetry and dry states*, (2008).
- Hajian, Hintermüller, Ulbrich: *Total variation diminishing schemes in optimal control of scalar conservation laws*, IMA J. Num. Anal., (2017).
- Giles, Ulbrich: *Convergence of linearized and adjoint approximations for discontinuous solutions of conservation laws. Part 1 & 2*, SINUM (2010).
- Gassner, Winters, Kopriva, *A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations*, Appl. Math. Comput. (2016).
- Wilcox, Stadler, Bui-Thanh, Ghattas, *Discretely exact derivatives for hyperbolic PDE-constrained optimization prob. discretized by the dG method*, JSC, (2015).