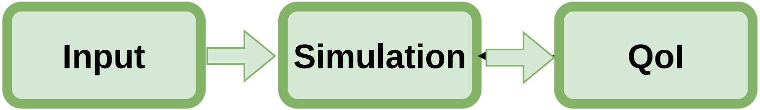
Learning to Integrate

O. G. Ernst, H. Gottschalk, T. Kowalewitz, P. Krüger
Institute of Mathematics, TU Berlin

CoMINDS Delft | 28. April 2024

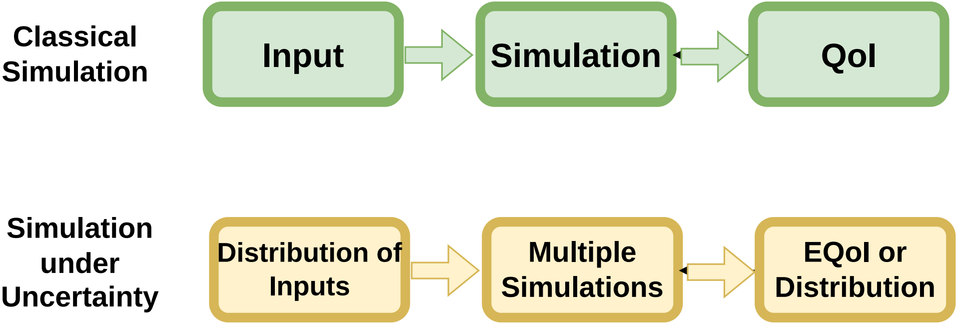
Uncertainty Quantification

Classical Simulation



Uncertainty Quantification

Technische Universität Berlin



Mathematics and Challenges of UQ

Technische Universität Berlin

- You have an input distribution μ on the parameter space Θ that has to be measured/modelled
- The simulation code $\mathrm{QoI} : \Theta \ni \theta \mapsto \Phi(\theta) = \mathrm{QoI}(\theta)$ is computationally expensive as it involves the set-up, solution and post-processing of PDE

$$\mathrm{QoI} \sim \Phi_* \mu \qquad \Phi_* \mu(A) = \mu(\Phi^{-1}(A)).$$

- The expensive simulation code prohibts the use of straight forward MC Methods
- In many cases, the space Θ is infinite diemensional and needs to be discretised.

Numerical Integration of Gaussian Distributions

- Let $\mu = N(0,1)$ standard normal distribution,

$$f_\mu(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2}.$$

- $H_n(\theta)$ n-the order Hermite polynomial, $\theta_1, \dots, \theta_n$ the zeropoints of H_n. Then,

$$\int_{\mathbb{R}} \mathrm{QoI}(\theta) f_\mu(\theta)\, \mathrm{d}\theta \approx \sum_{j=1}^{n} \omega_j \mathrm{QoI}(\theta_j).$$

- ω_j are called quadrature weights and θ_j quadrature points.
- Choice of ω_j make approximation exact for polynomials of deg. $\leq n-1$.

Tensor and Sparse Grid Quadratures

- $\mu = N(0, \mathbb{1})$, d-dimensional Normal distribution, $f_\mu(\theta) = \frac{1}{\sqrt{2\pi}^d} e^{-\frac{1}{2}\theta^\top \theta}$.
- We search for a quadrature which is exact on polynomials of degree $n-1$

$$p(\theta) = \sum_{|\alpha| \leq n-1} a_\alpha \theta^\alpha, \quad \theta^\alpha = \prod_{j=1}^{d} \theta_j^{\alpha_j}, \quad |\alpha| = \max\{\alpha_j\} \text{ or } |\alpha| = \sum_{j=1}^{d} \alpha_j.$$

- The linear system of equations requires $q(d,n) \sim n^d$ quadrature points for max degree and $q(d,n) \sim n^d/d!$ for the sum degree.

$$\int_{\mathbb{R}^d} p(\theta) f_\mu(\theta)\,\mathrm{d}\theta = \sum_{j=1}^{q(d,n)} \omega_j p(\theta_j), \quad \int_{\mathbb{R}^d} \mathrm{QoI}(\theta) f_\mu(\theta)\,\mathrm{d}\theta \approx \sum_{j=1}^{q(d,n)} \omega_j \mathrm{QoI}(\theta_j)$$

Error bound for tensor and sparse grid quadrature

- QoI $\in C^r \Rightarrow$
- For the tensor quadrature rule

$$E(n) = O\left(q(n,d)^{-\frac{r}{d}}\right)$$

- For the SG quadrature rule

$$E(n) = O\left(q(n,d)^{-r}\log(q(n,d))^{(d-1)(r+1)}\right)$$

Non Gaussian Distributions, Problem Statement

- **Q1:** What, if $f_\mu(\theta)$ is an involved distribution where quadrature weights ar not known?
- **Q2:** What, if $f_\mu(\theta)$ is even unknown and only i.i.d. samples $\Theta_l \sim f_\mu(\theta)\mathrm{d}\theta$ are known, $l = 1, \dots, N$?

Technische Universität Berlin

Using Transport Maps (Q1)

- A *transport map* $\phi : \mathbb{R}^d \to \mathbb{R}^d$ transports the *source measure* ν to the *target measure* μ is
$$\phi_* \nu = \mu.$$
- Taking $\nu = N(0, \mathbb{1})$, we obtain for any transport map ϕ
$$\int_{\mathbb{R}^d} \mathrm{QoI}(\theta) f_\mu(\theta)\, \mathrm{d}\theta = \int_{\mathbb{R}^d} \mathrm{QoI}(\phi(\theta))\, e^{-\frac{1}{2}\theta^\top \theta}\, \mathrm{d}\theta \approx \sum_{j=1}^{q(n,d)} \omega_j \mathrm{QoI}(\phi(\theta_j)).$$
- Poblems:
  a) ϕ is hard to compute
  b) The error bound depend on the nonlinearity (sup-norms of derivatives) of ϕ.

Existence of Transport Maps

- (Optimal) Transport Theory: ϕ exists as a bijective map if f_μ is atom free, e.g. as the Rosenblatt-Knothe map
- As the solution of the OT problem

$$\phi \in \arg\min_{\phi_*\nu=\mu} \int_{\mathbb{R}^d} |x-\phi(x)|^2 \mathrm{d}\nu(x)$$

- As a *flow endpoint* ϕ_1 of the time-temepdent vector field $v : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$,

$$\dot{\phi}^v(t,\theta) = v(\phi_t(\theta),t), \quad \phi_{t=0}(\theta) = \theta.$$

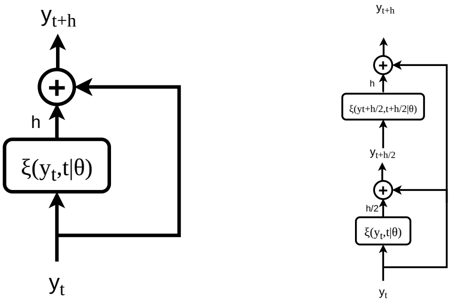
- The regularity of ϕ^v depends on the regularity og $f_\mu(\theta)$

Approximation of Flow Endpoints with Neural ODE

- Let v be the vector field such that ϕ_1^v is a transport map from ν to μ
- Let $\xi_w : \mathbb{R}^d \to [0,1]$ be a deep neural network depending on wights w and ϕ_1^ξ its flow endpoint
$$\dot{\phi}^\xi(t,\theta) = \xi_w(\phi_t(\theta), t), \quad \phi_{t=0}(\theta) = \theta.$$
- From the general theory of ODE we have with L lipshitz constent for v
$$\sup_{\theta \in K} \left| \phi_1^v(\theta) - \phi_2^\xi(\theta) \right| \leq \frac{1}{L}\left(e^L - 1\right) \|v - \xi_w\|_\infty$$
- For regular $v \in C^1$ or better, the right hand side can be made small by universal approximation if one incrases the arcitecture size and optimizes w.

Technische Universität Berlin

Numerical solutions to the ODE



- The flow enpoints of NeuralODE ϕ_1 themselves are not neural networks.
- **But:** the numerical solutions rto the neuralODE with Rungge-Kutta scheme are ResNet like neural networks.

Technische Universität Berlin

How to Learn a Transport Map?

- Express the log-likelihood of flow endpoints via Liouville's formula

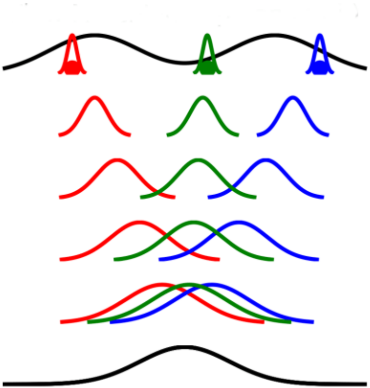
$$\log\left(f_{\phi^{\xi}_{1*}\nu}(\theta)\right) = -\frac{1}{2}\|(\phi^{\xi}_{1})^{-1}(\theta)\|^2 + \int_0^1 \mathrm{div}_{\theta'}\xi_w(\theta',t)|_{\theta'=\phi_t(\theta)}\,\mathrm{d}t + \mathrm{cst}.$$

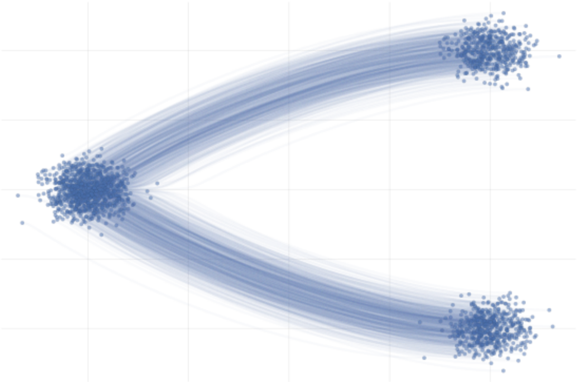
- Then train with the negative log likelihood of the observed data points $\{\Theta_j\}$

$$\hat{w} \in \arg\min_w \left\{-\frac{1}{n}\sum_{l=1}^{N}\log\left(f_{\phi^{\xi_w}_{1*}\nu}(\Theta_l)\right)\right\}$$

- The set $\hat{\phi} = \phi^{\xi_w}_1$ and dicretize.

Flow Matching (Lipman et al 2022)





Technische Universität Berlin

Training with Flow Matching

- Build a trivial probability path that transports a small gaussioan arounf the data Point Θ_j to a standard normal vector distribution. Each such probability pathh comes from a vector field $v(\theta|\Theta_j)$
- Train ξ_w as the vector field that minimizes (in w)

$$\mathbb{E}_{\Theta\sim f_\mu}\left[\|\xi_w \circ \phi^{v(\cdot|\Theta)} - v(\Theta)\circ\phi^{v(\cdot|\Theta)}\|_{L^2}\right] \approx \sum_{k=1}^{Q}\frac{1}{Q}\frac{1}{N}\sum_{l=1}^{N}\left|v(\phi_{k/Q}^{v(\cdot|\Theta_j)}|\Theta_j) - \xi_w(\phi_{k/Q}^{v(\cdot|\Theta_j)})\right|^2$$

Uncertainty for the Poisson Equation

Technische Universität Berlin

- The linear stationary diffusion problem on bounded domain $D \subseteq \mathbb{R}^d$

$$-\nabla \cdot (a \nabla u) = f, \quad u|_{\partial D_d} = 0, \qquad n \cdot \nabla u = g_n|_{\partial D_n}.$$

- $a(x, \omega) = T(Z(x, \omega)) > 0$ random coeff., eg $T(z) = \exp(z)$.
- $Z(x) = Z(x, \omega)$ random field

Technische Universität Berlin

Uncertainty for the Poisson Equation

- The linear stationary diffusion problem on bounded domain $D \subseteq \mathbb{R}^d$

$$-\nabla \cdot (a\nabla u) = f, \quad u|_{\partial D_d} = 0, \qquad n \cdot \nabla u = g_n|_{\partial D_n}.$$

- $a(x,\omega) = T(Z(x,\omega)) > 0$ random coeff., eg $T(z) = \exp(z)$.
- $Z(x) = Z(x,\omega)$ random field

Uncertainty for the Poisson Equation

- The linear stationary diffusion problem on bounded domain $D \subseteq \mathbb{R}^d$

$$-\nabla \cdot (a \nabla u) = f, \quad u|_{\partial D_d} = 0, \qquad n \cdot \nabla u = g_n|_{\partial D_n}.$$

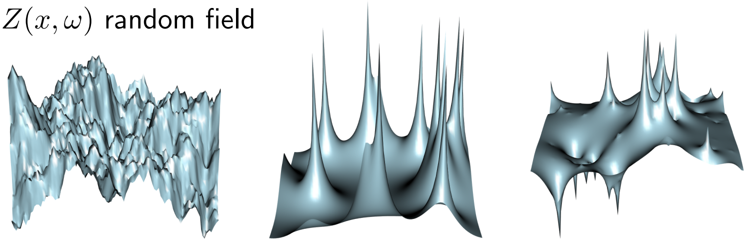
- $a(x, \omega) = T(Z(x, \omega)) > 0$ random coeff., eg $T(z) = \exp(z)$.
- $Z(x) = Z(x, \omega)$ random field

Uncertainty for the Poisson Equation

- The linear stationary diffusion problem on bounded domain $D \subseteq \mathbb{R}^d$

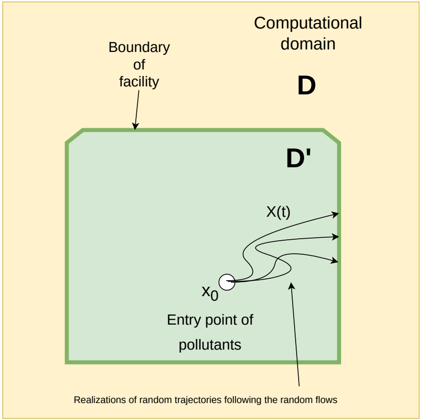
$$-\nabla \cdot (a\nabla u) = f, \quad u|_{\partial D_d} = 0, \qquad n \cdot \nabla u = g_n|_{\partial D_n}.$$

- $a(x,\omega) = T(Z(x,\omega)) > 0$ random coeff., eg $T(z) = \exp(z)$.
- $Z(x) = Z(x,\omega)$ random field



Technische Universität Berlin

QoI: Flow Modeling for a Waste Deposit Facility



Technische Universität Berlin

Physical Interpretation of the Poisson Equation

- $u(x)$ is ground pressure at location $x \in D$
- $a(x)$ is conductivity of ground material at location x
- Thus, the flow velocity $q(x)$ at location x is

$$q(x) = -a(x)\nabla u(x).$$

- QoI is e.g. first exit time from restricted region $D' \subseteq D$ around pollution location $x_0 \in D'$

$$\dot{X}(t) = q(X(t)), \quad X(0) = x_0, \quad \mathrm{QoI} = \inf\{t > 0 : X(t) \notin D'\}.$$

- As $q(x)$ is random through $a(x)$, also the QoI becomes a random variable!

Technische Universität Berlin

Generalized Random Fields

- Generalized random fields do not have ordinary functions as pathes, but generalized functions, i.e. (tempered) distributions.
- **Definition:** A generalized random field over a real topological vector space of functions V fulfills
  - $Z(f) ='' \int_{\mathbb{R}^d} Z(x)f(x)\mathrm{d}x\,''$ is a random variable $\forall f \in V$;
  - $Z(\gamma f + \beta g) = \gamma Z(f) + \beta Z(g)$ holds a.s. $\forall g, f \in V,\ \gamma, \beta \in \mathbb{R}$;
  - $f_n \to f$ in $V \Rightarrow Z(f_n) \to Z(f)$ in distribution.

Minlos Theroem

- **Definition:** A characteristic functional $\varphi : V \to \mathbb{C}$ fulfills
  - $\varphi(0) = 1$;
  - φ is continuous (at zero);
  - φ is positive definite $(\varphi(f_1 - f_j))_{i,j=1,\dots,n} \geq 0$.
- **Theorem: (Minlos)** Let V be a nuclear space, then the random fields are in one-to-one correspondence to the characteristic functionals via

$$\varphi(f) = \mathbf{E}\left[e^{iZ(f)}\right] = \int_{V'} e^{i\omega(f)} \mathrm{d}\mathbf{P}(\omega).$$

- Hence it is enough to write down a characteristic functional to construct a random field.

Lévy Noise Fields

- GWN - Gaussian white noise ($V = \mathscr{S}(\mathbb{R}^d)$)

$$\varphi_{\mathrm{GWN}}(f) = \mathbf{E}\left[e^{i\eta(f)}\right] = e^{-\frac{1}{2}\sigma^2 \int_{\mathbb{R}^d} f^2 \,\mathrm{d}x}$$

- Lévy noise field

$$\varphi_{\mathrm{LN}}(f) = \mathbf{E}\left[e^{i\eta(f)}\right] = e^{\int_{\mathbb{R}^d} \psi(f)\,\mathrm{d}x}$$

- $\psi(t)$ is a Lévy characteristic

$$\psi(t) = ibt - \frac{1}{2}\sigma^2 t^2 + \int_{\mathbb{R}\setminus\{0\}} \left(e^{ist} - 1 - ist\mathbb{1}_{|s|<1}\right)\,\mathrm{d}\nu(s)$$

- $b \in \mathbb{R}, \sigma^2 > 0$, ν measure s.t. $\int_{\mathbb{R}\setminus\{0\}} \min(1, |s|^2)\mathrm{d}\nu(s) < \infty$. (b, σ^2, ν) Lévy triplet.

Examples for Lévy Noise Fields

- We assume that $\int_{0<|s|<1} |s| \mathrm{d}\nu(s) < \infty$. We can reparametrize $b' = b - \int_{|s|<1} s \mathrm{d}\nu(s)$.

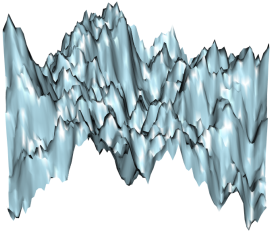
$$\psi(t) = ib't - \frac{1}{2}\sigma^2 t^2 + \int_{\mathbb{R}\setminus\{0\}} \left(e^{ist} - 1\right) \mathrm{d}\nu(s)$$

- $b' = 0$, $\nu = 0$ GWN
- $b' = 0$, $\sigma^2 = 0$ compound Poisson noise / Poisson point process
  - $\int_{0<|s|<1} \mathrm{d}\nu(s) < \infty$ finite activity, e.g. $\nu = \delta_1$ Poisson noise
  - $\int_{0<|s|<1} \mathrm{d}\nu(s) = \infty$ inifinite activity, e.g. $d\nu(s) = \lambda \frac{e^{-\gamma|s|}}{|s|}$ bi-gamma noise, $\lambda, \gamma > 0$.

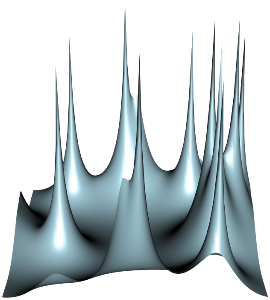
Smoothed Lévy Noise Fields

- The natural generalization from GWN to LN for random fields $Z(x)$ is via the SPPDE
$$(-\Delta + m^2)^{\alpha/2} Z(x) = \eta(x).$$
- When is this a conventional random field?
- **Theorem:** Let $\alpha > d + \max\{0, \frac{3d-12}{8}\}$, then $Z(x)$ has a.s. continuous paths.
- Thus for $a(x) = T(Z(x))$ with $T(z) > 0$ continuous, the solution $u(x)$ to the diffusion equation exists a.s. .

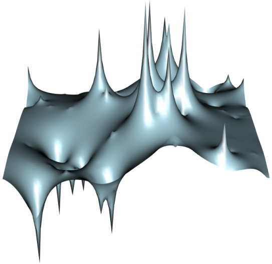
Examples for Smoothed Lévy Coefficient Fields



Gaussian random field



Poisson random field

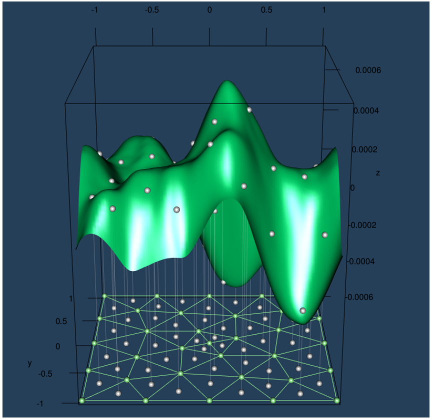


bi-gamma random field

- All fields have exactly the same covariance function!
- Does this thus have any effect on the distribution of the QoI?

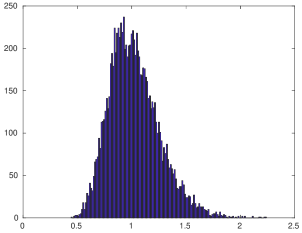
Numerical Simulation

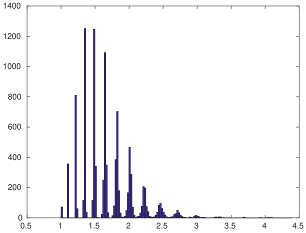
- Generate noise field on a grid
- FFT noise field
- Multiply with symbol $\frac{1}{(|k|^2+m^2)^\alpha}$
- FFT back
- Interpolate to FE grid quadrature points
- Set up FE with random coefficients and solve
- Postprocess QoI

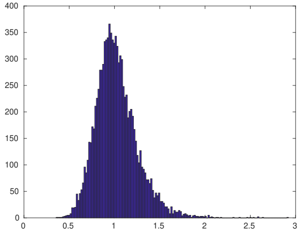


Technische Universität Berlin

QoI Distributions from 10 000 MC Simulations





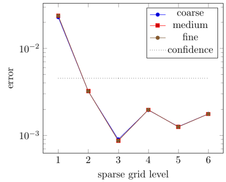


Gaussian QoI-distribution   Poisson QoI-distribution   bi-gamma QoI-distribution

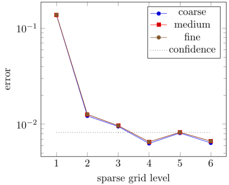
- Differences are not dramatic, but bi-gamma has a thicker lower tail as compared with the normal distribution (i.e. to the risky side)
- Poisson noise has clear cut-off on the 'risky' tail

Technische Universität Berlin

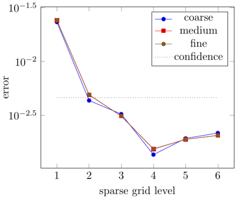
Results from the Learned Quadrature (9 modes)



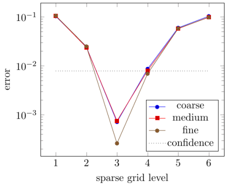
Gaussian



Gamma



Bigamma



Compound Poisson

Conclusion

Problems

- Uncertainty Quantification is a necessity in simulation of real life problems
- Distributions are rarely Gaussian
- Distribtions often are only given by data

Towards a solution

- One can learn transport maps using invertible normalizing flows
- One can thus learn quadrature rules
- But: the non-linearity of the transport map matters!
- Flow matching is an excellent candidate to keep derivatives low.