An introduction to Uncertainty Quantification

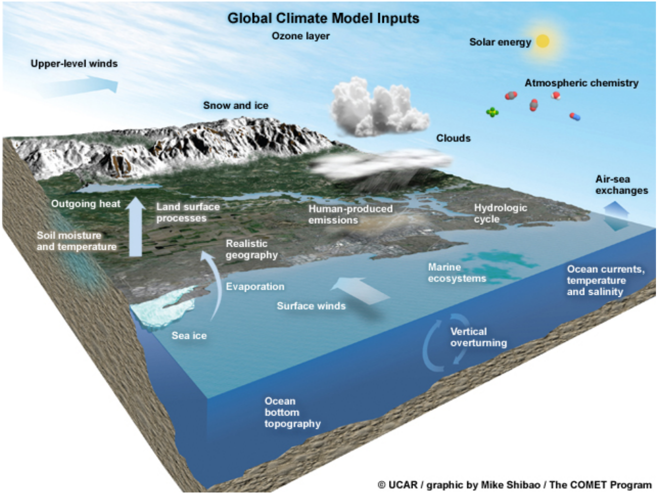
Olivier Zahm
olivier.zahm@inria.fr

INRIA-AIRSEA
https://team.inria.fr/airsea/en/

GdR MaNu,
October 27, 2023

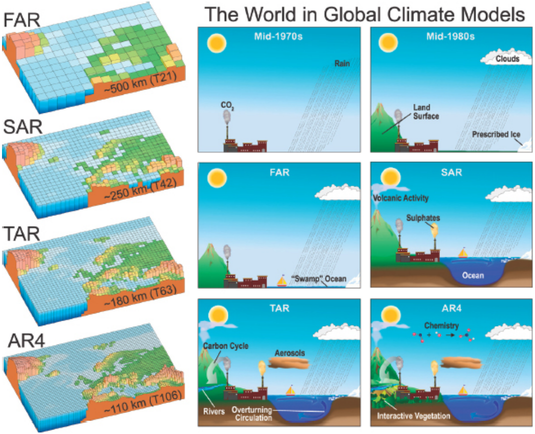
Motivating example: **climate change simulation**

Input



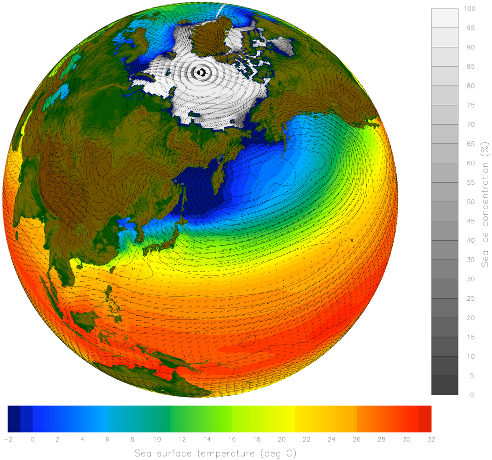
Motivating example: **climate change simulation**

Input



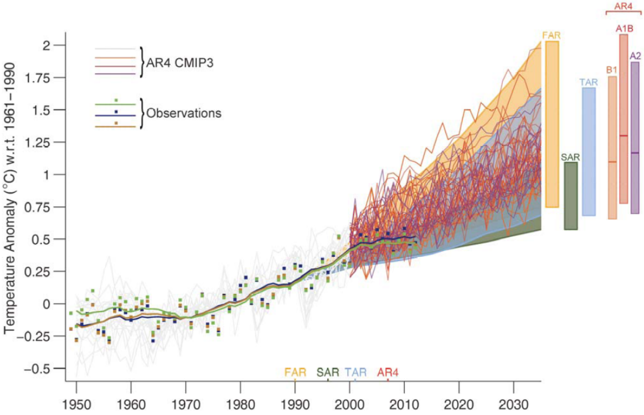
Motivating example: **climate change simulation**

Input ⟶ Output



Motivating example: **climate change simulation**

Input ⟶ Output

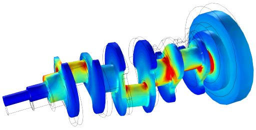


Another example: **computational mechanics**

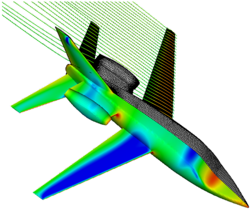
Input ⟶ Output

- Mean of the VonMises stress
- Probability of failure
- Lifespan
- ...

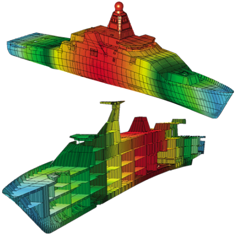
Cars



Airplanes



Ships

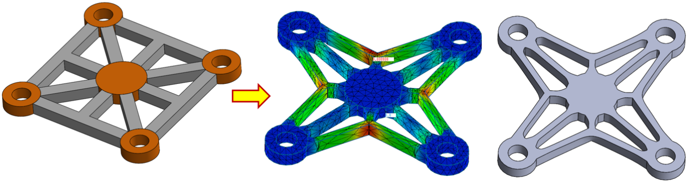


Another example: **computational mechanics**

Input → Output

- **Geometry**

- Mean of the VonMises stress
- Probability of failure
- Lifespan
- ...

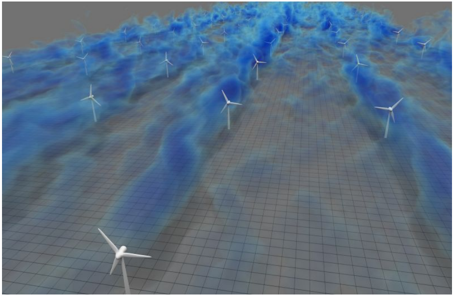


Another example: **computational mechanics**

Input ⟶ Output

- Geometry
- **External forcing**

- Mean of the VonMises stress
- Probability of failure
- Lifespan
- ...

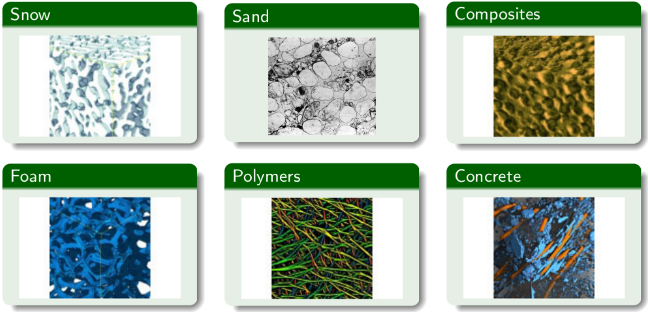


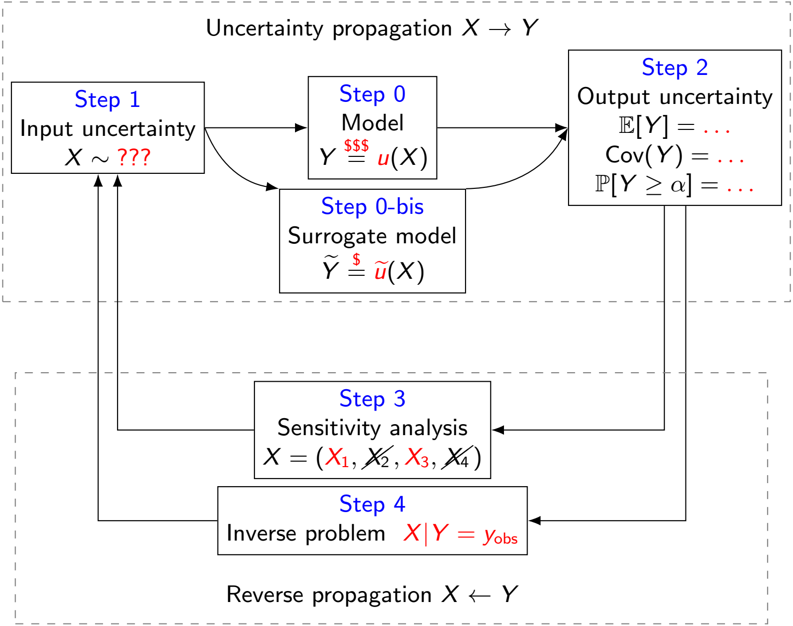
Another example: **computational mechanics**

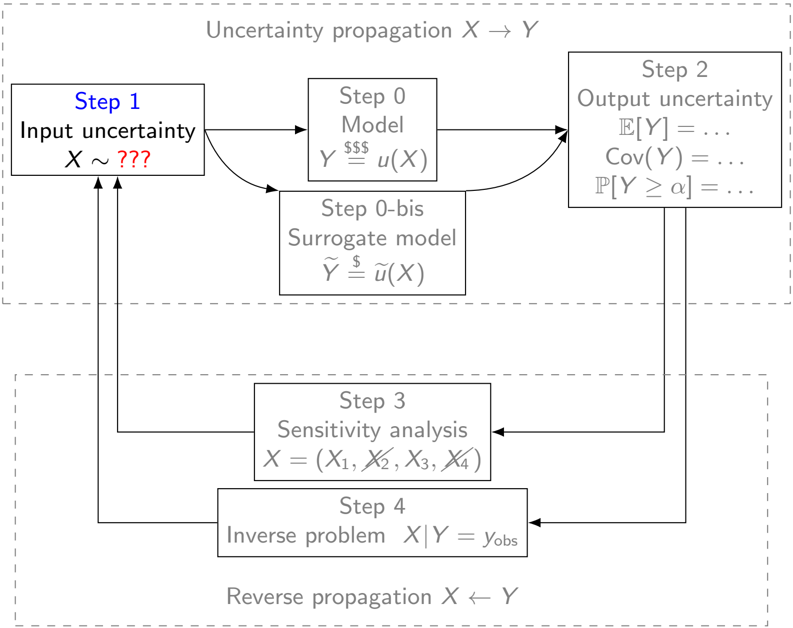
Input ⟶ Output

- Geometry
- External forcing
- **Material property**

- Mean of the VonMises stress
- Probability of failure
- Lifespan
- ...




Uncertainty propagation $X \to Y$
Step 1
Input uncertainty
$X \sim ???$
Step 0
Model
$Y \overset{\$\$\$}{=} u(X)$
Step 0-bis
Surrogate model
$\widetilde{Y} \overset{\$}{=} \widetilde{u}(X)$
Step 2
Output uncertainty
$\mathbb{E}[Y] = \dots$
$\mathrm{Cov}(Y) = \dots$
$\mathbb{P}[Y \geq \alpha] = \dots$
Step 3
Sensitivity analysis
$X = (X_1, \cancel{X_2}, X_3, \cancel{X_4})$
Step 4
Inverse problem $X|Y = y_{\text{obs}}$
Reverse propagation $X \leftarrow Y$


Uncertainty propagation $X \to Y$
Step 1
Input uncertainty
$X \sim$???
Step 0
Model
$Y \overset{\$\$\$}{=} u(X)$
Step 0-bis
Surrogate model
$\widetilde{Y} \overset{\$}{=} \widetilde{u}(X)$
Step 2
Output uncertainty
$\mathbb{E}[Y] = \dots$
$\mathrm{Cov}(Y) = \dots$
$\mathbb{P}[Y \geq \alpha] = \dots$
Step 3
Sensitivity analysis
$X = (X_1, \cancel{X_2}, X_3, \cancel{X_4})$
Step 4
Inverse problem $X|Y = y_{\mathrm{obs}}$
Reverse propagation $X \leftarrow Y$

Input uncertainty

Two types of uncertain inputs:

- **Stochastic uncertainties**. These variables exhibit **inherent variability** due to random phenomena (typically, a quantity subject to random fluctuations like wind, rain etc).
- **Epistemic uncertainties**. These variables have a value, but it is unknown to us due to a **lack of knowledge** (typically, a constant in a physical law).

In both cases, we model the uncertainties with the probability theory: we consider $X = (X_1, \dots, X_d)$ as a **continuous random vector** with probability density function π_X so that the probability of an event $A \subset \mathbb{R}^d$ is given by

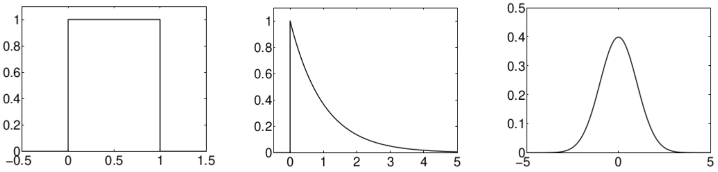
$$\mathbb{P}(X \in A) = \int_A \pi_X(x)\mathrm{d}x \quad \in [0, 1]$$

Remark: alternative modelling of uncertainties using fuzzy sets and possibility theory (Zadeh, 1978) or the theory of evidence (Dempster 1967, Shafer 1976)...

Independent input variables

$$\pi_X(x_1, \ldots, x_d) = \pi_{X_1}(x_1) \ldots \pi_{X_d}(x_d)$$

where π_{X_i} is the i-th marginal density.



(a) Uniform X_1 (b) Exponential X_2 (c) Gaussian X_3

We can identify π_X by **maximizing the entropy** (the "lack of information") under some prescribed contraints, like support, mean, variance...

To identify the density **from a sample** $\{X^{(1)}, X^{(2)}, \ldots\}$, we can

- compute histograms or use kernel methods (**non-parametric methods**)
- maximize the likelihood of the sample over a given class of densities (**parametric method**)

[Hastie, Tibshirani and Friedman: The Elements of Statistical Learning, Springer, 2001.]

Dependent input variables via:

- the copula of X

$$\pi_X(x_1,\dots,x_d) = \pi_{X_1}(x_1)\dots\pi_{X_d}(x_d)c(x_1,\dots,x_d)$$

- the conditional marginals of X (directed graphical model)

$$\pi_X(x_1,\dots,x_d) = \pi_{X_1}(x_1)\pi_{X_2|X_1}(x_2|x_1)\pi_{X_3|X_1,X_2}(x_3|x_1,x_2)\dots$$

- the conditional independence structure of X (undirected graphical model)

$$\pi_X(x_1,\dots,x_d) \propto \exp\left(V_1(x_1) + V_2(x_2) + V_{12}(x_1,x_2) + V_{13}(x_1,x_3) + \dots\right)$$

- a transport map $T : \mathbb{R}^d \to \mathbb{R}^d$, ideally invertible

$$\pi_X(x) = T_\sharp \pi_Z(x) \qquad \Leftrightarrow \qquad X = T(Z), \qquad Z \sim \pi_Z$$

- a hierarchical model $\pi_{X,Z}(x,z) = \pi_{X|Z}(x|z)\pi_Z(z)$ so that

$$\pi_X(x) = \int \pi_{X,Z}(x,z)\mathrm{d}z \qquad \Leftrightarrow \qquad \begin{cases} \text{first } Z \sim \pi_Z(\cdot) \\ \text{then } X \sim \pi_{X|Z}(\cdot|Z) \end{cases}$$

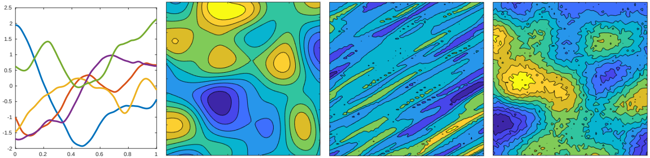
- ...

Most of the time, we use the multivariate Gaussian density $X \sim \mathcal{N}(m,\Sigma)$

$$\pi_X(x) \propto \exp\left(-\frac{1}{2}\|x-m\|^2_{\Sigma^{-1}}\right)$$

Input as a **random field** ($d = \infty$)

$$\underbrace{X = (X_1, \ldots, X_d)}_{\text{random vector}} \quad \hookrightarrow \quad \underbrace{X = \{X(s), s \in \Omega\}}_{\text{random field}}$$



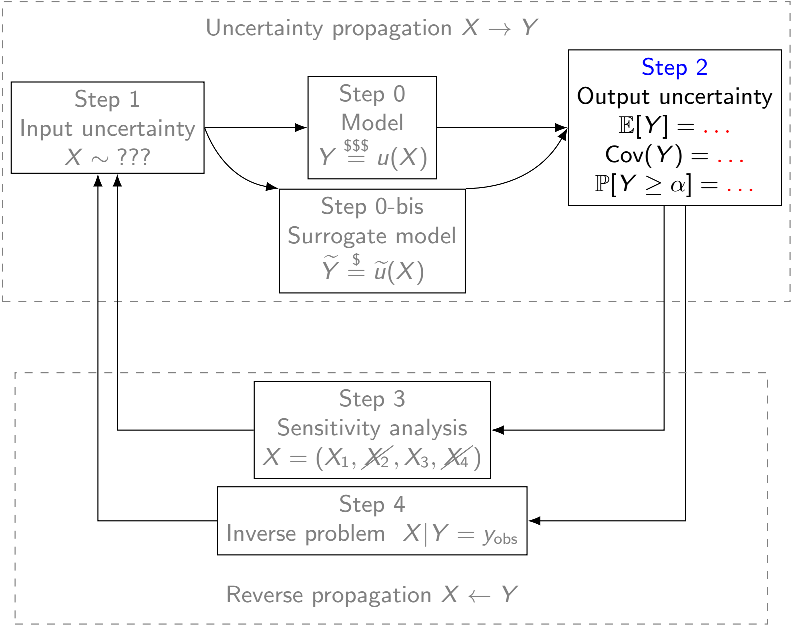
- Gaussian random field $X \sim \mathcal{N}(\mu, c)$ are such that, for any set of points $s_1, \ldots, s_k \in \Omega$, we have $(X(s_1), \ldots, X(s_n)) \sim \mathcal{N}(m, \Sigma)$ with

$$m = \begin{pmatrix} \mu(s_1) \\ \vdots \\ \mu(s_n) \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} c(s_1, s_1) & \cdots & c(s_1, s_n) \\ \vdots & \ddots & \vdots \\ c(s_n, s_1) & \cdots & c(s_n, s_n) \end{pmatrix}$$

  Valid bivariate functions $c(\cdot,\cdot)$ such that $\Sigma \succeq 0$ are called **kernels**.
- Karhunen-Loève decomposition (= Singular Value Decomposition)

$$X(s) = \sum_{i=1}^{\infty} \underbrace{\sigma_i}_{\to 0} \underbrace{X_i}_{\text{random}} \underbrace{\varphi_i(s)}_{\text{deterministic}} \quad \approx \quad X_d(s) = \sum_{i=1}^{d} \sigma_i X_i \varphi_i(s)$$

  Back to finite dimension, but in practice, $d \gg 1$ can be large...


Uncertainty propagation $X \to Y$
Step 1
Input uncertainty
$X \sim ???$
Step 0
Model
$Y \overset{\$\$\$}{=} u(X)$
Step 0-bis
Surrogate model
$\widetilde{Y} \overset{\$}{=} \widetilde{u}(X)$
Step 2
Output uncertainty
$\mathbb{E}[Y] = \dots$
$\mathrm{Cov}(Y) = \dots$
$\mathbb{P}[Y \geq \alpha] = \dots$
Step 3
Sensitivity analysis
$X = (X_1, \cancel{X_2}, X_3, \cancel{X_4})$
Step 4
Inverse problem $X|Y = y_{\mathrm{obs}}$
Reverse propagation $X \leftarrow Y$

Uncertainty propagation

Most of the time, the goal is to **compute an expectation** of the form

$$\mathbb{E}[\psi(Y)] = \int F(x)\,\pi_X(x)\mathrm{d}x, \qquad F(x) = \psi \circ u(x)$$

- $\psi(t) = t$: mean of Y, *i.e.* $\mathbb{E}[Y] = \int u(x)\pi_X(x)\mathrm{d}x$
- $\psi(t) = t^2$: variance of Y, *i.e.* $\mathrm{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$
- $\psi(t) = \mathbf{1}_{[\alpha,\infty)}(t)$: probability of exceeding a threshold α, *i.e.*

$$\mathbb{P}(Y \geq \alpha) = \int \mathbf{1}_{[\alpha,\infty)}(u(x))\pi_X(x)\mathrm{d}x$$

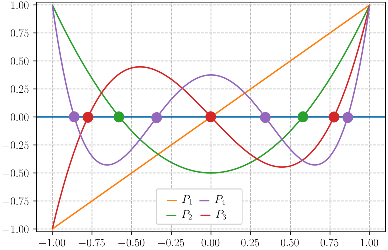
Unless in the **exceptional situation** where the integral can be computed analytically (see later), the expectation $\mathbb{E}[\psi(Y)]$ need to be **approximated numerically**...

Where to evaluate the model $u(x)$ in order to best approximate $\mathbb{E}[\psi(Y)]$?

Deterministic **Gaussian quadrature** in dimension $d = 1$

$$\int F(x)\,\pi_X(x)\mathrm{d}x \quad \approx \quad \sum_{i=1}^{n} \omega_i\, F(x^{(i)})$$

where $x^{(1)}, \ldots, x^{(n)}$ are the roots of the n-th orthogonal polynomial P_n of degree n such that $\int P_m P_n \pi_X \mathrm{d}x = \delta_{mn}$, and $\omega_i > 0$ the corresponding weights.



Legendre polynomials $P_1, \ldots, P_4$, orthogonal on $\pi_X = \mathcal{U}([-1, 1])$

- Exact for polynomials F with degree $\leq 2n - 1$
- If $F(x) = \psi \circ u(x)$ is **analytic**, quadrature error decays in $\mathcal{O}(\rho^n)$
- If we only have $F \in \mathcal{C}^1$, then error decays in $\mathcal{O}(1/n)$
- But $n \leftarrow n + 1$ requires re-evaluating u at new points (no recycling)

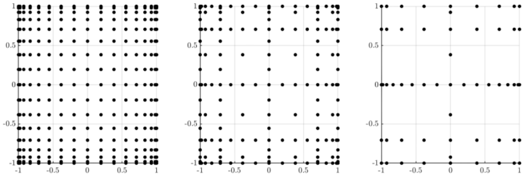
Deterministic quadrature **in dimension d>1**

Full tensorization requires $N = n^d$ model evaluations

$$\int F(x)\,\pi_X(x)\mathrm{d}x \quad \approx \quad \sum_{i_1=1}^{n} \dots \sum_{i_d=1}^{n} \omega_{i_1} \dots \omega_{i_d}\, F\left(x_1^{(\alpha_1)}, \dots, x_d^{(\alpha_d)}\right)$$

Sparse tensorization using Smolyak grid permits to avoid the exponential increase $d \mapsto n^d$... but still limited to reasonable dimensions $d = \mathcal{O}(50)$.

$$\int F(x)\,\pi_X(x)\mathrm{d}x \quad \approx \quad \sum_{\substack{\alpha \in \Lambda_N \subset \{1,\dots,n\}^d \\ \#\Lambda_N = N}} \omega_\alpha\, F\left(x_1^{(\alpha_1)}, \dots, x_d^{(\alpha_d)}\right)$$



Full tensorization Sparse tensorizations

If F admits an **holomorphic extension**, and if its Taylor coefficients are ℓ^p-summable for some $p < 1$, then error in $\mathcal{O}(1/N^{1+\varepsilon})$ with $\varepsilon = 2(\frac{1}{p} - 1) > 0$

[Zech and Schwab: Convergence rates of high dimensional Smolyak quadrature, ESAIM:M2AN (2020)]

Monte Carlo (MC) method

Draw N **independent** samples $X^{(1)}, \dots, X^{(N)}$ from π_X and

$$\mathrm{I} = \int F(x)\, \pi_X(x) \mathrm{d}x \quad \approx \quad \frac{1}{N} \sum_i^N F\left(X^{(i)}\right) = \widehat{\mathrm{I}}_N$$

- Constant weights: $\omega_i = 1/N$
- Unbiased estimator: $\mathbb{E}[\widehat{\mathrm{I}}_N] = \mathrm{I}$ for all N
- Converging estimator: $\widehat{\mathrm{I}}_N \underset{N \to \infty}{\longrightarrow} \mathrm{I}$ with probability 1
- Variance of the estimator

$$\mathrm{Var}(\widehat{\mathrm{I}}_N) \overset{\text{independance of the } X^{(i)}}{=} \frac{\mathrm{Var}(F(X))}{N}$$

- Relative quadratic error

$$\frac{\mathbb{E}[(\mathrm{I} - \widehat{\mathrm{I}}_N)^2]^{1/2}}{\mathrm{I}} = \frac{1}{\sqrt{N}} \frac{\sqrt{\mathrm{Var}(F(X))}}{\mathrm{I}}$$

The convergence is independent on the dimension, requires no regularity assumption on $F(x) = \psi \circ u(x)$, permits recycling when $N \leftarrow N + 1$... but convergence is terribly slow $\mathcal{O}(1/N^{1/2})$.

Bonus: **confidence intervals**

Let $\sigma = \mathrm{Var}(F(X))$. The **central limit theorem** states that

$$\mathbb{P}\left\{\frac{\widehat{\mathrm{I}}_N - \mathrm{I}}{\sigma/\sqrt{N}}\right\} \underset{N\to\infty}{\longrightarrow} \mathbb{P}\left[Z < t\right], \quad Z \sim \mathcal{N}(0,1)$$

Then, for N "sufficiently large", we have $\boxed{\widehat{\mathrm{I}}_N \approx \mathcal{N}(\mathrm{I}, \sigma/N)}$ and then

$$\mathbb{P}\left\{\mathrm{I} \in \left[\widehat{\mathrm{I}}_N - \frac{1.96\,\sigma}{\sqrt{N}}, \widehat{\mathrm{I}}_N + \frac{1.96\,\sigma}{\sqrt{N}}\right]\right\} \approx 0.95$$

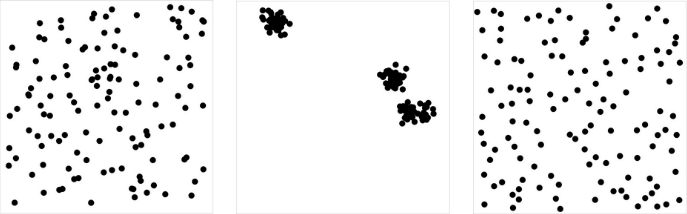
We can estimate σ via the (unbiased) estimator

$$\widehat{\sigma}_N = \left(\frac{1}{N-1}\sum_{i=1}^{N} F(X^{(i)})^2 - \widehat{\mathrm{I}}_N^2\right)^{1/2}$$

Take-away message: we cannot provide any interval which **surely** contains I... but we can give **confidence intervals** which we know they contains I with high probability.

Alternative: **Determinantal Point Process (DPP)**

Idea: construct a **stochastic Point Process (PP)** which exhibits **repulsion**.



(a) PP, independent (MC) (b) PP with attraction (c) DPP exhibits repulsion

DPP: given a kernel $c(\cdot,\cdot)$, the density of the sample $\mathcal{X}_N = \{X^{(1)}, \dots, X^{(N)}\}$ is

$$\pi(\mathcal{X}_N) \propto \det(K), \qquad K_{i,j} = c(X^{(i)}, X^{(j)})$$

- The points $X^{(i)}$ are **not independent**, and the weights ω_i are **not constant**
- Sample from a DPP is not trivial: you'll have a lot of fun to implement it!
- Recycling for $N \leftarrow N+1$? I don't think so...
- Convergence: $\mathcal{O}(1/N^{\frac{1+1/d}{2}})$ for $f \in \mathcal{C}^1([0,1]^d)$.

[Bardenet, Hardy: Monte Carlo with DPP, Ann. Appl. Probab. 2020]

Another alternative: **Quasi Monte Carlo (QMC)**

Idea: construct a **deterministic** sequence of points with **low discrepancy**.
For the uniform $\pi_X = \mathcal{U}([0,1]^d)$, the discrepancy of $\mathcal{X}_N = \{X^{(1)}, \dots, X^{(N)}\}$ is a measure of how well $\frac{\#\{B \cap X\}}{\#X}$ approximates the volume of any box B, that is

$$D(\mathcal{X}_N) = \sup_{B \in \textbf{boxes of } [0,1]^d} \left| \frac{\#\{B \cap X\}}{\#X} - \int_B \mathrm{d}\pi_X \right|$$

The **Koksma-Hlawka theorem** states

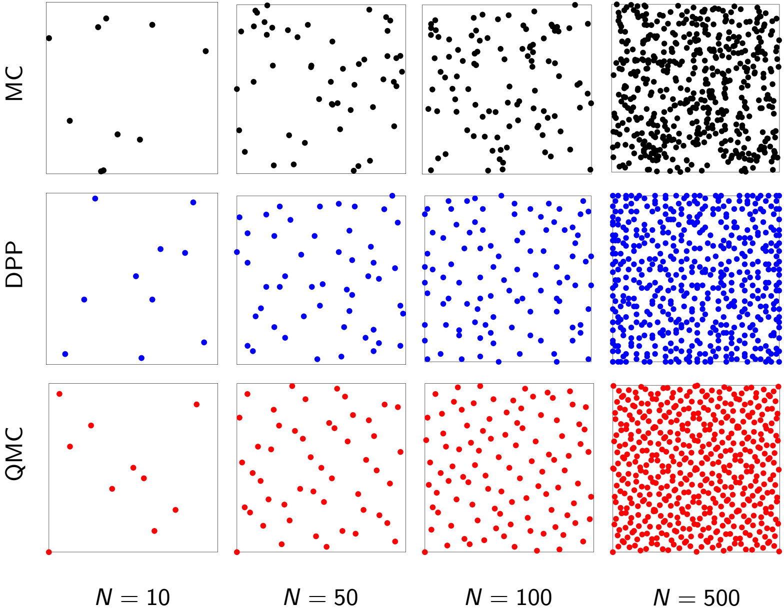
$$|\mathrm{I} - \widehat{\mathrm{I}}_N| \leq V(F) D(\mathcal{X}_N),$$

where $V(F)$ is the **Hardy–Krause variation** of F. The Halton sequence, the Sobol sequence, or the Faure sequence permits to construct $\mathcal{X}$ such that

$$D(\mathcal{X}_N) = \mathcal{O}\left\{ \frac{\log(N)^d}{N} \right\}$$

- "Almost" $\mathcal{O}(1/N)$ convergence!
- But the constant in $\mathcal{O}$ depends on d: works well in moderate dimension (comparable to sparse grids)

[Caflisch: Monte carlo and quasi-monte carlo methods, Acta numerica 7 (1998): 1-49.]


MC
DPP
QMC
N = 10
N = 50
N = 100
N = 500

Variance reduction for Monte Carlo

$$\frac{\mathbb{E}[(\mathrm{I} - \widehat{\mathrm{I}}_N)^2]^{1/2}}{\mathrm{I}} = \frac{1}{\sqrt{N}} \frac{\sqrt{\mathrm{Var}(F(X))}}{\mathrm{I}}$$

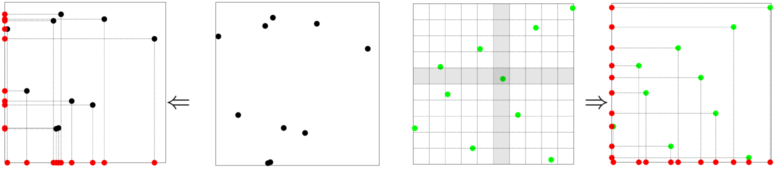
Instead of trying to improve the convergence rate $1/\sqrt{N}$ of Monte Carlo, **variance reduction techniques** aim at reducing the constant in front.

- Latin Hypercube Sampling
- Importance Sampling
- Control Variables
- ...

[Rubinstein and Kroese: Simulation and the Monte Carlo method, Wiley 2016]

Latin hypercube sample (LHS) for $\pi_X = \mathcal{U}([0,1]^d)$

Idea: create a sample which **represents well all 1D marginals** $X_1, \ldots, X_d$



(a) Marginals (b) MC (c) LHS (d) Marginals

The resulting estimator $\widehat{\mathrm{I}}_N^{LHS}$ is still **unbiased** and **convergent** with

$$\frac{\mathbb{E}[(\mathrm{I}-\widehat{\mathrm{I}}_N^{LHS})^2]^{1/2}}{\mathrm{I}} = \frac{1}{\sqrt{N}}\frac{\sqrt{\mathrm{Var}(F(X)-F_{\mathrm{add}}(X))}}{\mathrm{I}} + o\left(\frac{1}{\sqrt{N}}\right)$$

where $F_{\mathrm{add}}(X) = F_1(X_1) + \ldots + F_d(X_d)$ is the ℓ^2-best additive approximation to $F(X)$. No recycling for $N \leftarrow N+1$... still, super popular!

More generally, space filling designs consist typically in

$$\max_{X^{(1)},\ldots,X^{(N)}} \min_{i\neq j} \|X^{(i)} - X^{(j)}\|$$

[Stein: Large sample properties of simulations using LHS, Technometrics 1987]

[Pronzato: Minimax and maximin space-filling designs: some properties and methods for construction, J-SFdS 2017]

Importance Sampling using an importance density ρ

Idea: estimate

$$\mathrm{I} = \int F(x)\pi_X(x)\mathrm{d}x = \int F(x)\frac{\pi_X(x)}{\rho(x)}\rho(x)\mathrm{d}x$$

with

$$\widehat{\mathrm{I}}_N^{\mathrm{IS}} = \frac{1}{N}\sum_{i=1}^{N} F(X^{(i)})\frac{\pi_X(X^{(i)})}{\rho(X^{(i)})}, \qquad \text{where } X^{(i)} \sim \rho$$

This is an **unbiased** and **convergent** estimator with

$$\frac{\mathbb{E}[(\mathrm{I}-\widehat{\mathrm{I}}_N^{\mathrm{IS}})^2]^{1/2}}{\mathrm{I}} = \frac{1}{\sqrt{N}}\frac{\sqrt{\mathrm{Var}_{X\sim\rho}\left(\frac{F(X)\pi_X(X)}{\rho(X)}\right)}}{\mathrm{I}}$$

Observe that, if $F(X) \geq 0$, the optimal choice

$$\rho^{\mathrm{opt}}(x) = \frac{F(x)\pi_X(x)}{\mathrm{I}} \qquad \Rightarrow \qquad \frac{\mathbb{E}[(\mathrm{I}-\widehat{\mathrm{I}}_N^{\mathrm{IS}})^2]^{1/2}}{\mathrm{I}} = 0$$

... but ρ^{opt} depends on I: **sequential/adaptive** methods to approximate ρ^{opt}.

Illustration for **rare event estimation** $\mathbb{P}[u(X) \geq \alpha]$

For $F(x) = \mathbf{1}_{[\alpha,\infty)}(u(x))$ we have

$$\frac{\mathbb{E}[(\mathrm{I} - \widehat{\mathrm{I}}_N^{\mathsf{MC}})^2]^{1/2}}{\mathrm{I}} = \sqrt{\frac{1-\mathrm{I}}{N\,\mathrm{I}}}$$

We need at least $N = \mathcal{O}(\mathrm{I}^{-1})$ to hit the failure domaine $\{x : u(x) \geq \alpha\}$

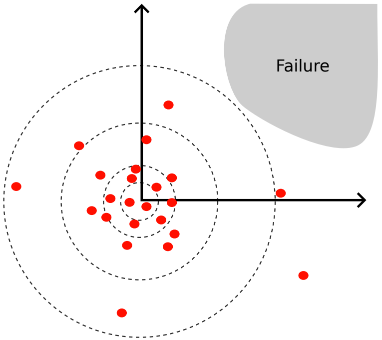


Illustration for **rare event estimation** $\mathbb{P}[u(X) \geq \alpha]$

For $F(x) = \mathbf{1}_{[\alpha,\infty)}(u(x))$ we have

$$\frac{\mathbb{E}[(\mathrm{I} - \widehat{\mathrm{I}}_N^{\mathrm{MC}})^2]^{1/2}}{\mathrm{I}} = \sqrt{\frac{1-\mathrm{I}}{N\,\mathrm{I}}}$$

We need at least $N = \mathcal{O}(\mathrm{I}^{-1})$ to hit the failure domaine $\{x : u(x) \geq \alpha\}$

A simple importance sampling scheme: first, find the most probable failure point x^* by solving

$$\max_{u(x)\geq\alpha} \pi_X(x)$$

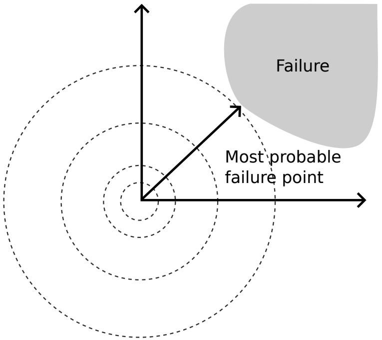


Illustration for **rare event estimation** $\mathbb{P}[u(X) \geq \alpha]$

For $F(x) = \mathbf{1}_{[\alpha,\infty)}(u(x))$ we have

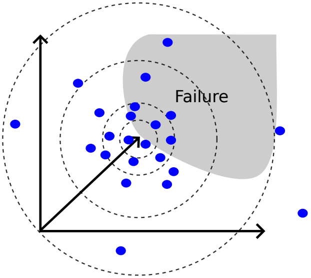
$$\frac{\mathbb{E}[(\mathrm{I} - \widehat{\mathrm{I}}_N^{MC})^2]^{1/2}}{\mathrm{I}} = \sqrt{\frac{1-\mathrm{I}}{N\,\mathrm{I}}}$$

We need at least $N = \mathcal{O}(\mathrm{I}^{-1})$ to hit the failure domaine $\{x : u(x) \geq \alpha\}$

A simple importance sampling scheme: first, find the most probable failure point x^* by solving

$$\max_{u(x)\geq\alpha} \pi_X(x)$$

Then use $\rho(x) = \pi_X(x - x^*)$ as an importance density.



Control variable for variance reduction

Idea: assuming we are given $\widetilde{F}(x)$ such that $\mathbb{E}[\widetilde{F}(X)]$ is known, we estimate

$$\mathrm{I} = \mathbb{E}[F(X)] = \mathbb{E}[F(X) - \widetilde{F}(X)] + \mathbb{E}[\widetilde{F}(X)]$$

with

$$\widehat{\mathrm{I}}_N^{\mathsf{CV}} = \left\{ \frac{1}{N} \sum_{i=1}^{N} F(X^{(i)}) - \widetilde{F}(X^{(i)}) \right\} + \mathbb{E}[\widetilde{F}(X)]$$

This is again an **unbiased** and **convergent** estimator with

$$\frac{\mathbb{E}[(\mathrm{I} - \widehat{\mathrm{I}}_N^{\mathsf{CV}})^2]^{1/2}}{\mathrm{I}} = \frac{1}{\sqrt{N}} \frac{\sqrt{\mathrm{Var}(F(X) - \widetilde{F}(X))}}{\mathrm{I}}$$

In practice, we *just* need to find a $\widetilde{F}(x) \approx F(x)$ such that

$$\mathrm{Var}(F(X) - \widetilde{F}(X)) \leq \mathrm{Var}(F(X))$$

or equivalently $\mathbb{E}[\|F(X) - \widetilde{F}(X)\|^2] \leq \mathbb{E}[\|F(X)\|^2]$ ($\rightarrow$ **surrogate models**)

Control variable and its variants...

- Replace $\widetilde{F}(x)$ with $\theta\widetilde{F}(x)$ and optimize over $\theta \in \mathbb{R}$:

$$\min_{\theta\in\mathbb{R}} \mathrm{Var}(F(X) - \theta\widetilde{F}(X)) = \mathrm{Var}(F(X)) - \frac{\mathrm{Cov}(F(X), \widetilde{F}(X))^2}{\mathrm{Var}(\widetilde{F}(X))}$$

- If $\mathbb{E}[\widetilde{F}(X)]$ is **unknown** but $x \mapsto \widetilde{F}(x)$ is **cheap-to-evaluate**, then:

$$\widehat{\mathrm{I}}_N^{\mathrm{CV}} = \left\{ \frac{1}{N} \sum_{i=1}^{N} F(X^{(i)}) - \widetilde{F}(X^{(i)}) \right\} + \left\{ \frac{1}{M} \sum_{i=N+1}^{N+M} \widetilde{F}(X^{(i)}) \right\}, \qquad M \gg N$$

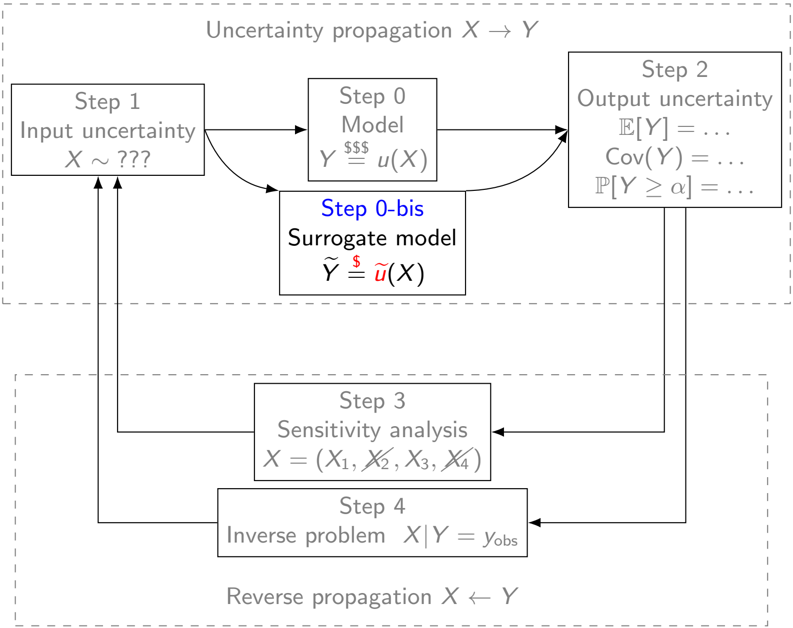
- Multiple control variables $\widetilde{F}_1(x), \widetilde{F}_2(x), \ldots$: use **telescoping sums**

$$\widehat{\mathrm{I}}_N^{\mathrm{ML}} = \left\{ \frac{1}{N_1} \sum_{i=1}^{N} F(X^{(i)}) - \widetilde{F}_1(X^{(i)}) \right\} + \left\{ \frac{1}{N_2} \sum_{i=N_1}^{N_1+N_2} \widetilde{F}_1(X^{(i)}) - \widetilde{F}_2(X^{(i)}) \right\} + \ldots$$

  Depending on the context, this is called **multi-level** or **multi-index** or **multi-fidelity**. In some cases, we know the optimal balance between the levels $N_1, N_2, \ldots$, see:

[Giles: Multilevel monte carlo methods, Acta numerica 2015]

[Peherstorfer, Willcox and Gunzburger: Survey of multifidelity methods in uncertainty propagation, inference, and optimization, Siam Review 2018]


Uncertainty propagation $X \to Y$
Step 1
Input uncertainty
$X \sim ???$
Step 0
Model
$Y \overset{\$\$\$}{=} u(X)$
Step 2
Output uncertainty
$\mathbb{E}[Y] = \dots$
$\mathrm{Cov}(Y) = \dots$
$\mathbb{P}[Y \geq \alpha] = \dots$
Step 0-bis
Surrogate model
$\widetilde{Y} \overset{\$}{=} \widetilde{u}(X)$
Step 3
Sensitivity analysis
$X = (X_1, \cancel{X_2}, X_3, \cancel{X_4})$
Step 4
Inverse problem $X|Y = y_{\text{obs}}$
Reverse propagation $X \leftarrow Y$

Uncertainty propagation via a **surrogate models**

Idea: **replace the model** u with an approximation $\widetilde{u}$ and Y with

$$\widetilde{Y} = \widetilde{u}(X)$$

Alternatively, use $\psi(\widetilde{Y})$ as a **control variable** for $\psi(Y)$.

- If $\widetilde{u}$ is **simple** (linear, quadratic) then analytic computation of $\mathbb{E}[\psi(\widetilde{Y})]$
- If $\widetilde{u}$ is **cheap to evaluate**, then use the preceding methods with $N \gg 1$

Constructing $\widetilde{u}$ is an art: depending on the context, such $\widetilde{u}$ are readily available (*e.g.* **crude mesh, simplified physics** etc). If not, there is a zoology of methods to construct $\widetilde{u}$ from either

- point evaluations of u,
- residual of the equation solved by u,
- prior knowledge on u,
- ...

Local approximation via Taylor expansion

Taylor expansion of $u(X)$ around $m = \mathbb{E}[X]$

$$\widetilde{Y} = u(m) + \nabla u(m)^\top (X - m)$$

We just need to compute $u(m)$ and $\nabla u(m)$, which requires at most $N = d + 1$ evaluations of the model (using finite differences). Permits to **rapidely sketch the trends** of Y via

$$\mathbb{E}[Y] \approx \mathbb{E}[\widetilde{Y}] = u(m)$$

and

$$\mathrm{Var}(Y) \approx \mathrm{Var}(\widetilde{Y}) = \nabla u(m)^\top \mathrm{Cov}(X) \nabla u(m)$$

Basic: works for models which can be **linearized**...

Taylor expansion for rare event estimation $\mathbb{P}[u(X) \geq \alpha]$

First, find the **most probable failure point** x^* by solving

$$\max_{u(x)\geq\alpha} \pi_X(x)$$

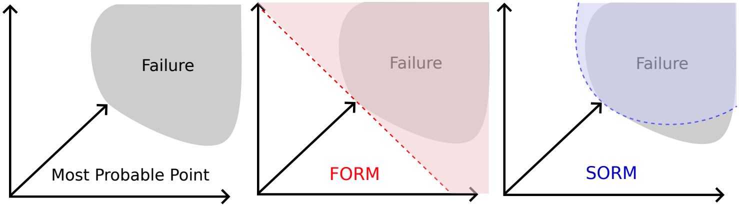
Then, compute $\mathbb{P}(\widetilde{Y} \geq \alpha)$ **analytically** with

$$\widetilde{Y} = \underbrace{u(x^*) + \nabla u(x^*)^\top (X - x^*)}_{\text{FORM}} + \underbrace{\frac{1}{2}(X - x^*)^\top \nabla^2 u(x^*)(X - x^*)}_{\text{SORM}}$$

- FORM (First-Order Reliability Method): we have

$$\mathbb{P}(\widetilde{Y} \geq \alpha) \overset{\pi_X = \mathcal{N}(0, I_d)}{=} \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{\|x^*\|}{\sqrt{2}}\right)$$

- SORM (Second-Order): $\mathbb{P}(\widetilde{Y} \geq \alpha) =$ Breitung's formula.



[Ditlevsen and Madsen: Structural reliability methods, Wiley 1996]

Towards **global approximation**: the curse of dimensionality

- **Q:** I want to construct an approximation $\widetilde{u}$ to $u(x_1, \dots, x_d)$ such that

$$\|u - \widetilde{u}\|_\infty \leq \varepsilon \|u\|_\infty$$

  How many point evaluations of u do I need?
- **A:** Well, if u is linear, then $N = d + 1$ evaluations are enough.
- **Q:** Okay... what if u is *just* extremely regular, say,

$$\sup_{\alpha \in \mathbb{N}^d} \left\| \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \right\|_\infty < \infty$$

- **A:** Sorry: for any algorithms you can ever think of, there exists such a u which would require at least

$$N \geq 2^{\lfloor d/2 \rfloor}$$

- **Q:** Come one! I remember that with polynomial interpolations, I can reach

$$\|u - \widetilde{u}_{\text{interpolation}}\|_\infty = \mathcal{O}(\rho^N)$$

- **A:** Sure! But the constant hidden in $\mathcal{O}$ depends in d. You'll need at least $N \geq 2^{\lfloor d/2 \rfloor}$ to be sure to reach the asymptotic regime.

[Novak and Woźniakowski: **Approximation of infinitely differentiable multivariate functions is intractable**, Journal of Complexity 2009]

Exploit some **low-dimensional structure** that u can have

- **Sparsity**

$$u(x) \approx \sum_{\alpha \in \Lambda_N} u_\alpha \varphi_\alpha(x), \qquad \#\Lambda_N = N$$

- **Low-rank structure**

$$u(x) \approx \sum_{i=1}^{r} u_1^i(x_1) \dots u_d^i(x_d)$$

- **Low-effective dimension**

$$u(x) \approx f(z_1, \dots, z_m), \quad \begin{cases} z = g(x) \\ g : \mathbb{R}^d \to \mathbb{R}^m \\ m \ll d \end{cases}$$

- **...**

Prototypical example: **parametrized elliptic PDE**

Find $u(x) \in H^1(\Omega)$ solution to

$$-\text{div}(\kappa(x)\nabla u(x)) = f \qquad \text{in } \Omega \subset \mathbb{R}^2 \text{ or } \mathbb{R}^3$$

where the diffusion coefficient writes

$$\kappa(x,s) = \kappa_0(s) + \sum_{i=1}^{\infty} x_i\, \kappa_i(s), \qquad \begin{cases} x_1, x_2, \ldots \in [-1,1] \\ s \in \Omega \end{cases}$$

Assume

$$(\|\kappa_i\|_\infty)_{i\geq 1} \in \ell^p \qquad \text{for some } p < 1$$

then there exists

$$\widetilde{u}(x,s) = \sum_{i=1}^{n} \varphi_i(x) v_i(s), \qquad \begin{cases} v_i \in H^1(\Omega) \\ \varphi_i \in L^\infty([-1,1]) \end{cases}$$

such that

$$\boxed{\sup_{x\in[-1,1]^{\mathbb{N}}} \|u(x) - \widetilde{u}(x)\|_{H^1(\Omega)} \leq C n^{-s}, \quad \text{where } s = p^{-1} - 1 > 0}$$

for some constant C: **no curse of dimensionality!**

[Cohen&DeVore: Approximation of high-dimensional parametric PDEs, Acta Numerica 2015]

Near-optimal approximations can be obtained using

- **sparse polynomial** expansions:

$$\widetilde{u}(x,s) = \sum_{\substack{\alpha \in \Lambda_n \\ \#\Lambda_n = n}} \varphi_\alpha(x) v_\alpha(s) \qquad \begin{cases} \varphi_\alpha(x) : \text{given multivariate polynomials} \\ \Lambda_n : \textbf{Greedy algorithm } \Lambda_{n+1} = \Lambda_n \cup \{\alpha^*_{n+1}\} \\ v_\alpha(s) : \text{least-squares, interpolation, ...} \end{cases}$$

- the **Reduced Basis** method:

$$\widetilde{u}(x,s) = \sum_{i=1}^{n} \varphi_i(x) u(x_i, s) \qquad \begin{cases} x_1, \dots, x_n : \textbf{Greedy algorithm } n \leftarrow n+1 \\ \varphi_i(x) : \text{Galerkin projection of } u(x) \text{ on} \\ \qquad \text{span}(u(x_1), \dots, u(x_n)) \end{cases}$$

Rozza, Huynh and Patera: Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive PDE Arch. Comput. Methods Eng. 2008

Blatman and Sudret: Adaptive sparse polynomial chaos expansion based on least angle regression JCP, 2011

Chkifa, Cohen and Schwab: Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs Journal de Mathématiques Pures et Appliquées, 2015

The **Reduced Basis method**

For $A(x) \in \mathbb{R}^{m\times m}$ and $b \in \mathbb{R}^m$ with $m \gg 1$, compute $u(x) \in \mathbb{R}^m$ solution to

$$A(x)u(x) = b$$

- **Offline phase:** compute $n \ll m$ solutions $u(x^{(1)}), \dots, u(x^{(n)})$ and
$$V_n = [u(x^{(n)}), \dots, u(x^{(n)})] \in \mathbb{R}^{m\times n}$$
- **Online phase:** given a new parameter x, compute the Galerkin projection $u_n(x)$ of $u(x)$ onto range(V_n) by computing $\widetilde{u}_n(x) \in \mathbb{R}^n$ solution to
$$[V_n^\top A(x) V_n]\widetilde{u}_n(x) = [V_n^\top b] \quad \Rightarrow \quad u_n(x) = V_n \widetilde{u}_n(x)$$

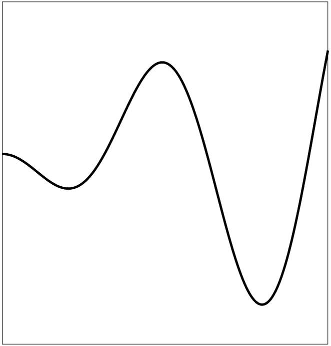
Remarks:

- $V_n \leftarrow \mathrm{qr}(V_n)$ for numerical stability
- Greedy enrichment $n \leftarrow n+1$ via $x^{(n+1)} \in \arg\max_x \|A(x)u_n(x) - b\|$
- If $A(x) = \sum_{i=1}^r c_i(x) A_i$ admits an affine parametric decomposition then

$$[V_n^\top A(x) V_n] = \sum_{i=1}^r c_i(x) \underbrace{[V_n^\top A_i V_n]}_{\substack{\text{precompute for} \\ \text{online efficieny}}}$$

Kriging: surrogate models via **Gaussian processes** (GP)

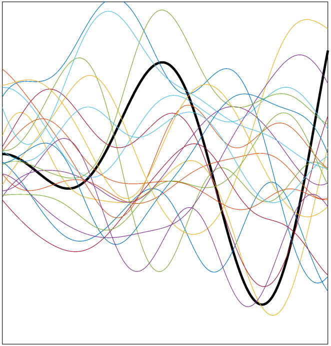
- Goal: approximate $u(x)$



Kriging: surrogate models via **Gaussian processes** (GP)

- Goal: approximate $u(x)$
- **Idea:** model u as a **realization of a GP** in the variable x

$$Z \sim \mathcal{N}(0, c)$$



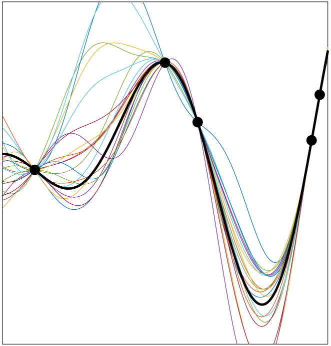
Kriging: surrogate models via **Gaussian processes** (GP)

- Goal: approximate $u(x)$
- **Idea:** model u as a **realization of a GP** in the variable x

$$Z \sim \mathcal{N}(0, c)$$

- Evaluate the model at $x^{\text{obs}} = \{x^{(1)}, \dots, x^{(N)}\}$ and condition Z on $Z^{\text{obs}} = u(x^{\text{obs}})$

$$Z | Z^{\text{obs}} \sim \mathcal{N}(m', c')$$



Kriging: surrogate models via **Gaussian processes** (GP)

- Goal: approximate $u(x)$
- **Idea:** model u as a **realization of a GP** in the variable x

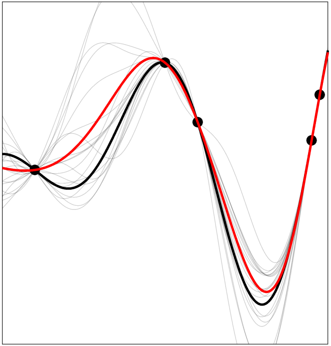
$$Z \sim \mathcal{N}(0, c)$$

- Evaluate the model at $x^{\text{obs}} = \{x^{(1)}, \ldots, x^{(N)}\}$ and condition Z on $Z^{\text{obs}} = u(x^{\text{obs}})$

$$Z | Z^{\text{obs}} \sim \mathcal{N}(m', c')$$

- Use the mean as a surrogate $\widetilde{u}(x) = m'(x)$ where

$$m'(x) = c(x, x^{\text{obs}})[c(x^{\text{obs}}, x^{\text{obs}})]^{-1} u(x^{\text{obs}})$$



Kriging: surrogate models via **Gaussian processes** (GP)

- Goal: approximate $u(x)$
- **Idea:** model u as a **realization of a GP** in the variable x

$$Z \sim \mathcal{N}(0, c)$$

- Evaluate the model at $x^{\text{obs}} = \{x^{(1)}, \dots, x^{(N)}\}$ and condition Z on $Z^{\text{obs}} = u(x^{\text{obs}})$

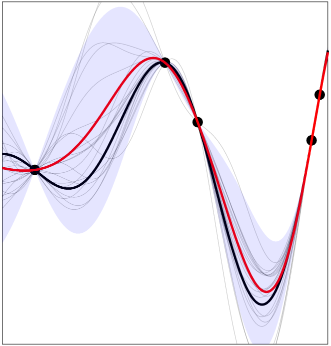
$$Z | Z^{\text{obs}} \sim \mathcal{N}(m', c')$$

- Use the mean as a surrogate $\widetilde{u}(x) = m'(x)$ where

$$m'(x) = c(x, x^{\text{obs}})[c(x^{\text{obs}}, x^{\text{obs}})]^{-1} u(x^{\text{obs}})$$

- Confidence intervals via the conditional variance c'

$$c'(x, y) = c(x, y) - c(x, x^{\text{obs}})[c(x^{\text{obs}}, x^{\text{obs}})]^{-1} c(x^{\text{obs}}, y)$$



Kriging: surrogate models via **Gaussian processes** (GP)

- Goal: approximate $u(x)$
- **Idea:** model u as a **realization of a GP** in the variable x

$$Z \sim \mathcal{N}(0, c)$$

- Evaluate the model at $x^{\mathrm{obs}} = \{x^{(1)}, \dots, x^{(N)}\}$ and condition Z on $Z^{\mathrm{obs}} = u(x^{\mathrm{obs}})$

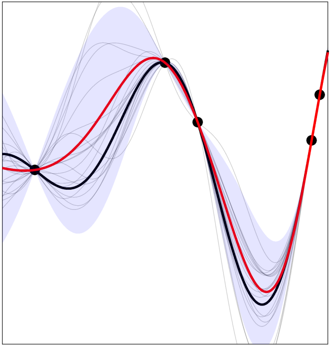
$$Z|Z^{\mathrm{obs}} \sim \mathcal{N}(m', c')$$

- Use the mean as a surrogate $\widetilde{u}(x) = m'(x)$ where

$$m'(x) = c(x, x^{\mathrm{obs}})[c(x^{\mathrm{obs}}, x^{\mathrm{obs}})]^{-1} u(x^{\mathrm{obs}})$$

- Confidence intervals via the conditional variance c'

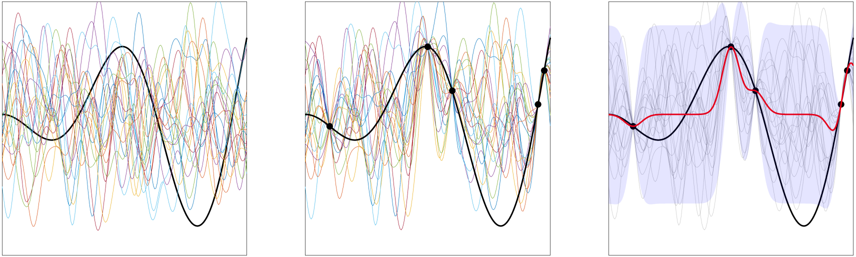
$$c'(x,y) = c(x,y) - c(x, x^{\mathrm{obs}})[c(x^{\mathrm{obs}}, x^{\mathrm{obs}})]^{-1} c(x^{\mathrm{obs}}, y)$$



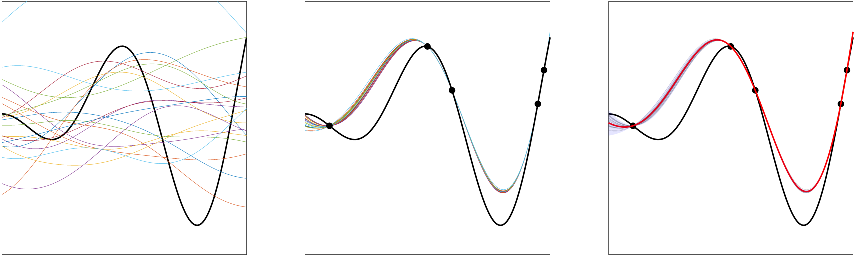
Remarks:

- There exists (infinitely) many ways to enrich $x^{\mathrm{obs}} \leftarrow x^{\mathrm{obs}} \cup \{x^{(N+1)}\}$.
- Requires solving a $N \times N$ linear system,
- $u \sim \mathcal{N}(0, c)$ is a very strong assumption (how to choose c?)
- Still, extremely popular!!

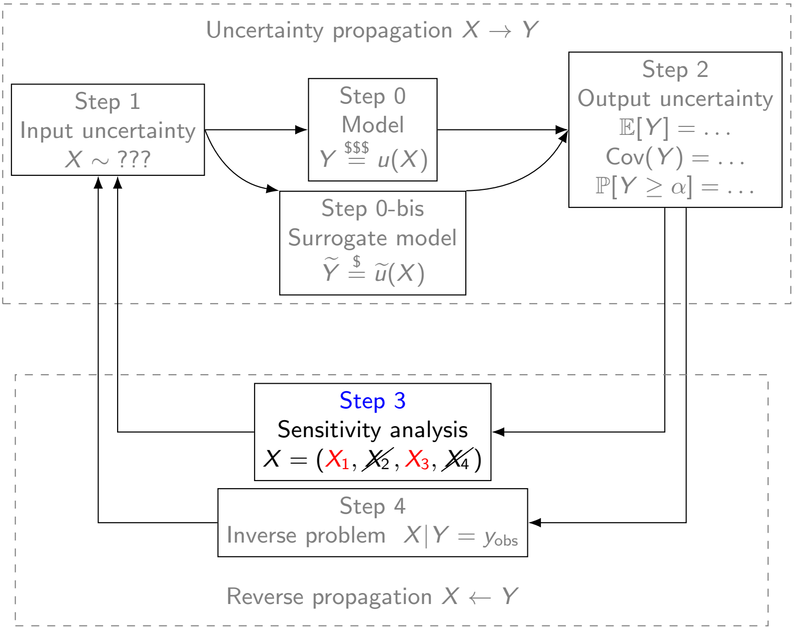
$c(x,y) = \exp(-\|x-y\|^2/2\ell^2)$ with a wrong length scale ℓ



$\ell = 0.5$ too small $\rightarrow$ "swiss cheese"



$\ell = 5$ too large $\rightarrow$ polynomial interpolation



Uncertainty propagation $X \to Y$
Step 1
Input uncertainty
$X \sim ???$
Step 0
Model
$Y \overset{\$\$\$}{=} u(X)$
Step 0-bis
Surrogate model
$\widetilde{Y} \overset{\$}{=} \widetilde{u}(X)$
Step 2
Output uncertainty
$\mathbb{E}[Y] = \dots$
$\mathrm{Cov}(Y) = \dots$
$\mathbb{P}[Y \geq \alpha] = \dots$
Step 3
Sensitivity analysis
$X = (X_1, X_2, X_3, X_4)$
Step 4
Inverse problem $X|Y = y_{\mathrm{obs}}$
Reverse propagation $X \leftarrow Y$

Global Sensitivity Analysis

$$Y = u(X_1, \dots, X_d)$$

- $X \sim \pi_X$: input parameter, typically with product density π_X
- Y: output of interest, generally scalar $Y \in \mathbb{R}$
- u: expensive numerical model

Goal: determine the relative influence of the inputs $X_1, \dots, X_d$ on Y. Formally, for any $\tau \subset \{1, \dots, d\}$, we want to define and compute a sensitivity index which measures how well

$$u(X_1, \dots, X_d) \approx f(X_{\tau_1}, \dots, X_{\tau_m})$$

Remark: super useful to construct low-dimensional meta models f later on!

[Da Veiga, Gamboa, Iooss, and Prieur: Basics and trends in sensitivity analysis SIAM 2021.]

The **function approximation** perspective

Let $L^2_{\pi_X}$ be the space of square-integrable functions endowed with the norm

$$\|u\|^2 = \int u(x)^2 \mathrm{d}\pi_X(x)$$

Expectations and conditional expectations are orthogonal projections in $L^2_{\pi_X}$:

- The constant $c \in \mathbb{R}$ which best approximates u in $L^2_{\pi_X}$ is the **expectation** $c = \mathbb{E}[u(X)]$

$$\min_{c\in\mathbb{R}} \|u - c\|^2 =: \mathrm{Var}(u(X))$$

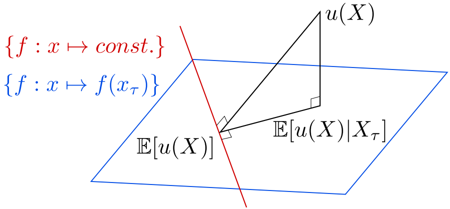
- For any $\tau \subset \{1, \dots, d\}$, the function $f : x \mapsto f(x_{\tau_1}, \dots, x_{\tau_m})$ which best approximates u in $L^2_{\pi_X}$

$$\min_{f:x\mapsto f(x_\tau)} \|u - f\|^2 =: \mathbb{E}[\mathrm{Var}(u(X)|X_\tau)]$$

is the **conditional expectation** $f(x_\tau) = \mathbb{E}[u(X)|X_\tau = x_\tau]$.

The **total variance formula**: Pythagorean theorem in $L^2_{\pi_X}$

$$
\begin{aligned}
\|u - \mathbb{E}[u(X)]\|^2 &= \|(u - \mathbb{E}[u(X)|X_\tau]) + (\mathbb{E}[u(X)|X_\tau] - \mathbb{E}[u(X)])\|^2 \\
&= \|u - \mathbb{E}[u(X)|X_\tau]\|^2 + \|\mathbb{E}[u(X)] - \mathbb{E}[u(X)|X_\tau]\|^2
\end{aligned}
$$



Put in statistical language:

$$
\mathrm{Var}(u(X)) = \underbrace{\mathbb{E}[\mathrm{Var}(u(X)|X_\tau)]}_{\min_{f : x \mapsto f(x_\tau)} \|u - f\|^2} + \mathrm{Var}(\mathbb{E}[u(X)|X_\tau]) \qquad (\star)
$$

Connection with the **Sobol' indices**

The **closed Sobol' indices** writes

$$S_\tau(u) := \frac{\mathrm{Var}(\mathbb{E}[u(X)|X_\tau])}{\mathrm{Var}(u(X))} \overset{(\star)}{=} 1 - \frac{\min_{f: x \mapsto f(x_\tau)} \|u - f\|^2}{\mathrm{Var}(u(X))}$$

$$\begin{aligned} S_\tau(u) \approx 1 \quad &\Leftrightarrow \quad u(X) \approx f(X_\tau) \\ &\Leftrightarrow \quad X_\tau \text{ ``explains'' well } Y = u(X) \end{aligned}$$

Similarly, the **total Sobol' indices** writes

$$T_\tau(u) := 1 - \frac{\mathrm{Var}(\mathbb{E}[u(X)|X_{-\tau}])}{\mathrm{Var}(u(X))} \overset{(\star)}{=} \frac{\min_{f: x \mapsto f(x_{-\tau})} \|u - f\|^2}{\mathrm{Var}(u(X))}$$

$$\begin{aligned} T_\tau(u) \approx 0 \quad &\Leftrightarrow \quad u(X) \approx f(X_{-\tau}) \\ &\Leftrightarrow \quad X_\tau \text{ is useless to ``explain'' } Y = u(X) \end{aligned}$$

Link with the **ANOVA** decomposition

Assuming π_X is a product density, the **AN**alysis **O**f **VA**riance of u reads

$$u(x) = u_0 + \sum_{i=1}^{d} u_i(x_i) + \sum_{i \neq j}^{d,d} u_{i,j}(x_i, x_j) + \sum_{i \neq j \neq k}^{d,d,d} u_{i,j,k}(x_i, x_j, x_k) + \dots$$

where all above terms are **pairwise orthogonals** in $L^2_{\pi_X}$.

- Closed Sobol' index
$$S_\tau(u) = \sum_{\alpha \subset \tau} \mathrm{Var}(u_\alpha)$$
- Total Sobol' index
$$T_\tau(u) = \sum_{\alpha \cap \tau \neq \emptyset} \mathrm{Var}(u_\alpha)$$
- Superset importance
$$\Upsilon^2_\tau(u) = \sum_{\alpha \supset \tau} \mathrm{Var}(u_\alpha)$$
- Shapley-Owen value
$$\phi_\tau(u) = \sum_{\alpha \supset \tau} \frac{\mathrm{Var}(u_\alpha)}{|\alpha| - |\tau| + 1}$$
- ...

Pick & freeze estimators of Sobol' indices

Assuming π_X is a product density, the following identities hold

$$\mathrm{Var}(\mathbb{E}[u(X)|X_\tau]) = \mathrm{Cov}(u(X), u(X_\tau, X'_{-\tau}))$$
$$\mathbb{E}[\mathrm{Var}(u(X)|X_{-\tau})] = \frac{1}{2}\mathbb{E}[(u(X) - u(X_{-\tau}, X'_\tau))^2]$$

where X' is an independent copy of X.

- Estimation of closed Sobol' indices

$$S_\tau(u) \approx \frac{\frac{1}{N}\sum_{i=1}^N u(X^{(i)})u(X_\tau^{(i)}, X'^{(i)}_{-\tau}) - \left(\frac{1}{N}\sum_{i=1}^N u(X^{(i)})\right)\left(\frac{1}{N}\sum_{i=1}^N u(X_\tau^{(i)}, X'^{(i)}_{-\tau})\right)}{\frac{1}{N}\sum_{i=1}^N u(X^{(i)})^2 - \left(\frac{1}{N}\sum_{i=1}^N u(X^{(i)})\right)^2}$$

- Estimation of total Sobol' indices

$$T_\tau(u) \approx \frac{\frac{1}{2N}\sum_{i=1}^N (u(X^{(i)}) - u(X_{-\tau}^{(i)}, X'^{(i)}_\tau))^2}{\frac{1}{N}\sum_{i=1}^N u(X^{(i)})^2 - \left(\frac{1}{N}\sum_{i=1}^N u(X^{(i)})\right)^2}$$

Remarks: Requires $2N$ model evaluations. No recycling possible for estimating the indices for another τ: estimating **all first order indices** $\#\tau = 1$ would require $(d+1)N$ evaluations.

Gradient-based global sensitivity analysis

Suppose we have access to

$$x \mapsto \nabla u(x) = \begin{pmatrix} \partial_1 u(x) \\ \vdots \\ \partial_d u(x) \end{pmatrix}$$

via *e.g.* adjoint models, automatic differentiation...

- $|\partial_i u(x)|$ gives a **local sensitivity measure** of the i-th variable around x.
- **Global sensitivity measure** can be obtained *e.g.* using the Derivative Based Sensitivity Measure (DGSM)

$$\nu_i(u) = \mathbb{E}[\partial_i u(X)^2]$$

- The Monte Carlo estimator requires N evaluations of ∇u to estimate simultaneously all $\nu_i(u)$'s:

$$\begin{pmatrix} \nu_1(u) \\ \vdots \\ \nu_d(u) \end{pmatrix} \approx \frac{1}{N} \sum_{i=1}^{N} \nabla u(X^{(i)})^{\circ 2}$$

- Assuming π_X is a product density and let $C(\pi_{X_i})$ be the Poincaré constant of X_i. Then

$$T_i(u) \leq C(\pi_{X_i}) \nu_i(u)$$

Active subspaces: rotation in the parameter space

Instead of $u(X) \approx f(X_\tau)$, we seek

$$u(X) \approx f(U_m^T X)$$

for some function $f : \mathbb{R}^m \to \mathbb{R}$ and some matrix $U_m \in \mathbb{R}^{d\times m}$ with $U_m^\top U_m = I_m$.

- The optimal f for a given U_m is the conditional expectation

$$\|u - \mathbb{E}[u(X)|U_m^T X]\|^2 = \min_{f:\mathbb{R}^m\to\mathbb{R}} \|u - f(U_m^T \cdot)\|^2$$

- Bound the error using subspace Poincaré inequality

$$\|u - \mathbb{E}[u(X)|U_m^T X]\|^2 \leq \overline{C}(\pi_X)(\mathbb{E}[\|\nabla u(X)\|^2] - \mathbb{E}[\|U_m^\top \nabla u(X)\|^2])$$

- Minimizing the bound yields the **active subspace**: $U_m = [v_1, \ldots, v_m]$ contains the m-largest eigenvectors of

$$H = \mathbb{E}\left[\nabla u(X) \nabla u(X)^\top\right] = \sum_{i=1}^{d} \lambda_i v_i v_i^\top$$

  and the error becomes

$$\|u - \mathbb{E}[u(X)|U_r^T X]\|^2 \leq \overline{C}(\pi_X) \sum_{i=m+1}^{d} \lambda_i$$

[Constantine, Dow and Wang: Active subspace methods in theory and practice: applications to kriging surfaces, SIAM-SISC 2014]

Two examples

Assume $u(x) = f(A_r^T x)$ is a **ridge function** for some $A_r \in \mathbb{R}^{d\times m}$. Since $\nabla u(x) = A_r \nabla f(A_r^T x)$, we have

$$H = \mathbb{E}\left[\nabla u(X) \nabla u(X)^\top\right] = A_r \mathbb{E}\left[\nabla f(A_r^T X) \nabla f(A_r^T X)^T\right] A_r^T$$

Then $\lambda = (\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$ and $\text{range}(U_m) = \text{range}(A_r)$.

Assume $u(x) = f(\|x\|)$ and $\pi_X(x) \propto \rho(\|x\|)$ are **isotropic functions**, then

$$H = \mathbb{E}\left[\nabla u(X) \nabla u(X)^\top\right] \propto I_d$$

No decay in the spectrum $\lambda = (1, \ldots, 1)$: no dimension reduction.

Extensions (part of my current research)

- Joint **input–output** reduction of $u : \mathbb{R}^d \to \mathbb{R}^m$ [Chen et al, 2023]

$$\mathbb{E}[\|u(X) - V_s f(U_m^\top X)\|^2] \leq \overline{\mathbb{C}}(\pi_X)(\mathbb{E}[\|\nabla u(X)\|_F^2] - \mathbb{E}[\|V_s^\top \nabla u(X) U_m\|_F^2])$$

where $\nabla u(x) \in \mathbb{R}^{m \times d}$ is the Jacobian of $u(x)$

- **Nonlinear version** [Bigoni et al, 2022]: for "any" $g : \mathbb{R}^d \to \mathbb{R}^r$ we have

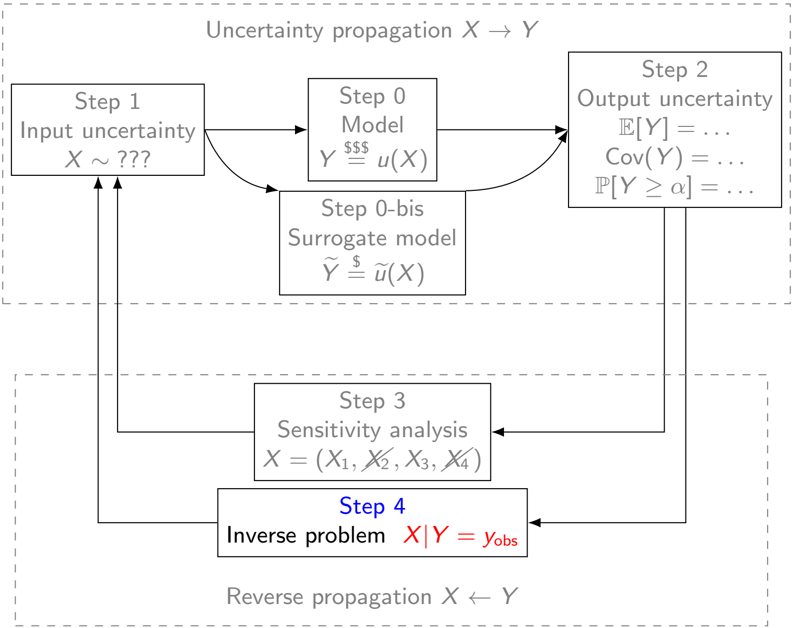
$$\mathbb{E}[(u(X) - f(g(X)))^2] \leq \overline{\mathbb{C}}(\pi_X|\mathcal{G})\, \mathbb{E}[\|\Pi_{\ker(\nabla g(X))} \nabla u\|^2]$$

where $\Pi_{\ker(\nabla g(X))}$ is the orthogonal projector onto $\ker(\nabla g(X))$.

  - Minimizing the RHS over g corresponds to **aligning** the Jacobian of g with the gradient of u.
  - The function g must have **path-connected level sets**, which is not trivial to impose, unless (work in progress [Verdière et al, 2023])

$$g(x) = (\varphi_1(x), \dots, \varphi_m(x)), \quad \varphi \in \mathrm{Diff}(\mathbb{R}^d; \mathbb{R}^d)$$

  - **Many connections with machine learning**: deep approximation, autoencoders, normalizing flows,...


Uncertainty propagation $X \to Y$
Step 1
Input uncertainty
$X \sim ???$
Step 0
Model
$Y \overset{\$\$\$}{=} u(X)$
Step 0-bis
Surrogate model
$\widetilde{Y} \overset{\$}{=} \widetilde{u}(X)$
Step 2
Output uncertainty
$\mathbb{E}[Y] = \dots$
$\mathrm{Cov}(Y) = \dots$
$\mathbb{P}[Y \geq \alpha] = \dots$
Step 3
Sensitivity analysis
$X = (X_1, \not{X_2}, X_3, \not{X_4})$
Step 4
Inverse problem $X|Y = y_{\mathrm{obs}}$
Reverse propagation $X \leftarrow Y$

Inverse problem

$$Y = u(X_1, \dots, X_d) + \varepsilon$$

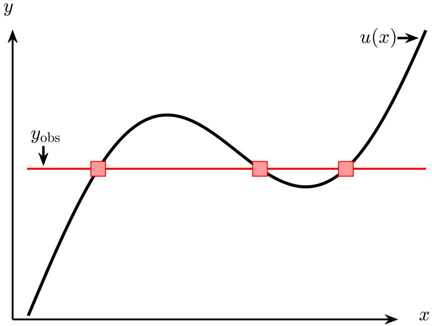
- $X \in \mathbb{R}^d$: input parameter
- $u : \mathbb{R}^d \to \mathbb{R}^m$: expensive computer model
- $Y \in \mathbb{R}^m$: observable output, corrupted by noise $\varepsilon \sim \mathcal{N}(0, \Sigma_{\text{obs}})$

Question: given an observation y_{obs} of Y, how to identify the parameter X which could have produced this observation?

The Bayesian perspective: from prior to posterior update

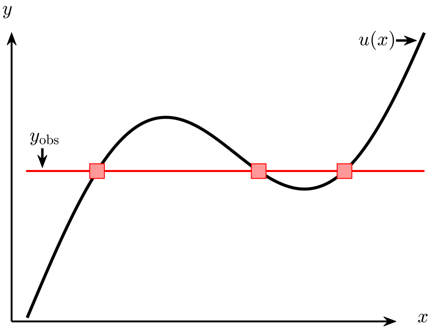
$\pi_X(x)$ $\longrightarrow$  $\longrightarrow$ $\pi_{X|Y}(x|y_{\text{obs}})$

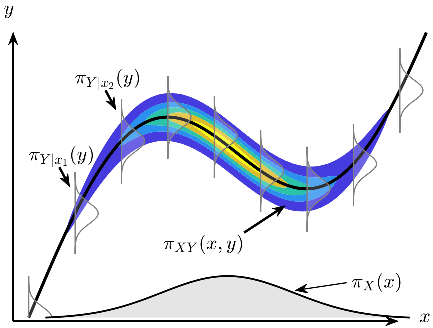
Variational **V.S.** Bayesian



$$\min_x \underbrace{\frac{1}{2}\|y_{\text{obs}} - u(x)\|^2_{\Sigma_{\text{obs}}^{-1}}}_{\text{data mismatch}} + \underbrace{\lambda \mathcal{R}(x)}_{\text{regularization}}$$

Variational **V.S.** Bayesian

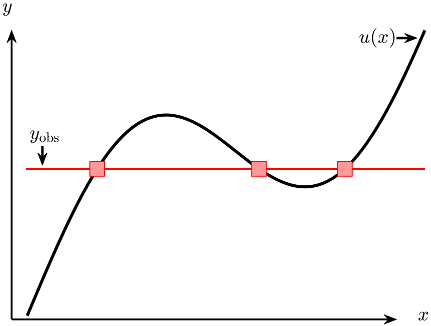




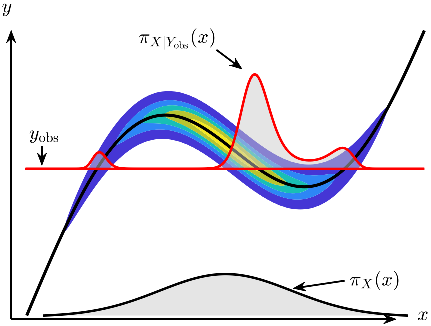
$$\min_{x} \underbrace{\frac{1}{2}\|y_{\text{obs}} - u(x)\|^2_{\Sigma_{\text{obs}}^{-1}}}_{\text{data mismatch}} + \underbrace{\lambda \mathcal{R}(x)}_{\text{regularization}}$$

$$\underbrace{\pi_{XY}(x,y)}_{\text{joint}} = \underbrace{\pi_{Y|X}(y|x)}_{\text{likelihood}} \underbrace{\pi_X(x)}_{\text{prior}}$$

Variational **V.S.** Bayesian



$$\min_{x} \quad \underbrace{\frac{1}{2}\|y_{\text{obs}} - u(x)\|^2_{\Sigma_{\text{obs}}^{-1}}}_{\text{data mismatch}} + \underbrace{\lambda \mathcal{R}(x)}_{\text{regularization}}$$

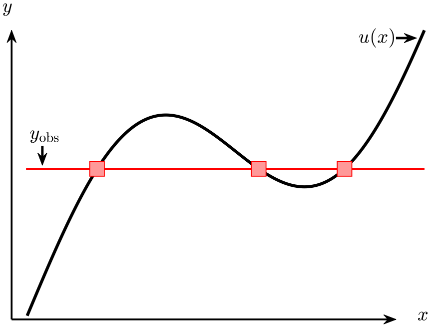


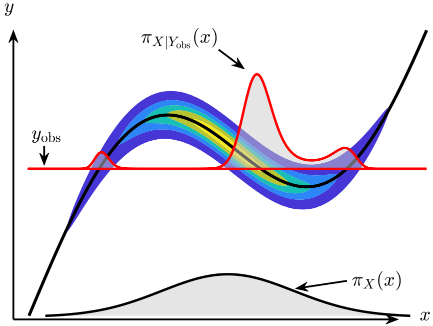
$$\underbrace{\pi_{XY}(x, y)}_{\text{joint}} = \underbrace{\pi_{Y|X}(y|x)}_{\text{likelihood}} \underbrace{\pi_X(x)}_{\text{prior}}$$

Given y_{obs}, the **posterior** is given by

$$\pi_{X|y_{\text{obs}}}(x) = \frac{\pi_{Y|X}(y_{\text{obs}}|x)\pi_X(x)}{\pi_Y(y_{\text{obs}})}$$

Variational **V.S.** Bayesian





$$\min_x \underbrace{\tfrac{1}{2}\|y_{\text{obs}} - u(x)\|^2_{\Sigma_{\text{obs}}^{-1}}}_{\text{data mismatch}} + \underbrace{\lambda \mathcal{R}(x)}_{\text{regularization}}$$

$$\underbrace{\pi_{XY}(x,y)}_{\text{joint}} = \underbrace{\pi_{Y|X}(y|x)}_{\text{likelihood}} \underbrace{\pi_X(x)}_{\text{prior}}$$

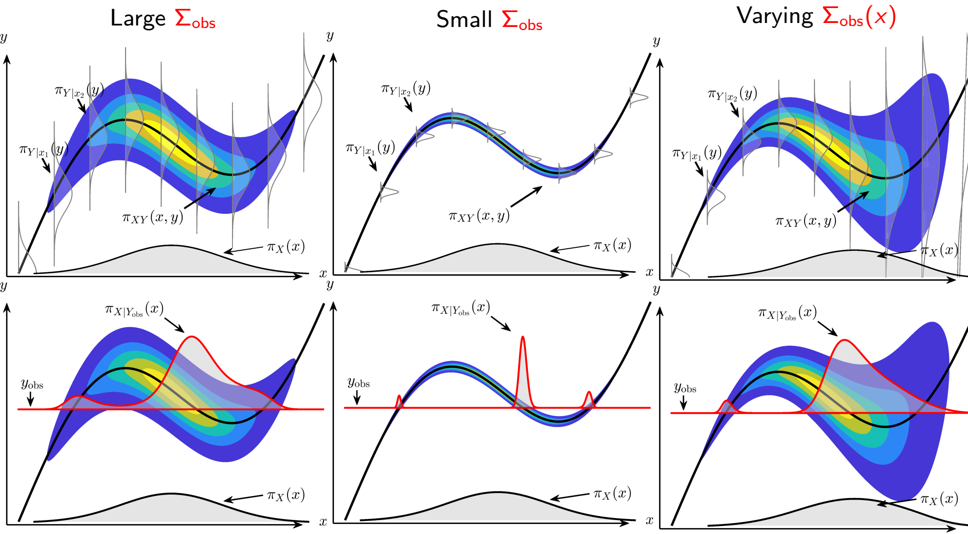
Given y_{obs}, the **posterior** is given by

$$\pi_{X|y_{\text{obs}}}(x) = \frac{\pi_{Y|X}(y_{\text{obs}}|x)\pi_X(x)}{\pi_Y(y_{\text{obs}})}$$

Conceptually different, but not that far:

$$\pi_{X|y_{\text{obs}}}(x) \propto \exp\left(-\frac{1}{2}\|y_{\text{obs}} - u(x)\|^2_{\Sigma_{\text{obs}}^{-1}} - \lambda \mathcal{R}(x)\right)$$

Importance of the **model error** $\varepsilon = \mathcal{N}(0, \Sigma_{\text{obs}})$



The case of **linear Gaussian** problems $u(x) = Ax$

Gaussian prior

$$\pi_X(x) \propto \exp\left(-\frac{1}{2}\|x - m\|^2_{\Sigma^{-1}}\right)$$

and **linear Gaussian** likelihood

$$\pi_{Y|X}(y|x) \propto \exp\left(-\frac{1}{2}\|y - Ax\|^2_{\Sigma_{\text{obs}}^{-1}}\right)$$

yield **Gaussian** posterior

$$\pi_{X|Y}(x|y_{\text{obs}}) \propto \exp\left(-\frac{1}{2}\|x - m_{\text{pos}}(y_{\text{obs}})\|^2_{\Sigma_{\text{pos}}^{-1}}\right)$$

where

$$\Sigma_{\text{pos}}^{-1} = \Sigma^{-1} + A^\top \Sigma_{\text{obs}}^{-1} A$$

$$m_{\text{pos}}(y_{\text{obs}}) = \Sigma_{\text{pos}}\Sigma^{-1} m + \underbrace{\Sigma_{\text{pos}} H^\top \Sigma_{\text{obs}}^{-1}}_{\text{Kalman Gain}} y_{\text{obs}}$$

Ensemble Kalman Filters (EnKF) methods for time-dependent data assimilation problems with $d \gg 1$: replace the above covariances with sampled covariances. Works well for nonlinear/nonGaussian filtering problems... and we don't really know why. Square Root EnKF, Ensemble Transform Kalman filter (ETKF), Extended Kalman Filter,...

The **Laplace approximation** of $\pi_{X|y_{\text{obs}}}$

Taylor expansions around the Maximum A Posteriori (MAP) point

$$x^{\text{MAP}} \in \arg\max_{x \in \mathbb{R}^d} \pi_{X|y_{\text{obs}}}(x)$$

Compute the Hessian at the MAP and

$$\widetilde{\Sigma}_{\text{pos}}^{-1} = -\nabla^2 \log \pi_{X|y_{\text{obs}}}(x^{\text{MAP}})$$

and then

$$\boxed{\pi_{X|y_{\text{obs}}} \approx \mathcal{N}(x^{\text{MAP}}, \widetilde{\Sigma}_{\text{pos}})}$$

Alternatively, if **Gaussian prior** + **Gaussian likelihood**

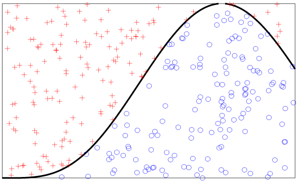
$$\pi_{X|y_{\text{obs}}}(x) \propto \exp\left(-\frac{1}{2}\|y_{\text{obs}} - u(x)\|^2_{\Sigma_{\text{obs}}^{-1}} - \frac{1}{2}\|x - m\|^2_{\Sigma^{-1}}\right)$$

then **linearize the model** $u(x) \approx u(x^{\text{MAP}}) + \nabla u(x^{\text{MAP}})(x - x^{\text{MAP}})$ and

$$\widetilde{\Sigma}_{\text{pos}}^{-1} = \Sigma^{-1} + \nabla u(x^{\text{MAP}})^\top \Sigma_{\text{obs}}^{-1} \nabla u(x^{\text{MAP}})$$

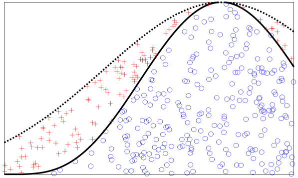
How to sample from a **nonGaussian** density $\pi_{X|y_{\text{obs}}}$?

Rejection method: draw points uniformly on $\text{supp}(\pi_{X|y_{\text{obs}}}) \times [0, \max(\pi_{X|y_{\text{obs}}})]$ and reject the points which have landed above the graph of $\pi_{X|y_{\text{obs}}}$



1. Draw $X \sim \mathcal{U}(\text{supp}(\pi_{X|y_{\text{obs}}}))$
2. Draw $Z \sim \mathcal{U}([0, \max(\pi_{X|y_{\text{obs}}})])$
3. If $Z \leq \pi_{X|y_{\text{obs}}}(X)$: accept
4. Otherwise, reject and $\rightarrow$ 1.

If the **acceptance rate** $\dfrac{\text{blue}}{\text{red} + \text{blue}}$ is too small, use a given $\rho_X \approx \pi_{X|y_{\text{obs}}}$ and



1. Draw $X \sim \rho_X$
2. Draw $Z \sim \mathcal{U}([0, \max \frac{\pi_{X|y_{\text{obs}}}}{\rho_X}])$
3. If $Z \leq \frac{\pi_{X|y_{\text{obs}}}(X)}{\rho_X(X)}$: accept
4. Otherwise, reject and $\rightarrow$ 1.

In practice $\max(\pi_{X|y_{\text{obs}}})$ and/or $\max(\pi_{X|y_{\text{obs}}}/\rho_X)$ **might not be accessible!!**

Markov Chain Monte Carlo (MCMC)

Idea: use an **iterative rejection scheme** to define a Markov chain

$$\pi_X \rightarrow X^{(0)} \rightarrow X^{(1)} \rightarrow X^{(2)} \rightarrow \dots \rightarrow X^{(\infty)} \sim \pi_{X|y_{\text{obs}}}$$

For a given a proposal density $\rho_X(\cdot|\cdot)$, compute $X^{(t)} \rightarrow X^{(t+1)}$ as follow:

1. Draw a **candidate** $X^\dagger \sim \rho_X(\cdot|X^t)$ compute the **acceptance probability**

$$\alpha(X^\dagger|X^t) = \min\left\{1; \frac{\pi_{X|y_{\text{obs}}}(X^\dagger)\rho_X(X^t|X^\dagger)}{\pi_{X|y_{\text{obs}}}(X^t)\rho_X(X^\dagger|X^t)}\right\}$$

2. Draw $Z \sim \mathcal{U}([0,1])$
3. If $Z \leq \alpha(X^\dagger|X^t)$, accept $X^{t+1} = X^\dagger$
4. Otherwise, reject $X^\dagger$ and $X^{t+1} = X^t$

This accept/reject step is called the **Metropolis-Hastings** correction: $\pi_{X|y_{\text{obs}}}$ can be known only **up to a multiplicative constant**! Under mild assumptions on $\rho_X(\cdot|\cdot)$, we have **convergence** $X^{(\infty)} \sim \pi_{X|y_{\text{obs}}}$. However, designing a proposal $\rho_X(\cdot|\cdot)$ which yields **fast convergence** is an art...

[Hastings: Monte Carlo sampling methods using Markov chains and their applications, 1970]

Some **popular proposals** for MCMC

- **Random walk (RW)** proposal

$$X^{\dagger} = X^{t} + \delta\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I_d)$$

- **Preconditioned Crank-Nicolson (pCN)** proposal

$$X^{\dagger} = \sqrt{1-\beta^2}X^{t} + \beta^2\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, P)$$

- **Metropolis-adjusted Langevin algorithm (MALA)**: based on the discretization of the Langevin SDE

$$X^{\dagger} = X^{t} + \nabla \log \pi_{X|y_{\text{obs}}}(X^{t})\Delta t + \sqrt{2\Delta t}\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I_d)$$

- **Hamiltonian Monte Carlo (HMC)** and its variant **No U-Turn Sampler (NUTS)**:

$$X^{\dagger} = x(L\delta t)$$

  where $x(t)$ solves the Hamilton's equations (gradient based) with initial position $x(0) = X^{t}$ and **random initial momentum**.

- ...

Importance sampling correction

$$\mathrm{I} = \int F(x)\pi_{X|y_{\text{obs}}}(x)\mathrm{d}x = \int F(x)\frac{\pi_{X|y_{\text{obs}}}(x)}{\pi_X(x)}\pi_X(x)\mathrm{d}x$$

Draw $X^{(1)}, \ldots, X^{(N)} \sim \pi_X(x)$ and

$$\mathrm{I}_N^{\text{IS}} = \frac{1}{N}\sum_{i=1}^{N} F(X^{(i)})\omega_i, \qquad \omega_i = \frac{\pi_{X|y_{\text{obs}}}(X^{(i)})}{\pi_X(X^{(i)})}$$

If $\pi_{X|y_{\text{obs}}}$ known up to a constant, then use **self-normalized** weights

$$\mathrm{I}_N^{\text{IS}} = \sum_{i=1}^{N} F(X^{(i)})\frac{\omega_i}{\omega_1 + \ldots + \omega_N}, \qquad \omega_i \propto \frac{\pi_{X|y_{\text{obs}}}(X^{(i)})}{\pi_X(X^{(i)})}$$

Weight degeneracy when π_X is too far from $\pi_{X|y_{\text{obs}}}$:

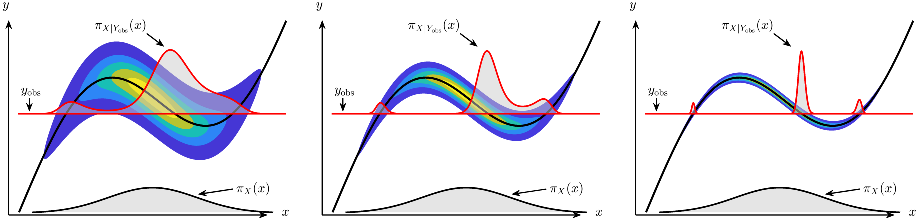
$$\text{Effective Sample Size } = \sum_{i=1}^{N}\left(\frac{\omega_i}{\omega_1 + \ldots + \omega_N}\right)^2 \longrightarrow 1$$

Importance sampling + **Sequential Monte Carlo (SMC)**

Consider a sequence of **bridging densities** with "increasing complexity"

$$\pi_X =: \rho_{X,0} \to \rho_{X,1} \to \ldots \to \rho_{X,L} := \pi_{X|y_{\text{obs}}}$$

For instance, modify the data noise Σ_{obs} ($\approx$ **tempering or annealing**)



Idea: use Important Sampling across two consecutive bridging densities.

Importance sampling + **Sequential Monte Carlo (SMC)**

Draw N **particles** from $X_0^{(1)}, \dots, X_0^{(N)} \sim \rho_{X,0}$ and

$$\rho_{X,1}(x) \approx \sum_{i=1}^{N} \delta_{X_0^{(i)}}(x) \frac{\omega_i^0}{\omega_1^0 + \dots + \omega_N^0}, \quad \omega_i^0 \propto \frac{\rho_{X,1}(X_0^{(i)})}{\rho_{X,0}(X_0^{(i)})}$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} \delta_{X_0^{(i)\prime}}(x)$$

where $X_0^{(1)\prime}, \dots, X_0^{(N)\prime}$ are **re-sampled** from $\{X_0^{(1)}, \dots, X_0^{(N)}\}$ with probability

$$\mathbb{P}(X_0' = X_0^{(i)}) = \frac{\omega_i^0}{\omega_1^0 + \dots + \omega_N^0}$$

Next, we draw $X_1^{(i)} \sim \rho_X(\cdot | X_0^{(i)\prime})$ according to some given proposal and

$$\rho_{X,2}(x) \approx \sum_{i=1}^{N} \delta_{X_1^{(i)}}(x) \frac{\omega_i^1}{\omega_1^1 + \dots + \omega_N^1}, \quad \omega_i^1 \propto (\text{some expression...})$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} \delta_{X_2^{(i)}}(x) \quad (\text{re-sample})$$

The rest $\ell \to \ell + 1$ follows.

[Del Moral, Doucet and Jasra: Sequential Monte Carlo Samplers 2006]

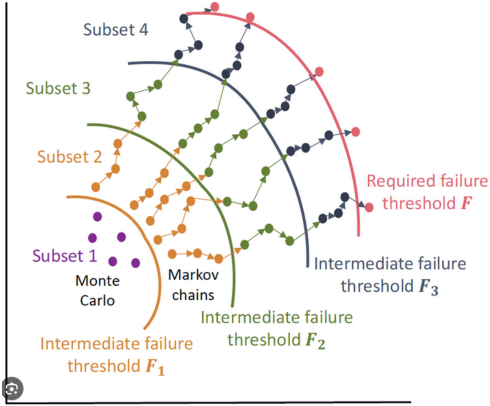
Sequential Monte Carlo (SMC) **for rare event**

$$\boxed{\pi_{X|\alpha}(x) \propto \mathbf{1}_{[\alpha,\infty)}(u(x))\pi_X(x)}$$

Sequential Monte Carlo (SMC) **for rare event**

$$\boxed{\pi_{X|\alpha}(x) \propto \mathbf{1}_{[\alpha,\infty)}(u(x))\pi_X(x)}$$

Increase the threshold α:



Conclusion: there are lots of things in UQ...

Questions?