Czech Academy of Sciences

ÚGN

VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE | DEPARTMENT OF APPLIED MATHEMATICS

Gradient-Informed Proposals for Surrogate-Assisted Bayesian Sampling

Simona Bérešová[1,2], Michal Béreš[1,2], Tomáš Luber[1], Stanislav Sysala[1]

May 18, 2026
HPCSE 2026

[1]Czech Academy of Sciences, Institute of Geonics
[2]VŠB – Technical University of Ostrava, FEECS, Department of Applied Mathematics

Co-funded by the European Union

INODIN

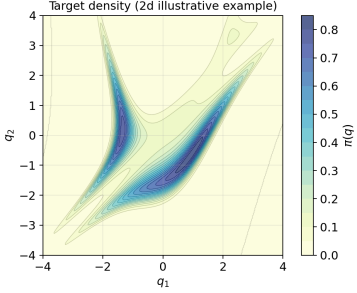
Co-funded by the European Union under the project INODIN (CZ.02.01.01/00/23_020/0008487)

Outline

1. Inverse problems with uncertainties
2. Markov Chain Monte Carlo sampling
3. Hamiltonian proposal
4. Neural network surrogate models
5. 45 dimensional geotechnical benchmark

- motivation: geological inverse problems with uncertain material properties
- we observe noisy in-situ measurements $y \in \mathbb{R}^m$
- the goal is to find unknown parameters $q \in \mathbb{R}^d$ of a forward model $G : \mathbb{R}^d \to \mathbb{R}^m$ that lead to the in-situ measurements (up to noise)
- Bayesian formulation combines **information from data** and **prior information** $\to$ the result is a posterior distribution (not a single estimate)

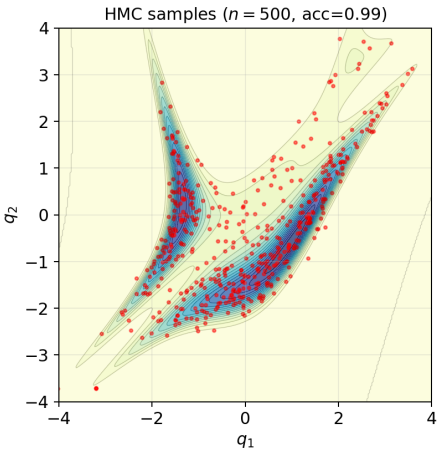
$$\pi(q) := \pi(q|y) \propto \pi_{\text{noise}}(y - G(q))\, \pi_{\text{prior}}(q)$$



Goal: generate samples from the posterior distribution

- not possible to generate independent samples, MCMC methods construct Markov chains of correlated samples
- the number of unknown parameters may be high (e.g. 10 to 100)
- computationally expensive - usually requires thousands to millions of evaluations of the forward model G

→ reduce correlation between samples → reduce number of required G evaluations



Metropolis-Hastings principle

1. start from the current state q_k
2. draw a proposal $q^* \sim P(\cdot|q_k)$
3. compute the acceptance probability

$$\alpha = \min\left(1, \frac{\pi(q^*)\, P(q_k|q^*)}{\pi(q_k)\, P(q^*|q_k)}\right)$$

4. accept the proposal with probability α; otherwise remain at q_k

Primary goal: **reduce correlation between samples**

1. the proposal q^* should be nearly independent of q_k
2. the acceptance probability α should be high

Secondary goal:

- keep the proposal cost low

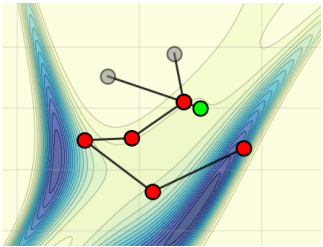
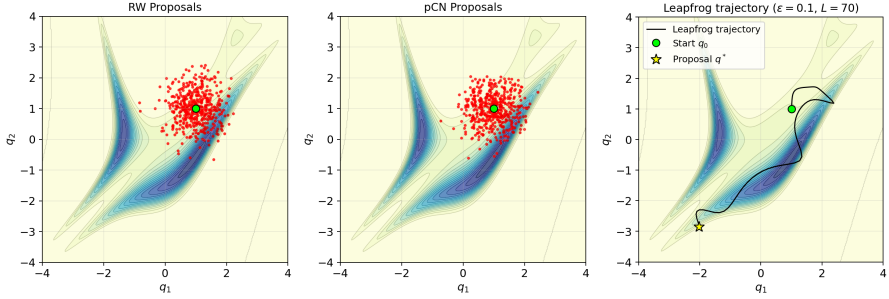


Illustration of proposal strategies



random walk: simple, but it leads to high correlation

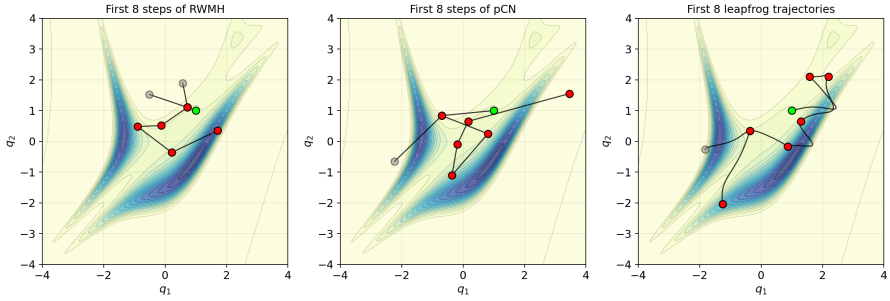
$$q^* = q_n + \eta; \;\; \eta \sim \mathcal{N}\left(0, \sigma^2 I\right)$$

pCN: preserves the Gaussian prior; remains stable as the dimension grows

$$q^* = \sqrt{1-\beta^2} q_n + \beta\eta; \;\; \eta \sim \mathcal{N}\left(0, C_{\text{prior}}\right)$$

Hamiltonian: longer moves (deterministic trajectory with randomly chosen momentum); respects the geometry

Illustration of proposal strategies (several steps)



random walk: simple, but it leads to high correlation

$$q^* = q_n + \eta; \;\; \eta \sim \mathcal{N}\left(0, \sigma^2 I\right)$$

pCN: preserves the Gaussian prior; remains stable as the dimension grows

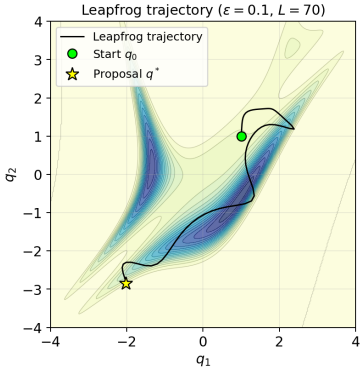
$$q^* = \sqrt{1-\beta^2} q_n + \beta\eta; \;\; \eta \sim \mathcal{N}\left(0, C_{\text{prior}}\right)$$

Hamiltonian: longer moves (deterministic trajectory with randomly chosen momentum); respects the geometry

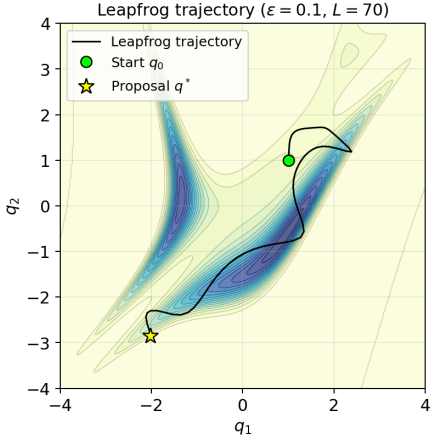
$U(q) = -\log \pi(q) =$ potential energy function ... *"hilly landscape"*

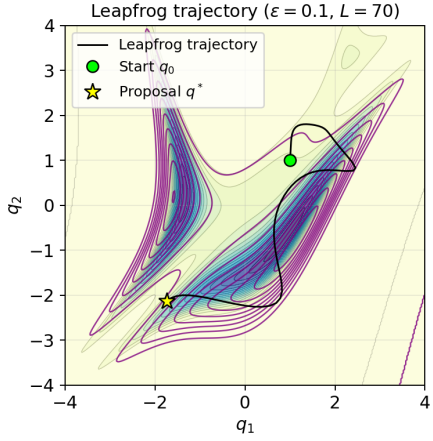
Give initial momentum to a "marble" $\rightarrow$ it moves up and down, converting kinetic energy into potential energy and back.

- the trajectory is computed numerically using the leapfrog method
  - requires $\nabla U(q)$
  - α is close to 1 $\rightarrow$ almost every proposal is accepted
- MH algorithm remains valid even if we use $\nabla \widetilde{U}(q)$ instead of $\nabla U(q)$
  - with inaccurate derivatives, α gets lower



$$\widetilde{\pi}(q) \propto \pi_{\text{noise}}\left(y - \widetilde{G}(q)\right) \pi_{\text{prior}}(q)$$





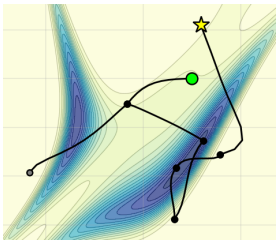
Delayed-acceptance (DA) scheme with subchains

1. start from the current state q_k
2. **draw a proposal q^* using an MH subchain targeting $\widetilde{\pi}$**
3. compute the delayed-acceptance correction α_{DA} so that the correct target distribution is preserved
4. accept the proposal with probability α_{DA}; otherwise remain at q_k

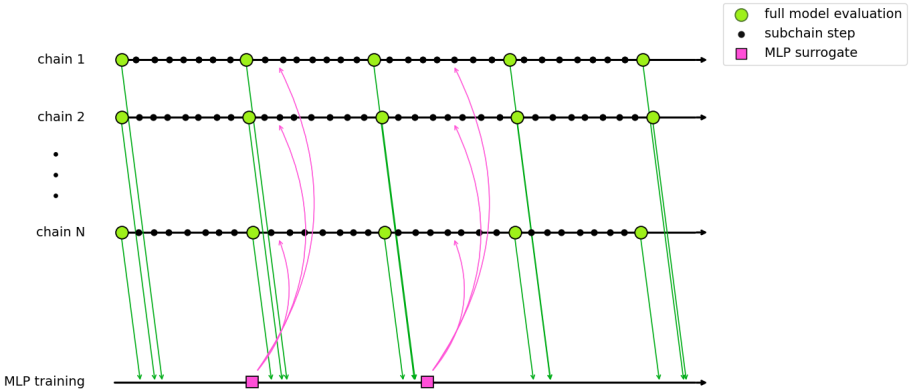
The subchain can use any proposal, e.g. Hamiltonian with surrogate derivatives.

- accurate $\widetilde{G}(q) \rightarrow$ high acceptance α_{DA}
- accurate $\nabla \widetilde{U}(q) \rightarrow$ high acceptance inside subchain

$\Rightarrow$ Surrogate model quality is crucial for sampling efficiency.

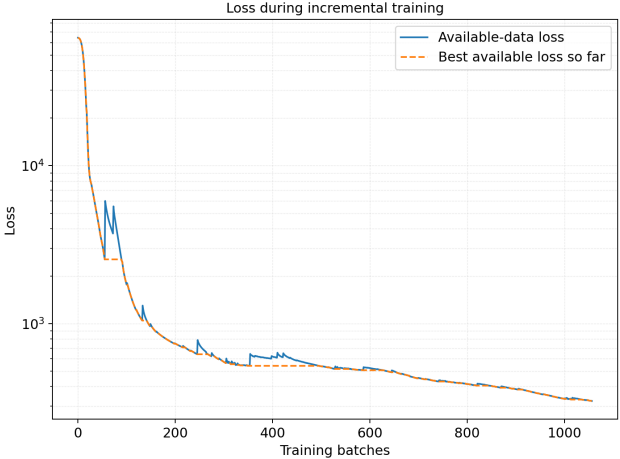


Surrogate model construction during sampling process

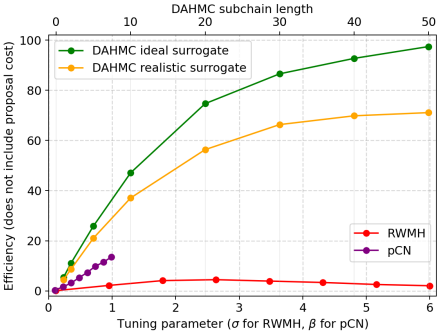


- several Markov chains are generated in parallel
- one additional process collects exact evaluations $(q_k, G(q_k))$ and constructs a shared neural network surrogate model (MLP)

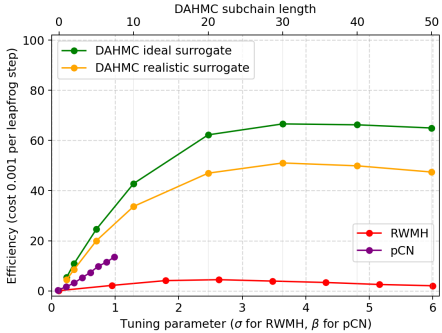
+ derivatives are available through automatic differentiation
− many hyperparameters must be tuned: architecture, normalization, optimizer, activation, learning rates, training batches, randomization, weight decay, ...



- **autocorrelation time** τ ... *how many samples we have to generate to obtain one almost uncorrelated "effective sample"*
- **cost per effective sample** $= \tau \cdot$ (cost per sample)
- **sampling efficiency** $= \frac{100}{\tau}$ ... *in 100 samples, how many are "effective"*



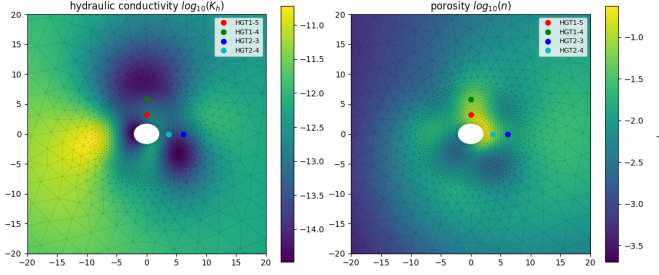
(a) cost ratio $\frac{\text{leapfrog step}}{G \text{ evaluation}} = 0$

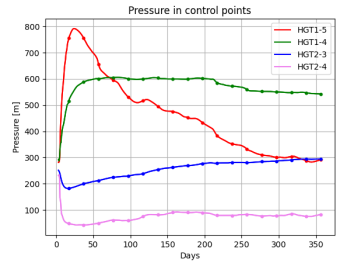


(b) cost ratio $\frac{\text{leapfrog step}}{G \text{ evaluation}} = 0.001$

hydraulic conductivity field represented by 20 KL coefficients
porosity field represented by 20 KL coefficients
drained bulk modulus 2 parameters - EDZ, outer
shear modulus 2 parameters - EDZ, outer
grain bulk modulus 1 parameter

$\to$ poroelasticity model G $\to$ outputs matching in-situ measurements (up to noise)





Algorithm	Subchain length	Proposal	Accept. rate	τ	Total* samples	of that effective
MH	–	RW ($\sigma = 0.1$)	1%	> 1000	34689	< 35
MH	–	RW ($\sigma = 0.05$)	8%	> 1000	34450	< 35
MH	–	pCN ($\beta = 0.05$)	9%	> 1000	34466	< 35
DAMH	100	pCN ($\beta = 0.05$)	65%	121	34850	288
DAMH	100	Hamilton ($L = 10$)	25%	**16.7**	27145	**1625**
DAMH	50	Hamilton ($L = 10$)	34%	**18.1**	32910	**1818**

*for the same computation time, all chains together

- 15 sampling processes + 1 collecting data, MLP trained on a GPU

- Despite using surrogate models and approximate derivatives, the DAMHC algorithm converges to the correct target distribution.
- Hamiltonian proposals can successfully reduce the correlation between samples.
- Sampling efficiency strongly depends on the accuracy of the surrogate model.

Thank you for your attention!

Co-funded by the European Union
MINISTRY OF EDUCATION, YOUTH AND SPORTS
INODIN

Co-funded by the European Union under the project INODIN (CZ.02.01.01/00/23_020/0008487)