



Faculty of
Computer Science

Data Science and
Business Analytics

Sampling for Bayesian Optimization

Presented by

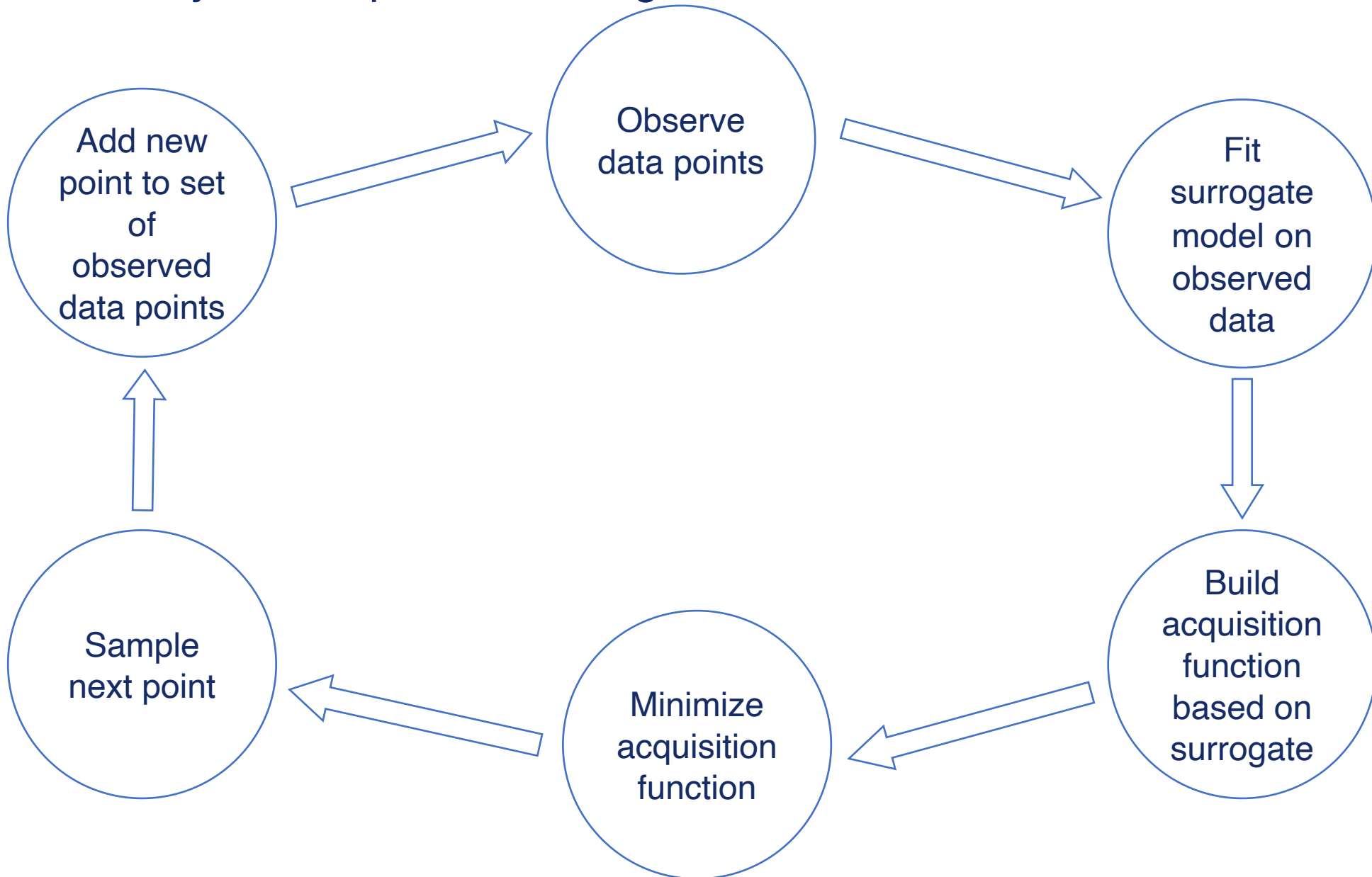
Nadezhda Vladimirovna Fomicheva,
3rd year student,
group 232, DSBA

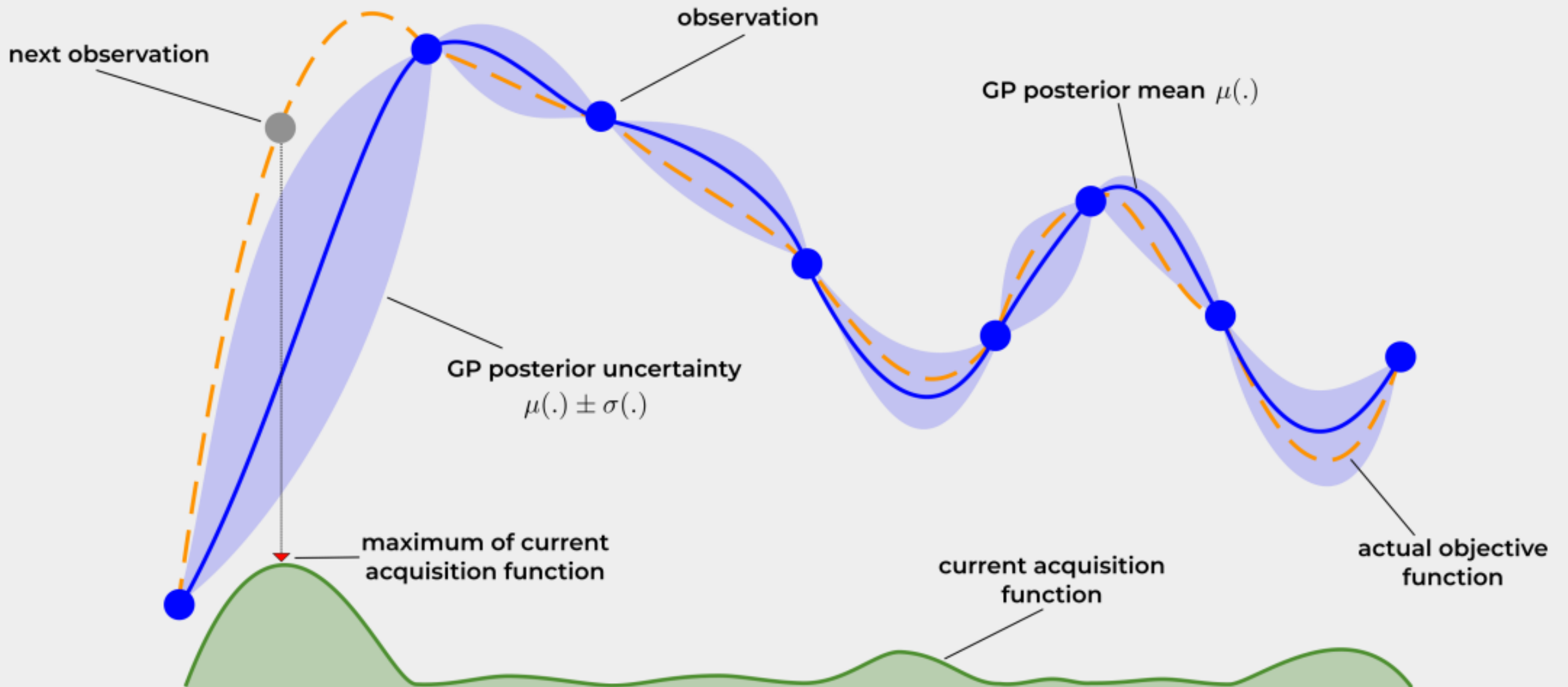
Moscow 2026

Project supervisor

Tigran Armenovich Ramazyan,
Junior Research Fellow,
Laboratory of Methods for Big Data Analysis,
HSE University







Gradient-based:

- Gradient Descent
- Adam

Evolutionary methods:

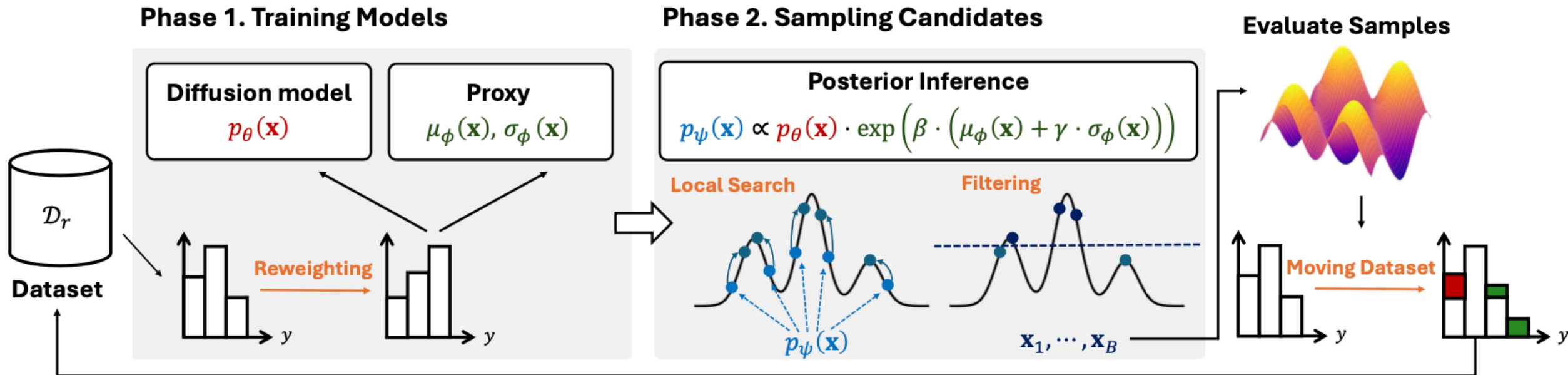
- Differential Evolution

Newton methods:

- L-BFGS-B
- BoTorch internal optimization

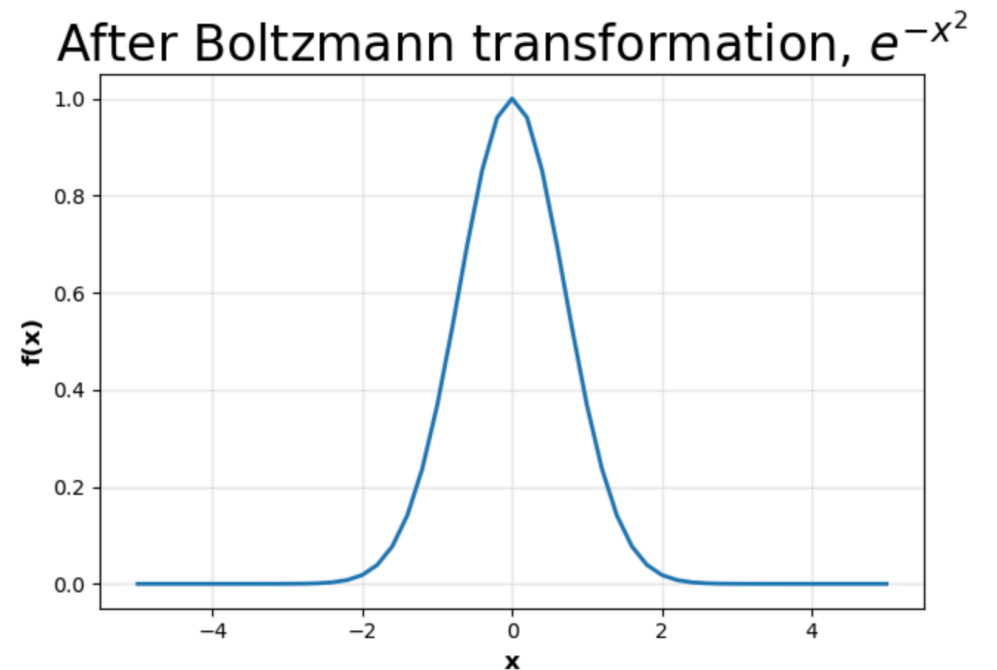
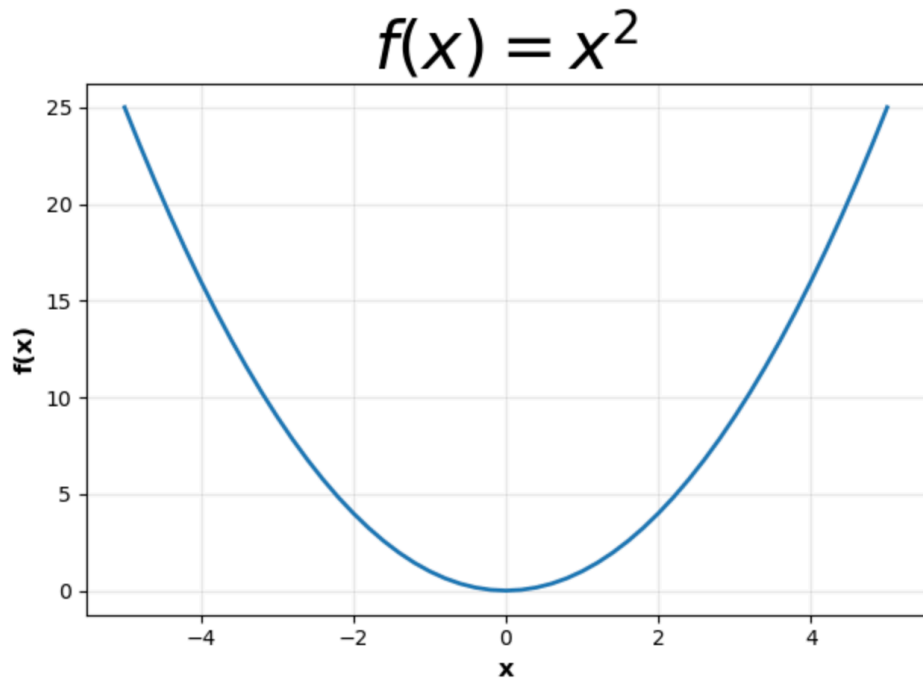
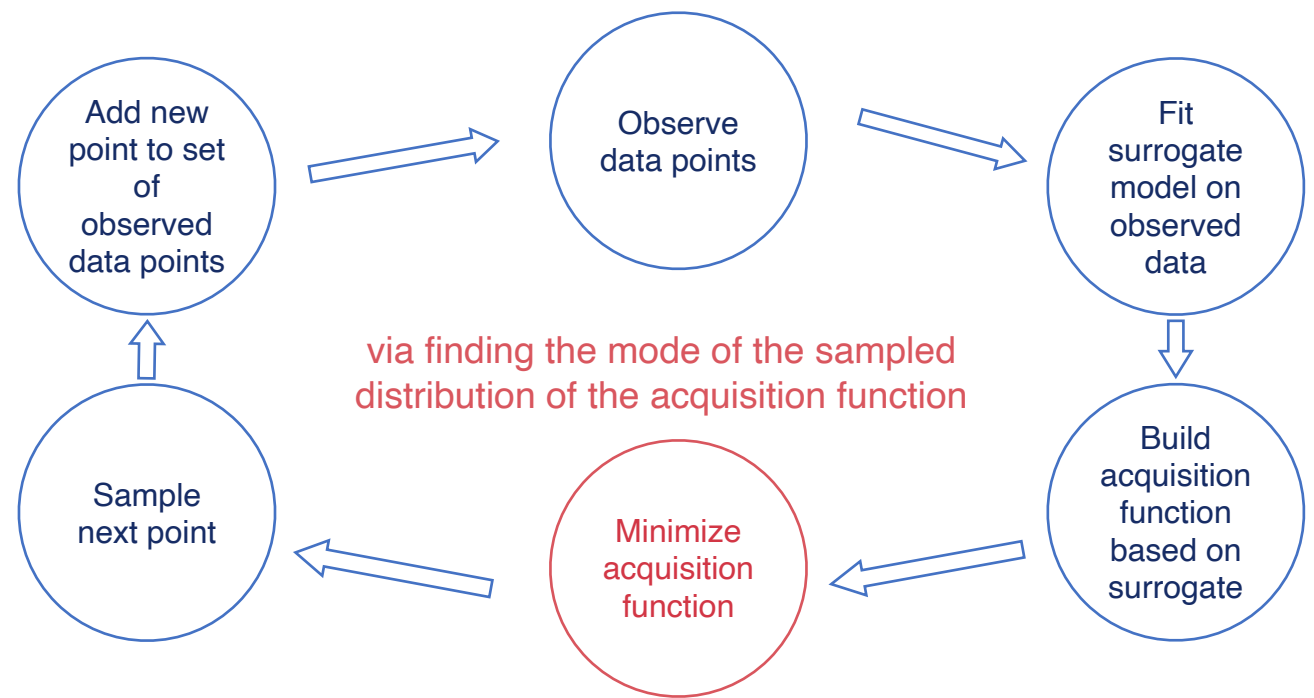
Using diffusion models:

- DiBo





The project aimed to create new approach **to make the process of minimizing the acquisition function more efficient** compared to the existing ones using MCMC algorithms.



Using **ULA**, one can achieve ε accuracy of finding the optimum of the acquisition function in $\mathcal{O}(d/\varepsilon)$, and in $\mathcal{O}(d^2 \ln(1/\varepsilon))$ using **MALA**, while other algorithms converge in $\Omega((1/\varepsilon)^d)$, where d is the number of dimensions.

These results can be obtained in the context, where the Lower Confidence Bound (LCB)

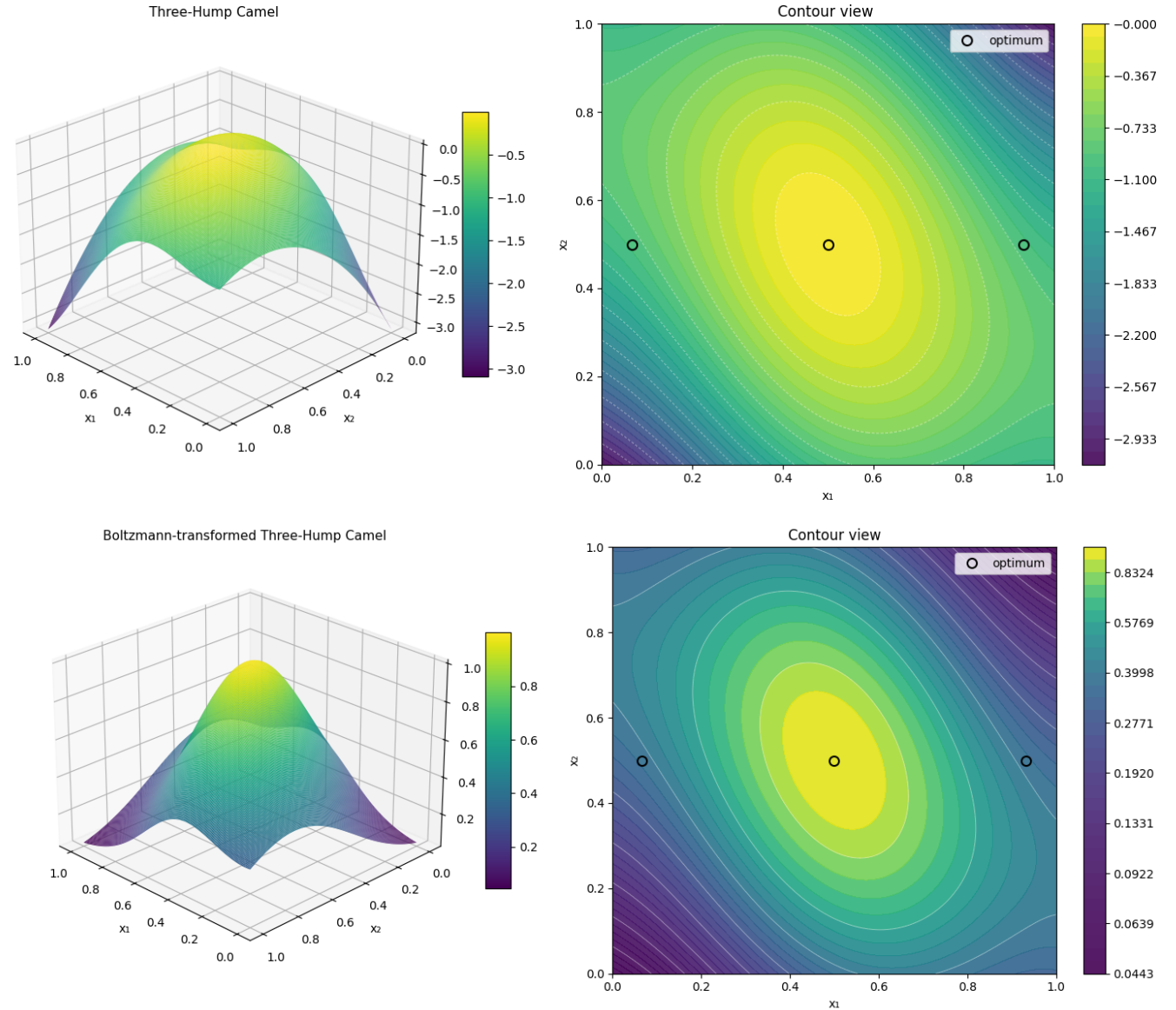
$$\alpha_t(x) = \mu_t(x) - \beta_t \sigma_t(x)$$

is used as the **acquisition function**,

and the Gaussian Process for the **surrogate model** is defined by a stationary RBF kernel:

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{\|x - x'\|^2}{2\ell^2}\right)$$

So, in the context of minimizing the acquisition functions, **Langevin dynamics appear to be the most prospective algorithms**.



Monte-Carlo Markov-Chain:

- ULA
- MALA
- ULMC
- HMC
- Gibbs

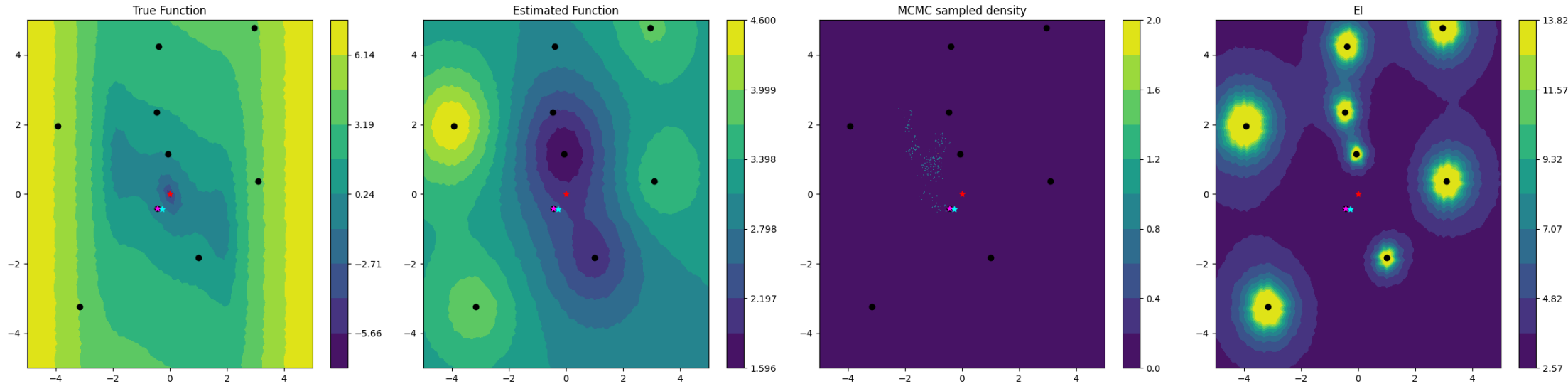
Tests

Synthetics functions:

- Three Hump Camel
- Six Hump Camel
- Branin
- Holder Table
- Hartmann
- Ackley
- Rastrigin
- Rosenbrock
- Styblinski-Tang

Real life problems:

- Tension Compression String
- Pressure Vessel
- Welded Beam SO
- Speed Reducer

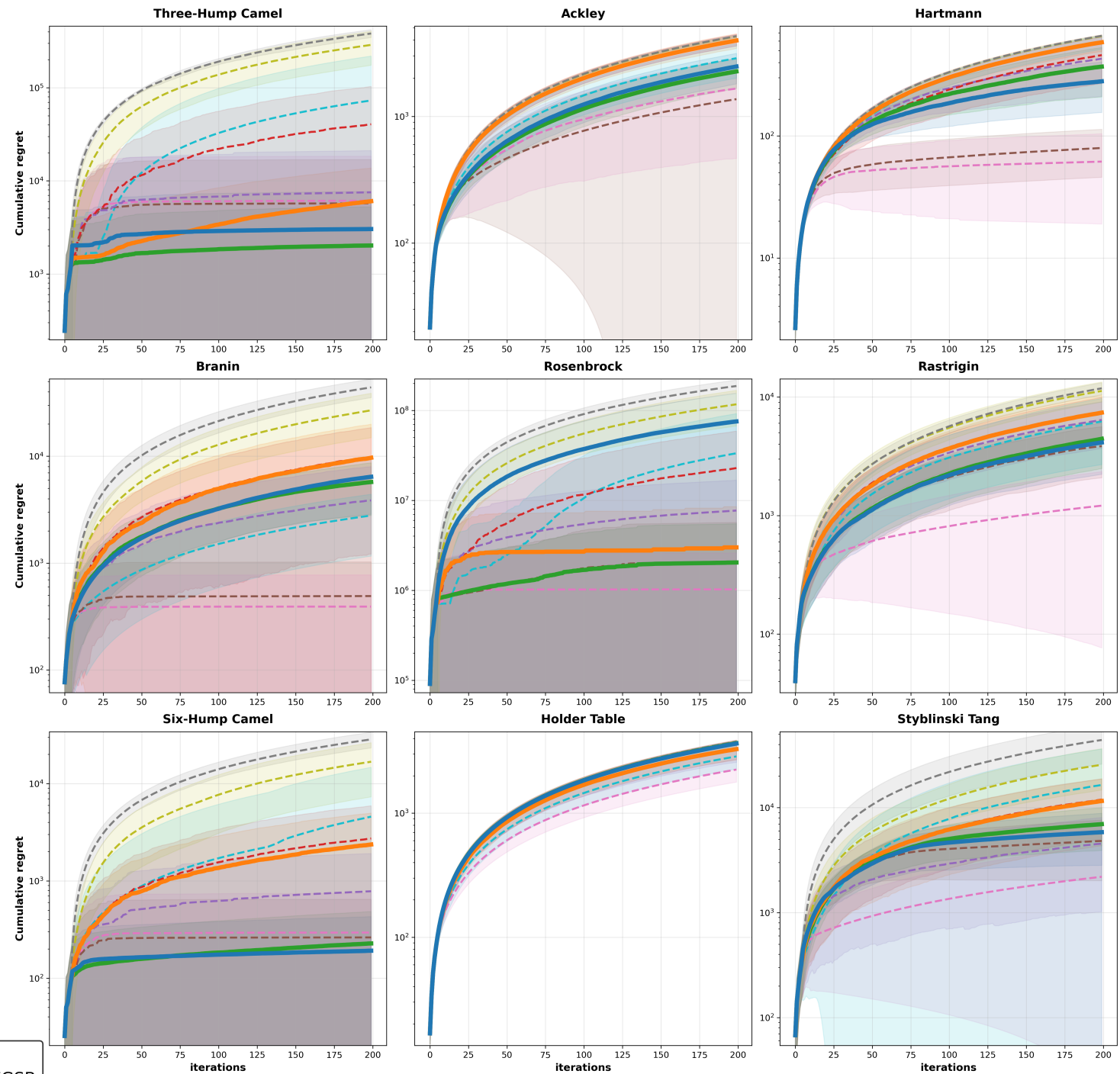


Results of the 5th iteration of the Bayesian Optimization with HMC sampling on the 2D Three-Hump Camel problem, 100 steps for burn-in phase and 200 for sampling, step size=1e-2



regret = estimated minimum – theoretical minimum

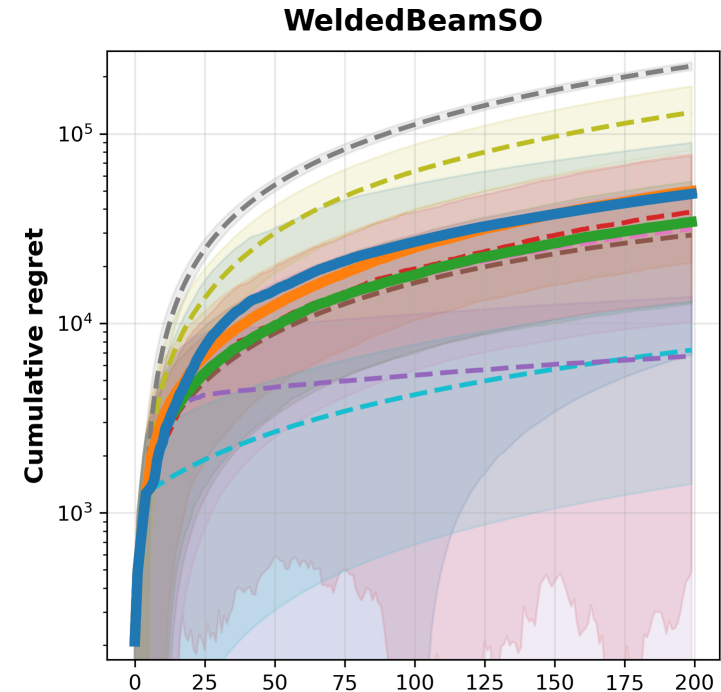
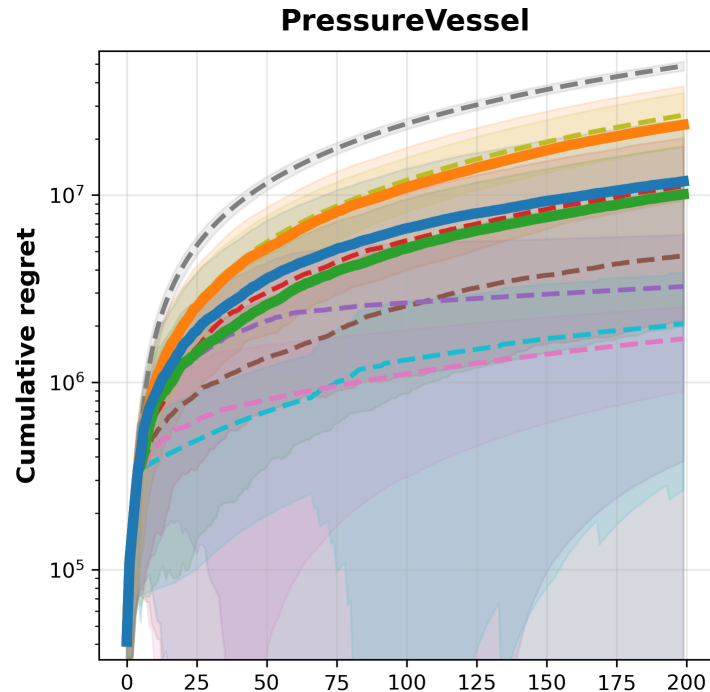
$$\text{cumulative regret} = \sum \text{regret}_i$$



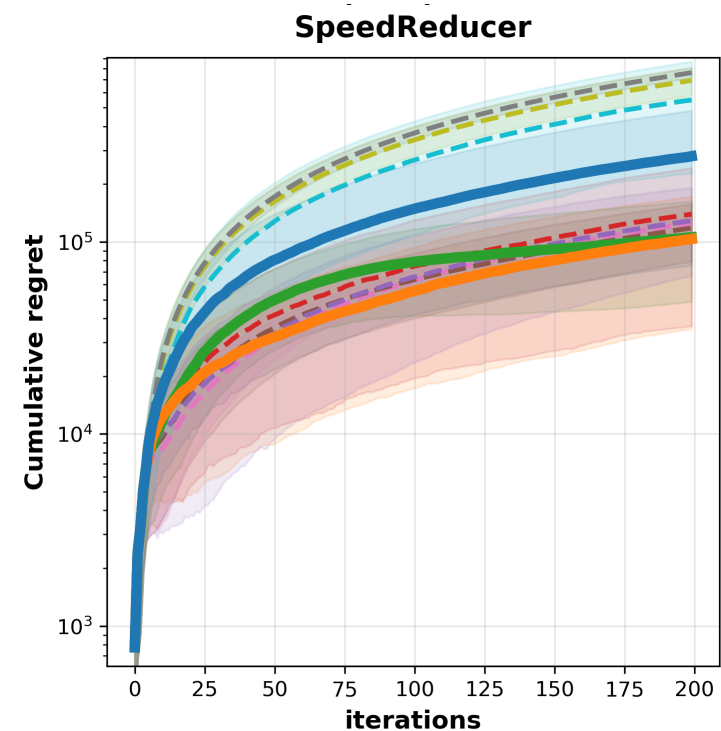
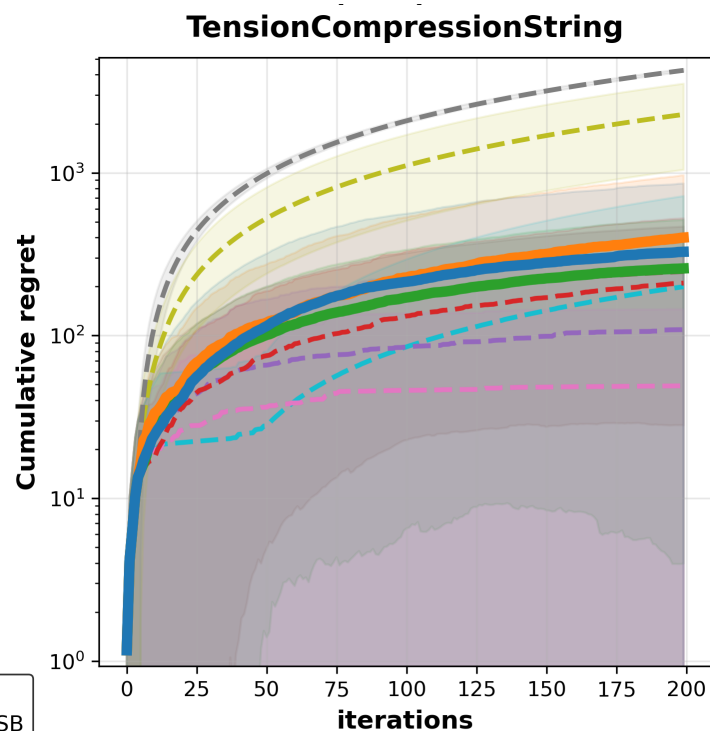
Mostly, Langevin dynamics algorithms do not outperform L-BFGS-B based ones, especially in low-dimensional cases.



And the performance preserves even
in the cases of real life problems.



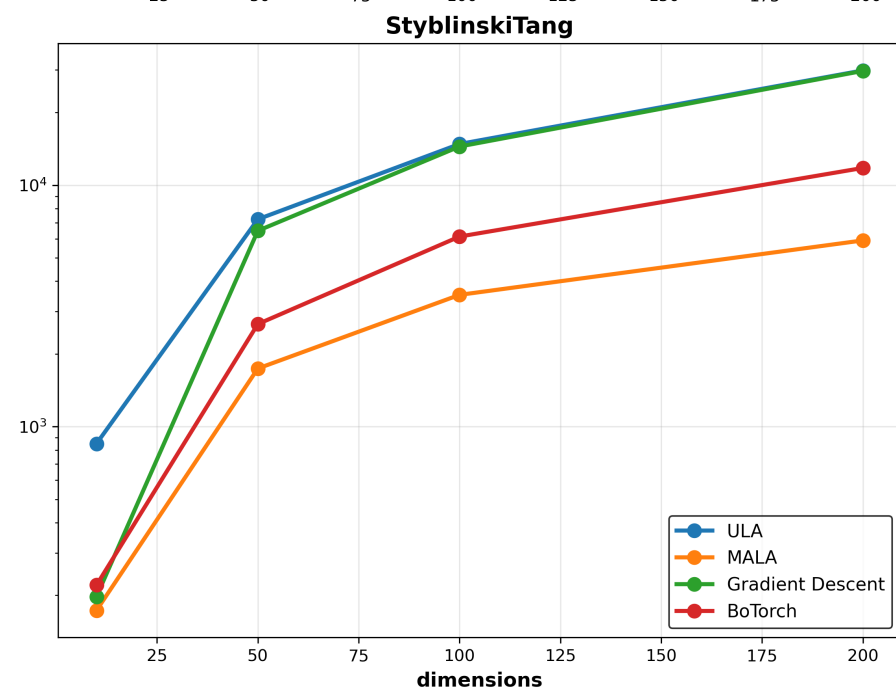
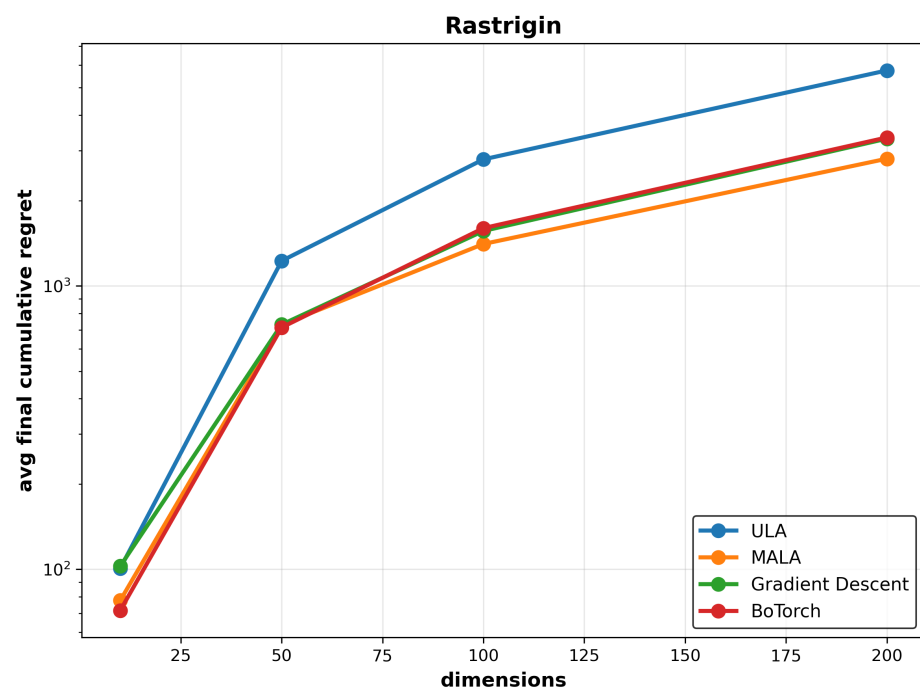
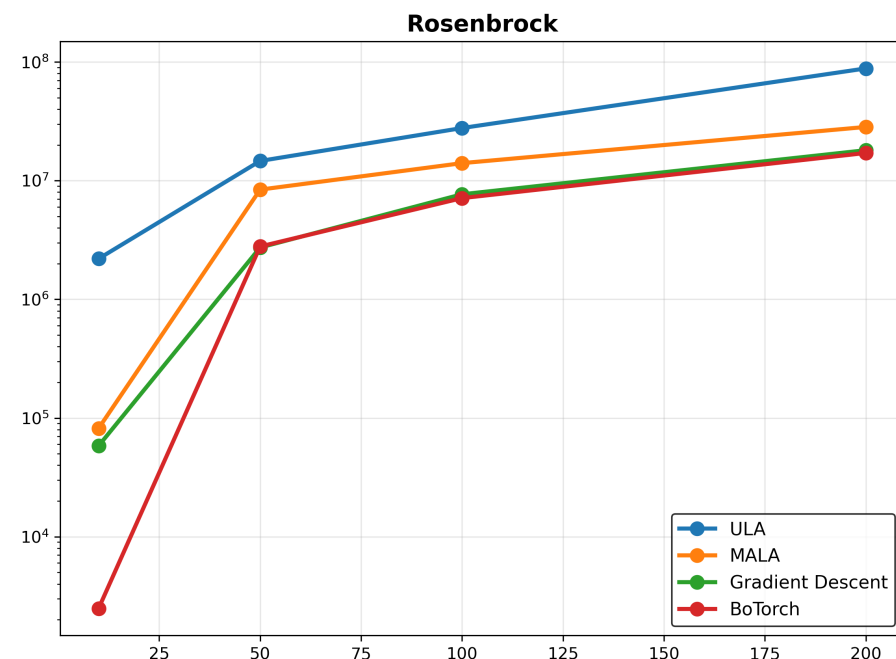
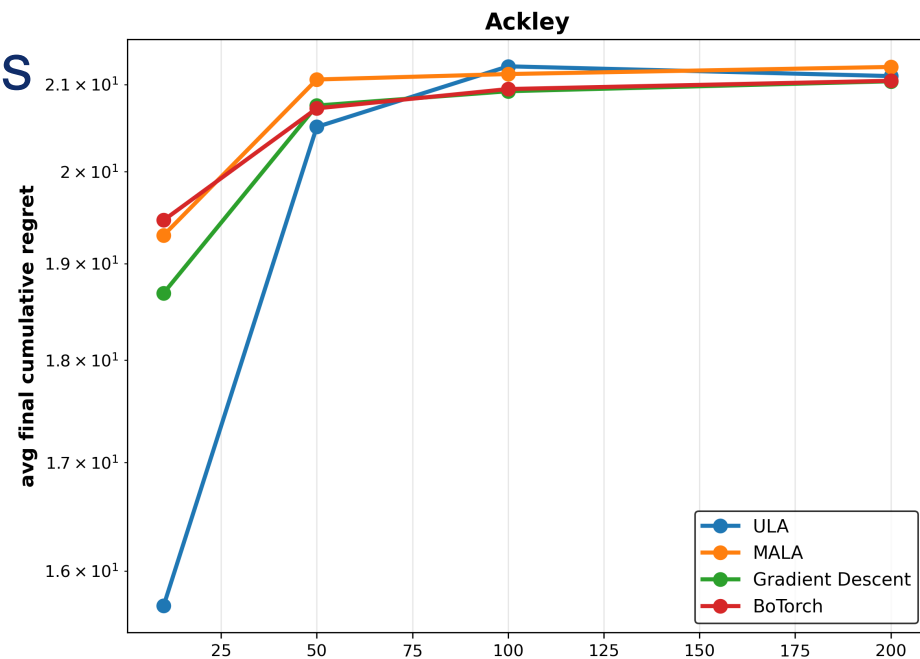
10





Results

Langevin dynamics outperform other algorithms on Rastrigin and Styblinski-Tang with the increase of dimensionalities, and show promising results on other tests.





- The **inner optimization** problem of **Bayesian optimization** is considered.
- We **replace** costly **optimization** of non-convex function **with MCMC**.
- **Langevin dynamics can be faster than optimization**, especially in high dimensions.





1. Implement the whole Bayesian Optimization algorithm

- Take Gaussian process (GP) specified by the RBF kernel and two types of mean modules as **the surrogate model**
- Put LCB as **the acquisition function**, with possibility to optimize it via various algorithms

2. Run the experiments using different MCMC algorithms with the posterior quadratic mean used for the GP

3. Run the experiments using benchmark algorithms using standard mean module for the GP

All algorithms were starting with the step size of $1e-3$ and were tested for 10 seeds for conclusive results.

For MCMC algorithms:

- 100 iterations were used for burn-in phases
- 200 iterations for sampling

For benchmark algorithms were used their default settings.



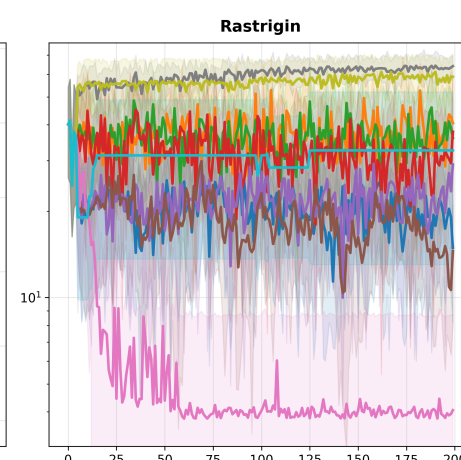
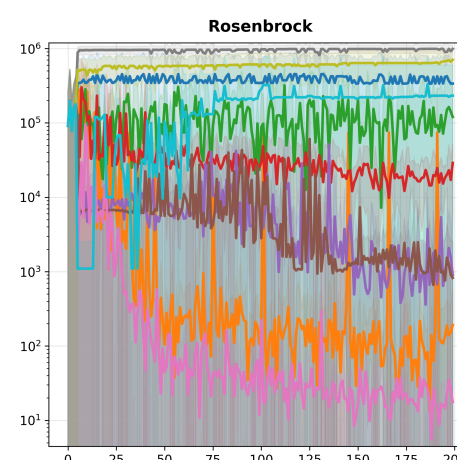
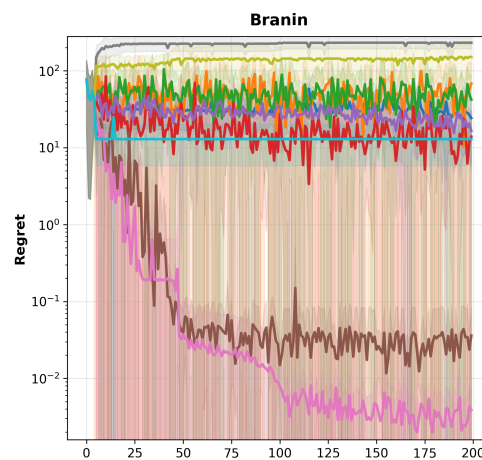
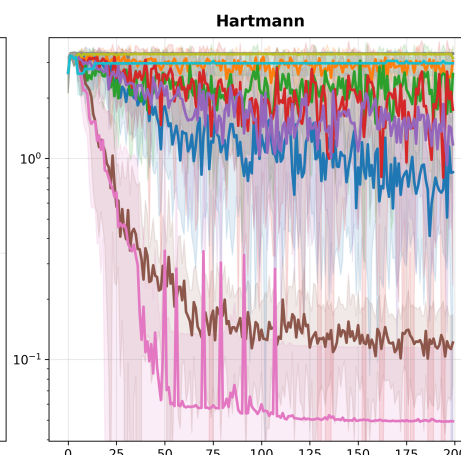
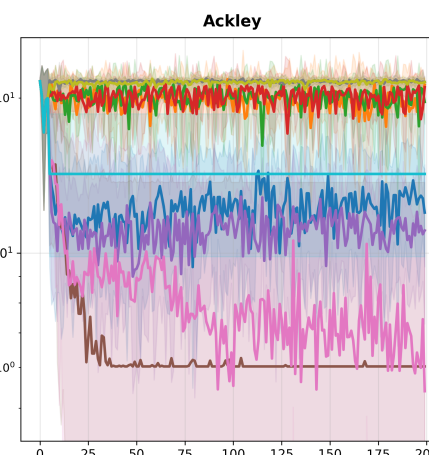
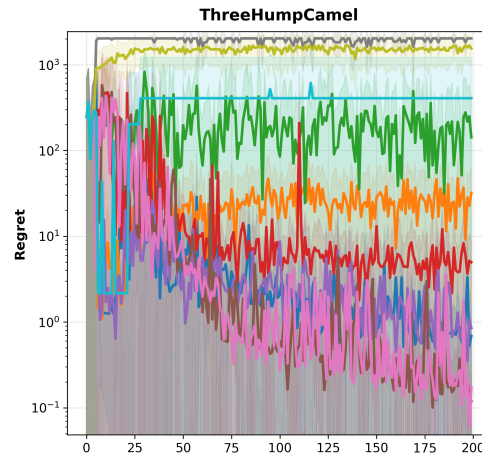
Faculty of
Computer Science

Appendix: Results



PressureVessel

WeldedBeamSO



TensionCompressionString

SpeedReducer

