Universitetet i Stavanger

TTK Institute for Theoretical Particle Physics and Cosmology | RWTH AACHEN UNIVERSITY

GPry: A package for fast Bayesian inference with Gaussian Processes

Cosmology from Home Conference 2022

```
pip install gpry
```

https://github.com/jonaselgammal/GPry

JONAS EL GAMMAL

JESUS TORRADO, CHRISTIAN FIDLER, NILS SCHÖNEBERG

1. Motivation

Parameter estimation with Bayesian inference: A typical example

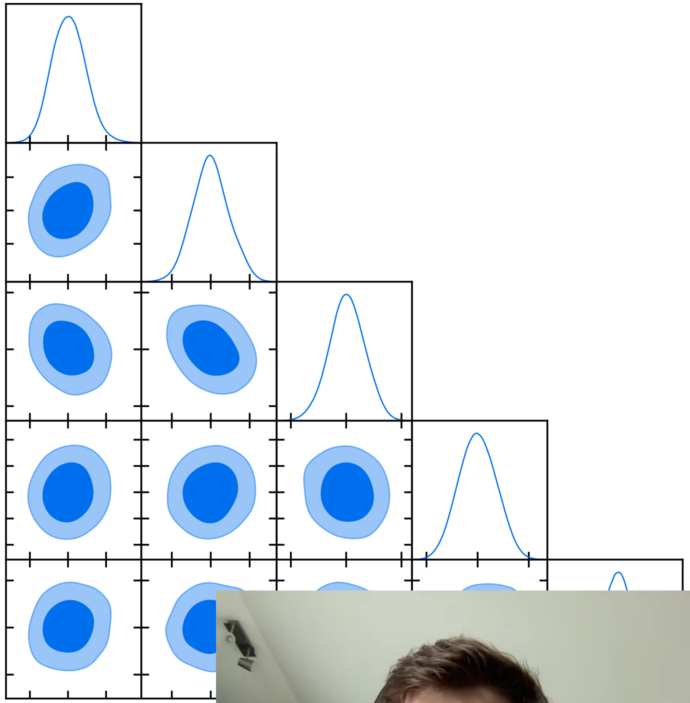
Physical parameters
- H_0
- Ω_b
- m_DM
- …

→ Theory code(s) → typically slow → (Log) Likelihood(s)

Data → Nuisance parameters
- Instruments
- Cuts
- Background
- …

→ typically fast → (Log) Likelihood(s)

(Log) Likelihood(s) → Monte Carlo Sampler →



1. Motivation

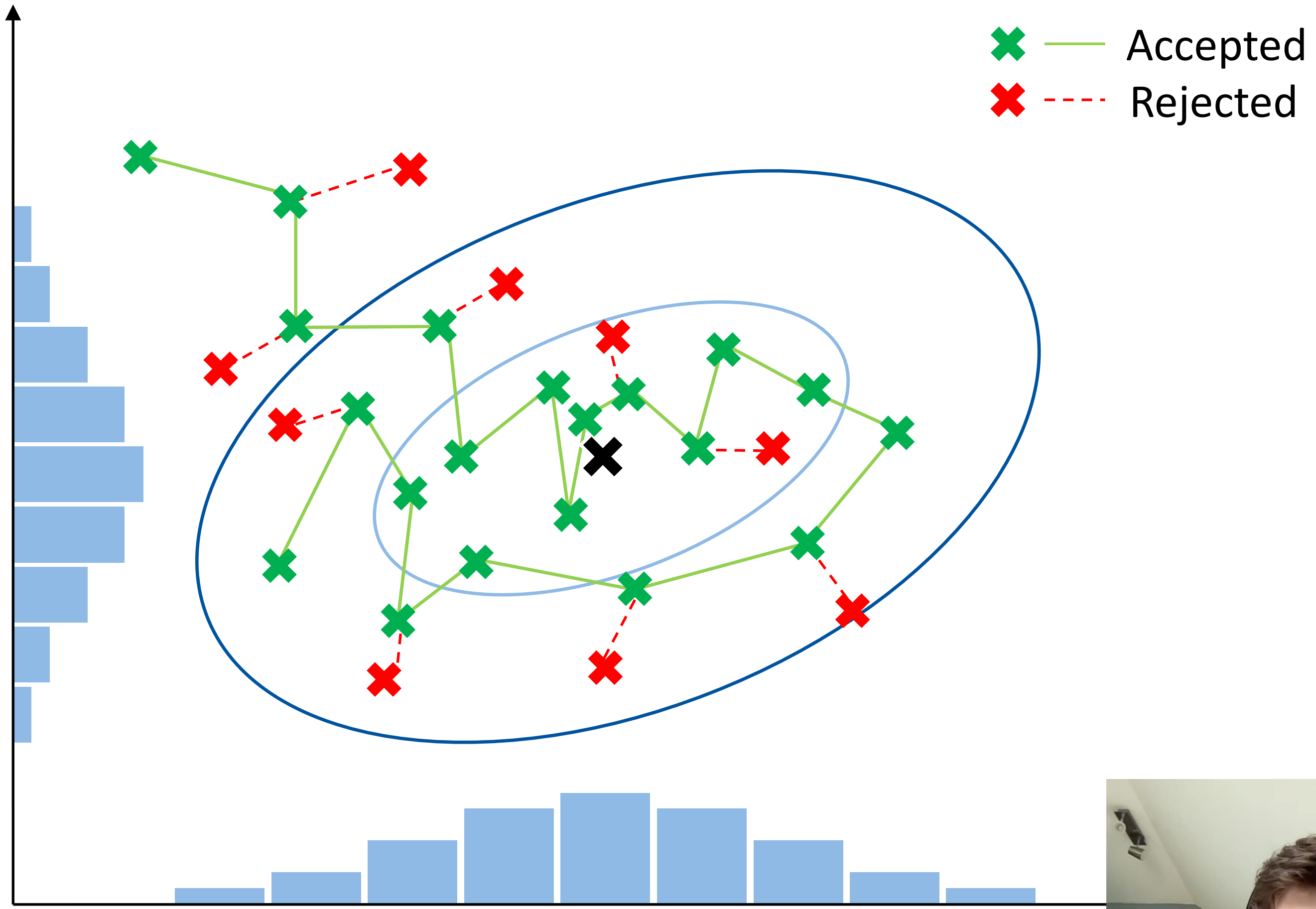
State of the art: MCMC

Advantages

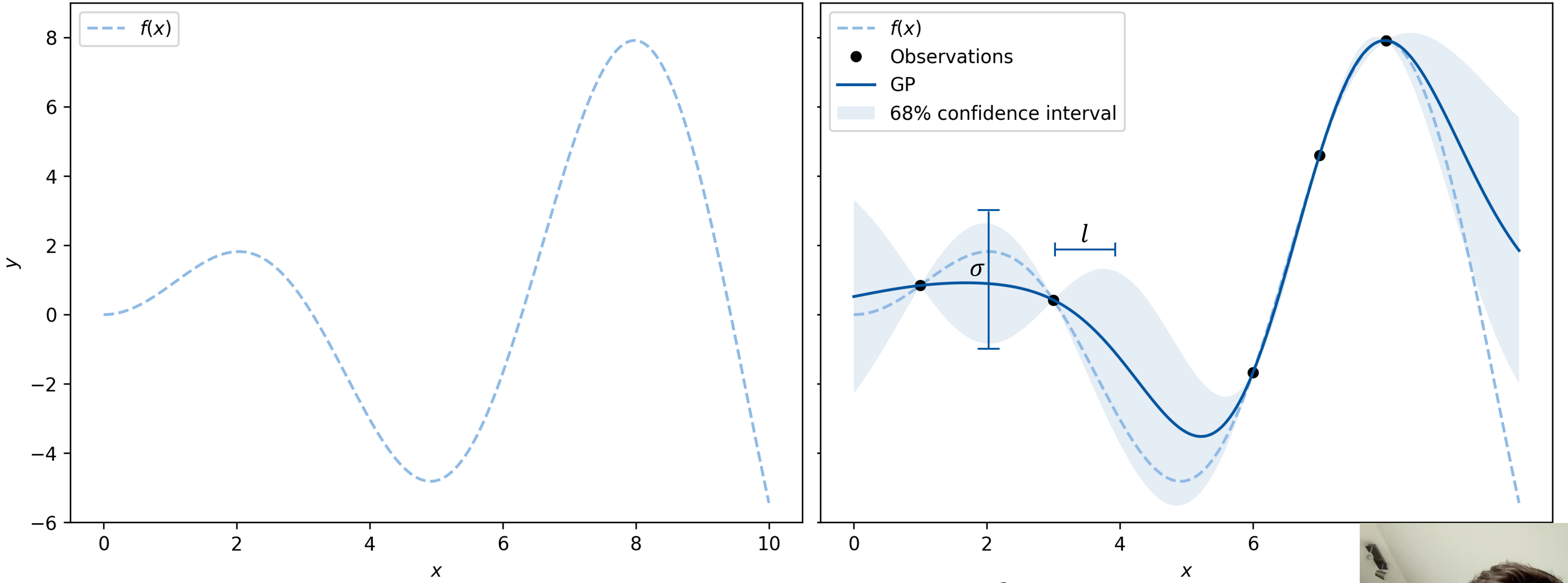
- Easy to implement
- Versatile

Disadvantages

- Only a fraction of Likelihood evaluations is used for estimating the posterior
- Discards the likelihood values and the position of rejected samples



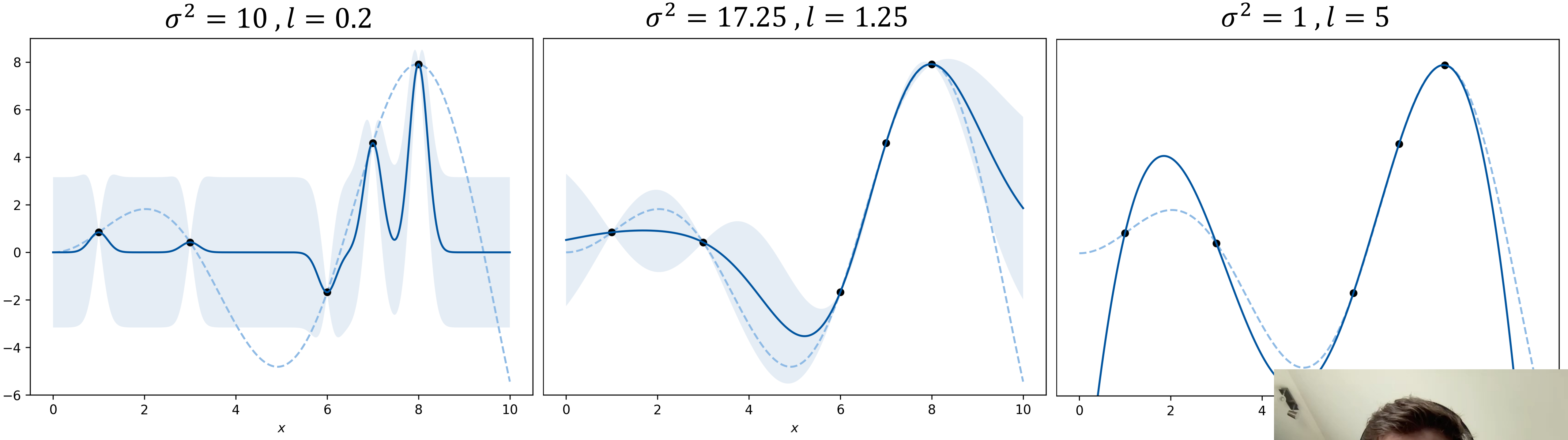
2. Gaussian Process Regression



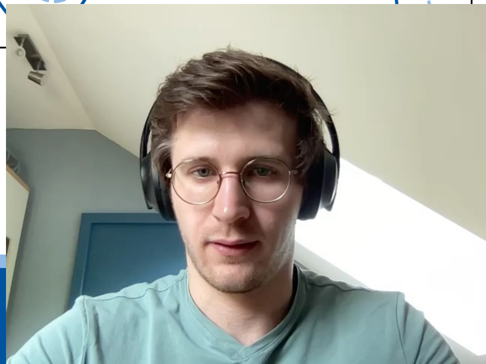
$$k(x, x') = \sigma^2 \cdot \exp\left(-\frac{|x - x'|^2}{2l^2}\right)$$

2. Gaussian Process Regression

$$k(x, x') = \sigma^2 \cdot \exp\left(-\frac{|x - x'|^2}{2l^2}\right)$$

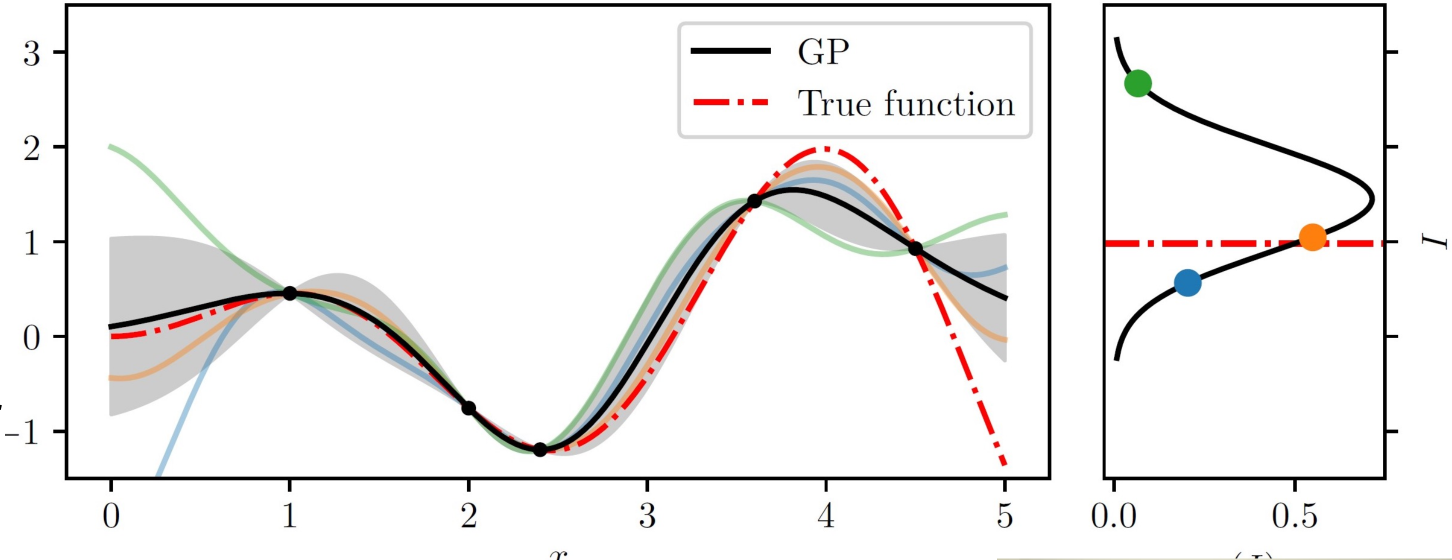


Can use MAP to get the best estimate



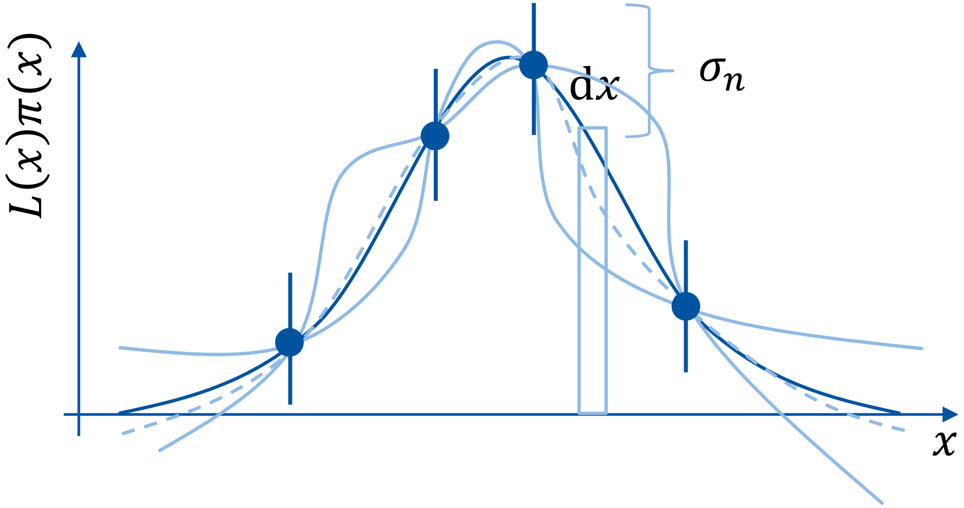
3. Bayesian quadrature

- To get marginalised quantities we want to integrate
  $$\int L(x)\pi(x)\,\mathrm{d}x$$
- With a GP we can get a model for
  $$L(x)\pi(x) \sim \mathcal{GP}(0, k(x, x'))$$
- We can integrate that model by integrating $\int \mu(x)\,\mathrm{d}x = \int \bar{f}(x)\,\mathrm{d}x$
- We can use $\mu(x)$ and $\sigma(x) = \sqrt{\mathrm{cov}(f_*(x,x))}$ to find the next most informative point to sample



3. Bayesian quadrature

⇒ At each step maximize an **acquisition function**



$L(x)\pi(x)$ is always positive

$$\Rightarrow a(x) = \mu(x) \cdot \sigma(x)$$

$L(x)\pi(x)$ has high dynamic range

⇒ Sample log-posterior:

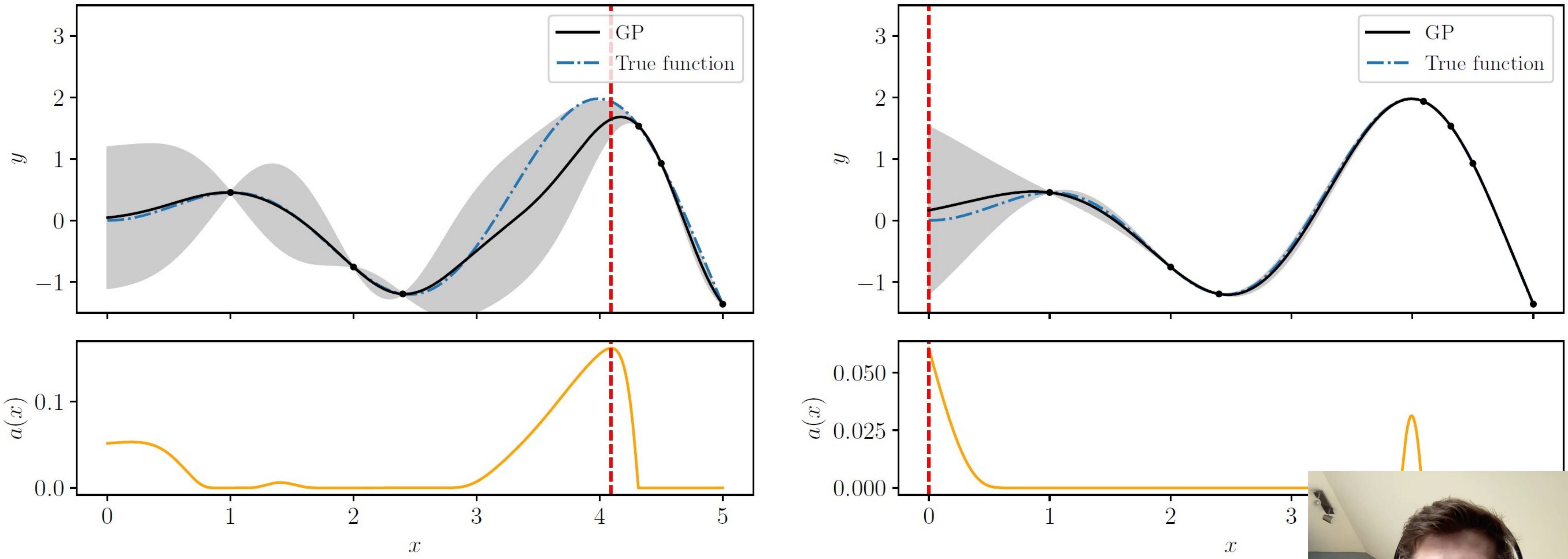
$$a(x) = \exp(2 \cdot \bar{\mu}) \cdot \sigma_{\bar{\mu}}(x)$$

$\bar{\mu} =$ Mean of GP fit to log-posterior

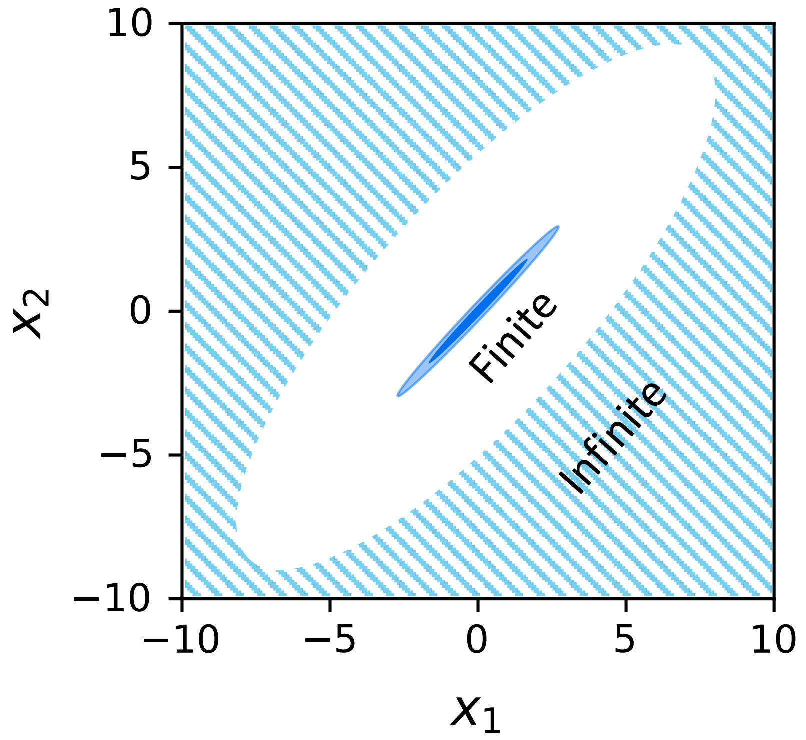
Correction factor ζ and statistical noise σ_n

$$a(x) = \exp(2\zeta \cdot \bar{\mu}) \cdot (\sigma_{\bar{\mu}}(x)$$

3. Bayesian quadrature



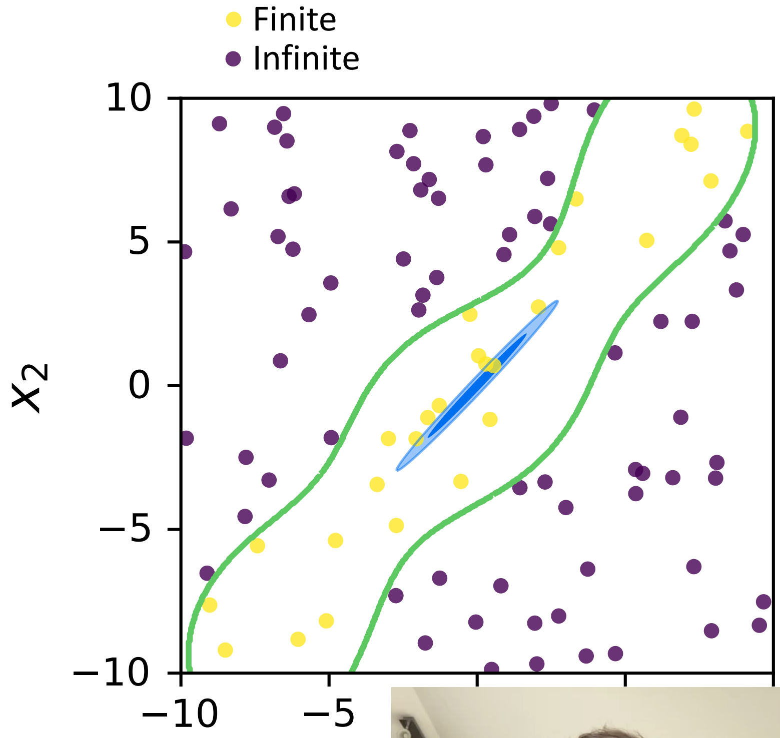
4. Region of interest



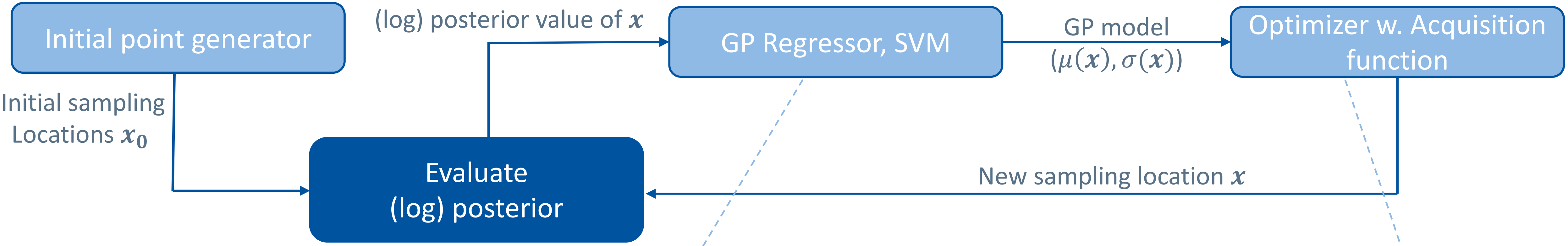
- $p(x) \to 0 \Rightarrow \log p(x) \to -\infty$
- No point in modelling regions far away from the maximum

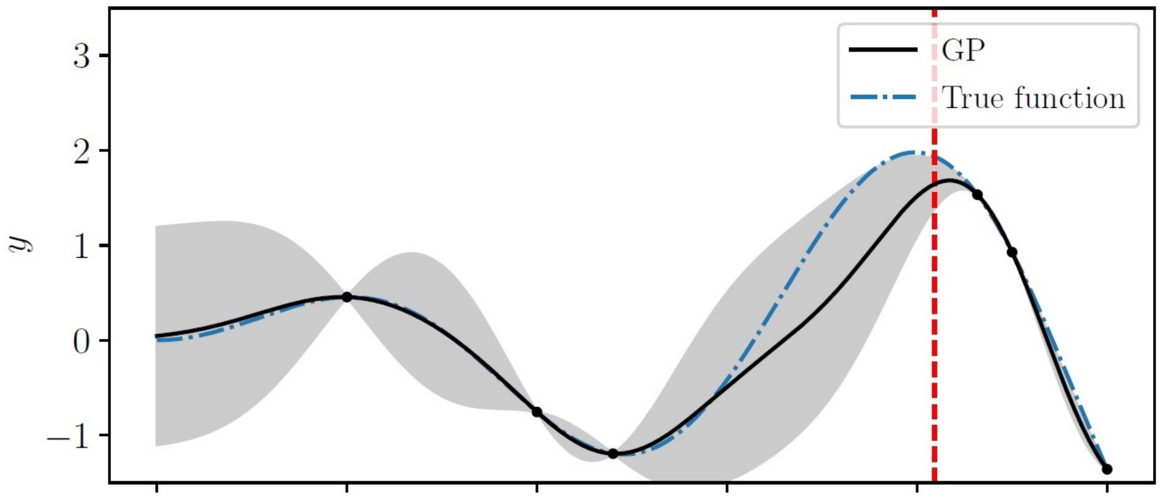
Solution: SVM Classifier

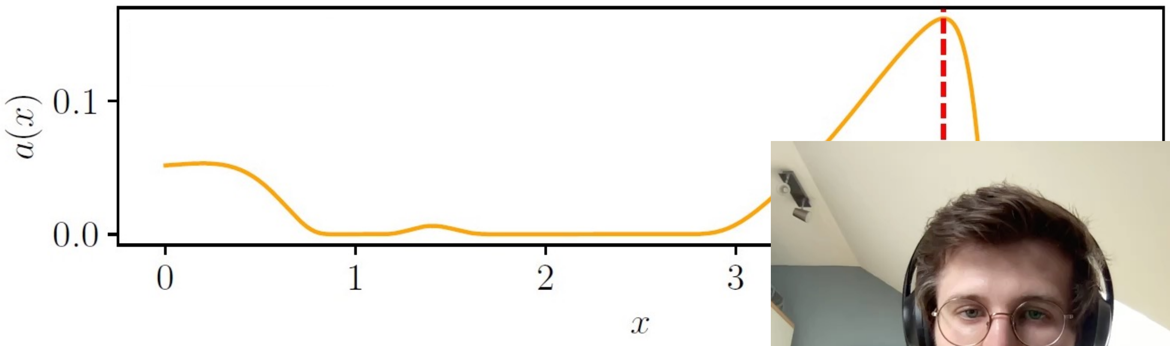
Multiply μ with $-\infty$ where SVM classifies as infinite



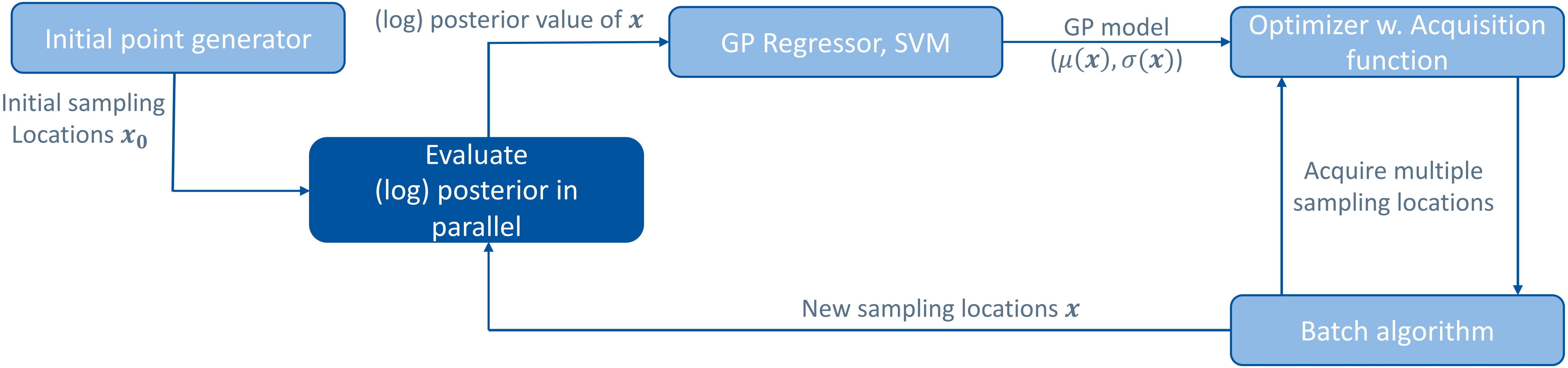
5. The Algorithm



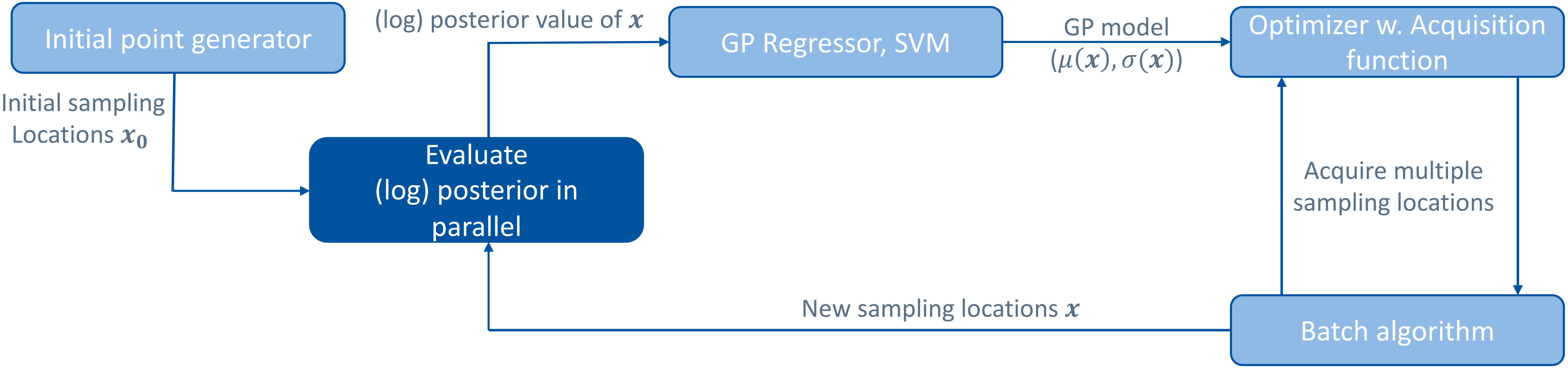




5. The Algorithm



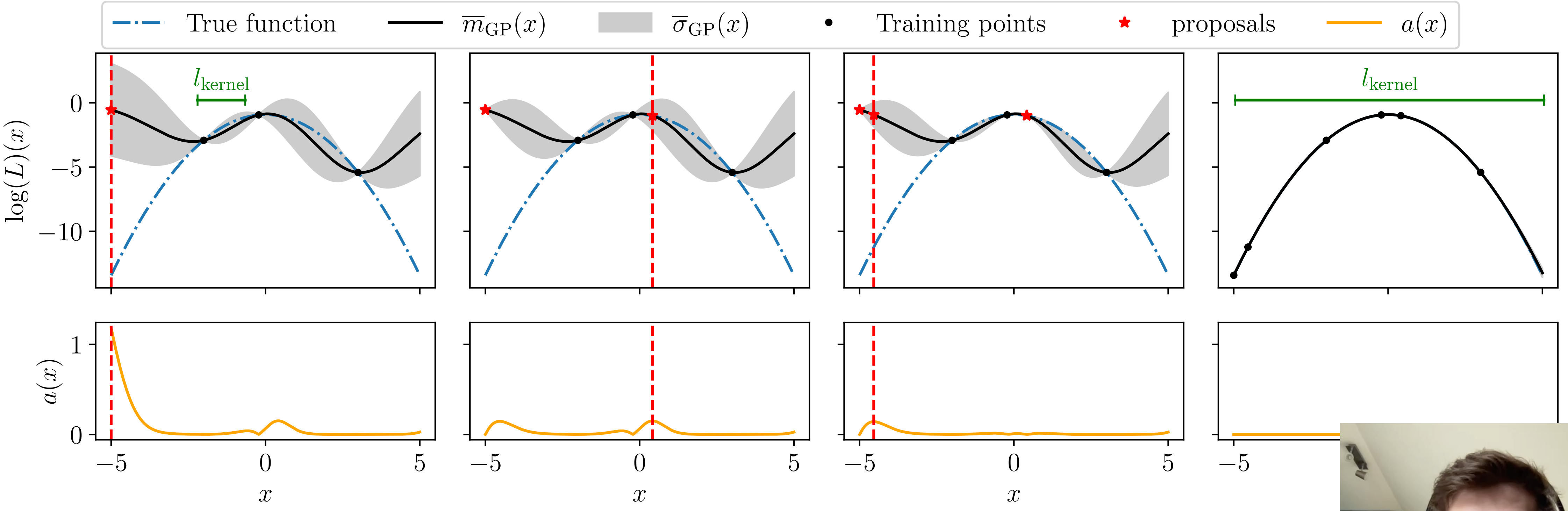
5. The Algorithm



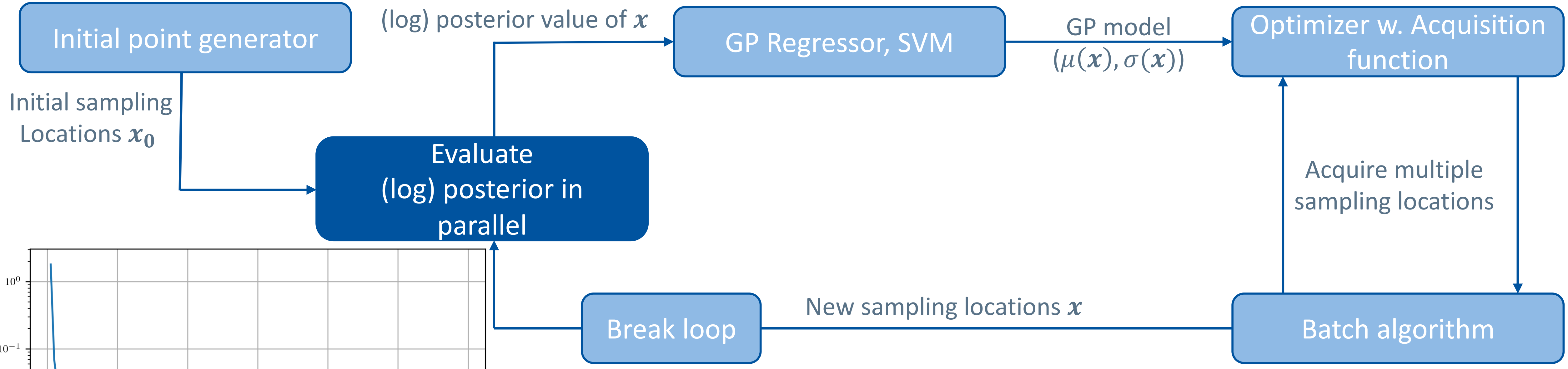
The Kriging believer algorithm:

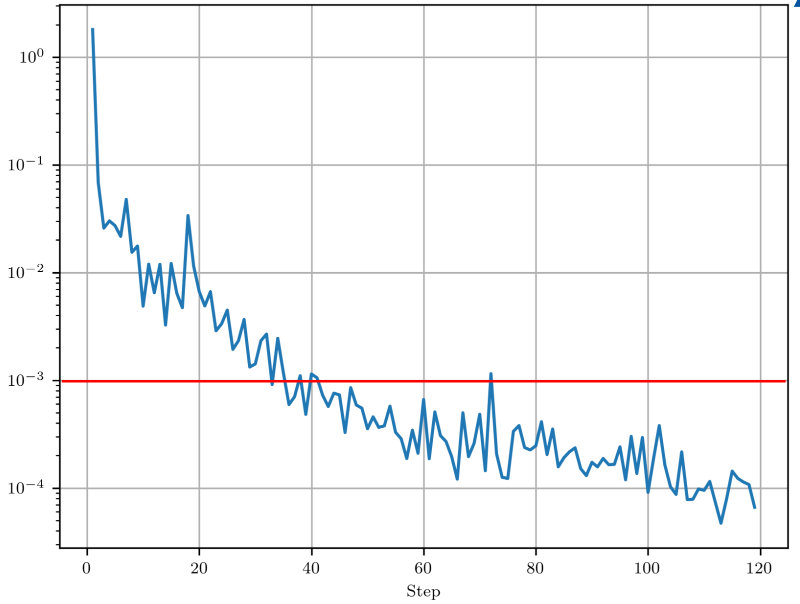
- Instead of evaluating p assume $p(x) \approx \mu_{\mathrm{GP}}(x)$ at next sampling location
- Recompute the GP regressor and acquisition fu this *lie*
- Find next point with active sampling

5. The Algorithm



5. The Algorithm



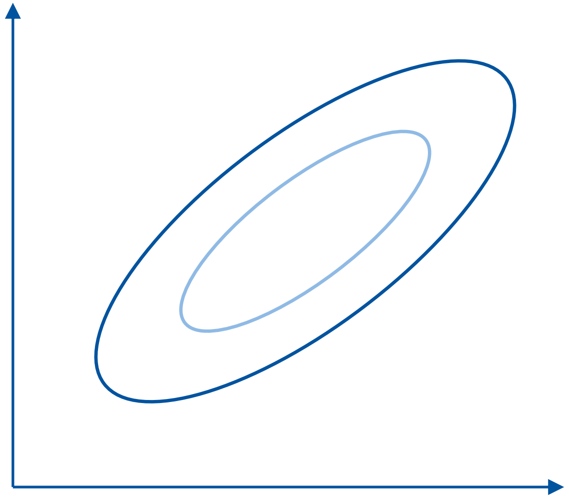


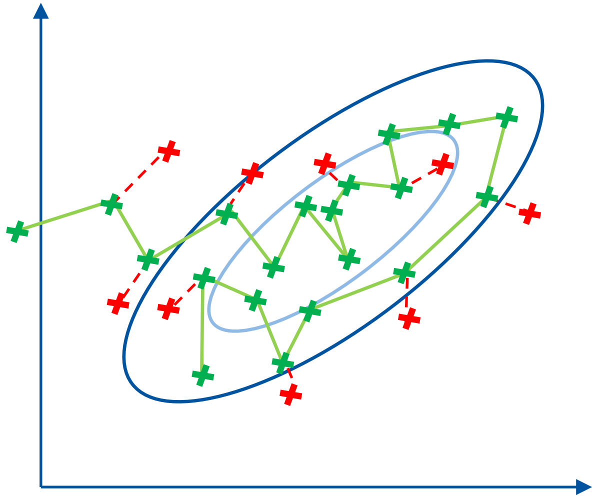
Break if

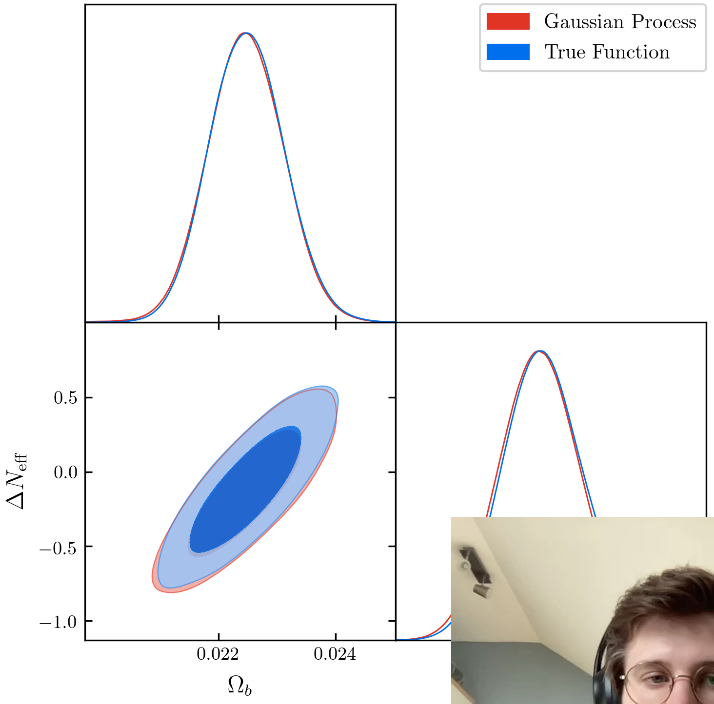
$$\left|\log p_{\mathrm{true}} - \log p_{\mathrm{predict}}\right| < \alpha + \beta * |\log p_{\max} - \log p_{\mathrm{true}}|$$

6. Marginalised quantities

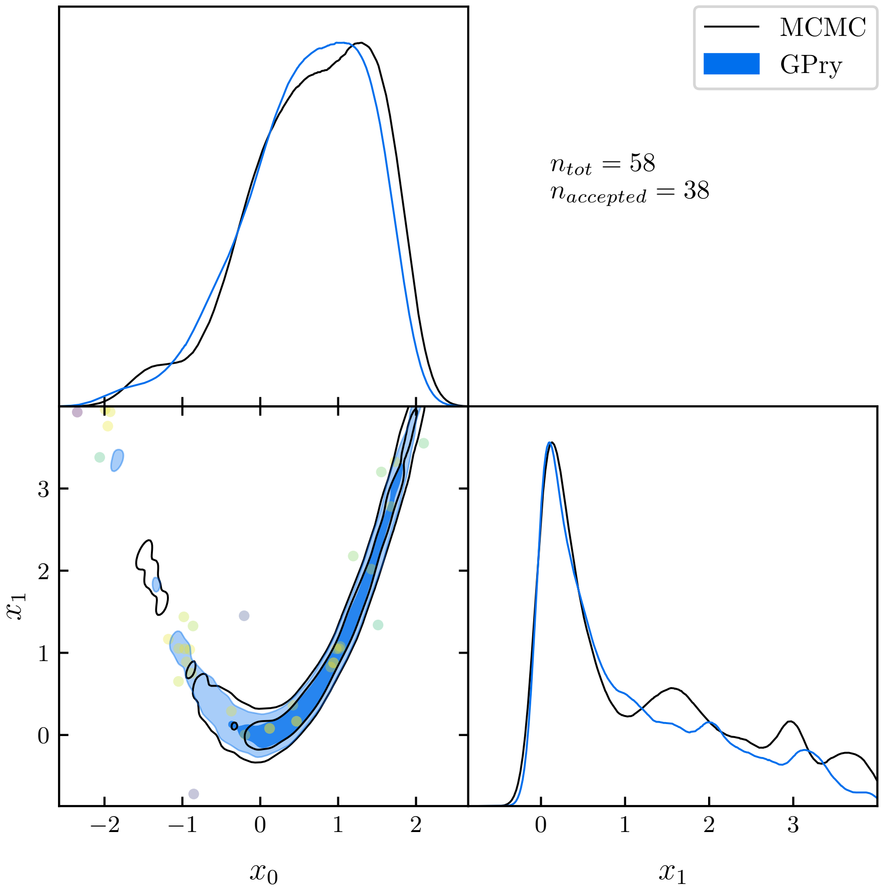
GP model $(\mu(x), \sigma(x))$ → MC sampler → Plotting etc.

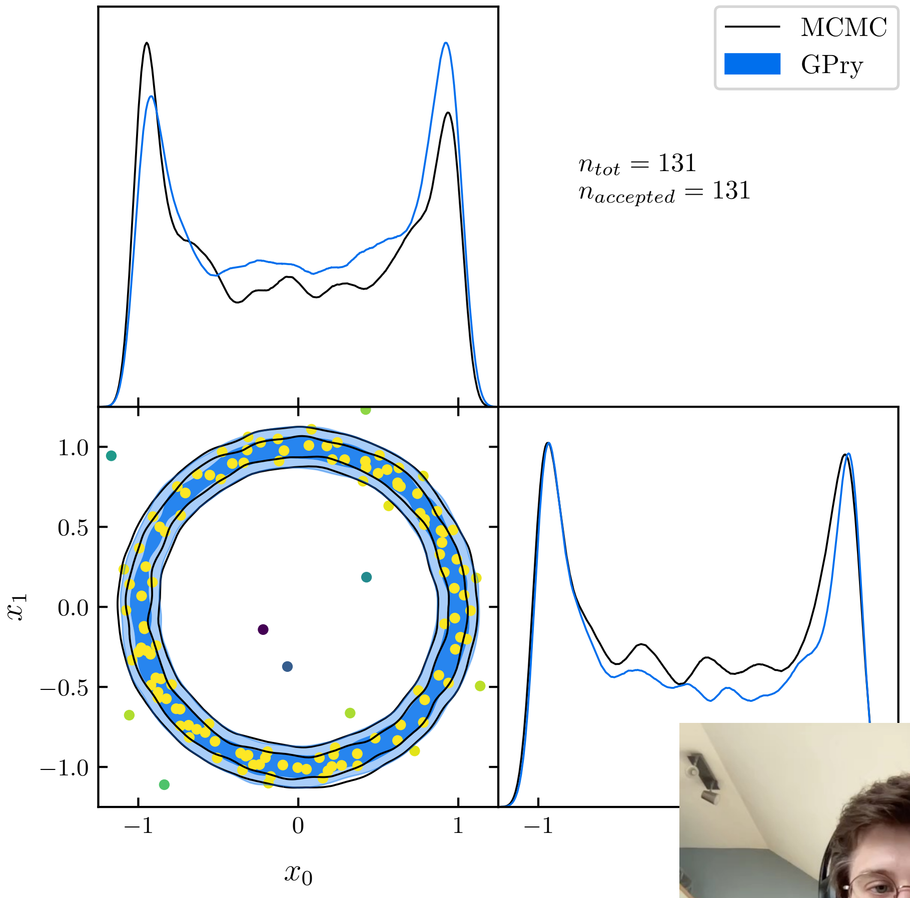




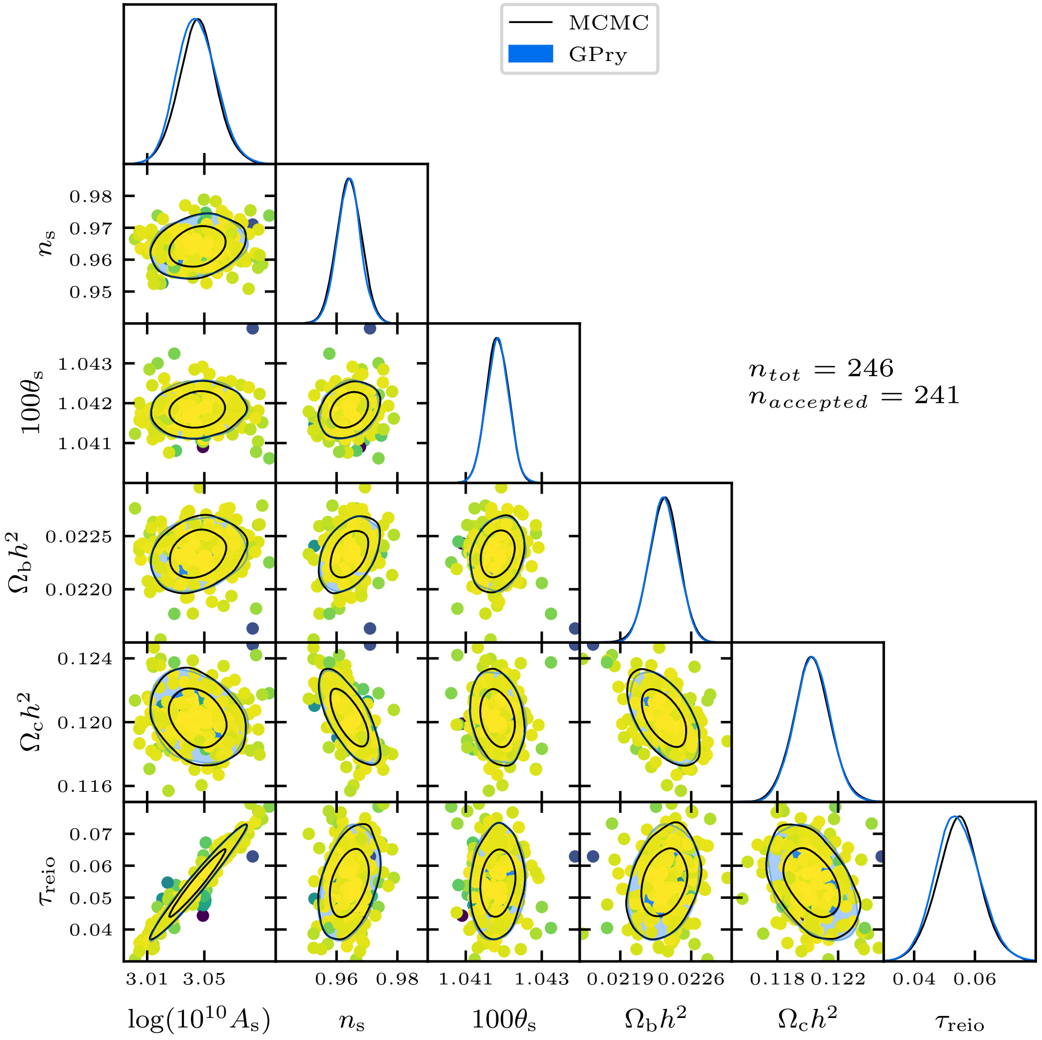


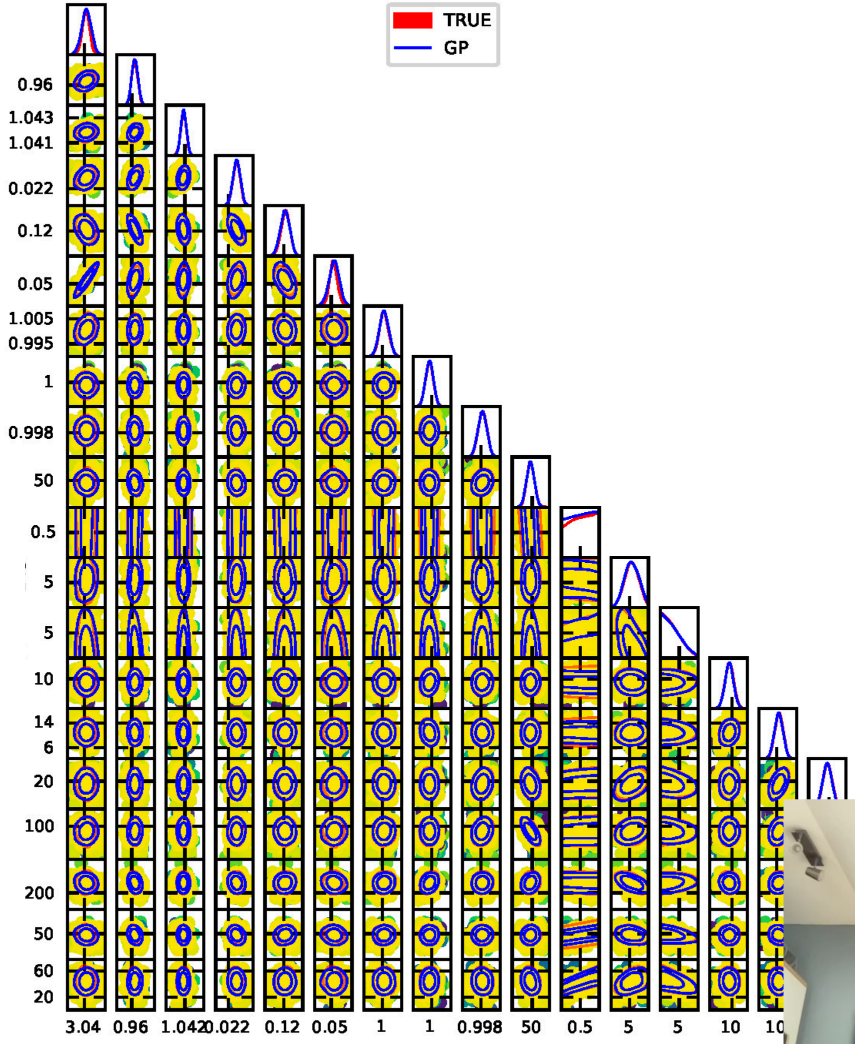
7. Experiments





7. Experiments



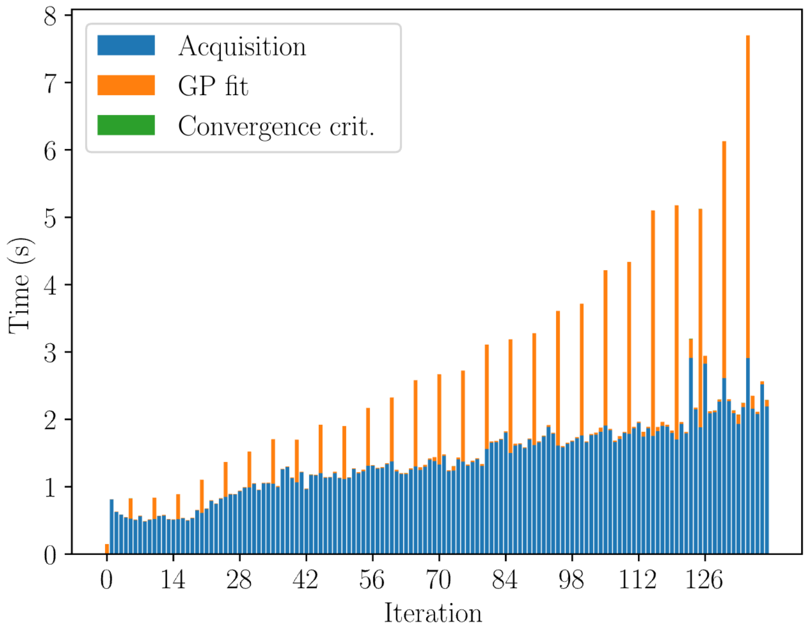


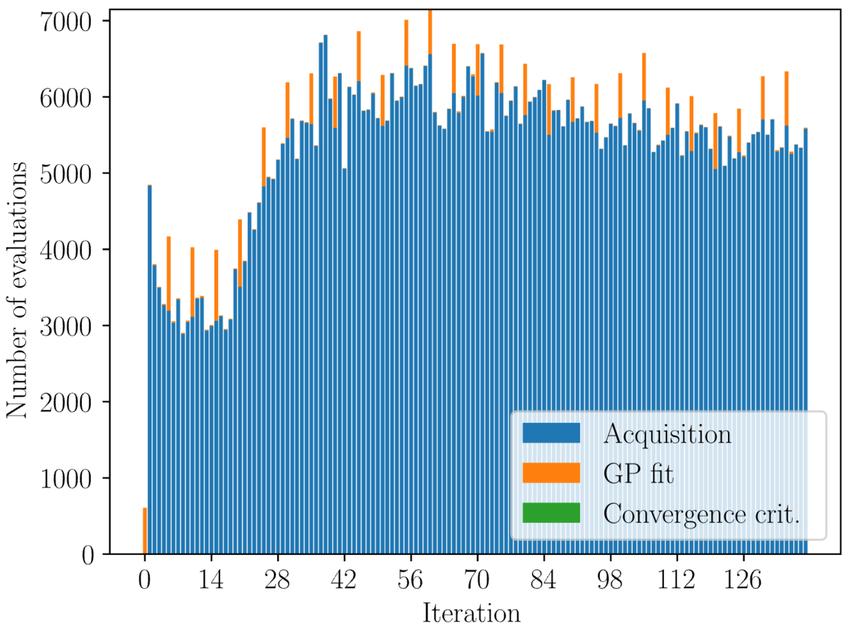
7. Overhead

8 dimensions

2 Kriging believer steps/iteration

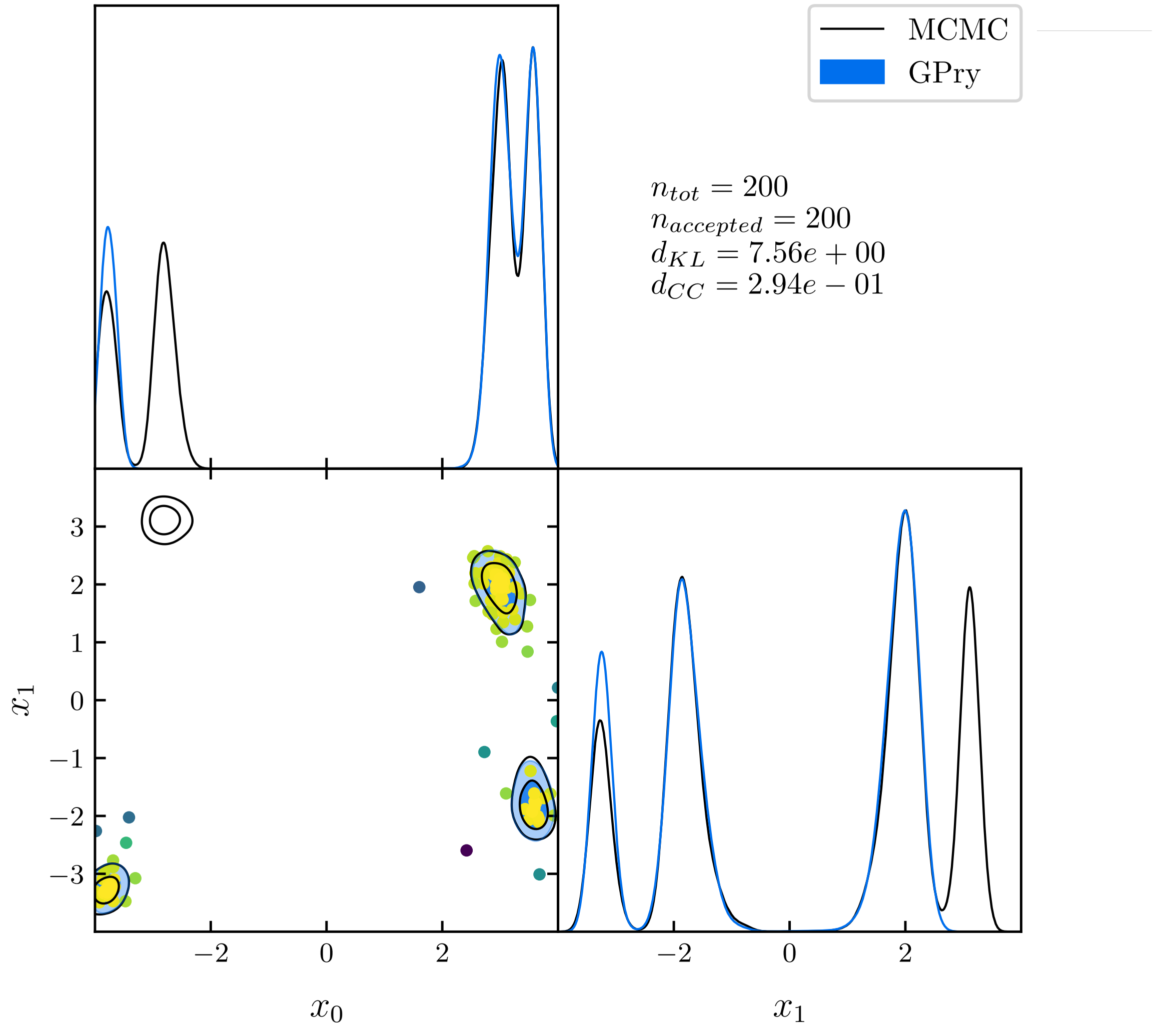
In total 300 accepted samples





Refitting GP hyperparameters requires many inversions of the kernel matrix, scales $\mathcal{O}(N_{\text{samples}}^3)$

8. Limitations



- Multimodality
- Overhead
- Limiting correlation lengths in kernel/preventing overfitting
- Dimensionality (limit $\approx 15\ d$)
- Vastly different correlation lengths

9. Practical info

PyPI	https://pypi.org/project/gpry/
GitHub	https://github.com/jonaselgammal/GPry
Read the Docs	https://gpry.readthedocs.io

Dependencies:
- Cobaya
- Scikit-learn
- mpi4py