Epistemic Max-Min with Bayesian Optimization

Ondrej Bajgar & Michael A. Osborne

1. Task

Setting

- We're interested in the **minimum** of an expensive-to-evaluate unknown $f : \mathcal{X} \to \mathbb{R}$
- Surrogate model $\mathcal{M}(D_t)$ encodes our current knowledge of f based on data $\mathcal{D}_t = \mathcal{D}_0 \cup \{(x_1, y_1), ..., (x_t, y_t)\}$
- We sequentially select query points x_t and receive observations $y_t = f(x_t)$
- The usual goal in Bayesian optimization is to query as low a value as possible or reduce uncertainty in location or value of the minimum.

Our goal: Robust minimization

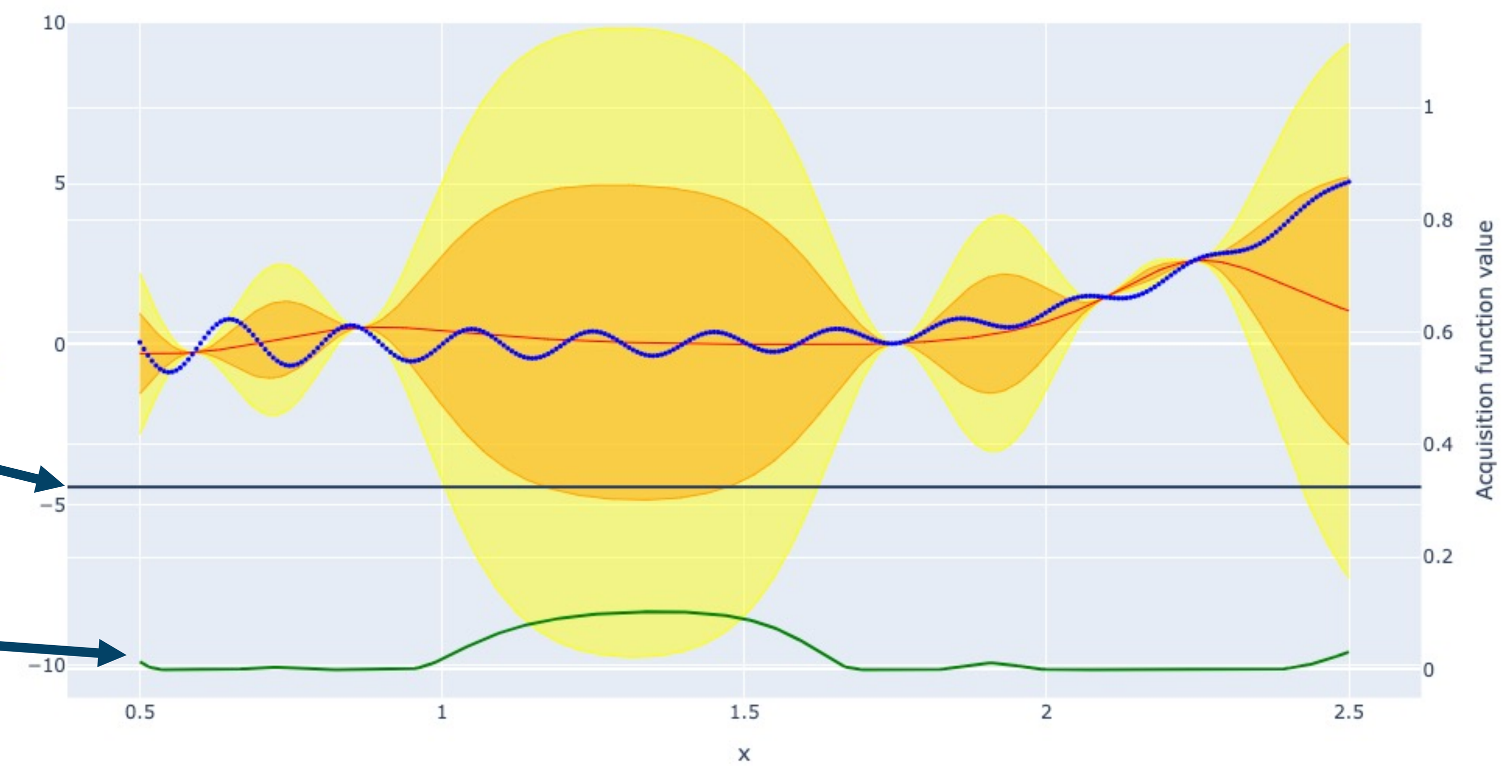
By choosing suitable query points, maximize the ε-quantile of the minimum (a probabilistic lower bound) with respect to the surrogate model: $\max_{x_1,...,x_T} Q_\epsilon(\min_x f(x) | f \sim \mathcal{M}(\mathcal{D}_T))$

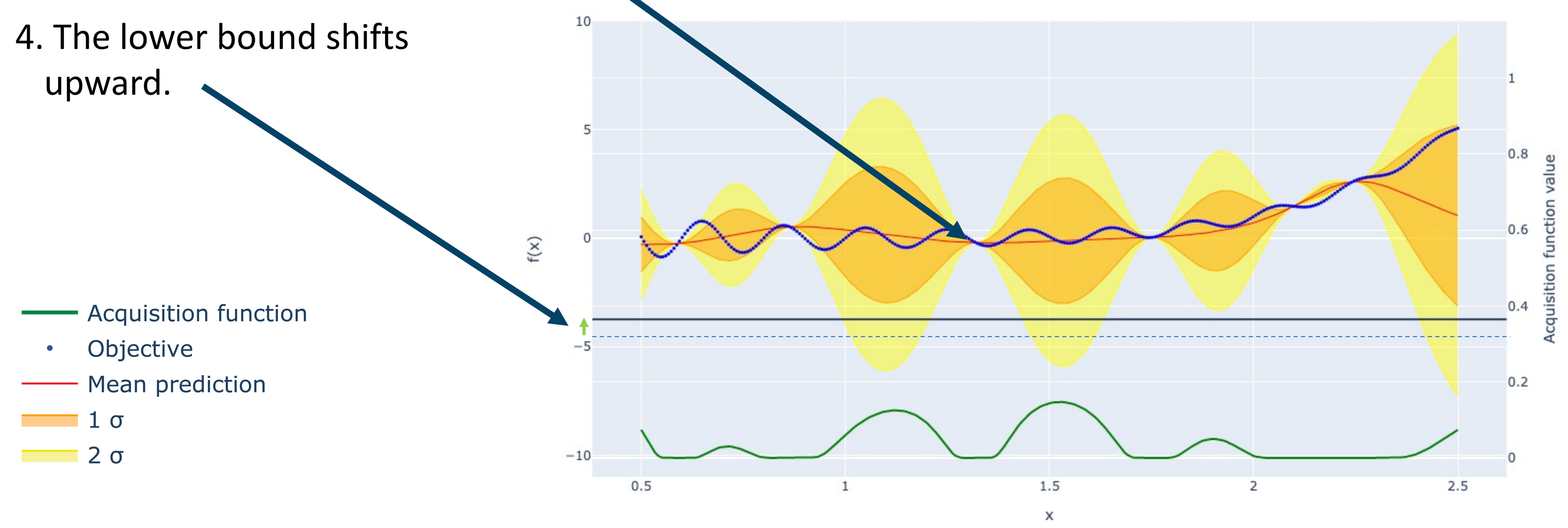
Possible application: DoS attack prevention

- We are responsible for the performance *f(x)* of a system, which depends on inputs x.
- f is a-priori unknown, but we can run costly experiments to reduce uncertainty.
- We want to create a conservative estimate of worst-case performance.
- If worst-case performance is too poor, we must spend resources to increase f.
- Increasing f is costly, so we want the worst-case performance bound to be as tight as possible.

2. Illustration

1. A probabilistic lower bound on the value of the minimum is estimated using the current surrogate model.
2. The acquisition function estimates how promising each point is to evaluate.
3. The maximizer of the acquisition function is chosen to update the surrogate model.
4. The lower bound shifts upward.





3. Algorithm

- Ideal myopic acquisition function: $\alpha_t^{\text{ideal}}(x_t) = \mathbb{E}_{y_t|x_t}[Q_\epsilon(\min_x f(x) | \mathcal{D}_{t-1} \cup \{(x_t, y_t)\})]$ (generally intractable)
- Our approximate acquisition function: $\alpha_t(x_t) = \Phi\left(\bar{y}_{\epsilon,t-1}; \mu_{\mathcal{D}_{t-1}}(x_t), \sigma_{\mathcal{D}_{t-1}}(x_t)\right)$

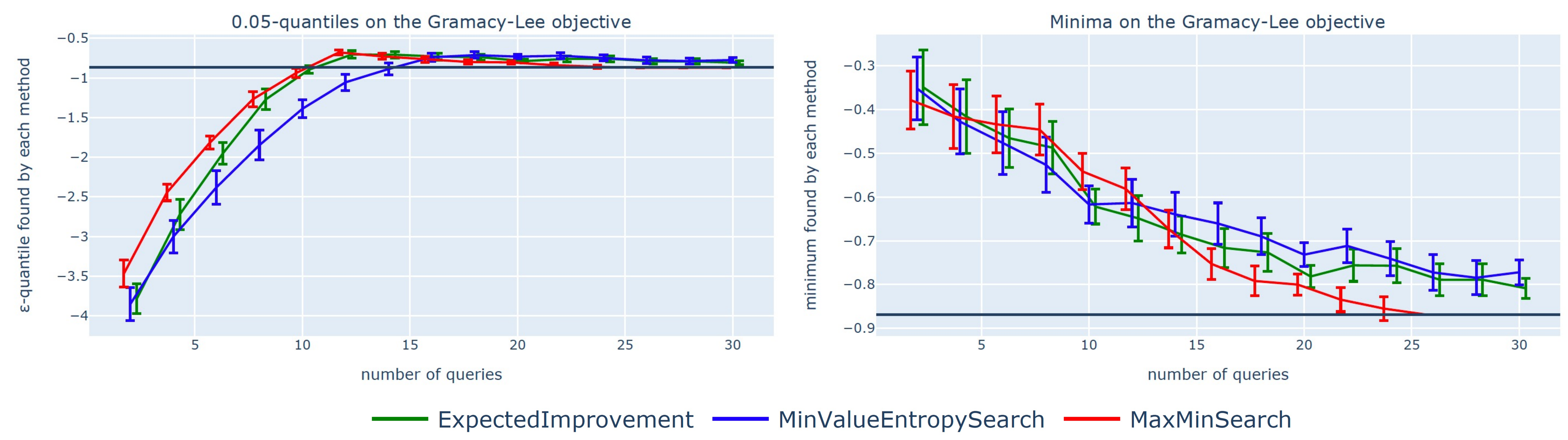
  where
  - Φ is the cdf of the standard Gaussian (when using a Gaussian process surrogate model)
  - $\bar{y}_{\epsilon,t-1}$ is the current probabilistic lower bound (ε-quantile of the min)
  - $\mu_{\mathcal{D}_{t-1}}(x), \sigma_{\mathcal{D}_{t-1}}(x)$ are the mean and standard deviation of the Gaussian process at x
- Intuition:
  - In every time step, select the point with most probability mass below the current lower bound

Algorithm 1: Max-min BO

Result: $\bar{y}_{\epsilon,T}$
Given $\mathcal{D}_0, \epsilon \in (0,1)$, ;
for $t = 1, ..., T$ **do**
  # Find previous quantile lower bound on min:
  $\bar{y}_{\epsilon,t-1} = Q_\epsilon(\min_x f(x) | f \sim \mathcal{M}(\mathcal{D}_{t-1}))$;
  # Choose next evaluation point:
  $x_t = \operatorname{argmax}_x \Phi(y_{\epsilon,t-1}; \mu_{\mathcal{D}_{t-1}}(x), \sigma_{\mathcal{D}_{t-1}}(x))$;
  # Evaluate and add to the dataset:
  $y_t = f(x_t)$;
  $\mathcal{D}_t = D_{t-1} \cup \{(x_t, y_t)\}$;
end
$\bar{y}_{\epsilon,T} = Q_\epsilon(\min_x f(x) | f \sim \mathcal{M}(\mathcal{D}_T))$;

4. Experiments

- Experiments on synthetic objective functions bring mixed results.
  - Positive example: comparison against 2 standard acquisition functions on the 1-dimensional Gramacy-Lee objective brings faster initial improvements in the lower bound (left)



- We are working on experiments in a more realistic RegEx DoS setting.
- Good calibration of the Gaussian process with respect to estimating the lower bound is crucial and can be difficult to achieve if the minimum significantly deviates from "typical" objective function values
  - This currently makes the method fail on several difficult-to-optimize synthetic objectives.
    -> Calibration of Gaussian process minima is a possible topic for future work.