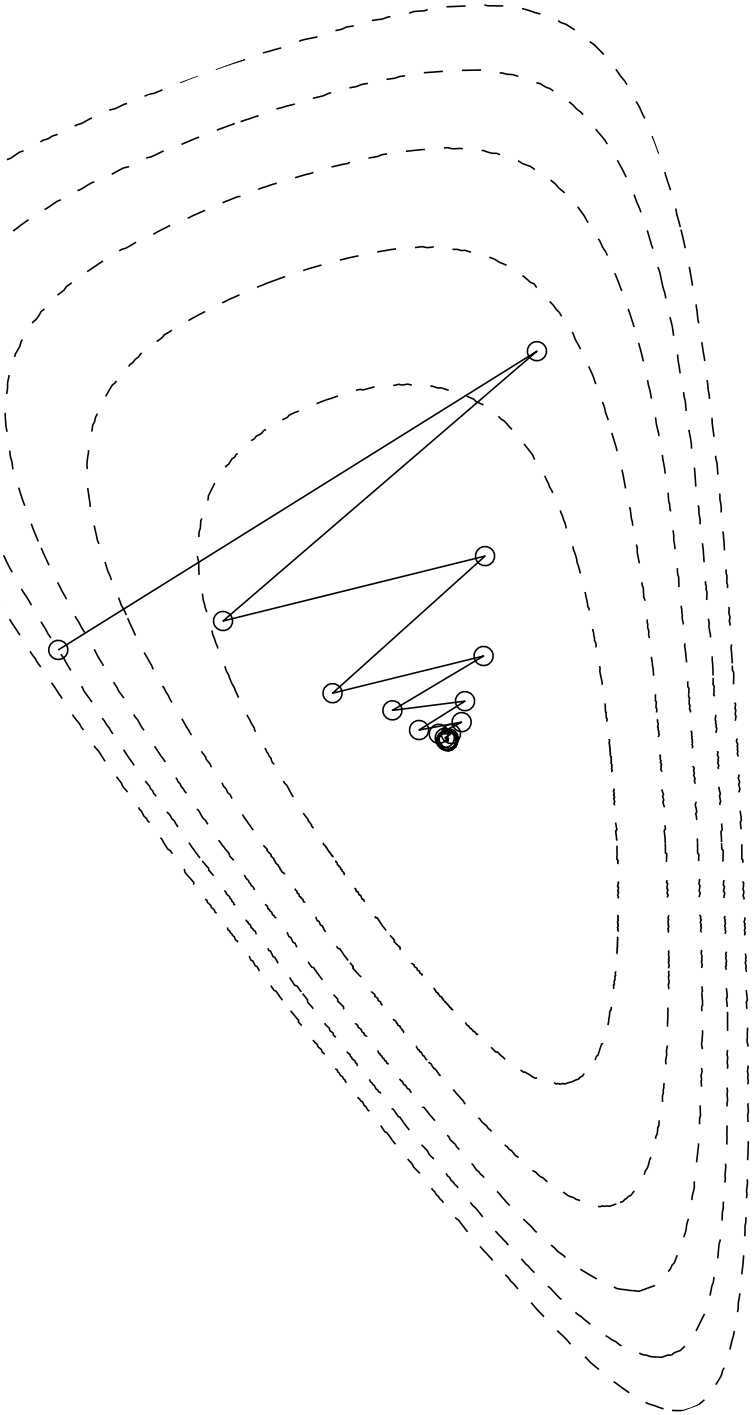


Introduction to Optimization

Global Optimization

Marc Toussaint
U Stuttgart

Global Optimization

- Is there an optimal way to optimize (in the Blackbox case)?
- Is there a way to find the *global* optimum instead of only local?

Core references

Journal of Global Optimization **13:** 455–492, 1998.

Efficient Global Optimization of Expensive Black-Box Functions

DONALD R. JONES[1], MATTHIAS SCHONLAU[2,*] and WILLIAM J. WELCH[3,**]

- Jones, D., M. Schonlau, & W. Welch (1998). Efficient global optimization of expensive black-box functions. Journal of Global Optimization 13, 455-492.
- Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. Journal of Global Optimization 21, 345-383.
- Poland, J. (2004). Explicit local models: Towards optimal optimization algorithms. Technical Report No. IDSIA-09-04.

More up-to-date – very nice GP-UCB introduction

Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger

***Abstract*—Many applications require optimizing an unknown, noisy function that is expensive to evaluate. We formalize this task as a multiarmed bandit problem, where the payoff function is either sampled from a Gaussian process (GP) or has low norm in a reproducing kernel Hilbert space. We resolve the important open problem of deriving regret bounds for this setting, which imply novel convergence rates for GP optimization. We analyze an intuitive Gaussian process upper confidence bound (GP-UCB) algorithm, and bound its cumulative regret in terms of maximal information gain, establishing a novel connection between GP optimization and experimental design. Moreover, by bounding the latter in terms of operator spectra, we obtain explicit sublinear regret bounds for many commonly used covariance functions. In some important cases, our bounds have surprisingly weak dependence on the dimensionality. In our experiments on real sensor data, GP-UCB compares favorably with other heuristical GP optimization approaches.**

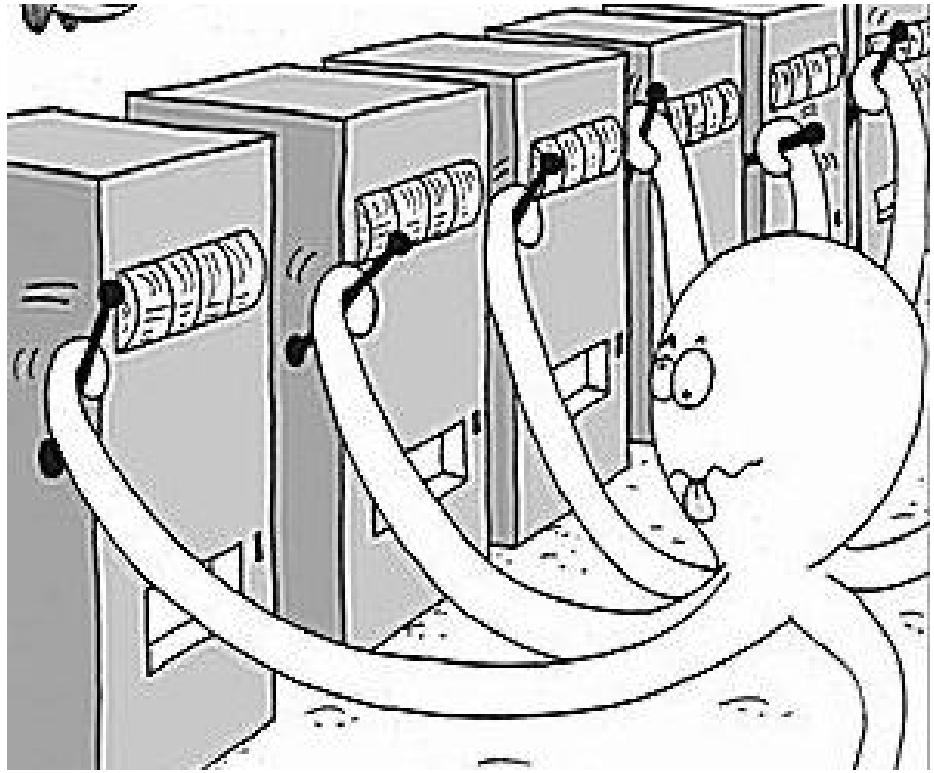
cumulative reward by optimally balancing exploration and exploitation, and experimental design [5], where the function is to be explored globally with as few evaluations as possible, for example, by maximizing information gain. The challenge in both approaches is twofold: we have to estimate an unknown function f from noisy samples, and we must optimize our estimate over some high-dimensional input space. For the former, much progress has been made in machine learning through kernel methods and Gaussian process (GP) models [6], where smoothness assumptions about f are encoded through the choice of kernel in a flexible nonparametric fashion. Beyond Euclidean spaces, kernels can be defined on diverse domains such as spaces of graphs, sets, or lists.

We are concerned with GP optimization in the multiarmed bandit setting, where f is sampled from a GP distribution or has

Outline

- Play a game
- Multi-armed bandits & Upper Confidence Bound (UCB)
- Optimization as infinite bandits; GPs as response surfaces
- Standard criteria:
  - Upper Confidence Bound (UCB)
  - Maximal Probability of Improvement (MPI)
  - Expected Improvement (EI)

Multi-armed bandits



- There are n machines.
  Each machine has an *average* reward f_i – but you don't know the f_i's.

What do you do?

Multi-armed bandits

- Let $a_t \in \{1, .., n\}$ be the choice of machine at time t
  Let $y_t \in \{0, 1\}$ be outcome with mean $\langle y_t \rangle = f_{a_t}$

- A **policy** or strategy maps all the history to a new action:

$$\pi : \ [(a_1, y_1), (a_2, y_2), ..., (a_{t\text{-}1}, y_{t\text{-}1})] \mapsto a_t$$

- Example objectives: find a policy π that

$$\max \left\langle \sum_{t=1}^{T} y_t \right\rangle$$

  or

$$\max \langle y_T \rangle$$

  or other variants.

Exploration vs. Exploitation

- Such kinds of problems appear in many contexts
  (Global Optimization, AI, Reinforcement Learning, etc)

- In simple domains (standard MDPs), actions influence the (external) world state $\rightarrow$ actions navigate through the state space

  In learning domains, **actions influence your knowledge** $\rightarrow$ actions navigate through state and belief space

  In multi-armed bandits, the bandits usually do not have an internal state variable – they are the same every round.

Exploration vs. Exploitation

- The “knowledge” can be represented as the full history

$$h_t \;=\; [(a_1, y_1), (a_2, y_2), ..., (a_{t\text{-}1}, y_{t\text{-}1})]$$

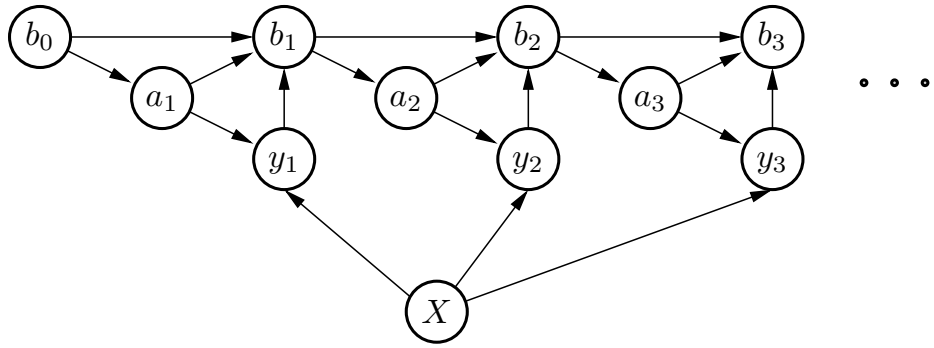
  or, in the Bayesian thinking, as belief

$$b_t = P(X|h_t) = \frac{P(h_t|X)}{P(h_t)} P(X)$$

  where X is all the (unknown) properties of the world

- In the multi-armed bandit case:
  $X = (f_1, .., f_n)$
  $b_t = P(X|h_t) = \prod_i \mathcal{N}(f_i|\hat{y}_{i,t}, \sigma_{i,t})$ (if bandits are Gaussian)

Navigating through Belief Space



- Maximizing for $\langle y_3 \rangle$ requires to have a "good" b_2
- Actions a_1 and a_2 should be planned to achieve best possible b_2
- Action a_3 then greedily chooses machine with highest $\hat{y}_{i,2}$

- Exploration: Choose the next action a_t to $\min \langle H(b_t) \rangle$
- Exploitation: Choose the next action a_t to $\max \langle y_t \rangle$
- Maximizing for $\langle y_T \rangle$ (or similar) requires exploration and exploitation

Such policies can in principle be computed $\rightarrow$ POMDPs (or Lai & Robbins)

But in the following we discuss more efficient 1-step criteria

Upper Confidence Bound (UCB) selection

1: Initializaiton: Play each machine once
2: **repeat**
3: Play the machine i that maximizes $\hat{y}_i + \sqrt{\frac{2 \ln n}{n_i}}$
4: **until**

$\hat{y}_i$ is the average reward of machine i so far

n_i is how often machine i has been played so far

$n = \sum_i n_i$ is the number of rounds so far

(The $\ln n$ makes this work also for non-Gaussian bandits, e.g. heavy-tailed.)

See `lane.compbio.cmu.edu/courses/slides_ucb.pdf` for a summary of Auer et al.

UCB algorithms

- UCB algorithms determine a **confidence interval** such that

$$\hat{y}_i - \sigma_i < f_i < \hat{y}_i + \sigma_i$$

  with high probability.
  UCB chooses the upper bound of this confidence interval
  Strong theory on efficiency of this method in comparision to optimal

- UCB methods are also used for planning:
  Upper Confidence Bounds for Trees (UCT)

How exactly is this related to global optimization?

Global Optimization = infinite bandits

- In global optimization $f(x)$ defines a "reward" for every $x \in \mathbb{R}^n$
  – Instead of a finite number of actions a_t we now have x_t

- Optimal Optimization could be defined as: find a π that

$$\min \left\langle \sum_{t=1}^{T} f(x_t) \right\rangle$$

  or

$$\min \langle f(x_T) \rangle$$

- In principle we know what an optimal optimization algorithm would have to do – it is just computationally infeasible (in general)

Gaussian Processes as belief

- Assume we have a history

$$h_t \;=\; [(x_1, y_1), (x_2, y_2), ..., (x_{t\text{-}1}, y_{t\text{-}1})]$$

- Gaussian Processes are a Machine Learning method that
  – provides a **mean** estimate $\hat{f}(x)$ (*response surface*)
  – provides a variance estimate $\sigma^2(x) \leftrightarrow$ confidence intervals

- Caveat: One needs to make assumptions about the kernel (e.g., how smooth the function is)

1-step criteria based on GPs

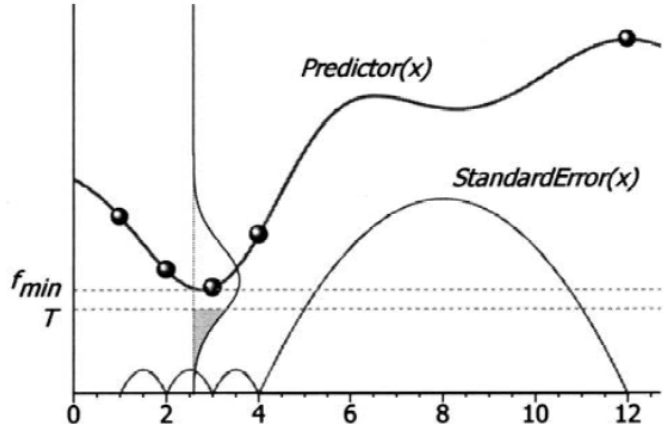


Figure 14. Using kriging, we can estimate the probability that sampling at a given point will 'improve' our solution, in the sense of yielding a value that is equal or better than some target T.

- Maximize Probability of Improvement (MPI)

$$x_t = \operatorname*{argmax}_x \int_{-\infty}^{y^*} \mathcal{N}(y|\hat{f}(x), \sigma(x))$$

- Maximize Expected Improvement (EI)

$$x_t = \operatorname*{argmax}_x \int_{-\infty}^{y^*} \mathcal{N}(y|\hat{f}(x), \sigma(x))\ (y^* - y)$$

- Maximize UCB

$$x_t = \operatorname*{argmax}_x \hat{f}(x) + \beta_t \sigma(x)$$

[Often, $\beta_t = 1$ is chosen. UCB theory allows for better choices. See Srinivas et al.]

Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger

Abstract—**Many applications require optimizing an unknown, noisy function that is expensive to evaluate. We formalize this task as a multiarmed bandit problem, where the payoff function is either sampled from a Gaussian process (GP) or has low norm in a reproducing kernel Hilbert space. We resolve the important open problem of deriving regret bounds for this setting, which imply novel convergence rates for GP optimization. We analyze an intuitive Gaussian process upper confidence bound (GP-UCB) algorithm, and bound its cumulative regret in terms of maximal information gain, establishing a novel connection between GP optimization and experimental design. Moreover, by bounding the latter in terms of operator spectra, we obtain explicit sublinear regret bounds for many commonly used covariance functions. In some important cases, our bounds have surprisingly weak dependence on the dimensionality. In our experiments on real sensor data, GP-UCB compares favorably with other heuristical GP optimization approaches.**

cumulative reward by optimally balancing exploration and exploitation, and experimental design [5], where the function is to be explored globally with as few evaluations as possible, for example, by maximizing information gain. The challenge in both approaches is twofold: we have to estimate an unknown function f from noisy samples, and we must optimize our estimate over some high-dimensional input space. For the former, much progress has been made in machine learning through kernel methods and Gaussian process (GP) models [6], where smoothness assumptions about f are encoded through the choice of kernel in a flexible nonparametric fashion. Beyond Euclidean spaces, kernels can be defined on diverse domains such as spaces of graphs, sets, or lists.

We are concerned with GP optimization in the multiarmed bandit setting, where f is sampled from a GP distribution or has

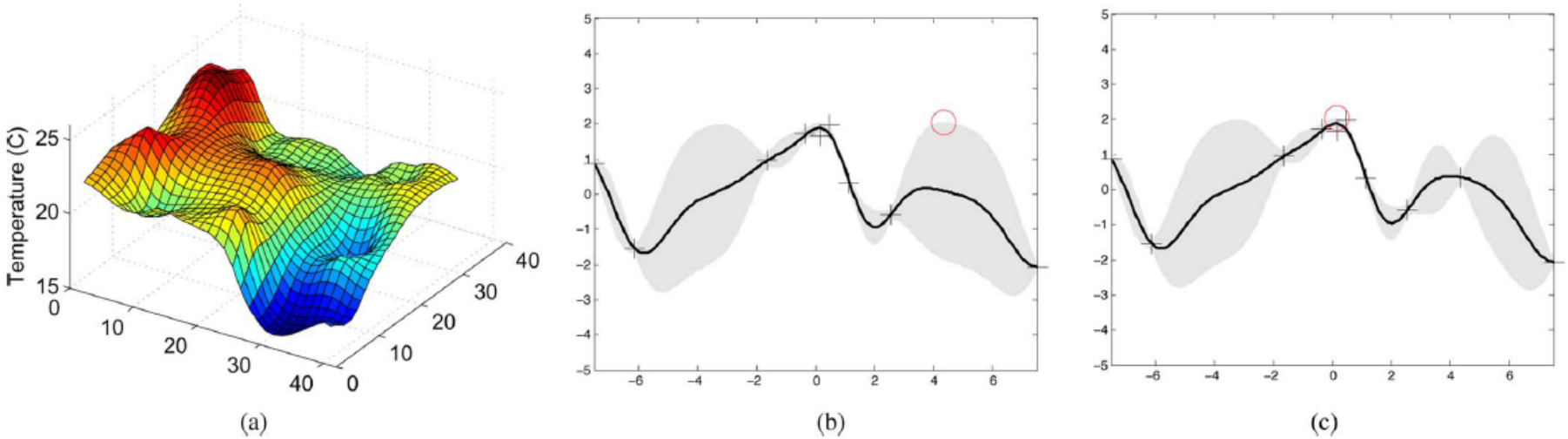


Fig. 2. (a) Example of temperature data collected by a network of 46 sensors at Intel Research Berkeley. (b) and (c) Two iterations of the GP-UCB algorithm. The dark curve indicates the current posterior mean, while the gray bands represent the upper and lower confidence bounds which contain the function with high probability. The "+" mark indicates points that have been sampled before, while the "o" mark shows the point chosen by the GP-UCB algorithm to sample next. It samples points that are either (b) uncertain or have (c) high posterior mean.

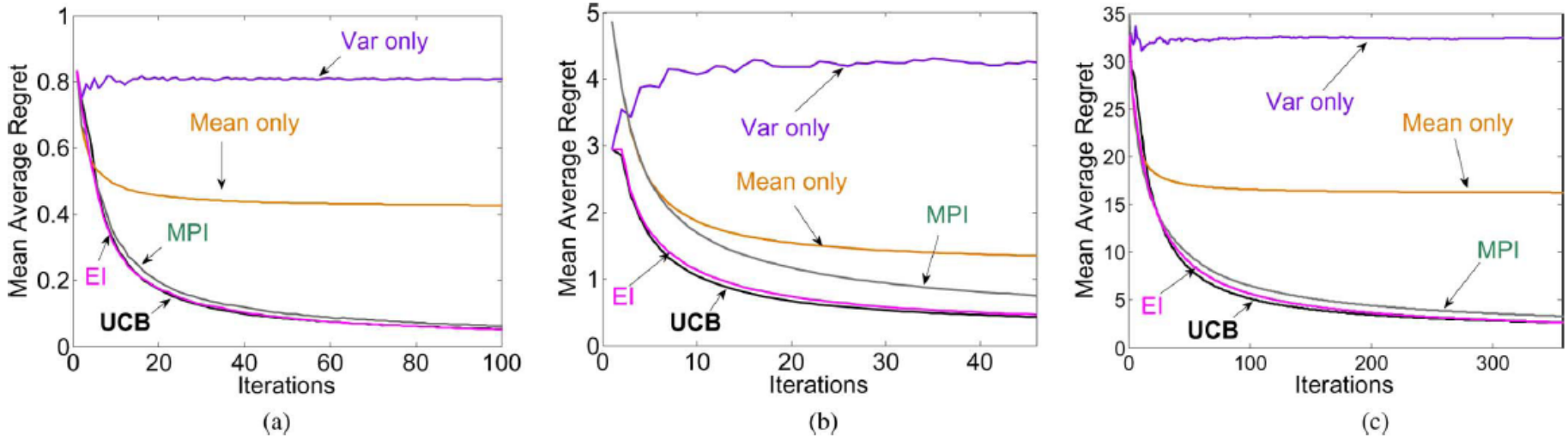


Fig. 6. Mean average regret: GP-UCB and various heuristics on (a) synthetic and (b, c) sensor network data.

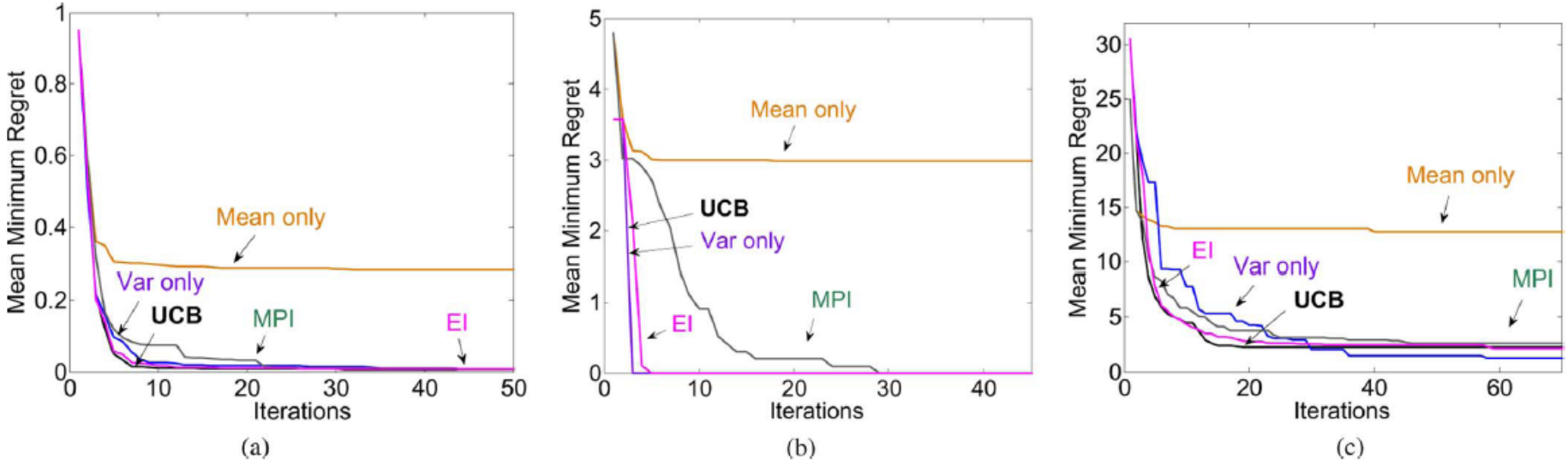


Fig. 7. Mean minimum regret: GP-UCB and various heuristics on (a) synthetic, and (b, c) sensor network data.

Global Optimization

- Given data, we compute a **belief** over $f(x)$
- The belief expresses mean estimate $\hat{f}(x)$ and confidence $\sigma(x)$
  – Use Gaussian Processes or other Bayesian ML methods.
- Optimal Optimization would imply planning in belief space
- Efficient Global Optimization uses 1-step criteria
  – Upper Confidence Bound (UCB)
  – Maximal Probability of Improvement (MPI)
  – Expected Improvement (EI)

- Global Optimization with **gradient** information
  $\rightarrow$ Gaussian Processes with derivative observations