

BAYESIAN OPTIMIZATION OVERVIEW

Marco Forgione

¹IDSIA Dalle Molle Institute for Artificial Intelligence SUPSI-USI, Lugano, Switzerland

December 10, 2019

Bayesian Optimization

Overview

One of the best *off-the-shelf* global optimization algorithms

- Iteratively updates a **stochastic surrogate model** of the unknown $J(\theta)$ via Bayesian inference. Typically, a Gaussian Process (GP)
- Balances **exploitation** and **exploration** by optimizing an **acquisition function** $A(\theta)$ instead of the surrogate model directly
- The acquisition function $A(\theta)$ favors points with estimated **good performance** \rightarrow exploitation and/or **high variance** \rightarrow exploration
- The acquisition function $A(\theta)$ is (relatively) cheap to evaluate. It is a mathematical object!

Bayesian Optimization

Overview

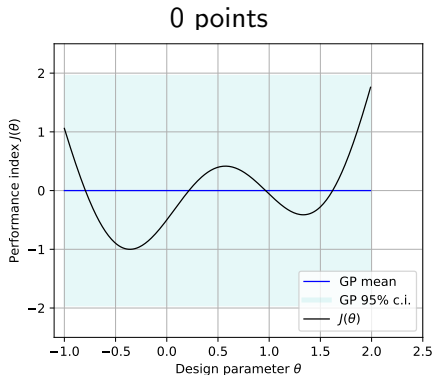
One of the best *off-the-shelf* global optimization algorithms

- Iteratively updates a **stochastic surrogate model** of the unknown $J(\theta)$ via Bayesian inference. Typically, a Gaussian Process (GP)
- Balances **exploitation** and **exploration** by optimizing an **acquisition function** $A(\theta)$ instead of the surrogate model directly
- The acquisition function $A(\theta)$ favors points with estimated **good performance** \rightarrow exploitation and/or **high variance** \rightarrow exploration
- The acquisition function $A(\theta)$ is (relatively) cheap to evaluate. It is a mathematical object!

Bayesian Optimization

Gaussian Process

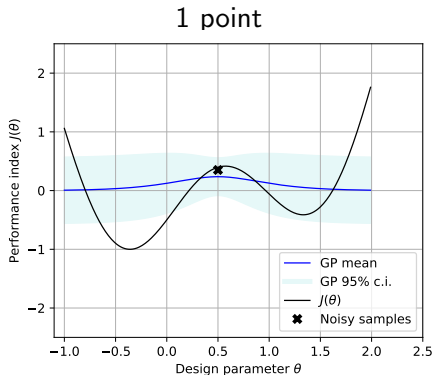
- The function $J(\theta)$ assumed Gaussian with **prior** mean $E[J(\theta)] = \mu(\theta)$ and covariance $\text{cov}[J(\theta_1), J(\theta_2)] = \kappa(\theta_1, \theta_2)$.
- The **posterior** mean and covariance given a new observation (θ_i, J_i) is obtained in closed form



Bayesian Optimization

Gaussian Process

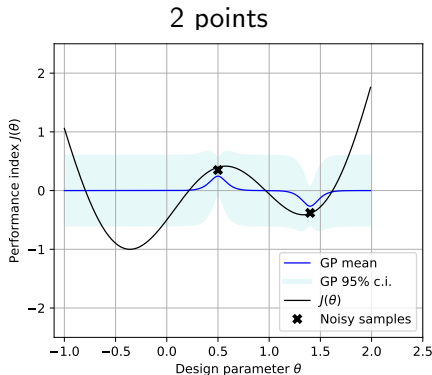
- The function $J(\theta)$ assumed Gaussian with **prior** mean $E[J(\theta)] = \mu(\theta)$ and covariance $\text{cov}[J(\theta_1), J(\theta_2)] = \kappa(\theta_1, \theta_2)$.
- The **posterior** mean and covariance given a new observation (θ_i, J_i) is obtained in closed form



Bayesian Optimization

Gaussian Process

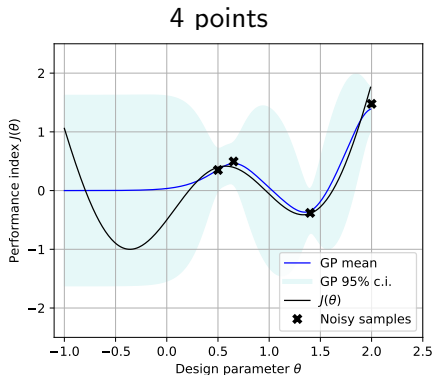
- The function $J(\theta)$ assumed Gaussian with **prior** mean $E[J(\theta)] = \mu(\theta)$ and covariance $\text{cov}[J(\theta_1), J(\theta_2)] = \kappa(\theta_1, \theta_2)$.
- The **posterior** mean and covariance given a new observation (θ_i, J_i) is obtained in closed form



Bayesian Optimization

Gaussian Process

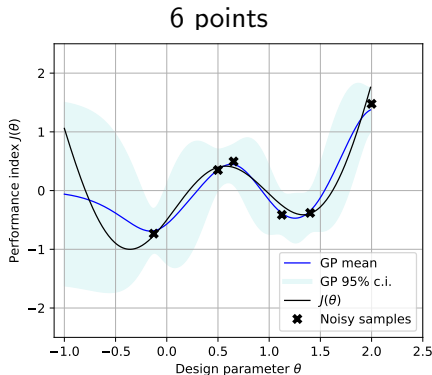
- The function $J(\theta)$ assumed Gaussian with **prior** mean $E[J(\theta)] = \mu(\theta)$ and covariance $\text{cov}[J(\theta_1), J(\theta_2)] = \kappa(\theta_1, \theta_2)$.
- The **posterior** mean and covariance given a new observation (θ_i, J_i) is obtained in closed form



Bayesian Optimization

Gaussian Process

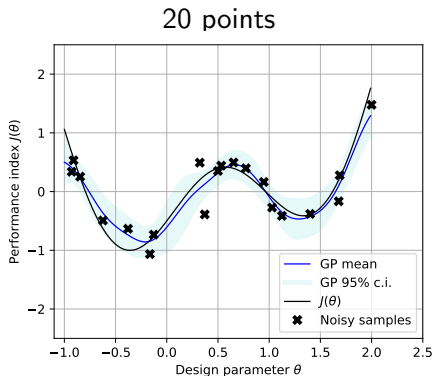
- The function $J(\theta)$ assumed Gaussian with **prior** mean $E[J(\theta)] = \mu(\theta)$ and covariance $\text{cov}[J(\theta_1), J(\theta_2)] = \kappa(\theta_1, \theta_2)$.
- The **posterior** mean and covariance given a new observation (θ_i, J_i) is obtained in closed form



Bayesian Optimization

Gaussian Process

- The function $J(\theta)$ assumed Gaussian with **prior** mean $E[J(\theta)] = \mu(\theta)$ and covariance $\text{cov}[J(\theta_1), J(\theta_2)] = \kappa(\theta_1, \theta_2)$.
- The **posterior** mean and covariance given a new observation (θ_i, J_i) is obtained in closed form



Bayesian Optimization

Acquisition function

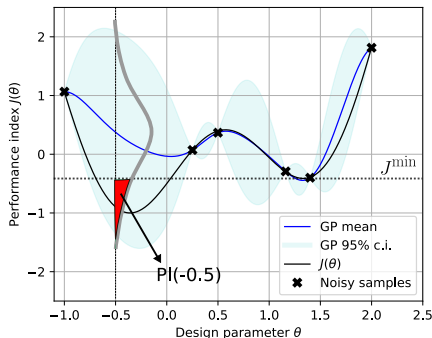
The GP provides the **probability distribution** of $J(\theta)$ for each parameter θ . This probability is used to define an **acquisition function**, e.g.,

Probability of Improvement

Expected improvement

$$A(\theta) = \text{PI}(\theta) = p(J(\theta) \leq J^{\min})$$

$$A(\theta) = \text{EI}(\theta) = \mathbb{E}[\max(0, J^{\min} - J(\theta))]$$



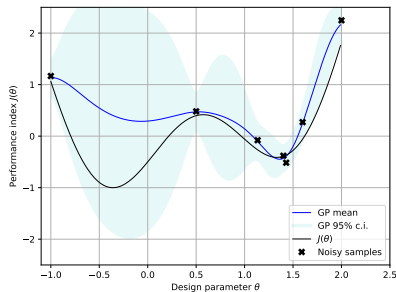
Bayesian Optimization

Overview

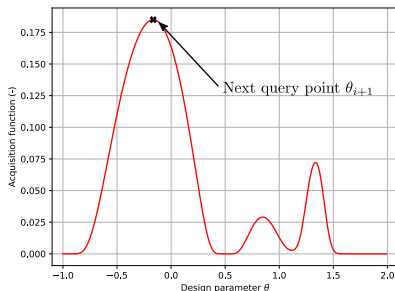
Steps of BO: for $i = 1, 2, \dots, i_{\max}$

- 1 **Execute** experiment with θ_i , measure $J_i = J(\theta_i) + e_i$
- 2 **Update** the GP model $\theta \rightarrow J(\theta)$ with (θ_i, J_i)
- 3 **Construct** acquisition function $A(\theta)$
- 4 **Maximize** $A(\theta)$ to obtain next query point θ_{i+1}

GP at iteration i



$A(\theta)$ at iteration i

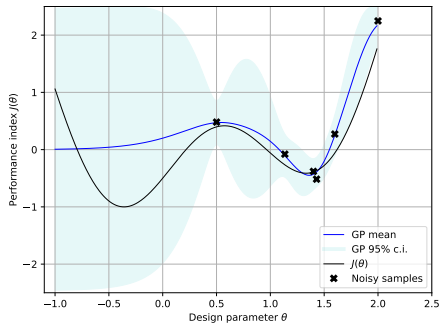


Bayesian Optimization

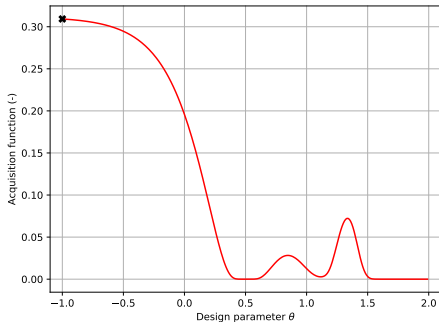
Example

iteration 6

GP fit



$A(\theta) = \text{EI}(\theta)$

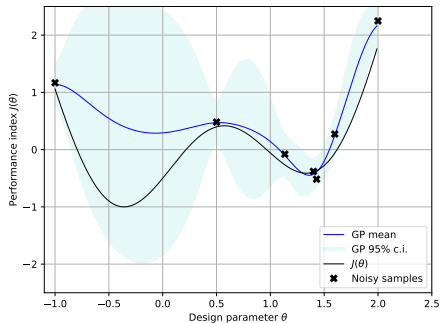


Bayesian Optimization

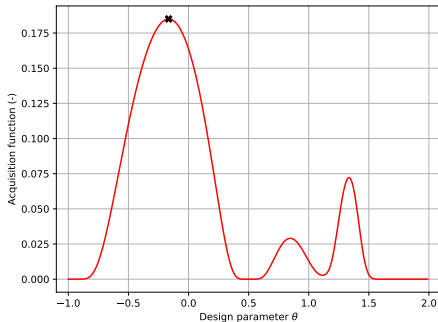
Example

iteration 7

GP fit



$A(\theta) = \text{EI}(\theta)$

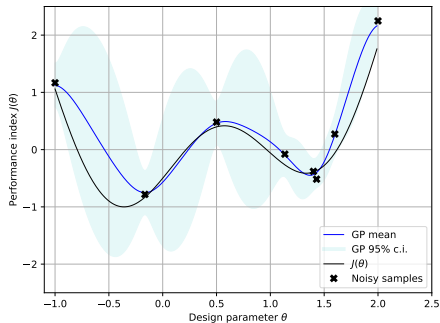


Bayesian Optimization

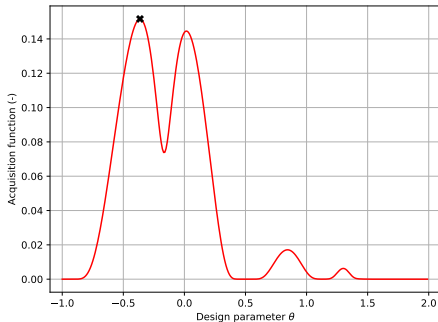
Example

iteration 8

GP fit



$A(\theta) = \text{EI}(\theta)$

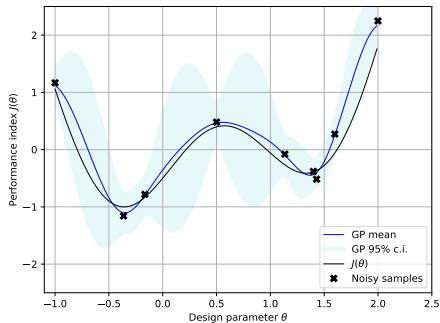


Bayesian Optimization

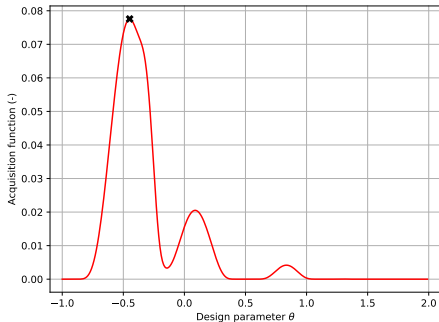
Example

iteration 9

GP fit



$A(\theta) = \text{EI}(\theta)$

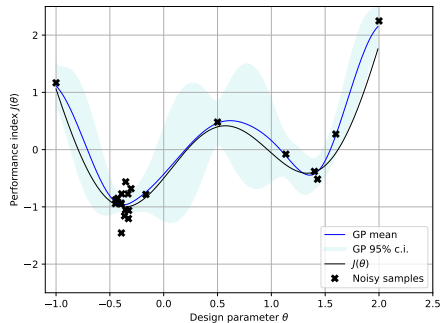


Bayesian Optimization

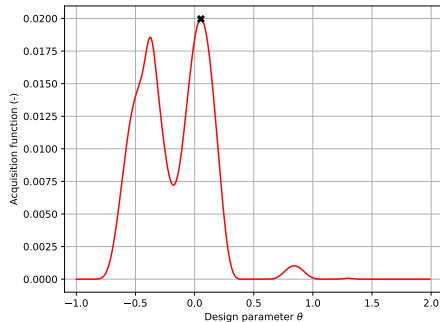
Example

iteration 20

GP fit



$A(\theta) = \text{EI}(\theta)$

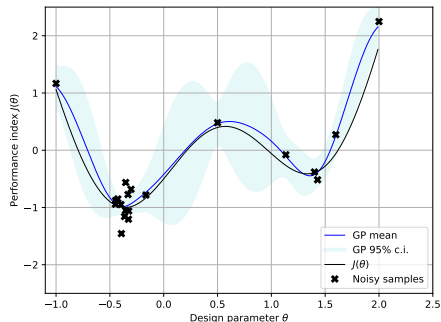


Bayesian Optimization

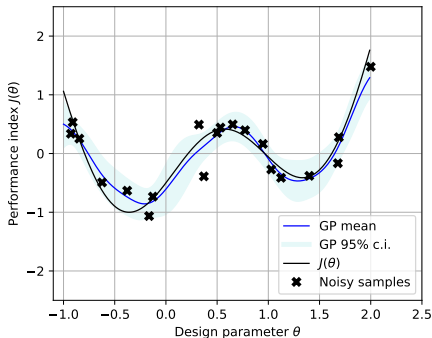
Example

iteration 20

Bayesian Optimization



Random sampling



Thank you.
Questions?