Optimally switching between local and Bayesian optimization

Introduction of Bayesian optimization

Existing Problem

Detail in this paper

Results and conclusion

Introduction

- Materials design and discovery:

Choosing the chemical structure , composition, or processing conditions of a material to meet design criteria.

For example:

We know shear modulus of material compounds belonging to the family of M_2AX phases.

How to find the combinations of atoms to meet this ?

Introduction

- Traditional approach: trial and error.

Choose some material design we think work well

↓

Synthesize and test the material in physical experiments

↓

Use what we learn from these experiments to try next.

Introduction

- Bayesian Optimization for the problem:

  BO considers this problem parameterized by a k-dimensional vector x. That is, suppose that the space of materials designs in space R^k

- BO :

  BO is concerned with the global optimization of a black box function , considered to be expensive to evaluate.

Framework

Algorithm 1 Basic pseudo-code for Bayesian optimization

Place a Gaussian process prior on f

Observe f at n_0 points according to an initial space-filling experimental design. Set $n = n_0$.

while $n \leq N$ **do**

  Update the posterior probability distribution on f using all available data

  Let x_n be a maximizer of the acquisition function over x, where the acquisition function is computed using the current posterior distribution.

  Observe $y_n = f(x_n)$.

  Increment n

end while

Return a solution: either the point evaluated with the largest $f(x)$, or the point with the largest posterior mean.

Gaussian processes

Two main acquisition functions

- Expected Improvement :

    Choose the point with the greatest improvement in expectation on the best value observed so far.

- Predictive Entropy Search:

    Choose the location expected to yield the greatest change in the information content of the GP model.

Stopping Criterion

- 1.Fix the number of iterations;
- 2.Fix a computational budget;
- 3.Probability of improvement on the incumbent solution.

Separating Global and Local Regret

- BO aim to minimize the difference between the function value at the final recommended and the value at the true global.

$$
\begin{aligned}
\text{Regret} &= \mathbb{E}(\hat{y} - y_*) \\
&= \mathbb{E}(\hat{y} - y_i) + \mathbb{E}(y_i - y_*) \\
&= \mathbb{E}(\hat{y} - y_i) + \mathbb{E}(\max(y_i - y_o, 0)) \\
&= R_{\text{local}} + R_{\text{global}} \,,
\end{aligned}
$$

- $\hat{y}$ *denotes final recommended value*;
- y_* *denotes true global value*;
- y_i *denotes the minimum value of* S *and* y_0 *denotes minimum value outside* S.

Separating Global and Local Regret

- R_{local} *represents the difference between our candidate point and the local value.*
- R_{global} *represents the difference between global point and the local value.*

$$
\begin{aligned}
\text{Regret} &= \mathbb{E}(\hat{y} - y_*) \\
&= \mathbb{E}(\hat{y} - y_i) + \mathbb{E}(y_i - y_*) \\
&= \mathbb{E}(\hat{y} - y_i) + \mathbb{E}(\max(y_i - y_o, 0)) \\
&= R_{\text{local}} + R_{\text{global}} \, ,
\end{aligned}
$$

- $if\ y_* = y_i; R_{global} = 0;$
- $if\ y_* = y_o; R_{global} > 0.$

Multiple Acquisition Functions

- Four distinct modes:
  1. Random Initialization;
  2. Bayesian Optimization;
  3. Global Regret Reduction;
  4. Local Optimization.

Global Regret Reduction

- Direct our efforts towards reducing the probability of any other local minima which take lower values than our incumbent solution existing , reduction the global regret. So use a modified form of EI:

$$\alpha_{\text{GRR}} = (\mathbb{E}(y_i) - \mu)\Phi\left(\frac{\mathbb{E}(y_i) - \mu}{\sigma}\right) + \sigma\phi\left(\frac{\mathbb{E}(y_i) - \mu}{\sigma}\right)$$

Local Optimization

- BFGS algorithm for local optimization;

$$y_k = g_{k+1} - g_k, \delta_k = x_{k+1} - x_k$$

$$\delta_k = H_k^{-1} y_k$$

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T \delta_k} - \frac{B_k \delta_k \delta_k^T B_k}{\delta_k^T B_k \delta_k}$$

Local Optimization

- After global regret reduction, we have a both sufficiently high certainty that GP posterior is close to a minimum of the objective;

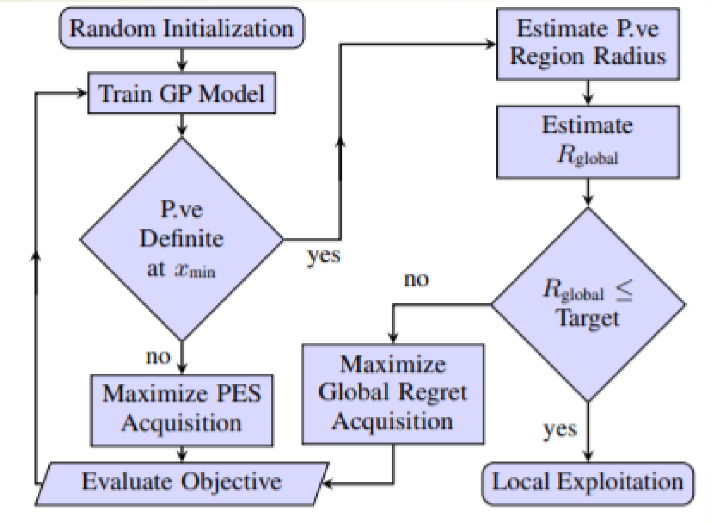
- BFGS algorithm for local optimization;

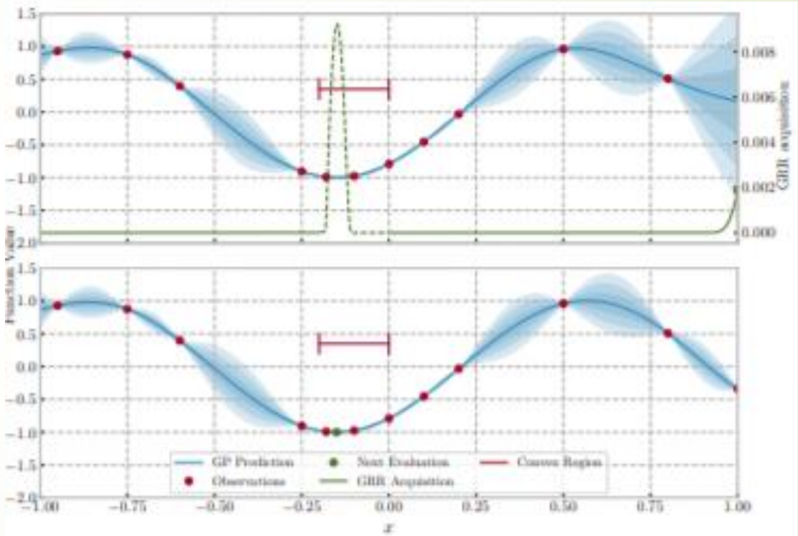
$$f(x) = \frac{1}{2}x^T H x + x^T g + c$$

H is positive definite matrix, $H = CC^T$ $by\ Cholesky\ decomposition\ of\ H$.

$$g(z) = \frac{1}{2}z^T z + z^T R^T g + c$$

Switching Flowchart





Identifying a Convex Region

- Convexity is characterized by Hessian matrix of the objective. We require for all x,

$$x^T H x > 0$$

We need to construct:

1. A method to determine the probability that the Hessian is positive definite at any point;

2. A method to determine, the largest region centered on the current posterior minimum with a required probability of being convex.

Convexity at a point

- Making use of Cholesky decomposition to determine if a matrix is positive definite.

Algorithm 1 Positive Definite Test

Input: location x, tolerance ϵ
$\texttt{G} \leftarrow \texttt{GP\_model}$
$\texttt{Hmean}, \texttt{Hvar} \leftarrow \texttt{G.infer\_Hessian}(x)$
$\texttt{PVEcount} \leftarrow 0$
$n \leftarrow \frac{1}{\epsilon} - 2$
for $i = 1...n$ **do**
  $h \leftarrow \texttt{draw\_Gaussian}(\texttt{Hmean}, \texttt{Hvar})$
  $h^* \leftarrow \texttt{remove\_boundary\_elements}(h)$
  if $\texttt{Cholesky}(h^*) \neq$ **FAIL then**
    $\texttt{PVEcount} \leftarrow \texttt{PVEcount} + 1$
  end if
end for
$p \leftarrow \frac{PVEcount+1}{n+2}$
Return $p \geq 1 - \epsilon$

Radius of a convex region

- Find the hypersphere center at minimum posterior with the greatest possible radius R_{max}.

$$R_u(u) = \underset{PD(\hat{x}+ru)=1}{\arg\max}\ r$$

- Binary search

Radius of a convex region

Algorithm 2 Positive Definite Sphere Radius

Input: center $x_{\min}$, number of directions, n_u tolerance ϵ
$u \leftarrow$ `random_unit_vector`
$x_{\text{edge}} \leftarrow$ `dist_to_domain_boundary`
$\hat{R} \leftarrow \|x_{\min} - x_{\text{edge}}\|$
for $i = 1 \ldots n_u$ **do**
  if $D(x + \hat{R}u) = 0$ **then**
    $\hat{R} \leftarrow$ `binarysearch`$(u, \hat{R})$
  end if
  $u \leftarrow$ `random_unit_vector`
end for
Return $\hat{R}$

Global regret

- Definition

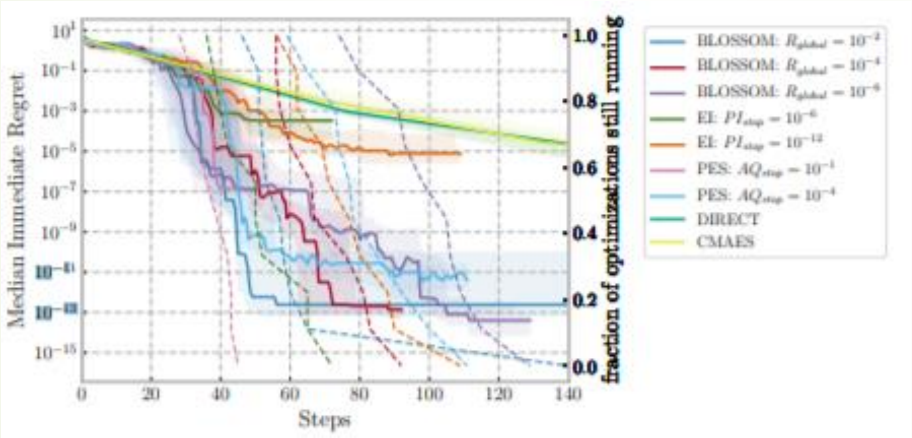
$$R_{global} = \iint_{y_i \times y_o} \max(y_i - y_o, 0)\, p(y_i) p(y_o)\, dy_i\, dy_o$$

- Consider $y_i \sim N(\mu_i, \delta_i^2)$.

$$\begin{aligned}R_{\text{global}} &= \int_{y_o} \int_{y_i = y_o}^{+\infty} \max(y_i - y_o, 0) p(y_i)\, dy_i\, p(y_o) dy_o \\ &= \int_{y_o} \left[(\mu_i - y_o) \Phi\left(\frac{\mu_i - y_o}{\sigma_i}\right) \right. \\ &\qquad \left. + \sigma_i \phi\left(\frac{\mu_i - y_o}{\sigma_i}\right) \right] p(y_o) dy_o\end{aligned}$$

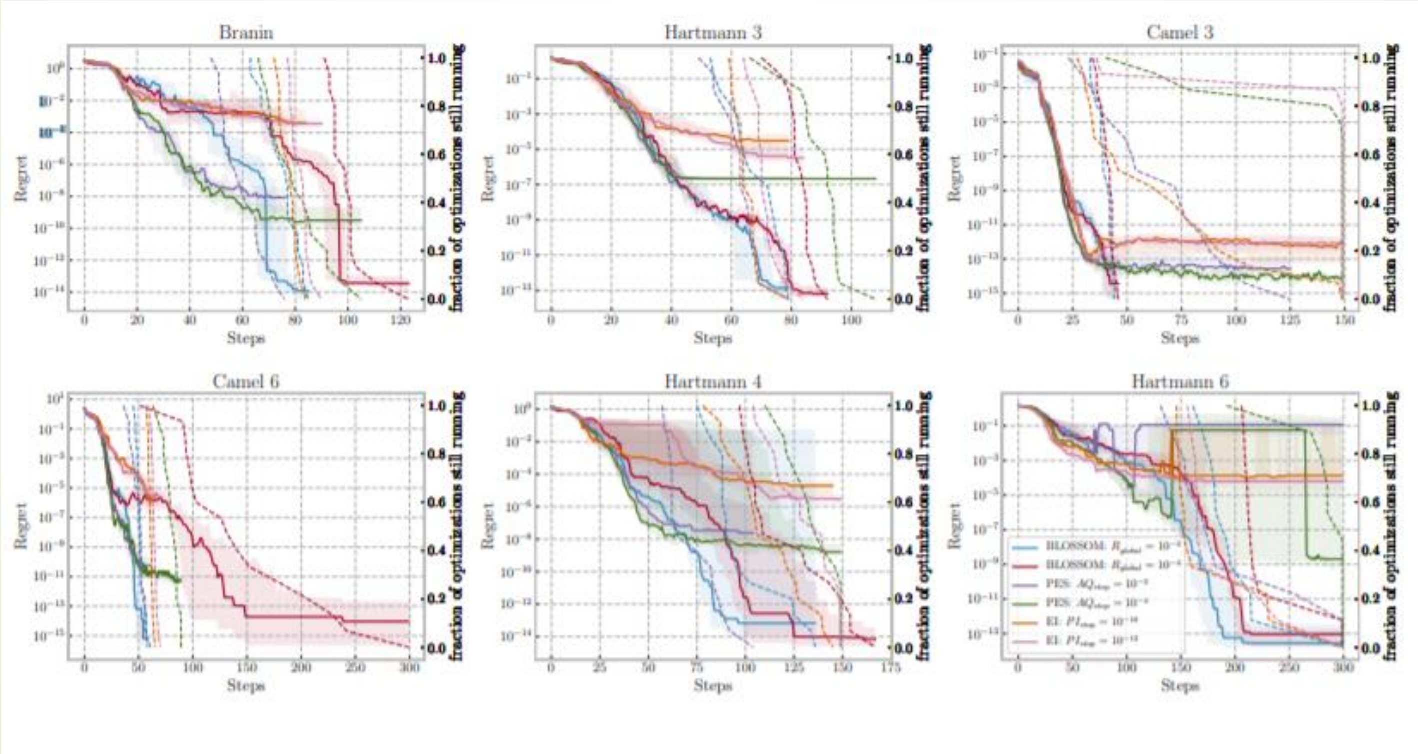
Results

- In-Model Objectives:



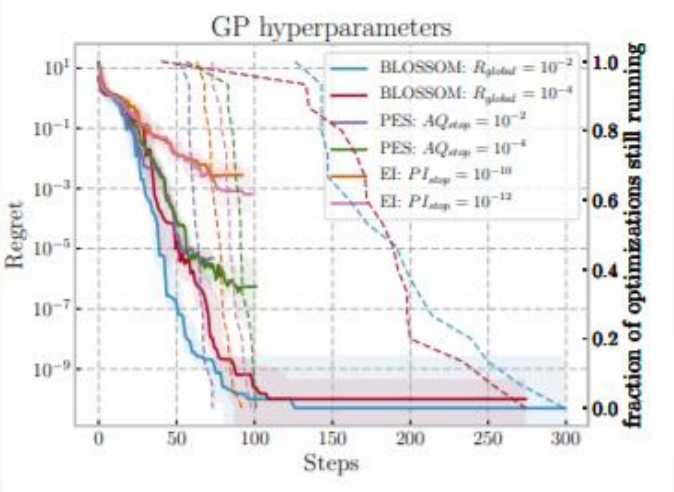
Results

- Common Benchmark Functions



Results

- GP Hyperparameter Optimization



Conclusion

- 1. Avoid the poor local convergence of Gaussian Process methods.
- 2. Be able to halt optimization once a specified value of global regret has been achieved.