Energetic Variational Gaussian Process Regression for Computer Experiments

Lulu Kang, Yuanxing Cheng, Yiwei Wang, Chun Liu
Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616

Abstract

Gaussian process (GP) regression model is a popular surrogate modeling approach for computer experiments. It provides both accurate prediction and inference for the computer experimental data. Both frequentist and Bayesian framework can be used to perform the estimation and inference tasks for GP. In this chapter, we build the GP model via variational inference, or more specifically, the newly proposed energetic variational inference by [1]. Following the GP model assumption, we first derive the posterior distributions of the parameters for GP model with non-zero mean functions. The energetic variational inference method is used to generate samples of the posterior distributions. More importantly, it can bridge the Bayesian sampling and optimization and provides a much more computationally efficient solution. With a normal prior on the mean component of the GP model, shrinkage estimation is also applied to the parameters to achieve variable selection of the mean function. The performance of the proposed GP model is illustrated via several examples.

Keyword: Gaussian process regression, Energetic variational approach, Bayesian inference

Gaussian Process Regression: Bayesian Approach

- Denote $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$ as the data from computer experiments, where $\boldsymbol{x}_i \in \Omega \subseteq \mathbb{R}^d$ are the ith experimental input values and $y_i \in \mathbb{R}$ is the corresponding output. Here we consider the case of univariate response.
- Gaussian process regression is built on the following model assumption of the response,

$$y_i = \mathbf{g}(\boldsymbol{x}_i)^\top \boldsymbol{\beta} + Z(\boldsymbol{x}_i) + \epsilon_i, \quad i = 1, \dots, n, \tag{1}$$

  - $\mathbf{g}(\boldsymbol{x})$ is a p-dim vector of user-specified basis functions and $\boldsymbol{\beta}$ are the linear coefficients combining the basis functions.
  - The noise ϵ_i's are IID following $\mathcal{N}(0, \sigma^2)$ and also independent of the other stochastic ingredients of (1).
  - The Gaussian process prior means that the stochastic function $Z(\boldsymbol{x})$ is a Gaussian process $Z(\cdot) \sim GP(0, \tau^2 K)$, which indicates $\mathbb{E}[Z(\boldsymbol{x})] = 0$ and

$$\mathrm{Cov}\left[Z(\boldsymbol{x}_1), Z(\boldsymbol{x}_2)\right] = \tau^2 K(\boldsymbol{x}_1, \boldsymbol{x}_2).$$

  - Here τ^2 is a constant variance and $K(\cdot,\cdot;\boldsymbol{\omega}) : \Omega \times \Omega \mapsto \mathbb{R}_+$ is the correlation function of the stochastic process with hyperparameters $\boldsymbol{\omega}$.
  - For any Gaussian process, the correlation function should be a symmetric positive definite kernel function. For instance, if we use the Gaussian kernel function, then

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{\omega}) = \exp\left\{-\sum_{j=1}^d \omega_j (x_{1j} - x_{2j})^2\right\},$$

  with $\boldsymbol{\omega} \in \mathbb{R}^d$ and $\boldsymbol{\omega} \geq 0$.
- In terms of response y, it follows a Gaussian Process with the following mean and covariance,

$$\mathbb{E}[y(\boldsymbol{x})] = \mathbf{g}(\boldsymbol{x})^\top \boldsymbol{\beta}, \quad \forall \boldsymbol{x} \in \Omega \tag{2}$$

$$\mathrm{Cov}\left[y(\boldsymbol{x}_1), y(\boldsymbol{x}_2)\right] = \tau^2 K(\boldsymbol{x}_1, \boldsymbol{x}_2; \omega) + \sigma^2 \delta(\boldsymbol{x}_1, \boldsymbol{x}_2), \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega, \tag{3}$$

$$= \tau^2 \left[K(\boldsymbol{x}_1, \boldsymbol{x}_2; \omega) + \eta \delta(\boldsymbol{x}_1, \boldsymbol{x}_2)\right], \tag{4}$$

  where $\delta(\boldsymbol{x}_1, \boldsymbol{x}_2) = 1$ if $\boldsymbol{x}_1 = \boldsymbol{x}_2$ and 0 otherwise, and $\eta = \sigma^2/\tau^2$.
- η is interpreted as the noise-to-signal ratio. For deterministic computer experiments, the noise component is not part of the model, i.e., $\sigma^2 = 0$. However, a nugget effect, which is a small η value, is usually included in the covariance function to avoid singularity of the covariance matrix. The unknown parameter values are $\theta = (\boldsymbol{\beta}, \boldsymbol{\omega}, \tau^2, \eta)$.

Prior and Posterior Distribution

We assume the following prior distributions for the parameters,

$$\boldsymbol{\beta} \sim \mathcal{MVN}_p(\mathbf{0}, \nu^2 \boldsymbol{R}) \tag{5}$$

$$\omega_i \overset{\text{i.i.d.}}{\sim} \mathrm{Gamma}(a_\omega, b_\omega), \text{ for } i = 1, \dots, d \tag{6}$$

$$\tau^2 \sim \text{Inverse-}\chi^2(df_{\tau^2}), \tag{7}$$

$$\eta \sim \mathrm{Gamma}(a_\eta, b_\eta). \tag{8}$$

In practice, using non-informative prior $p(\boldsymbol{\beta}) \propto 1$ also gives promising result.

- Based on the data, the sampling distribution is

$$\boldsymbol{y}_n | \boldsymbol{\theta} \sim \mathcal{MVN}_n\left(\boldsymbol{G}\boldsymbol{\beta}, \tau^2(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)\right), \tag{9}$$

  where $\boldsymbol{y}_n$ is the vector of y_i's and $\boldsymbol{G}$ is a matrix of rows and each row is $\mathbf{g}(\boldsymbol{x}_i)^\top$. The matrix $\boldsymbol{K}_n$ is the $n \times n$ kernel matrix and $K_n[i,j] = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and is symmetric and positive definite, and $\boldsymbol{I}_n$ is an $n \times n$ identity matrix.
- The pdf of $\boldsymbol{y}_n | \boldsymbol{\theta}$ is

$$p(\boldsymbol{y}_n|\boldsymbol{\theta}) \propto (\tau^2)^{-\frac{n}{2}} \det(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1/2} \exp\left(-\frac{1}{2\tau^2}(\boldsymbol{y}_n - \boldsymbol{G}\boldsymbol{\beta})^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} (\boldsymbol{y}_n - \boldsymbol{G}\boldsymbol{\beta})\right). \tag{10}$$

- The joint posterior distribution of all parameters is

$$p(\boldsymbol{\theta}|\boldsymbol{y}_n) \propto p(\boldsymbol{\theta}) p(\boldsymbol{y}_n|\boldsymbol{\theta}) \tag{11}$$

$$\propto p(\boldsymbol{\beta}) \left(\prod_{j=1}^d p(\omega_i)\right) p(\tau^2) p(\eta) p(\boldsymbol{y}_n|\boldsymbol{\theta}). \tag{12}$$

- Due to the conditional conjugacy,

$$\boldsymbol{\beta} | \boldsymbol{y}_n, \boldsymbol{\omega}, \tau^2, \eta \sim \mathcal{MVN}_d\left(\hat{\boldsymbol{\beta}}_n, \Sigma_{\boldsymbol{\beta}|n}\right),$$

  where

$$\Sigma_{\boldsymbol{\beta}|n} = \left[\frac{1}{\tau^2} \boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G} + \frac{1}{\nu^2} \boldsymbol{R}^{-1}\right]^{-1}$$

$$\hat{\boldsymbol{\beta}}_n = \Sigma_{\boldsymbol{\beta}|n} \frac{(\boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{y}_n)}{\tau^2}$$

- if with non-informative prior for $\boldsymbol{\beta}$, we have

$$\Sigma_{\boldsymbol{\beta}|n} = \tau^2 \left[\boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G}\right]^{-1}$$

$$\hat{\boldsymbol{\beta}}_n = \left[\boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G}\right]^{-1} \left[\boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1}\right] \boldsymbol{y}_n$$

Estimate τ

$$\begin{aligned} p(\boldsymbol{\omega}, \tau^2, \eta | \boldsymbol{y}_n) &\propto \int p(\boldsymbol{\beta}) p(\boldsymbol{y}_n|\boldsymbol{\theta}) p(\boldsymbol{\omega}) p(\tau^2) p(\eta) \, \mathrm{d}\boldsymbol{\beta} \\ &\propto \det(\Sigma_{\boldsymbol{\beta}|n})^{1/2} \exp\left[-\frac{1}{2} \hat{\boldsymbol{\beta}}_n^\top \Sigma_{\boldsymbol{\beta}|n}^{-1} \hat{\boldsymbol{\beta}}_n - \frac{1}{2\tau^2} \boldsymbol{y}_n^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{y}_n\right] (\tau^2)^{-n/2} \det(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1/2} p(\tau^2) p(\boldsymbol{\omega}) p(\eta) \end{aligned} \tag{13}$$

Define

$$\begin{aligned} s_n^2 &= \tau^{-2} \hat{\boldsymbol{\beta}}_n^\top \Sigma_{\boldsymbol{\beta}|n}^{-1} \hat{\boldsymbol{\beta}}_n + \boldsymbol{y}_n^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{y}_n \\ &= \boldsymbol{y}_n^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \left[\boldsymbol{G}\left(\boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G}\right)^{-1} \boldsymbol{G}^\top + (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)\right] (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{y}_n. \end{aligned}$$

Then $p(\boldsymbol{\omega}, \tau^2, \eta | \boldsymbol{y}_n)$ is

$$p(\boldsymbol{\omega}, \tau^2, \eta | \boldsymbol{y}_n) \propto (\tau^2)^{-\left[\frac{1}{2}(df_{\tau^2} + n - p) + 1\right]} \exp\left(-\frac{1 + s_n^2}{2\tau^2}\right) \det(\boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G})^{-1/2} \det(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1/2} p(\boldsymbol{\omega}) p(\eta).$$

Due to the conditional conjugacy, we can see the conditional posterior distribution of $\tau^2 | \boldsymbol{\omega}, \eta, \boldsymbol{y}_n$ is $\tau^2 | \boldsymbol{\omega}, \eta, \boldsymbol{y}_n \sim \text{Scaled Inverse-}\chi^2(df_{\tau^2} + n - p, \hat{\tau}^2)$ where $\hat{\tau}^2 = \left(1 + s_n^2\right) / (df_{\tau^2} + n - p)$.

Posterior Distribution

Using the informative prior $\boldsymbol{\beta} \sim \mathcal{MVN}_p(\mathbf{0}, \nu^2 \boldsymbol{R})$, we obtain the posterior distribution (14) in which $\boldsymbol{\beta}$ has been integrated out, and thus it saves the computation involved in posterior sampling stage.

$$\begin{aligned} p(\boldsymbol{\omega}, \tau^2, \eta | \boldsymbol{y}_n) \propto{} & \det(\Sigma_{\boldsymbol{\beta}|n})^{1/2} \exp\left[-\frac{1}{2} \hat{\boldsymbol{\beta}}_n^\top \Sigma_{\boldsymbol{\beta}|n}^{-1} \hat{\boldsymbol{\beta}}_n + \frac{1}{2\tau^2} \boldsymbol{y}_n^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{y}_n\right] \det(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1/2} \\ & \times (\tau^2)^{-n/2} p(\tau^2) p(\boldsymbol{\omega}) p(\eta). \end{aligned} \tag{14}$$

Using the non-informative prior $p(\boldsymbol{\beta}) \propto 1$, we obtain the posterior distribution (15) in which both $\boldsymbol{\beta}$ and τ^2 are integrated out.

$$p(\boldsymbol{\omega}, \eta | \boldsymbol{y}_n) \propto (\hat{\tau}^2)^{-\frac{1}{2}(df_{\tau^2} + n - p)} \det(\boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G})^{-1/2} \det(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1/2} p(\boldsymbol{\omega}) p(\eta). \tag{15}$$

Given the parameters $(\boldsymbol{\omega}, \tau^2, \eta)$, the posterior predictive distribution of $y(\boldsymbol{x})$ at any query point $\boldsymbol{x}$ is the following normal distribution.

$$y(\boldsymbol{x}) | \boldsymbol{y}_n, \boldsymbol{\omega}, \tau^2, \eta \sim \mathcal{N}(\hat{\mu}(\boldsymbol{x}), \sigma_n^2(\boldsymbol{x})), \tag{16}$$

where

$$\hat{\mu}(\boldsymbol{x}) = \mathbf{g}(\boldsymbol{x})^\top \hat{\boldsymbol{\beta}}_n + K(\boldsymbol{x}, \mathcal{X}_n)(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} (\boldsymbol{y}_n - \boldsymbol{G}\hat{\boldsymbol{\beta}}_n), \tag{17}$$

$$\sigma_n^2(\boldsymbol{x}) = \tau^2 \left\{1 - K(\boldsymbol{x}, \mathcal{X}_n)(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} K(\mathcal{X}_n, \boldsymbol{x}) + \boldsymbol{c}(\boldsymbol{x})^\top \left[\boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G}\right]^{-1} \boldsymbol{c}(\boldsymbol{x})\right\}, \tag{18}$$

$$\boldsymbol{c}(\boldsymbol{x}) = \mathbf{g}(\boldsymbol{x}) - \boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} K(\mathcal{X}_n, \boldsymbol{x}),$$

$$K(\boldsymbol{x}, \mathcal{X}_n) = K(\mathcal{X}_n, \boldsymbol{x})^\top = [K(\boldsymbol{x}, \boldsymbol{x}_1), \dots, K(\boldsymbol{x}, \boldsymbol{x}_n)]$$

For the sake of computational stability, we use an alternative formula for $\hat{\mu}(\boldsymbol{x})$ by first plugging in $\hat{\beta}_n$ into $\hat{\mu}(\boldsymbol{x})$.

$$\hat{\mu}(\boldsymbol{x}) = \left(\left(\mathbf{g}(\boldsymbol{x})^\top - K(\boldsymbol{x}, \mathcal{X}_n)(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G}\right) \left(\boldsymbol{G}^\top (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G}\right)^{-1} \boldsymbol{G}^\top + K(\boldsymbol{x}, \mathcal{X}_n)\right) (\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{y}_n$$

We can use cholesky decomposition on $\boldsymbol{K}_n + \eta \boldsymbol{I}_n$ where we define $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{K}_n + \eta \boldsymbol{I}_n$. Then we can obtain higher accuracy in calculating $(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{G}$ by solving $\boldsymbol{\gamma}_1$ from the following equations.

$$\begin{aligned} \boldsymbol{L}^\top \boldsymbol{\alpha} &= \eta \boldsymbol{G} \\ \boldsymbol{L} \boldsymbol{\gamma}_1 &= \boldsymbol{\alpha} / \eta \end{aligned} \tag{19}$$

Similar method is used on $(\boldsymbol{K}_n + \eta \boldsymbol{I}_n)^{-1} \boldsymbol{y}_n = \boldsymbol{\gamma}_2$. Then the final formula reads,

$$\hat{\mu}(\boldsymbol{x}) = \left(\left(1 - K(\boldsymbol{x}, \mathcal{X}_n) \boldsymbol{\gamma}_1\right) \left(\boldsymbol{G}^\top \boldsymbol{\gamma}_1\right)^{-1} \boldsymbol{G}^\top + K(\boldsymbol{x}, \mathcal{X}_n)\right) \boldsymbol{\gamma}_2$$

Sampling procedure and EVI

The sampling procedure is

1. Generate posterior samples from $p(\boldsymbol{\omega}, \tau^2, \eta | \boldsymbol{y}_n)$ or $p(\boldsymbol{\omega}, \eta | \boldsymbol{y}_n)$, depending on which prior is used for $\boldsymbol{\beta}$.
2. Based on the posterior samples of $(\boldsymbol{\omega}, \tau^2, \eta)$, generate the conditional posterior distribution for $\boldsymbol{\beta}$. If non-informative prior is used, generate the conditional distribution of $\tau^2 | \boldsymbol{\omega}, \eta$.
3. Based on the posterior samples of $(\boldsymbol{\omega}, \tau^2, \eta)$ (or $(\boldsymbol{\omega}, \eta)$ for non-informative prior), generate the posterior predictive distribution of any query point $y(\boldsymbol{x}) | \boldsymbol{y}_n, \boldsymbol{\omega}, \tau^2, \eta$. For non-informative prior, the posterior predictive distribution $y(\boldsymbol{x}) | \boldsymbol{y}_n, \boldsymbol{\omega}, \eta$ is the same except τ^2 is replaced by $\hat{\tau}^2$.

Energetic variational inference method is introduced here. Let $\boldsymbol{\phi}_t$ be the dynamic flow map $\boldsymbol{\phi}_t : \mathbb{R}^d \to \mathbb{R}^d$ that continuously transforms the d-dimensional distribution from an initial distribution toward the target one and we require the map $\boldsymbol{\phi}_t$ to be smooth and one-to-one. Then define functional $\mathcal{F}(\boldsymbol{\phi}_t)$ to be (14) or (15). The EVI start with energy dissipation law:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}(\boldsymbol{\phi}_t) = -\triangle(\boldsymbol{\phi}_t, \dot{\boldsymbol{\phi}}_t), \tag{20}$$

where $\triangle(\boldsymbol{\phi}_t, \dot{\boldsymbol{\phi}}_t)$ is a user-specified functional representing the rate of energy dissipation, and $\dot{\boldsymbol{\phi}}_t$ is the derivative of $\boldsymbol{\phi}_t$ with time t. The dissipation functional should satisfy $\triangle(\boldsymbol{\phi}_t, \dot{\boldsymbol{\phi}}_t) \geq 0$ so that $\mathcal{F}(\boldsymbol{\phi}_t)$ decreases with time. As discussed in [?], there are many ways to specify $\triangle(\boldsymbol{\phi}_t, \dot{\boldsymbol{\phi}}_t)$ and the simplest among them is a quadratic functional in terms of $\dot{\boldsymbol{\phi}}_t$,

$$\triangle(\boldsymbol{\phi}_t, \dot{\boldsymbol{\phi}}_t) = \int_{\Omega_t} \rho_{[\boldsymbol{\phi}_t]} \|\dot{\boldsymbol{\phi}}_t\|_2^2 \, \mathrm{d}\boldsymbol{x},$$

where $\rho_{[\boldsymbol{\phi}_t]}$ denotes the pdf of the current distribution which is the initial distribution transformed by $\boldsymbol{\phi}_t$, Ω_t is the current support, and $\|\boldsymbol{a}\|_2 = \boldsymbol{a}^\top \boldsymbol{a}$ for $\forall \boldsymbol{a} \in \mathbb{R}^d$. Then the EVI reads:

$$\frac{\delta \frac{1}{2} \triangle}{\delta \dot{\boldsymbol{\phi}}_t} = -\frac{\delta \mathcal{F}}{\delta \boldsymbol{\phi}_t} \implies \rho_{[\boldsymbol{\phi}_t]} \dot{\boldsymbol{\phi}}_t = -\frac{\delta \mathcal{F}}{\delta \boldsymbol{\phi}_t}.$$

EVI(contin.)

We then do discretization on space parameter, called it "approximation-then-variation" approach and obtain the particle version of the PDE.

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}_h(\{\boldsymbol{x}_i(t)\}_{i=1}^N) = -\triangle_h(\{\boldsymbol{x}_i(t)\}_{i=1}^N, \{\dot{\boldsymbol{x}}_i(t)\}_{i=1}^N). \tag{21}$$

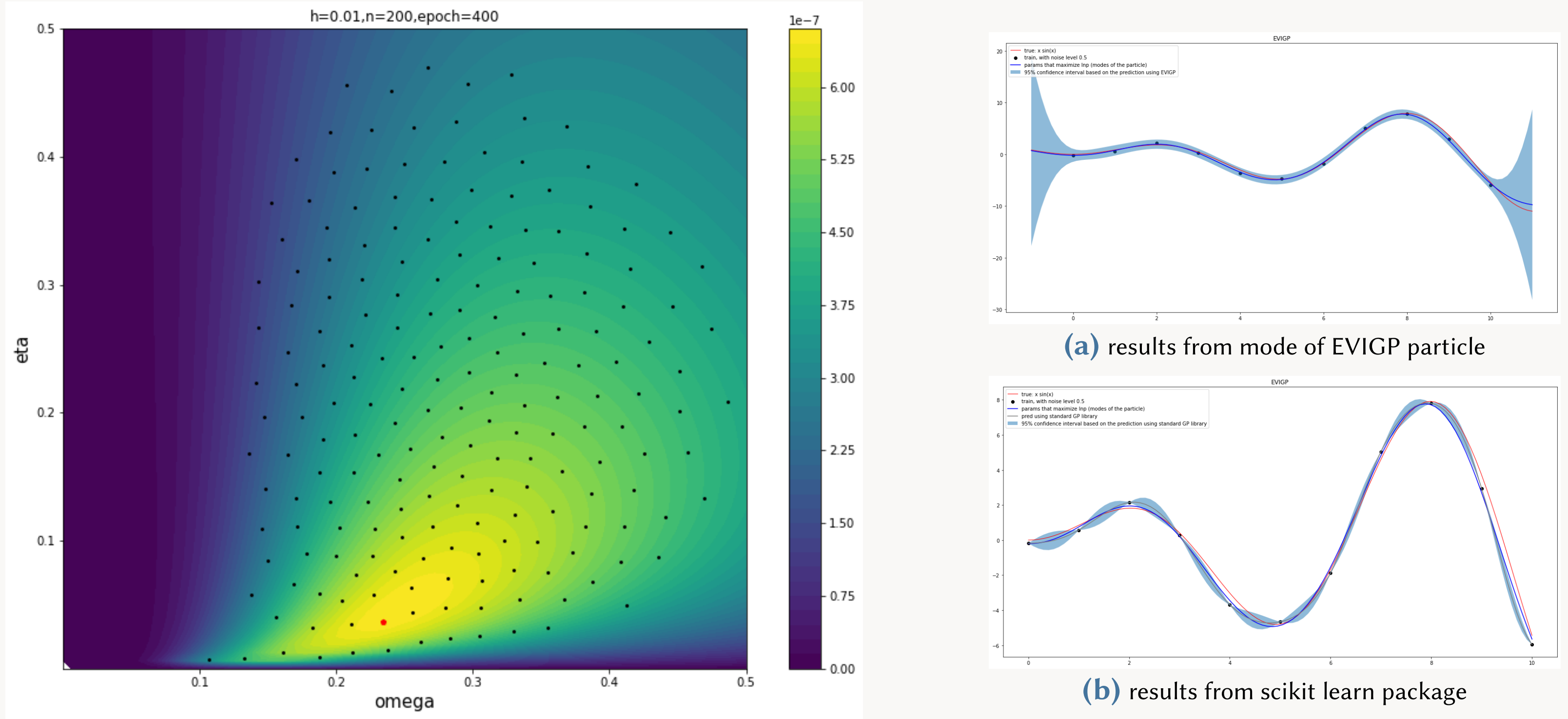
Here $\{\boldsymbol{x}(t)\}_{i=1}^N$ is the locations of N particles at time t and $\dot{\boldsymbol{x}}_i(t)$ is the derivative of $\boldsymbol{x}_i$ with t, and thus is the velocity of the ith particle as it moves toward the target distribution. The subscript h of $\mathcal{F}$ and $\triangle$ denotes the bandwidth parameter of the kernel function used in the kernelization operation. Applying the variational steps to (21), we obtain the dynamics of decreasing $\mathcal{F}$ at the particle level,

$$\frac{\delta \frac{1}{2} \triangle_h}{\delta \dot{\boldsymbol{x}}_i(t)} = -\frac{\delta \mathcal{F}_h}{\delta \boldsymbol{x}_i}, \quad \text{for } i = 1, \dots, N. \tag{22}$$

Backward-Euler is then applied to solve this ODE, which gives particles representing the samples $\boldsymbol{\theta}$s from the posterior $\log p(\omega, \tau^2, \eta \mid \boldsymbol{y}_n)$ or $\log p(\omega, \eta \mid \boldsymbol{y}_n)$.

In accordance with the frequentist approach, we can then pick the mode from those samples that maximize the posterior probability and make predictions according to (17).

2D toy example: $y = x\sin(x)$



(a) results from mode of EVIGP particle

(b) results from scikit learn package

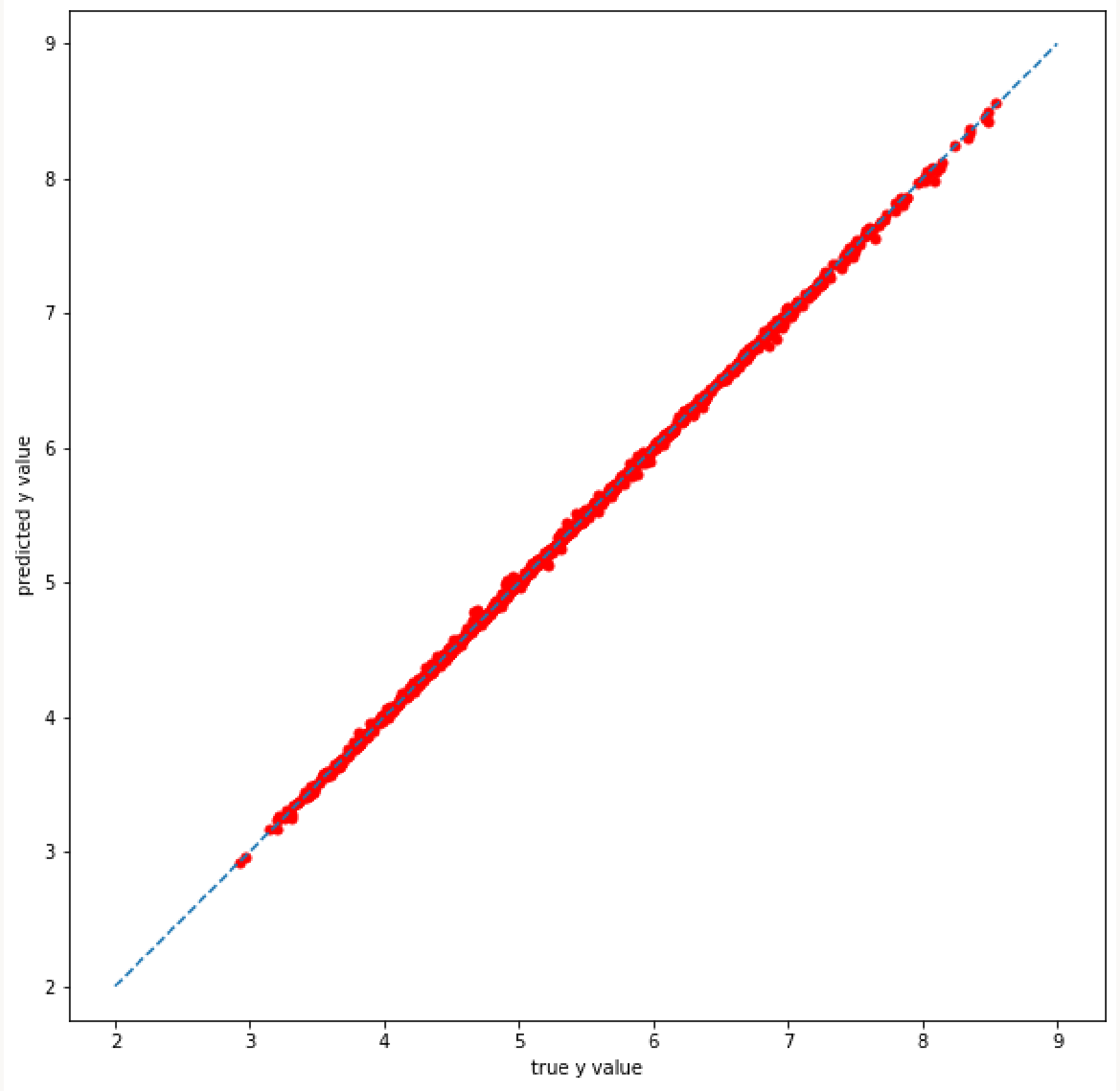
high dimension function

We compare the regression results for borehole function and OTL-circuit function to those from R package GPfit and mlegp, corresponding to constant mean and linear mean in the GP model, respectively. Here the metric standardized root mean squared percentage error (rmspe) is defined as

$$\sqrt{\frac{1}{\mathrm{N_{test}}} \sum \left(\mathrm{y_{pred}}/\mathrm{y_{true}} - 1\right)^2} \Big/ \mathrm{std}(\mathrm{y_{true}})$$

function name	borehole	OTC-Circuit
EVIGP (constant mean)	0.000267	0.00386
gpfit	0.0224	0.0131
EVIGP (linear mean)	0.000235	0.00263
mlegp	0.0118	0.00884

And an example of QQ-plot



Reference

Wang, Y., Chen, J., Liu, C., and Kang, L.
Particle-based energetic variational inference.
Statistics and Computing (04 2020).

Yuanxing Cheng
ycheng46@hawk.iit.edu
Illinoise Institute of Technology