Bayesian Optimization on Manifolds via Graph Gaussian Processes

Hwanwoo Kim

University of Chicago

Joint work with D.Sanz-Alonso, R.Yang

2023 SIAM OP

June 2, 2023

Problem of Interest

We want to solve

$$\max_{x \in \mathcal{M}} f(x)$$

where

- an evaluation of the objective is computationally expensive.
- an evaluation can be corrupted by random noise.
- an explicit form of the objective is unknown.
- $\mathcal{M}$ is smooth, connected, compact, and boundary-free.
- geometric information of the manifold such as chart, tangent space, Riemannian gradient, retraction map are unknown.
- a point cloud $\{x_1, \cdots, x_N\} \subset \mathcal{M} \subset \mathbb{R}^m, m \geq 2$ is available.

Problem of Interest

We want to solve

$$\max_{x \in \mathcal{M}} f(x)$$

where

- an evaluation of the objective is computationally expensive.
- an evaluation can be corrupted by random noise.
- an explicit form of the objective is unknown.
- $\mathcal{M}$ is smooth, connected, compact, and boundary-free.
- geometric information of the manifold such as chart, tangent space, Riemannian gradient, retraction map are unknown.
- a point cloud $\{x_1, \cdots, x_N\} \subset \mathcal{M} \subset \mathbb{R}^m, m \geq 2$ is available.

Bayesian Optimization I

Iterate the following steps:

1. Given a set of data $\mathcal{D} = \{(x_\ell, f(x_\ell) + \epsilon)\}$ composed of noisy function evaluations with evaluation locations, build a surrogate for the objective.
2. Choose an additional query location for the function evaluation by maximizing an acquisition function.
   - Exploration: a point that the surrogate is the most uncertain of.
   - Exploitation: a point that attains a large surrogate function value.
   - Acquisition Function: a function that ensures the maximum balances exploration and exploitation.
3. Repeat the previous two steps for the total L number of iterations.
4. Find the optimum of the surrogate among the observations.

Bayesian Optimization II: Illustration

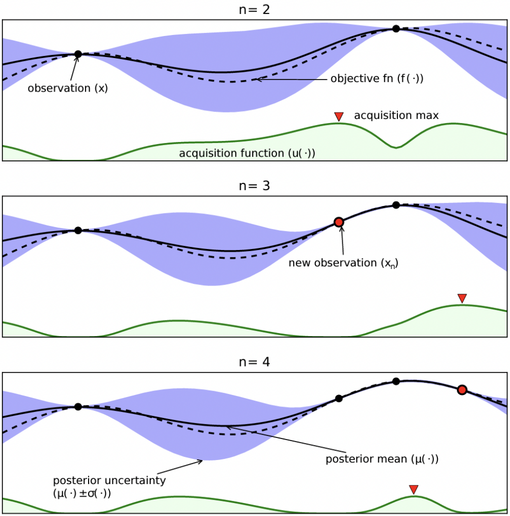

Figure: Illustration of the Bayesian optimization from [Shahriari et al., 2015]

Bayesian Optimization: GP Surrogate

1. Given a set of data composed of noisy function evaluations with evaluation locations, build a surrogate for the objective.
   - put a Gaussian process prior to the objective function.
   - derive a posterior distribution of the objective function conditioning on all function evaluations obtained so far.
   - use the posterior mean $\mu(x|\mathcal{D})$ as a surrogate for the objective function.
   - Quantify uncertainties in the surrogate through the posterior variance $\sigma(x|\mathcal{D})$.

Bayesian Optimization: Acquisition Function

2. Choose an additional query location for the function evaluation by maximizing an acquisition function.
   - Exploration: a point that the surrogate is the most uncertain of.
   - Exploitation: a point that attains a large surrogate function value.
   - Acquisition function: a function that ensures the maximum balances exploration and exploitation.

$\Rightarrow$ Upper Confidence Bound acquisition function:

$$\mu(x|\mathcal{D}) + B_\ell \cdot \sigma(x|\mathcal{D}),$$

where B_ℓ is a parameter controlling the level of exploration.

Reminder: Problem of Interest

We want to solve

$$\max_{x \in \mathcal{M}} f(x)$$

where

- an evaluation of the objective is computationally expensive.
- an evaluation can be corrupted by random noise.
- an explicit form of the objective is unknown (possibly nonconvex, nonsmooth).
- $\mathcal{M}$ is smooth, connected, compact, and boundary-free.
- geometric information of the manifold such as chart, tangent space, Riemannian gradient, retraction map are unknown.
- a point cloud $\{x_1, \cdots, x_N\} \subset \mathcal{M} \subset \mathbb{R}^m, m \geq 2$ is available.

Matérn-type Gaussian process on a manifold

If we know the closed manifold, define a Gaussian process on $\mathcal{M}$ given by

$$N(0, (\tau I - \Delta_{\mathcal{M}})^{-s})$$

where $\Delta_{\mathcal{M}}$ is the Laplace-Beltrami operator on $\mathcal{M}$, τ controls the inverse length scale and s is a sample smoothness parameter. By Karhunen-Loeve expansion, samples are of the form:

$$u^{\mathrm{Ma}} = \sum_{i=1}^{\infty} (\tau + \lambda_i)^{-\frac{s}{2}} \zeta_i \phi_i,$$

where $(\lambda_i, \phi_i)_{i=1}^{\infty}$ are the eigenpairs of $-\Delta_{\mathcal{M}}$ and ζ_i is a standard Gaussian random variable [Sanz-Alonso and Yang, 2022, Borovitskiy et al., 2020].

Reminder: Problem of Interest

We want to solve

$$\max_{x \in \mathcal{M}} f(x)$$

where

- an evaluation of the objective is computationally expensive.
- an evaluation can be corrupted by random noise.
- an explicit form of the objective is unknown (possibly nonconvex, nonsmooth).
- $\mathcal{M}$ is smooth, connected, compact, and boundary-free.
- geometric information of the manifold such as chart, tangent space, Riemannian gradient, retraction map are unknown.
- a point cloud $\{x_1, \cdots, x_N\} \subset \mathcal{M} \subset \mathbb{R}^m, m \geq 2$ is available.

Graph Laplacian

Given a point cloud $\{x_1, \cdots, x_N\}$ on a manifold, the unnormalized graph Laplacian is given by

$$\Delta_N = D - W,$$

where W is a similarity matrix and

$$D = \mathsf{diag}(D_{11}, \cdots, D_{NN}), \quad D_{ii} = \sum_{j=1}^{N} W_{ij}.$$

Graph-Gaussian process on a point cloud

Approximate the Matérn-type Gaussian process on $\mathcal{M}$ through a graph-Gaussian process on $\{x_1, \cdots, x_N\}$, given by

$$N(0, (\tau I_N + \Delta_N)^{-s}),$$

where τ, s respectively control inverse length scale and smoothness of samples. Samples from the graph-Gaussian process are of the form

$$u_N^{\mathrm{Ma}} = \sum_{i=1}^{k_N} (\tau + \lambda_{N,i})^{-\frac{s}{2}} \zeta_i \phi_{N,i},$$

where $(\lambda_{N,i}, \phi_{N,i})$ are the eigenpairs of the graph Laplacian Δ_N and $k_N \leq N$ is a truncation level [Sanz-Alonso and Yang, 2022].

Assumptions for the approximation result

- Suppose we use the pairwise similarity given by

$$W_{ij} = \frac{2(m+2)}{N\nu_m h_N^{m+2}} \mathbf{1}\{|x_i - x_j| < h_N\},$$

  where ν_m is the volume of m-dimensional unit ball and h_N is a graph connectivity parameter.

- The graph connectivity parameter h_N decays at a rate of $\left(\frac{\log N}{N}\right)^{\frac{1}{2m}}$.

- The truncation level k_N grows at a rate of $\left(\frac{N}{\log N}\right)^{\frac{1}{4s-6m+2}}$ with an assumption $s > \frac{3}{2}m - \frac{1}{2}$.

Approximation Result

Theorem (Approximation)

Assume the point cloud $\{x_1, \cdots, x_N\}$ are i.i.d. samples from the uniform distribution on $\mathcal{M}$ with vol($\mathcal{M}$) $= 1$, then with probability $1 - O(N^{-c})$ for some $c > 0$, there exists $T_N : \mathcal{M} \to \{x_1, \cdots, x_N\}$ with $T_N(x_i) = x_i$ such that

$$\mathbb{E}||u_N^{Ma} \circ T_N - u^{Ma}||_\infty \leq C \left(\frac{\log N}{N} \right)^{\frac{1}{4m}} =: \epsilon_N,$$

for some universal constant $C > 0$.

- Suppose the objective f is a realization of u^{Ma} and denote its restriction on the point cloud as f_N. If u_N is a realization of u_N^{Ma}, then we have
$$||u_N - f_N||_\infty \leq \delta^{-1} \epsilon_N,$$
with probability $1 - \delta$.
- More detailed analyses are provided in the paper [Kim et al., 2022].

Simple Regret Bound

Theorem (Regret Bound)

Suppose the objective function f is a realization of u^{Ma}. Applying the Bayesian optimization algorithm with graph-Gaussian process with $B_\ell = \sqrt{2\log\left(\frac{\pi^2\ell^2 N}{6\delta}\right)} + \frac{\epsilon_N\sqrt{\ell-1}}{\delta\sigma}$ yields the following simple regret:

$$\max_{z\in\mathcal{M}} f(z) - \max_{z\in\mathcal{Z}_L} f(z) = \tilde{O}\left((L^{-\frac{1}{2}} + \epsilon_N)\sqrt{k_N}\right)$$

where $\mathcal{Z}_L = \{z_1, \cdots, z_L\}$ is the set of all query points after L iterations.

- For a fixed N with $L \to \infty$, the simple regret will decrease towards the approximation error.
- As both N and L tend to infinity while $L << N$, one can recover the global maximizer of f.

Cow Manifold: No Chart

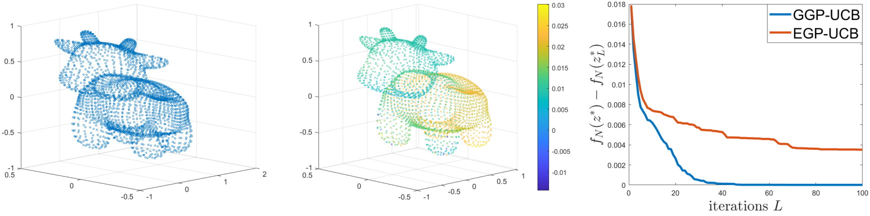

Figure: (a) Point cloud with $N = 2000$. (b) A random sample $f_{\bar{N}}$ from $u_{\bar{N}}^{\mathrm{Ma}}$ with $\tau_* = 5, s_* = 2.5, \bar{N} = 2930$; values of $f_{\bar{N}}$ vary smoothly along the point cloud. (c) Comparison of simple regrets as a function of L between GGP-UCB and EGP-UCB. The results are averaged over 50 trials.

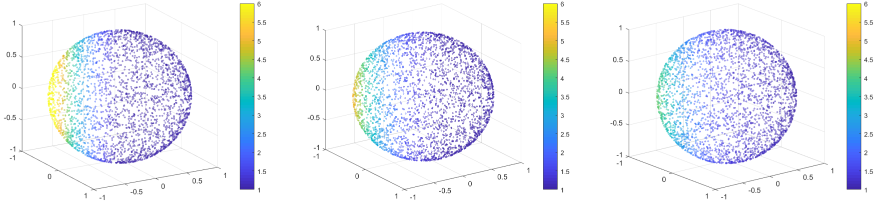

Figure: (a) Initial heat over the point cloud of size $N = 3000$. (b) Noisy corrupted heat configuration at $t = 0.25$. (c) Noisy corrupted heat configuration at $t = 0.4$.

Heat Inversion: Expensive Function evaluations II

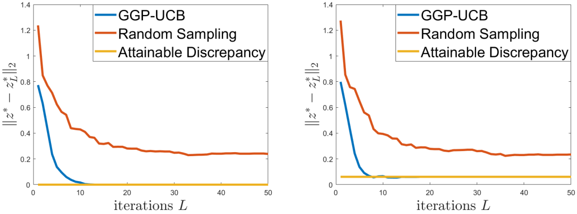

Figure: Comparison of $\|z^* - z_L^*\|_2$ between GGP-UCB and random sampling for (a) $t = 0.25$ and (b) $t = 0.4$. The results are averaged over 50 trials.

Conclusion

- We introduced a graph Gaussian process defined on a point cloud to approximate the Matérn-type Gaussian process on a manifold.
- Maximum value obtained through BO based on graph Gaussian process over the point cloud gets closer to the maximum value of the objective function on a manifold as the size of point cloud and iteration of BO increases, for some class of objective functions f.
- More clever way to leverage geometric structure in acquisition function maximization would improve efficiency.
- Looking for cool applications!

References

Borovitskiy, V., Terenin, A., Mostowsky, P., et al. (2020).
Matérn gaussian processes on riemannian manifolds.
Advances in Neural Information Processing Systems, 33:12426–12437.

Kim, H., Sanz-Alonso, D., and Yang, R. (2022).
Optimization on manifolds via graph gaussian processes.
arXiv preprint arXiv:2210.10962.

Sanz-Alonso, D. and Yang, R. (2022).
The spde approach to matérn fields: Graph representations.
Statistical Science, 37(4):519–540.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and De Freitas, N. (2015).
Taking the human out of the loop: A review of bayesian optimization.
Proceedings of the IEEE, 104(1):148–175.