Global Optimization with Parametric Function Approximation

Chong Liu CHONGLIU@CS.UCSB.EDU
Yu-Xiang Wang YUXIANGW@CS.UCSB.EDU
Department of Computer Science, University of California, Santa Barbara, CA 93106, USA

Abstract

We consider the problem of global optimization with noisy function evaluation oracles — a well-motivated problem useful for various applications ranging from hyper-parameter tuning to new material design. Existing work relies on Gaussian processes or other nonparametric family, which suffers from the curse of dimensionality. In this paper, we assume having access to a differentiable parametric family that contains the unknown function. We show that under mild assumptions, and an optimization oracle that solves the maximum-likelihood estimate, an exploration-first strategy has cumulative regret of $T^{2/3}$, and an Upper Confidence Bound (UCB) exploration algorithm enjoys a regret of $T^{1/2}$ where T is the time horizon. Our simulation shows the effectiveness of our algorithm compared with classical Bayesian optimization approaches.

1. Introduction

Consider the following optimization problem. Let $f : \mathcal{X} \to \mathbb{R}$ be an underlying non-convex function. Our goal is to find a global solution to, w.l.o.g., the maximization problem, i.e.,

$$f^* = \max_{x \in \mathcal{X}} f(x). \tag{1}$$

To learn about $f(x)$, the learner relies on zeroth-order noisy function observations, i.e., at round t, learner selects a point $x_t \in \mathcal{X}$ and receives a noisy function value y_t,

$$y_t = f(x_t) + \epsilon_t,\ \epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \tag{2}$$

where ϵ_t is the Gaussian noise. We define the cumulative regret to measure the performance. After round T, the cumulative regret is defined as $R_T = \sum_{t=1}^{T} f^* - f(x_t)$. An algorithm $\mathcal{A}$ is said to be a no-regret algorithm if $\lim_{T\to\infty} R_T(\mathcal{A})/T = 0$.

In this paper, we propose the Global Optimization via Upper Confidence Bound (GO-UCB) algorithm. The motivation is to use a parametric function f_w where $w \in \mathcal{W}$ to approximate the true function f and solve the global optimization problem. The algorithm works closely with the parameter class $\mathcal{W}$. The optimization problem is said to be "realizable" if the true function sits in the function class controlled by $\mathcal{W}$, "non-realizable" otherwise. From algorithm design, GO-UCB has two phases: (1) Phase I: Passively and uniformly query n data points. (2) Phase II: Actively query $T - n$ data points.

The goal of Phase I to sufficiently explore the function and make sure the estimated parameter $\hat{w}$ is somewhat close to the true parameter w^*. However, in Phase II, queries are made differently. One key observation is that parametric function $f_w(x)$ is equivalent to $f_x(w)$ which is the function of w parameterized by x. Therefore, the algorithm should be able to switch information between parameter class $\mathcal{W}$ and function domain $\mathcal{X}$. Specifically, at round t, the key idea is to embed all information from previous rounds $1, ..., t-1$ into the estimated $\hat{w}_t$ and then use it to actively query points from $\mathcal{X}$. Phase II of GO-UCB closely follows this idea; it takes advantage of the nice properties of the neighboring region of w^* resulted by Phase I and builds the UCB in $\mathcal{X}$, i.e., the bound of $|f_{\hat{w}}(x) - f_{w^*}(x)|, \forall x \in \mathcal{X}$. Then the algorithm optimizes the objective function $f_{w^*}(x)$ via UCB.

To estimate $\hat{w}$ efficiently, throughout this paper, we assume the algorithm has access to the Maximum Likelihood Estimation (MLE) oracle that is able to return a estimated $\hat{w}$ for true w^*. In detail, at round t after observing a dataset $\{(x_i, y_i)\}_{i=1}^{t-1}$,

$$\hat{w}_t \leftarrow \mathrm{MLE}_{w \in \mathcal{W}}(x_1, y_1, ..., x_{t-1}, y_{t-1}). \tag{3}$$

Then the next query point x_t is selected based on $\hat{w}_t$. By assuming access to MLE oracle, it allows us to avoid tediously enumerating all possible parameters in the parameter class and establish the UCB without kernels.

In summary, our main contributions are: (1) We study the important but challenging global non-convex optimization problem and propose the GO-UCB algorithm to solve it with parametric function approximation. (2) GO-UCB does *not* rely on Gaussian process used in Bayesian optimization which suffers from the curse of dimensionality. With only zeroth noisy feedback, our algorithm is able to solve the global optimization problem where the objective function is non-convex, black-boxed, and even high-dimensional and non-continuous. (3) We prove that GO-UCB converges to the global optima with cumulative regret at the order of $\tilde{O}(\sqrt{T\log(T)})$ where T is the time horizon, which is better than the $\tilde{O}(T^{2/3})$ rate achieved by the exploration-first algorithm. (4) Our simulation shows that GO-UCB works better than classical Bayesian optimization methods in both realizable and non-realizable settings.

2. Preliminaries

2.1 Notations

We use $[m]$ to denote the set $\{1, 2, ..., m\}$. The algorithm queries n points in Phase I and $T-n$ points in Phase II, so T is the total time horizon which is indexed by t. Let $\mathcal{X} \in \mathbb{R}^{d_x}$ denote the function domain, w.l.o.g. $\mathcal{Y} = [0, F] \in \mathbb{R}$ denote the function range where F is a constant, and $\mathcal{W} \in \mathbb{R}^{d_w}$ denote the parameter class. An parametric function f_w is a twice differentiable function w.r.t. w mapping from $\mathcal{X}$ to $\mathcal{Y}$. Let $L(w) = \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}_{(x,y)\sim\mathcal{D}_i}(f_x(w) - y)^2$ denote the expected loss function where $\mathcal{D}_i$ denotes the data generating distribution at round i. For a vector x, its ℓ_p norm is denoted by $\|x\|_p = (\sum_{i=1}^{d} |x_i|^p)^{1/p}$ for $1 \leq p < \infty$. For a vector x and a square matrix A, define $\|x\|_A^2 = x^\top Ax$. Throughout this paper, we use standard big O notations; and to improve the readability, we use $\tilde{O}$ to hide poly-logarithmic factors.

2.2 Assumptions

Assumption 1 (MLE oracle) *At round i, let $p_w(y_i|x_i)$ denote the probability of observing y_i conditioning on x_i parametrized by w. The MLE oracle returns the estimated $\hat{w}$ for true w^* after t rounds of observations, i.e.,* $\hat{w} \leftarrow \mathrm{argmax}_{w\in\mathcal{M}} \sum_{i=1}^{t} \log p_w(y_i|x_i)$.

Remark 2 *Following Agarwal et al. (2020), we make this assumption as a method towards practical algorithm that avoids enumerating all possible parameters in $\mathcal{W}$. In fact, the MLE oracle assumption is common in literature (Agarwal et al., 2014; Du et al., 2019; Misra et al., 2020) and it can be approximated in practice.*

Assumption 3 (Parameter class) *The parameter class $\mathcal{W}$ is finite and realizable, i.e., $|\mathcal{W}| < \infty$ and the true parameter $w^* \in \mathcal{W}$.*

Assumption 4 (Loss function) *At each round $t \in [T]$, $\forall (x, y) \sim \mathcal{D}_t, \ell(w) = \mathbb{E}[(f_x(w) - y)^2]$ satisfies locally self-concordance, μ-strongly convexity, and α, γ-growth condition at w^*,*

$$\forall w \in \mathcal{W}, \min \left\{ \frac{\mu}{2} \|w - w^*\|_2^2, \frac{\alpha}{2} \|w - w^*\|_2^\gamma \right\} \leq \ell(w) - \ell(w^*). \tag{4}$$

Remark 5 *This assumption may appear to be strong, but it does not require convexity except in a local neighborhood near the global optimal w^*. It also does not limit the number of spurious local minima, as the global γ growth condition only gives a mild lower bound as we move away from w^*. Our results works even if γ is a small constant < 1. Local self-concordance is needed for technical reasons, but again it is only required near w^*. Example 4 of Zhang et al. (2017) lists some self-concordant function examples.*

Assumption 6 (Objective function) *Let G, C be constants. The objective function $f_x(w)$ is G-Lipschitz and C-smooth (C-gradient Lipschitz), i.e., $\forall x \in \mathcal{X}$,*

$$\|\nabla f_x(w)\|_2 \leq G, \forall w \in \mathcal{W}, \tag{5}$$

$$f_x(w_1) \leq f_x(w_2) + (w_1 - w_2)^\top \nabla f_x(w_2) + \frac{C}{2} \|w_1 - w_2\|_2^2, \forall w_1, w_2 \in \mathcal{W}. \tag{6}$$

3. Main Results

In this section, we derive the upper bound on $\|\hat{w} - w^*\|_2$ via a MLE lemma and then build the UCB for active queries. Next, we show GO-UCB algorithm and prove its regret analysis.

3.1 MLE Oracle and Guarantees

Lemma 7 (MLE lemma (adapted from Agarwal et al. (2020))) *Suppose Assumption 1 3 hold. Fix $\delta \in (0, 1)$, after round t, then w.p. $> 1 - \delta$, the MLE-estimated $\hat{w}$ satisfies*

$$\frac{1}{t} \sum_{i=1}^{t} \mathbb{E}_{x \sim \mathcal{D}_i} (f_x(w^*) - f_x(\hat{w}))^2 \leq \frac{4\sigma^2 \log(|\mathcal{W}|/\delta)}{t}. \tag{7}$$

Theorem 8 (MLE guarantee) *Suppose Assumption 1 3 & 4 hold. Fix $\delta \in (0, 1)$, after sampling t points where t satisfies*

$$t \geq \max \left\{ \frac{2^{3+\gamma} \sigma^2 \log(|\mathcal{W}|/\delta)}{\alpha \mu^{\gamma/2}}, \frac{16\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu^2} \right\}. \tag{8}$$

Then w.p. $> 1 - \delta$, the MLE-estimated $\hat{w}$ satisfies $\|\hat{w} - w^*\|_2^2 \leq \frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}$.

Remark 9 *Under mild assumptions, the main MLE guarantee shows that $\|\hat{w} - w^*\|_2$ converges at the order of $\tilde{O}(\sqrt{1/t})$. Note the bound itself doesn't depend on anything from previous rounds except $\hat{w}$. Detailed proofs are shown in Appendix C D.*

3.2 Our Algorithm

We first build the upper confidence bound and then present our GO-UCB algorithm.

Theorem 10 (Upper confidence bound) *Suppose Assumption 1 3 4 6 hold. After uniformly querying n data points in Phase I where*

$$n \geq \max\left\{\frac{2^{3+\gamma}\sigma^2\log(|\mathcal{W}|/\delta)}{\alpha\mu^{\gamma/2}}, \frac{16\sigma^2\log(|\mathcal{W}|/\delta)}{\mu^2}\right\}. \tag{9}$$

Then at round t in Phase II, fix $\delta \in (0,1)$, w.p. $> 1-\delta$, $\forall x \in \mathcal{X}$,

$$|f_x(\hat{w}_t) - f_x(w^*)| \leq \|\nabla f_x(\hat{w}_t)\|_2\sqrt{\frac{512\sigma^2\log(|\mathcal{W}|/\delta)}{\mu t}} + \frac{256C\sigma^2\log(|\mathcal{W}|/\delta)}{\mu t}. \tag{10}$$

Remark 11 *Note t is indexing over number of queries in both Phase I and II and the sample comlexity requirement (eq. (9)) is the same as eq. (8). The theorem says that for all $x \in \mathcal{X}$, $f_{\hat{w}_t}(x)$ converges to $f_{w^*}(x)$ at the rate of $\tilde{O}(\sqrt{1/t})$. Also, the second term in Theorem 10 converges faster than the first term.*

The Global Optimization via Upper Confidence Bound (GO-UCB) algorithm is shown in Algorithm 1. Step 1-4 of Phase I are pretty standard which are doing uniform sampling. In Phase II, Step 2 queries the MLE oracle for estimated parameter $\hat{w}_t$. Then in Step 3, first term $f_x(\hat{w}_t)$ is the function parameterized by $\hat{w}_t$ and the second and third terms measure the uncertainty of the current estimation. Note $\nabla f_x(\hat{w}_t)$ is the gradient of function w.r.t. $\hat{w}_t$ parameterized by x, which can be calculated for each $x \in \mathcal{X}$.

Algorithm 1 GO-UCB

```
Input: Time horizon $T$, noise parameter $\sigma$, MLE oracle, uniform distribution $\mathcal{U}$, constant $C$.
Phase I (Passive query):
1: Set $n \geq \max\{\frac{2^{3+\gamma}\sigma^2\log(|\mathcal{W}|/\delta)}{\alpha\mu^{\gamma/2}}, \frac{16\sigma^2\log(|\mathcal{W}|/\delta)}{\mu^2}\}$.
2: for $t = 1, ..., n$ do
3:     Sample $x_t \sim \mathcal{U}(\mathcal{X})$.
4:     Observe $y_t = f(x_t) + \epsilon_t$.
5: end for
Phase II (Active query):
1: for $t = n+1, ..., T$ do
2:     MLE query: $\hat{w}_t \leftarrow \mathrm{argmax}_{w \in \mathcal{W}} \sum_{i=1}^{t-1} \log p(y_i|x_i, w, \sigma)$.
3:     Select $x_t = \mathrm{argmax}_{x \in \mathcal{X}} f_x(\hat{w}_t) + \|\nabla f_x(\hat{w}_t)\|_2\sqrt{\frac{512\sigma^2\log(|\mathcal{W}|/\delta)}{\mu t}} + \frac{256C\sigma^2\log(|\mathcal{W}|/\delta)}{\mu t}$.
4:     Observe $y_t = f(x_t) + \epsilon_t$.
5: end for
Output: $\hat{x} = \mathrm{argmax}_{x \in \{x_1, ..., x_T\}} y(x)$.
```

3.3 Regret Analysis

Theorem 12 (Cumulative regret of GO-UCB) *Suppose Assumption 1 3 4 6 hold. After uniformly querying n data points in Phase I where $n \geq \max\{\frac{2^{3+\gamma}\sigma^2 \log(|\mathcal{W}|/\delta)}{\alpha\mu^{\gamma/2}}, \frac{16\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu^2}\}$. Then fix $\delta \in (0,1)$, after T rounds in total, w.p. $> 1-\delta$,*

$$R_T \leq nF + 2G\sqrt{T-n}\sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)\log(T/n)}{\mu}} + \frac{512C\sigma^2 \log(|\mathcal{W}|/\delta)\log(T/n)}{\mu}, \quad (11)$$

where $\alpha, \gamma, \mu, F, G, C$ are constants. Moreover, by choosing $n = \Theta(\sqrt{T})$, $R_T \leq \tilde{O}(\sqrt{T\log(T)})$.

Cumulative regret of GO-UCB is $\tilde{O}(\sqrt{T})$, which is the same rate achieved by GP-UCB. However, no Gaussian process assumption is needed in our algorithm.

Remark 13 (Choice of n) *The success of GO-UCB relies on the careful choice of n. The choice of n plays two role here. First, as a sufficiently large number in eq. (9), the choice of n guarantees that the MLE-estimated $\hat{w}$ lies in the neighboring region of w^* of the loss function $L(w)$ with high probability. The neighboring region of w^* has nice properties, e.g., strong convexity, which allow us to transfer the predictive MLE bound into the ℓ_2-distance between $\hat{w}$ and w^*. Second, in Phase I we are doing uniform sampling therefore the cumulative regret can only be bounded by $O(n)$. The choice of n makes sure the cumulative regret is smaller than the one incurred in Phase II.*

Remark 14 (Bonus term) *In Step 3 of Algorithm 1, UCB relies on $\|\nabla f_x(\hat{w}_t)\|_2$, which can be understood as a* bonus *term. Here in our analysis, we simply assume it is smaller than some constant G, however, if it has faster convergence than constant, it leads to an improved cumulative regret bound, potentially $\tilde{O}(\log(T))$. See further discuss in Appendix E.*

For comparison, we also include a cumulative regret bound of the exploration-first algorithm, which also has two phases. In Phase I, it is the same as GO-UCB, and then in Phase II, it relies on the convergence bound of $\hat{w}_n$, rather than $\hat{w}_t$, to actively query $T-n$ points. Note $\hat{w}_n$ doesn't change after Phase I.

Lemma 15 (Cumulative regret of exploration-first algorithm) *Suppose Assumption 1 3 4 6 hold. In Phase I the number of queries n satisfies eq. (9). Then fix $\delta \in (0,1)$, after T rounds in total, w.p. $> 1-\delta$, $R_T \leq nF + 2(T-n)G\sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu n}}$. Moreover, by choosing $n = \Theta(T^{2/3})$, $R_T \leq \tilde{O}(T^{2/3})$.*

The exploration-first algorithm is also a no-regret algorithm that achieves the $\tilde{O}(T^{2/3})$ cumulative regret after careful choice of n. However, it is worse than the $\tilde{O}(\sqrt{T\log(T)})$ rate of GO-UCB. It shows the effectiveness of using UCB in Phase II.

4. Simulation

In this section, we present simulation results on two 1-dimensional functions:

$$f_1 = 1 + \frac{1}{1+e^{-(x+1)}}, \quad f_2 = \sin(x/4), \quad (12)$$

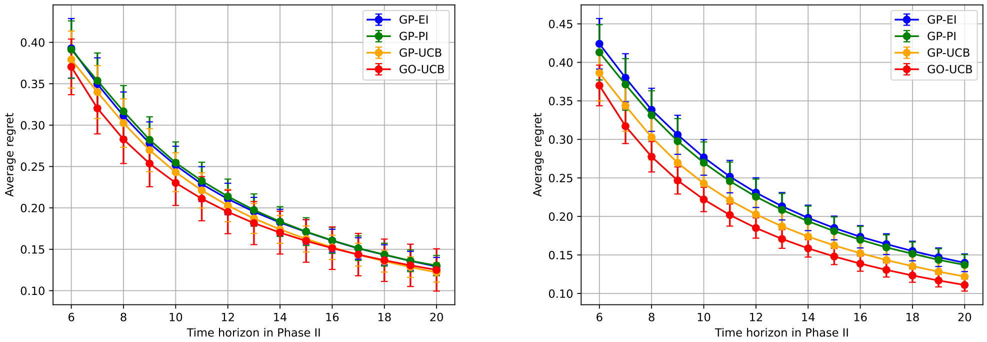
where the domain is $\mathcal{X} = [-2\pi, 2\pi]$. We compare our algorithm GO-UCB with three Bayesian optimization methods: GP-EI (Jones et al., 1998), GP-PI (Kushner, 1964), and GP-UCB (Srinivas et al., 2010). The GP-EI, GP-PI, and GP-UCB are implemented by the scikit-optimize package (Head et al., 2021) with default settings under the BSD license.

Our model is a 2-layer neural network parameterized by $w = [w_1, b_1, w_2, b_2] \in \mathbb{R}^4$:

$$\hat{f} = \frac{w_2}{1 + e^{-(w_1 x + b_1)}} + b_2, \tag{13}$$

which has two linear layers and a sigmoid activation function. Therefore, optimizing f_1 is a realizable task because when $\hat{f} = f_1$ if $w = [1, 1, 1, 1]$. Optimizing f_2 is a non-realizable task.

To run GO-UCB in practice, we set a constant $U = 10$ and use the following surrogate criterion to select x_t at the round t: $x_t = \mathrm{argmax}_{x \in \mathcal{X}} f_x(\hat{w}_t) + \sigma \|\nabla f_x(\hat{w}_t)\|_2 \sqrt{\frac{U}{t}} + \frac{U}{t}$. The MLE oracle is approximated by gradient descent algorithm with 500 iterations. Phase I number of data points N is set to be 5 and total time horizon T is set as 20. The whole simulation is repeated for 100 times, and the average regret (the lower the better) in Phase II is reported by mean and error bar in the following figures. The error bar is measured by Wald's test with 98% confidence. From Figure 1, we learn that in both realizable and non-realizable tasks, our algorithm GO-UCB performs significantly better than all other Bayesian optimization approaches. Surprisingly, GO-UCB performs even better (with error bar gaps) in non-realizable tasks where non-parametric Bayesian optimization is expected to do well. Though there are only 15 rounds in Phase II, we can still observe that error bars are becoming smaller as the number of round goes up. Among Bayesian optimization methods, GP-UCB uses the same UCB idea as ours and it performs similarly as our method.



(a) Realizable case (f_1)

(b) Non-realizable case (f_2)

Figure 1: Algorithm comparison between GP-EI, GP-PI, GP-UCB, and GO-UCB.

5. Discussion

We include a section discussing related work in Appendix A and a section showing all auxiliary lemmas in Appendix B. Future work includes understanding the bonus term $\|\nabla f_x(\hat{w})\|$ and doing more experiments on high dimensional functions optimization.

References

Alekh Agarwal, Daniel Hsu, Satyen Kale, John Langford, Lihong Li, and Robert Schapire. Taming the monster: A fast and simple algorithm for contextual bandits. In *International Conference on Machine Learning (ICML'14)*, 2014.

Alekh Agarwal, Sham Kakade, Akshay Krishnamurthy, and Wen Sun. Flambe: Structural complexity and representation learning of low rank mdps. In *Advances in Neural Information Processing Systems 33 (NeurIPS'20)*, 2020.

Naman Agarwal, Zeyuan Allen-Zhu, Brian Bullins, Elad Hazan, and Tengyu Ma. Finding approximate local minima faster than gradient descent. In *Symposium on Theory of Computing (STOC'17)*, 2017.

Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12(46):1655–1695, 2011.

Adam D Bull. Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 12(10), 2011.

Xu Cai and Jonathan Scarlett. On lower bounds for standard and robust gaussian process bandit optimization. In *International Conference on Machine Learning (ICML'21)*, 2021.

Simon Du, Akshay Krishnamurthy, Nan Jiang, Alekh Agarwal, Miroslav Dudik, and John Langford. Provably efficient rl with rich observations via latent state decoding. In *International Conference on Machine Learning (ICML'19)*, 2019.

Dylan Foster and Alexander Rakhlin. Beyond ucb: Optimal and efficient contextual bandits with regression oracles. In *International Conference on Machine Learning (ICML'20)*, 2020.

Peter Frazier, Warren Powell, and Savas Dayanik. The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4):599–613, 2009.

Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

Peter I Frazier and Jialei Wang. Bayesian optimization for materials design. In *Information science for materials discovery and design*, pages 45–75. Springer, 2016.

Steve Hanneke et al. Theory of disagreement-based active learning. *Foundations and Trends® in Machine Learning*, 7(2-3):131–309, 2014.

Elad Hazan, Adam Klivans, and Yang Yuan. Hyperparameter optimization: a spectral approach. In *International Conference on Learning Representations (ICLR'18)*, 2018.

Tim Head, Manoj Kumar, Holger Nahrstaedt, Gilles Louppe, and Iaroslav Shcherbatyi. scikit-optimize. https://scikit-optimize.github.io, 2021.

Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

Kirthevasan Kandasamy, Willie Neiswanger, Jeff Schneider, Barnabas Poczos, and Eric P Xing. Neural architecture search with bayesian optimisation and optimal transport. In *Advances in neural information processing systems 31 (NeurIPS'18)*, 2018.

Krishna Reddy Kesari and Jean Honorio. First order methods take exponential time to converge to global minimizers of non-convex functions. In *International Symposium on Information Theory (ISIT'21)*, 2021.

Harold J Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, 8(1):97–106, 1964.

Yingkai Li, Yining Wang, and Yuan Zhou. Nearly minimax-optimal regret for linearly parameterized bandits. In *Annual Conference on Learning Theory (COLT'19)*, 2019.

Cédric Malherbe and Nicolas Vayatis. Global optimization of lipschitz functions. In *International Conference on Machine Learning (ICML'17)*, 2017.

Dipendra Misra, Mikael Henaff, Akshay Krishnamurthy, and John Langford. Kinematic state abstraction and provably efficient rich-observation reinforcement learning. In *International conference on machine learning (ICML'20)*, 2020.

AHG Rinnooy Kan and Gerrit T Timmer. Stochastic global optimization methods part i: Clustering methods. *Mathematical programming*, 39(1):27–56, 1987a.

AHG Rinnooy Kan and Gerrit T Timmer. Stochastic global optimization methods part ii: Multi level methods. *Mathematical Programming*, 39(1):57–78, 1987b.

Daniel Russo and Benjamin Van Roy. Eluder dimension and the sample complexity of optimistic exploration. In *Advances in Neural Information Processing Systems 26 (NeurIPS'13)*, 2013.

Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy gaussian process bandit optimization. In *Annual Conference on Learning Theory (COLT'17)*, 2017.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

Shubhanshu Shekhar and Tara Javidi. Significance of gradient information in bayesian optimization. In *International Conference on Artificial Intelligence and Statistics (AISTATS'21)*, 2021.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *International Conference on Machine Learning (ICML'10)*, 2010.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

Linnan Wang, Rodrigo Fonseca, and Yuandong Tian. Learning search space partition for black-box optimization using monte carlo tree search. In *Advances in Neural Information Processing Systems 33 (NeurIPS'20)*, 2020.

Yining Wang, Sivaraman Balakrishnan, and Aarti Singh. Optimization of smooth functions with noisy observations: Local minimax rates. In *Advances in Neural Information Processing Systems 31 (NeurIPS'18)*, 2018.

Christopher K Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning.* MIT Press, 2006.

Jian Wu, Matthias Poloczek, Andrew G Wilson, and Peter Frazier. Bayesian optimization with gradients. In *Advances in neural information processing systems 30 (NeurIPS'17)*, 2017.

Lijun Zhang, Tianbao Yang, Jinfeng Yi, Rong Jin, and Zhi-Hua Zhou. Improved dynamic regret for non-degenerate functions. In *Advances in Neural Information Processing Systems 30 (NeurIPS'17)*, 2017.

Heyang Zhao, Dongruo Zhou, Jiafan He, and Quanquan Gu. Bandit learning with general function classes: Heteroscedastic noise and variance-dependent regret bounds. *arXiv preprint arXiv:2202.13603*, 2022.

Appendix A. Related Work

In this section, we briefly review related work to our paper. Global non-convex optimization is an important problem that can be found in a lot of research communities and real-world applications, e.g., optimization (Rinnooy Kan and Timmer, 1987a,b), machine learning (Bubeck et al., 2011; Malherbe and Vayatis, 2017), hyperparameter tuning (Hazan et al., 2018), neural architecture search (Kandasamy et al., 2018; Wang et al., 2020), and material discovery (Frazier and Wang, 2016).

Generally speaking, solving a global non-convex optimization is NP-hard. Bayesian optimization (Shahriari et al., 2015; Frazier, 2018) or Gaussian process bandit optimization (Cai and Scarlett, 2021) is one line of research focusing on zeroth-order optimization with noisy feedback. The objective function is usually modeled by Gaussian Process (Williams and Rasmussen, 2006) using some kernels and then acquisition functions are used to *actively* selected data points to query. Popular choices of acquisition functions are Upper Confidence Bound (UCB) (Srinivas et al., 2010), expected improvement (Jones et al., 1998; Bull, 2011), knowledge gradient (Frazier et al., 2009), probability of improvement (Kushner, 1964), and Thompson sampling (Thompson, 1933). Among them, GP-UCB (Srinivas et al., 2010) is the closest to our paper because our algorithm also *actively* selects data points in a UCB style but the construction of UCB in our paper is different since we are not working with kernels. Scarlett et al. (2017) proves lower bounds on regret for noisy Gaussian process bandit optimization. One drawback of Bayesian optimization is that it suffers from curse of dimensionality.

Without Gaussian process, Wang et al. (2018) solves the smooth function optimization in a candidate removal way which resembles the disagreement-based active learning (Hanneke et al., 2014). There is also a line of research on dependent arm bandit problems. Li et al. (2019) studies the linearly parameterized contextual bandit and then Foster and Rakhlin (2020) goes beyond the linear function. All of these are contextual bandit settings which is different to our work. Russo and Van Roy (2013) studies the sample complexity of multi-arm bandit and propose the Eluder dimension to capture the dependence among action arms. Recent work (Zhao et al., 2022) builds upon it and captures contextual linear bandits and generalized linear bandits as special cases. By contrast, our work considers non-convex function with noisy zeroth order feedback.

A collection of research papers considers the global non-convex optimization with first-order feedback where gradient information is available (Agarwal et al., 2017). Wu et al. (2017) and Shekhar and Javidi (2021) studies Bayesian optimization with gradient feedback. Kesari and Honorio (2021) shows that first-order methods take exponential time to converge to global optima.

The MLE oracle assumption was introduced in reinforcement learning literature (Agarwal et al., 2014) which allows algorithm designer to avoid tedious enumerations over all parameters in the large parameter class. Because of this, MLE oracle assumption becomes a common assumption later in reinforcement. See Du et al. (2019); Misra et al. (2020) and Agarwal et al. (2020). However, to the best of our knowledge, this assumption has never been used in global non-convex optimization before. In this paper, we show that with MLE oracle assumption, we can propose a UCB-style no-regret algorithm that efficiently solves the problem with parametric function approximation.

Appendix B. Auxiliary Lemmas

In this section, we list auxiliary lemmas that are used in the main paper.

Lemma 16 (Lemma 24 of Agarwal et al. (2020)) *Let $D = \{(x_i, y_i)\}_{i=1}^t$ be a dataset of t examples, and let $D' = \{(x_i', y_i')\}_{i=1}^t$ be a tangent sequence where $x_i' \sim \mathcal{D}_i(x_{1:i-1}, y_{1:i-1})$ and $y_i' \sim p(\cdot|x_i')$. Let $Q(p, D) = \sum_{i=1}^t q(p, (x_i, y_i))$ be a function that decomposes additively across examples where q is any function, and let $\hat{p}(D)$ be any estimator taking as input random variable D and with range $\mathcal{P}$. Then*

$$\mathbb{E}_D[\exp(Q(\hat{p}(D), D) - \log \mathbb{E}_{D'} \exp(Q(\hat{p}(D), D')) - \log |\mathcal{P}|)] \leq 1. \tag{14}$$

Lemma 17 (Hessian of self-concordant function (Lemma 8 of Zhang et al. (2017))) *Let $f(x)$ be a self-concordant function, and $\|h\|_{\nabla^2 f(x)} = \sqrt{h^\top \nabla^2 f(x) h}$. Then, for a given point x and for any h with $\|h\|_{\nabla^2 f(x)} \leq 1$, we have*

$$(1 - \|h\|_{\nabla^2 f(x)})^2 \nabla^2 f(x) \preceq \nabla^2 f(x+h) \preceq \frac{\nabla^2 f(x)}{(1 - \|h\|_{\nabla^2 f(x)})^2}. \tag{15}$$

Lemma 18 (KL divergence between Gaussian distributions) *Let $p(x), q(x)$ denote the pdf of $\mathcal{N}(\mu_1, \sigma_1), \mathcal{N}(\mu_2, \sigma_2)$, respectively. Then the KL divergence $\mathrm{KL}(p, q)$ between them is*

$$\mathrm{KL}(p, q) = -\frac{1}{2} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} + \log\Big(\frac{\sigma_2}{\sigma_1}\Big). \tag{16}$$

Proof By definiton of KL divergence,

$$\begin{aligned}
\mathrm{KL}(p, q) &= \int p(x) \log\Big(\frac{p(x)}{q(x)}\Big) \mathrm{d}x & (17)\\
&= \int \frac{1}{\sqrt{2\pi}\sigma_1} \exp\Big(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\Big)\Big(\frac{(x-\mu_2)^2}{2\sigma_2^2} + \frac{1}{2}\log(2\pi) - \log(\sigma_2)\\
&\quad - \frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{1}{2}\log(2\pi) - \log(\sigma_1)\Big)\mathrm{d}x & (18)\\
&= \int \frac{1}{\sqrt{2\pi}\sigma_1} \exp\Big(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\Big)\Big(\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2} + \log\Big(\frac{\sigma_2}{\sigma_1}\Big)\Big)\mathrm{d}x & (19)\\
&= \mathbb{E}_p\Big[\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2} + \log\Big(\frac{\sigma_2}{\sigma_1}\Big)\Big] & (20)\\
&= \frac{\mathbb{E}_p[(x-\mu_1+\mu_1-\mu_2)^2]}{2\sigma_2^2} - \frac{\mathbb{E}_p[(x-\mu_1)^2]}{2\sigma_1^2} + \log\Big(\frac{\sigma_2}{\sigma_1}\Big) & (21)\\
&= \frac{\mathbb{E}_p[(x-\mu_1)^2 + 2(x-\mu_1)(\mu_1-\mu_2) + (\mu_1-\mu_2)^2]}{2\sigma_2^2} - \frac{1}{2} + \log\Big(\frac{\sigma_2}{\sigma_1}\Big) & (22)\\
&= \frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2} + \log\Big(\frac{\sigma_2}{\sigma_1}\Big). & (23)
\end{aligned}$$

where eq. (18) is due to pdf function of Gaussian distribution, eq. (20) is due to definition of expectation, and eq. (22) (23) are due to definition of variance. ■

Appendix C. Intermediate Results of MLE Guarantees

Corollary 19 *Suppose Assumption 1 3 hold. Fix $\delta \in (0,1)$, after round t, then w.p. $> 1-\delta$, the MLE-estimated $\hat{w}$ satisfies*

$$L(\hat{w}) - L(w^*) \leq \frac{4\sigma^2 \log(|\mathcal{W}|/\delta)}{t}. \tag{24}$$

Proof Plug in the definition of loss function $L(w)$ and we have

$$\begin{aligned}
L(\hat{w}) - L(w^*) &= \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}_{\mathcal{D}_i}(f_x(\hat{w}) - y)^2 - (f_x(w^*) - y)^2 && (25)\\
&= \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}_{\mathcal{D}_i}(f_x(\hat{w}) - f_x(w^*) + f_x(w^*) - y)^2 - (f_x(w^*) - y)^2 && (26)\\
&= \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}_{\mathcal{D}_i}(f_x(\hat{w}) - f_x(w^*))^2 + 2(f_x(\hat{w}) - f_x(w^*))(f_x(w^*) - y) && (27)\\
&= \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}_{\mathcal{D}_i}(f_x(\hat{w}) - f_x(w^*))^2 + 2(f_x(\hat{w}) - f_x(w^*))\epsilon_i && (28)\\
&= \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}_{\mathcal{D}_i}(f_x(\hat{w}) - f_x(w^*))^2 && (29)\\
&\leq \frac{4\sigma^2 \log(|\mathcal{W}|/\delta)}{t}, && (30)
\end{aligned}$$

where eq. (28) (29) are due to Gaussian noise model (eq. (2)) and eq. (30) holds because of Lemma 7. ■

Lemma 20 *Suppose Assumption 1 3 hold. Fix $\delta \in (0,1)$, after round t, then w.p. $> 1-\delta$, the MLE-estimated $\hat{w}$ satisfies*

$$\|\hat{w} - w^*\|^2_{\nabla^2 L(\tilde{w})} \leq \frac{8\sigma^2 \log(|\mathcal{W}|/\delta)}{t}, \tag{31}$$

where $\tilde{w}$ is a point on the line segment between $\hat{w}$ and w^*.

Proof Use Taylor expansion on loss function $L(\hat{w})$ at w^*,

$$L(\hat{w}) = L(w^*) + (\hat{w} - w^*)^\top \nabla L(w^*) + \frac{1}{2}\|\hat{w} - w^*\|^2_{\nabla^2 L(\tilde{w})}, \tag{32}$$

where $\tilde{w}$ is on the line segment between $\hat{w}$ and w. Rearrange the equation,

$$\frac{1}{2}\|\hat{w} - w^*\|^2_{\nabla^2 L(\tilde{w})} = L(\hat{w}) - L(w^*) + (w^* - \hat{w})^\top \nabla L(w^*). \tag{33}$$

By Assumption 4, we know that $L(w^*)$ is convex function. Due to the first order optimal condition of convex function, $(w^* - \hat{w})^\top \nabla L(w^*) \leq 0$. Therefore,

$$\frac{1}{2}\|\hat{w} - w^*\|^2_{\nabla^2 L(\tilde{w})} \leq L(\hat{w}) - L(w^*). \tag{34}$$

The proof completes by plugging in Corollary 19. ■

Lemma 21 *Suppose Assumption 1 3 4 hold, then $L(w^*)$ is a μ-strongly convex function. Also, fix $\delta \in (0,1)$, after sampling t points where t satisfies*

$$t \geq \max\Big\{\frac{2^{3+\gamma}\sigma^2 \log(|\mathcal{W}|/\delta)}{\alpha\mu^{\gamma/2}}, \frac{16\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu^2}\Big\}, \tag{35}$$

where α, γ, μ are constants, then w.p. $> 1-\delta$, the MLE-estimated $\hat{w}$ satisfies

$$\|\hat{w} - w^*\|_{\nabla^2 L(w^*)} \leq \frac{1}{2}. \tag{36}$$

Proof By definition of $L(w^*)$ and Assumption 4, $L(w^*)$ is a μ-strongly convex function, then $\lambda_{\min}(\nabla^2 L(w^*)) \geq \mu$. Therefore,

$$\|\hat{w} - w^*\|^2_{\nabla^2 L(w^*)} \leq \frac{\|\hat{w} - w^*\|_2^2}{\lambda_{\min}(\nabla^2 L(w^*))} \leq \frac{\|\hat{w} - w^*\|_2^2}{\mu}. \tag{37}$$

Next, we discuss two cases of relationships between $\hat{w}$ and w^*.

Case I. If $\hat{w}$ is far away from w^*, by growth condition (Assumption 4) and Corollary 19,

$$\frac{\alpha}{2}\|\hat{w} - w^*\|_2^\gamma \leq L(\hat{w}) - L(w^*) \leq \frac{4\sigma^2 \log(|\mathcal{W}|/\delta)}{t}, \tag{38}$$

$$\|\hat{w} - w^*\|_{\nabla^2 L(w^*)} \leq \frac{\|\hat{w} - w^*\|_2}{\sqrt{\mu}} \leq \frac{1}{\sqrt{\mu}}\Big(\frac{8\sigma^2 \log(|\mathcal{W}|/\delta)}{\alpha t}\Big)^{\frac{1}{\gamma}}. \tag{39}$$

Therefore, set $\frac{1}{\sqrt{\mu}}(\frac{8\sigma^2 \log(|\mathcal{W}|/\delta)}{\alpha t})^{\frac{1}{\gamma}} \leq \frac{1}{2}$ will result in a sample complexity bound on t, i.e.,

$$t \geq \frac{2^{3+\gamma}\sigma^2 \log(|\mathcal{W}|/\delta)}{\alpha\mu^{\gamma/2}}. \tag{40}$$

Case II. If $\hat{w}$ is close to w^*, by local strong convexity (Assumption 4) and Corollary 19,

$$\mu\|\hat{w} - w^*\|_2^2 \leq L(\hat{w}) - L(w^*) \leq \frac{4\sigma^2 \log(|\mathcal{W}|/\delta)}{t} \tag{41}$$

$$\|\hat{w} - w^*\|_{\nabla^2 L(w^*)} \leq \frac{1}{\sqrt{\mu}}\|\hat{w} - w^*\|_2 \leq \sqrt{\frac{4\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu^2 t}}. \tag{42}$$

Therefore, set $\sqrt{\frac{4\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu^2 t}} \leq \frac{1}{2}$ will also lead to a sample complexity bound on n, i.e.,

$$t \geq \frac{16\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu^2}. \tag{43}$$

Combining eq. (40) (43) completes the proof. ■

Corollary 22 *Suppose Assumption 1 3 4 hold. Fix $\delta \in (0,1)$, after sampling t points where t satisfies eq. (35), then w.p. $> 1-\delta$, the MLE-estimated $\hat{w}$ satisfies*

$$\lambda_{\min}(\nabla^2 L(\hat{w})) \geq \frac{\mu}{4}. \tag{44}$$

Proof As a result of Assumption 4 and Lemma 17 in Appendix B,

$$(1 - \|\hat{w} - w^*\|_{\nabla^2 L(w^*)})^2 \nabla^2 L(w^*) \preceq \nabla^2 L(\hat{w}). \tag{45}$$

Plugging in Lemma 21 completes the proof. ■

Lemma 23 *Suppose Assumption 1 3 4 hold. Fix $\delta \in (0,1)$, after sampling t points where t satisfies eq. (35), then w.p. $> 1-\delta$, the MLE-estimated $\hat{w}$ satisfies*

$$\|\hat{w} - w^*\|^2_{\nabla^2 L(\hat{w})} \leq \frac{128\sigma^2 \log(|\mathcal{W}|/\delta)}{t}. \tag{46}$$

Proof As a result of Assumption 4 and Lemma 17 in Appendix B, $L(w)$ is also a self-concordant function at w^*, which means

$$(1 - \|\tilde{w} - w^*\|_{\nabla^2 L(w^*)})^2 \nabla^2 L(w^*) \preceq \nabla^2 L(\tilde{w}), \tag{47}$$

$$(1 - \|\hat{w} - w^*\|_{\nabla^2 L(w^*)})^2 \nabla^2 L(\hat{w}) \preceq \nabla^2 L(w^*). \tag{48}$$

Put them together,

$$(1 - \|\tilde{w} - w^*\|_{\nabla^2 L(w^*)})^2 (1 - \|\hat{w} - w^*\|_{\nabla^2 L(w^*)})^2 \nabla^2 L(\hat{w}) \preceq \nabla^2 L(\tilde{w}), \tag{49}$$

$$(1 - \|\hat{w} - w^*\|_{\nabla^2 L(w^*)})^4 \nabla^2 L(\hat{w}) \preceq \nabla^2 L(\tilde{w}), \tag{50}$$

where the last inequality holds because $\tilde{w}$ lies on the line between $\hat{w}$ and w^* and $\nabla^2 L(w^*) \succ 0$ by Assumption 4. Multiply $\hat{w} - w^*$ to both sides and we get

$$\begin{aligned} \|\hat{w} - w^*\|^2_{\nabla^2 L(\hat{w})} &\leq \frac{8\sigma^2 \log(|\mathcal{W}|/\delta)}{n(1 - \|\hat{w} - w^*\|_{\nabla^2 L(w^*)})^4} \\ &\leq \frac{128\sigma^2 \log(|\mathcal{W}|/\delta)}{t}, \end{aligned}$$

where the first inequality holds because of Lemma 20 and the second inequality is due to Lemma 21. ■

Appendix D. Missing Proofs in Main Paper

Lemma 24 (Restatement of Lemma 7) *Suppose Assumption 1 3 hold. Fix $\delta \in (0,1)$, after round t, then w.p. $> 1-\delta$, the MLE-estimated $\hat{w}$ satisfies*

$$\frac{1}{t}\sum_{i=1}^{t} \mathbb{E}_{x\sim\mathcal{D}_i}(f_x(w^*) - f_x(\hat{w}))^2 \leq \frac{4\sigma^2 \log(|\mathcal{W}|/\delta)}{t}. \tag{51}$$

Proof Combine Lemma 16 in Appendix B with the Chernoff method and we obtain an exponential tail bound: w.p. $> 1-\delta$,

$$-\log \mathbb{E}_{D'} \exp(Q(\hat{p}(D), D')) \leq -Q(\hat{p}(D), D) + \log(|\mathcal{P}|/\delta), \tag{52}$$

$$-\log \mathbb{E}_{D'} \exp(Q(\hat{p}(D), D')) \leq -Q(\hat{p}(D), D) + \log(|\mathcal{W}|/\delta), \tag{53}$$

where the second inequality is due to the fact each parameter $w \in \mathcal{W}$ determines a probability function $p \in \mathcal{P}$. Now we set

$$Q(\hat{p}, D) = \sum_{i=1}^{t} -\frac{1}{2} \log \Big(\frac{p_{w^*}(y_i|x_i)}{p_{\hat{w}}(y_i|x_i)} \Big), \tag{54}$$

where $p_{w^*}(\cdot|x), p_{\hat{w}}(\cdot|x)$ denote the probability density function of $\mathcal{N}(f_x(w^*), \sigma^2), \mathcal{N}(f_x(\hat{w}), \sigma^2)$, respectively. With this choice, RHS of eq. (53) is

$$\sum_{i=1}^{t} \frac{1}{2} \log \Big(\frac{p_{w^*}(y_i|x_i)}{p_{\hat{w}}(y_i|x_i)} \Big) + \log(|\mathcal{W}|/\delta) \leq \log(|\mathcal{W}|/\delta), \tag{55}$$

since $p_{\hat{w}}$ is the empirical maximum likelihood estimator. And LHS of eq. (53) is

$$\begin{aligned} & -\log \mathbb{E}_{D'} \Bigg[\exp \Bigg(\sum_{i=1}^{t} -\frac{1}{2} \log \Big(\frac{p_{w^*}(y_i|x_i)}{p_{\hat{w}}(y_i|x_i)} \Big) \Bigg) \Bigg| D \Bigg] \\ = & -\sum_{i=1}^{t} \log \mathbb{E}_{x,y\sim\mathcal{D}_i} \exp \Bigg(-\frac{1}{2} \log \Big(\frac{p_{w^*}(y|x)}{p_{\hat{w}}(y|x)} \Big) \Bigg), \end{aligned} \tag{56}$$

where eq. (56) holds because $p_{\hat{w}}$ is independent of dataset D'. Then LHS of eq. (53) becomes

$$\begin{aligned} -\sum_{i=1}^{t} \log \mathbb{E}_{x,y\sim\mathcal{D}_i} \sqrt{\frac{p_{\hat{w}}(y|x)}{p_{w^*}(y|x)}} &= -\sum_{i=1}^{t} \mathbb{E}_{x,y\sim\mathcal{D}_i} \log \sqrt{\frac{p_{\hat{w}}(y|x)}{p_{w^*}(y|x)}} && (57) \\ &= -\sum_{i=1}^{t} \int_{x\sim\mathcal{D}_i} p_{w^*}(\cdot|x) \log \sqrt{\frac{p_{\hat{w}}(\cdot|x)}{p_{w^*}(\cdot|x)}} \mathrm{d}x && (58) \\ &= \frac{1}{2} \sum_{i=1}^{t} \int_{x\sim\mathcal{D}_i} p_{w^*}(\cdot|x) \log \frac{p_{w^*}(\cdot|x)}{p_{\hat{w}}(\cdot|x)} \mathrm{d}x && (59) \\ &= \frac{1}{2} \sum_{i=1}^{t} \mathbb{E}_{x\sim\mathcal{D}_i} \mathrm{KL}(\mathcal{N}(f_x(w^*), \sigma^2) || \mathcal{N}(f_x(\hat{w}), \sigma^2)) && (60) \\ &= \frac{1}{2} \sum_{i=1}^{t} \mathbb{E}_{x\sim\mathcal{D}_i} \frac{(f_x(w^*) - f_x(\hat{w}))^2}{2\sigma^2}, && (61) \end{aligned}$$

where eq. (58) is due to definition of expectation, eq. (60) is due to definition of KL divergence, and eq. (61) is due to Lemma 18 is Appendix B. Therefore, combining LHS and RHS of eq. (53), we have

$$\frac{1}{4t\sigma^2} \sum_{i=1}^{t} \mathbb{E}_{x\sim\mathcal{D}_i} (f_x(w^*) - f_x(\hat{w}))^2 \leq \frac{\log(|\mathcal{W}|/\delta)}{t}. \tag{62}$$

The proof completes by rearrangement. ■

Theorem 25 (Restatement of Theorem 8) *Suppose Assumption 1 3 4 hold. Fix $\delta \in (0,1)$, after sampling t points where t satisfies eq. (8), then w.p. $> 1-\delta$, the MLE-estimated $\hat{w}$ satisfies*

$$\|\hat{w} - w^*\|_2^2 \le \frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}. \tag{63}$$

Proof First we use minimum eigenvalue to upper bound $\|\hat{w} - w^*\|_2^2$:

$$\|\hat{w} - w^*\|_2^2 \le \frac{\|\hat{w} - w^*\|_{\nabla^2 L(\hat{w})}^2}{\lambda_{\min}(\nabla^2 L(\hat{w}))}. \tag{64}$$

Then the proof is completed by Corollary 22 and Lemma 23. ■

Theorem 26 (Restatement of Theorem 10) *Suppose Assumption 1 3 4 6 hold. After uniformly querying n data points in Phase I where*

$$n \ge \max\left\{\frac{2^{3+\gamma}\sigma^2 \log(|\mathcal{W}|/\delta)}{\alpha\mu^{\gamma/2}}, \frac{16\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu^2}\right\}. \tag{65}$$

Then at round t in Phase II, fix $\delta \in (0,1)$, w.p. $> 1-\delta$, $\forall x \in \mathcal{X}$,

$$|f_x(\hat{w}_t) - f_x(w^*)| \le \|\nabla f_x(\hat{w}_t)\|_2 \sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}} + \frac{256C\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}. \tag{66}$$

Proof With the C-smoothness Assumption (Assumption 6), then $\forall x \in \mathcal{X}$,

$$\begin{aligned} |f_x(\hat{w}_t) - f_x(w^*)| &\le |(\hat{w}_t - w^*)^\top \nabla f_x(\hat{w}_t)| + \frac{C}{2}\|\hat{w}_t - w^*\|_2^2 && (67) \\ &\le \|\hat{w}_t - w^*\|_2 \|\nabla f_x(\hat{w}_t)\|_2 + \frac{C}{2}\|\hat{w}_t - w^*\|_2^2, && (68) \end{aligned}$$

where the last inequality holds due to Holder's inequality. The proof completes by plugging in Theorem 8. ■

Lemma 27 (Instant regret) *Suppose Assumption 1 3 4 6 hold. In Phase I the number of uniform queries n satisfies eq. (9). Then fix $\delta \in (0,1)$, at round t in Phase II, w.p. $> 1-\delta$, $\forall x \in \mathcal{X}$,*

$$r_t \le 2\|\nabla f_x(\hat{w}_t)\|_2 \sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}} + \frac{512C\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}. \tag{69}$$

Proof By definition of instant regret,

$$\begin{aligned}
r_t &= f_{w^*}(x^*) - f_{w^*}(x_t) && (70)\\
&\leq f_{\hat{w}}(x^*) - f_{w^*}(x_t) + \|\nabla f_x(\hat{w}_t)\|_2 \sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}} + \frac{256C\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t} && (71)\\
&\leq f_{\hat{w}}(x_t) - f(x_t) + \|\nabla f_x(\hat{w}_t)\|_2 \sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}} + \frac{256C\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t} && (72)\\
&\leq 2\|\nabla f_x(\hat{w}_t)\|_2 \sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}} + \frac{512C\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}, && (73)
\end{aligned}$$

where the first and third inequalities are due to Theorem 10 and the second inequality is due to the selection criterion of x_t. ■

Theorem 28 (Restatement of Theorem 12) *Suppose Assumption 1 3 4 6 hold. After uniformly querying n data points in Phase I where*

$$n \geq \max\left\{\frac{2^{3+\gamma}\sigma^2 \log(|\mathcal{W}|/\delta)}{\alpha\mu^{\gamma/2}}, \frac{16\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu^2}\right\}. \quad (74)$$

Then fix $\delta \in (0,1)$, after T rounds in total, w.p. $> 1-\delta$,

$$R_T \leq nF + 2G\sqrt{T-n}\sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)\log(T/n)}{\mu}} + \frac{512C\sigma^2 \log(|\mathcal{W}|/\delta)\log(T/n)}{\mu}, \quad (75)$$

where $\alpha, \gamma, \mu, F, G, C$ are constants. Moreover, by choosing $n = \Theta(\sqrt{T})$, $R_T \leq \tilde{O}(\sqrt{T\log(T)})$.

Proof By definition of cumulative regret, plug in Lemma 27,

$$\begin{aligned}
R_T &= \sum_{t=1}^{n} r_t + \sum_{t=n+1}^{T} r_t && (76)\\
&\leq nF + \sum_{t=n+1}^{T} 2\|\nabla f_x(\hat{w}_t)\|_2 \sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}} + \frac{512C\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t} && (77)\\
&\leq nF + 2\underbrace{\sqrt{\sum_{t=n+1}^{T} \|\nabla f_x(\hat{w}_t)\|_2^2}}_{(i)}\underbrace{\sqrt{\sum_{t=n+1}^{T} \frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}}}_{(ii)} + \underbrace{\sum_{t=n+1}^{T} \frac{512C\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}}_{(iii)}, && (78)
\end{aligned}$$

where the last inequality follows the Cauchy-Schwarz inequality. Next, we bound these three terms separately.

In (i), by Assumption 6,

$$(i) = \sqrt{\sum_{t=n+1}^{T} \|\nabla f_x(\hat{w}_t)\|_2^2} \leq \sqrt{\sum_{t=n+1}^{T} G^2} \leq G\sqrt{T-n}. \tag{79}$$

To bound (ii), we use $\sum_{i=a}^{b} 1/i \leq \log(b/a)$ and we have

$$(ii) = \sqrt{\sum_{t=n+1}^{T} \frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t}} \leq \sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)\log(T/n)}{\mu}}. \tag{80}$$

Again, we use the same trick to bound (iii),

$$(iii) = \sum_{t=n+1}^{T} \frac{512C\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu t} \leq \frac{512C\sigma^2 \log(|\mathcal{W}|/\delta)\log(T/n)}{\mu}. \tag{81}$$

Therefore,

$$\begin{aligned} R_T &\leq nF + 2G\sqrt{T-n}\sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)\log(T/n)}{\mu}} + \frac{512C\sigma^2 \log(|\mathcal{W}|/\delta)\log(T/n)}{\mu} && (82) \\ &\leq \tilde{O}(n + \sqrt{(T-n)\log(T/n)} + \sqrt{\log(T/n)}). && (83) \end{aligned}$$

When n is chosen as $n = \Theta(\sqrt{T})$, the cumulative regret bound is

$$R_T \leq \tilde{O}(\sqrt{T\log(T)}). \tag{84}$$

■

Lemma 29 (Restatement of Lemma 15) *Suppose Assumption 1 3 4 6 hold. In Phase I the number of queries n satisfies eq. (9). Then fix $\delta \in (0,1)$, after T rounds in total, w.p. $> 1-\delta$,*

$$R_T \leq nF + 2(T-n)G\sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu n}}. \tag{85}$$

Moreover, by choosing $n = \Theta(T^{2/3})$, $R_T \leq \tilde{O}(T^{2/3})$.

Proof Follow the definition of cumulative regret,

$$
\begin{aligned}
R_T &= \sum_{t=1}^{n} r_t + \sum_{t=n+1}^{T} r_t && (86)\\
&\leq nF + \sum_{t=n+1}^{T} f_{w^*}(x^*) - f_{w^*}(\hat{x}_t) && (87)\\
&\leq nF + \sum_{t=n+1}^{T} f_{w^*}(x^*) - f_{\hat{w}}(x^*) + f_{\hat{w}}(x^*) - f_{\hat{w}}(\hat{x}_t) + f_{\hat{w}}(\hat{x}_t) - f_{w^*}(\hat{x}_t) && (88)\\
&\leq nF + \sum_{t=n+1}^{T} f_{w^*}(x^*) - f_{\hat{w}}(x^*) + f_{\hat{w}}(\hat{x}_t) - f_{w^*}(\hat{x}_t), && (89)
\end{aligned}
$$

where F is a constant bounding the function range and the second inequality is due to optimal choice of $\hat{x}_t$ in UCB. Then by Lipschitz assumption (Assumption 6),

$$
\begin{aligned}
R_T &\leq nF + 2(T-n)G\|\hat{w} - w^*\|_2 && (90)\\
&\leq nF + 2(T-n)G\sqrt{\frac{512\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu n}} && (91)\\
&\leq \tilde{O}\Big(n + \frac{T}{\sqrt{n}} - \sqrt{n}\Big) && (92)
\end{aligned}
$$

The bound is minimized by setting $n = \Theta(T^{2/3})$, which leads to

$$R_T \leq \tilde{O}(T^{2/3}). \tag{93}$$

■

Appendix E. Discussion on the *Bonus* term

Define the empirical Hessian matrix of the estimated parameter $\hat{w}$ as follows.

$$\hat{H}(\hat{w}) = \nabla^2 \hat{L}(\hat{w}) = \frac{1}{t}\sum_{i=1}^{t} \nabla^2 (f_{x_i}(\hat{w}) - y_i)^2. \tag{94}$$

Following Theorem 10, there is another way to establish an upper confidence bound.

Lemma 30 (Upper confidence bound) *Suppose Assumption 1 3 4 6 hold. After uniformly querying n data points in Phase I where*

$$n \geq \max\Big\{\frac{2^{3+\gamma}\sigma^2 \log(|\mathcal{W}|/\delta)}{\alpha\mu^{\gamma/2}}, \frac{16\sigma^2 \log(|\mathcal{W}|/\delta)}{\mu^2}\Big\}. \tag{95}$$

Then at round t in Phase II, fix $\delta \in (0,1)$, w.p. $> 1-\delta$, $\forall x \in \mathcal{X}$,

$$|f_x(\hat{w}_t) - f_x(w^*)| \leq \|\hat{w}_t - w^*\|_{\hat{H}} \|\nabla f_x(\hat{w}_t)\|_{\hat{H}^{-1}} + \frac{C\|\hat{w}_t - w^*\|_{\hat{H}}^2}{2\sqrt{\lambda_{\min}(\hat{H})}}. \tag{96}$$

Proof With the C-smoothness Assumption (Assumption 6), then $\forall x \in \mathcal{X}$,

$$|f_x(\hat{w}_t) - f_x(w^*)| \leq |(\hat{w}_t - w^*)^\top \nabla f_x(\hat{w}_t)| + \frac{C}{2}\|\hat{w}_t - w^*\|_2^2 \tag{97}$$

$$\leq \|\hat{w}_t - w^*\|_{\hat{H}} \|\nabla f_x(\hat{w}_t)\|_{\hat{H}^{-1}} + \frac{C\|\hat{w}_t - w^*\|_{\hat{H}}^2}{2\sqrt{\lambda_{\min}(\hat{H})}}, \tag{98}$$

where the second inequality holds due to Holder's inequality. ■

With Lemma 30, at round t, if we are able to derive $B_U(t)$ to upper bound $\|\hat{w} - w^*\|_{\hat{H}}$ and $B_L(t)$ to lower bound $\lambda_{\min}(\hat{H})$ where both $B_U(t)$ and $B_L(t)$ depend on t, we can design a new acquisition function:

$$x_t = \operatorname*{argmax}_{x \in \mathcal{X}} f_x(\hat{w}_t) + B_U(t)\|\nabla f_x(\hat{w}_t)\|_{\hat{H}^{-1}} + \frac{CB_U^2(t)}{2\sqrt{B_L(t)}}. \tag{99}$$

Unlike the acquisition function in Step 3 of Algorithm 1 where bonus term $\|\nabla f_x(\hat{w}_t)\|$ only depends on $\hat{w}_t$, in this new acquisition function (eq. (99)), the new bonus term $\|\nabla f_x(\hat{w}_t)\|_{\hat{H}^{-1}}$ depends on not only $\hat{w}_t$ but also $\hat{H}$. By definition of $\hat{H}$, it contains rich information from all previous rounds. Potentially, it is able to result in faster convergence of the new GO-UCB algorithm.