HIGH DIMENSIONAL BAYESIAN OPTIMIZATION WITH REINFORCED TRANSFORMER DEEP KERNELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Bayesian Optimization (BO) has proved to be an invaluable technique for efficient, high-dimensional optimization. Gaussian process (GP) surrogates and dynamic acquisition functions have enabled BO to perform well in challenging high dimensional optimization due to its sample efficiency and uncertainty modeling. In earlier research, Reinforcement Learning was introduced to improve optimization performance on both single function optimization and *few-shot* multi-objective optimization. However, until now, even few-shot techniques treat each objective as independent optimization tasks, failing to exploit the similarities shared between objectives. In this paper, we combine recent developments in Deep Kernel Learning (DKL) and attention-based Transformer models to improve the modeling powers of GP surrogates with meta-learning. We propose a novel method for improving meta-learning BO surrogates by incorporating attention mechanisms into DKL, empowering the surrogates to adapt to contextual information gathered during the BO process. We combine this Transformer Deep Kernel with a learned acquisition function using Soft Actor-Critic Reinforcement Learning to aid in exploration. This Reinforced Transformer Deep Kernel (RTDK) approach produces state-of-the-art results on various high dimensional optimization problems.

1 INTRODUCTION

Sample efficient and high dimensional optimization is at the core of many industrial and scientific processes with applications including material design (Zhang et al., 2020), physics (Carr et al., 2016), synthetic chemistry and biology (Shields et al.; Barnes et al., 2011), and hyper-parameter optimization (Snoek et al., 2012). In many cases, we are not just interested in optimizing a single function or process but in optimizing many related problems, each for a slightly different configuration of a core process. Bayesian Optimization (BO) is a ubiquitous technique that has proven to be very promising across all of the domains listed above and is often the standard for sample-efficient block-box optimization (Frazier, 2018). However, Bayesian Optimization relies on the design of a *surrogate model* which estimates the optimization objective in yet unexplored regions, as well as an *acquisition function* for effectively exploring the optimization domain. Both of these components typically require domain knowledge to adapt to specific optimization objectives, and this design is critical to BO's performance on challenging domains.

In this work, we aim to improve Bayesian Optimization's generalizability and sample efficiency by introducing deep learning methods into both these components. We use contextual deep learning models to learn a general surrogate model which can effectively meta-learn the objective from very few samples. Additionally, we improve the acquisition function through Reinforcement Learning to dynamically adapt the exploration-exploitation trade-off.

We evaluate these methods on both discrete and continuous optimization domains, focusing on families of high-dimensional continuous optimization objectives. We show a significant improvement in both sample efficiency to reach an optimization target and a better final optimization value.

2 RELATED WORK

Reinforcement Learning Acquisitions Reinforcement learning (RL) approaches to Bayesian optimization have recently shown promising results, especially for discrete objectives. MetaBO

presents a seminal framework for interpreting the acquisition function within Bayesian optimization as a reinforcement learning policy, to be trained with policy gradient methods (Volpp et al., 2020). This work has been further generalized in Hsieh et al. (2021) to allow for few-shot q-learning instead of a discrete policy optimization. These approaches present useful frameworks for tackling small sample Bayesian optimization, but they primarily focus on the acquisition function, leaving the surrogate model as a traditional Gaussian process. Additionally, both approaches rely on a discrete grid to perform optimization, approximating continuous domain optimization with a quasi-random hierarchical grid.

Model-Based RL Model based reinforcement learning presents a method for improving the sample efficiency of RL algorithms by estimating the environment's dynamics and allowing the policy to train from the information learned by such a model (Sutton, 1991). Additionally, it has been recently shown that the model does not necessarily have to predict the exact dynamics but can instead predict a latent representation of the dynamics, which may include more information than the observations (Oh et al., 2017). We are interested in incorporating the sample-efficiency of model-based RL methods to improve black-box optimization methods.

Soft Actor-Critic Reinforcement learning in continuous action spaces presents several unique challenges compared to discrete action spaces such as exploration and policy representation. Since sample efficiency is one of the primary focuses of this work, we use soft actor-critic (SAC) (Haarnoja et al., 2018), an off-policy RL algorithm, as our foundational algorithm. We will extend SAC to include the improvement from model based methods and combine its actor, critic, and model to construct a robust acquisition function for Bayesian Optimization.

Attention and Transformers Attention mechanisms (Luong et al., 2015) provide a method for deep neural networks to modify their activations in response to a set of contextual vectors. This allows us to condition a parameterized neural network on an arbitrary number of unordered vectors. Attention has been used in various network architectures to achieve state-of-the-art results in many applications including natural language processing (Vaswani et al., 2017) and computer vision Dosovitskiy et al. (2021). We will use the contextual embeddings of transformers to assist in rapidly optimizing black box functions from observations on similarly structured objectives.

3 Bayesian Optimization

Bayesian Optimization (BO) corresponds to a general set of techniques for optimizing a black-box function $f(x) : \mathbb{X} \to \mathbb{Y}$ by: fitting a *surrogate model* to estimate function values across the optimization domain; and employing an *acquisition function* based on the surrogate to select promising query points (Frazier, 2018).

The surrogate model, $\hat{f}(x; x_{train}, y_{train}) = \hat{f}(x; \mathcal{D})$, provides a probabilistic estimate of the objective across the entire optimization domain given a sparse sample of points $x_{train} = \{x_1, x_2, \ldots, x_k\}$ where the objective is known $y_{train} = \{y_1, y_2, \ldots, y_k\}$. Surrogates typically provide both a mean estimate of the function value $\mu(x; \mathcal{D})$, as well as an uncertainty for that estimate in the form of a variance $\sigma^2(x; \mathcal{D})$.

After fitting the surrogate on the observed dataset $\mathcal{D}$, the acquisition function, $\mathcal{A}(x; \mathcal{D})$, defines a score for selecting the next query point x_{query}. Bayesian optimization selects queries by maximizing the acquisition $x_{query} = \arg\max_{x \in \mathbb{X}} \mathcal{A}(x; \mathcal{D})$. Therefore, the acquisition must be responsible for balancing exploration to ensure a global optimum across the domain and exploitation to optimize locally within a promising region. One common acquisition function, especially with Gaussian Process surrogates, is the Expected Improvement (EI) criterion (Jones et al., 1998).

In this work, we aim to improve both aspects of BO by introducing deep neural networks to both the surrogate and acquisition functions while maintaining the generality of the BO approach. These improvements will minimize required training data while ensuring that these methods work in both continuous and discrete optimization domains.

4 CONDITIONAL DEEP KERNEL SURROGATE

The Gaussian Process (GP) (Rasmussen & Williams, 2005) is a fundamental architecture for BO surrogate models (Frazier, 2018). GPs estimate the objective with a Gaussian posterior over functions fitting the observed data: $\hat{f}(y|x;\mathcal{D}) \sim \mathcal{N}(\mu(x;\mathcal{D}), \sigma^2(x;\mathcal{D}))$. We use a learned mean component which is parameterized by a linear transformation of the input $\mu_W(x) = Wx$. In general, both the mean and covariance components may have parameters that must be learned. The parameters of a Gaussian process are trained to maximize the log-likelihood of the training dataset $\sum_{x,y\in\mathcal{D}} \log P(\hat{f}(y;x,\mathcal{D}) = y)$.

The GP covariance is determined by a *Kernel*, $K(x_1, x_2)$, which defines the distance between any two points within a high dimensional (sometimes infinite) manifold. This kernel representation, along with the Gaussian likelihood, defines an analytical posterior distribution over the optimization domain given a set of observed function values. The choice of kernel function is crucial for well-fitting GPs, and many domains have specially designed kernel functions to fit domain-specific data.

4.1 COMBINATION KERNEL

In the interest of generality, we want a consistent architecture that works reasonably well across various domains. A common kernel in traditional GPs is the RBF, which corresponds to a dot-product in an infinite dimensional manifold and may be defined as an infinite sum of polynomial kernels (Rasmussen & Williams, 2005).

$$K_{RBF}(x_1, x_2) \propto \sum_{k=1}^{\infty} \frac{\langle x_1, x_2 \rangle^k}{k!}$$

Taking inspiration from this representation, we choose to instead approximate this high dimensional embedding up to a certain power, K, learning the coefficients, α_k, as part of the kernel's parameters. We call this kernel the *Combination* up to K, or $C(K)$, kernel.

$$K_{C(K)}(x_1, x_2) = \sum_{k=1}^{K} \frac{\alpha_k \langle x_1, x_2 \rangle^k}{k!} \tag{1}$$

We find this kernel provides additional learning potential over a fixed RBF kernel. A hyperparameter selects the maximum polynomial.

4.2 DEEP KERNEL LEARNING

Deep Kernel Learning (DKL) (Wilson et al., 2015; Ober et al., 2021) extends the learning capability of the GP by mapping the original optimization domain $x \in \mathbb{X}$ into a new domain via a parameterized transformation such as a deep neural network $z = g_\theta(x) \in g_\theta(\mathbb{X})$. The underlying Gaussian process is then trained on these embedding inputs after neural network reprocessing: $\hat{f}(y|z; g(\mathcal{D})) \sim \mathcal{N}(\mu(z), \sigma^2(z))$. The GP parameters are optimized via log-likelihood minimization, and the gradients are passed along to the embedding g_θ, optimizing θ through back-propagation.

We find that DKL sometimes experiences numerical instability, especially when training for Bayesian Optimization tasks, which typically have very little data. We, therefore, introduce an additional parameterized neural network, $v_\phi(x)$, to explicitly estimate the diagonal components of the covariance, while using a normalized variant (Rasmussen & Williams, 2005) of the kernel, $K(x_1, x_2)$, to estimate the non-diagonal components. This allows us to use a flexible kernel while avoiding large values due to a poorly conditioned embedding network.

$$K_{normalized}(x_1, x_2) = \exp v_\phi(x_1) \cdot \exp v_\phi(x_2) \cdot \frac{K(x_1, x_2)}{K(x_1, x_1)K(x_2, x_2)}$$

4.3 TRANSFORMER DEEP KERNEL LEARNING (TDKL)

Transformers present a mechanism for adding arbitrary context sequences to condition neural network activations. Crucially for BO, since the context does not necessarily require a unique target, it

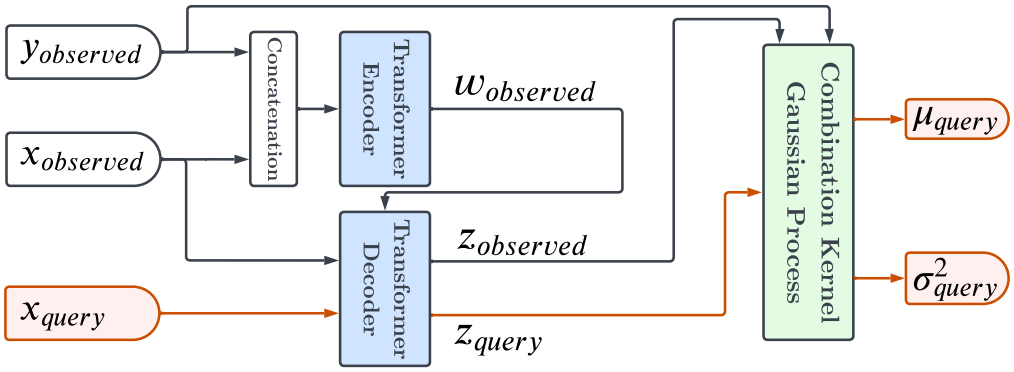


Figure 1: Block Diagram of Transformer Deep Kernel Learning Gaussian Process. The red path indicates the prediction path for the query point. The black paths indicate the contextual information.

may include additional information that is not present in traditional BO observations. Specifically, we include not just the previous sample points $x_{observed}$ in the context but also the known function values for those points $y_{observed}$. This mechanism has been shown to be sufficient for Bayesian inference by itself (Müller et al., 2021) and we follow a similar framework for conditioning a DKL embedding on previously observed data.

Formally, we extend the DKL framework to include a conditioning term on the embedding network, $z_{query} = g_\theta(x_{query}|x_{observed}, y_{observed})$, where g is a sequence-to-sequence transformer encoder-decoder model (Vaswani et al., 2017). The observed data, $(x_{observed}, y_{observed}) = \{(x_1, y_1), (x_2, y_2), \ldots, (x_K, y_K)\}$, is first fed through the transformer encoder to produce the latent encoded sequence $w_{observed} = \{w_1, w_2, \ldots, w_K\}$. This is used as the keys and values for the decoder, whereas the target sequence $x_{query} = \{x_{K+1}, x_{K+2}, \ldots, x_N\}$ is used as the query for the decoder. The output sequence, $z_{query} = \{z_{K+1}, z_{K+2}, \ldots, z_N\}$, represents a conditional embedding of the query locations. We also produce the conditional embedding of the original observed locations, $z_{observed} = \{z_1, z_2, \ldots, z_K\}$, to condition the downstream Gaussian Process. The output distribution is parameterized by our GP, $\hat{f}(y|z_{query}; z_{observed}, y_{observed})$. See Figure 1 for a flow diagram for all inputs. Similarly to (Müller et al., 2021), we remove the temporal embedding from the input to ensure that the transformer is invariant to sequence order. Additionally, we ensure that query points do not attend to each other by enforcing a diagonal attention mask on the decoder query sequence.

We refer to this complete surrogate model - employing a transformer embedding, learned pointwise variances, and a combination base kernel - as the Transformer Deep Kernel Learning (TDKL) surrogate. Unlike (Müller et al., 2021), we focus this architecture on sample efficiency, relying on the Gaussian process mechanism to encode the uncertainty in our predictions. This design introduces an inductive bias towards smooth functions, which may be represented by our combination kernel GP within the embedding space, but we find that this assumption does limit our performance due to the learning potential of the transformer.

5 Soft Actor-Critic Acquisition

We turn our focus to the BO acquisition function. The TDKL described above provides a robust surrogate model which can estimate a function given a sufficiently robust set of samples. To collect such samples, we need a policy that will sufficiently explore the domain. For this purpose, we will employ a soft actor-critic (SAC) reinforcement learning agent (Haarnoja et al., 2018). We will represent the Bayesian optimization problem as a Markov decision process and use this formulation to train a novel model-based soft actor-critic agent

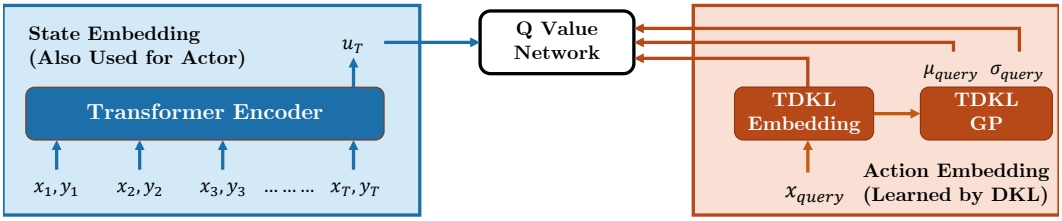


Figure 2: Flow diagram for the critic network, showing how the state-action pair is embedded and processed into a Q-Value estimate.

5.1 Optimization Environment

Following the representations presented in MetaBO (Volpp et al., 2020) and FSAF (Hsieh et al., 2021), the state space for the problem at any time t of the optimization process is defined as the collection of points and values in the BO trajectory, $s = (x_{observed}, y_{observed}) = \{(x_1, y_1), (x_2, y_2), \ldots, (x_t, y_t)\}$. The action space corresponds to the space of possible locations where we can sample defined by the input domain $\mathbb{X}$, and action represents the next candidate point, $a_t = x_{t+1} \in \mathbb{X}$. We define the reward in terms of the (approximate) regret, $r_{t+1} = -log(f^* - f(x_{t+1}))$, where f^* is the estimate true optimum for the function. We consider finite-length trajectories determined by the budget for the number of steps in each trajectory $t < T_{max}$.

In general, each trajectory may originate from a different underlying objective. This enables meta-learning between similar objectives as both the agent and model must perform well across many different trajectories. We may also collect additional trajectories from the same environment if we wish to continue optimizing the given objective. Sampled trajectories are stored in a replay buffer, separated by their objective variant.

5.2 Model-Based Soft Actor-Critic

Following the soft actor-critic framework, our agent consists of two learned Q-value network $q_1(s, a)$ and $q_2(s, a)$, a conservative estimate of the q value $q(s, a) = \min\{q_1(s, a), q_2(s, a)\}$, and a probabilistic policy $\pi(s)$. We extend SAC to a model-based reinforcement learning method by introducing the learned surrogate model into the framework.

Surrogate Model The surrogate model $\hat{f}$ will be trained on function samples from the replay buffer. After sampling a trajectory $s = \{(x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T)\}$, we shuffle this trajectory and arbitrarily split the x and y values into *observation* and *query* datasets using a uniform random splitting pivot $M \sim \mathcal{U}(1, T-1)$. These datasets, $\mathcal{D}_{observed} = \{(x_1, y_1), \ldots, (x_M, y_M)\}$ and $\mathcal{D}_{query} = \{(x_{M+1}, y_{M+1}), \ldots, (x_T, y_T)\}$ are then used to optimize the surrogate on purely observed data.

$$\mathcal{L}_{model} = \sum_{(x_{query}, y_{query}) \in \mathcal{D}_{query}} - \log P(\hat{f}(y; x_{query}, \mathcal{D}_{observed}) = y_{query}) \tag{2}$$

Notice that this loss function differs from the traditional GP optimization because the training dataset which conditions our TDKL is different than the prediction dataset. This is to ensure that the TDKL learns a general representation regardless of which objective originated the trajectory. Optimizing this loss with a stochastic gradient-descent optimizer, sampling a different trajectory after each step, presents a cheap method for meta-learning GP models on a variety of objectives.

Critics We train the dual critic networks using the standard entropy-corrected Bellman update described in (Haarnoja et al., 2018). However, we wish to include information learned by the surrogate in the critic networks to improve sample efficiency. We accomplish this by exploiting the learned embeddings of the TDKL and the GP estimates of the objective.

The state, s, is embedded using another transformer architecture (Vaswani et al., 2017). This time, the transformer is acting only as an encoder to transform the variable-length observations, $s =$

$\{(x_1, y_1), \ldots, (x_T, y_T)\}$, into a fixed-size latent embedding. We encode the sequence into a latent sequence using the transformer encoder $\{u_1^{critic}, \ldots, u_T^{critic}\} = TransformerEncoder(s)$ and then simply take the final latent vector, u_T^{critic}, as the fixed-length embedding.

The action, a, is encoded by passing the suggested point through the TDKL to extract both the embedding and function estimates from the surrogate model. $w_a = g_\theta(a|s))$ represents the embedding from the TDKL transformer g_θ, and $\mu_a, \sigma_a^2 = \hat{f}(a|w_a; s)$ are the surrogate model estimates for the objective at the action location.

The Q-networks through the state-action representation, $q(u_T^{critic}, w_a, \mu_a, \sigma_a^2)$, allow information to be shared between the model and critics. Note that TDKL parameters are treated as constant w.r.t the critic, and the gradient is not passed to the TDKL when optimizing the Bellman loss. A diagram of the critic architecture is presented in Figure 2.

Actor The actor network, $\pi(s)$, uses the same architecture as the state encoding from the Q-networks (The left-hand component in Figure 2). We use a separately trained transformer encoder to construct a fixed-length embedding for the state representation. $u_T^{actor} \in TransformerEncoder(s)$. Following the methods described in (Haarnoja et al., 2018), we use a tanh-squashed normal for the actor distribution and apply an affine transform to fit the desired optimization domain.

5.3 Acquisition Exploration in Continuous Domains

We find that relying purely on the actor for selecting good actions while exploring performs poorly on the small sample counts found in Bayesian optimization. Therefore, instead of sampling the action directly from $a \sim \pi(s)$ like in traditional SAC, we would like to incorporate the Q networks into the policy.

To do this, we take inspiration from Boltzmann exploration (Cesa-Bianchi et al., 2017), a common exploration technique in discrete environments. This involves constructing a policy that will sample proportional to a Boltzmann distribution based on the Q-network, $P(a|s) \propto \exp Q(s, a)$. We perform this kind of Boltzmann sampling when optimizing discrete domains using the Softmax function, but this approach fails for continuous optimization domains.

Extending this to the continuous domain can be difficult because we cannot generally sample from arbitrary functions in high dimensions. However, if we believe that our actor generally learns the landscape of our Q-function, then we can sample from the actor in order to assist with generating samples from the Boltzmann Q using importance sampling.

First, we sample a large batch of actions from the policy, $\{a_1, a_2, \ldots, a_N\} \sim \pi(s)$, where N is typically in the thousands. Then, we compute the importance weights of each sampled points w.r.t the Boltzmann Q, $w_i = \frac{exp(Q(s,a_i))}{P(\pi(s)=a_i)}$. Finally, construct an empirical distribution over $\{a_1, a_2, \ldots, a_N\}$ using these importance weights and sample an action a such that $P(a = a_i) = \frac{w_i}{\sum_{k=1}^{N} w_k}$. As long as the actor has a non-zero probability of sampling anywhere in the optimization domain, then this process will produce samples from exactly the Boltzmann Q distribution as $N \to \infty$. In practice, to trade off memory and computation time, we use a value of $N = 1024$

While this process is quite slow, the priority on sample efficiency in the BO environment allows us to spend more computation time selecting each action when collecting objective samples. We use this sampling technique to select actions during inference. Unfortunately, this process is too slow to execute when performing gradient updates, so we use the regular $\pi(s)$ policy with the reparameterization trick to compute the actor loss during SAC training.

6 Experiments

We evaluate the RTDK Bayesian optimization approach on a variety of test functions. First, we confirm that this fully learnable approach can still achieve reliable results in simpler discrete optimization tasks. Then we explore the effectiveness of this fully continuous approach on high dimensional optimization problems.

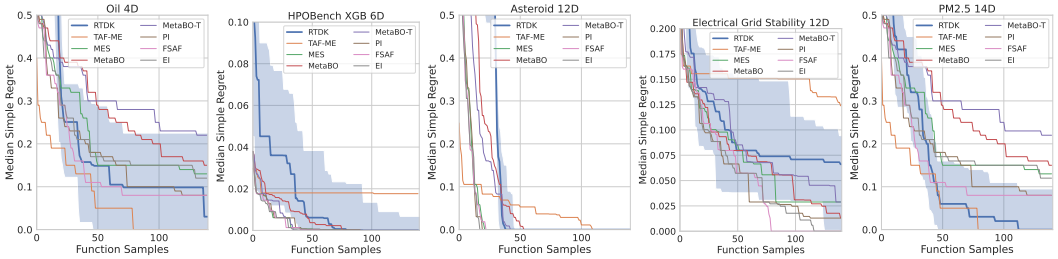


Figure 3: A comparison of different optimization methods on discrete domain hyper-parameter optimization objectives. TDKL results include a shaded region representing the IQR (25% - 75%) of achieved regret values.

We use the RTDK surrogate with continuous SAC acquisition as the model of study for all experiments. The RTDK uses a transformer DKL embedding and a combination kernel with 5 components $C(5)$. RTDK evaluation differs slightly from other BO methods because the transformer models must be trained on example trajectories. Therefore, we must present the model with similarly sized trajectories during both training and evaluation. To accommodate longer evaluation trajectories, we split the BO run into "sub-trajectories" of length 50, resetting the surrogate after 50 steps. This allows us to only train on trajectories up to length 50. Each sub-trajectory also begins by taking 5 uniform random function samples to initialize the RTDK surrogate. Additionally, we add a large action noise to the first sub-trajectory in order to encourage the RTDK to explore during this initial phase. We examine the effect of these sub-trajectories in Figure 6.

6.1 Discrete Optimization Baseline

We evaluate RTDK, as well as a variety of baselines, on discrete, real-world optimization tasks. In order aid in direct comparison, we evaluate this method on the same set of optimization problems and methods as Hsieh et al. (2021).

We compare against baseline acquisition based on a traditional Gaussian process including Expected Improvement (EI) (Močkus, 1975), Probability of Improvement (PI)(Kushner, 1964), Max-value entropy search (MES) (Wang & Jegelka, 2017), and TAF-ME (Wistuba et al., 2018). We also compare against deep learning acquisition functions FSAF (Hsieh et al., 2021) and MetaBO (Volpp et al., 2020). We allow the models which can perform meta-learning to pre-train on 250 function samples: 5 trajectories of 50 samples each. We then evaluated all of the methods on 36 additional variants and plot the median regret values for these evaluation runs in Figure 3.

RTDK had to be adapted to perform discrete optimization by limiting the acquisition function to only valid discrete locations. The SAC RL agent was still trained as a continuous acquisition function. This presents an opportunity to improve this method in the future since there is specialization of SAC to discrete action spaces (Christodoulou, 2019). However, we wanted to keep the method consistent across all experiments and employed the same continuous SAC for all cases.

We find that RTDK performs on par with other methods, showing very good performance on the higher dimensional PM2.5 data-set. We do find that baseline methods do perform better in lower dimensional tasks. We think this can be explained by the larger number of parameters and aggressive SAC exploration in RTDK. This may assist in challenging domains, but it can decrease performance in simpler domains. We nevertheless display reasonable performance on these simpler problems even without parameter tuning.

6.2 High Dimensional Continuous Domains

We perform a similar comparison of optimization methods in the continuous domain. For these experiments, we used the 10-dimensional variant of the Powell function for a direct comparison to Hsieh et al. (2021). We also evaluate on high dimensional variations of the Thompson Problem (F.R.S., 1904). This optimization objective consists of placing N electrons on the surface of a unit sphere, with the objective of minimizing the potential energy between all pairs of electrons.

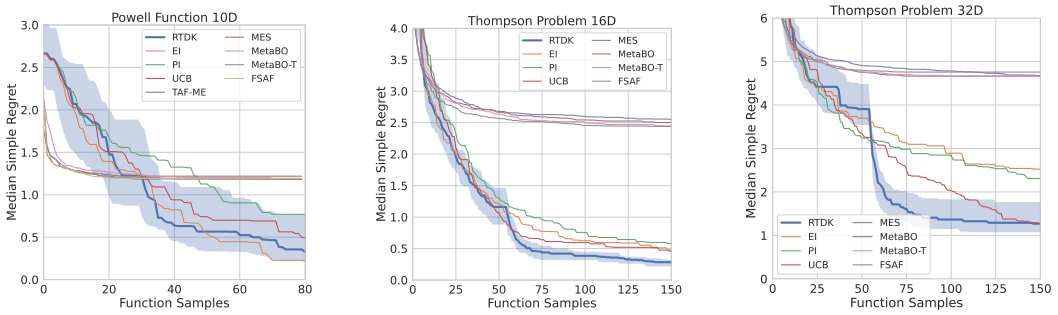


Figure 4: A comparison of different optimization methods on medium dimensional continuous optimization tasks. The final plot shows a comparison of FSAF across different dimensional optimization tasks.

We parameterize this function in $4N$ dimensions, storing the sin and cos of both the polar and azimuthal angle of each electron on the sphere. We construct the 16, 32, 48, and 64 dimensional variants of the function by taking random D-dimensional slices through an overarching $N = 32$ Thompson problem. Each slice is treated as a variant of the function, allowing us to make a near infinite amount of objectives with similar characteristics.

For this test, we continue using the discrete grid-based approach described in (Hsieh et al., 2021; Volpp et al., 2020) for baseline methods FSAF, MetaBO, MES, and TAF. This is because generalizing these methods to the continuous domain directly is challenging and Hsieh et al. (2021) recommends using a discrete grid of quasi-random Sobol samples for approximating continuous optimization. However, other baseline methods - EI, PI, and UCB - may be adapted for use with continuous Gaussian processes and optimization schemes. Therefore, we use a continuous Bayesian optimization routine for these baselines, along with the RTDK model.

Figure 4 presents results on the lower-dimensional continuous optimization tasks, demonstrating the limitations of the grid-based approximation for real-world problems. We find that TDKL consistently appears in the top scorers across these optimization domains, with better separation for the Thompson problem because it features more opportunities for meta-learning.

The combined benefits of the transformer surrogate and SAC acquisition truly shine through on the higher dimensional optimization problems presented in Figure 5. We found that the discrete methods failed to optimize within these high dimensional domains due to the exponential growth in required grid size to densely cover the domain. We also find that RTDK starts to break off from the baseline methods and effectively meta-learns on the 64-dimensional Thompson problem. Moreover, we observe a discrete phase transition between the first 50-sample sub-trajectory and the later sub-trajectories for the RTDK model. Due to the heightened exploration, we find that the first 50 samples are sub-optimal for the RTDK method before rapidly jumping to a better solution once the second sub-trajectory begins. We plan to look into this behavior to allow for a smoother transition, potentially improving sample efficiency.

6.3 Ablation Study

Sub-Trajectory Length We find that longer sub-trajectories assist with achieving lower final regret on the 10 Dimensional Powell function (Figure 6.a). However, this comes with a slight hit to convergence time, taking longer to achieve low regret earlier during the optimization. Additionally, sub-trajectories larger than 50 samples risked running out of memory as the transformer architecture memory requirements scale as n^2 with respect to the sequence length. However, we find that performance plateaus after length 30, so these longer trajectories are not necessary for optimization. For the purposes of the experiments, we used a trajectory length of 50 to ensure the lowest possible regret with reasonable memory usage. However, in practice, lower lengths may be used to improve performance time and improve performance in very sample-limited situations.

Single-Function DKL Architecture In Figure 6.b, we evaluate the effect of the DKL structure on optimizing the 10-dimensional Powell function, which changes very little with different variants

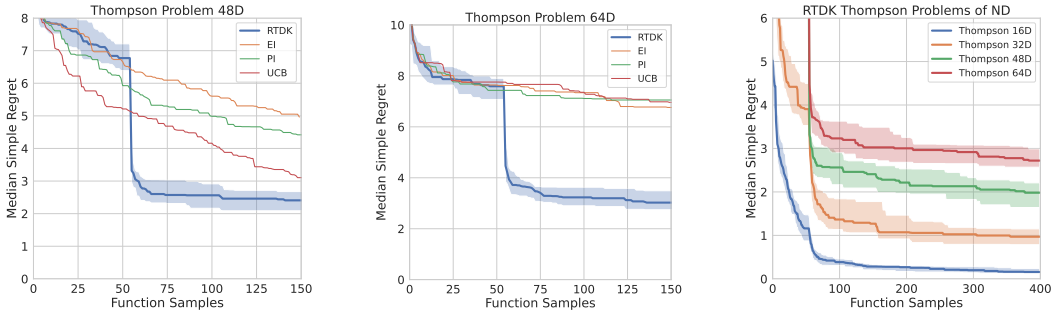
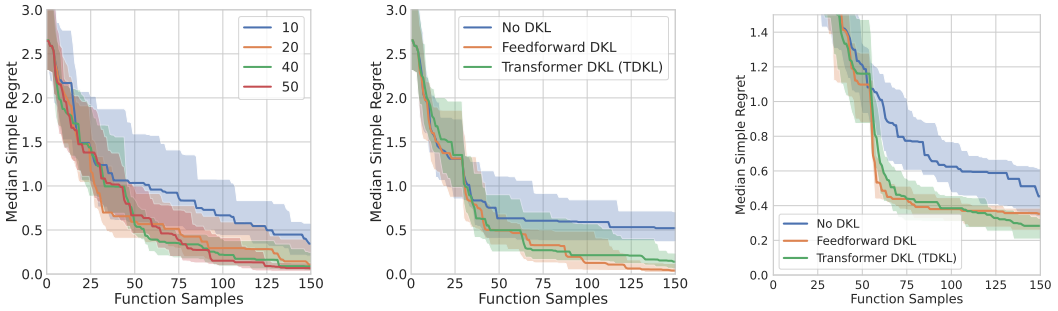


Figure 5: A comparison of different optimization methods on very high dimensional continuous optimization tasks. The final plot shows a comparison of RTDK across different dimensional Thompson optimization runs.



(a) Comparison of sub-trajectory length on regret in longer episodes.

(b) Comparison of DKL architecture in single-function optimization (Powell 10D).

(c) Comparison of DKL architecture in meta-learning optimization (Thompson 16D).

Figure 6: Ablation studies on TDKL model enabling and disabling various components.

(simple shift and translation). Overall, we find that the deep kernel architecture performs better than a simple Gaussian process. However, in this simpler case, the feed-forward, unconditional DKL performs slightly better. This is likely because it has fewer parameters and a simpler gradient than the transformer architecture. However, the difference is small so the transformer is still safe to use in unknown situations where the user is unsure of the degree to which different objectives are related to each other.

Meta-Learning DKL Architecutre We see better separation when evaluating the three different DKL approaches on the meta-learning task for the 16-dimensional Thompson problem. This objective has more meta-learning because each objective is a projection of a 128-dimensional Thompson problem projected onto a random 96-dimensional plane. This results in a wide variety of possible objective landscapes. We find that the Transformer DKL achieves lower final regret values when compared to the feed-forward approach.

7 Conclusion

We present novel contributions to two important aspects of black-box Bayesian optimization. We improve the surrogate model through contextual transformer deep kernel learning and a normalized combination base kernel, extending BO methods to higher dimensional problems such as the 64-dimensional Thompson problem. We design a model-based soft actor-critic to train an acquisition function with reinforcement learning, extending RL Bayesian optimization methods to continuous domains and problems. This two-fold approach could extend deep-learning enhanced Bayesian optimization methods to high dimensional, challenging black-box optimization in the physical sciences and machine learning.

REFERENCES

Chris P. Barnes, Daniel Silk, Xia Sheng, and Michael P. H. Stumpf. Bayesian design of synthetic biological systems. *Proceedings of the National Academy of Sciences*, 108(37):15190–15195, 2011. doi: 10.1073/pnas.1017972108. URL https://www.pnas.org/doi/abs/10.1073/pnas.1017972108.

Shane Carr, Roman Garnett, and Cynthia Lo. Basc: Applying bayesian optimization to the search for global minima on potential energy surfaces. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 898–907, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/carr16.html.

Nicolò Cesa-Bianchi, Claudio Gentile, Gábor Lugosi, and Gergely Neu. Boltzmann exploration done right. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6287–6296, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Petros Christodoulou. Soft actor-critic for discrete action settings. *CoRR*, abs/1910.07207, 2019. URL http://arxiv.org/abs/1910.07207.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Peter I. Frazier. A tutorial on bayesian optimization, 2018. URL https://arxiv.org/abs/1807.02811.

J.J. Thomson F.R.S. Xxiv. on the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 7(39):237–265, 1904. doi: 10.1080/14786440409463107. URL https://doi.org/10.1080/14786440409463107.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1861–1870. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/haarnoja18b.html.

Bing-Jing Hsieh, Ping-Chun Hsieh, and Xi Liu. Reinforced few-shot acquisition function learning for bayesian optimization. In *Conference on Neural Information Processing Systems*, 2021. URL https://proceedings.neurips.cc/paper/2021/file/3fab5890d8113d0b5a4178201dc842ad-Paper.pdf.

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, Dec 1998. ISSN 1573-2916. doi: 10.1023/A:1008306431147. URL https://doi.org/10.1023/A:1008306431147.

H. J. Kushner. A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise. *Journal of Basic Engineering*, 86(1):97–106, 03 1964. ISSN 0021-9223. doi: 10.1115/1.3653121. URL https://doi.org/10.1115/1.3653121.

Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1412–1421, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/D15-1166. URL https://aclanthology.org/D15-1166.

J. Močkus. On bayesian methods for seeking the extremum. In G. I. Marchuk (ed.), *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*, pp. 400–404, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg. ISBN 978-3-540-37497-8.

Samuel Müller, Noah Hollmann, Sebastian Pineda Arango, Josif Grabocka, and Frank Hutter. Transformers can do bayesian-inference by meta-learning on prior-data. In *Fifth Workshop on Meta-Learning at the Conference on Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=h9yIMMjRoje.

Sebastian W. Ober, Carl E. Rasmussen, and Mark van der Wilk. The Promises and Pitfalls of Deep Kernel Learning. *arXiv e-prints*, art. arXiv:2102.12108, February 2021.

Junhyuk Oh, Satinder Singh, and Honglak Lee. Value prediction network. 2017. URL https://arxiv.org/pdf/1707.03497.pdf.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 11 2005. ISBN 9780262256834. doi: 10.7551/mitpress/3206.001.0001. URL https://doi.org/10.7551/mitpress/3206.001.0001.

Benjamin J. Shields, Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I. Alvarado, Jacob M. Janey, Ryan P. Adams, and Abigail G. Doyle. Bayesian reaction optimization as a tool for chemical synthesis. *Nature*, 590(7844). doi: 10.1038/s41586-021-03213-y. URL https://par.nsf.gov/biblio/10231959.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf.

Richard S. Sutton. Dyna, an integrated architecture for learning, planning, and reacting. *SIGART Bull.*, 2(4):160–163, jul 1991. ISSN 0163-5719. doi: 10.1145/122344.122377. URL https://doi.org/10.1145/122344.122377.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Michael Volpp, Lukas P. Fröhlich, Kirsten Fischer, Andreas Doerr, Stefan Falkner, Frank Hutter, and Christian Daniel. Meta-learning acquisition functions for transfer learning in bayesian optimization. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=ryeYpJSKwr.

Zi Wang and Stefanie Jegelka. Max-value entropy search for efficient Bayesian optimization. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3627–3635. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/wang17e.html.

Andrew Gordon Wilson, Zhiting Hu, Ruslan Salakhutdinov, and Eric P. Xing. Deep kernel learning. *CoRR*, abs/1511.02222, 2015. URL http://arxiv.org/abs/1511.02222.

Martin Wistuba, Nicolas Schilling, and Lars Schmidt-Thieme. Scalable gaussian process-based transfer surrogates for hyperparameter optimization. *Mach. Learn.*, 107(1):43–78, jan 2018. ISSN 0885-6125. doi: 10.1007/s10994-017-5684-y. URL https://doi.org/10.1007/s10994-017-5684-y.

Yichi Zhang, Daniel W. Apley, and Wei Chen. Bayesian optimization for materials design with mixed quantitative and qualitative variables. *Scientific Reports*, 10(1):4924, Mar 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-60652-9. URL https://doi.org/10.1038/s41598-020-60652-9.