A Experiments

Here we report the results of REMBO (Wang et al., 2013) and SMAC (Hutter et al., 2011) on the synthetic data in Section 5.

A.1 Other Algorithms

For REMBO, the best parameter we try is to assume the number of underlying lower dimensions as $D/4$. The simple regret we get is 76.6 and 93.4 for $D = 50$ and $D = 100$, respectively. When we increase the number of lower dimensions, however, the performance degrades. For instance, when we use 25 dimensions for $D = 50$, the regret is 424. We address this result to imperfect optimization of acquisition function. Moreover, it is not surprising that the regret increases when we increase D. If the low dimensional embedding does not exist, we lose more information for larger D.

We also study the random-forests-based algorithm (SMAC) (Hutter et al., 2011) by using the code provided by the authors [1]. Hutter et al. (2011) demonstrate SMAC has good performance on solving combinatorial problem with 76 dimensions. However, in our synthetic data, the average regret after 1000 iterations for $D = 50$ and $D = 100$ are 130.5 and 460.1 respectively, which is generally worse than than GP-based methods. Although more iterations will result in better performance, it worths further studying to improve this method.

B Technical Proofs

In the following, we use superscript to indicate the index of the vector, matrix and tensor.

B.1 Proof of Theorem 2

Theorem 1. *Kandasamy et al. (2015) Suppose f is constructed by sampling $f^{(j)} \sim \mathcal{GP}(\mathbf{0}, \kappa^{(j)})$ for $j = 1, \ldots, M$ and then adding them. Let all kernels $\kappa^{(j)}$ satisfy certain smooth and bounded conditions Kandasamy et al. (2015). If we maximize the acquisition function $\widetilde{\varphi}_t$ to within $\tilde{O}(t^{-1/2})$ accuracy at time step t and choose $\beta_t = \tilde{O}(d \log t)$,* **Add-GP-UCB** *attains simple regret $S_T \in \tilde{O}\left(\sqrt{D\gamma_T \log T/T}\right)$ with high probability.*

Since f is projected-additive function on $\mathbf{x}$, g is additive on the projected data $\mathbf{z} = \mathbf{W}^\top \mathbf{x}$. Then we could apply Theorem 1 directly to completes the proof.

[1] http://www.cs.ubc.ca/labs/beta/Projects/SMAC/

B.2 Proof of Proposition 4

By mean value theorem, there exists $\mathbf{0} \preceq \mathbf{z}' \preceq \mathbf{z}$ such that $\tilde{f}(\mathbf{z}) = \tilde{f}(\mathbf{0}) + \nabla \tilde{f}(\mathbf{0})^\top \mathbf{z} + \frac{1}{2}\mathbf{z}^\top H(\mathbf{z}')\mathbf{z}$, where $H(\mathbf{z}') = \nabla^2 f(\mathbf{z}')$. We construct g by $g(\mathbf{z}) = \sum_{d=1}^{D} \frac{1}{D} f(\mathbf{0}) + (\nabla f(\mathbf{z}))^{(d)} \mathbf{z}^{(d)} + \frac{1}{2}(H(\mathbf{z}')^{(d,d)}(\mathbf{z}^{(i)})^2)$. Since each element in $H(\mathbf{z}')$ is bounded by Assumption 3, $|\tilde{f}(\mathbf{z}) - g(\mathbf{z})| = O(\|\mathbf{z}\|^2)$.

B.3 Proof of Proposition 4

By mean value theorem, there exists $\mathbf{0} \preceq \mathbf{z}' \preceq \mathbf{z}$ such that $\tilde{f}(\mathbf{z}) = \tilde{f}(\mathbf{0}) + \nabla \tilde{f}(\mathbf{0})^\top \mathbf{z} + \frac{1}{2}\mathbf{z}^\top H(\mathbf{0})\mathbf{z} + \frac{1}{6}T(\mathbf{z}') \times_1 \mathbf{z} \times_2 \mathbf{z} \times_3 \mathbf{z}$, where $T(\mathbf{z}')$ is the tensor of the third derivatives of $\tilde{f}(\mathbf{z}')$. Let denote the SVD of $H(\mathbf{0})$ as $H(\mathbf{0}) = \mathbf{U\Sigma U}^\top$, where $\mathbf{UU}^\top = \mathbf{I}$ and $\mathbf{\Sigma}$ is diagonal. Then

$$\begin{aligned}\tilde{f}(\mathbf{z}) &= \tilde{f}(\mathbf{0}) + \nabla \tilde{f}(\mathbf{0})^\top \mathbf{UU}^\top \mathbf{z} + \tfrac{1}{2}\mathbf{z}^\top \mathbf{U\Sigma U}^\top \mathbf{z} + \\ &\quad \tfrac{1}{6}T(\mathbf{z}') \times_1 (\mathbf{UU}^\top \mathbf{z}) \times_2 (\mathbf{UU}^\top \mathbf{z}) \times_3 (\mathbf{UU}^\top \mathbf{z}) \\ &= \tilde{f}(\mathbf{0}) + \mathbf{g}^\top \tilde{\mathbf{z}} + \frac{\tilde{\mathbf{z}}^\top \mathbf{\Sigma} \tilde{\mathbf{z}}}{2} + \frac{\tilde{T}(\mathbf{z}') \times_1 \tilde{\mathbf{z}} \times_2 \tilde{\mathbf{z}} \times_3 \tilde{\mathbf{z}}}{6},\end{aligned}$$

where $\mathbf{g} = \mathbf{U}^\top \nabla \tilde{f}(\mathbf{0})$, and $\tilde{T}(\mathbf{z}') = T(\mathbf{z}') \times_1 \mathbf{U} \times_2 \mathbf{U} \times_3 \mathbf{U}$. Then we construct h as $h(\mathbf{z}) = \sum_{d=1}^{D} \frac{1}{D}\tilde{f}(\mathbf{0}) + \mathbf{g}^{(d)}\tilde{\mathbf{z}}^{(d)} + \frac{1}{2}\mathbf{\Sigma}^{(d,d)}(\tilde{\mathbf{z}}^{(i)})^2 + \frac{1}{6}\tilde{T}(\tilde{\mathbf{z}}')^{(d,d,d)}(\tilde{\mathbf{z}}^{(i)})^3$. Since T is bounded by Assumption 3, and $\|\mathbf{U}\| = 1$, so $\tilde{T}$ is still bounded. Therefore, $|\tilde{f}(\mathbf{z}) - h(\mathbf{z})| = O(\|\mathbf{z}\|^3)$.

B.4 Proof of Corollary 8

Using the same proof of Proposition 4 by replacing $\mathbf{0}$ with $\mathbf{z}_*$ and using the decomposition $-\mathbf{QQ}^\top$ completes the proof.

B.5 Proof of Theorem 6

The proof is based on the following lemmas from Srinivas et al. (2010). Here we use $\tilde{\mu}_t(\mathbf{x})$ to denote the mean function based on the biased $\tilde{y}$, and $\mu_t(\mathbf{x})$ to denote the mean function based on y.

Lemma 2. *Srinivas et al. (2010) Set $\beta_t = \tilde{O}(d \log t)$. Then $|g(\mathbf{x}_t) - \mu_{t-1}(\mathbf{x}_t)| \leq \beta_t^{1/2}\sigma_{t-1}(\mathbf{x}_t)$ and $g(\mathbf{x}_*) \leq \mu_{t-1}(\mathbf{x}_t) + \beta_t^{1/2}\sigma_{t-1}(\mathbf{x}_t) + \frac{1}{t^2}$ with high probability.*

Lemma 3. *Set $\beta_t = \tilde{O}(d \log t)$. Then $|g(\mathbf{x}_t) - \tilde{\mu}_{t-1}(\mathbf{x}_t)| \leq \beta_t^{1/2}\sigma_{t-1}(\mathbf{x}_t) + C\epsilon$ and $g(\mathbf{x}_*) \leq \tilde{\mu}_{t-1}(\mathbf{x}_t) + \beta_t^{1/2}\sigma_{t-1}(\mathbf{x}_t) + C\epsilon + \frac{1}{t^2}$ with high probability.*

Proof. Applying Lemma 3 and $|f(\mathbf{x}) - g(\mathbf{x})| \leq \epsilon$ completes the proof. □

Lemma 4. *Set $\beta_t = \tilde{O}(d \log t)$. With high probability, the regret is bounded as follows: $r_t \leq 2\beta_t^{1/2}\sigma_{t-1}(\mathbf{x}_t) + 2C\epsilon + \frac{1}{t^2}$.*

Proof. By Lemma 3, we have $g(\mathbf{x}_*) \leq \tilde{\mu}_{t-1}(\mathbf{x}_t) + \beta_t^{1/2}\sigma_{t-1}(\mathbf{x}_t) + C\epsilon + \frac{1}{t^2}$. Therefore,

$$\begin{aligned} r_t &= g(\mathbf{x}_*) - g(\mathbf{x}_t) \\ &\leq \tilde{\mu}_{t-1}(\mathbf{x}_t) + \beta_t^{1/2}\sigma_{t-1}(\mathbf{x}_t) + C\epsilon + 1/t^2 - g(\mathbf{x}_t) \\ &\leq 2\beta_t^{1/2}\sigma_{t-1}(\mathbf{x}_t) + 2C\epsilon + 1/t^2, \end{aligned}$$

which completes the proof. □

Lemma 5. *Srinivas et al. (2010) Set $\beta_t = \tilde{O}(d\log t)$, with high probability, $\sum_{t=1}^{T} 2\beta_t^{1/2}\sigma_{t-1}(\mathbf{x}_t) \leq \sqrt{C_1 T\beta_T\gamma_T}$, where C_1 is a constant.*

Then by Lemma 4 and Lemma 5, the simple regret is bounded by

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} r_t &\leq \sqrt{\frac{C_1\beta_T\gamma_T}{T}} + 2C\epsilon + \sum_{t=1}^{T}\frac{1}{t^2} \\ &= \tilde{O}(\sqrt{\frac{d\gamma_T}{T}} + \epsilon) \end{aligned}$$

B.6 Proof of Corollary 7

Let $\mathbf{u}_* = \operatorname{argmax}_{\mathbf{x}} f(\mathbf{x})$ and $\mathbf{v}_* = \operatorname{argmax}_{\mathbf{x}} g(\mathbf{x})$. Since $|f(\mathbf{x}) - g(\mathbf{x})| \leq \epsilon$, we $f(\mathbf{u}_*) - g(\mathbf{v}_*) \leq f(\mathbf{u}_*) - g(\mathbf{u}_*) \leq \epsilon$. Combining with Theorem 6, the simple regret on f is bounded by

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} f(\mathbf{u}_*) - f(\mathbf{x}_t) &\leq \frac{1}{T}\sum_{t=1}^{T} g(\mathbf{v}_*) + C\epsilon - g(\mathbf{x}_t) + C\epsilon \\ &= 2C\epsilon + \frac{1}{T}\sum_{t=1}^{T} g(\mathbf{v}_*) - g(\mathbf{x}_t) \\ &= \tilde{O}(\sqrt{\frac{d\gamma_T}{T}} + \epsilon), \end{aligned}$$

with high probability.

References

Hutter, F., Hoos, H. H., and Leyton-Brown, K. (2011). Sequential model-based optimization for general algorithm configuration. In *Proceedings of the 5th International Conference on Learning and Intelligent Optimization*.

Kandasamy, K., Schneider, J., and Póczos, B. (2015). High Dimensional Bayesian Optimisation and Bandits via Additive Models. In *International Conference on Machine Learning*.

Srinivas, N., Krause, A., Kakade, S., and Seeger, M. (2010). Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. In *International Conference on Machine Learning*.

Wang, Z., Zoghi, M., Hutter, F., Matheson, D., and de Freitas, N. (2013). Bayesian Optimization in High Dimensions via Random Embeddings. In *International Joint Conference on Artificial Intelligence*.