Distributionally Robust Bayesian Quadrature Optimization: Supplementary

APPENDIX A

In this appendix, we provide a detailed proof of Theorem 2 in the main text about a sublinear bound on the Bayesian regret of the DRBQO algorithm. For simplicity, we focus on the case where the decision space $\mathcal{X}$ and the distributional uncertainty set $\mathcal{P}_{n,\rho}$ are finite. The results can be extended to infinite sets using the discretization trick as in Srinivas et al. [2010].

Notations and conventions. Unless explicitly specified otherwise, we denote a conditional distribution $P(.|x) \in \mathcal{P}_{n,\rho}, \forall x \in \mathcal{X}$ by P, i.e., $P \in \mathcal{P}_{n,\rho} \times \mathcal{X}$. Recall the definition of the quadrature functional in the main text as

$$g(f, x, P) = \int P(w|x) f(x, w) dw,$$

for any $x \in \mathcal{X}$ and $P \in \mathcal{P}_{n,\rho} \times \mathcal{X}$. Let $x^* \in \arg\max_{x \in \mathcal{X}} \min_{P \in \mathcal{P}_{n,\rho}} \mathbb{E}_{P(w)}[f(x, w)]$, and $P^*(.|x) = \arg\min_{P \in \mathcal{P}_{n,\rho}} \mathbb{E}_{P(w)}[f(x, w)], \forall x \in \mathcal{X}$. Since f is a stochastic process (a GP in our case), x^* and P^* are also random variables. The DRBQO algorithm π^{DRBQO} maps at a time step t the history $H_t = (x_1, w_1, P_1, ..., x_{t-1}, w_{t-1}, P_{t-1})$ to a new decision $(x_t, w_t) \in \mathcal{X} \times S_n$ and conditional distribution $P_t \in \mathcal{P}_{n,\rho} \times \mathcal{X}$ as presented in line 2-3 of Algorithm 1 in the main text. The practical implementation of Algorithm 1 samples (x_t, P_t) as follows: $x_t \in \arg\max_{x \in \mathcal{X}} \min_{P \in \mathcal{P}_{n,\rho}} \mathbb{E}_{P(w)}[\tilde{f}_t(x, w)]$, and $P_t(.|x) = \arg\min_{P \in \mathcal{P}_{n,\rho}} \mathbb{E}_{P(w)}[\tilde{f}_t(x, w)], \forall x \in \mathcal{X}$ where $\tilde{f}_t$ is a function sample of f at time t from its posterior GP.

Lemma 1. *For any sequence of deterministic functions $\{U_t : \mathcal{X} \times \mathcal{P}_{n,\rho} \times \mathcal{X} \to \mathbb{R} | t \in \mathbb{N}\}$,*

$$\begin{aligned}
&BayesRegret(T, \pi^{DRBQO}) \\
&= \mathbb{E} \sum_{t=1}^{T} \left[U_t(x_t, P_t) - g(f, x_t, P_t)\right] \\
&+ \mathbb{E} \sum_{t=1}^{T} \left[g(f, x^*, P^*) - U_t(x^*, P^*)\right],
\end{aligned}$$

for all $T \in \mathbb{N}$.

Proof. Given H_t, π^{DRBQO} samples (x_t, P_t) according to the probability they are optimal, i.e., $(x_t, P_t) \sim Pr(x^*, P^*|H_t)$. Thus, conditioned on H_t, (x^*, P^*) and (x_t, P_t) are identically distributed. As a result, given a deterministic function U_t, we have $\mathbb{E}[U_t(x^*, P^*)] = \mathbb{E}[U_t(x_t, P_t)]$. Therefore,

$$\begin{aligned}
&\mathbb{E}\left[g(f, x^*, P^*) - g(f, x_t, P_t)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[g(f, x^*, P^*) - g(f, x_t, P_t)\right] | H_t\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[U_t(x_t, P_t) - g(x_t, P_t)\right] | H_t\right] \\
&+ \mathbb{E}\left[\mathbb{E}\left[g(f, x^*, P^*) - U_t(x^*, P^*)\right] | H_t\right] \\
&= \mathbb{E}\left[U_t(x_t, P_t) - g(f, x_t, P_t)\right] \\
&+ \mathbb{E}\left[g(f, x^*, P^*) - U_t(x^*, P^*)\right].
\end{aligned}$$

□

Lemma 2. *Let $X \sim \mathcal{N}(\mu, \sigma^2)$.*

1. *For all $\beta \geq 0$, we have*

$$Pr\{|X - \mu| > \beta^{1/2}\sigma\} \leq e^{-\beta/2}.$$

2. *If $\mu \leq 0$, then*

$$\mathbb{E}[\max\{X, 0\}] = \frac{\sigma}{\sqrt{2\pi}} e^{\frac{-\mu^2}{2\sigma^2}}.$$

3. *For all $a \leq b$, we have*

$$\mathbb{E}[X | a < X < b] = \mu - \sigma^2 \frac{p(a) - p(b)}{\phi(a) - \phi(b)},$$

where $p(x)$ and $\phi(x)$ denote the density function and cumulative distribution function of X, respectively.

Proof. The results are simple properties of normal distributions. □

Lemma 3. *Given $H_t, \forall t \in \mathbb{N}$, let $\sigma_t^2(x, w) := C_t(x, w; x, w)$ be the variance of $f(x, w)$. Then, for all P, all x and for $w^* = \arg\max_{w \in S_n} \sigma_t^2(x, w)$, we have*

$$\sigma_t^2(x, P) = Var[g(f, x, P)|H_t] \leq \sigma_t^2(x, w^*).$$

Proof. It follows from a simple property of posterior covariance that

$$\sigma_t^2(x,P) = \sum_{w,w'} P(w|x)P(w'|x)C_t(x,w;x,w')$$
$$\le \sum_{w,w'} P(w|x)P(w'|x)C_t(x,w;x,w)$$
$$\le \sum_{w,w'} P(w|x)P_t(w'|x)\sigma_{t-1}^2(x,w^*)$$
$$= \sigma_t^2(x,w^*).$$

□

Lemma 4. *If $U_t(x,P) = \mu_{t-1}(x,P) + \sqrt{\beta_t}\sigma_{t-1}(x,P)$ where*

$$\mu_{t-1}(x,P) := \int P(w|x)\mu_{t-1}(x,w)dw,$$
$$\sigma_{t-1}^2(x,P) :=$$
$$\int\int C_{t-1}(x,w;x,w')P(w|x)P(w'|x)dwdw',$$

and $\beta_t = 2\log\frac{(t^2+1)|\mathcal{X}||\mathcal{P}_{n,\rho}|}{\sqrt{2\pi}}$, then

$$\mathbb{E}\sum_{t=1}^T [g(f,x^*,P^*) - U_t(x^*,P^*)] \le 1,$$

for all $T \in \mathbb{N}$.

Proof. The trick is to concentrate on the non-negative terms of the expectation. These non-negative terms can be bounded due to the specific choice of upper confidence bound U_t.

Note that for any deterministic conditional distribution $P \in \mathcal{P}_{n,\rho} \times \mathcal{X}$, we have $g(f,x,P) \sim \mathcal{N}(\mu_{t-1}(x,P), \sigma_{t-1}^2(x,P))$, i.e., $g(f,x,P) - U_t(x,P) \sim \mathcal{N}(-\sqrt{\beta_t}\sigma_{t-1}(x,P), \sigma_{t-1}^2(x,P))$. It thus follows from Lemma 2.2 that:

$$\mathbb{E}[\max\{g(f,x,P) - U_t(x,P), 0\}|H_t]$$
$$= \frac{\sigma_{t-1}(x,P)}{\sqrt{2\pi}}\exp(\frac{-\beta_t}{2})$$
$$= \frac{\sigma_{t-1}(x,P)}{(t^2+1)|\mathcal{X}||\mathcal{P}_{n,\rho}|} \le \frac{1}{(t^2+1)|\mathcal{X}||\mathcal{P}_{n,\rho}|}.$$

The final inequality above follows from Lemma 3 and from the assumption that $\sigma_0(x,w) \le 1, \forall x, w$, i.e.,

$$\sigma_{t-1}(x,P) \le \sigma_{t-1}(x,w^*) \le \sigma_0(x,w^*) \le 1,$$

where $w^* = \arg\max_w C_{t-1}(x,w;x,w)$.

Therefore, we have

$$\mathbb{E}\sum_{t=1}^T [g(f,x^*,P^*) - U_t(x^*,P^*)]$$
$$\le \mathbb{E}\sum_{t=1}^T \mathbb{E}[\max\{g(f,x^*,P^*) - U_t(x^*,P^*), 0\}|H_t]$$
$$\le \mathbb{E}\sum_{t=1}^T \sum_{x\in\mathcal{X}} \sum_{P\in\mathcal{P}_{n,\rho}} \mathbb{E}[\max\{g(f,x,P) - U_t(x,P), 0\}]$$
$$\le \sum_{t=1}^\infty \sum_{x\in\mathcal{X}} \sum_{P\in\mathcal{P}_{n,\rho}} \frac{1}{(t^2+1)|\mathcal{X}||\mathcal{P}_{n,\rho}|} = 1.$$

□

Lemma 5. *Given the definition of the maximum information gain γ_T as in Srinivas et al. [2010], we have*

$$\mathbb{E}\sum_{t=1}^T [U_t(x_t,P_t) - g(f,x_t,P_t)]$$
$$\le \frac{(\sqrt{\beta_T}+B)\sqrt{2\pi}}{|\mathcal{X}||\mathcal{P}_{n,\rho}|} + 2\gamma_T\sqrt{(1+2\rho)n}(1+\sigma^{-2})^{-1}$$
$$+ 2\sqrt{T\gamma_T(1+\sigma^{-2})^{-1}\log\frac{(1+T^2)|\mathcal{X}||\mathcal{P}_{n,\rho}|}{\sqrt{2\pi}}},$$

for all $T \in \mathbb{N}$.

Proof. Now we bound the first term

$$L := \mathbb{E}\sum_{t=1}^T [U_t(x_t,P_t) - g(f,x_t,P_t)]$$
$$= \mathbb{E}\sum_{t=1}^T \mathbb{E}[J(x_t,H_t)|x_t,H_t],$$

where

$$J(x_t,H_t) = \mathbb{E}[U_t(x_t,P_t) - g(f,x_t,P_t)|H_t,x_t].$$

While the second term of the Bayesian regret of DR-BQO can be bounded as in Lemma 4 by adopting the techniques from Russo and Roy [2014], bounding L in DRBQO is non-trivial. This is because $P_t(.|x)$ is also a random process on the simplex given H_t. Thus, $g(f,x_t,P_t)|H_t$ does not follow a GP as in the standard Quadrature formulae. In addition, we do not have a closed form of $\mathbb{E}[g(f,x_t,P_t)|H_t]$. We overcome this difficulty by decomposing J into several terms that can be bounded more easily and leveraging the mild assumptions of f in the problem setup.

Given (H_t,x_t), we are interested in bounding $J(x_t,H_t)$. The main idea for bounding this term is that we decompose the range $\mathbb{R}$ of the random variable $f(x_t,w), \forall w$

into three disjoint sets:

$$A_t(w) = \left\{ f(x_t, w) \middle| |f(x_t, w) - \mu_{t-1}(x_t, w)| \leq \sqrt{\beta_t}\sigma_{t-1}(x_t, w) \right\},$$

$$B_t(w) = \left\{ f(x_t, w) \middle| \mu_{t-1}(x_t, w) - f(x, w) > \sqrt{\beta_t}\sigma_{t-1}(x_t, w) \right\},$$

$$C_t(w) = \left\{ f(x_t, w) \middle| \mu_{t-1}(x_t, w) - f(x, w) < -\sqrt{\beta_t}\sigma_{t-1}(x_t, w) \right\},$$

for all $w \in \Omega$. Note that $A_t(w) \cup B_t(w) \cup C_t(w) = \mathbb{R}, \forall w$. We also denote $\bar{A}_t(w) = \mathbb{R} \backslash A_t(w) = B_t(w) \cup C_t(w), \forall w$.

Since f is bounded on A_t, there exists P_t^* such that

$$P_t^*(.|x) = \arg\max_{P \in \mathcal{P}_{n,\rho}} \{U_t(x, P) - g(f, x, P) | f \in A_t\},$$

for all $x \in \mathcal{X}$.

Using the equation above, we decompose $J(x_t, H_t)$ as

$$\begin{aligned}
&J(x_t, H_t) = \mathbb{E}[U_t(x_t, P_t) - g(f, x_t, P_t)|H_t, x_t] \\
&= \mathbb{E}_{f \in A_t}[U_t(x_t, P_t) - g(f, x_t, P_t)|H_t, x_t] \\
&+ \mathbb{E}_{f \in \bar{A}_t}[U_t(x_t, P_t) - g(f, x_t, P_t)|H_t, x_t] \\
&\leq \mathbb{E}_{f \in A_t}[U_t(x_t, P_t^*) - g(f, x_t, P_t^*)|H_t, x_t] \\
&+ \mathbb{E}_{f \in \bar{A}_t}[U_t(x_t, P_t) - g(f, x_t, P_t)|H_t, x_t] \\
&= \mathbb{E}[U_t(x_t, P_t^*) - g(f, x_t, P_t^*)|H_t, x_t] \\
&+ \mathbb{E}_{f \in \bar{A}_t}[U_t(x_t, P_t) - g(f, x_t, P_t)|H_t, x_t] \\
&- \mathbb{E}_{f \in \bar{A}_t}[U_t(x_t, P_t^*) - g(f, x_t, P_t^*)|H_t, x_t] \\
&= J_1 + J_2 + J_3,
\end{aligned}$$

where

$$\begin{aligned}
J_1 &= \mathbb{E}[U_t(x_t, P_t^*) - g(f, x_t, P_t^*)|H_t, x_t], \\
J_2 &= \mathbb{E}_{f \in \bar{A}_t}[U_t(x_t, P_t) - g(f, x_t, P_t)|H_t, x_t], \\
J_3 &= \mathbb{E}_{f \in \bar{A}_t}[g(f, x_t, P_t^*) - U_t(x_t, P_t^*)|H_t, x_t].
\end{aligned}$$

It follows from Lemma 3 and from the selection of w_t for the highest posterior variance in the DRBQO algorithm (Algorithm 1 in the main text) that for all P, we have

$$\begin{aligned}
&\sigma_{t-1}^2(x_t, P) = \sum_{w, w'} P(w|x) P(w'|x) C_{t-1}(x_t, w; x_t, w') \\
&\leq \sigma_{t-1}^2(x_t, w_t).
\end{aligned}$$

Note that given (H_t, x_t), w_t is deterministic.

For J_1, we have

$$\begin{aligned}
J_1 &= \mathbb{E}[U_t(x_t, P_t^*) - g(f, x_t, P_t^*)|H_t, x_t] \\
&= U_t(x_t, P_t^*) - \mathbb{E}[g(f, x_t, P_t^*)|H_t, x_t] \\
&= U_t(x_t, P_t^*) - \mu_{t-1}(x_t, P_t^*) \\
&= \sqrt{\beta_t}\sigma_{t-1}(x_t, P_t^*) \\
&\leq \sqrt{\beta_t}\sigma_{t-1}(x_t, w_t).
\end{aligned}$$

For J_2, we have

$$\begin{aligned}
J_2 &= \mathbb{E}_{f \in \bar{A}_t}[U_t(x_t, P_t) - g(f, x_t, P_t)|H_t, x_t] \\
&= \mathbb{E}_{f \in \bar{A}_t}[\sqrt{\beta_t}\sigma_{t-1}(x_t, P_t)|H_t, x_t] \\
&+ \mathbb{E}_{f \in B_t}[\sum_w P_t(w)(\mu_{t-1}(x_t, w) - f(x_t, w)|H_t, x_t] \\
&+ \mathbb{E}_{f \in C_t}[\sum_w P_t(w)(\mu_{t-1}(x_t, w) - f(x_t, w)|H_t, x_t] \\
&\leq \mathbb{E}_{f \in \bar{A}_t}[\sqrt{\beta_t}\sigma_{t-1}(x_t, w_t)] \\
&+ \mathbb{E}_{f \in B_t}[\sum_w P_t(w)(\mu_{t-1}(x_t, w) - f(x_t, w)|H_t, x_t] \\
&\leq \sqrt{\beta_t}\sigma_{t-1}(x_t, w_t) e^{-\beta_t/2} \\
&+ \mathbb{E}_{f \in B_t}[\sum_w P_t(w)(\mu_{t-1}(x_t, w) - f(x_t, w)|H_t, x_t] \\
&\leq \sqrt{\beta_t}\sigma_{t-1}(x_t, w_t) e^{-\beta_t/2} \\
&+ \mathbb{E}_{f \in B_t} \sqrt{\sum_w (\mu_{t-1}(x_t, w) - f(x_t, w))^2 \sum_w P_t^2(w)} \\
&\leq \sqrt{\beta_t}\sigma_{t-1}(x_t, w_t) e^{-\beta_t/2} \\
&+ \mathbb{E}_{f \in B_t} \sqrt{\sum_w (\mu_{t-1}(x_t, w) - f(x_t, w))^2 \frac{1+2\rho}{n}} \\
&\leq \sqrt{\beta_t}\sigma_{t-1}(x_t, w_t) e^{-\beta_t/2} \\
&+ \sqrt{\frac{1+2\rho}{n}} \mathbb{E}_{f \in B_t} \sum_w (\mu_{t-1}(x_t, w) - f(x_t, w)) \\
&= \sqrt{\beta_t}\sigma_{t-1}(x_t, w_t) e^{-\beta_t/2} \\
&+ \sqrt{\frac{1+2\rho}{n}} \sum_w (\mu_{t-1}(x_t, w) - \mathbb{E}_{f \in B_t}[f(x_t, w)]) \\
&= \sqrt{\beta_t}\sigma_{t-1}(x_t, w_t) e^{-\beta_t/2} \\
&+ \sqrt{\frac{1+2\rho}{n}} \sum_w \sigma_{t-1}^2(a_t, w) \kappa(x_t, w) \\
&\leq \sqrt{\beta_t}\sigma_{t-1}(x_t, w_t) e^{-\beta_t/2} + \sqrt{\frac{1+2\rho}{n}} \sum_w \sigma_{t-1}^2(a_t, w) \\
&\leq \sqrt{\beta_t}\sigma_{t-1}(x_t, w_t) e^{-\beta_t/2} + \sqrt{n(1+2\rho)} \sigma_{t-1}^2(a_t, w_t),
\end{aligned}$$

where

$$\kappa(x_t, w) := \frac{p(\mu_{t_1}(x_t, w) - \sqrt{\beta_t}\sigma_{t-1}(x_t, w))}{\phi(\mu_{t_1}(x_t, w) - \sqrt{\beta_t}\sigma_{t-1}(x_t, w))} \leq 1,$$

and $p(.)$ and $\phi(.)$ denote the density function and the cumulative distribution function of the Gaussian distribution $\mathcal{N}(\mu_{t-1}(x_t, w), \sigma^2_{t-1}(x_t, w)), \forall w$. Here, the third inequality follows from the Cauchy-Schwartz inequality; the fourth inequality follows from the bound of the χ^2 ball on the distributions in it; the fifth inequality follows from that fact that $\mu_{t-1}(x, w) - f(x, w) \geq \sqrt{\beta_t}\sigma_{t-1}(x, w) \geq 0$; and the final equation follows from Lemma 2.3.

For J_3, we have

$$\begin{aligned}
J_3 &= \mathbb{E}_{f \in \bar{A}_t}[g(f, x_t, P_t^*) - U_t(x_t, P_t^*)|H_t, x_t] \\
&= \mathbb{E}_{f \in \bar{A}_t}[g(f, x_t, P_t^*)|H_t, x_t] + \mathbb{E}_{f \in \bar{A}_t}[-U_t(x_t, P_t^*)] \\
&= \mathbb{E}_{f \in \bar{A}_t}[-U_t(x_t, P_t^*)] \\
&= \mathbb{E}_{f \in \bar{A}_t}[-\mu_{t-1}(x_t, P_t^*) - \sqrt{\beta_t}\sigma_{t-1}(x_t, P_t^*)] \\
&\leq \mathbb{E}_{f \in \bar{A}_t}[-\mu_{t-1}(x_t, P_t^*)] \\
&\leq \mathbb{E}_{f \in \bar{A}_t}[B] \\
&\leq Be^{-\beta_t/2}.
\end{aligned}$$

Here, the second equation follows from the property that $\mathbb{E}_{f \in \bar{A}_t}[f(x_t, w)] = 0$ since $f(x_t, w) \sim \mathcal{N}(\mu_{t-1}(x_t, w), \sigma^2_{t-1}(x_t, w)), \forall w$, and $\bar{A}_t(w)$ is a symmetric region in $\mathbb{R}$ with respect to (but not including) the line $x = \mu_{t-1}(x_t, w), \forall w$; the first inequality follows the non-negativity of the posterior variance $\sigma_{t-1}(x_t, P_t^*)$; the second inequality follows from that the posterior mean $\mu_{t-1}(x, w)$ of a GP is in the RKHS associated with kernel k of the GP, thus is bounded above by B by the mild assumption in the problem setup; and the final inequality follows from Lemma 2.1.

Combining these results, we can finally bound the first term of the Bayesian regret of DRBQO,

$$\begin{aligned}
L &= \mathbb{E}\sum_{t=1}^{T}\mathbb{E}[J(x_t, H_t)|x_t, H_t] \\
&\leq \mathbb{E}\sum_{t=1}^{T}\sqrt{\beta_t}\sigma_{t-1}(x_t, w_t) + \mathbb{E}\sum_{t=1}^{T} Be^{-\beta_t/2} \\
&+ \mathbb{E}\sum_{t=1}^{T}\sqrt{\beta_t}\sigma_{t-1}(x_t, w_t)e^{-\beta_t/2} \\
&+ \mathbb{E}\sum_{t=1}^{T}\sqrt{n(1+2\rho)}\sigma^2_{t-1}(a_t, w_t) \\
&\leq \mathbb{E}\sqrt{T\beta_T}\sqrt{\sum_{t=1}^{T}\sigma^2_{t-1}(x_t, w_t)} \\
&+ (B + \sqrt{\beta_T})\sum_{t=1}^{\infty}\frac{\sqrt{2\pi}}{(1+t^2)|\mathcal{X}||\mathcal{P}_{n,\rho}|} \\
&+ \sqrt{n(1+2\rho)}\mathbb{E}\sum_{t=1}^{T}\sigma^2_{t-1}(a_t, w_t) \\
&\leq \sqrt{T\beta_T}\sqrt{2(1+\sigma^{-2})^{-1}\gamma_T} + \frac{(\sqrt{\beta_T} + B)\sqrt{2\pi}}{|\mathcal{X}||\mathcal{P}_{n,\rho}|} \\
&+ \sqrt{n(1+2\rho)}2(1+\sigma^{-2})^{-1}\gamma_T,
\end{aligned}$$

where γ_T is the maximum information gain defined in Srinivas et al. [2010], and we also use the following inequality of the maximum information gain

$$\sum_{t=1}^{T}\sigma^2_{t-1}(x_t, w_t) \leq 2(1+\sigma^{-2})^{-1}\gamma_T.$$

□

Theorem 2 is a direct consequence of Lemma 1, Lemma 4 and Lemma 5.

Upper bounds on the information gain. For completeness, we include here the upper bounds for the information gains γ_T which are derived from Srinivas et al. [2010]:

Kernel type	**Information gain γ_T**
Linear	$\mathcal{O}(d\log T)$
Squared exponential	$\mathcal{O}((\log T)^{d+1})$
Matérn with $\nu > 1$	$\mathcal{O}(T^{d(d+1)/(2\nu+d(d+1))}\log T)$

where $d \in \mathbb{N}$ is the dimension of the search domain.

APPENDIX B.1

In this appendix, we provide derivation details of Proposition 1 in the main text.

Consider the constrained optimization problem

$$\min_{p \in \mathcal{P}_{n,\rho}}\sum_{i=1}^{n} p_i l_i. \tag{1}$$

This is a convex optimization problem which forms the Lagrangian:

$$\begin{aligned}
L(p, \lambda, \eta, \zeta) = p^T l - \lambda\left(\rho - \frac{1}{2n}\sum_{i=0}^{n}(np_i - 1)^2\right) \\
- \eta(1 - \sum_{i=1}^{n} p_i) - \sum_{i=1}^{n}\zeta_i p_i, \quad (2)
\end{aligned}$$

where $p \in \mathbb{R}^n, \lambda \geq 0$, $\eta \in \mathbb{R}$, and $\zeta \in \mathbb{R}^n_+$. The KKT conditions for the primal problem (1) are:

$$\begin{cases} l_i + \lambda(np_i - 1) + \eta - \zeta_i = 0, \forall 1 \leq i \leq n & (3) \\ \lambda\left(2\rho + 1 - n\|p\|_2^2\right) = 0 & (4) \\ n\|p\|_2^2 \leq 2\rho + 1 & (5) \\ \lambda \geq 0 & (6) \\ \sum_{i=1}^{n} p_i = 1 & (7) \\ \zeta_i p_i = 0, \forall 1 \leq i \leq n & (8) \\ \zeta_i \geq 0, \forall 1 \leq i \leq n. & (9) \end{cases}$$

We can see that the strong duality holds because the primal problem in (1) satisfies the Slater's condition; therefore the KKT conditions are the necessary and sufficient conditions for the primal optimal solution. It follows from Equations (3), (8), and (9) that:

$$\lambda n p_i = (-l_i - \eta)_+ := \max\{-l_i - \eta, 0\}, \tag{10}$$

which, combined with Equation (7), implies that:

$$n\lambda = \sum_{i=1}^{n} (-l_i - \eta)_+ \tag{11}$$

From Equation (11), we have:

$$\eta = \frac{-\sum_{i \in A} l_i - n\lambda}{|A|}, \tag{12}$$

where $A = \{i : l_i + \eta \leq 0\}$. Note that $|A| \geq 1$ because otherwise $p_i = 0, \forall 1 \leq i \leq n$ which contradicts Equation (7). We then plug Equation (12) and (10) into Equation (5) to solve for λ. Note that $\|p\|_2^2$ is decreasing in λ, thus we can bisect to find the optimal λ within its bound. We can easily obtain a bound on λ from Equation (5):

$$0 \leq \lambda \leq \max\left\{\frac{-l_{min} + \sum_{i=1}^{n} l_i}{\sqrt{1+2\rho} - 1}, \frac{-l_{min} + l_{max}}{\sqrt{1+2\rho}}\right\},$$

where $l_{min} = \min_{1 \leq i \leq n} l_i$, and $l_{max} = \max_{1 \leq i \leq n} l_i$.

The optimal distribution $\arg\min_{p \in \mathcal{P}_{n,\rho}} \sum_{i=1}^{n} p_i l_i$ is not constant, but rather a function of l. Thus, its gradients with respect to some parameter ψ must be computed from those of l. This becomes straightforward when we have solved (p_i, λ, η) in terms of l as in the results above:

$$\begin{cases} \frac{\partial p_i}{\partial \psi} = \frac{-1}{n\lambda^2}(-l_i - \eta)\frac{\partial \lambda}{\partial \psi} + \frac{1}{n\lambda}\left(-\frac{\partial l_i}{\partial \psi} - \frac{\partial \eta}{\partial \psi}\right) \\ |A|\frac{\partial \eta}{\partial \psi} = -\sum_{i \in S} \frac{\partial l_i}{\partial \psi} - n\frac{\partial \lambda}{\partial \psi} \\ \sum_{i \in S} p_i \frac{\partial p_i}{\partial \psi} = 0. \end{cases}$$

APPENDIX B.2

In this appendix, we present the details of the bisection search in Algorithm 2 for computing the ρ-robust distributions for line 3 of Algorithm 1 in the main text.

Algorithm 2: Bisection search

Input: $p(\lambda)$ computed in Proposition 1 in the main text, $\epsilon \geq 0$

$\lambda_{min} = 0$
$\lambda_{max} = \max\left\{\frac{-l_{min}+\sum_{i=1}^{n} l_i}{\sqrt{1+2\rho}-1}, \frac{-l_{min}+l_{max}}{\sqrt{1+2\rho}}\right\}$
$\lambda = \lambda_{min}$
while $\lambda_{max} - \lambda_{min} > \epsilon$ **do**
    $\lambda = \frac{1}{2}(\lambda_{max} + \lambda_{min})$;
    if $n\|p(\lambda)\|_2^2 > 2\rho + 1$ **then**
        $\lambda_{min} = \lambda$
    end
    else
        $\lambda_{max} = \lambda$
    end
end

Output: $\lambda, p(\lambda)$

APPENDIX B.3

For the details of derivation for posterior sampling (a.k.a Thompson sampling), see Appendix A of Hernández-Lobato et al. [2014].

APPENDIX C

In this appendix, we provide some more experimental results of DRBQO on synthetic and real-world problems.

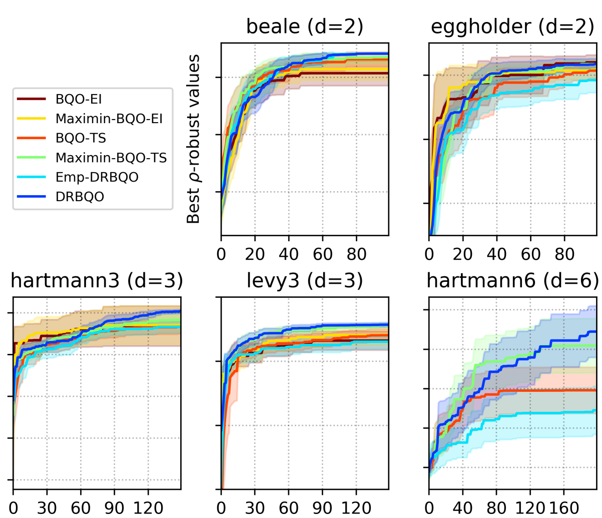

Figure 6: The performance of DRBQO and the baselines on the expected reformulation of various synthetic functions. Here we use $n = 10$ and the best ρ values are calculated with $\rho = 1.0$. DRBQO achieves higher ρ-robust values than the BQO baselines in almost all the tested functions.

Synthetic functions. The task in this experiment is to maximize $\mathbb{E}_{w \in \mathcal{N}(0,1)}[f(x,w)]$ where f is a standard synthetic function such as Beale, Eggholder, Hartmann and Levy, x is normalized to the unit cube and $f(x,w) := f(x+w)$. The performance metric used in this experiment is the ρ-robust values $\min_{P \in \mathcal{P}_{n,\rho}} \mathbb{E}_{P(w)}[f(x,w)]$. Here we use $n = 10$ and $\rho = 1.0$. We repeat the experiment 30 times and report the average mean and the 96% confidence interval for each evaluation metric. The result is presented in Figure 6. The result shows that DRBQO achieves higher ρ-robust values than the baseline methods in all these functions except that in EggHolder function, DRBQO is compatable with BQO-EI but outperforms the other algorithms.

Cross-validation hyperparameter tuning for SVM. We use glass and connectionist bench classification datasets from UCI machine learning repository. [1] The glass dataset contains 214 samples describing glass properties in 10 features. The task associated with the glass dataset is to classify an example into one of 7 classes. The connectionist bench dataset contains 208 samples each of which has 60 attributes. The task in the connectionist bench dataset is to classify whether sonar signals bounced off a metal cylinder or a roughly cylindrical rock. Each of the datasets is split into the training and test sets with the ratio of 80 : 20. The training set is further split into $n = 5$ folds for this experiment.

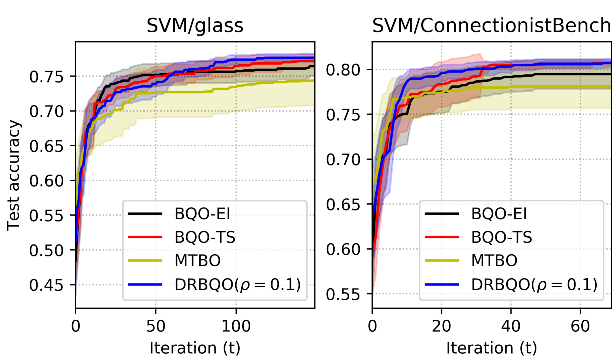

Figure 7: The test classification accuracy of SVM on glass and connectionist bench dataset tuned by DRBQO and the BQO baselines. In this example, we use $n = 5$.

Support vector machine (SVM) is a simple machine learning algorithm for classification problems. SVMs with RBF kernels have two hyperparameters: the misclassification trade-off C and the RBF hyperparameter γ. We tuned these two hyperparameters in this example.

[1] http://archive.ics.uci.edu/ml

The performance metric for this experiment is the classification accuracy of SVM in the test set. We repeat the experiment 30 times and report the average mean and the 96% confidence interval for each evaluation metric. The result is presented in Figure 7. In this example, DRBQO outperforms the baselines.

Appendix D: Further discussion

In this appendix, we provide further discussion of our proposed framework suggested by our anonymous reviewers during the review phase.

Selection of ρ in DRBQO. ρ is a problem-dependent hyperparameter and depends on the variance of $f(x,w)$ along w. Intuitively the higher the variance, more conservative we would like to be by setting the larger ρ value in the range of $[0, (n-1)/2]$. If there is no prior knowledge of the variance, we can heuristically perform grid search for ρ in $[0, (n-1)/2]$.

Other potential applications of our proposed framework. Our work considers the problem of learning to optimize under uncertain contexts where the context distribution itself is misspecified due to limited data. In particular, our proposed method can be applied to any problem where apart from the main inputs (i.e., optimization variables x), an additional factor (i.e., w) also affects the function output but this additional factor is not under our control. We list two specific applications here:

- As the first application, we are considering our method for alloy design where alloying element powders are slightly impure (in practice it is hard to find 100% pure element powder). The intended mixture composition of various elements is x while the impurities are w in our formulation. It is only possible to obtain a limited set of samples of the impurity values through measuring the exact alloy composition via X-ray or optical spectroscopy as the impurity measuring process is expensive. In this scenario, our proposed method aims to find an alloy composition that has the highest strength while being distributionally robust to impurities.

- Another example is robust control in reinforcement learning where our goal is to learn a policy (i.e., x) that is both optimal and robust to the unknown environment variables (i.e., w) e.g. some unobserved state features determined randomly by the environment (e.g., [Paul et al., 2018]). Here $f(x,w)$ is the goodness of policy x for an environment variable w. Measuring $f(x,w)$ for a given policy x is computationally expensive. While current works assume knowledge of the true environment distribution $P_0(w)$, it is typically unknown except a few of its samples that are obtained via previous catastrophic experiences with the environment.

Extensions to other divergence measures beyond χ^2 divergence. We have focused on χ^2 divergence mainly for simplicity. Our algorithmic and theoretical results can be potentially extended to f-divergence (including χ^2, KL and Hellinger) that requires the involved distribution to have the same support as the nominal distribution $\hat{P}$. Regarding the algorithmic extension for f-divergence, since f in f-divergence is convex, the surrogate DRO still reduces to convenient KKT conditions (as the strong duality still holds). Regarding the theoretical extension for f-divergence, the sublinear convergence rate in Theorem 2 remains valid because in our analysis the distribution-dependent term $\sum_i p_i^2$ is always bounded above by 1 (though in the case of χ^2 divergence, this bound can be tighter as shown in our proof of Theorem 2). The current form of our framework cannot however be extended to divergences that are defined for distributions of continuous support such as Wasserstein because our analysis relies on the assumption of finite support for the distributional uncertainty set. This assumption is however very mild in practice because if one of the involved distributions is not discrete, computing the Wasserstein distance becomes intractable even with the simplest scenario where one distribution is uniform while the other is discrete with two atoms. In practice, we can usually avoid this intractability by discretizing the support via discrete distributions for the distributional uncertainty set, and thus can leverage our analytical insights.

Discussion of the derived Bayesian regret bound of DRBQO. The Bayesian bound of DRBQO in Theorem 2 of our main paper is of order $\sqrt{T\gamma_T \log((1+T^2)|\mathcal{X}||\mathcal{P}_{n,\rho}|)}$ which matches the standard upper bounds (up to an extra log constant $\log|\mathcal{P}_{n,\rho}|$) established in Russo and Roy [2014] and Srinivas et al. [2010]. The extra log constant in our bound accounts for an additional decision space $\mathcal{P}_{n,\rho}$ for the context distribution in our problem. To our knowledge, the standard bound above is one of the best known upper bounds for GP optimization (e.g., Scarlett et al. [2017] establishes a lower bound for GP optimization suggesting that the standard bound above is near-optimal (w.r.t. the established lower bound) for the square exponential kernel).

References

José Miguel Hernández-Lobato, Matthew W. Hoffman, and Zoubin Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. In *NIPS*, pages 918–926, 2014.

Supratik Paul, Konstantinos I. Chatzilygeroudis, Kamil Ciosek, Jean-Baptiste Mouret, Michael A. Osborne, and Shimon Whiteson. Alternating optimisation and quadrature for robust control. In *AAAI*, pages 3925–3933. AAAI Press, 2018.

Daniel Russo and Benjamin Van Roy. Learning to optimize via posterior sampling. *Math. Oper. Res.*, 39(4):1221–1243, 2014.

Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy gaussian process bandit optimization. In *COLT*, volume 65 of *Proceedings of Machine Learning Research*, pages 1723–1742. PMLR, 2017.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *ICML*, pages 1015–1022. Omnipress, 2010.