Distributional Robust Bayesian Optimization with φ-divergences

Hisham Husain Vu Nguyen Anton van den Hengel

Summary: (i) A theoretical result showing DRO-BO breaks down into a finite dimensional optimization problem (ii) A practical algorithm to solve DRO-BO (iii) A regret analysis that holds for most choices of φ.

Distributional Robust Bayesian Optimization

$$\mathcal{U}_D = \{Q : D(P,Q) \le \varepsilon\}$$

$$\max_x \min_{Q \in \mathcal{U}_D} \mathbb{E}_{Q(c)}\left[f(x,c)\right]$$

$D = \text{MMD}$ →

Kernel Distributional Robust Bayesian Optimization

1. Computationally expensive: Requires solving the minimax directly
2. Limited to the setting when contexts are finite.

Can we solve the DRO-BO more efficiently with another choice of D?

($\varphi-$)Distributional Robust Bayesian Optimization

Pick $\mathrm{D}_\varphi(P,Q) = \int_{\mathscr{C}} \varphi\left(\frac{dP}{dQ}\right) dQ$

- Total Variation
- Kullback-Leibler Divergence
- χ^2-divergence

$$B^t_\phi(p) := \left\{q \in \Delta(\mathscr{C}) : \mathrm{D}_\phi(q, p_t) \le \varepsilon_t\right\}$$

DRO-BO (Theorem 1)

$$\max_{x \in \mathscr{X}} \inf_{q \in B^t_\phi(p)} \mathbb{E}_{c \sim q}[f(x,c)] = \max_{x \in \mathscr{X}, \lambda \ge 0, b \in \mathbb{R}} \left(b - \lambda\epsilon_t - \lambda \mathbb{E}_{p_t(c)}\left[\varphi^\star\left(\frac{b - f(x,c)}{\lambda}\right)\right]\right)$$

← $D = \chi^2$

χ^2-divergence

$$\sup_{x \in \mathscr{X}} \left(\mathbb{E}_{p_t(c)}[f(x,c)] - \sqrt{\epsilon_t \cdot \mathrm{Var}_{p_t(c)}[f(x,c)]}\right)$$

$D = \text{TV}$ →

Total Variation Distance

$$\sup_{x \in \mathscr{X}} \left(\mathbb{E}_{p_t(c)}[f(x,c)] - \frac{\epsilon_t}{2}\left(\sup_{c \in \mathscr{C}} f(x,c) - \inf_{c \in \mathscr{C}} f(x,c)\right)\right)$$

Regret Analysis

$$R_T(\varphi) \le \frac{\sqrt{8T\beta_T\gamma_T}}{\log(1 + \sigma_f^{-2})} + M \sum_{t=1}^{T} \Gamma_\varphi(\varepsilon_t)$$

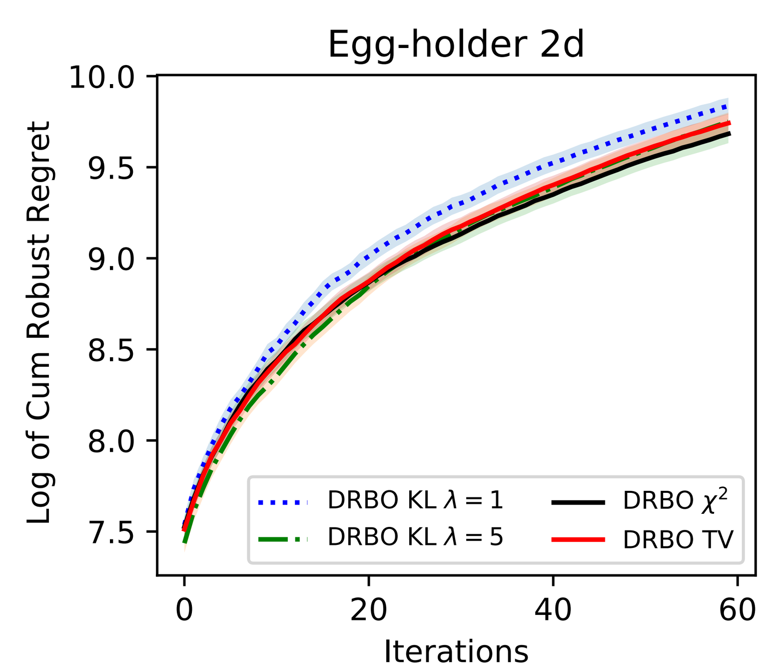
$M = \sup_{(x,c) \in \mathscr{X} \times \mathscr{C}} \left|f(x,c)\right| < \infty$

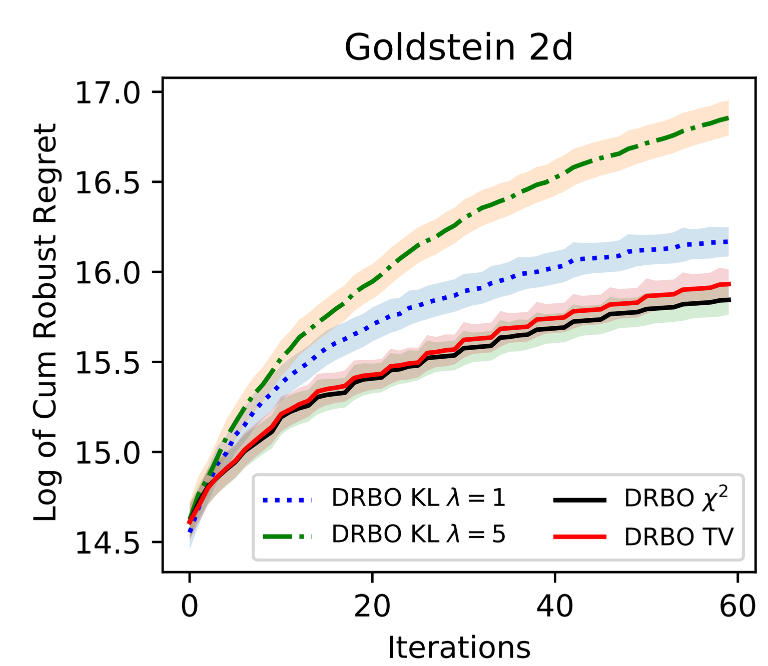
$f \in$ RKHS

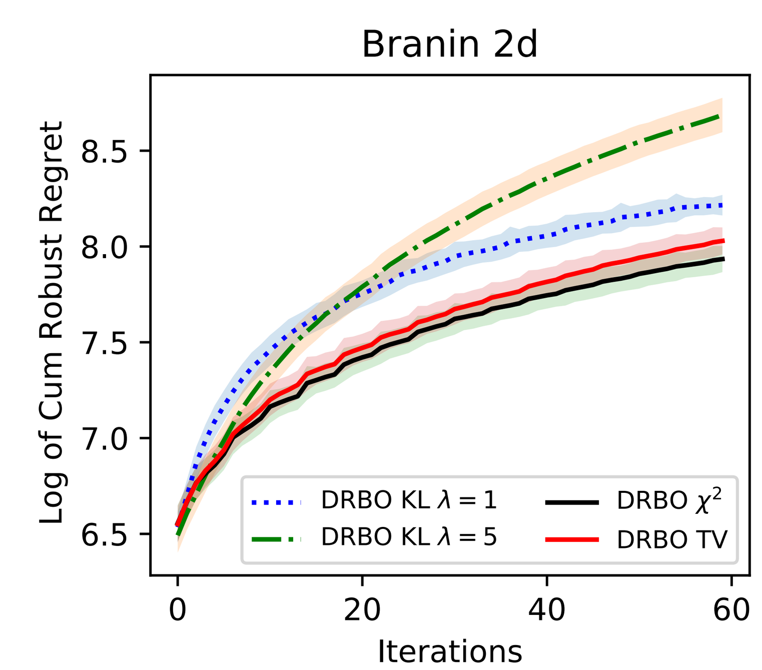
$\varphi : \mathbb{R} \to (-\infty, \infty]$

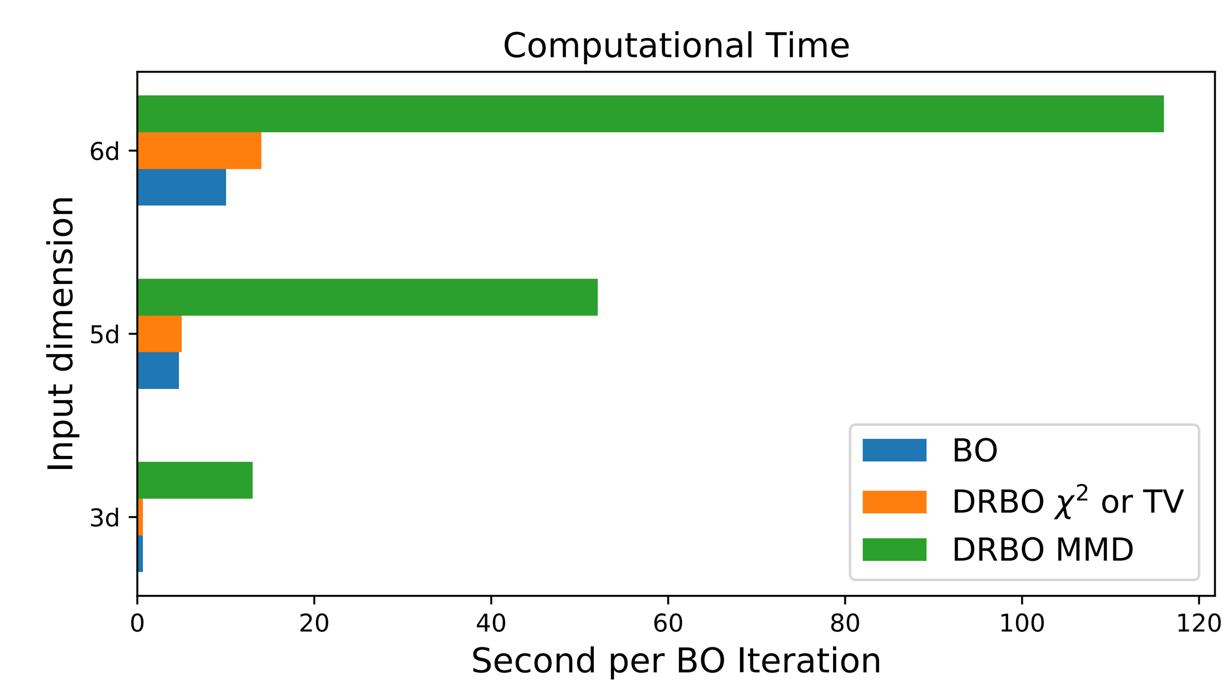
$\Gamma_\varphi : [0, \infty) \to \mathbb{R}$

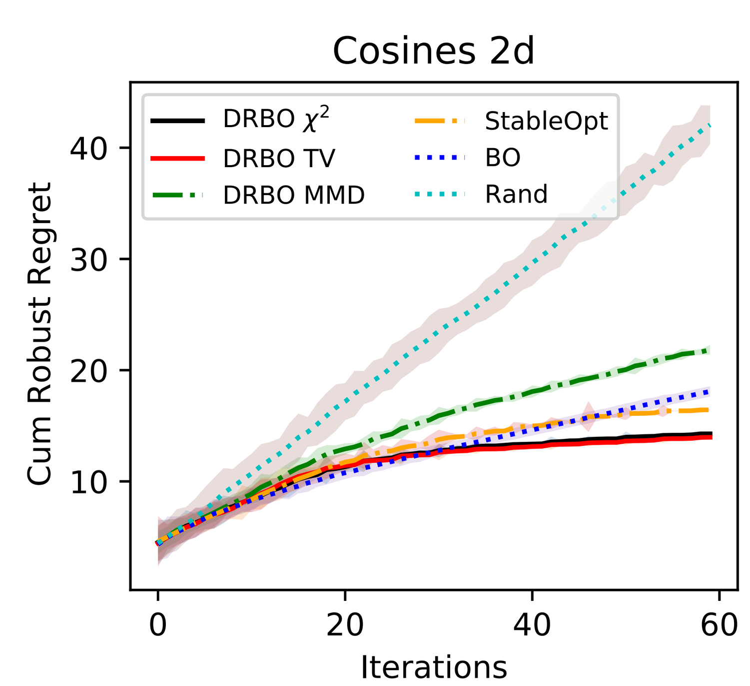
$\mathrm{TV}(p,q) \le \Gamma_\varphi(\mathrm{D}_\varphi(p,q))$

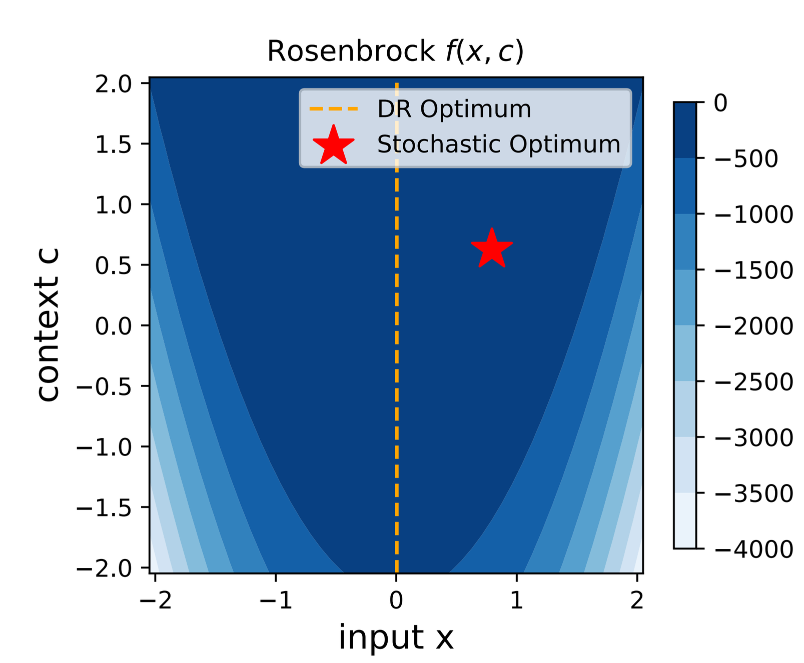


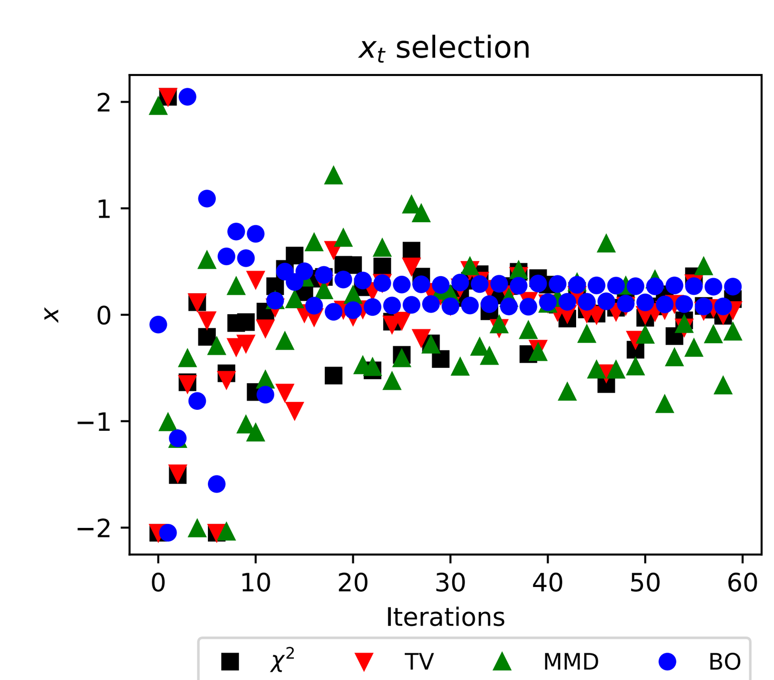


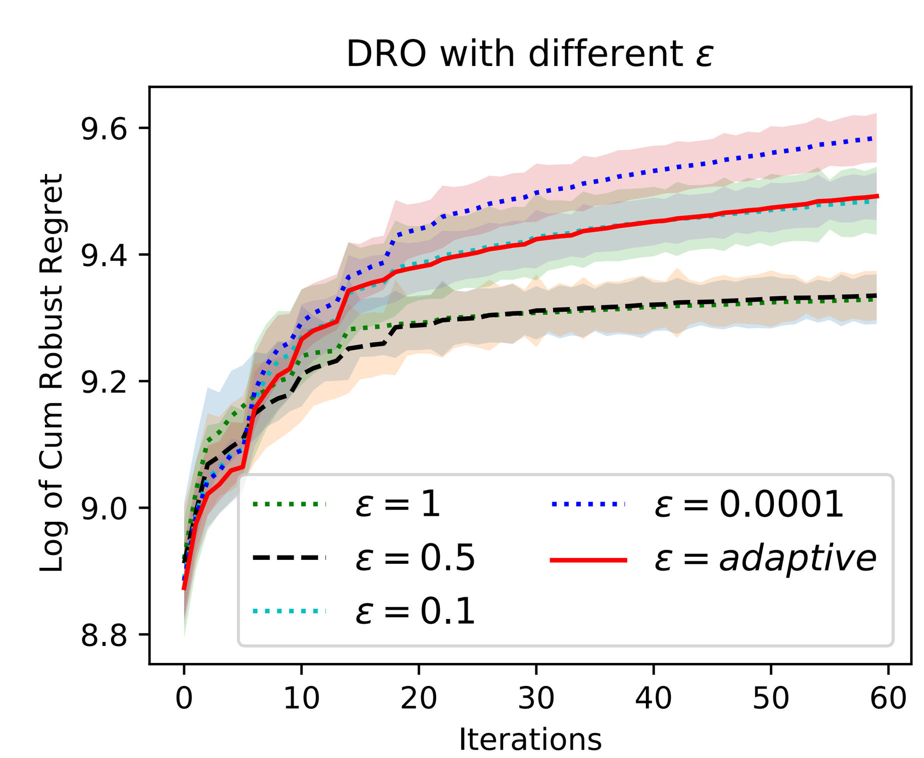












hushisha@amazon.com