Optimal Scalarizations for Sublinear Hypervolume Regret

Qiuyi (Richard) Zhang [1]

Abstract

Scalarization is a general technique that can be deployed in any multiobjective setting to reduce multiple objectives into one, such as recently in RLHF for training reward models that align human preferences. Yet some have dismissed this classical approach because linear scalarizations are known to miss concave regions of the Pareto frontier. To that end, we aim to find simple nonlinear scalarizations that can explore a diverse set of k objectives on the Pareto frontier, as measured by the dominated hypervolume. We show that hypervolume scalarizations with uniformly random weights are surprisingly optimal for provably minimizing the hypervolume regret, achieving an optimal sublinear regret bound of $O(T^{-1/k})$, with matching lower bounds that preclude any algorithm from doing better asymptotically. As a theoretical case study, we consider the multiobjective stochastic linear bandits problem and demonstrate that by exploiting the sublinear regret bounds of the hypervolume scalarizations, we can derive a novel non-Euclidean analysis that produces improved hypervolume regret bounds of $\widetilde{O}(dT^{-1/2} + T^{-1/k})$. We support our theory with strong empirical performance of using simple hypervolume scalarizations that consistently outperforms both the linear and Chebyshev scalarizations, as well as standard multiobjective algorithms in bayesian optimization, such as EHVI.

1. Introduction

Optimization objectives in modern AI systems are becoming more complex with many different components that must be combined to perform precise tradeoffs in machine learning models. Starting from standard ℓ_p regularization objectives in regression problems (Kutner et al., 2005) to increasingly multi-component losses used in reinforcement learning (Sutton et al., 1998) and deep learning (LeCun et al., 2015), many of these single-objective problems are phrased as a scalarized form of an inherently multiobjective problem.

Practitioners often vary the weights of the scalarization method, with the main goal of exploring the entire *Pareto frontier*, which is the set of optimal objectives that cannot be simultaneously improved. First, one chooses some weights $\lambda \in \mathbb{R}^k$ and scalarization functions $s_\lambda(y) : \mathbb{R}^k \to \mathbb{R}$ that convert k multiple objectives $F(a) := (f_1(a), ..., f_k(a))$ over some parameter space $a \in \mathcal{A} \subseteq \mathbb{R}^d$ into a single-objective scalar. Optimization is then applied to this family of single-objective functions $s_\lambda(F(x))$ for various λ and since we often choose s_λ to be monotonically increasing in all coordinates, $x_\lambda = \arg\max_{a\in\mathcal{A}} s_\lambda(F(a))$ is on the Pareto frontier and the various choices of λ recovers an approximation to the Pareto frontier (Paria et al., 2018).

Due to its simplicity of use, many have turned to a heuristic-based scalarization strategy to pick the family of scalarizer and weights, which efficiently splits the multi-objective optimization into numerous single "scalarized" optimizations (Roijers et al., 2013). Linear scalarizations with varying weights are often used in multi-objective optimization problems, such as in multi-objective reinforcement learning to combine task reward with the negative action norm (Abdolmaleki et al., 2021) or in RLHF to align responses with human preferences (Ouyang et al., 2022). Furthermore, some works have proposed piecewise linear scalarizations inspired by economics (Busa-Fekete et al., 2017), while for multi-armed bandits, scalarized knowledge gradient methods empirically perform better with non-linear scalarizations (Yahyaa et al., 2014). Other works have come up with novel scalarizations that perform better empirically in some settings (Aliano Filho et al., 2019; Schmidt et al., 2019). In general, previous works have tried to do comparisons between different scalarizations but with varying conclusions (Kasimbeyli et al., 2019).

However, the appeal of using scalarizations in multiobjective optimization largely declined as linear scalarizations are shown to be provably incapable of exploring the full Pareto frontier (Boyd and Vandenberghe, 2004; Emmerich and Deutz, 2018). This has led to a flurry of recent developments in specific multi-objective algorithms tailored

[1]Google Deepmind. Correspondence to: Qiuyi Zhang <qiuyiz@google.com>.

The Many Facets of Preference Learning Workshop at the International Conference on Machine Learning (ICML), Honolulu, Hawaii, USA, 2023.

to specific settings such as ParEgo (Knowles, 2006) and MOEAD (Zhang and Li, 2007) for black-box optimization or multivariate iteration for reinforcement learning (Yang et al., 2019). Furthermore, many adaptive reweighting strategies have been proposed in order to target or explore the full Pareto frontier, which have connections to gradient-based multi-objective optimization; however these strategies are much more complicated to implement and produce higher runtimes due to the addition logic (Lin et al., 2019; Abdolmaleki et al., 2021). This begs the question of

Are simple scalarization methods at all competitive and if so, how would one optimally choose them?

To judge the effectiveness of an multiobjective optimizer, a natural and widely used metric to measure progress is the *hypervolume indicator*, which is the volume of the dominated portion of the Pareto set (Zitzler and Thiele, 1999; Shah and Ghahramani, 2016). The hypervolume metric has become a gold standard because it has strict Pareto compliance meaning that if set A is a subset of B and B has at least one Pareto point not in A, then the hypervolume of B is greater than that of A. Therefore, it is of no surprise that multiobjective optimization methods often use hypervolume related metrics for progress tracking or acquisition optimization, such as the Expected Hypervolume Improvement (EHVI) or its differentiable counterpart (Daulton et al., 2020; Hupkens et al., 2015; Emmerich and Deutz, 2018).

In previous works, we note that the notion of optimality becomes varied. Previous work by (Lu et al., 2019) proved Pareto regret bounds of $O(d\sqrt{T})$, but that only guarantees recovery of a single point close to the Pareto frontier. Some works minimize a notion of distance to the Pareto frontier, such as the ℓ_∞ norm (Auer et al., 2016), although such approaches work in the finite multi-arm bandit setting which mandates at least a pull of each arm. Some recent works provide sub-linear hypervolume regret bounds which guarantees convergence to the full Pareto frontoer; however, they are exponential in k and its analysis only applies to a specially tailored algorithm that requires an unrealistic classification step (Zuluaga et al., 2013). Most relevant is recent work by (Golovin and Zhang, 2020) that introduces random hypervolume scalarizations and when combined with our generalization bounds, one can directly derive a $O(k^2 d/\sqrt{T} + T^{-1/O(k)})$ convergence bound for Gaussian Process bandits.

1.1. Our Contributions

We show, perhaps surprisingly, that a simple ensemble of hypervolume scalarizations, first introduced in (Golovin and Zhang, 2020), are theoretically optimal to minimize hypervolume regret and are empirically competitive for general multiobjective optimization. Specifically, we show that the hypervolume scalarization has sharp level curves that allows for the targeting of a specific part of the Pareto frontier, without any convexity assumptions or the need for adaptively changing weights. Theoretically, we show that exploring the Pareto frontier by choosing T maximizers of randomly weighted hypervolume scalarizations achieves a sublinear hypervolume regret rate of $O(T^{-1/k})$, where T is the number of points sampled. Our proofs follow from novel arguments that combine the Lipschitz properties of the hypervolume scalarizations with classic metric entropy bounds for L-Lipschitz functions in $\mathbb{R}^k$.

We observe that our derived sublinear hypervolume regret rate of the hypervolume scalarization holds for any Pareto frontier, regardless of the inherent multiobjective function F or the underlying optimizer. Therefore, we emphasize that analyzing these model-agnostic rates can be a general theoretical tool to compare and analyze the effectiveness of proposed multiobjective algorithms. In fact, although many scalarizers will search the entire Pareto frontier as $T \to \infty$, the rate at which this convergence occurs can differ significantly, implying that this framework paves the road for a theoretical standard by which to judge the effectiveness of advanced strategies, such as adaptively weighted scalarizations. On the other hand, we show surprisingly that no multiobjective algorithm, whether scalarized, adaptive, or not, can beat the optimal hypervolume regret rates of applying single-objective optimization with the hypervolume scalarization.

To accomplish this, we prove novel lower bounds showing one cannot hope for a better convergence rate due to the exponential nature of our regret, for any set of T points. Specifically, we show that the hypervolume regret of any algorithm after T actions is at least $\Omega(T^{-1/k})$, demonstrating the necessity of the $O(T^{-1/k})$ term up to small constants in the denominator. As a corollary, we leverage the sublinear regret properties of hypervolume scalarization to transfer our lower bounds to the more general setting of scalarized Bayes regret. Together, we demonstrate that for general multiobjective optimzation, finding maximas of the hypervolume scalarizations with a uniform weight distribution optimally finds the Pareto frontier asymptotically.

Theorem 1 (Informal Restatement of Theorem 6 and Theorem 7)**.** *Let $\mathbf{Y}_T = \{y_1, ..., y_T\}$ be a set of T points in $\mathbb{R}^k$ such that $y_i \in \arg\max_{y \in \mathcal{Y}} s^{\text{HV}}_{\lambda_i}(y)$ with $\lambda_i \sim \mathcal{S}_+$ randomly drawn i.i.d. from an uniform distribution and s^{HV} are hypervolume scalarizations. Then, the hypervolume regret satisfies*

$$\mathcal{HV}(\mathcal{Y}^\star) - \mathcal{HV}(\mathbf{Y}_T) = O(T^{-\frac{1}{k}})$$

where $\mathcal{Y}^\star$ is the Pareto frontier and $\mathcal{HV}$ is the hypervolume function. Furthermore, any algorithm for choosing these T points must suffer hypervolume regret of at least $\Omega(T^{-\frac{1}{k}})$.

Next, we use a novel non-Euclidean analysis to prove im-

proved hypervolume regret bounds for our theoretical toy model: the classic *stochastic linear bandit* setting. For any scalarization and weight distribution, we propose a new scalarized algorithm (Algorithm 1) for multiobjective stochastic linear bandit that combines uniform exploration and exploitation via an UCB approach to provably obtain scalarized Bayes regret bounds, which we then combine with the hypervolume scalarization to derive optimal hypervolume regret bounds. Specifically, for any scalarization s_λ, we show that our algorithm in the linear bandit setting has a scalarized Bayes regret bound of $\widetilde{O}(L_p k^{1/p} d T^{-1/2} + T^{-1/k})$, where L_p is the Lipschitz constant of the $s_\lambda(\cdot)$ in the ℓ_p norm. Finally, by using hypervolume scalarizations and exploiting their ℓ_∞-smoothness, we completely remove the dependence on the number of objectives, k, which had a polynomial dependence in previous regret bound given by (Golovin and Zhang, 2020).

Theorem 2 (Informal Restatement of Theorem 8)**.** *Let $\mathbf{A}_T \subseteq \mathcal{A}$ be the actions generated by T rounds of Algorithm 1, then our hypervolume regret is bounded by:*

$$\mathcal{HV}_z(\Theta^\star \mathcal{A}) - \mathcal{HV}_z(\Theta^\star \mathbf{A}_T) \leq \widetilde{O}(dT^{-\frac{1}{2}} + T^{-\frac{1}{k}})$$

Guided by our theoretical analysis, we empirically evaluate a diverse combination of scalarizations and weight distributions with our proposed algorithm for multiobjective linear bandits. Our experiments show that for some settings of linear bandits, in spite of a convex Pareto frontier, applying linear or Chebyshev scalarizations naively with various weight distributions leads to suboptimal hypervolume progress, especially when the number of objective increase to exceed $k \geq 5$. This is because the non-uniform curvature of the Pareto frontier, exaggerated by the curse of dimensionality and combined with a stationary weight distribution, hinders uniform progress in exploring the frontier. Although one can possibly adapt the weight distribution to the varying curvature of the Pareto frontier when it is convex, we suggest remediating the issue by simply adopting the use of non-linear scalarizations that are more robust to the choice of weight distribution and are theoretically sound.

For general multiobjective optimization, we perform empirical comparisons on BBOB benchmarks for biojective functions in a bayesian optimization setting, using classical Gaussian Process models (Williams and Rasmussen, 2006). When comparing EHVI with hypervolume scalarization approaches, we find that EHVI tends to limit its hypervolume gain by over-focusing on the central portion of the Pareto frontier, whereas the hypervolume scalarization encourages a diverse exploration of the extreme ends. From our broader analysis, we recommend the use of hypervolume scalarizations as a simple, general, efficient, non-adaptive method to perform various multiobjective optimization, even in complex settings, such as reinforcement learning. We believe that this is especially relevant given the modern era of learning algorithms that commonly makes tradeoffs between multiple objectives such as fairness, privacy, latency.

2. Problem Setting and Notation

We assume, for sake of normalization, that $\|\Theta_i^\star\| \leq 1$ and that $\|a_t\| \leq 1$, where $\|\cdot\|$ denotes the ℓ_2 norm unless otherwise stated. Other norms that are used include the classical ℓ_p norms $\|\cdot\|_p$ and matrix norms $\|x\|_{\mathbf{M}} = x^\top \mathbf{M} x$ for a positive semi-definite matrix $\mathbf{M}$. For a scalarization function $s_\lambda(x)$, s_λ is L_p-Lipschitz with respect to x in the ℓ_p norm on $\mathcal{X}$ if for $x_1, x_2 \in \mathcal{X}$, $|s_\lambda(x_1) - s_\lambda(x_2)| \leq L_p \|x_1 - x_2\|_p$, and analogously for λ. We let $\mathcal{S}_+^{k-1} = \{y \in \mathbb{R}^k \,|\, \|y\| = 1, y > 0\}$ be the sphere in the positive orthant and by abuse of notation, we also let $y \sim \mathcal{S}_+^{k-1}$ denote that y is drawn uniformly on $\mathcal{S}_+^{k-1}$.

For two outputs $y, z \in \mathcal{Y} \subseteq \mathbb{R}^k$, we say that y is *Pareto-dominated* by z if $y \leq z$ and there exists j such that $y_j < z_j$, where $y \leq z$ is defined for vectors element-wise. A point is *Pareto-optimal* if no point in the output space $\mathcal{Y}$ can dominates it. Let $\mathcal{Y}^\star$ denote the set of Pareto-optimal points (objectives) in $\mathcal{Y}$, which is also known as the *Pareto frontier*. Our main progress metrics for multiobjective optimization is given by the standard hypervolume indicator. For $S \subseteq \mathbb{R}^k$ compact, let $\mathrm{vol}(S)$ be the regular hypervolume of S with respect to the standard Lebesgue measure.

Definition 3. *For $\mathcal{Y} \subseteq \mathbb{R}^k$, we define the (dominated)* **hypervolume indicator** *of $\mathcal{Y}$ with respect to reference point z as: $\mathcal{HV}_z(\mathcal{Y}) = \mathrm{vol}(\{x \,|\, x \geq z, x \text{ is dominated by some } y \in \mathcal{Y}\})$*

We can formally phrase our optimization objective as trying to rapidly minimize the hypervolume (psuedo-)regret. Let $\mathcal{A}$ be our action space and for some general multi-objective function F, let $\mathcal{Y}$ be the image of $\mathcal{A}$ under F. Let $\mathbf{A}_T$ be any matrix T actions and let $\mathbf{Y}_T = F(\mathbf{A}_T) \subseteq \mathbb{R}^k$ be the k objectives corresponding. Then, the hypervolume regret of actions $\mathbf{A}_T$, with respect to the reference point z, is given by: $\text{Hypervolume-Regret}(\mathbf{A}_T) = \mathcal{HV}_z(\mathcal{Y}^\star) - \mathcal{HV}_z(\mathbf{Y}_T)$

2.1. Scalarizations for Multiobjective Optimization

For multiobjective optimization, we generally only consider *monotone* scalarizers that have the property that if $y > z$, then $s_\lambda(y) > s_\lambda(z)$ for all λ. Note this implies that an unique optimal solution to the scalarized optimization is on the Pareto frontier. A common scalarization used widely in practice is the linear scalarization: $s_\lambda^{\text{LIN}}(y) = \lambda^\top y$ for some chosen positive weights $\lambda \in \mathbb{R}^k$. By Lagrange duality and hyperplane separation of convex sets, one can show that any convex Pareto frontier can be characterized fully by an optimal solution for some weight settings.

However, linear scalarizations cannot recover the non-convex regions of Pareto fronts since the linear level curves can only be tangent to the Pareto front in the protruding convex regions (see Figure 1). To overcome this drawback, another scalarization that is proposed is the Chebyshev scalarization: $s_\lambda^{\text{CS}}(y) = \min_i \lambda_i y_i$. Indeed, one can show that the sharpness of the scalarization, due to its minimum operator, can discover non-convex Pareto frontiers (Emmerich and Deutz, 2018).

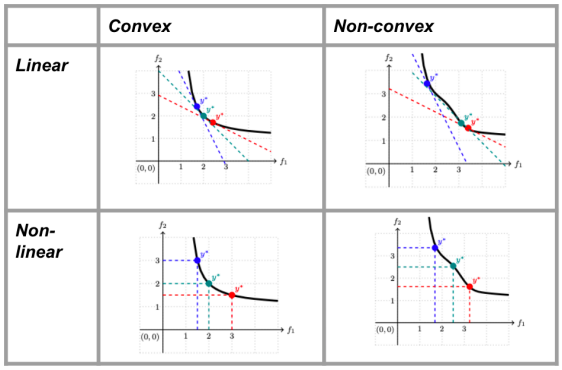


Figure 1. Comparisons of the scalarized minimization solutions with various weights for a multiobjective problem with convex and non-convex Pareto fronts. Different colors represent different weights; the dots are scalarized optima and the corresponding dotted lines represent level curves. Linear scalarization does not have an optima in the concave region of the Pareto front for any set of weights, but the non-linear scalarization, with its sharper level curves, can discover the whole Pareto front (Emmerich and Deutz, 2018).

3. Hypervolume Scalarizations

In this section, we show the utility and optimality of a related scalarization known as the hypervolume scalarization, $s_\lambda^{\text{HV}}(y) = \min_i (y_i/\lambda_i)^k$ that was introduced in (Golovin and Zhang, 2020). First, observe that this scalarization allows you to target a specific part of the Pareto frontier, which eliminates the need of adaptive targeting techniques that heuristically update parameters of the optimization objectives. The visualization of the non-linear level curves of the scalarization provides intuition that our scalarization targets the portion of the Pareto frontier in the direction of λ for any $\lambda > 0$ (see Figure 2), as we can show that the tangent point of the level curves of the scalarization is always on the vector in the direction of λ.

Lemma 4. *For any point $y^\star$ on the Pareto frontier of any set $\mathcal{Y}$ that lies in the positive orthant, there exists $\lambda > 0$ such that $y^\star = \arg\max_{y \in \mathcal{Y}} s_\lambda^{\text{HV}}(y)$. Furthermore, for any $\alpha, \lambda > 0$ such that $\alpha\lambda$ is on the Pareto frontier, then $\alpha\lambda \in \arg\max_{y \in \mathcal{Y}} s_\lambda^{\text{HV}}(y)$.*

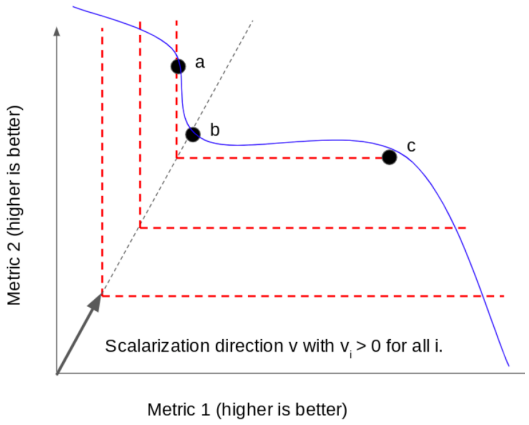


Figure 2. The dotted red lines represent the level curves of the hypervolume scalarization with weights $\lambda = v$. Hence, the scalarization is able to dig into the Pareto frontier and discover b, whereas the linear scalarization would prefer a or c. Furthermore, the optima is exactly the Pareto point that is in the direction of v.

Furthermore, this scalarization additionally has the special property that the expected scalarized value under a uniform weight distribution on $\mathcal{S}_+^{k-1}$ gives the dominated hypervolume, up to a constant scaling factor. Intuitively, this lemma says that the optima of the hypervolume scalarization over some uniform choice of weights will be sufficiently diverse for any Pareto set so as to capture its dominated hypervolume.

Lemma 5 ((Golovin and Zhang, 2020)). *Let $\mathbf{Y}_T = \{y_1, ..., y_T\}$ be a set of T points in $\mathbb{R}^k$. Then, the hypervolume with respect to a reference point z is given by:*

$$\mathcal{HV}_z(\mathbf{Y}_T) = c_k \mathop{\mathbf{E}}_{\lambda \sim \mathcal{S}_+^{k-1}} \left[\max_{y \in \mathbf{Y}_T} s_\lambda^{\text{HV}}(y - z) \right]$$

where $c_k = \frac{\pi^{k/2}}{2^k\, \Gamma(k/2+1)}$ is a constant depending only on k.

While this lemma captures useful properties of the scalarization in the infinite limit, we supplement it by showing that finite asymptotic bounds on the strongly sublinear convergence rate of using this scalarization in optimization. In fact, while many scalarizations will eventually explore the whole Pareto frontier in the infinite limit, the rate at which the exploration improves the hypervolume is not known, and in the worst case might be exponentially slow. We show that the simple procedure of optimizing hypervolume scalarizations with a uniform weight distribution is a *sublinear hypervolume regret* multiobjective algorithm in that it satisfies $O(T^{-\epsilon})$ hypervolume regret convergence rates when $\mathcal{Y}$ is known. We note that this rate is agnostic of the underlying

optimization algorithm or objective function, meaning this is a general property of the scalarization.

Our novel proof of convergence uses a symmetry argument and exploits the Lipschitz properties of s_λ^{HV} to derive generalization bounds via metric entropy. Proving smoothness properties of our hypervolume scalarizations for any $\lambda > 0$ with λ normalized on the unit sphere is non-obvious as $s_\lambda^{\text{HV}}(y)$ depends inversely on λ_i so when λ_i is small, s_λ^{HV} might change wildly. However, the crucial observation is that λ_i being small makes it unlikely that it becomes the minimum coordinate, implying that it is not contributing to the scalarized value or its rate of change.

Theorem 6 (Sublinear Hypervolume Regret)**.** *Let $\mathbf{Y}_T = \{y_1, ..., y_T\}$ be a set of T points in $\mathbb{R}^k$ such that $y_i \in \arg\max_{y \in \mathcal{Y}} s_{\lambda_i}^{\text{HV}}(y)$ with $\lambda_i \sim \mathcal{S}_+$ i.i.d. drawn. Then with probability at least $1-\delta$ over the randomness of λ_i, the hypervolume of $\mathbf{Y}_T$ with respect to a reference point z satisfies sublinear hypervolume regret bounds*

$$\mathcal{HV}_z(\mathcal{Y}^\star) - \mathcal{HV}_z(\mathbf{Y}_T) = O(T^{-\frac{1}{k+1}} + \sqrt{\ln(1/\delta)}T^{-\frac{1}{2}})$$

3.1. Lower Bounds and Optimality

The dominating factor in our derived convergence rate is the $O(T^{-1/(k+1)})$ term and we show that this cannot be improved. Over all subsets $\mathbf{Y}_T \subseteq \mathcal{Y}$ of size T, note that our optimal convergence rate is given by the the subset that maximizes the dominated hypervolume of $\mathbf{Y}_T$, although finding this is in fact a NP-hard problem due to reduction to set cover. By constructing a lower bound via a novel packing argument, we show that even this optimal set would incur at least $\Omega(T^{-1/k})$ regret, implying that our convergence rates, derived from generalization rates when empirically approximating the hypervolume, are optimal.

Theorem 7 (Hypervolume Regret Lower Bound)**.** *There exists a setting of linear objective parameters $\Theta^\star$ and $\mathcal{A} = \{a : \|a\| = 1\}$ such that for any actions $\mathbf{A}_T$, the hypervolume regret at $z = 0$ after T rounds is*

$$\mathcal{HV}_z(\Theta^\star\mathcal{A}) - \mathcal{HV}_z(\Theta^\star\mathbf{A}_T) = \Omega(T^{-\frac{1}{k-1}})$$

3.2. Multiobjective Stochastic Linear Bandits

We propose a simple scalarized algorithm for linear bandits and provide a novel ℓ_p analysis of the hypervolume regret that removes the polynomial dependence on k in the scalarized regret bounds. When combined with the ℓ_∞ sharpness of the hypervolume scalarization, this analysis gives an optimal $O(d/\sqrt{T})$ bound on the scalarized regret, up to $\log(k)$ factors. This $\log$ dependence on k is perhaps surprising but is justified information theoretically since each objective is observed separately. Our scalarized algorithm works despite of noise in the observations, which makes it difficult to even statistically infer measures of hypervolume progress. Many of the notation and intermediate theorems in this section are given in the Appendix.

Algorithm 1: EXPLOREUCB$(T, \mathcal{D}, s_\lambda)$: Scalarized UCB for Linear Bandits

Input : number of maximum actions T, weight distribution $\mathcal{D}$, scalarization s_λ

Initialize iteration counter $n = 1$
repeat
  Play action $e_i \in \mathcal{E}$ for $i \equiv n \mod d$ and increment $n \leftarrow n + 1$
  Let C_{ti} be the confidence ellipsoid for Θ_i and let $UCB_i(a) = \max_{\theta \in C_i} \theta^\top a$
  Sample $\lambda \sim \mathcal{D}$ and play action that maximizes $a^* = \text{argmax}_{a \in \mathcal{A}}\, s_\lambda(UCB_i(a))$
until *number of actions exceed T*

Theorem 8 (HyperVolume Regret of EXPLOREUCB)**.** *Let $z \in \mathbb{R}^k$ be a reference point such that over all $a \in \mathcal{A}$, $B = \min_a \Theta^\star a - z$ is positive. Then, with constant probability, running Algorithm 1 with $s_\lambda^{\text{HV}}(y)$ and $\mathcal{D} = \mathcal{S}_+$ gives hypervolume regret bound of $\mathcal{HV}_z(\Theta^\star\mathcal{A}) - \mathcal{HV}_z(\Theta^\star\mathbf{A}_T) \leq O\left(c_k \frac{(B+2)^k}{Bk^{k/2-1}} d\sqrt{\frac{\log(kT)}{T}} + c_k \frac{(B+2)^{2k+1}}{Bk^{k-1/2}T^{1/(k+1)}}\right)$*

For k constant, the hypervolume regret satisfies

$$\mathcal{HV}_z(\Theta^\star\mathcal{A}) - \mathcal{HV}_z(\Theta^\star\mathbf{A}_T) \leq \widetilde{O}\left(dT^{-1/2} + T^{-\frac{1}{k+1}}\right)$$

4. Experiments

In this section, we empirically justify our theoretical results by running Algorithm 1 with multiple scalarizations and weight distributions in different settings of the multiobjective stochastic linear bandits. Our empirical results highlight the advantage of the hypervolume scalarization with uniform weights in maximizing the diversity and hypervolume of the resulting Pareto front when compared with other scalarizations and weight distributions, especially when there are a mild number of output objectives k. Our experiments are not meant to show that scalarizations is the only or even the best way to solve multiobjective optimization; rather, it is a simple yet competitive baseline when optimal scalarizations and weight distributions are chosen for solving multiobjective optimization in a variety of settings.

4.1. Stochastic Linear Bandits

We compare the three widely types of scalarizations that were previously mentioned: the linear, Chebyshev, and the

hypervolume scalarization. Note that we use slightly altered form of our hypervolume scalarization as $s_\lambda^{\text{HV}}(y) = \min_i y_i/\lambda_i$, which is a simply a monotone transform of the proposed scalarization and does not inherently affect the optimization. We set our reference point to be $\mathbf{z} = -\mathbf{2}$ in k dimension space, since our action set of $\mathcal{A} = \{a : \|a\| = 1\}$ and our norm bound on $\Theta^\star$ ensures that our rewards are in $[-1, 1]$. In conjuction with the scalarizer, we use our weight distribution $\mathcal{D} = \mathcal{S}_+$, which samples vectors uniformly across the unit sphere. In addition, we also compare this with the bounding box distribution methods that were suggested by (Paria et al., 2018), which samples from the uniform distribution from the min to the max each objective and requires some prior knowledge of the range of each objective (Hakanen and Knowles, 2017). We run our algorithm with inherent dimension $d = 4$ for $T = 100, 200$ rounds with $k = 2, 6, 10$.

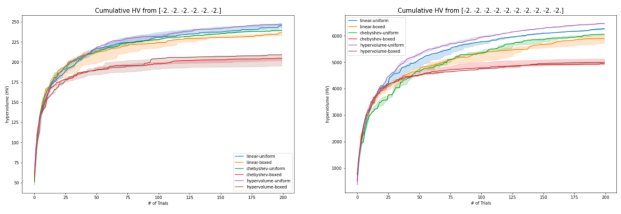


Figure 3. Comparisons of the cumulative hypervolume plots with some anti-correlated θ. When the output dimension increase, there is a clearer advantage to using the hypervolume scalarization over the linear and Chebyshev scalarization. We find that the boxed weight distribution does consistently worse than the uniform distribution.

As expected, we find the hypervolume scalarization consistently outperforms the Chebyshev and linear scalarizations, with linear scalarization as the worst performing (see Figure 3). Note that when we increase the output dimension of the problem by setting $k = 10$, the hypervolume scalarization shows a more distinct advantage. The boxed distribution approach of (Paria et al., 2018) does not seem to fare well and consistently performs worse than its uniform counterpart. While linear scalarization provides relatively good performance when the number of objective $k \leq 5$, it appears that as the number of objectives increase in multi-objective optimization, more care needs to be put into the design of scalarization and their weights due to the curse of dimensionality, since the regions of non-uniformity will exponentially increase.

4.2. BBOB Functions

We empirically demonstrate the competitiveness of hypervolume scalarizations for Bayesian Optimization by comparing them to the popular BO method of EHVI (Hupkens et al., 2015). Running our proposed multiobjective algorithms on the Black-Box Optimization Benchmark (BBOB) functions, which can be paired up into multiple bi-objective optimization problems (Tušar et al., 2016). Our goal is to use a wide set of non-convex benchmarks to supplement our experiments on our simple toy example of linear bandits. For scalarized approaches, we use hypervolume scalarizations with the scalarized UCB algorithm (Golovin and Zhang, 2020) with a constant standard deviation multiplier of 1.8 and all algorithms with use a Gaussian Process as the underlying model with a standard Matérn kernel that is tuned via ARD (Williams and Rasmussen, 2006). Our objectives are given by BBOB functions, which are usually non-negative and are minimized. The input space is always a compact hypercube $[-5, 5]^n$ and the global minima is often at the origin. For bi-objective optimization, given two different BBOB functions f_1, f_2, we attempt to maximize the hypervolume spanned by $(-f_1(x_i), -f_2(x_i))$ over choices of inputs x_i with respect to the reference point $(-5, -5)$. We normalize each function due to the drastically different ranges and add random observation noise, as well as applying vectorized shifts.

We run each of our algorithms in dimensions $d = 8, 16, 24$ and optimize for 160 iterations with 5 repeats. From our results, we see that both EHVI and UCB with hypervolume scalarizations are competitive on the BBOB problems but the scalarized UCB algorithm seems to be able to explore the extreme ends of the Pareto frontier, whereas EHVI tends to favor points in the middle (see Figure 4). From our experiments, this trend appears to be consistent across different functions and is more prominent as the input dimensions d increase, as shown by additional plots given in the appendix.

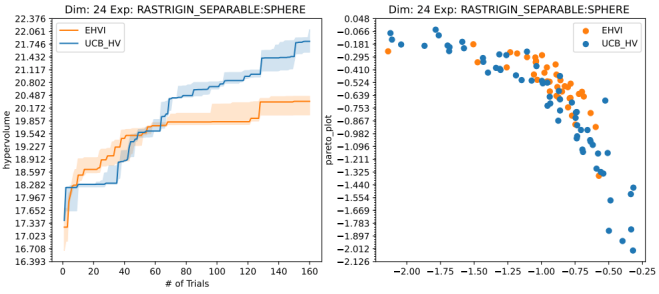


Figure 4. Comparisons of the hypervolume indicator and the optimization fronts with BBOB functions. The left plot tracks the dominated hypervolume as a function of trials that were evaluated. The blue/orange dots represent the frontier points of the UCB-HV/EHVI algorithms respectively, over 5 repeats. Especially in high dimensions, EHVI tends overly concentrate on points in the middle of the frontier, limiting its hypervolume gain, while hypervolume scalarizations produce more diverse points.

References

Michael H Kutner, Christopher J Nachtsheim, John Neter, William Li, et al. *Applied linear statistical models*, volume 5. McGraw-Hill Irwin Boston, 2005.

Richard S Sutton, Andrew G Barto, et al. *Introduction to reinforcement learning*, volume 2. MIT press Cambridge, 1998.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

Biswajit Paria, Kirthevasan Kandasamy, and Barnabás Póczos. A flexible framework for multi-objective bayesian optimization using random scalarizations. *arXiv preprint arXiv:1805.12168*, 2018.

Diederik M Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research*, 48:67–113, 2013.

Abbas Abdolmaleki, Sandy H Huang, Giulia Vezzani, Bobak Shahriari, Jost Tobias Springenberg, Shruti Mishra, Dhruva TB, Arunkumar Byravan, Konstantinos Bousmalis, Andras Gyorgy, et al. On multi-objective policy optimization as a tool for reinforcement learning. *arXiv preprint arXiv:2106.08199*, 2021.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

Róbert Busa-Fekete, Balázs Szörényi, Paul Weng, and Shie Mannor. Multi-objective bandits: Optimizing the generalized gini index. In *International Conference on Machine Learning*, pages 625–634. PMLR, 2017.

Saba Q Yahyaa, Madalina M Drugan, and Bernard Manderick. The scalarized multi-objective multi-armed bandit problem: An empirical study of its exploration vs. exploitation tradeoff. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 2290–2297. IEEE, 2014.

Angelo Aliano Filho, Antonio Carlos Moretti, Margarida Vaz Pato, and Washington Alves de Oliveira. An exact scalarization method with multiple reference points for bi-objective integer linear optimization problems. *Annals of Operations Research*, pages 1–35, 2019.

Marie Schmidt, Anita Schöbel, and Lisa Thom. Min-ordering and max-ordering scalarization methods for multi-objective robust optimization. *European Journal of Operational Research*, 275(2):446–459, 2019.

Refail Kasimbeyli, Zehra Kamisli Ozturk, Nergiz Kasimbeyli, Gulcin Dinc Yalcin, and Banu Icmen Erdem. Comparison of some scalarization methods in multiobjective optimization. *Bulletin of the Malaysian Mathematical Sciences Society*, 42(5):1875–1905, 2019.

Stephen Poythress Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004. ISBN 0-521-83378-7.

Michael TM Emmerich and André H Deutz. A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural computing*, 17(3):585–609, 2018.

Joshua Knowles. Parego: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 10(1):50–66, 2006.

Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on evolutionary computation*, 11(6):712–731, 2007.

Runzhe Yang, Xingyuan Sun, and Karthik Narasimhan. A generalized algorithm for multi-objective reinforcement learning and policy adaptation. *Advances in Neural Information Processing Systems*, 32, 2019.

Xi Lin, Hui-Ling Zhen, Zhenhua Li, Qing-Fu Zhang, and Sam Kwong. Pareto multi-task learning. In *Advances in Neural Information Processing Systems 32*, pages 12037–12047. Curran Associates, Inc., 2019. URL http://papers.nips.cc/paper/9374-pareto-multi-task-learning.pdf.

Eckart Zitzler and Lothar Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE transactions on Evolutionary Computation*, 3(4):257–271, 1999.

Amar Shah and Zoubin Ghahramani. Pareto frontier learning with expensive correlated objectives. In *International conference on machine learning*, pages 1919–1927. PMLR, 2016.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. *Advances in Neural Information Processing Systems*, 33: 9851–9864, 2020.

Iris Hupkens, André Deutz, Kaifeng Yang, and Michael Emmerich. Faster exact algorithms for computing expected hypervolume improvement. In *international conference on evolutionary multi-criterion optimization*, pages 65–79. Springer, 2015.

Shiyin Lu, Guanghui Wang, Yao Hu, and Lijun Zhang. Multi-objective generalized linear bandits. *arXiv preprint arXiv:1905.12879*, 2019.

Peter Auer, Chao-Kai Chiang, Ronald Ortner, and Madalina Drugan. Pareto front identification from stochastic bandit feedback. In *Artificial intelligence and statistics*, pages 939–947. PMLR, 2016.

Marcela Zuluaga, Guillaume Sergent, Andreas Krause, and Markus Püschel. Active learning for multi-objective optimization. In *International Conference on Machine Learning*, pages 462–470, 2013.

Daniel Golovin and Qiuyi Zhang. Random hypervolume scalarizations for provable multi-objective black box optimization. *arXiv preprint arXiv:2006.04655*, 2020.

Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

Jussi Hakanen and Joshua D Knowles. On using decision maker preferences with parego. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 282–297. Springer, 2017.

Tea Tušar, Dimo Brockhoff, Nikolaus Hansen, and Anne Auger. Coco: the bi-objective black box optimization benchmarking (bbob-biobj) test suite. *ArXiv e-prints*, 2016.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Jack Kiefer and Jacob Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960.

Peter L Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482, 2002.

Lee-Ad Gottlieb, Aryeh Kontorovich, and Robert Krauthgamer. Adaptive metric dimensionality reduction. *Theoretical Computer Science*, 620:105–118, 2016.

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24:2312–2320, 2011.

A. Additional Notation for Section 3.2

For various scalarizations and weight distributions, an related measure of progress that attempts to capture the requirement of diversity in the Pareto front is scalarized Bayes regret for some scalarization function s_λ. For some fixed scalarization with weights λ, $s_\lambda : \mathbb{R}^k \to \mathbb{R}$, we can define the instantaneous scalarized (psuedo-)regret as $r(s_\lambda, a_t) = \max_{a \in \mathcal{A}} s_\lambda(\Theta^\star a) - s_\lambda(\Theta^\star a_t)$

Since the scalarized regret is only a function of a single action a_t, it fails to capture the variety of solutions that we would optimize for in the Pareto frontier. To capture some notion of diversity, we must define progress with respect to a set of past actions $\mathbf{A}_t$. Generalizing the scalarized regret above, we can formulate the Bayes regret as an average of the scalarized regret over some distribution of non-negative weight vectors, $\lambda \sim \mathcal{D}$. Specifically, we define the *(scalarized) Bayes regret* with respect to a set of actions $\mathbf{A}_t$ to be: $BR(s_\lambda, \mathbf{A}_t) = \mathbf{E}_{\lambda \sim \mathcal{D}}[\max_{a \in \mathcal{A}} s_\lambda(F(a)) - \max_{a \in \mathbf{A}_t} s_\lambda(F(a))] = \mathbf{E}_{\lambda \sim \mathcal{D}}[\min_{a \in \mathbf{A}_t} r(s_\lambda, a)]$

Unlike previous notions of Bayes regret in literature, we are actually calculating the Bayes regret of a reward function that is maximized with respect to an entire set of actions $\mathbf{A}_t$. Specifically, by maximizing over all previous actions, this captures the notion that during multi-objective optimization our Pareto set is always expanding. We will see later that this novel definition is the right one, as it generalizes to the multi-objective setting in the form of hypervolume regret.

For theory, we use the classic *stochastic linear bandit* setting. For the single-objective setting, in round $t = 1, 2, ..., T$, the learner chooses an action $a_t \in \mathbb{R}^d$ from the action set $\mathcal{A}$ and receives a reward $y_t = \langle \theta^\star, a_t \rangle + \xi_t$ where ξ_t is i.i.d. 1-sub-Gaussian noise and $\theta^\star \in \mathbb{R}^d$ is the unknown true parameter vector. In the *multi-objective stochastic linear bandit* setting, the learner instead receives a vectorized reward $y_t = \Theta^\star a_t + \xi_t$, where $\Theta^\star \in \mathbb{R}^{k \times d}$ is now a matrix of k true parameters and $\xi_t \in \mathbb{R}^k$ is a vector of independent 1-sub-Gaussian noise. We also denote $\mathbf{A}_t \in \mathbb{R}^{d \times t}$ to be the history action matrix, whose i-th column is a_i, the action taken in round i. Similarly, $\mathbf{y}_t$ is defined analogously. Finally, for sake of analysis, we assume that $\mathcal{A}$ contains an isotropic set of actions and specifically, there is $\mathcal{E} \subset \mathcal{A}$ with size $|\mathcal{E}| = O(d)$ such that $\sum_i e_i e_i^\top \succeq \frac{1}{2}\mathbf{I}$, where $\succeq$ denotes the PSD ordering on symmetric matrices. This assumption is not restrictive, as it can be relaxed by using optimal design for least squares estimators (Lattimore and Szepesvári, 2020) and the Kiefer-Wolfowitz Theorem (Kiefer and Wolfowitz, 1960), which guarantees the existence and construction of an uniform exploration basis of size $O(\text{poly}(d))$.

B. Additional Theorems for Section 3.2

By using the confidence ellipsoids given by the UCB algorithm, we can determine each objective parameter $\Theta_i^\star$, up to a small error. To bound the scalarized regret, we utilize the ℓ_p smoothness of s_λ, L_p, to reduce the dependence on k to be $O(k^{1/p})$, which effectively removes the polynomial dependence on k when $p \to \infty$. This is perhaps not surprising, since each objective is observed independently and fully, so the information gain scales with the number of objectives.

Lemma 9. *Consider running* ExploreUCB *(Algorithm 1) for $T > \max(k, d)$ iterations and for T even, let a_T be the action that maximizes the scalarized UCB in iteration $T/2$. Then, with probability at least $1 - \delta$, the instantaneous scalarized regret can be bounded by*

$$r(s_\lambda, a_T) \leq 10 k^{\frac{1}{p}} L_p d \sqrt{\frac{\log(k/\delta) + \log(T)}{T}}$$

where L_p is the ℓ_p-Lipschitz constant for $s_\lambda(\cdot)$.

Finally, we connect the expected Bayes regret with the empirical average of scalarized regret via uniform convergence properties of all functions of the form $f(\lambda) = \max_{a \in \mathbf{A}} s_\lambda(\Theta^\star a)$. By using s_λ^{HV} and setting $p = \infty$, we derive our final fast hypervolume regret rates for stochastic linear bandits, which is the combination of the scalarized regret rates and the hypervolume regret rates. Note that our analysis improve the scalarized regret rates by removing the polynomial dependence on k.

Theorem 10. *Assume that for any $a \in \mathcal{A}$, $|s_\lambda(\Theta^\star a)| \leq B$ for some B and s_λ is L_λ-Lipschitz with respect to the ℓ_2 norm in λ. With constant probability, the Bayes regret of running Algorithm 1 at round T can be bounded by*

$$BR(s_\lambda, \mathbf{A}_T) \leq O\left(k^{\frac{1}{p}} L_p d \sqrt{\frac{\log(kT)}{T}} + \frac{B L_\lambda}{T^{\frac{1}{k+1}}} \right)$$

C. Missing Proofs

Proof of Lemma 4. Let $\lambda = y^\star / \|y^\star\|$. Note that $\lambda > 0$ since $y^\star$ is in the positive orthant and for the sake of contradiction, assume there exists z such that $s_\lambda(z) > s_\lambda(y^\star)$. However, note that for any i, $\frac{z_i}{\lambda_i} \geq \min_i \frac{z_i}{\lambda_i} > \min_i \frac{y_i^\star}{\lambda_i} = \frac{y_i^\star}{\lambda_i}$, where the last line follows since $y_i^\star / \lambda_i = \|y^\star\|$ for all i by construction. Therefore, we conclude that $y^\star < z$, contradicting that $y^\star$ is Pareto optimal.

Finally, note that if $\alpha\lambda$ is on the Pareto frontier, then we see that $\min_i \alpha\lambda_i / \lambda_i = \alpha$ and furthermore, this min value is achieved for all i. Therefore, since $\alpha\lambda$ is on the Pareto frontier, any other point $y \in \mathcal{Y}$ has some coordinate j such that $y_j < \alpha\lambda_j$, which implies that $\min_i y_i / \lambda_i < \alpha$. □

Proof of 6. Let us first decompose

$$\mathcal{HV}_z(\mathcal{Y}^\star) - \mathcal{HV}_z(\mathcal{Y}_T) \leq |\mathcal{HV}_z(\mathcal{Y}^\star) - \sum_{i=1}^T \max_{y\in\mathcal{Y}} s_{\lambda_i}(y)| + |\sum_{i=1}^T \max_{y\in\mathcal{Y}} s_{\lambda_i}(y) - \mathcal{HV}_z(\mathcal{Y}_T)|$$

$$\leq |\mathcal{HV}_z(\mathcal{Y}) - \sum_i \max_{y\in\mathcal{Y}} s_{\lambda_i}(y)| + |\sum_i \max_{y\in\mathcal{Y}_T} s_{\lambda_i}(y) - \mathcal{HV}_z(\mathcal{Y}_T)|$$

where the second inequality exploits the fact that $y_i \in \arg\max_{y\in\mathcal{Y}} s_{\lambda_i}(y)$. We proceed to bound both parts separately and we show that it suffices to prove uniform concentration of the empirical sum to the expectation, which is the hypervolume by Lemma 5.

Let $f_{\mathbf{Y}}(\lambda_i) = \max_{y\in\mathbf{Y}} s_{\lambda_i}(y)$. We let $\mathcal{F} = \{f_{\mathbf{Y}} : \mathbf{Y} \subseteq \mathcal{A}\}$ be our class of functions over all possible output sets $\mathbf{Y}$. We will first demonstrate uniform convergence by bounding the complexity of $\mathcal{F}$. Specifically, by generalization bounds from Rademacher complexities (Bartlett and Mendelson, 2002), over choices of $\lambda_i \sim \mathcal{D}$, we know that with probability $1-\delta$, for all $\mathbf{Y}$, we have the bound

$$\left|\mathop{\mathbf{E}}_{\lambda\sim\mathcal{D}}[f_{\mathbf{Y}}] - \frac{1}{m}\sum_{i=1}^m f_{\mathbf{Y}}(\lambda_i)\right| \leq R_m(\mathcal{F}) + \sqrt{\frac{8\ln(2/\delta)}{m}}$$

where $R_m(\mathcal{F}) = \mathbf{E}_{\lambda_i\sim\mathcal{D},\sigma_i}\left[\sup_{f\in\mathcal{F}} \frac{2}{m}\sum_i \sigma_i f(\lambda_i)\right]$, where σ_i are i.i.d. ± 1 Rademacher variables.

To bound $R_m(\mathcal{F})$, we appeal to Dudley's integral formulation that allows us to use the metric entropy of $\mathcal{F}$ to bound

$$R_m(\mathcal{F}) \leq \inf_{\alpha>0}\left(4\alpha + 12\int_\alpha^\infty \sqrt{\frac{\log(\mathcal{N}(\epsilon,\mathcal{F},\|\cdot\|_2))}{m}}\,d\epsilon\right)$$

where $\mathcal{N}$ denotes the standard covering number for $\mathcal{F}$ under the ℓ_2 function norm metric over $\lambda \in \mathcal{D}$.

Since $\mathcal{D}$ is the uniform distribution over $\mathcal{S}_+$, this induces a natural ℓ_∞ function norm metric on $\mathcal{F}$ that is $\|f\|_\infty = \sup_{\lambda\in\mathcal{S}_+}|f(\lambda)|$. By Lemma 14, $s_\lambda(y)$ is L_λ Lipschitz with respect to the Euclidean norm in λ. Note that since the maximal operator preserves Lipschitzness, $f_{\mathbf{Y}}(\lambda)$ is also L_λ-Lipschitz with respect to $\lambda \in \mathbb{R}^k$ for any $\mathbf{Y}$. Since $\mathcal{F}$ contains L_λ-Lipschitz functions in $\mathbb{R}^k$, we can bound the metric entropy via a covering of λ via a Lipschitz covering argument (see Lemma 4.2 of (Gottlieb et al., 2016)), so we have

$$\log(\mathcal{N}(\epsilon,\mathcal{F},\|\cdot\|_2)) \leq \log(\mathcal{N}(\epsilon,\mathcal{F},\|\cdot\|_\infty)) \leq (4L_\lambda/\epsilon)^k\log(8/k)$$

Finally, we follow the same Dudley integral computation of Theorem 4.3 of (Gottlieb et al., 2016) to get that

$$R_m(\mathcal{F}) \leq \inf_{\alpha>0}\left(4\alpha + 12\int_\alpha^2 \sqrt{\frac{(4L_\lambda/\epsilon)^k\log(8/k)}{T}}\,d\epsilon\right)$$

$$= O(L_\lambda/m^{1/(k+1)})$$

Therefore, we conclude that with probability at least $1-\delta$ over the independent choices of $\lambda_i \sim \mathcal{D}$, for all $\mathbf{Y}$ and setting $m = T$

$$\left|\mathop{\mathbf{E}}_{\lambda\sim\mathcal{D}}\left[\max_{a\in\mathbf{Y}} s_\lambda(\Theta^\star a)\right] - \frac{1}{T}\sum_{i=1}^T \max_{a\in\mathbf{Y}} s_{\lambda_i}(\Theta^\star a)\right|$$

$$\leq O\left(\frac{BL_\lambda}{T^{1/(k+1)}}\right) + \sqrt{\frac{8\ln(2/\delta)}{T}}$$

Finally, we conclude by using Lemma 5 to replace the expectation by the hypervolume and by setting $\mathbf{Y} = \mathcal{Y}, \mathcal{Y}_T$ respectively. □

C.1. Proofs of Lipschitz Properties

We utilize the fact that if s_λ is differentiable everywhere except for a finite set, bounding Lipschitz constants is equivalent to bounding the dual norm $\|\nabla s_\lambda\|_q$, where $1/p + 1/q = 1$, which follows from mean value theorem, which we state as Proposition 11.

Proposition 11. *Let $f : \mathcal{X} \to \mathbb{R}$ be a continuous function that is differentiable everywhere except on a finite set, then if $\|\nabla f(x)\|_q \leq L_p$ for all $x \in \mathcal{X}$, $f(x)$ is L_p-Lipshitz with respect to the ℓ_p norm.*

Lemma 12. *Let $s_\lambda(y) = \lambda^\top y$ be the linear scalarization with $\|\lambda\| \leq 1$ and $\|y\|_\infty \leq 1$. Then, we may bound $L_p \leq \max(1, k^{1/2-1/p})$ and $L_\lambda \leq \sqrt{k}$ and $|s_\lambda| \leq \sqrt{k}$.*

Proof of Lemma 12. Since $\nabla_\lambda s_\lambda(x) = y$, we use Proposition 11 to bound $L_\lambda \leq \max_y \|y\| \leq \sqrt{k}\|y\|_\infty = \sqrt{k}$. Similarly, since $\nabla_y s_\lambda(y) = \lambda$, we may bound for $p \leq 2$, $L_p \leq \|\lambda\|_q \leq \|\lambda\| \leq 1$ for $1/p + 1/q = 1$ and for $p \geq 2$, we may use Holder's inequality to bound $L_p \leq \|\lambda\|_q \leq k^{1/q-1/2}\|\lambda\| \leq k^{1/2-1/p}$.

To bound the absolute value of s_λ, note $s_\lambda(y) = \lambda^\top y \leq \sqrt{k}$ for all since $\|y\|_2 \leq \sqrt{k}\|y\|_\infty \leq \sqrt{k}$. □

Lemma 13. *Let $s_\lambda(y) = \min_i \lambda_i y_i$ be the Chebyshev scalarization with $\|\lambda\| \leq 1$ and $\|y\|_\infty \leq 1$. Then, we may bound $L_p \leq 1$ and $L_\lambda \leq 1$ and $|s_\lambda| \leq 1/\sqrt{k}$.*

Proof of Lemma 13. For a specific λ, y, let $i^\star$ be the optimal index of the minimization. Then, the gradient $\nabla_\lambda s_\lambda(x)$ is simply zero in every coordinate except at $i^\star$, where it is

$y_{i^\star}$. Therefore, since we can only have a finite number of discontinuities due to monotonicity, we use Proposition 11 to bound $L_\lambda \le y_{i^\star} \le 1$. Similarly, since $\nabla_y s_\lambda(y)$ has only one non-zero coordinate except at $i^\star$, which is $\lambda_{i^\star}$, we may bound for $L_q \le \lambda_{i^\star} \le 1$.

To bound the absolute value of s_λ, note that there must exists $\lambda_i < 1/\sqrt{k}$ as $\|\lambda\| \le 1$. Thus, $\min_i \lambda_i y_i < 1/\sqrt{k}$ for $\|y\|_\infty \le 1$. □

Lemma 14. *Let $s_\lambda(y) = \min_i (y_i/\lambda_i)^k$ be the hypervolume scalarization with $\|\lambda\| = 1$ and $0 < B_l \le y_i \le B_u$. Then, we may bound $L_p \le \frac{B_u^k}{B_l k^{k/2-1}}$ and $L_\lambda \le \frac{B_u^{k+1}}{B_l k^{(k-1)/2}}$ and $|s_\lambda| \le \frac{B_u^k}{k^{k/2}}$.*

Proof of Lemma 14. For a specific λ, y, let $i^\star$ be the optimal index of the minimization. Then, the gradient $\nabla_\lambda s_\lambda(x)$ is simply zero in every coordinate except at $i^\star$, which in absolute value is $k(y_{i^\star}/\lambda_{i^\star})^k(1/\lambda_{i^\star})$.

Let j be the index such that λ_j is maximized and since $\|\lambda\| = 1$, we know that $\lambda_j \ge 1/\sqrt{k}$. Therefore, we see that since $y_{i^\star}/\lambda_{i^\star} \le y_j/\lambda_j \le y_j/\sqrt{k}$, we conclude that $1/\lambda_{i^\star} \le (B_u/B_l)/\sqrt{k}$.

Therefore, using Proposition 11, we have $L_\lambda \le k(y_{i^\star}/\lambda_{i^\star})^k(1/\lambda_{i^\star}) \le k(B_u/\sqrt{k})^k \frac{(B_u/B_l)}{\sqrt{k}} = \frac{B_u^{k+1} k^{(k+1)/2}}{B_l k^{(k-1)/2}}$

And similarly, since $\nabla_y s_\lambda(y)$ has only one non-zero coordinate except at $i^\star$, which is $k(y_{i^\star}/\lambda_{i^\star})^{k-1}(1/\lambda_{i^\star})$, we may bound for

$$L_q \le k(y_{i^\star}/\lambda_{i^\star})^{k-1}(1/\lambda_{i^\star}) \le k(B_u/\sqrt{k})^{k-1}\frac{(B_u/B_l)}{\sqrt{k}} \le \frac{B_u^k}{B_l k^{k/2-1}}$$

To bound the absolute value of s_λ, note that $s_\lambda(y) \le (\frac{y_j}{\lambda_j})^k \le \frac{B_u^k}{k^{k/2}}$. □

C.2. Proofs for Linear Bandits

The following lemma about the UCB ellipsoid is borrowed from the original analysis of linear bandits.

Lemma 15 ((Abbasi-Yadkori et al., 2011)). *Consider the least squares estimator $\widehat{\theta}_t = (\mathbf{M}_t)^{-1}\mathbf{A}_t^\top \mathbf{y}_t$, where the covariance matrix of the action matrix is $\mathbf{M}_t = \mathbf{A}_t^\top \mathbf{A}_t + \lambda \mathbf{I}$, then with probability $1-\delta$,*

$$\|\widehat{\theta}_t - \theta^*\|_{\mathbf{M}_t} \le \sqrt{\lambda}\|\theta^*\| + \sqrt{2\log(\frac{1}{\delta}) + d\log(T/\lambda)}$$

Proof of Lemma 9. Let $\widehat{\Theta}_T$ be the least squares estimate of the true parameters after observing $(\mathbf{A}_T, \mathbf{y}_T)$. Since the noise ξ_t in each objective is independent and 1-sub-Gaussian, by Lemma 15, if we let $\mathbf{M}_T = \mathbf{A}_T^\top \mathbf{A}_T + \lambda \mathbf{I}$, then with regularization $\lambda = 1$

$$\|\widehat{\Theta}_{Ti} - \Theta_i^\star\|_{\mathbf{M}_T} \le 1 + \sqrt{2\log(k/\delta) + d\log(T)} := D_T$$

holds with probability at least $1 - \delta/k$. Note that this describes the confidence ellipsoid, $C_{Ti} = \{\theta \in \mathbb{R}^d : \|\widehat{\Theta}_{Ti} - \theta_i\|_{\mathbf{M}_T} \le D_T\}$ for Θ_{Ti}.

By the definition of the UCB maximization of a_t, we see $a_t, \widetilde{\Theta}_t = \text{argmax}_{a\in\mathcal{A}} \max_{\theta_i \in C_{Ti}} s_\lambda(\Theta_i^\top a)$. Note that since $\Theta^\star \in \mathbf{C}_T$, we can bound the instantaneous scalarized regret as:

$$r(s_\lambda, a_t) = \max_{a\in\mathcal{A}} s_\lambda(\Theta^\star a) - s_\lambda(\Theta^\star a_t) \le s_\lambda(\widetilde{\Theta}_t a_t) - s_\lambda(\Theta^\star a_t)$$

By the Lipschitz smoothness condition, we conclude that $r(s_\lambda, a_t) \le L_p\|(\widetilde{\Theta_t} - \Theta^\star)a_t\|_p$.

To bound the desired ℓ_p norm, first note that by triangle inequality, $\|\widetilde{\Theta_t} - \Theta^\star\|_{\mathbf{M}_T} \le 2D_T$. Since we apply uniform exploration every other step and $\sum_i e_i e_i^\top \succeq \frac{1}{2}\mathbf{I}$ for $e_i \in \mathcal{E}$ with size $|\mathcal{E}| = d$, we conclude that $\mathbf{M}_T \succeq \frac{T}{5d}\mathbf{I}$. Therefore, we conclude that $\|\widehat{\Theta}_{Ti} - \Theta_i^\star\| \le 5\sqrt{d/T}D_T := E_T$ with probability at least $1 - \delta/k$. Since $\|a_t\| \le 1$, we conclude by Cauchy-Schwarz, that $|(\widehat{\Theta}_{Ti} - \Theta_i^\star)a_t| \le E_T$. Together with our Lipschitz condition, we conclude that

$$r(s_\lambda, a_t) \le k^{1/p} L_p E_T \le 10k^{1/p} L_p d\sqrt{(\log(k/\delta) + \log(T))/T}$$

. □

Proof of Theorem 10. For any set of actions $\mathbf{A} \subseteq \mathcal{A}$, we define $f_\mathbf{A}(\lambda) = \max_{a\in\mathbf{A}} s_\lambda(\Theta^\star a)$. We let $\mathcal{F} = \{f_\mathbf{A} : \mathbf{A} \subseteq \mathcal{A}\}$ be our class of functions over all possible action sets and for any Bayes regret bounds, we will first demonstrate uniform convergence by bounding the complexity of $\mathcal{F}$. Specifically, by generalization bounds from Rademacher complexities (Bartlett and Mendelson, 2002), over choices of $\lambda_i \sim \mathcal{D}$, we know that with probability $1-\delta$, for all $\mathbf{A}$, we have the bound

$$\left|\mathop{\mathbf{E}}_{\lambda\sim\mathcal{D}}[f_\mathbf{A}] - \frac{1}{m}\sum_{i=1}^m f_\mathbf{A}(\lambda_i)\right| \le R_m(\mathcal{F}) + \sqrt{\frac{8\ln(2/\delta)}{m}}$$

where $R_m(\mathcal{F}) = \mathbf{E}_{\lambda_i\sim\mathcal{D},\sigma_i}\left[\sup_{f\in\mathcal{F}} \frac{2}{m}\sum_i \sigma_i f(\lambda_i)\right]$, where σ_i are i.i.d. ± 1 Rademacher variables.

To bound $R_m(\mathcal{F})$, we appeal to Dudley's integral formulation that allows us to use the metric entropy of $\mathcal{F}$ to bound

$$R_m(\mathcal{F}) \leq \inf_{\alpha>0}\left(4\alpha + 12\int_{\alpha}^{\infty}\sqrt{\frac{\log(\mathcal{N}(\epsilon,\mathcal{F},\|\cdot\|_2))}{m}}\,d\epsilon\right)$$

where $\mathcal{N}$ denotes the standard covering number for $\mathcal{F}$ under the ℓ_2 function norm metric over $\lambda \in \mathcal{D}$.

Since $\mathcal{D}$ is the uniform distribution over $\mathcal{S}_+$, this induces a natural ℓ_∞ function norm metric on $\mathcal{F}$ that is $\|f\|_\infty = \sup_{\lambda\in\mathcal{S}_+}|f(\lambda)|$. Since $s_\lambda(\Theta^\star a)$ is L_λ Lipschitz with respect to the Euclidean norm in λ. Note that since the maximal operator preserves Lipschitzness, $f_{\mathbf{A}}(\lambda)$ is also L_λ-Lipschitz with respect to $\lambda \in \mathbb{R}^k$. Since $\mathcal{F}$ contains L_λ-Lipschitz functions in $\mathbb{R}^k$, we can bound the metric entropy via a covering of λ via a Lipschitz covering argument (see Lemma 4.2 of (Gottlieb et al., 2016)), so we have

$$\log(\mathcal{N}(\epsilon,\mathcal{F},\|\cdot\|_2)) \leq \log(\mathcal{N}(\epsilon,\mathcal{F},\|\cdot\|_\infty)) \leq (4L_\lambda/\epsilon)^k\log(8/k)$$

Finally, we follow the same Dudley integral computation of Theorem 4.3 of (Gottlieb et al., 2016) to get that

$$R_m(\mathcal{F}) \leq \inf_{\alpha>0}\left(4\alpha + 12\int_{\alpha}^{2}\sqrt{\frac{(4L_\lambda/\epsilon)^k\log(8/k)}{m}}\,d\epsilon\right)$$

$$= O(L_\lambda/m^{1/(k+1)})$$

Therefore, we conclude that with probability at least $1-\delta$ over the independent choices of $\lambda_i \sim \mathcal{D}$, for all $\mathbf{A}$,

$$\left|\mathop{\mathbf{E}}_{\lambda\sim\mathcal{D}}\left[\max_{a\in\mathbf{A}} s_\lambda(\Theta^\star a)\right] - \frac{1}{m}\sum_{i=1}^{m}\max_{a\in\mathbf{A}} s_{\lambda_i}(\Theta^\star a)\right|$$

$$\leq O\left(\frac{BL_\lambda}{m^{1/(k+1)}}\right) + \sqrt{\frac{8\ln(2/\delta)}{m}}$$

Finally, note that for T even, with constant probability,

$$\begin{aligned}
BR(s_\lambda,\mathbf{A}_t) &= \mathop{\mathbf{E}}_{\lambda\sim\mathcal{D}}[r(s_\lambda,\mathbf{A}_t)] \\
&= \mathop{\mathbf{E}}_{\lambda\sim\mathcal{D}}[\max_{a\in\mathcal{A}} s_\lambda(\Theta^\star a) - \max_{a\in\mathbf{A}_T} s_\lambda(\Theta^\star a)] \\
&\leq \frac{1}{T/2}\sum_{i=1}^{T/2}\left[\max_{a\in\mathcal{A}} s_{\lambda_i}(\Theta^\star a) - \max_{a\in\mathbf{A}_T} s_{\lambda_i}(\Theta^\star a)\right] \\
&\quad + O\left(\frac{BL_\lambda}{T^{1/(k+1)}}\right) \\
&\leq \frac{1}{T/2}\sum_{i=1}^{T/2}\left[\max_{a\in\mathcal{A}} s_{\lambda_i}(\Theta^\star a) - s_{\lambda_i}(\Theta^\star a_{2i})\right] \\
&\quad + O\left(\frac{BL_\lambda}{T^{1/(k+1)}}\right) \\
&\leq \frac{1}{T/2}\sum_{i=1}^{T/2} r(s_{\lambda_i},a_{2i}) + O\left(\frac{BL_\lambda}{T^{1/(k+1)}}\right) \\
&\leq O\left(k^{1/p}L_p d\sqrt{\frac{\log(kT)}{T}} + \frac{BL_\lambda}{T^{1/(k+1)}}\right)
\end{aligned}$$

where the last line used Lemma 9 with $\delta = 1/T^2$ and applied a union bound over all $O(T)$ iterations. □

Proof of Theorem 8. Note that by Lemma 5, we connect the Bayes regret to the hypervolume regret for $\mathcal{D}$:

$$\mathcal{HV}_z(\Theta^\star\mathcal{A}) - \mathcal{HV}_z(\Theta^\star\mathbf{A}_t) = c_k \mathop{\mathbf{E}}_{\lambda\sim\mathcal{D}}[\max_{a\in\mathcal{A}} s_\lambda(\Theta^\star a) - \max_{a\in\mathbf{A}_T} s_\lambda(\Theta^\star a)]$$

where $s_\lambda(y) = \min_i(y_i - z_i/\lambda_i)^k$.

Note that since $\|\Theta^\star a\|_\infty \leq 1$ for all $a \in \mathcal{A}$ and B is maximal, we have $B \leq \Theta^\star a - z \leq B+2$. Therefore, we conclude by Lemma 14 that s_λ is Lipschitz with

$$L_p \leq \frac{(B+2)^k}{Bk^{k/2-1}}, L_\lambda \leq \frac{(B+2)^{k+1}}{Bk^{(k-1)/2}}, |s_\lambda| \leq \frac{(B+2)^k}{k^{k/2}}$$

Finally, we combine this with Theorem 10 with $p=\infty$ as the optimal choice of p (since L_p does not depend on p) to get our desired bound on hypervolume regret. □

Proof of Theorem 7. We let $\mathcal{A} = \{a : \|a\| \leq 1\}$ be the unit sphere and $\Theta_i^\star = e_i$ be the unit vector directions. Note that in this case the Pareto frontier is exactly $\mathcal{S}_+^{k-1}$.

Consider a uniform discretization of the Pareto front by taking an ϵ grid with respect to each angular component with respect to the polar coordinates. Let $p_1,\ldots,p_m$ be the center (in terms of each of the $k-1$ angular dimensions)

in the $m = \Theta((1/\epsilon)^{k-1})$ grid elements. We consider the output $\mathbf{y}_T = \Theta^\star \mathbf{A}_T$ and assume that for some grid element i, it contains none of the T outputs $\mathbf{y}_T$. Since our radial component $r = 1$, by construction of our grid in the angular component, we deduce that $\min_t \|y_t - p_i\|_\infty > \epsilon/10$ by translating polar to axis-aligned coordinates.

Let $\epsilon' = \epsilon/10$. Assume also that $\frac{1}{k} < p_i < 1 - \frac{1}{k}$. Next, we claim that the hypercube from $p_i - \epsilon'/k^2$ to p_i is not dominated by any points in $\mathbf{Y}_T$. Assume otherwise that there exists y_t such that $y_t \geq p_i - \epsilon'/k^2$. Now, this combined with the fact that since $\min_t \|y_t - p_i\|_\infty > \epsilon'$ implies that there must exist a coordinate such that $y_{tj} \geq p_j + \epsilon'$.

However, this implies that

$$\begin{aligned}\sum_{i=1}^{k} y_{ti}^2 &\geq \sum_{i \neq j} (p_i - \epsilon'/k^2)^2 + (p_j + \epsilon')^2 \\ &\geq \sum_i p_i^2 - 2(\epsilon'/k^2) \sum_{i \neq j} p_i + 2\epsilon' p_j > 1\end{aligned}$$

where the last inequality follows since $\sum_i p_i < 1/\sqrt{k}$ and $p_j > \frac{1}{k}$ by assumption. However, this contradicts that $\|y_t\| \leq 1$, so it follows that $p_i - \epsilon'$ is not dominated.

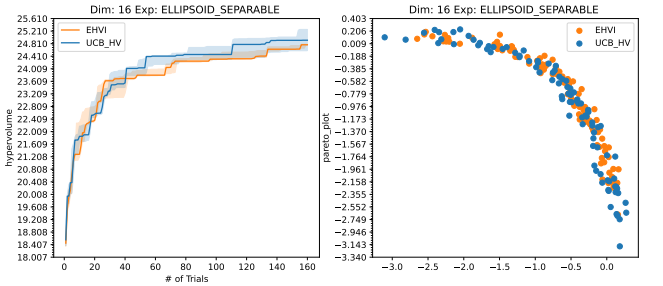
Therefore, for any grid element such that $p_i > 1/k$, if there is no y_t in the grid, we must have a hypervolume regret of at least $\Omega(\epsilon'^k) = \Omega(\epsilon^k)$ be simply consider the undominated hypervolume from p_i to $p_i - \epsilon'$, which lies entirely within the grid element. In fact, since there are $\Theta((1/\epsilon)^{k-1})$ such grid elements satisfying $p_i > 1/k$, we see that if $T < O((1/\epsilon)^{k-1})$, by pigeonhole, there must be a hypervolume regret of at least $\Omega((1/\epsilon)^{k-1}\epsilon^k) = \Omega(\epsilon)$

Therefore, for any $1/2 > \epsilon > 0$, $\mathcal{HV}_z(\Theta^\star \mathcal{A}) - \mathcal{HV}_z(\Theta^\star \mathbf{A}_T) < \epsilon$ implies that $T = \Omega((1/\epsilon)^{k-1})$. Rearranging shows that

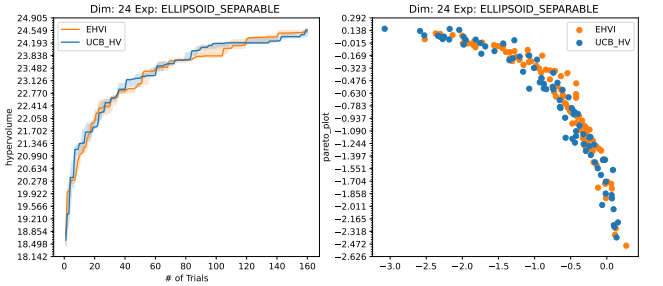
$$\mathcal{HV}_z(\Theta^\star \mathcal{A}) - \mathcal{HV}_z(\Theta^\star \mathbf{A}_T) = \Omega(T^{-1/(k-1)})$$

□

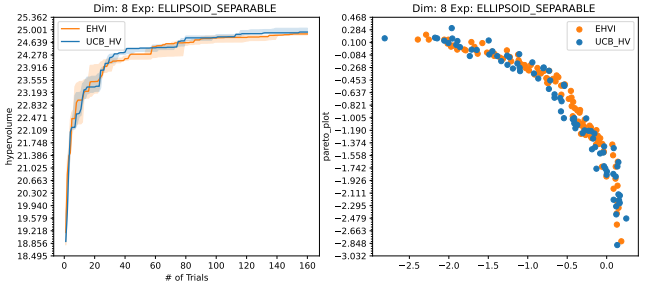
D. Figures


Dim: 16 Exp: ELLIPSOID_SEPARABLE
EHVI
UCB_HV
hypervolume
# of Trials

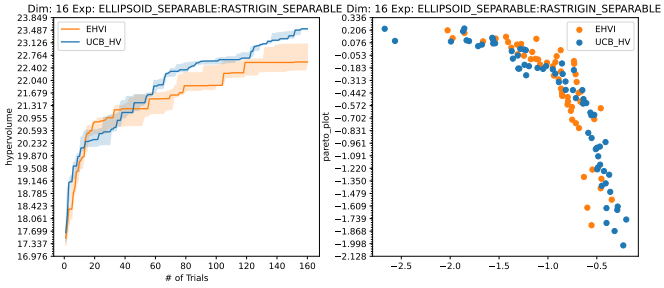
Dim: 16 Exp: ELLIPSOID_SEPARABLE
EHVI
UCB_HV
pareto_plot


Dim: 24 Exp: ELLIPSOID_SEPARABLE
EHVI
UCB_HV
hypervolume
# of Trials

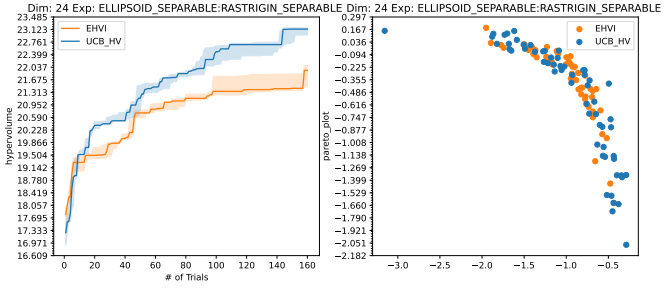
Dim: 24 Exp: ELLIPSOID_SEPARABLE
EHVI
UCB_HV
pareto_plot


Dim: 8 Exp: ELLIPSOID_SEPARABLE
EHVI
UCB_HV
hypervolume
# of Trials

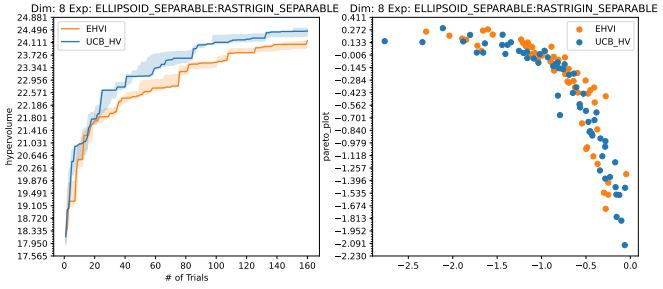
Dim: 8 Exp: ELLIPSOID_SEPARABLE
EHVI
UCB_HV
pareto_plot


Dim: 16 Exp: ELLIPSOID_SEPARABLE:RASTRIGIN_SEPARABLE
EHVI
UCB_HV
hypervolume
# of Trials

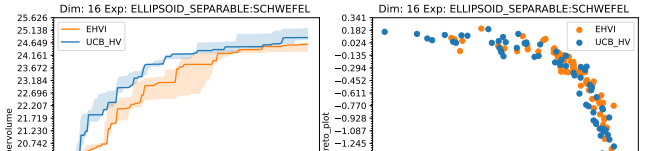
Dim: 16 Exp: ELLIPSOID_SEPARABLE:RASTRIGIN_SEPARABLE
EHVI
UCB_HV
pareto_plot


Dim: 24 Exp: ELLIPSOID_SEPARABLE:RASTRIGIN_SEPARABLE
EHVI
UCB_HV
hypervolume
# of Trials

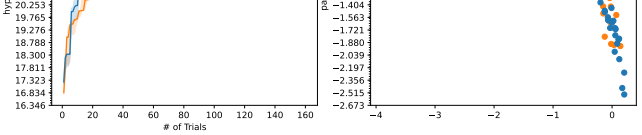
Dim: 24 Exp: ELLIPSOID_SEPARABLE:RASTRIGIN_SEPARABLE
EHVI
UCB_HV
pareto_plot

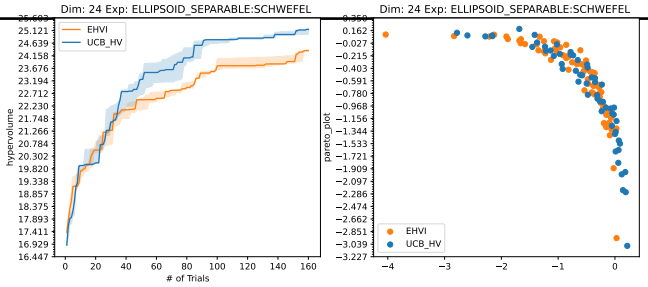

Dim: 8 Exp: ELLIPSOID_SEPARABLE:RASTRIGIN_SEPARABLE
EHVI
UCB_HV
hypervolume
# of Trials

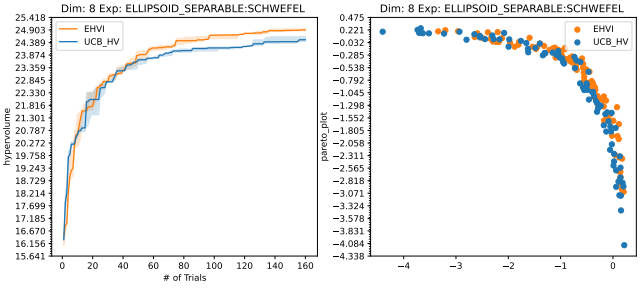
Dim: 8 Exp: ELLIPSOID_SEPARABLE:RASTRIGIN_SEPARABLE
EHVI
UCB_HV
pareto_plot

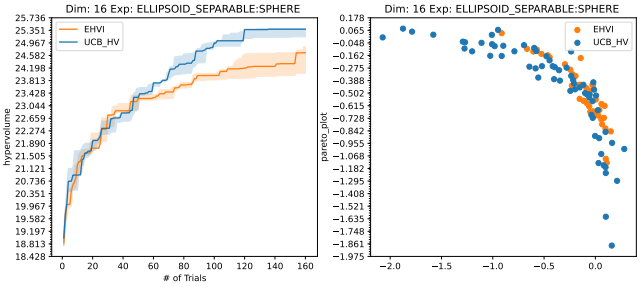

Dim: 16 Exp: ELLIPSOID_SEPARABLE:SCHWEFEL
EHVI
UCB_HV

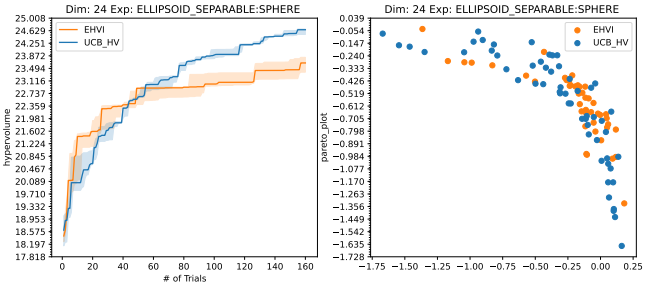
Dim: 16 Exp: ELLIPSOID_SEPARABLE:SCHWEFEL
EHVI
UCB_HV

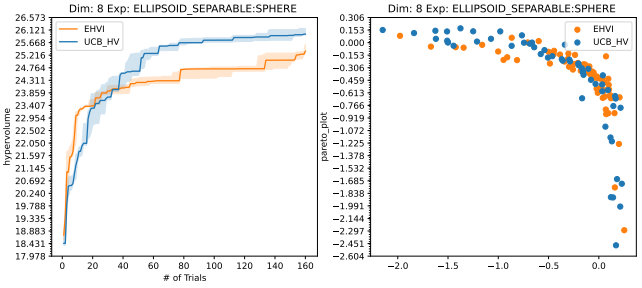

hypervolume
# of Trials
pareto_plot

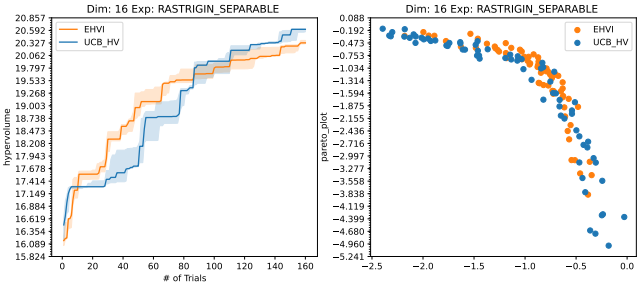

Dim: 24 Exp: ELLIPSOID_SEPARABLE:SCHWEFEL
EHVI
UCB_HV
hypervolume
# of Trials
pareto_plot

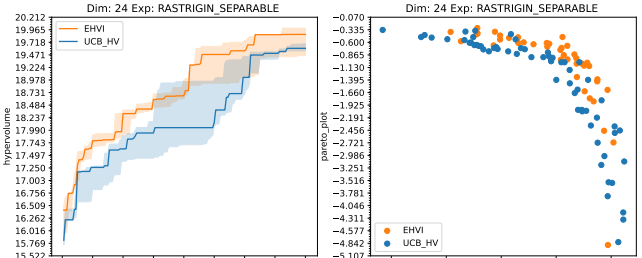

Dim: 8 Exp: ELLIPSOID_SEPARABLE:SCHWEFEL
EHVI
UCB_HV
hypervolume
# of Trials
pareto_plot


Dim: 16 Exp: ELLIPSOID_SEPARABLE:SPHERE
EHVI
UCB_HV
hypervolume
# of Trials
pareto_plot


Dim: 24 Exp: ELLIPSOID_SEPARABLE:SPHERE
EHVI
UCB_HV
hypervolume
# of Trials
pareto_plot


Dim: 8 Exp: ELLIPSOID_SEPARABLE:SPHERE
EHVI
UCB_HV
hypervolume
# of Trials
pareto_plot


Dim: 16 Exp: RASTRIGIN_SEPARABLE
EHVI
UCB_HV
hypervolume
# of Trials
pareto_plot


Dim: 24 Exp: RASTRIGIN_SEPARABLE
EHVI
UCB_HV
hypervolume
# of Trials
pareto_plot