A. Missing Proofs

Proof of Lemma 5(Hypervolume as Scalarization).
Without loss of generality, let $z = 0$ be the origin and consider computing the volume of a rectangle with corners at the origin and at $y = (y_1, ..., y_k) \geq 0$, but doing so in polar coordinates centered at z. Given a direction $v \in \mathcal{S}_+^{k-1}$ with $v_i \geq 0$ for all i, and $\|v\| = 1$, lets suppose a ray in the direction of v exits the rectangle at a point $p = cv$ so that $\|p\| = c$. We claim that $\|p\| = \min_i(y_i/v_i)$. Note that this claim holds in any norm; however for the eventual dominated hypervolume calculation to hold, we use the ℓ_2 norm.

Note that by definition of p we have $p_j \leq y_j$ for all j and there must exist i such that $p_i = y_i$. Also $p_j/v_j = c = p_i/v_i$ for any i, j since $p = cv$. It follows that $c \leq y_j/v_j$ for all j and $c = y_i/v_i$, which proves $\|p\| = \min_i(y_i/v_i)$.

Integrating in polar coordinates, we can approximate an volume via radial slivers of the circle, which for a radius r sweeping through angles $d\theta$ have an volume proportional to $r^k d\theta$. Hence, the volume of the of the rectangle is

$$\text{vol}(R) = c_k \int_{v \in \mathcal{S}_+^{k-1}} \min_i \left(\frac{y_i}{v_i}\right)^k d\theta(v)$$

under a uniform measure θ, where the c_k is a constant that depends only on the dimension.

So far we assumed $y \geq 0$. Now, if any $y_i < 0$, then our total dominated hypervolume is zero and $\min_i \frac{y_i}{v_i} < 0$. So, by changing our scalarization slightly, we can account for any y and the volume of the rectangle with respect to the origin is given by:

$$\text{vol}(R) = c_k \int_{v \in \mathcal{S}_+^{k-1}} \min_i \left(\max(0, y_i/v_i)\right)^k d\theta(v)$$

By definition of dominated hypervolume, note that $\mathcal{HV}_z(Y) = \text{vol}(S)$ where

$$S = \{x \mid x \geq z, x \text{ is dominated by some } y \in Y\}.$$

Since S is simply the union of rectangles at $y_1, ..., y_m$ and note that wherever p exists S, the length of p is the maximal over all rectangles and so

$$\|p\| = \max_{y \in Y} \min_i \max(0, y_i/v_i).$$

Repeating the argument gives:

$$\text{vol}(S) = c_k \int_{v \in \mathcal{S}_+^{k-1}} \max_{y \in Y} \left[\min_i \left(\max(0, y_i/v_i)\right)^k\right] d\theta(v)$$

To calculate c_k, we simply evaluate the hypervolume of the k-dimensional ball in the positive orthant. If the ball has radius and is centered at the origin. In this case we get $\int_{v \in \mathcal{S}_+^{k-1}} r^k d\mu(v) = r^k$ and $c_k \int_{v \in \mathcal{S}_+^{k-1}} r^k d\mu(v) = V_k(r)/2^k$ where $V_k(r)$ is defined as the volume of the k-dimensional ball of radius r, which is $\pi^{k/2} r^k \,/\, \Gamma(k/2+1)$. The formula for c_k then follows from some basic algebra. □

Proof of Lemma 6 (Hypervolume Concentration). Recall $s_\lambda(y) = \min_i(\max(0, y_i/\lambda_i))^k$ and $\|\lambda\| = 1$. Note that there must exists i such that $\lambda_i \geq k^{-1/2}$ and therefore $(\max(0, y_{i^*}/\lambda_{i^*}))^k \leq (Bk^{1/2})^k$ where i^* is the index that minimizes $\max(0, y_i/\lambda_i)^k$. Note that the gradient of $s_\lambda(y)$, if non-zero, is $ky_{i^*}^{k-1}/\lambda_{i^*}^k$ and therefore, the Lipschitz constant is bounded by $k(Bk^{1/2})^k = B^k k^{1+k/2}$.

Since $0 \leq s_\lambda(y - z) \leq B^k k^{k/2}$ for any λ and y, we conclude by standard Chernoff bounds that if weight vectors λ_j are independent samples,

$$\Pr\left(|\mathbb{E}_{\lambda \sim \mathcal{S}_+^{k-1}}[\max_{y \in Y} s_\lambda(y - z)] - \frac{1}{s}\sum_j \max_{y \in Y} s_{\lambda_j}(y - z)| \geq \epsilon\right)$$
$$\leq 2\exp(-2s\epsilon^2/(B^{2k}k^k))$$

Therefore, choosing $s = O(B^{2k}k^k \log(1/\delta)/\epsilon^2)$ samples from $\mathcal{S}_+^{k-1}$ bounds the failure probability by δ and using Lemma 5, our result follows. □

Proof of Theorem 9 (General Regret Bounds). WLOG, let $z = 0$. Let $X_T = \bigcup_i \left\{x_{\lambda_i}^{(t)}\right\}_{t=1}^T$. Then, for $\lambda_1, ..., \lambda_l$, by the guarantees of $\mathcal{A}$, we deduce that

$$\frac{1}{l}\sum_{i=1}^{l}\left[\max_{x \in \mathcal{X}} s_{\lambda_i}(F(x)) - \max_{x \in X_T} s_{\lambda_i}(F(x))\right] \leq \epsilon_T$$

By Lemma 6, we see that for the Pareto frontier Y^*, we have concentration to the desired hypervolume:

$$\left|\frac{1}{c_k}\mathcal{HV}_z(Y^*) - \frac{1}{l}\sum_i \max_{x \in \mathcal{X}} s_{\lambda_i}(x)\right| \leq \epsilon$$

when $l = O(B^{2k}k^k \log(1/\delta)/\epsilon^2)$ with probability $1 - \delta$. We would like to apply the same lemma to also show that our empirical estimate is close to $\mathcal{HV}_z(X_T)$. However, since X_T depends on λ_i, this requires a union bound and we proceed with a ϵ-net argument.

We assume that $F(\mathcal{X}) \subseteq [0, B]^k$ and let us divide the hypercube into a grid with spacing Δ. Then, there are

$O((B/\Delta)^k)$ lattice points on the grid. Consider the Pareto frontier of any set of points, call it S. Out of the $(B/\Delta)^k$ small hypercubes of volume Δ^k in grid, note that by the monotonicity property of the frontier, S intersects at most $(2B/\Delta)^{k-1}$ small hypercubes.

Therefore, we can find a set S_u consisting of at most $(2B/\Delta)^{k-1}$ lattice points on the grid such that $|\mathcal{HV}_z(S) - \mathcal{HV}_z(S_u)| \leq (2B)^k\Delta$. This can be done by simply looking at each small hypercube that has intersection with S and choosing the lattice point that increases the dominated hypervolume. Since each small hypercube has volume Δ^k and there are at most $(2B/\Delta)^{k-1}$ hypercubes, the total hypervolume increased is at most $2^k\Delta$.

Finally, there are at most $(B/\Delta)^{k(2B/\Delta)^{k-1}}$ choices of S_u, so to apply a union bound over all possible sets S_u, we simply choose $l = O(B^{2k}k^{k+1}(2B/\Delta)^{k-1}\log(B/\Delta)/\epsilon^2)$ so that for any possible S_u, we use Lemma 6 to deduce that with high probability,

$$\left|\frac{1}{c_k}\mathcal{HV}_z(S_u) - \frac{1}{l}\sum_i \max_{x\in S_u} s_{\lambda_i}(x)\right| \leq \epsilon$$

Since $\mathcal{HV}_z(S)$ is close to $\mathcal{HV}_z(S_u)$, we conclude that for any S,

$$\left|\frac{1}{c_k}\mathcal{HV}_z(S) - \frac{1}{l}\sum_i \max_{x\in S} s_{\lambda_i}(x)\right| \leq \epsilon + \frac{(2B)^k\Delta}{c_k}$$

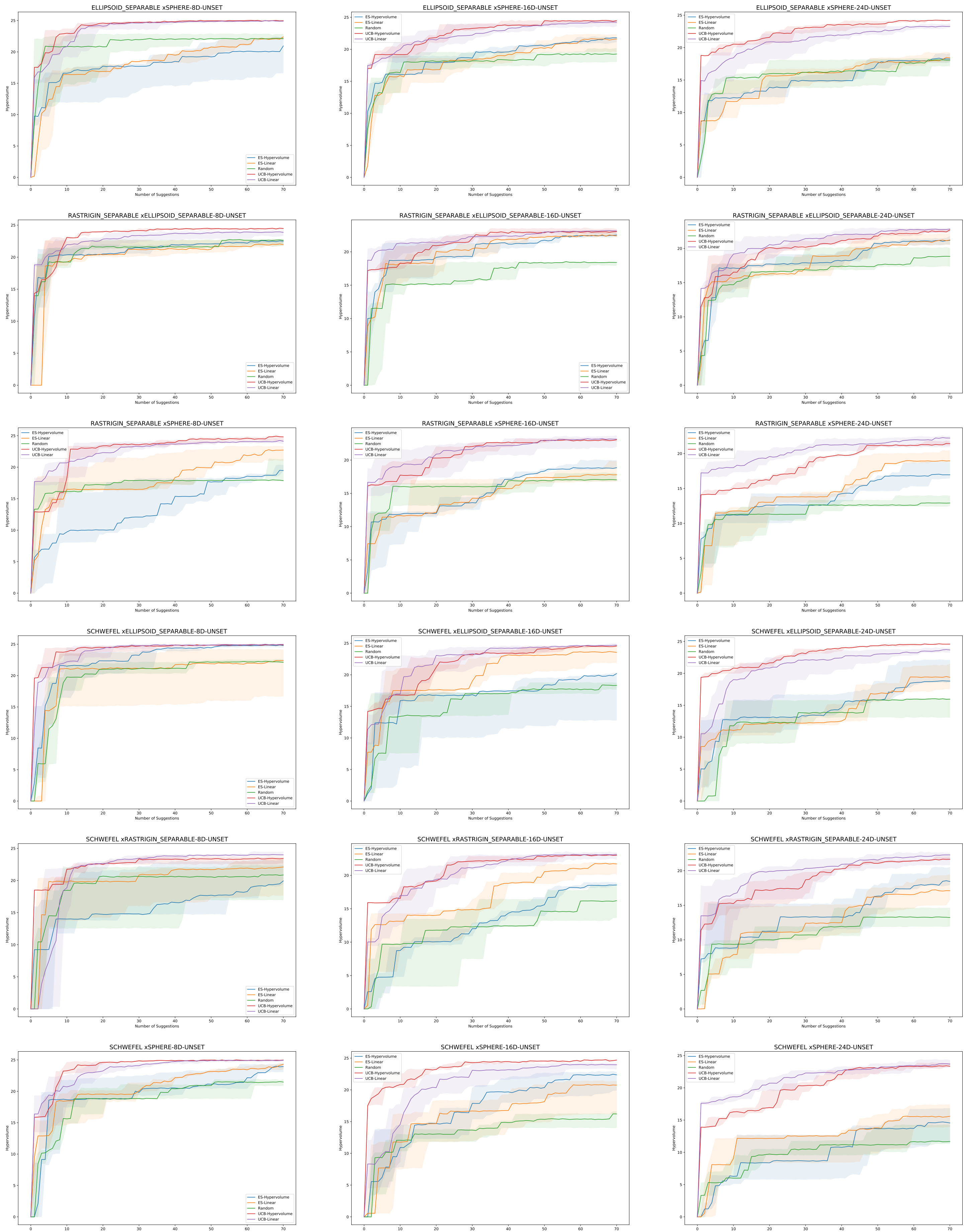
Together, we conclude that

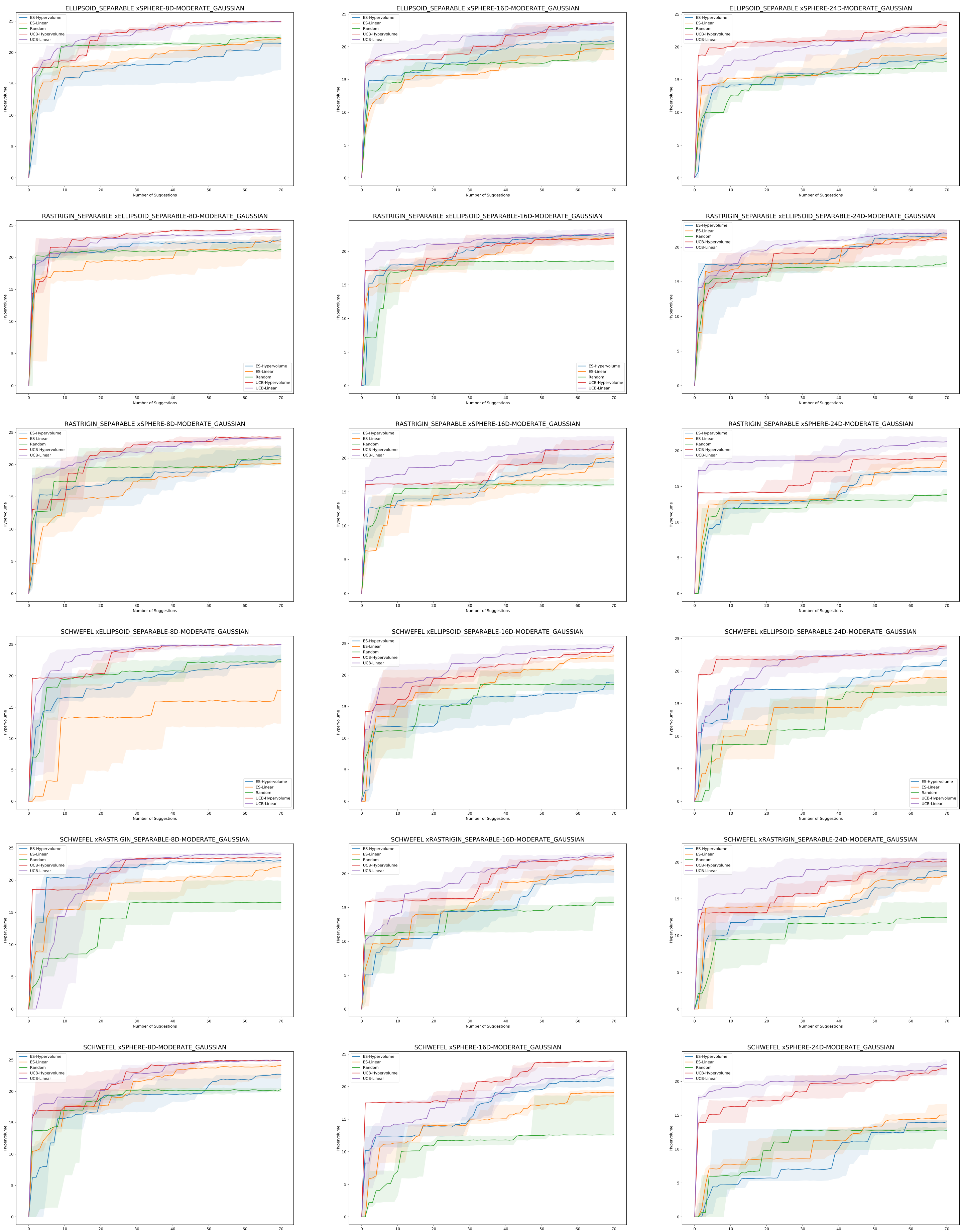
$$|\mathcal{HV}_z(F(X_T)) - \mathcal{HV}_z(Y^*)| \leq c_k\epsilon_T + 2c_k\epsilon + (2B)^k\Delta$$

By choosing $\epsilon = \epsilon_T$ and $\Delta = c_k\epsilon_T(2B)^{-k}$, we conclude.

□

B. Figures


ELLIPSOID_SEPARABLE xSPHERE-8D-UNSET
ELLIPSOID_SEPARABLE xSPHERE-16D-UNSET
ELLIPSOID_SEPARABLE xSPHERE-24D-UNSET
RASTRIGIN_SEPARABLE xELLIPSOID_SEPARABLE-8D-UNSET
RASTRIGIN_SEPARABLE xELLIPSOID_SEPARABLE-16D-UNSET
RASTRIGIN_SEPARABLE xELLIPSOID_SEPARABLE-24D-UNSET
RASTRIGIN_SEPARABLE xSPHERE-8D-UNSET
RASTRIGIN_SEPARABLE xSPHERE-16D-UNSET
RASTRIGIN_SEPARABLE xSPHERE-24D-UNSET
SCHWEFEL xELLIPSOID_SEPARABLE-8D-UNSET
SCHWEFEL xELLIPSOID_SEPARABLE-16D-UNSET
SCHWEFEL xELLIPSOID_SEPARABLE-24D-UNSET
SCHWEFEL xRASTRIGIN_SEPARABLE-8D-UNSET
SCHWEFEL xRASTRIGIN_SEPARABLE-16D-UNSET
SCHWEFEL xRASTRIGIN_SEPARABLE-24D-UNSET
SCHWEFEL xSPHERE-8D-UNSET
SCHWEFEL xSPHERE-16D-UNSET
SCHWEFEL xSPHERE-24D-UNSET
ES-Hypervolume
ES-Linear
Random
UCB-Hypervolume
UCB-Linear
Hypervolume
Number of Suggestions


ELLIPSOID_SEPARABLE xSPHERE-8D-MODERATE_GAUSSIAN
ELLIPSOID_SEPARABLE xSPHERE-16D-MODERATE_GAUSSIAN
ELLIPSOID_SEPARABLE xSPHERE-24D-MODERATE_GAUSSIAN
RASTRIGIN_SEPARABLE xELLIPSOID_SEPARABLE-8D-MODERATE_GAUSSIAN
RASTRIGIN_SEPARABLE xELLIPSOID_SEPARABLE-16D-MODERATE_GAUSSIAN
RASTRIGIN_SEPARABLE xELLIPSOID_SEPARABLE-24D-MODERATE_GAUSSIAN
RASTRIGIN_SEPARABLE xSPHERE-8D-MODERATE_GAUSSIAN
RASTRIGIN_SEPARABLE xSPHERE-16D-MODERATE_GAUSSIAN
RASTRIGIN_SEPARABLE xSPHERE-24D-MODERATE_GAUSSIAN
SCHWEFEL xELLIPSOID_SEPARABLE-8D-MODERATE_GAUSSIAN
SCHWEFEL xELLIPSOID_SEPARABLE-16D-MODERATE_GAUSSIAN
SCHWEFEL xELLIPSOID_SEPARABLE-24D-MODERATE_GAUSSIAN
SCHWEFEL xRASTRIGIN_SEPARABLE-8D-MODERATE_GAUSSIAN
SCHWEFEL xRASTRIGIN_SEPARABLE-16D-MODERATE_GAUSSIAN
SCHWEFEL xRASTRIGIN_SEPARABLE-24D-MODERATE_GAUSSIAN
SCHWEFEL xSPHERE-8D-MODERATE_GAUSSIAN
SCHWEFEL xSPHERE-16D-MODERATE_GAUSSIAN
SCHWEFEL xSPHERE-24D-MODERATE_GAUSSIAN
ES-Hypervolume
ES-Linear
Random
UCB-Hypervolume
UCB-Linear
Hypervolume
Number of Suggestions