Appendix

Proof for Proposition 2

Proposition 3 (Exploration Bounds for TER)**.** *Suppose every action is corresponded with at least one efficiency record within the sliding window of size w and τ is the parameter for softmax decisions, the probability of taking the action with the not-the-highest mean normalized record (exploration) in the fragment of length w satisfies the bounds*

$$p_{explore} \in \left[\frac{e^{1/\tau}}{|A| - 1 + e^{1/\tau}}, \frac{|A| - 1}{(|A| - 1) \cdot e^{1/\tau} + e^{\frac{w-|A|}{\tau(w-|A|+1)}}} \right) \quad (3)$$

Proof. After the normalization of the controller, the highest reward within the sliding window is 1 and the lowest is 0. We first stretch the reward stream fragment corresponding to the sliding window into a $|A| \times w$ rectangle by putting the rewards of each action on the corresponding row, as shown in Fig. 1. Making softmax decisions within the fragment does not care about the orders of rewards, thus an equivalence class of stretched fragments can be obtained by permutations on the fragment before stretching. Suppose that a_1 is the best performing action, we can show the upper bound can be obtained within the class equivalent to

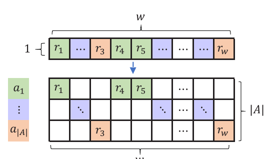
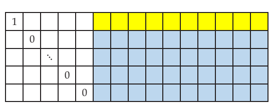
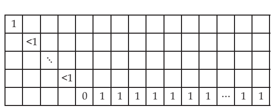


Figure 1: Demonstration for stretching fragments.



where we can put 1 in the yellow cells and 0 in the cyan cells but at most one element in each column. Similarly, we can get the lower bound within the class equivalent to



where "< 1" represents normalized rewards that are infinitely close to 1 but still less than 1 and thus the lower bound cannot be reached.

The probability bounds for exploration can be derived directly by subtracting the exploitation bounds. □

Principled Guidelines for Practical Use

If we have a preferred interval for the probability of exploration or exploitation, we can inversely locate the potential combinations of the hyperparameters. For example, if there are totally 3 heuristics to be articulated, $|A| = 3$, we can first constrain w by $w \geq 4$ and simplify the exploitation bound according to Proposition 1. Given a preferred exploitation probability interval, (p_{min}, p_{max}), we can solve τ using p_{max} since only τ is involved in the upper-bound. Then, we use the solved τ and p_{min} to get w.

Details of Articulated Heuristics

Table 6: Three Heuristics used for Experiments

Name	LS
Costs	$25 \times D$
Details	Local Search strategy used in MTS (Tseng and Chen 2008), MOS (LaTorre, Muelas, and Pena 2012), resembling the trajectory search based algorithms. The same as the cofiguration in MOS: Initial step size is $0.2 mean(\boldsymbol{u} - \boldsymbol{l})$, where $\boldsymbol{l}$ and $\boldsymbol{u}$ are the lower and upper box constraints respectively. Minimal step size is 1×10^{-15}.
Name	CC
Costs	$75 \times D$
Details	Cooperative Coevolution with random grouping (Omidvar et al. 2010) and SaNSDE (Yang, Tang, and Yao 2008) as optimizer, resembling the DECC family. A robust and classical configuration: The mean of NP is set to be 15, group size is 50 and 250 generations is assigned for each group.
Name	GS
Costs	$25 \times D$
Details	Global Search that applies SHADE (Tanabe and Fukunaga 2014) on all dimensions of the problem, resembling direct optimization strategies, *e.g.* CSO (Cheng and Jin 2015). The same configuration as SHADE-ILS (Molina, LaTorre, and Herrera 2018): NP is set to be 50, iterate for $D/2$ generations.

Hyperparameter Selection and Guideline

We want to constrain the probability of exploitation to be at least 0.5 and at most 0.99. Using the bounds we have obtained, we can get three sets of solutions of $\langle \tau, w \rangle$: $\langle 1/5, 5 \rangle$, $\langle 1/6, 6 \rangle$, $\langle 1/7, 7 \rangle$ (considering only integer fractions). Then we test on the three combinations of hyperparameters, whose results are presented in Table 7. From the t-test results, the hyperparameter pairs $\langle 1/5, 5 \rangle$ and $\langle 1/6, 6 \rangle$ performed similarly well. In the experiments, we select the $\langle 1/5, 5 \rangle$ setting, since smaller w leads to quicker adaptation.

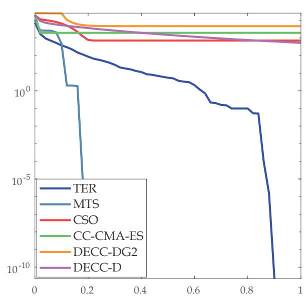
Table 7: Performance with Different Hyperparameters

LSO13	⟨1/5, 5⟩ mean	⟨1/5, 5⟩ std	⟨1/6, 6⟩ mean	⟨1/6, 6⟩ std	⟨1/7, 7⟩ mean	⟨1/7, 7⟩ std
f_1	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
f_2	8.69E+00	2.78E+00	1.20E+01	2.54E+00	1.28E+01	4.70E+00
f_3	9.83E-13	5.27E-14	8.51E-13	5.78E-14	1.05E-12	5.35E-14
f_4	6.98E+08	2.51E+08	6.04E+08	2.02E+08	6.27E+08	2.99E+08
f_5	2.68E+06	4.38E+05	2.72E+08	5.12E+05	2.70E+06	6.04E+05
f_6	4.44E+04	3.42E+04	7.14E+04	1.92E+04	9.48E+04	1.87E+04
f_7	1.58E+05	3.68E+04	1.34E+05	5.02E+04	2.46E+05	7.81E+04
f_8	1.24E+11	9.29E+10	2.61E+11	1.51E+11	7.16E+11	5.66E+11
f_9	2.36E+08	2.61E+07	2.25E+08	4.38E+07	2.52E+08	4.64E+07
f_{10}	7.64E+05	5.90E+05	8.18E+05	5.79E+05	8.00E+05	2.63E+05
f_{11}	3.33E+07	1.08E+07	2.76E+07	8.55E+06	4.45E+07	3.48E+07
f_{12}	6.27E+02	2.11E+02	6.69E+02	2.09E+02	4.32E+02	2.57E+02
f_{13}	1.14E+07	2.20E+06	1.07E+07	2.24E+06	1.05E+07	1.13E+07
f_{14}	4.35E+07	6.85E+06	4.17E+07	7.80E+06	4.29E+07	1.00E+07
f_{15}	3.66E+06	2.32E+05	3.63E+06	2.01E+06	4.25E+06	9.17E+05
t-test	< / ≈ / >		3/9/3		6/8/1	

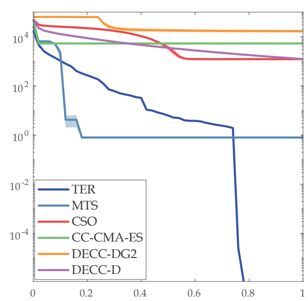
Color indicators are added for each test case. The greener, the better performance.

Optimization Curves for Reproduced Experiments

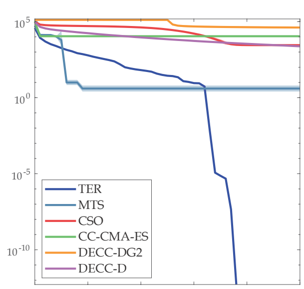
We present a representative set of optimization curves for the scalability tests on f_4 of LSO08 in Fig. 2. The curves only include those that we are able to implement or reproduce.



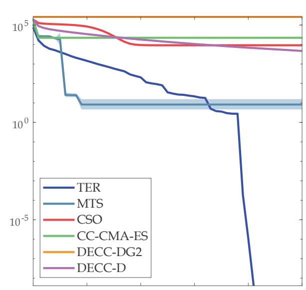
(a) D = 1000



(b) D = 2500



(c) D = 5000

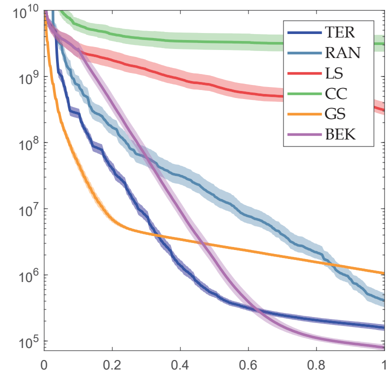


(d) D = 10000

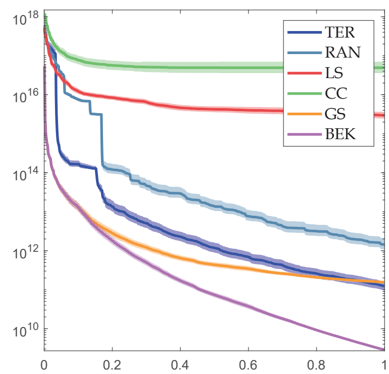
Figure 2: f_4 of LSO08.

Optimization Curves for Baseline Comparison

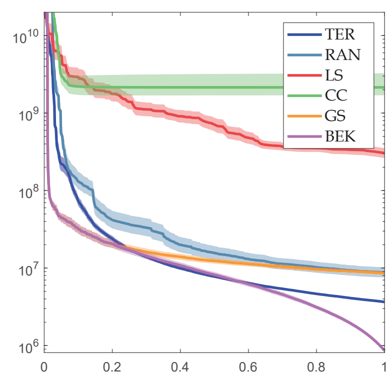
We present some representative sets of optimization curves for the baseline comparison in Fig. 3.



(a) f_7 of LSO13



(b) f_8 of LSO13



(c) $f_1$5 of LSO13

Figure 3: Selected curves for the baseline comparison.