APPENDIX

1 The AvgHExp3 Mean and Variance

According to [1], Exp3 uses the total estimated reward $\hat{S}_{t-1,i}$ to evaluate the reward of arms. Although $\hat{S}_{t-1,i}$ is an unbiased estimator, its variance is very high. To decrease variance, AvgHExp3 uses the average estimated reward $\hat{S}'_{t-1,\omega_i}$ to replace $\hat{S}_{t-1,i}$ according to Theorem 11.

Theorem 11 *The average estimated reward $\hat{S}'_{t-1,\omega_i}$ used by AvgHExp3 is an unbiased estimator, and its variance is $1/(t-1)^2$ of Exp3's, i.e., $D\left[\hat{S}'_{t-1,\omega_i}\right] = \frac{1}{(t-1)^2} D\left[\hat{S}_{t-1,i}\right]$.*

Proof. According to the definition of $\hat{S}'_{t-1,\omega_i}$, i.e., $\hat{S}'_{t-1,\omega_i} = \frac{1}{t-1}\sum_{s=1}^{t-1}\hat{R}_{s,\omega_i}$, where $\hat{R}_{s,\omega_i}$ is the estimated reward of ω_i in iteration t. The expectation of $\hat{S}'_{t-1,\omega_i}$ is

$$E\left[\hat{S}'_{t-1,\omega_i}\right] = E\left[\frac{1}{t-1}\sum_{s=1}^{t-1}\hat{R}_{s,\omega_i}\right] = \frac{1}{t-1}\sum_{s=1}^{t-1} E\left[\hat{R}_{s,\omega_i}\right]. \tag{1}$$

Since $\hat{R}_{s,\omega_i}$ is an unbiased estimator [1], $E\left[\hat{R}_{t,\omega_i}\right] = R_{t,\omega_i}$. So $E\left[\hat{S}'_{t-1,\omega_i}\right] = \frac{1}{t-1}\sum_{s=1}^{t-1} R_{s,\omega_i}$, which means $\hat{S}'_{t-1,\omega_i}$ is an unbiased estimator of the average actual reward in the subspace ω_i.

The variance of $\hat{S}'_{t-1,\omega_i}$ is

$$\begin{aligned} D\left[\hat{S}'_{t-1,\omega_i}\right] &= E\left[\left(\hat{S}'_{t-1,\omega_i}\right)^2\right] - E^2\left[\hat{S}'_{t-1,\omega_i}\right] \\ &= E\left[\left(\frac{1}{t-1}\sum_{s=1}^{t-1}\hat{R}_{s,\omega_i}\right)^2\right] - \left(\frac{1}{t-1}\sum_{s=1}^{t-1}R_{s,\omega_i}\right)^2 \\ &= \frac{1}{(t-1)^2}\left(E\left[\left(\sum_{s=1}^{t-1}\hat{R}_{s,\omega_i}\right)^2\right] - \left(\sum_{s=1}^{t-1}R_{s,\omega_i}\right)^2\right) \\ &= \frac{1}{(t-1)^2} D\left[\hat{S}_{t-1,i}\right] \end{aligned}. \tag{2}$$

The theorem shows that, compared with Exp3, AvgHExp3 that uses the average estimated reward can decrease the Exp3's variance as the number (t) of GVAE-ABGEP running iterations increases.

2 Optimal Value of Parameter α

The parameter $\alpha \in [0, +\infty]$ in AvgHExp3 significantly affects the probability distribution $\mathcal{P}_t$ of selecting subspaces. If $\alpha = 0$, according to the importance weight equation,

$$P_{t,\omega_i} = \frac{e^{\alpha \hat{S}_{t-1,\omega_i}}}{\sum_{j=1}^{k} e^{\alpha \hat{S}_{t-1,\omega_j}}}. \tag{3}$$

$P_{t,\omega_i} = e^{\alpha \hat{S}_{t-1,\omega_i}} / \sum_{j=1}^{k} e^{\alpha \hat{S}_{t-1,\omega_j}} = e^{0 \hat{S}_{t-1,\omega_i}} / \sum_{j=1}^{k} e^{0 \hat{S}_{t-1,\omega_j}} = 1/k$. If α is large or even close to positive infinity, $P_{t,\omega_i} = e^{\alpha \hat{S}_{t-1,\omega_i}} / \sum_{j=1}^{k} e^{\alpha \hat{S}_{t-1,\omega_j}} = 1/\sum_{j=1}^{k} e^{\alpha\left(\hat{S}_{t-1,\omega_j} - \hat{S}_{t-1,\omega_i}\right)} \approx 0$ (except for the subspace ω_j with the best reward, $P_{t,\omega_j} \approx 1$). Therefore, if α is too large or small, excluding the subspace with $P_{t,\omega_i} \approx 1$, the probability of selecting each subspace will be the same, and AvgHExp3 loses its meaning. Moreover, GVAE-ABGEP will be close to classical GEPs.

From the perspective of balancing the subspace exploration and the subspace exploitation, the expected regret $R_n(\pi, r_t)$ of AvgHExp3 should be as small as possible. For determining α, we try to use α to limit the upper bound of $R_n(\pi, r_t)$ [1]. According to Theorem 21, the optimal value of parameter α is $\sqrt{\frac{logk + n(n-1)}{2nk}}$.

Theorem 21 *For limiting the upper bound of the regret reward $R_n(\pi, r_t)$, the optimal value of α is $\sqrt{\frac{logk + n(n-1)}{2nk}}$ when $\alpha \leq t$.*

Proof. The following proof is similar to the proof of Theorem 11.1 in [1]. However, the difference between Theorem 11.1 [1] and Theorem 21 is the estimated reward $\hat{R}$. The classical Exp3 [1] uses the total estimated reward while AvgHExp3 uses the average estimated reward.

For any subspace ω_i, the regret after the round n is

$$R_{n,\omega_i} = \sum_{t=\alpha}^{n+\alpha} R_{t,\omega_i} - E\left[\sum_{t=\alpha}^{n+\alpha} R_t\right]. \tag{4}$$

According to Equation 1 and the definition of $\hat{S}'_{t-1,\omega_i}$, i.e., $\hat{S}'_{t-1,\omega_i} = \frac{1}{t-1} \sum_{s=1}^{t-1} \hat{R}_{s,\omega_i}$, $\sum_{t=\alpha}^{n+\alpha} R_{t,\omega_i} = E\left[\sum_{t=\alpha}^{n+\alpha} \hat{R}_{t,\omega_i}\right] = E\left[n \hat{S}'_{n,\omega_i}\right]$. Meanwhile, according to the definition of expectation, $E\left[R_t\right]$ is expanded into $E\left[R_t\right] = \sum_{i=1}^{k} P_{t,\omega_i} R_{t,\omega_i} = \sum_{i=1}^{k} P_{t,\omega_i} E\left[\hat{R}_{t,\omega_i}\right]$. So Equation 4 can be computed as the follow equation.

$$
\begin{aligned}
R_{n,\omega_i} &= \sum_{t=\alpha}^{n+\alpha} R_{t,\omega_i} - E\left[\sum_{t=\alpha}^{n+\alpha} R_t\right] \\
&= nE\left[\hat{S}'_{n,\omega_i}\right] - E\left[\sum_{t=\alpha}^{n+\alpha}\sum_{i=1}^{k} P_{t,\omega_i} E\left[\hat{R}_{t,\omega_i}\right]\right] \\
&= nE\left[\hat{S}'_{n,\omega_i} - \frac{1}{n}\sum_{t=\alpha}^{n+\alpha}\sum_{i=1}^{k} P_{t,\omega_i}\hat{R}_{t,\omega_i}\right]
\end{aligned} \tag{5}
$$

For the following inequality

$$
e^{\alpha \hat{S}'_{n,\omega_i}} \leq \sum_{j=1}^{k} e^{\alpha \hat{S}'_{n,\omega_j}}, \tag{6}
$$

the right side of Equation 6 is expanded into the form of continuous multiplication in Inequality 7. Since $\hat{S}'_{0,\omega_j} = 0$ at the initialization of AvgHExp3, $e^{\alpha \hat{S}'_{0,\omega_j}} = 1$. So, Inequality 6 is finally transformed into Inequality 7.

$$
\begin{aligned}
e^{\alpha \hat{S}'_{n,\omega_i}} &\leq \sum_{j=1}^{k} e^{\alpha \hat{S}'_{0,\omega_j}} \frac{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{1,\omega_j}}}{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{0,\omega_j}}} \cdots \frac{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{n,\omega_j}}}{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{n-1,\omega_j}}} \\
&\leq k \prod_{t=1}^{n} \frac{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{t,\omega_j}}}{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{t-1,\omega_j}}}
\end{aligned} \tag{7}
$$

According to the definition of $\hat{S}'_{t,\omega_j}$ and Equation 3, the continuous multiplication term in Inequation 7 can be simplified to

$$
\begin{aligned}
\frac{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{t,\omega_j}}}{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{t-1,\omega_j}}} &= \sum_{j=1}^{k} \frac{e^{\alpha \frac{t-1}{t} \hat{S}'_{t-1,\omega_j}} \cdot e^{\alpha \frac{1}{t} \hat{R}_{t,\omega_j}}}{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{t-1,\omega_j}}} \\
&\leq \sum_{j=1}^{k} \frac{e^{\alpha \hat{S}'_{t-1,\omega_j}} \cdot e^{\alpha \frac{1}{t} \hat{R}_{t,\omega_j}}}{\sum_{j=1}^{k} e^{\alpha \hat{S}'_{t-1,\omega_j}}} \\
&= \sum_{j=1}^{k} P_{t,\omega_j} e^{\frac{\alpha}{t} \hat{R}_{t,\omega_j}}
\end{aligned} \tag{8}
$$

Since $e^x \leq n + \frac{1}{n}x + 2x^2$ when $x \leq 1$, and according to the following Equation

$$
\hat{R}_{t,\omega_j} = 1 - \frac{r_{t,a_t}}{P_{t,\omega_j}} = 1 - \frac{\mathrm{II}\left\{a_t = i\right\} r_{t,a_t}}{P_{t,\omega_j}} \tag{9}
$$

where $r_{t,a_t} \geq 0$, so $\frac{\alpha}{t}\hat{R}_{t,\omega_j} \leq 1$, the right side of Equation 8 can be bounded as

$$\sum_{j=1}^{k} P_{t,\omega_j} e^{\frac{\alpha}{t}\hat{R}_{t,\omega_j}} \leq n + \frac{1}{n}\sum_{j=1}^{k} P_{t,\omega_j}\frac{\alpha}{t}\hat{R}_{t,\omega_j} + 2\sum_{j=1}^{k} P_{t,\omega_j}\frac{\alpha^2}{t^2}\hat{R}_{t,\omega_j}^2. \tag{10}$$

Since $1 + x \leq e^x$ when $x \in R$, Inequality 10 can be further rewritten as

$$\sum_{j=1}^{k} P_{t,\omega_j} e^{\frac{\alpha}{t}\hat{R}_{t,\omega_j}} \leq e^{n-1+\frac{1}{n}\sum_{j=1}^{k} P_{t,\omega_j}\frac{\alpha}{t}\hat{R}_{t,\omega_j}+2\sum_{j=1}^{k} P_{t,\omega_j}\frac{\alpha^2}{t^2}\hat{R}_{t,\omega_j}^2} \tag{11}$$

Combining Inequality 7, Inequality 8 and Inequality 11,

$$\begin{aligned} e^{\alpha\hat{S}'_{n,\omega_i}} &\leq k e^{n(n-1)+\frac{1}{n}\alpha\sum_{t=1}^{n}\sum_{j=1}^{k}\frac{1}{t}P_{t,\omega_j}\hat{R}_{t,\omega_j}+2\alpha^2\sum_{t=1}^{n}\sum_{j=1}^{k}\frac{1}{t^2}P_{t,\omega_j}\hat{R}_{t,\omega_j}^2} \\ &\leq k e^{n(n-1)+\frac{1}{n}\alpha\sum_{t=1}^{n}\sum_{j=1}^{k}P_{t,\omega_j}\hat{R}_{t,\omega_j}+2\alpha^2\sum_{t=1}^{n}\sum_{j=1}^{k}P_{t,\omega_j}\hat{R}_{t,\omega_j}^2} \end{aligned} \tag{12}$$

Taking logarithm on both sides of Equation 12 and dividing by α,

$$\hat{S}'_{n,\omega_i} - \frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{k} P_{t,\omega_j}\hat{R}_{t,\omega_j} \leq \frac{\log k}{\alpha} + \frac{n(n-1)}{\alpha} + 2\alpha\sum_{t=1}^{n}\sum_{j=1}^{k} P_{t,\omega_j}\hat{R}_{t,\omega_j}^2 \tag{13}$$

According to $E\left[\sum_{t=1}^{n}\sum_{j=1}^{k} P_{t,\omega_j}\hat{R}_{t,\omega_j}^2\right] \leq nk$ in [1], The expectation of Inequality 13 can be further modified as

$$E\left[\hat{S}_{n,\omega_i} - \frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{k} P_{t,\omega_j}\hat{R}_{t,\omega_j}\right] \leq \frac{\log k}{\alpha} + \frac{n(n-1)}{\alpha} + 2\alpha nk \tag{14}$$

Combining Equation 5 and Inequality 14,

$$R_{n,\omega_i} \leq \frac{n\log k + n^2(n-1)}{\alpha} + 2\alpha n^2 k \tag{15}$$

According to the inequality $a + b \geq 2\sqrt{ab}$, the right side of Inequality 15 is

$$\frac{n\log k + n^2(n-1)}{\alpha} + 2\alpha n^2 k \geq 2\sqrt{2n^2k(n\log k + n^2(n-1))}. \tag{16}$$

For limit the right side of Inequality 15,

$$\frac{n\log k + n^2(n-1)}{\alpha} = 2\alpha n^2 k. \tag{17}$$

So the optimal value of α is $\sqrt{\dfrac{logk + n(n-1)}{2nk}}$.

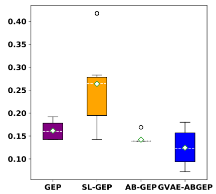
3 More Experimental Details

Table 1: Symbolic Regression Benchmarks

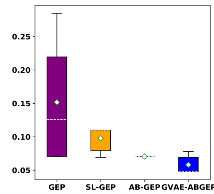
number	name	formula	dataset
F1	Nguyen-4	$x^6+x^5+x^4+x^3+x^2+x$	$\mathrm{U}[-1,1,20]$
F2	Nguyen-3	$x^5+x^4+x^3+x^2+x$	$\mathrm{U}[-1,1,20]$
F3	Nguyen-2	$x^4+x^3+x^2+x$	$\mathrm{U}[-1,1,20]$
F4	Nguyen-1	x^3+x^2+x	$\mathrm{U}[-1,1,20]$
F5	Koza-3	$x^6-2x^4+x^2$	$\mathrm{U}[-1,1,20]$
F6	Koza-2	x^5-2x^3+x	$\mathrm{U}[-1,1,20]$
F7	Nguyen-5	$\sin\left(x^2\right)\cos(x)-1$	$\mathrm{U}[-1,1,20]$
F8	Nguyen-6	$\sin(x)+\sin\left(x+x^2\right)$	$\mathrm{U}[-1,1,20]$
F9	Nguyen-7	$\ln(x+1)+\ln\left(x^2+1\right)$	$\mathrm{U}[0,2,20]$
F10	Keijzer-1	$0.3x\sin(2\pi x)$	$\mathrm{E}[-1,1,0.1]$
F11	Vladislavleva-2	$e^{-x}x^3(\cos x\sin x)\left(\cos x\sin^2 x-1\right)$	$\mathrm{E}[0.05,10,0.1]$
F12	Keijzer-8	$\sqrt{x}$	$\mathrm{U}[0,4,20]$
F13	Nguyen-9	$\sin(x)+\sin\left(y^2\right)$	$\mathrm{U}[0,1,20]$
F14	Nguyen-10	$2\sin(x)\cos(y)$	$\mathrm{U}[-1,1,100]$
F15	Keijzer-11	$xy+\sin((x-1)(y-1))$	$\mathrm{U}[-3,3,100]$
F16	Korns-2	$\frac{1}{1+x^{-4}}+\frac{1}{1+y^{-4}}$	$\mathrm{E}[-5,5,0.4]$
F17	Keijzer-12	$x^4-x^3+\frac{y^2}{2}-y$	$\mathrm{U}[-3,3,20]$
F18	Vladislavleva-3	$e^{-x}x^3(\cos x\sin x)\left(\cos x\sin^2 x-1\right)(y-5)$	$x:\mathrm{E}[0.05,10,0.1]$ $y:\mathrm{E}[0.05,10.05,2]$

Table 2: Parameters of Algorithms

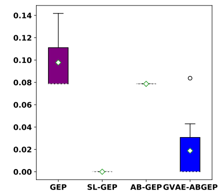
Name	Parameter	Value
GEP	function set	$+,-,\times,\div$,sin,cos,$ln(\vert n\vert)$,exp
	population size	100
	head length	10
	chromosome length	21
	max generations	1000
	crossover Rate	0.7
	mutation rate	0.1
SL-GEP	function set	$+,-,\times,\div$,sin,cos,$ln(\vert n\vert)$,exp
	population size	100
	head length	10
	chromosome length	21
	max generations	1000
	number of ADFs	2
	head length of ADFs	3
AB-GEP	function set	$+,-,\times,\div$,sin,cos,$ln(\vert n\vert)$,exp
	population size	100
	head length	10
	chromosome length	21
	max generations	1000
	crossover Rate	0.7
	mutation rate	0.1
	number of subspaces	[81,100]
GVAE-ABGEP	function set	$+,-,\times,\div$,sin,cos,$ln(\vert n\vert)$,exp
	population size	100
	head length	10
	chromosome length	21
	max generations	1000
	crossover Rate	0.7
	mutation rate	0.1
	number of subspaces	[81,100]
	learning rate	$\sqrt{\frac{logk+n(n-1)}{2nk}}$



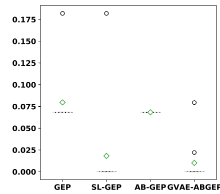
(a) F1



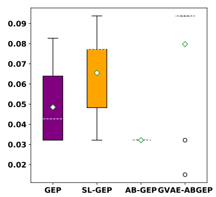
(b) F2



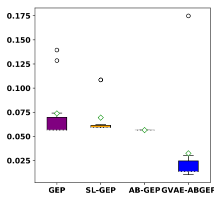
(c) F3



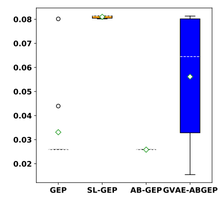
(d) F4



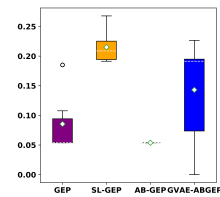
(e) F5



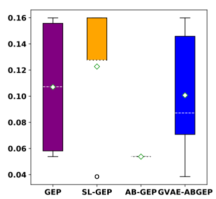
(f) F6



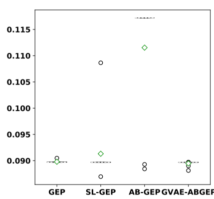
(g) F7



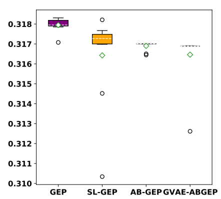
(h) F8



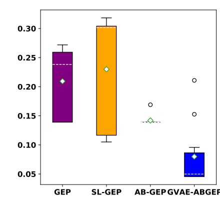
(i) F9



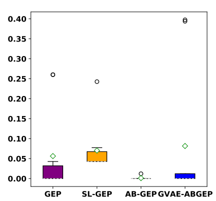
(j) F10



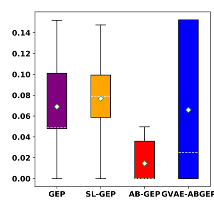
(k) F11



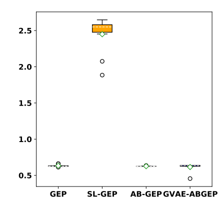
(l) F12



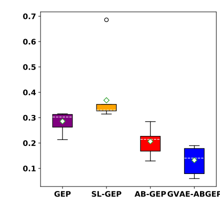
(m) F13



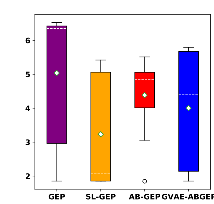
(n) F14



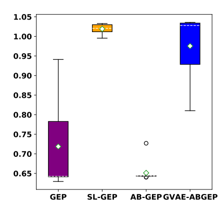
(o) F15



(p) F16

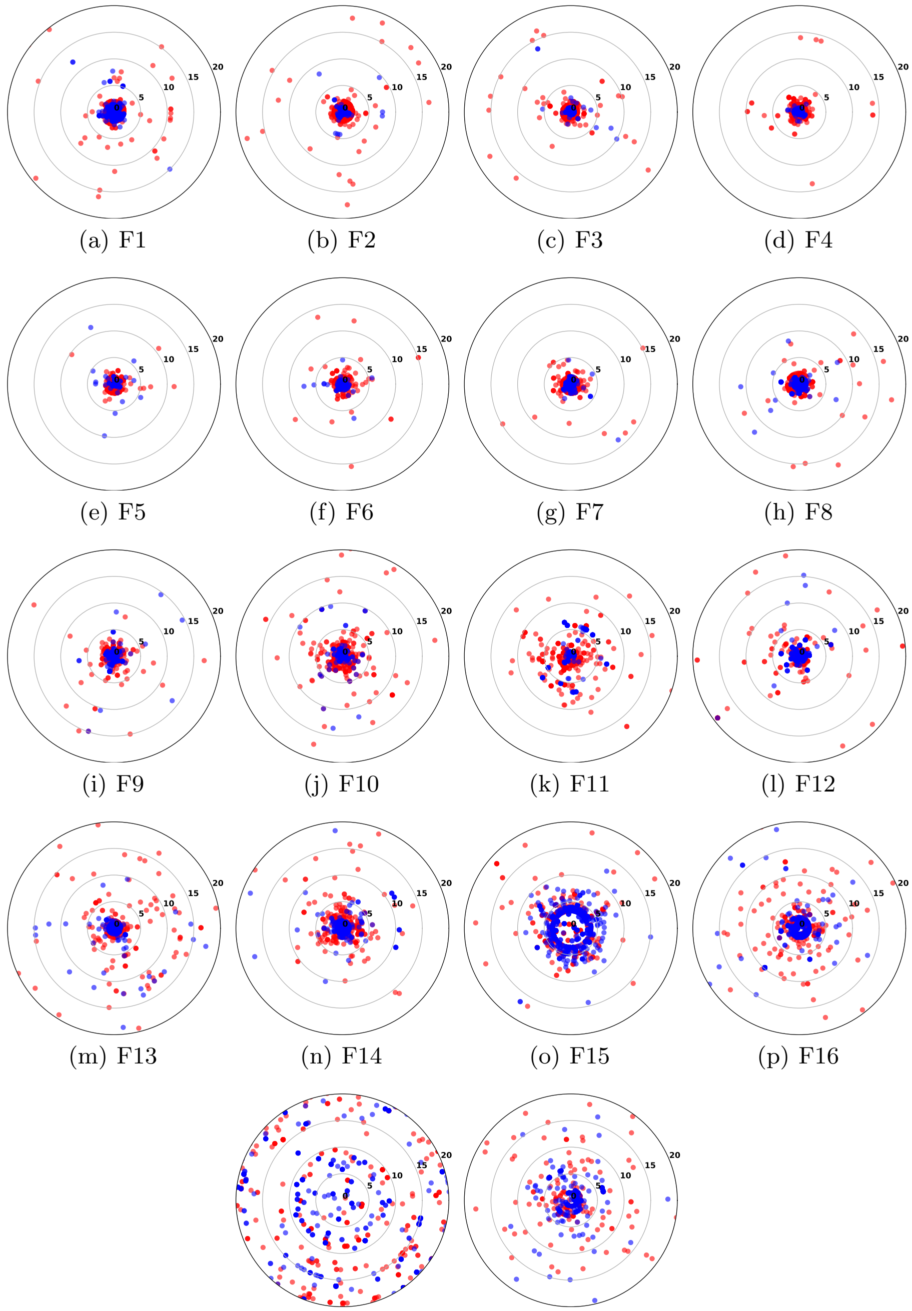


(q) F17



(r) F18

Fig. 1: Fitness comparison on all SRB benchmarks.



(a) F1 (b) F2 (c) F3 (d) F4

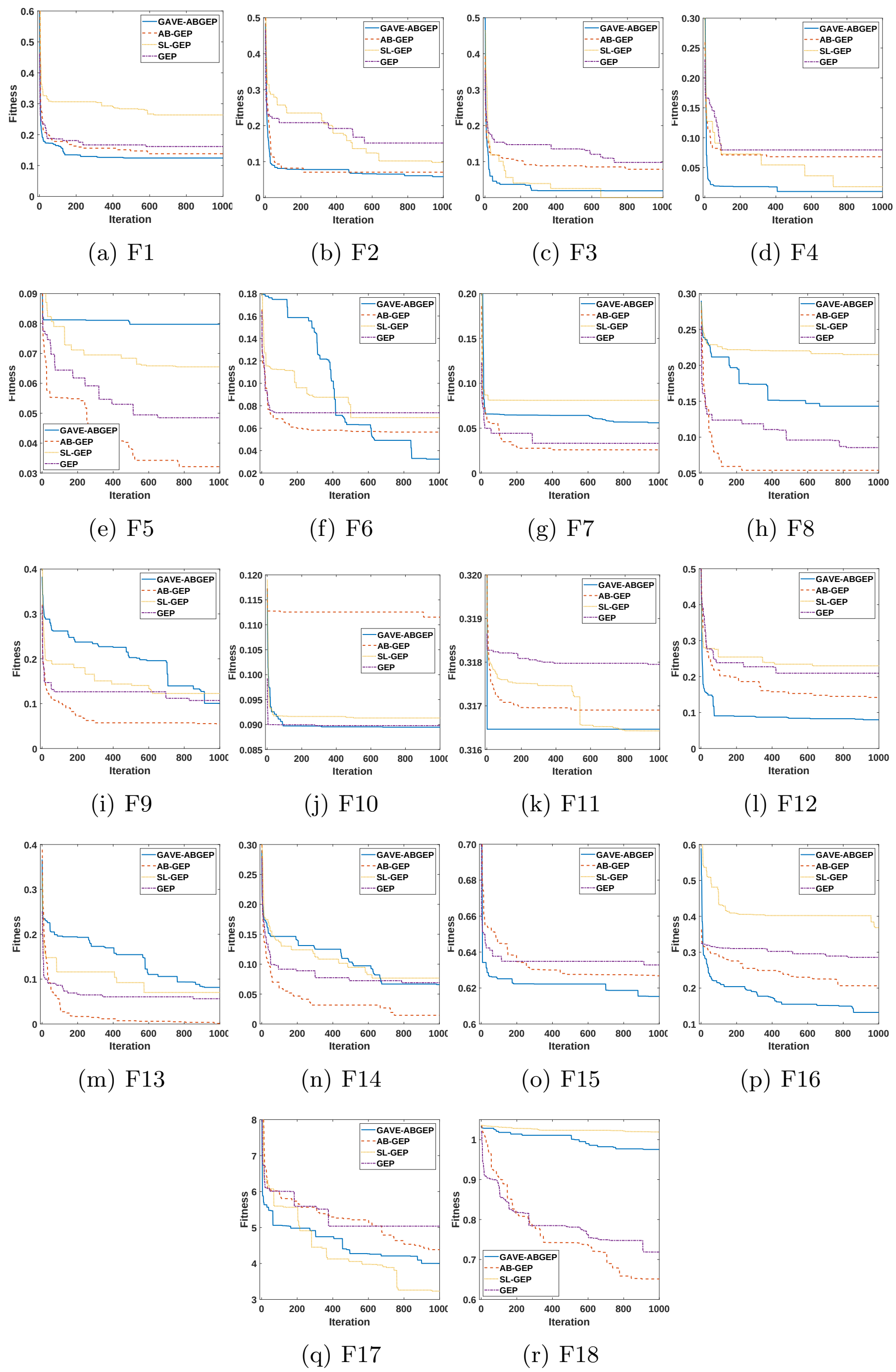
(e) F5 (f) F6 (g) F7 (h) F8

(i) F9 (j) F10 (k) F11 (l) F12

(m) F13 (n) F14 (o) F15 (p) F16

(q) F17 (r) F18

Fig. 2: Individual Fitness distribution on all SRB benchmarks. A blue (or red) point represents the fitness of a GVAE-ABGEP (or an AB-GEP) individual recorded every 100 generations.



(a) F1 (b) F2 (c) F3 (d) F4
(e) F5 (f) F6 (g) F7 (h) F8
(i) F9 (j) F10 (k) F11 (l) F12
(m) F13 (n) F14 (o) F15 (p) F16
(q) F17 (r) F18

Fig. 3: Convergence comparison on all SRB benchmarks.

References

1. Lattimore, T., Szepesvári, C.: Bandit algorithms. Cambridge University Press (2020)