Trading off Quality and Uncertainty through Multi-Objective Optimisation in Batch Bayesian Optimisation (Supplementary Material)

Chao Jiang and Miqing Li*

School of Computer Science, University of Birmingham

cxj249@student.bham.ac.uk, m.li.8@bham.ac.uk

In this supplementary document, Sections 1 and 2 first provide a detailed explanation of the TOPSIS and the test problems, respectively. Sections 3, 4, and 5 then present the batch optimisation outcomes and convergence trajectories for all considered batch sizes $q \in \{5, 10, 20\}$, addressing both synthetic and robot pushing optimisation problems. Specifically, these sections include comparisons with state-of-the-art methods (including noisy experiments and runtime comparison), an ablation study, and a sensitivity analysis.

*Corresponding author.

1 TOPSIS

In this work, we consider a multi-criteria decision analysis technique, i.e. the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) (Hwang and Yoon, 1981; Lai et al., 1994), which has been frequently used in the multi-criteria decision analysis area and multi-objective optimisation area (e.g., in Opricovic and Tzeng (2004); Cinelli et al. (2020, 2022); Gu et al. (2023); Petropoulos et al. (2024)). TOPSIS chooses a solution from a set of solutions on the basis of how far they are from the ideal point and the negative-ideal point in the Euclidean space, in order to achieve a good trade-off between multiple objectives. Algorithm 1 gives the procedure of TOPSIS.

Algorithm 1: Steps of TOPSIS

Input: Weights $W = \{w_j\}_{j=1}^{n}$, where w_j is the jth weight corresponding to the jth objective; decision matrix $(g_{ij})_{m\times n}$, that consists of m solutions and n objectives, with the intersection of each solution and objective denoted as g_{ij}.

Output: The best solution

1 Normalise the decision matrix $(g_{ij})_{m\times n}$:
$r_{ij} = \frac{g_{ij}}{\sqrt{\sum_{k=1}^{m} g_{kj}^2}}, \quad i = 1, 2, ..., m, \quad j = 1, 2, ..., n$

2 Weight the normalised decision matrix $(r_{ij})_{m\times n}$:
$v_{ij} = r_{ij} \times w_j, \quad i = 1, 2, ..., m, \quad j = 1, 2, ..., n$
where w_j is the weight corresponding to the jth objective.

3 Determine the ideal point A^+ and the negative-ideal point A^-:
$A^+ = \{\min_{i\in\{1,2,...,m\}} v_{ij}\}_{j=1}^{n} = \{v_j^+\}_{j=1}^{n}$
$A^- = \{\max_{i\in\{1,2,...,m\}} v_{ij}\}_{j=1}^{n} = \{v_j^-\}_{j=1}^{n}$

4 Calculate the Euclidean distances:
$d_i^+ = \sqrt{\sum_{j=1}^{n}(v_{ij} - v_j^+)^2}, \; i = 1, 2, ..., m$
$d_i^- = \sqrt{\sum_{j=1}^{n}(v_{ij} - v_j^-)^2}, \; i = 1, 2, ..., m$

5 Calculate the relative closeness to the ideal point:
$c_i = \frac{d_i^+}{d_i^+ + d_i^-}, \; i = 1, 2, ..., m$

6 Determine the best solution by identifying the minimum c_i, where $i = 1, 2, ..., m$.

2 Details of Test Problems

We consider 14 well-known synthetic and real-world problems (e.g., the active learning problem in robotic pushing (Wang and Jegelka, 2017)), as detailed in Table 1.

Test problem	d
WangFreitas (Wang and de Freitas, 2014)	1
BraninForrester (Sobester et al., 2008)	2
Branin*	2
Eggholder*	2
GoldsteinPrice*	2
SixHumpCamel*	2
Hartmann6*	6
Ackley*	2, 10
Griewank *	2, 10
GSobol (González et al., 2016)	10
Push (Wang and Jegelka, 2017)	4, 8

Table 1: Test problems with the corresponding dimension d. Apart from the practical problem Push, expressions of the problems can be found at `https://www.sfu.ca/~ssurjano/optimization.html`, which we marked with "$*$". For the rest, the expressions can be found in the corresponding references.

3 Comparison with Well-Established Methods

Tables 2 and 4 present the results (mean and standard deviation) of the regrets obtained for batch sizes of $q = 10$ and $q = 20$, respectively, and those that are statistically equivalent to the proposed POEE according to the Wilcoxon signed-rank test (Sidney, 1957). Additionally, the comparative results between POEE and the peer algorithms are summarised in Tables 3 and 5, respectively.

Method	WangFreitas (1)		BraninForrester (2)		Branin (2)		Eggholder (2)		GoldsteinPrice (2)		SixHumpCamel (2)		Hartmann6 (6)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	**6.81e-08**	**1.69e-07**	**5.15e-08**	**8.14e-08**	2.54e-06	3.06e-06	**2.38e+01**	**2.85e+01**	**7.31e-02**	**1.31e-01**	**2.74e-07**	**2.99e-07**	**2.00e-02**	**4.45e-02**
RS	1.45e-01	2.66e-01(−)	5.13e-01	4.64e-01(−)	1.90e-01	1.77e-01(−)	1.44e+02	6.78e+01(−)	8.15e+00	6.29e+00(−)	4.76e-02	3.86e-02(−)	1.00e+00	3.18e-01(−)
qEI	1.29e-07	3.05e-07(∼)	9.18e-06	1.55e-05(−)	3.35e-06	2.90e-06(∼)	1.27e+02	6.70e+01(−)	8.12e-01	6.54e-01(−)	1.91e-05	4.78e-05(−)	4.52e-02	6.94e-02(−)
LP	1.53e+00	8.46e-01(−)	4.00e-05	1.30e-04(−)	1.37e-05	6.30e-05(∼)	4.88e+01	4.67e+01(−)	1.34e+00	8.90e-01(−)	2.11e-06	4.03e-06(−)	3.99e-02	5.78e-02(−)
PLAyBOOK	1.60e+00	8.00e-01(−)	6.88e-06	1.01e-05(−)	3.13e-06	3.92e-06(∼)	3.45e+01	4.93e+01(∼)	8.07e-01	9.50e-01(−)	7.03e-06	9.40e-06(−)	6.34e-02	6.70e-02(−)
KB	4.05e-04	1.97e-03(−)	6.23e-08	9.57e-08(∼)	**6.16e-07**	**7.95e-07(+)**	3.82e+01	2.93e+01(−)	1.26e-01	3.67e-01(∼)	5.87e-07	1.13e-06(∼)	2.81e-02	5.16e-02(−)
TS	1.40e+00	9.17e-01(−)	5.96e-04	1.63e-04(−)	1.17e-04	8.26e-05(−)	4.65e+01	2.39e+01(−)	4.71e-01	5.89e-01(−)	1.98e-04	1.43e-04(−)	1.07e-01	4.38e-02(−)
ϵS-PF	1.67e+00	7.45e-01(−)	7.77e-05	1.93e-04(−)	4.25e-05	7.64e-05(−)	1.79e+02	1.45e+02(−)	1.24e+01	1.84e+01(−)	2.72e-02	1.47e-01(−)	3.01e-02	5.45e-02(−)
AEGiS	6.93e-07	5.10e-07(−)	1.66e-03	2.24e-04(−)	6.17e-04	5.05e-04(−)	4.58e+01	2.76e+01(−)	1.65e+00	1.58e+00(−)	2.22e-04	1.04e-05(−)	3.21e-02	5.16e-02(−)
GIBBON	2.06e-01	5.99e-01(−)	3.08e-04	5.39e-04(−)	8.41e-04	1.44e-03(−)	1.62e+02	1.33e+02(−)	6.36e+00	7.15e+00(−)	1.53e-02	4.34e-02(−)	2.33e-02	4.99e-02(−)
MACE	1.73e+00	6.80e-01(−)	3.72e-05	1.16e-04(−)	1.81e-06	1.64e-06(∼)	1.41e+02	1.28e+02(−)	2.22e+00	3.25e+00(−)	1.45e-06	2.11e-06(−)	3.82e-02	7.33e-02(−)
Gupta	1.87e+00	4.99e-01(−)	3.05e-01	1.64e+00(−)	1.75e-05	4.68e-05(−)	1.37e+02	1.07e+02(−)	7.64e-01	9.48e-01(−)	1.34e-06	2.30e-06(−)	3.70e-02	5.61e-02(−)
TuRBO	2.16e+00	7.86e-01(−)	1.38e+01	1.83e+01(−)	1.95e-01	5.59e-01(−)	4.11e+02	1.91e+02(−)	1.21e+02	2.80e+02(−)	7.07e-01	1.12e+00(−)	9.20e-01	6.08e-01(−)

Method	Ackley (2)		Griewank (2)		Ackley (10)		Griewank (10)		GSobol (10)		Push4 (4)		Push8 (8)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	**4.85e-01**	**2.53e+00**	**2.13e-01**	**1.28e-01**	**4.53e+00**	**5.06e+00**	9.23e-01	1.10e-01	2.75e+03	4.32e+03	**3.09e-02**	**3.15e-02**	**1.25e+00**	**1.49e+00**
RS	6.28e+00	1.81e+00(−)	1.05e+00	5.67e-01(−)	1.87e+01	1.00e+00(−)	8.28e+01	1.95e+01(−)	3.15e+03	2.78e+03(∼)	4.16e-01	2.73e-01(−)	3.43e+00	2.28e+00(−)
qEI	4.19e+00	2.58e+00(−)	1.17e+00	5.05e-01(−)	1.89e+01	8.02e-01(−)	9.24e+01	2.31e+01(−)	2.53e+03	1.90e+03(∼)	2.06e-01	1.66e-01(−)	1.59e+00	1.56e+00(∼)
LP	8.64e-01	2.59e+00(−)	8.30e-01	1.39e+00(−)	1.27e+01	3.68e+00(−)	4.61e+00	1.98e+00(−)	**7.27e+02**	**1.42e+03(+)**	1.35e-01	1.27e-01(−)	1.93e+00	1.68e+00(−)
PLAyBOOK	1.43e+00	3.79e+00(−)	8.23e-01	2.08e+00(−)	1.59e+01	2.37e+00(−)	5.51e+00	3.49e+00(−)	1.33e+03	1.44e+03(∼)	1.11e-01	1.08e-01(−)	2.55e+00	1.93e+00(−)
KB	5.06e-01	3.38e-01(−)	3.09e-01	1.83e-01(∼)	7.21e+00	6.81e+00(∼)	1.03e+00	4.36e-02(−)	3.61e+03	5.08e+03(∼)	1.39e-01	6.50e-02(−)	1.75e+00	1.26e+00(∼)
TS	8.11e-01	2.57e+00(−)	3.66e-01	1.55e-01(−)	1.26e+01	2.34e-01(−)	1.71e+01	1.34e+00(−)	7.82e+03	6.91e+03(−)	3.08e-01	1.42e-01(−)	3.05e+00	1.62e+00(−)
ϵS-PF	2.37e+00	3.86e+00(∼)	5.40e-01	2.22e+00(−)	1.66e+01	4.30e+00(−)	**7.38e-01**	**2.74e-01(+)**	1.61e+03	2.73e+03(∼)	5.45e-02	1.18e-01(∼)	2.07e+00	1.69e+00(−)
AEGiS	5.63e-01	2.80e-01(−)	3.86e-01	2.20e-01(−)	1.64e+01	2.13e+00(−)	1.88e+00	4.08e-01(−)	2.43e+03	2.62e+03(∼)	2.81e-01	1.88e-01(−)	2.22e+00	1.35e+00(−)
GIBBON	4.18e+00	5.86e+00(−)	4.85e-01	3.74e-01(−)	1.89e+01	8.12e-01(−)	8.46e+01	2.10e+01(−)	3.45e+03	3.02e+03(∼)	1.90e-01	2.70e-01(−)	3.09e+00	2.08e+00(−)
MACE	2.41e+00	4.41e+00(−)	2.46e+00	5.98e+00(−)	1.74e+01	1.62e+00(−)	8.12e-01	1.70e-01(+)	1.10e+03	2.39e+03(+)	1.33e-01	2.28e-01(−)	2.86e+00	1.90e+00(−)
Gupta	2.30e+00	4.76e+00(−)	1.79e+00	5.43e+00(∼)	1.81e+01	1.36e+00(−)	8.04e-01	1.97e-01(+)	8.31e+02	2.91e+03(+)	9.71e-02	1.54e-01(−)	3.07e+00	2.09e+00(−)
TuRBO	1.61e+01	4.05e+00(−)	4.17e+00	5.28e+00(−)	1.92e+01	5.05e-01(−)	2.89e+01	1.57e+01(−)	1.61e+03	1.87e+03(∼)	4.26e-01	9.17e-01(∼)	3.55e+00	2.57e+00(−)

Table 2: The performance outcomes of the proposed POEE and nine state-of-the-art methods were evaluated across all test problems using a batch size of $q = 10$. The method demonstrating the best mean performance is denoted in bold. The symbols "+", "$\sim$", and "$-$" indicate that the method is statistically better than, equivalent to, and worse than POEE respectively.

	RS	qEI	LP	PLAyBOOK	KB	TS	ϵS-PF	AEGiS	GIBBON	MACE	Gupta	TuRBO
POEE	13/ 1/ 0	10/ 4/ 0	12/ 1/ 1	11/ 3/ 0	6/ 7/ 1	14/ 0/ 0	10/ 3/ 1	13/ 1/ 0	13/ 1/ 0	11/ 1/ 2	11/ 1/ 2	12/ 2/ 0

Table 3: The summary of statistical outcomes, as detailed in Table 2. Here, the left, median, and right numbers denote the counts of test problems where the POEE was statistically superior, equivalent, or inferior to the peer algorithm, respectively.

Method	WangFreitas (1)		BraninForrester (2)		Branin (2)		Eggholder (2)		GoldsteinPrice (2)		SixHumpCamel (2)		Hartmann6 (6)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	**7.30e-08**	**2.56e-07**	6.39e-06	1.33e-05	3.14e-06	5.37e-06	3.73e+01	2.96e+01	1.59e-01	3.05e-01	**1.40e-06**	**2.54e-06**	**2.33e-02**	**5.29e-02**
RS	1.00e-01	1.37e-01(−)	5.62e-01	4.74e-01(−)	1.55e-01	1.61e-01(−)	1.23e+02	7.01e+01(−)	6.72e+00	7.81e+00(−)	7.98e-02	7.47e-02(−)	9.30e-01	3.09e-01(−)
qEI	1.44e-07	3.13e-07(−)	2.20e-05	2.59e-05(−)	1.32e-05	1.95e-05(−)	1.41e+02	6.41e+01(−)	1.60e+00	9.52e-01(−)	4.72e-05	8.63e-05(−)	6.44e-02	7.74e-02(−)
LP	1.33e+00	9.43e-01(−)	3.24e-05	1.08e-04(∼)	7.94e-06	1.99e-05(∼)	5.55e+01	6.95e+01(∼)	3.94e+00	3.04e+00(−)	1.01e-05	1.63e-05(−)	3.37e-02	5.40e-02(−)
PLAyBOOK	1.40e+00	9.17e-01(−)	4.11e-05	7.23e-05(−)	3.59e-05	1.35e-04(−)	3.41e+01	3.41e+01(∼)	4.09e+00	3.04e+00(−)	3.71e-03	6.14e-03(−)	5.93e-02	6.94e-02(−)
KB	1.92e-05	7.52e-05(−)	**1.22e-07**	**2.21e-07(+)**	**7.04e-07**	**4.72e-07(+)**	**3.34e+01**	**2.73e+01(∼)**	**1.34e-02**	**2.25e-02(+)**	2.51e-06	7.37e-06(∼)	2.70e-02	4.96e-02(−)
TS	8.00e-01	9.80e-01(−)	1.20e-03	1.75e-04(−)	6.47e-04	1.97e-05(−)	3.42e+01	2.52e+01(∼)	9.51e-01	9.11e-01(−)	4.51e-04	1.30e-04(−)	9.80e-02	2.83e-02(−)
ϵS-PF	1.73e+00	6.78e-01(−)	3.17e-03	1.29e-02(−)	2.19e-04	4.98e-04(−)	2.43e+02	1.62e+02(−)	1.38e+01	1.86e+01(−)	5.45e-02	2.04e-01(−)	3.41e-02	6.27e-02(∼)
AEGiS	2.05e-06	1.95e-06(−)	1.27e-03	7.54e-04(−)	6.05e-04	2.71e-04(−)	5.34e+01	2.37e+01(−)	1.95e+00	1.76e+00(−)	4.06e-04	1.99e-04(−)	2.63e-02	5.39e-02(∼)
GIBBON	6.67e-02	3.59e-01(−)	3.87e-01	1.68e+00(−)	4.55e-04	6.09e-04(−)	1.83e+02	1.23e+02(−)	1.85e+01	2.27e+01(−)	9.52e-02	1.40e-01(−)	5.97e-02	9.69e-02(−)
MACE	1.87e+00	4.99e-01(−)	1.10e-04	2.64e-04(−)	9.41e-05	3.00e-04(−)	1.93e+02	1.68e+02(−)	1.58e+01	1.58e+01(−)	2.19e-03	3.52e-03(−)	5.62e-02	8.59e-02(−)
Gupta	1.80e+00	6.00e-01(−)	3.05e-01	1.64e+00(−)	6.24e-05	1.65e-04(−)	1.51e+02	1.47e+02(−)	8.51e+00	1.47e+01(−)	1.58e-04	3.02e-04(−)	5.31e-02	6.53e-02(−)
TuRBO	2.48e+00	1.00e+00(−)	2.03e+01	2.04e+01(−)	1.74e+00	3.27e+00(−)	4.11e+02	1.91e+02(−)	2.93e+02	7.59e+02(−)	9.52e-01	1.79e+00(−)	1.48e+00	6.20e-01(−)
Method	**Ackley (2)**		**Griewank (2)**		**Ackley (10)**		**Griewank (10)**		**GSobol (10)**		**Push4 (4)**		**Push8 (8)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	**6.03e-02**	**1.91e-01**	**2.26e-01**	**1.22e-01**	**4.47e+00**	**4.86e+00**	1.02e+00	3.58e-02	5.22e+03	8.46e+03	**4.07e-02**	**5.89e-02**	**1.77e+00**	**1.48e+00**
RS	6.44e+00	1.91e+00(−)	1.15e+00	4.21e-01(−)	1.89e+01	6.57e-01(−)	8.71e+01	2.43e+01(−)	5.61e+03	5.07e+03(∼)	4.31e-01	2.48e-01(−)	3.76e+00	2.01e+00(−)
qEI	4.25e+00	2.26e+00(−)	1.09e+00	5.68e-01(−)	1.91e+01	4.96e-01(−)	8.84e+01	2.10e+01(−)	4.78e+03	4.04e+03(∼)	2.12e-01	1.44e-01(−)	1.98e+00	1.46e+00(∼)
LP	1.56e-01	4.49e-01(−)	4.72e-01	2.24e-01(−)	1.32e+01	3.32e+00(−)	8.47e+00	6.54e+00(−)	**1.13e+03**	**1.56e+03(+)**	1.23e-01	1.70e-01(−)	2.29e+00	1.85e+00(−)
PLAyBOOK	2.41e+00	4.41e+00(−)	8.05e-01	1.32e+00(−)	1.62e+01	1.96e+00(−)	7.90e+00	5.77e+00(−)	1.71e+03	2.10e+03(∼)	2.23e-01	4.73e-01(−)	2.97e+00	1.87e+00(−)
KB	8.46e-01	7.20e-01(−)	3.33e-01	1.34e-01(−)	7.93e+00	5.96e+00(−)	1.57e+00	2.84e-01(−)	4.21e+03	6.23e+03(∼)	1.48e-01	7.94e-02(−)	1.80e+00	1.18e+00(∼)
TS	8.28e-01	1.99e-01(−)	3.32e-01	1.32e-01(−)	1.45e+01	3.89e-01(−)	1.51e+01	6.24e-01(−)	5.87e+03	5.17e+03(∼)	2.89e-01	1.42e-01(−)	2.80e+00	1.71e+00(−)
ϵS-PF	6.51e+00	7.11e+00(−)	1.98e+00	6.72e+00(∼)	1.56e+01	5.03e+00(−)	**5.93e-01**	**2.75e-01(+)**	2.39e+03	4.44e+03(∼)	1.43e-01	2.21e-01(−)	2.44e+00	2.02e+00(−)
AEGiS	7.39e-01	5.64e-01(−)	3.19e-01	1.70e-01(∼)	1.74e+01	2.61e+00(−)	3.53e+00	1.99e+00(−)	3.85e+03	3.35e+03(∼)	2.94e-01	1.70e-01(−)	2.49e+00	1.76e+00(−)
GIBBON	4.33e+00	5.35e+00(−)	1.56e+00	2.34e+00(−)	1.88e+01	7.56e-01(−)	9.02e+01	2.22e+01(−)	4.60e+03	3.80e+03(∼)	1.75e-01	2.73e-01(−)	3.41e+00	2.17e+00(−)
MACE	8.74e+00	7.86e+00(−)	5.57e+00	8.65e+00(−)	1.80e+01	1.41e+00(−)	7.64e-01	2.04e-01(+)	2.65e+03	6.70e+03(+)	4.04e-01	7.08e-01(−)	3.53e+00	2.17e+00(−)
Gupta	6.85e+00	7.88e+00(−)	1.90e+00	3.60e+00(−)	1.84e+01	9.30e-01(−)	8.22e-01	1.71e-01(+)	2.31e+03	7.26e+03(+)	3.81e-01	5.99e-01(−)	3.92e+00	2.47e+00(−)
TuRBO	1.63e+01	4.02e+00(−)	9.88e+00	9.26e+00(−)	1.92e+01	4.27e-01(−)	6.81e+01	2.40e+01(−)	3.39e+03	6.77e+03(∼)	6.64e-01	1.03e+00(−)	4.06e+00	2.63e+00(−)

Table 4: The performance outcomes of the proposed POEE and nine state-of-the-art methods were evaluated across all test problems using a batch size of $q = 20$. The method demonstrating the best mean performance is denoted in bold. The symbols "+", "∼", and "−" indicate that the method is statistically better than, equivalent to, and worse than POEE respectively.

	RS	qEI	LP	PLAyBOOK	KB	TS	ϵS-PF	AEGiS	GIBBON	MACE	Gupta	TuRBO
POEE	13/ 1/ 0	12/ 2/ 0	10/ 3/ 1	12/ 2/ 0	7/ 4/ 3	12/ 2/ 0	10/ 3/ 1	11/ 3/ 0	13/ 1/ 0	12/ 0/ 2	12/ 0/ 2	13/ 1/ 0

Table 5: The summary of statistical outcomes, as detailed in Table 4. Here, the left, median, and right numbers denote the counts of test problems where the POEE was statistically superior, equivalent, or inferior to the peer algorithm, respectively.

Method	WangFreitas (1)		BraninForrester (2)		Branin (2)		Eggholder (2)		GoldsteinPrice (2)		SixHumpCamel (2)		Hartmann6 (6)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	9.79e-03	3.20e-02	1.99e-04	1.57e-04	2.59e-04	2.53e-04	3.51e+01	2.81e+01	4.67e-01	7.23e-01	**4.41e-04**	**4.19e-04**	5.24e-02	7.69e-02
qEI	**7.52e-06**	**2.82e-05(+)**	3.27e-04	4.27e-04(∼)	3.04e-04	2.84e-04(∼)	6.18e+01	5.30e+01(−)	7.25e-01	5.75e-01(−)	6.89e-04	6.92e-04(−)	5.74e-02	8.39e-02(∼)
LP	2.27e-01	2.78e-01(−)	4.32e-04	4.02e-04(∼)	5.14e-03	9.66e-03(−)	7.60e+01	6.45e+01(−)	1.74e+00	1.59e+00(−)	1.98e-02	1.79e-02(−)	8.52e-01	3.69e-01(−)
PLAyBOOK	2.03e-01	2.43e-01(−)	3.71e-04	2.40e-04(−)	1.00e-02	2.89e-02(−)	6.40e+01	5.30e+01(−)	2.02e+00	2.02e+00(−)	1.69e-02	1.50e-02(−)	9.84e-01	4.56e-01(−)
KB	7.45e-01	8.99e-01(−)	**1.02e-04**	**9.51e-05(+)**	1.77e-04	1.62e-04(∼)	**3.25e+01**	**2.78e+01(∼)**	**1.44e-01**	**3.62e-01(+)**	5.41e-04	5.78e-04(∼)	6.77e-02	5.68e-02(∼)
TS	2.98e-04	5.92e-04(∼)	2.52e-04	2.24e-04(∼)	**4.56e-05**	**5.39e-05(+)**	5.04e+01	2.86e+01(∼)	1.67e-01	2.76e-01(∼)	6.41e-04	5.23e-04(∼)	8.90e-02	1.30e-02(−)
ϵS-PF	1.13e+00	9.91e-01(−)	9.27e-04	1.78e-03(−)	1.03e-02	3.38e-02(−)	1.95e+02	1.57e+02(−)	4.54e+00	8.46e+00(−)	5.96e-02	1.79e-01(−)	4.52e-01	3.84e-01(−)
AEGiS	2.87e-01	6.37e-01(−)	4.55e-04	2.92e-04(−)	4.17e-04	3.30e-04(−)	4.12e+01	2.63e+01(∼)	1.73e+00	1.51e+00(−)	2.07e-02	7.12e-02(−)	4.22e-01	3.38e-01(−)
GIBBON	6.67e-01	9.43e-01(∼)	1.18e-03	1.16e-03(−)	1.16e-03	1.14e-03(−)	8.54e+01	6.50e+01(−)	4.28e+00	5.17e+00(−)	4.04e-03	5.13e-03(−)	**4.43e-02**	**4.97e-02(∼)**
MACE	8.78e-01	8.92e-01(−)	3.44e-04	3.53e-04(∼)	7.06e-02	2.82e-01(−)	4.89e+01	3.84e+01(∼)	4.92e-01	5.37e-01(∼)	2.08e-02	5.15e-02(−)	4.68e-02	8.07e-02(+)
Gupta	3.48e-01	6.72e-01(−)	5.88e-01	3.16e+00(∼)	4.36e-02	1.30e-01(−)	9.55e+01	1.04e+02(−)	2.37e+00	3.18e+00(−)	3.39e-02	9.53e-02(−)	9.57e-01	3.15e-01(−)
Method	**Ackley$_2$(2)**		**Griewank$_2$(2)**		**Ackley$_1$0(10)**		**Griewank$_1$0(10)**		**GSobol (10)**		**Push4 (4)**		**Push8 (8)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	**6.93e-02**	**2.38e-01**	2.05e-01	1.34e-01	**9.87e+00**	**7.65e+00**	9.67e-01	1.06e-01	3.35e+03	3.83e+03	**1.14e-01**	**1.25e-01**	2.40e+00	1.84e+00
qEI	2.54e+00	1.06e+00(−)	2.83e-01	1.32e-01(−)	1.90e+01	5.75e-01(−)	1.04e+02	1.51e+01(−)	1.44e+03	1.42e+03(+)	1.74e-01	1.46e-01(∼)	2.91e+00	1.62e+00(∼)
LP	5.57e+00	2.60e+00(−)	1.15e+00	1.85e+00(−)	1.85e+01	8.80e-01(−)	5.43e+01	2.44e+01(−)	2.27e+03	2.24e+03(∼)	3.80e-01	2.21e-01(−)	3.39e+00	1.80e+00(−)
PLAyBOOK	5.96e+00	2.23e+00(−)	8.60e-01	5.72e-01(−)	1.87e+01	6.44e-01(−)	5.39e+01	1.86e+01(−)	1.89e+03	1.97e+03(∼)	4.43e-01	2.76e-01(−)	3.51e+00	1.98e+00(−)
KB	7.78e-01	2.68e+00(−)	3.12e-01	1.28e-01(−)	1.16e+01	6.79e+00(∼)	1.03e+00	4.35e-02(−)	2.53e+03	2.68e+03(∼)	1.32e-01	8.26e-02(∼)	**1.36e+00**	**1.08e+00(+)**
TS	2.62e-01	1.36e-01(−)	3.97e-01	1.84e-01(−)	1.30e+01	7.69e-01(−)	1.40e+01	3.34e-01(−)	7.66e+03	8.49e+03(∼)	2.68e-01	1.86e-01(−)	3.02e+00	1.67e+00(∼)
ϵS-PF	5.64e+00	2.33e+00(−)	1.64e-01	1.86e-01(∼)	1.82e+01	2.79e+00(−)	8.32e-01	1.02e-01(+)	1.73e+03	2.82e+03(+)	1.29e-01	2.72e-01(∼)	2.54e+00	2.05e+00(∼)
AEGiS	3.88e-01	3.31e-01(−)	3.25e-01	1.12e-01(−)	1.79e+01	2.08e+00(−)	3.26e+00	1.64e+00(−)	3.20e+03	3.71e+03(∼)	1.64e-01	1.01e-01(−)	2.74e+00	1.50e+00(∼)
GIBBON	6.94e-01	1.26e+00(−)	**7.37e-02**	**3.60e-02(+)**	1.91e+01	4.51e-01(−)	8.97e+01	1.78e+01(−)	3.68e+03	3.87e+03(∼)	1.36e-01	7.95e-02(∼)	2.44e+00	1.95e+00(∼)
MACE	4.44e+00	3.42e+00(−)	4.26e+00	7.13e+00(−)	1.79e+01	1.39e+00(−)	**8.20e-01**	**1.57e-01(+)**	2.87e+03	6.71e+03(+)	3.66e-01	7.34e-01(∼)	3.15e+00	2.09e+00(−)
Gupta	3.84e+00	2.68e+00(−)	1.83e+00	4.18e+00(−)	1.90e+01	4.77e-01(−)	8.84e-01	5.89e-02(+)	**2.09e+02**	**3.60e+02(+)**	3.72e-01	3.01e-01(−)	3.76e+00	2.49e+00(−)

Table 6: The performance outcomes of the proposed POEE and nine state-of-the-art methods were evaluated across all test problems using a batch size of $q = 5$ with noise level $\sigma_\epsilon = 0.1$. The method demonstrating the best mean performance is denoted in bold. The symbols "+", "∼", and "−" indicate that the method is statistically better than, equivalent to, and worse than POEE respectively.

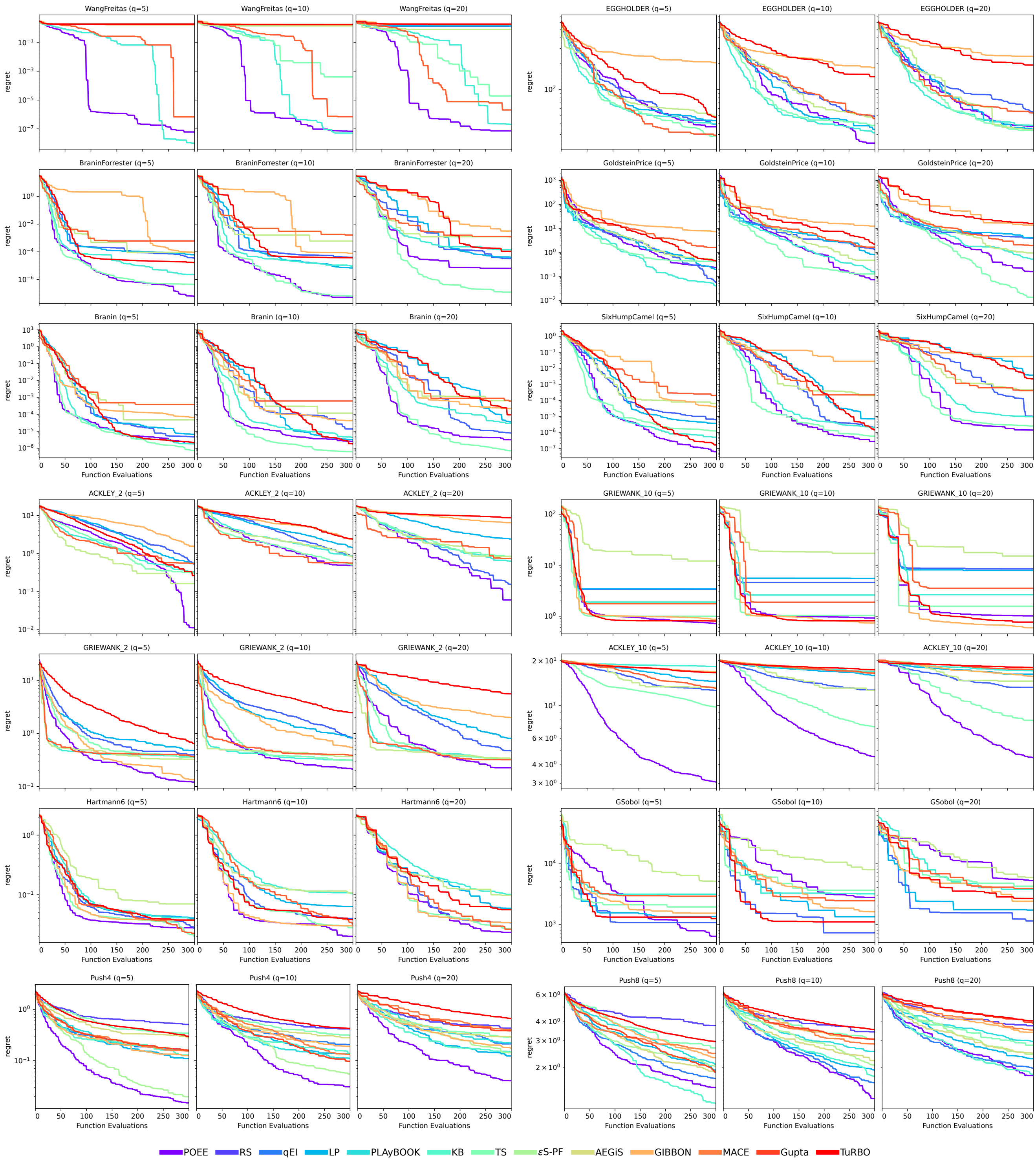

Figure 1: The convergence trajectories of the proposed POEE and the nine counterpart methods throughout the optimisation process on the 14 synthetic and practical problems. Each coloured line illustrates the mean difference between the true optimum and the best function value obtained over 30 independent runs.

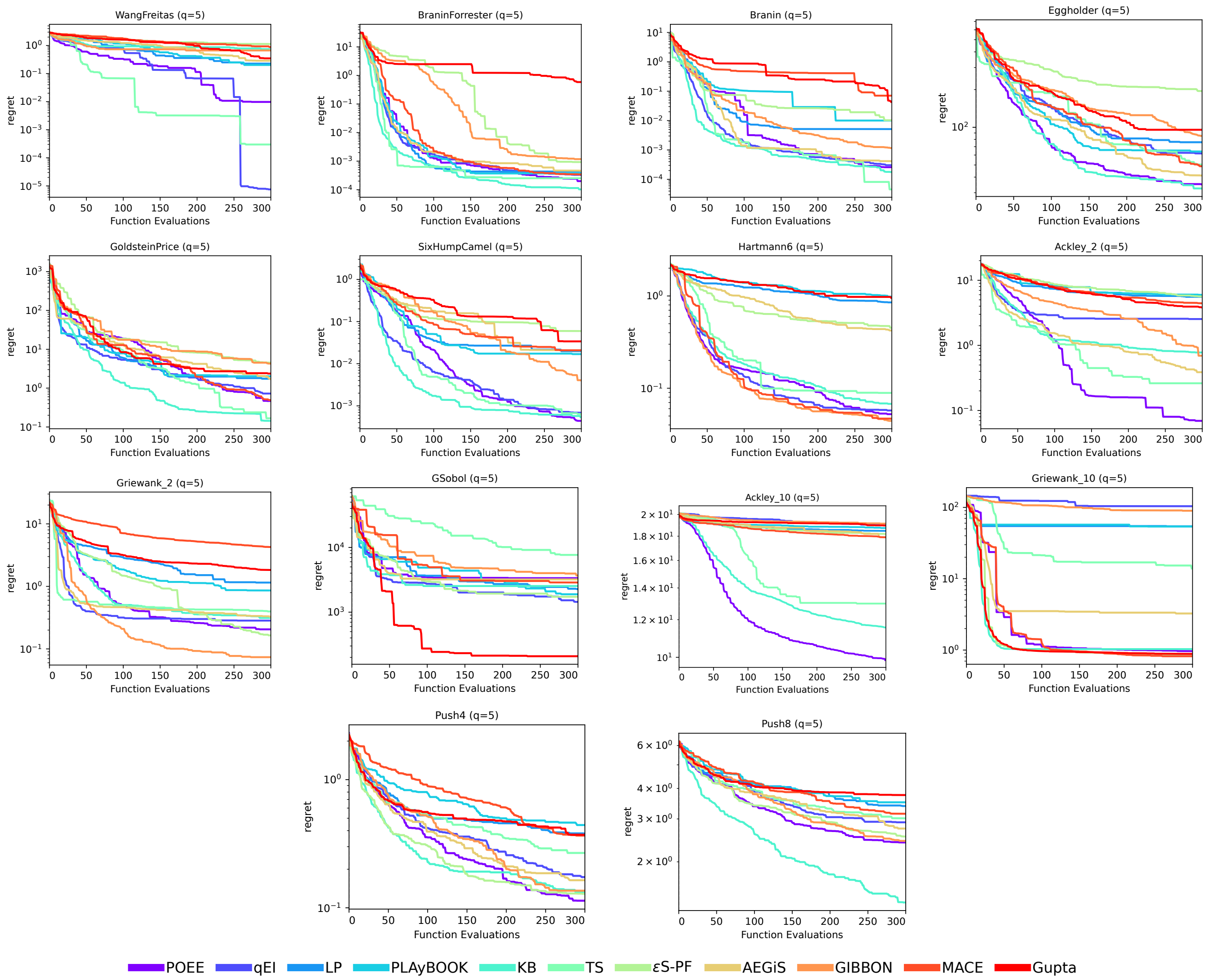

Figure 2: Convergence trajectories of the ten algorithms during the optimisation process for the batch size $q = 5$ with noise level $\sigma_\epsilon = 0.1$, in which each coloured line represents the mean difference between the best value obtained and the problem's real optimum through 30 independent runs (after initial LHS sampling, represented by the dashed grey line).

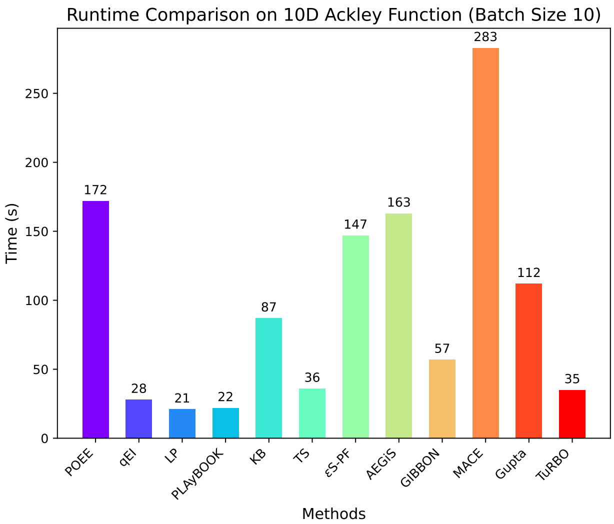

Figure 3: Runtime comparison of all the methods (except random search) for selecting a batch of solutions on the 10D Ackley function (batch size of 10).

Figure 3 gives the runtime comparison of all the methods (except random search) for selecting a batch of solutions on the 10D Ackley function (batch size of 10). As shown, MACE needs the longest runtime of 283 seconds, followed by POEE with a runtime of 172 seconds. LP and PLAyBOOK are the fastest, with runtimes of 21 and 22 seconds, respectively. Notably, the computational cost of POEE, though in general slower than most of the others, is acceptable (compared to the usual expensive evaluation of a solution involved in Bayesian optimisation).

4 Ablation Study

We conducted the ablation study to see how the three important components of the proposed POEE, i.e., 1) the dynamically updated Pareto front, 2) the first selection of the most exploitative solution, and 3) the utilisation of TOPSIS, contribute to the performance.

We considered four variants of the proposed POEE: $\text{POEE}_{\text{r\_PF}}$, $\text{POEE}_{\text{r\_uPF}}$, $\text{POEE}_{\text{no\_TOPSIS}}$, and $\text{POEE}_{\text{no\_exploit}}$. $\text{POEE}_{\text{r\_PF}}$ adopts a static approach, randomly selecting q batch solutions from the unchanged quality-uncertainty Pareto front. $\text{POEE}_{\text{r\_uPF}}$, in contrast, iteratively randomly selects q batch solutions from the dynamically updated Pareto front. $\text{POEE}_{\text{no\_TOPSIS}}$ initially selects the most exploitative solution and then iteratively randomly selects $q-1$ batch solutions from the dynamically updated Pareto front. Finally, $\text{POEE}_{\text{no\_exploit}}$ selects q solutions by TOPSIS from the dynamically updated Pareto front without using the extreme exploitation solution.

Tables 7, 9, and 11 present the results (mean and standard deviation) of the regrets obtained for batch sizes of $q = 5$, $q = 10$, and $q = 20$, respectively. Additionally, the comparative results between POEE and its variants are summarised in Tables 8, 10, and 12, respectively. Figure 4 presents the convergence trajectories of the POEE and its variants across the 14 test problems.

As can be seen from Tables 8, 10, and 12, $\text{POEE}_{\text{r\_PF}}$ performs poorly on all the batch sizes, likely due to random selection from a static Pareto front. These results may stem from the random selection from a static Pareto front. In contrast, $\text{POEE}_{\text{r\_uPF}}$, using a dynamically updated Pareto front, shows improved effectiveness, highlighting the importance of dynamic updates. However, random selection, even from an updated Pareto front, is still not deemed to be a good approach. In contrast, $\text{POEE}_{\text{r\_uPF}}$, which incorporates the dynamically updated Pareto front, demonstrates improved effectiveness. This comparison stresses the significance of the dynamic Pareto front in enhancing the algorithm's adaptability and performance. However, the practice of randomly selecting solutions, even from a dynamically updated Pareto front, cannot still be deemed as a good approach.

$\text{POEE}_{\text{no\_TOPSIS}}$ improves $\text{POEE}_{\text{r\_uPF}}$ by initially targeting the most exploitative solution on the Pareto front, demonstrating the value of a "greedy" approach. However, its reliance on random selection results in lower performance than POEE. $\text{POEE}_{\text{no\_exploit}}$ performs worse than POEE, indicating the importance of incorporating the most exploitative solution, as used in our method. In general, each component of our POEE method plays an important role in striking a good balance between exploitation and exploration.

$\text{POEE}_{\text{no\_TOPSIS}}$ enhances the efficacy of $\text{POEE}_{\text{r\_uPF}}$ by adopting an initial selection approach that targets the extreme exploitation solution on the Pareto front. This methodological refinement illustrates the effectiveness of incorporating a "greedy" approach within the algorithm's framework. However, $\text{POEE}_{\text{no\_TOPSIS}}$'s reliance on random selection from the Pareto front results in accessible performance compared with the POEE. This outcome highlights the limitations inherent in random selection strategies, even when coupled with an initial greedy approach. $\text{POEE}_{\text{no\_exploit}}$ demonstrates inferior performance compared to the original POEE, stressing the importance of incorporating the most exploitative solution as utilised in our method. Overall, each component in our POEE method plays a major role in striking a good balance between exploitation and exploration.

Method	**WangFreitas (1)**		**BraninForrester (2)**		**Branin (2)**		**EGGHOLDER (2)**		**GoldsteinPrice (2)**		**SixHumpCamel (2)**		**Hartmann6 (6)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	6.08e-08	1.23e-07	**5.98e-08**	**9.60e-08**	**1.86e-06**	**1.77e-06**	**3.69e+01**	**2.99e+01**	**2.36e-01**	**8.79e-01**	**6.29e-08**	**5.86e-08**	2.79e-02	5.06e-02
POEE$_{\text{r_PF}}$	6.08e-01	9.12e-01(−)	6.10e-04	8.92e-04(−)	1.07e-03	1.94e-03(−)	1.06e+02	1.03e+02(−)	9.78e-01	8.26e-01(−)	4.08e-03	1.53e-02(−)	4.51e-02	5.77e-02(−)
POEE$_{\text{r_uPF}}$	2.86e-01	6.80e-01(−)	2.46e-03	9.14e-03(−)	1.87e-03	3.92e-03(−)	1.50e+02	1.31e+02(−)	7.07e-01	6.03e-01(−)	1.39e-03	4.36e-03(−)	**2.76e-02**	**4.84e-02(+)**
POEE$_{\text{no_TOPSIS}}$	4.67e-01	8.46e-01(∼)	2.22e-06	2.86e-06(−)	4.80e-06	8.96e-06(∼)	1.40e+02	1.25e+02(−)	5.63e-01	8.92e-01(∼)	9.47e-06	1.73e-05(−)	3.33e-02	5.50e-02(−)
POEE$_{\text{no_exploit}}$	**5.42e-08**	**1.21e-07(∼)**	3.71e-05	1.02e-04(−)	3.88e-05	1.23e-04(∼)	7.20e+01	8.58e+01(−)	1.35e+00	5.85e+00(∼)	1.60e-05	3.37e-05(−)	3.24e-02	5.33e-02(−)
Method	**ACKLEY (2)**		**GRIEWANK (2)**		**ACKLEY (10)**		**GRIEWANK (10)**		**GSobol (10)**		**PUSH4 (4)**		**PUSH8 (8)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	1.12e-02	7.87e-03	**1.21e-01**	**1.16e-01**	**3.06e+00**	**3.27e+00**	**7.19e-01**	**1.93e-01**	**6.35e+02**	**6.21e+02**	**1.50e-02**	**1.87e-02**	1.48e+00	1.81e+00
POEE$_{\text{r_PF}}$	6.05e-02	1.15e-01(−)	3.50e-01	2.11e-01(−)	5.72e+00	5.54e+00(−)	5.05e+00	3.21e+00(−)	1.18e+03	2.86e+03(∼)	5.00e-02	3.61e-02(−)	1.44e+00	1.28e+00(∼)
POEE$_{\text{r_uPF}}$	3.00e-02	6.82e-02(∼)	3.46e-01	2.21e-01(−)	1.04e+01	6.27e+00(−)	2.00e+00	7.31e-01(−)	4.14e+03	5.76e+03(−)	5.26e-02	4.64e-02(−)	1.86e+00	1.95e+00(∼)
POEE$_{\text{no_TOPSIS}}$	**3.72e-03**	**4.19e-03(+)**	2.44e-01	1.31e-01(−)	6.92e+00	7.04e+00(−)	8.39e-01	1.52e-01(−)	4.51e+03	5.99e+03(−)	2.37e-02	3.42e-02(∼)	1.86e+00	1.99e+00(∼)
POEE$_{\text{no_exploit}}$	7.37e-03	6.81e-03(+)	1.79e-01	1.44e-01(−)	1.01e+01	8.50e+00(−)	8.04e-01	1.57e-01(∼)	8.44e+02	8.80e+02(∼)	3.68e-02	7.04e-02(∼)	**1.31e+00**	**1.79e+00(∼)**

Table 7: The performance outcomes of the proposed POEE and its four variants were evaluated across all the test problems using a batch size of $q = 5$. The method demonstrating the best mean performance is highlighted in bold. The symbols "+", "∼", and "−" indicate that the method is statistically better than, equivalent to, and worse than POEE respectively.

	POEE$_{\text{r_PF}}$	POEE$_{\text{r_uPF}}$	POEE$_{\text{no_TOPSIS}}$	POEE$_{\text{no_exploit}}$
POEE	12/ 2/ 0	11/ 2/ 1	8/ 5/ 1	6/ 7/ 1

Table 8: The summary of statistical outcomes, as detailed in Table 7. Here, the left, median, and right numbers denote the counts of test problems where the POEE was statistically superior, equivalent, or inferior to the peer algorithm, respectively.

Method	**WangFreitas (1)**		**BraninForrester (2)**		**Branin (2)**		**EGGHOLDER (2)**		**GoldsteinPrice (2)**		**SixHumpCamel (2)**		**Hartmann6 (6)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	6.81e-08	1.69e-07	**5.15e-08**	**8.14e-08**	**2.54e-06**	**3.06e-06**	**2.38e+01**	**2.85e+01**	**7.31e-02**	**1.31e-01**	**2.74e-07**	**2.99e-07**	**2.00e-02**	**4.45e-02**
POEE$_{\text{r_PF}}$	2.69e-01	6.79e-01(−)	1.18e-02	1.98e-02(−)	3.50e-03	6.16e-03(−)	1.10e+02	1.16e+02(−)	4.36e+00	5.88e+00(−)	1.88e-03	3.86e-03(−)	4.53e-02	5.55e-02(−)
POEE$_{\text{r_uPF}}$	3.33e-01	7.45e-01(−)	3.92e-04	6.98e-04(−)	1.36e-04	2.35e-04(−)	8.34e+01	8.80e+01(−)	3.99e-01	5.45e-01(−)	8.42e-05	2.63e-04(−)	3.01e-02	5.09e-02(−)
POEE$_{\text{no_TOPSIS}}$	3.33e-01	7.45e-01(−)	6.44e-06	1.17e-05(−)	4.52e-06	5.90e-06(∼)	1.21e+02	1.57e+02(−)	1.07e+00	1.51e+00(−)	3.30e-06	7.27e-06(−)	2.44e-02	4.85e-02(−)
POEE$_{\text{no_exploit}}$	**3.82e-08**	**7.69e-08(∼)**	1.83e-05	3.25e-05(−)	1.84e-05	4.17e-05(−)	5.26e+01	5.61e+01(−)	7.75e-01	2.88e+00(−)	6.02e-06	1.40e-05(−)	2.86e-02	5.07e-02(−)
Method	**ACKLEY (2)**		**GRIEWANK (2)**		**ACKLEY (10)**		**GRIEWANK (10)**		**GSobol (10)**		**PUSH4 (4)**		**PUSH8 (8)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	4.85e-01	2.53e+00	**2.13e-01**	**1.28e-01**	**4.53e+00**	**5.06e+00**	**9.23e-01**	**1.10e-01**	2.75e+03	4.32e+03	**3.09e-02**	**3.15e-02**	1.25e+00	1.49e+00
POEE$_{\text{r_PF}}$	8.87e-02	1.03e-01(+)	3.52e-01	2.64e-01(−)	6.35e+00	5.19e+00(−)	6.08e+00	3.11e+00(−)	**3.76e+02**	**7.10e+02(+)**	6.48e-02	6.29e-02(−)	2.13e+00	2.08e+00(−)
POEE$_{\text{r_uPF}}$	4.46e-02	1.68e-01(∼)	2.52e-01	1.90e-01(∼)	7.71e+00	6.67e+00(−)	1.74e+00	4.80e-01(−)	5.37e+03	6.77e+03(−)	5.15e-02	6.37e-02(−)	2.03e+00	1.95e+00(−)
POEE$_{\text{no_TOPSIS}}$	**2.38e-03**	**2.36e-03(+)**	2.78e-01	1.55e-01(∼)	6.36e+00	6.88e+00(∼)	9.79e-01	5.95e-02(−)	8.88e+03	1.21e+04(−)	4.62e-02	5.41e-02(∼)	1.57e+00	1.66e+00(∼)
POEE$_{\text{no_exploit}}$	1.01e-02	9.61e-03(+)	2.56e-01	1.25e-01(∼)	7.81e+00	7.87e+00(∼)	9.51e-01	9.79e-02(∼)	2.59e+03	4.19e+03(∼)	3.36e-02	4.41e-02(∼)	**1.22e+00**	**1.40e+00(∼)**

Table 9: The performance outcomes of the proposed POEE and its four variants were evaluated across all the test problems using a batch size of $q = 10$. The method demonstrating the best mean performance is highlighted in bold. The symbols "+", "∼", and "−" indicate that the method is statistically better than, equivalent to, and worse than POEE respectively.

	POEE$_{\text{r_PF}}$	POEE$_{\text{r_uPF}}$	POEE$_{\text{no_TOPSIS}}$	POEE$_{\text{no_exploit}}$
POEE	12/ 0/ 2	12/ 2/ 0	8/ 5/ 1	6/ 7/ 1

Table 10: The summary of statistical outcomes, as detailed in Table 9. Here, the left, median, and right numbers denote the counts of test problems where the POEE was statistically superior, equivalent, or inferior to the peer algorithm, respectively.

Method	**WangFreitas (1)**		**BraninForrester (2)**		**Branin (2)**		**EGGHOLDER (2)**		**GoldsteinPrice (2)**		**SixHumpCamel (2)**		**Hartmann6 (6)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	7.30e-08	2.56e-07	**6.39e-06**	**1.33e-05**	**3.14e-06**	**5.37e-06**	**3.73e+01**	**2.96e+01**	**1.59e-01**	**3.05e-01**	**1.40e-06**	**2.54e-06**	**2.33e-02**	**5.29e-02**
$\text{POEE}_{\text{r\_PF}}$	4.04e-01	7.98e-01(−)	6.27e-02	1.43e-01(−)	1.54e-02	2.30e-02(−)	1.14e+02	1.34e+02(−)	1.50e+01	1.22e+01(−)	1.53e-03	3.03e-03(−)	5.07e-02	7.35e-02(−)
$\text{POEE}_{\text{r\_uPF}}$	4.00e-01	8.00e-01(−)	5.56e-05	6.10e-05(−)	2.58e-05	3.92e-05(−)	1.15e+02	9.70e+01(−)	1.78e-01	1.79e-01(−)	5.59e-06	1.05e-05(−)	2.54e-02	4.81e-02(∼)
$\text{POEE}_{\text{no\_TOPSIS}}$	4.02e-01	7.99e-01(−)	9.65e-06	1.52e-05(∼)	8.53e-06	1.06e-05(−)	7.28e+01	8.23e+01(∼)	4.71e+00	5.74e+00(−)	1.41e-06	1.36e-06(∼)	2.57e-02	4.80e-02(−)
$\text{POEE}_{\text{no\_exploit}}$	**6.83e-08**	**2.51e-07(∼)**	2.61e-05	3.75e-05(−)	4.88e-05	9.56e-05(−)	6.58e+01	8.39e+01(∼)	2.12e+00	6.41e+00(−)	8.77e-06	1.61e-05(−)	3.31e-02	5.42e-02(∼)

Method	**ACKLEY (2)**		**GRIEWANK (2)**		**ACKLEY (10)**		**GRIEWANK (10)**		**GSobol (10)**		**PUSH4 (4)**		**PUSH8 (8)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	6.03e-02	1.91e-01	**2.26e-01**	**1.22e-01**	**4.47e+00**	**4.86e+00**	**1.02e+00**	**3.58e-02**	5.22e+03	8.46e+03	**4.07e-02**	**5.89e-02**	1.77e+00	1.48e+00
$\text{POEE}_{\text{r\_PF}}$	1.28e+00	3.10e+00(−)	2.56e-01	1.36e-01(∼)	9.50e+00	4.88e+00(−)	6.41e+00	3.30e+00(−)	**2.22e+03**	**7.29e+03(+)**	1.56e-01	2.25e-01(−)	2.23e+00	1.64e+00(∼)
$\text{POEE}_{\text{r\_uPF}}$	1.06e-02	1.30e-02(+)	2.77e-01	1.50e-01(∼)	5.41e+00	4.78e+00(−)	1.70e+00	3.97e-01(−)	5.93e+03	6.74e+03(∼)	4.48e-02	3.20e-02(∼)	3.08e+00	2.19e+00(−)
$\text{POEE}_{\text{no\_TOPSIS}}$	**4.24e-03**	**9.43e-03(+)**	3.08e-01	1.58e-01(−)	5.84e+00	6.19e+00(∼)	1.03e+00	6.58e-02(−)	7.36e+03	9.80e+03(∼)	5.65e-02	6.11e-02(∼)	1.89e+00	1.90e+00(∼)
$\text{POEE}_{\text{no\_exploit}}$	1.39e-02	1.53e-02(+)	2.85e-01	1.53e-01(−)	7.28e+00	7.07e+00(−)	1.03e+00	5.38e-02(∼)	7.55e+03	2.13e+04(∼)	1.19e-01	2.10e-01(−)	**1.66e+00**	**1.73e+00(∼)**

Table 11: The performance outcomes of the proposed POEE and its four variants were evaluated across all the test problems using a batch size of $q = 20$. The method demonstrating the best mean performance is highlighted in bold. The symbols "+", "∼", and "−" indicate that the method is statistically better than, equivalent to, and worse than POEE respectively.

	$\text{POEE}_{\text{r\_PF}}$	$\text{POEE}_{\text{r\_uPF}}$	$\text{POEE}_{\text{no\_TOPSIS}}$	$\text{POEE}_{\text{no\_exploit}}$
POEE	11/ 2/ 1	9/ 4/ 1	6/ 7/ 1	7/ 6/ 1

Table 12: The summary of statistical outcomes, as detailed in Table 11. Here, the left, median, and right numbers denote the counts of test problems where the POEE was statistically superior, equivalent, or inferior to the peer algorithm, respectively.

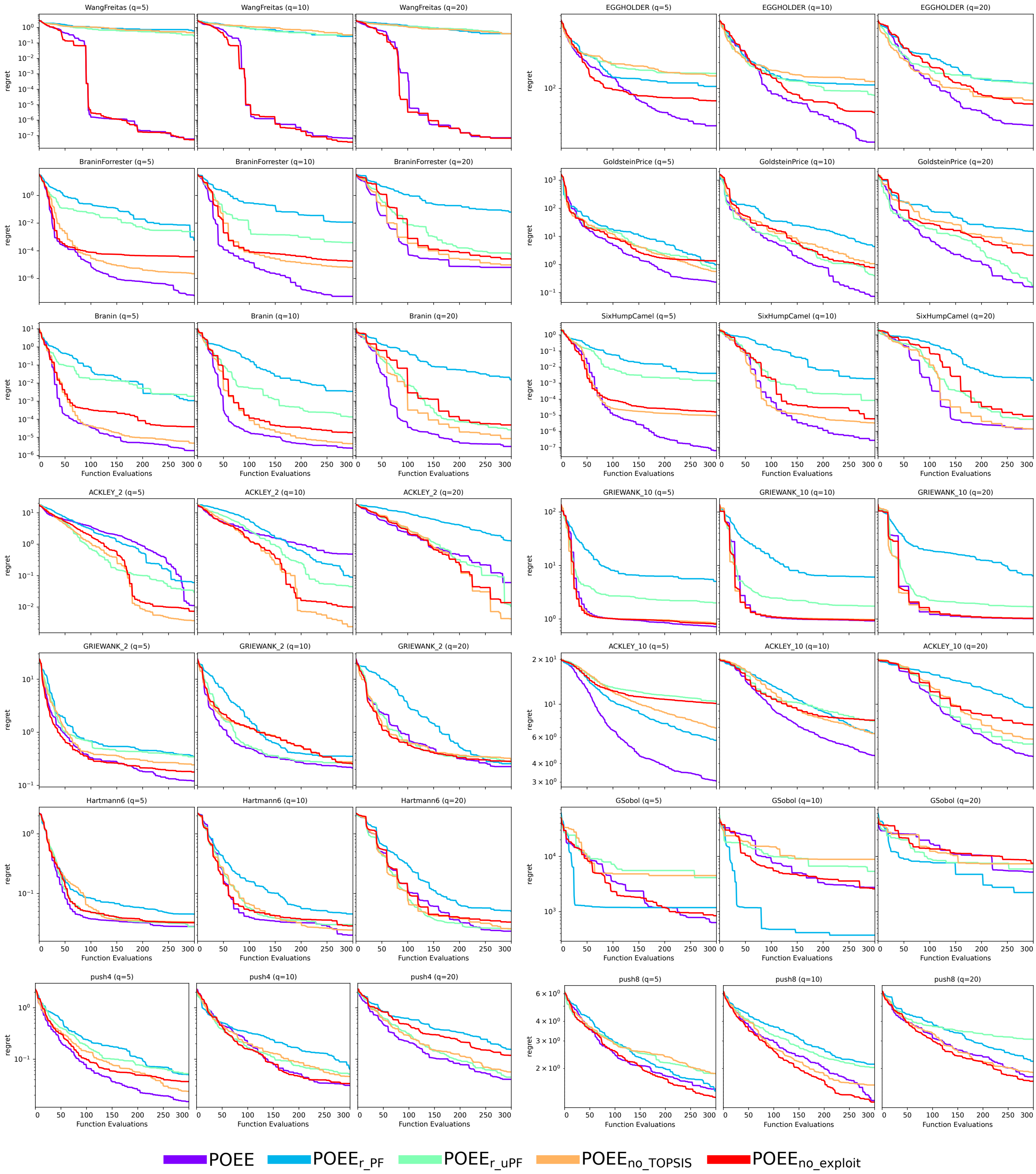

Figure 4: The convergence trajectories of POEE and its four variants throughout the optimisation process on the 14 synthetic and practical problems. Each coloured line illustrates the mean difference between the true optimum and the best function value obtained over 30 independent runs.

5 Sensitivity Analysis

Tables 13, 15, and 17 present the results (mean and standard deviation) of the regrets obtained for batch sizes of $q = 5$, $q = 10$, and $q = 20$, respectively, and those that are statistically equivalent to the proposed POEE according to the Wilcoxon signed-rank test (Sidney, 1957). Additionally, the comparative results between POEE and its variants are summarised in Tables 14, 16, and 18, respectively. Figure 5 presents the convergence trajectories of the five algorithms across the 14 test problems.

Method	WangFreitas (1)		BraninForrester (2)		Branin (2)		EGGHOLDER (2)		GoldsteinPrice (2)		SixHumpCamel (2)		Hartmann6 (6)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	6.08e-08	1.23e-07	**5.98e-08**	**9.60e-08**	**1.86e-06**	**1.77e-06**	**3.69e+01**	**2.99e+01**	2.36e-01	8.79e-01	**6.29e-08**	**5.86e-08**	**2.79e-02**	**5.06e-02**
POEE$_{28}$	**9.02e-09**	**1.42e-08(∼)**	2.94e-05	3.48e-05(−)	3.57e-06	3.84e-06(∼)	4.76e+01	3.58e+01(∼)	7.80e-01	1.01e+00(−)	3.84e-07	6.03e-07(−)	4.85e-02	8.94e-02(−)
POEE$_{55}$	1.27e+00	9.64e-01(−)	1.07e-05	1.71e-05(−)	1.98e-06	2.94e-06(∼)	5.01e+01	5.06e+01(∼)	**4.17e-02**	**8.31e-02(∼)**	6.16e-06	2.77e-05(−)	3.25e-02	5.35e-02(−)
POEE$_{64}$	1.60e+00	8.00e-01(−)	6.79e-05	1.61e-04(−)	5.24e-05	1.21e-04(−)	7.22e+01	8.56e+01(−)	3.27e+00	8.86e+00(∼)	1.88e-05	3.40e-05(−)	3.19e-02	5.27e-02(−)
POEE$_{82}$	1.60e+00	8.00e-01(−)	3.14e+00	9.42e+00(−)	5.11e-05	1.23e-04(−)	1.59e+02	1.17e+02(−)	1.09e+00	4.73e+00(∼)	1.36e-01	3.04e-01(−)	3.21e-02	5.30e-02(−)

Method	ACKLEY (2)		GRIEWANK (2)		ACKLEY (10)		GRIEWANK (10)		GSobol (10)		PUSH4 (4)		PUSH8 (8)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	1.12e-02	7.87e-03	**1.21e-01**	**1.16e-01**	**3.06e+00**	**3.27e+00**	7.19e-01	1.93e-01	6.35e+02	6.21e+02	**1.50e-02**	**1.87e-02**	1.48e+00	1.81e+00
POEE$_{28}$	7.35e-01	1.20e+00(−)	2.01e-01	1.39e-01(−)	1.02e+01	8.29e+00(−)	7.05e-01	1.57e-01(∼)	2.93e+03	3.69e+03(−)	6.84e-02	1.57e-01(∼)	8.36e-01	1.11e+00(∼)
POEE$_{55}$	5.27e-01	1.03e+00(−)	1.42e-01	9.06e-02(∼)	1.07e+01	8.03e+00(−)	**7.00e-01**	**1.86e-01(∼)**	1.07e+03	2.23e+03(∼)	8.76e-02	2.03e-01(−)	**8.25e-01**	**1.13e+00(+)**
POEE$_{64}$	**3.54e-03**	**2.91e-03(+)**	1.41e-01	8.61e-02(∼)	8.56e+00	8.62e+00(∼)	8.61e-01	1.27e-01(−)	8.40e+02	8.82e+02(∼)	2.40e-02	4.83e-02(∼)	1.59e+00	1.84e+00(∼)
POEE$_{82}$	3.02e+00	3.23e+00(−)	1.68e+00	2.13e+00(−)	1.00e+01	8.75e+00(−)	8.59e-01	1.18e-01(−)	**4.61e+02**	**4.58e+02(∼)**	6.58e-02	1.81e-01(∼)	2.53e+00	2.38e+00(−)

Table 13: The performance outcomes of the proposed POEE with different weight combinations were evaluated across all the test problems using a batch size of $q = 5$. The method demonstrating the best mean performance is highlighted in bold. The symbols "+", "∼", and "−" indicate that the method is statistically better than, equivalent to, and worse than POEE respectively.

	POEE$_{28}$	POEE$_{55}$	POEE$_{64}$	POEE$_{82}$
POEE	8/ 6/ 0	7/ 6/ 1	7/ 6/ 1	11/ 3/ 0

Table 14: The summary of statistical outcomes, as detailed in Table 13. Here, the left, median, and right numbers denote the counts of test problems where the POEE was statistically superior, equivalent, or inferior to the peer algorithm, respectively.

Method	WangFreitas (1)		BraninForrester (2)		Branin (2)		EGGHOLDER (2)		GoldsteinPrice (2)		SixHumpCamel (2)		Hartmann6 (6)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	6.81e-08	1.69e-07	**5.15e-08**	**8.14e-08**	**2.54e-06**	**3.06e-06**	**2.38e+01**	**2.85e+01**	7.31e-02	1.31e-01	**2.74e-07**	**2.99e-07**	**2.00e-02**	**4.45e-02**
POEE$_{28}$	**4.80e-08**	**7.73e-08(∼)**	2.22e-05	3.49e-05(−)	1.59e-05	3.89e-05(−)	3.59e+01	3.14e+01(−)	7.57e-01	9.94e-01(−)	1.35e-06	2.08e-06(−)	7.76e-02	1.10e-01(−)
POEE$_{55}$	6.67e-01	9.43e-01(−)	1.24e-05	3.05e-05(−)	5.78e-06	6.79e-06(−)	3.91e+01	3.72e+01(∼)	**6.92e-02**	**1.36e-01(∼)**	2.60e-06	4.71e-06(−)	2.91e-02	5.12e-02(−)
POEE$_{64}$	1.47e+00	8.84e-01(−)	1.82e-05	2.41e-05(−)	1.76e-05	4.18e-05(∼)	8.83e+01	1.22e+02(−)	9.33e-01	2.95e+00(−)	7.27e-06	1.47e-05(−)	3.19e-02	5.28e-02(−)
POEE$_{82}$	1.47e+00	8.84e-01(−)	3.14e+00	9.42e+00(−)	5.40e-05	1.43e-04(−)	1.46e+02	1.38e+02(−)	1.38e+00	5.62e+00(∼)	5.44e-02	2.04e-01(−)	3.19e-02	5.27e-02(−)

Method	ACKLEY (2)		GRIEWANK (2)		ACKLEY (10)		GRIEWANK (10)		GSobol (10)		PUSH4 (4)		PUSH8 (8)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	4.85e-01	2.53e+00	2.13e-01	1.28e-01	**4.53e+00**	**5.06e+00**	9.23e-01	1.10e-01	2.75e+03	4.32e+03	**3.09e-02**	**3.15e-02**	1.25e+00	1.49e+00
POEE$_{28}$	4.88e-01	1.04e+00(∼)	3.36e-01	2.02e-01(∼)	7.83e+00	6.95e+00(−)	**8.74e-01**	**1.77e-01(∼)**	5.92e+03	1.23e+04(∼)	1.00e-01	1.25e-01(−)	**8.18e-01**	**1.11e+00(∼)**
POEE$_{55}$	**3.88e-01**	**8.66e-01(∼)**	**2.09e-01**	**2.08e-01(∼)**	7.71e+00	7.19e+00(−)	9.34e-01	1.03e-01(∼)	1.54e+03	2.51e+03(∼)	5.71e-02	1.03e-01(∼)	1.36e+00	1.73e+00(∼)
POEE$_{64}$	6.49e-01	1.99e+00(−)	5.41e-01	6.82e-01(−)	7.32e+00	8.25e+00(∼)	9.55e-01	1.06e-01(∼)	2.29e+03	4.02e+03(∼)	8.21e-02	2.97e-01(∼)	1.29e+00	1.16e+00(∼)
POEE$_{82}$	3.42e+00	3.95e+00(−)	3.89e+00	3.57e+00(−)	8.60e+00	8.43e+00(∼)	9.63e-01	8.59e-02(−)	**3.63e+02**	**4.58e+02(+)**	1.47e-01	3.79e-01(∼)	2.48e+00	1.81e+00(−)

Table 15: The performance outcomes of the proposed POEE with different weight combinations were evaluated across all the test problems using a batch size of $q = 10$. The method demonstrating the best mean performance is highlighted in bold. The symbols "+", "∼", and "−" indicate that the method is statistically better than, equivalent to, and worse than POEE respectively.

	POEE$_{28}$	POEE$_{55}$	POEE$_{64}$	POEE$_{82}$
POEE	9/ 5/ 0	6/ 8/ 0	8/ 6/ 0	10/ 3/ 1

Table 16: The summary of statistical outcomes, as detailed in Table 15. Here, the left, median, and right numbers denote the counts of test problems where the POEE was statistically superior, equivalent, or inferior to the peer algorithm, respectively.

Method	**WangFreitas (1)**		**BraninForrester (2)**		**Branin (2)**		**EGGHOLDER (2)**		**GoldsteinPrice (2)**		**SixHumpCamel (2)**		**Hartmann6 (6)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	**7.30e-08**	**2.56e-07**	**6.39e-06**	**1.33e-05**	**3.14e-06**	**5.37e-06**	**3.73e+01**	**2.96e+01**	**1.59e-01**	**3.05e-01**	**1.40e-06**	**2.54e-06**	**2.33e-02**	**5.29e-02**
POEE$_{28}$	8.00e-08	1.23e-07$(-)$	2.58e-05	2.35e-05$(-)$	6.90e-05	8.66e-05$(-)$	4.27e+01	3.72e+01$(\sim)$	2.88e+00	2.40e+00$(-)$	3.66e-06	4.30e-06$(-)$	1.31e-01	1.45e-01$(-)$
POEE$_{55}$	4.67e-01	8.46e-01$(-)$	1.96e-05	3.47e-05$(-)$	2.34e-05	3.47e-05$(-)$	4.47e+01	6.78e+01$(\sim)$	2.28e-01	4.42e-01$(\sim)$	5.08e-06	5.80e-06$(-)$	3.49e-02	5.69e-02$(-)$
POEE$_{64}$	6.00e-01	9.16e-01$(-)$	4.91e-05	8.26e-05$(-)$	4.05e-05	9.30e-05$(-)$	9.30e+01	1.29e+02$(-)$	3.58e+00	8.09e+00$(-)$	2.72e-02	1.47e-01$(-)$	3.26e-02	5.38e-02$(\sim)$
POEE$_{82}$	1.27e+00	9.64e-01$(-)$	2.09e+00	7.84e+00$(-)$	3.52e-05	5.05e-05$(-)$	1.79e+02	1.89e+02$(-)$	2.88e+00	6.41e+00$(-)$	5.44e-02	2.04e-01$(-)$	3.21e-02	5.30e-02$(-)$

Method	**ACKLEY (2)**		**GRIEWANK (2)**		**ACKLEY (10)**		**GRIEWANK (10)**		**GSobol (10)**		**PUSH4 (4)**		**PUSH8 (8)**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
POEE	**6.03e-02**	**1.91e-01**	**2.26e-01**	**1.22e-01**	**4.47e+00**	**4.86e+00**	1.02e+00	3.58e-02	5.22e+03	8.46e+03	**4.07e-02**	**5.89e-02**	1.77e+00	1.48e+00
POEE$_{28}$	1.85e+00	3.15e+00$(-)$	6.87e-01	1.21e+00$(-)$	8.33e+00	6.51e+00$(-)$	**1.01e+00**	**5.41e-02$(\sim)$**	6.64e+03	1.31e+04$(\sim)$	1.36e-01	1.15e-01$(-)$	**1.10e+00**	**1.27e+00$(+)$**
POEE$_{55}$	1.47e+00	3.93e+00$(-)$	6.33e-01	1.31e+00$(\sim)$	8.62e+00	6.94e+00$(-)$	1.01e+00	4.59e-02$(\sim)$	1.82e+03	3.25e+03$(\sim)$	1.43e-01	2.21e-01$(-)$	1.31e+00	1.06e+00$(\sim)$
POEE$_{64}$	3.36e+00	6.35e+00$(\sim)$	1.51e+00	1.94e+00$(-)$	5.61e+00	6.61e+00$(\sim)$	1.02e+00	8.22e-02$(\sim)$	9.43e+03	2.26e+04$(\sim)$	5.09e-02	6.04e-02$(\sim)$	1.90e+00	1.86e+00$(\sim)$
POEE$_{82}$	4.09e+00	5.17e+00$(-)$	7.60e+00	7.05e+00$(-)$	9.07e+00	7.64e+00$(-)$	1.03e+00	3.72e-02$(\sim)$	**5.99e+02**	**7.80e+02$(+)$**	3.09e-01	5.83e-01$(\sim)$	2.70e+00	1.58e+00$(-)$

Table 17: The performance outcomes of the proposed POEE with different weight combinations were evaluated across all the test problems using a batch size of $q = 20$. The method demonstrating the best mean performance is highlighted in bold. The symbols "$+$", "$\sim$", and "$-$" indicate that the method is statistically better than, equivalent to, and worse than POEE respectively.

	POEE$_{28}$	POEE$_{55}$	POEE$_{64}$	POEE$_{82}$
POEE	10/ 3/ 1	8/ 6/ 0	7/ 7/ 0	11/ 2/ 1

Table 18: The summary of statistical outcomes, as detailed in Table 17. Here, the left, median, and right numbers denote the counts of test problems where the POEE was statistically superior, equivalent, or inferior to the peer algorithm, respectively.

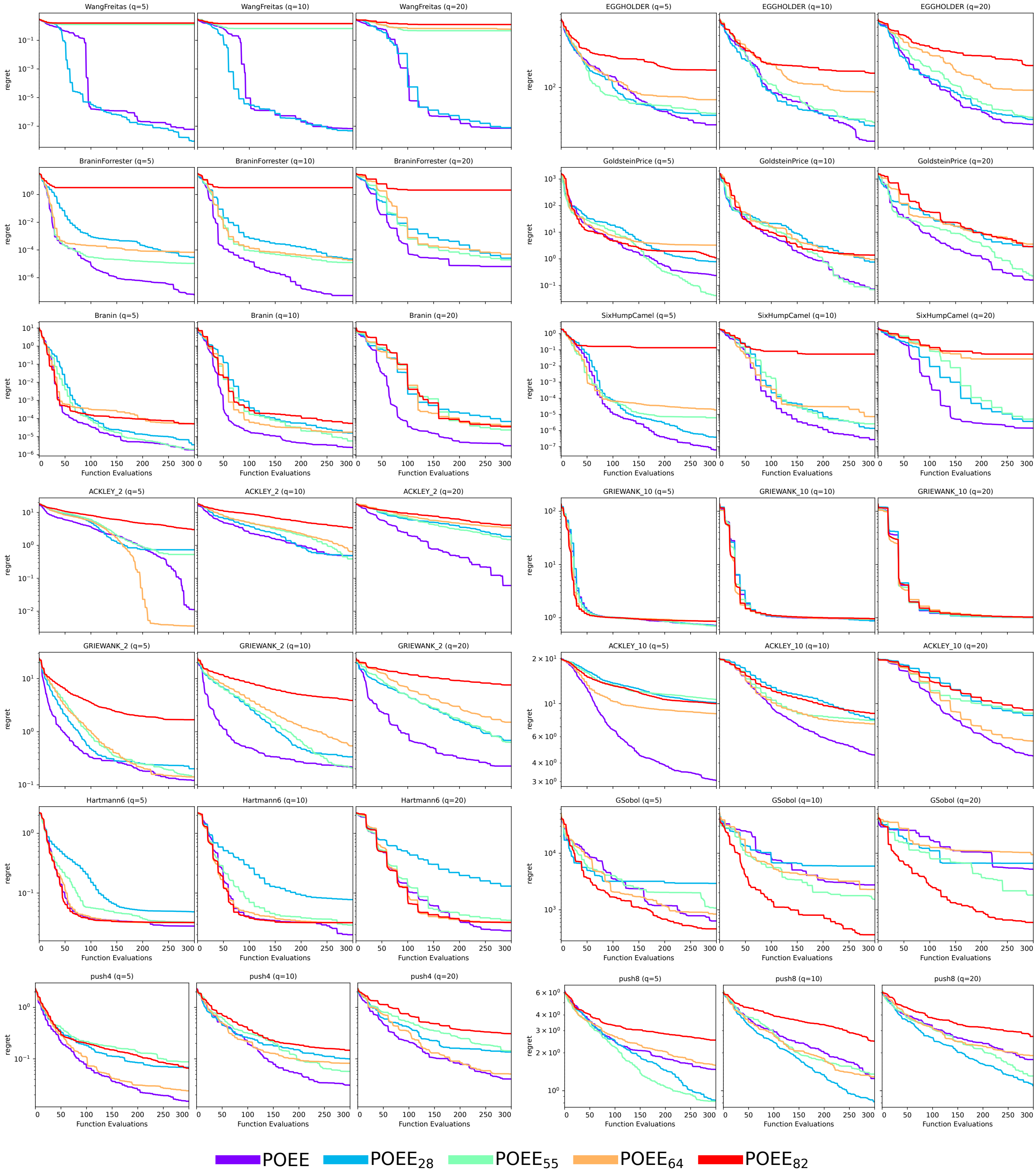

Figure 5: The convergence trajectories of POEE with different weight combinations throughout the optimisation process on the 14 synthetic and practical problems. Each coloured line illustrates the mean difference between the true optimum and the best function value obtained over 30 independent runs.

References

Cinelli, M., Kadziński, M., Gonzalez, M., and Słowiński, R. (2020). How to support the application of multiple criteria decision analysis? let us start with a comprehensive taxonomy. *Omega*, 96:102261.

Cinelli, M., Kadziński, M., Miebs, G., Gonzalez, M., and Słowiński, R. (2022). Recommending multiple criteria decision analysis methods with a new taxonomy-based decision support system. *European Journal of Operational Research*, 302(2):633–651.

González, J., Dai, Z., Hennig, P., and Lawrence, N. (2016). Batch Bayesian optimization via local penalization. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51, pages 648–657. PMLR.

Gu, X., Li, M., Shen, L., Tang, G., Ni, Q., Peng, T., and Shen, Q. (2023). Multi-objective evolutionary optimisation for prototype-based fuzzy classifiers. *IEEE Transactions on Fuzzy Systems*, 31(5):1703–1715.

Hwang, C.-L. and Yoon, K. (1981). *Multiple attribute decision making and applications*. Springer.

Lai, Y.-J., Liu, T.-Y., and Hwang, C.-L. (1994). TOPSIS for MODM. *European Journal of Operational Research*, 76(3):486–500.

Opricovic, S. and Tzeng, G.-H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research*, 156(2):445–455.

Petropoulos, F., Laporte, G., Aktas, E., Alumur, S. A., Archetti, C., Ayhan, H., Battarra, M., Bennell, J. A., Bourjolly, J.-M., Boylan, J. E., et al. (2024). Operational research: methods and applications. *Journal of the Operational Research Society*, 75(3):423–617.

Sidney, S. (1957). Nonparametric statistics for the behavioral sciences. *The Journal of Nervous and Mental Disease*, 125(3):497.

Sobester, A., Forrester, A., and Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons.

Wang, Z. and de Freitas, N. (2014). Theoretical analysis of Bayesian optimisation with unknown Gaussian process hyper-parameters. *arXiv preprint arXiv:1406.7758*.

Wang, Z. and Jegelka, S. (2017). Max-value entropy search for efficient Bayesian optimization. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3627–3635. PMLR.