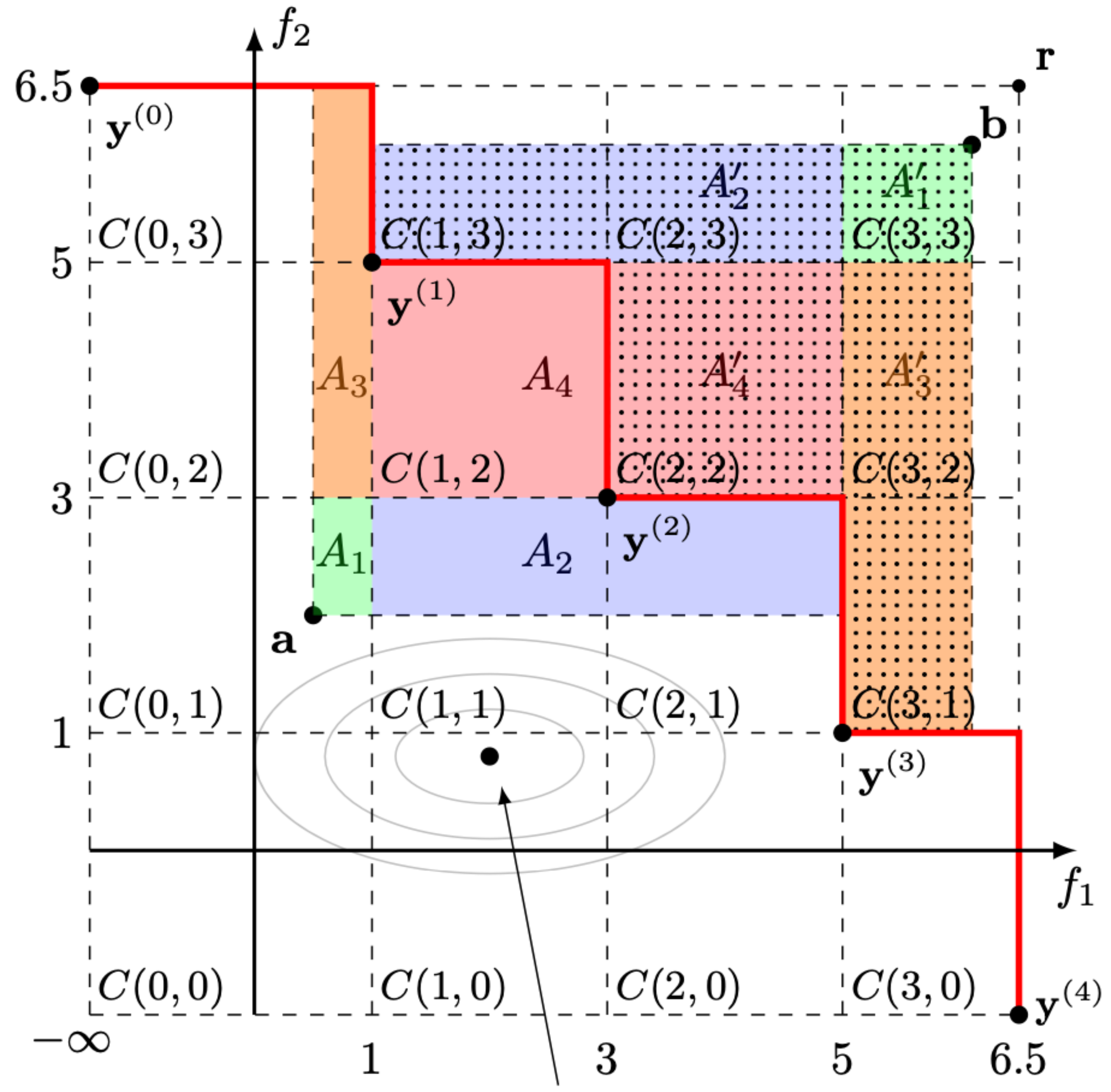
Probability Distribution of Hypervolume Improvement in Bi-objective Bayesian Optimization

Hao Wang[1], Kaifeng Yang[2], Michael Affenzeller[2]

[1] Leiden University, LIACS, Leiden, The Netherlands
[2] University of Applied Sciences Upper Austria, Hagenberg, Austria
kaifeng.yang@fh-hagenberg.at

Contributions:

1. We derive the exact expression of *the probability distribution of hypervolume improvement* under Gaussian randomness;
2. We propose a novel acquisition function for multi-objective Bayesian optimization: *ε-Probability of Hypervolume Improvement* (ε-PoHVI);
3. We experimentally validate the benefit of ε-PoHVI over other acquisition functions.



$$\mathbf{y} \mid \mathcal{D}, \mathbf{x} \sim \mathcal{N}(\hat{\mathbf{f}}(\mathbf{x}), \mathrm{diag}(\hat{k}_1(\mathbf{x},\mathbf{x}), \hat{k}_2(\mathbf{x},\mathbf{x})))$$

Problem formulation:

- Multi-objective optimization – minimizing

$$\mathbf{f} = (f_1, \ldots, f_m), f_i \colon \mathcal{X} \to \mathbb{R}, i \in [1..m]$$

- Given a data set of the functions $\mathrm{X} = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\}, \mathrm{Y} = \{\mathbf{f}(\mathbf{x}^1), \ldots, \mathbf{f}(\mathbf{x}^n)\}$

we have the predictive distribution of the function value via Gaussian process modeling:

$$\mathbf{f}(\mathbf{x}) \mid \mathrm{X}, \mathrm{Y} \sim \mathcal{N}(\hat{\mathbf{f}}(\mathbf{x}), \mathrm{diag}(\hat{k}_1(\mathbf{x},\mathbf{x}), \hat{k}_2(\mathbf{x},\mathbf{x})))$$

- Let $\mathbf{y} = \mathbf{f}(\mathbf{x})$, the hypervolume improve of $\mathbf{y}$ to a Pareto approximation set $\mathcal{P}$ is a random variable:

$$\Delta(\mathbf{y}; \mathcal{P}, \mathbf{r}) = \mathrm{HV}(\mathcal{P} \cup \{\mathbf{y}\}, \mathbf{r}) - \mathrm{HV}(\mathcal{P}, \mathbf{r})$$

- We aim to obtain the exact distribution of HVI.

Our approach:

1. Partition the objective space into quadratically number cells w.r.t. the points in $\mathcal{P}$
2. Conditioning $\mathbf{y}$ on a cell, the HVI function can always be expressed as:

$$\Delta(\mathbf{y})|_{C(i,j)} = \lambda(A_1) + \lambda(A_2) + \lambda(A_3) + \lambda(A_4).$$

3. Expressing each above term allows us to obtain the CDF of HVI conditioning on a cell:

$$\mathrm{CDF}^{(i,j)}_{\Delta^+(\mathbf{y})}(\delta) = \int_{L_1 L_2}^{p} \mathrm{PDF}_{z_1 z_2}(x)\,\mathrm{d}x$$
$$= D_1 D_2 \left(G + \int_{\alpha(p)}^{\beta(p)} \phi_{\mu'_1, \sigma_1}(\zeta) \Phi_{\mu'_2, \sigma_2}\left(\frac{p}{x}\right) \mathrm{d}x \right)$$

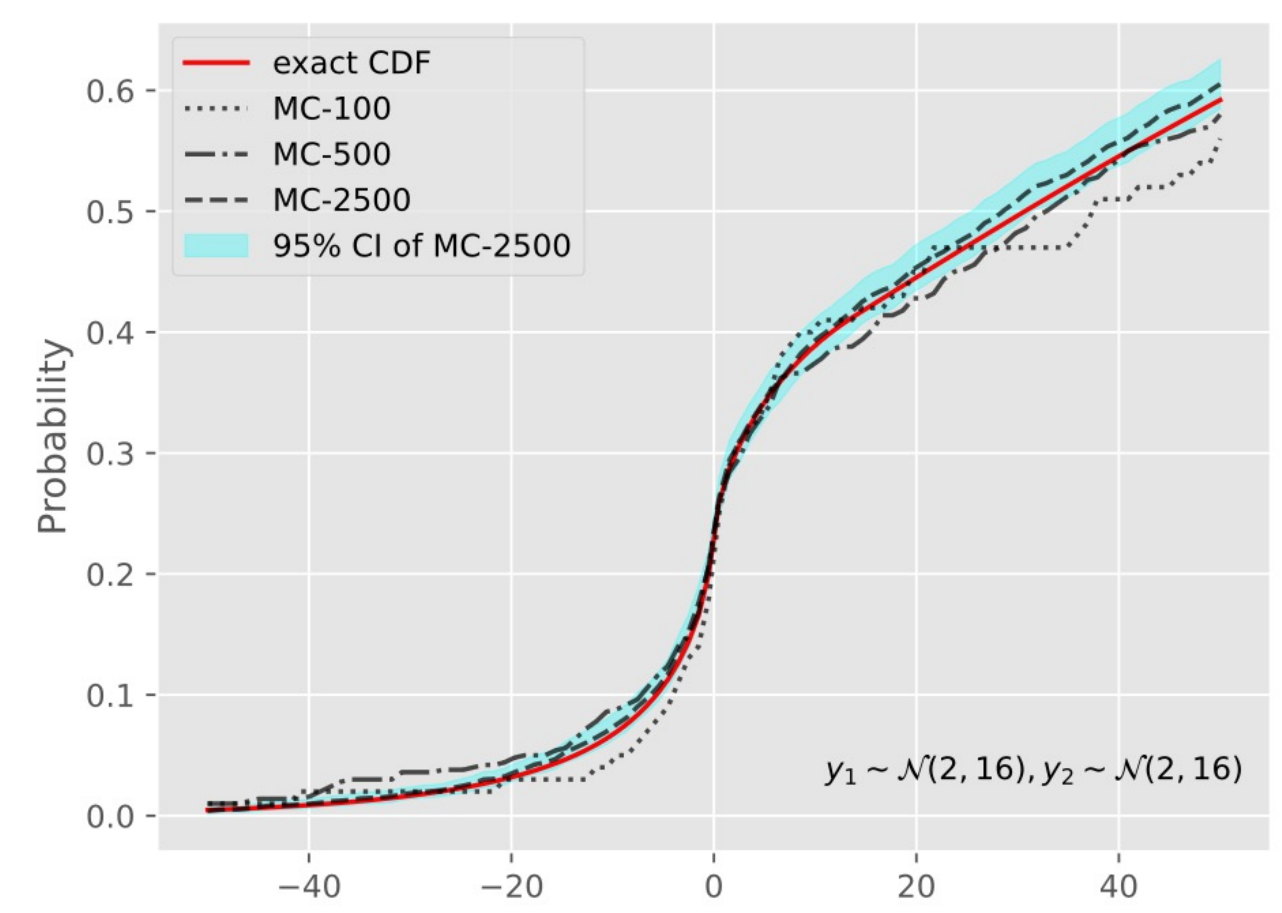
4. PDF of HVI can be obtained by marginalizing the conditional PDF over all cells.

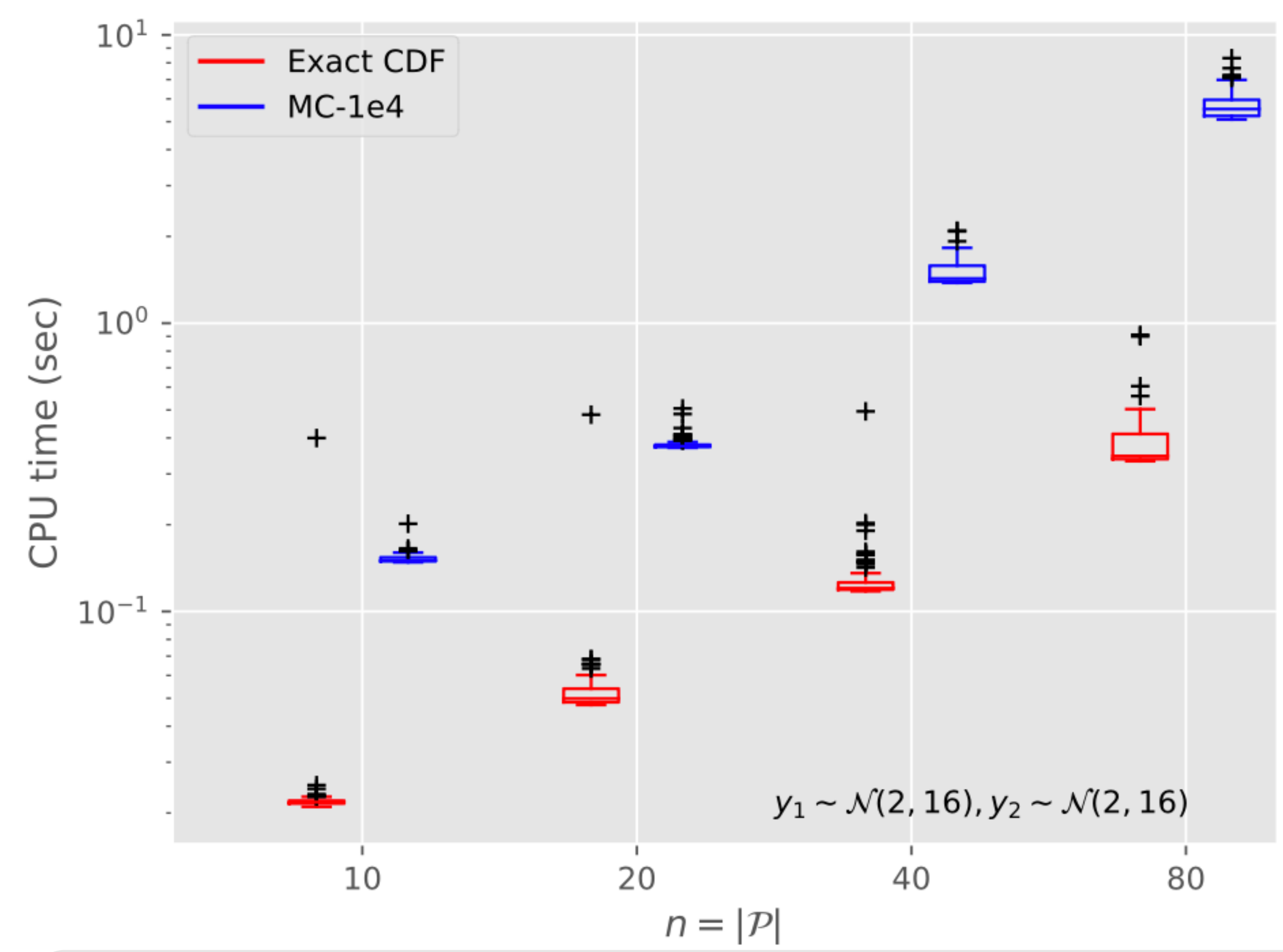
Novel acquisition function: ε-Probability of Hypervolume Improvement

The probability that a point $\mathbf{x} \in \mathcal{X}$ gives rise to at least ε hypervolume improvement:

$$\varepsilon\text{-PoHVI}(\mathbf{x}; \mathcal{P}, \mathbf{r}, \varepsilon) = 1 - \mathrm{CDF}_{\Delta(\mathbf{y})|\mathcal{D},\mathbf{x}}(\varepsilon)$$

Numerical Comparison to Monte Carlo method





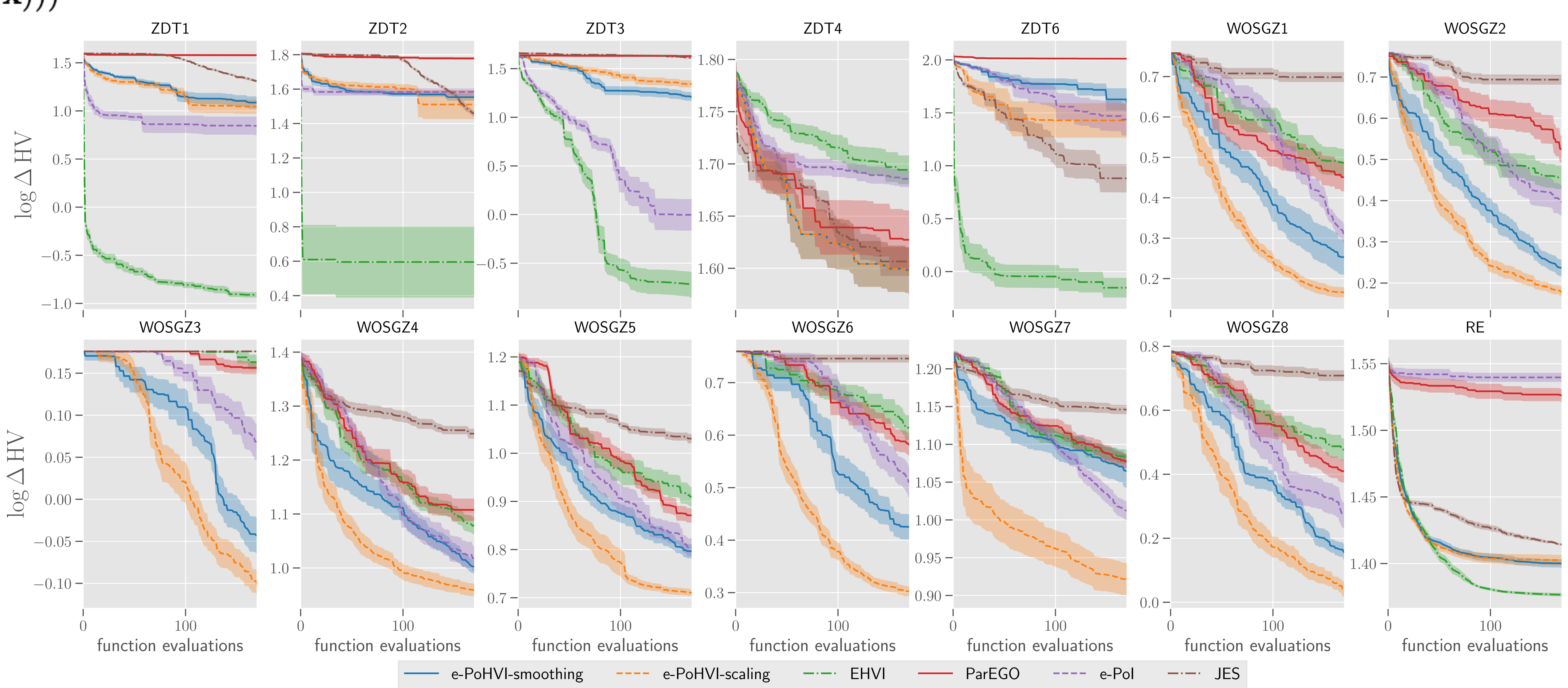
Experimental Results

- ε-PoHVI compared to expected hypervolume improvement (EHVI), ParEGO, ε-PoI, and Joint Entropy Search (JES)
- Benchmark functions: ZDT1-4, ZDT6, WOSGZ1-8, and a real-world problem (RE)
- 15 repetitions for each method; 170 iterations of BO
- Initial DoE of size
- Two variants of ε-PoHVI:
  - ε-PoHVI-scaling:

    $\varepsilon_t = \varepsilon_0 \exp(-ct), \varepsilon_0 = 0.05, c = 0.02$
  - ε-PoHVI-smoothing:

    $\varepsilon_t = \alpha(\mathrm{HV}(\mathcal{P}_t, \mathbf{r}) - \mathrm{HV}(\mathcal{P}_{t-1}, \mathbf{r})) + (1-\alpha)\varepsilon_{t-1}$

    $\alpha = 0.5, \varepsilon_0 = 0.05$



Key Takeaways:

- The exact expression of HVI's distribution is *computationally much faster* than the MC method;
- The exact expression is *much more accurate* than the MC method;
- HVI's distribution *is beneficial when the Gaussian process model has high uncertainty**: ε-PoHVI function greatly improves the empirical convergence speed on WOSGZs
- *ZDT functions are less complicated than WOSGZs, and the Gaussian process model has a larger uncertainty/variance in the latter case;

Future work:

- The exact expression of HVI's distribution for more than two-objectives
- Also for correlated Gaussian distributions
- Use the distribution function for parallelization