Amortized Multi-Objective Optimization Across Tasks with Generative Solution Modeling

WEI TINGYANG
College of Computing and Data Science,
Nanyang Technological University

Supervisor: Prof. Ong Yew-Soon
Co-supervisor: Dr. Tan Puay Siew, SIMTech, A*STAR.

2026.06.06

Research Goal

- Few Shot Optimization
  - $argmin_{x \in \mathcal{X}} f(x), \boxed{C_{max} = C}$

“Few Shot” means budget C can be really small.

- How?

Multi-Task Optimization[1]

\- $argmin_{x_m \in \mathcal{X}_m} f_m(x_m), \boxed{m \in [M]}$

Solve M optimization tasks simultaneously

Why? ***Assuming*** *M problems share certain similarities, we use their* ***solution distributions and surrogate models*** *as transferable knowledge to accelerate convergence.*

Parametric Multi-Task Optimization[2]

\- $argmin_{x \in \mathcal{X}} f(x, \theta), \boxed{\forall \theta \in \Theta}$

Each θ characterizes a task, so the space Θ specifies an infinite number of tasks

We assume access to ***an external task space $\boldsymbol{\Theta}$*** *and investigate how leveraging* ***task parameters θ as side information（额外的信息）*** *can enhance knowledge transfer and accelerate optimization convergence.*

[1]. Tingyang Wei, Jiao Liu, Abhishek Gupta, Puay Siew Tan, Yew-Soon Ong, “Bayesian Forward-Inverse Transfer for Multiobjective Optimization” International Conference on Parallel Problem Solving from Nature 2024.
[2]. Tingyang Wei, Jiao Liu, Abhishek Gupta, Puay Siew Tan, Yew-Soon Ong, “(θl, θu)-Parametric Multi-Task Optimization: Joint Search in Solution and Infinite Task Spaces” IEEE Transactions on Evolutionary Computation (2025).

Why Parametric Multi-Task Optimization?

- $argmin_{x \in \mathcal{X}} f(x, \theta), \forall \theta \in \Theta$

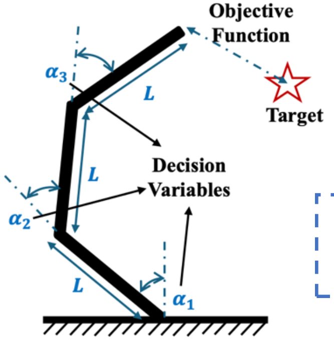
Opportunities:

- Explicit task parameters θ provide a structured way to quantify task similarity, contributing to faster optimization convergence and more efficient knowledge transfer.

Application Example:

- In robotic control, θ can represent different morphologies. Optimizing control policy x for diverse θ allows robots to unexpected environments.

  - Similar θ entails similar x
  - Backing up x-s for diverse θ can enable adaptation



$x: (\alpha_1, \alpha_2, \alpha_3)$

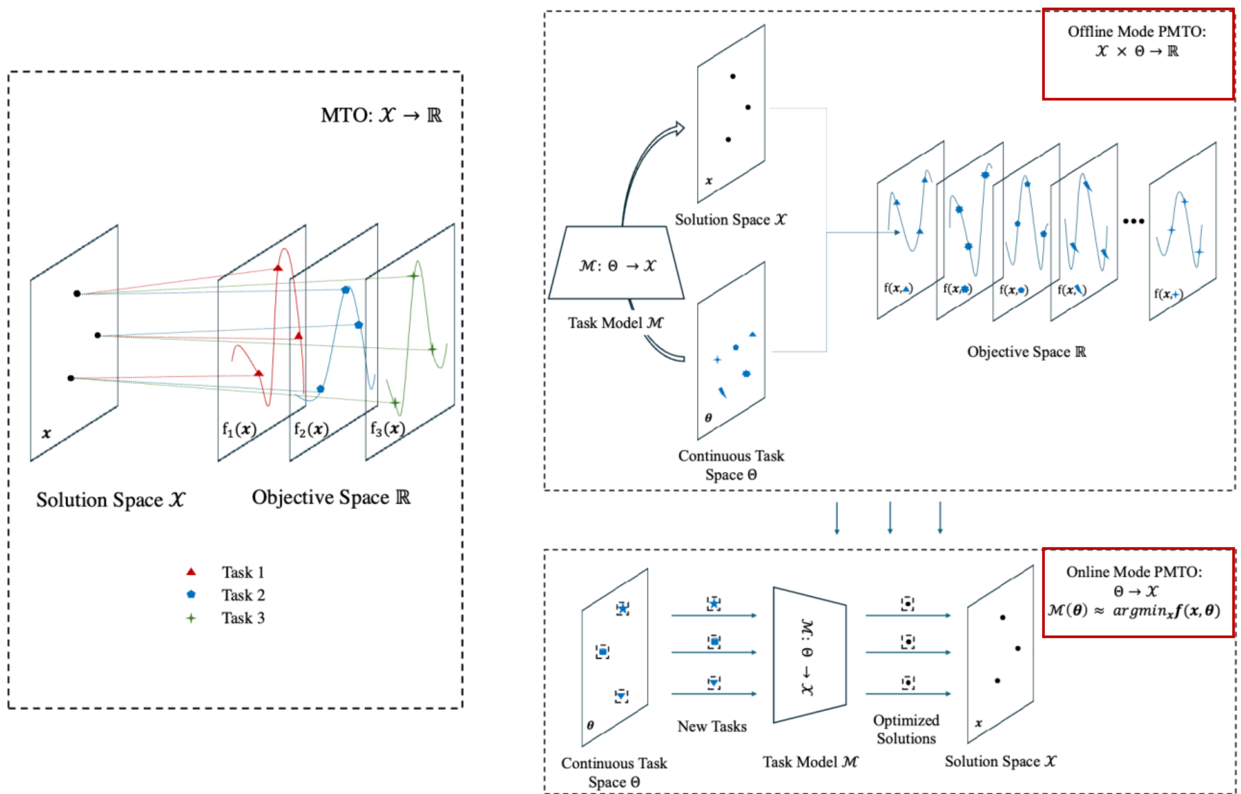
θ: (L, α_{max})

$$\forall \theta \in \Theta,\quad \mathcal{M}(\theta) = argmin_{x \in \mathcal{X}} f(x, \theta)$$

Parametric Multi-Task Optimization (Infinite Tasks)

$(\boldsymbol{\theta}_l, \boldsymbol{\theta}_u)$-Parametric Multi-Task Optimization: Joint Search in Solution and Infinite Task Spaces

Tingyang Wei, Jiao Liu, Abhishek Gupta, *Senior Member, IEEE*, Puay Siew Tan, and Yew-Soon Ong, *Fellow, IEEE*



(a) Multi-task Optimization

(b) Parametric Multi-task Optimization

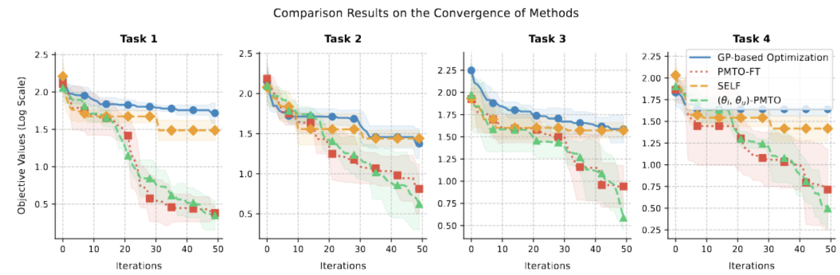
Tingyang Wei, Jiao Liu, Abhishek Gupta, Puay Siew Tan, Yew-Soon Ong, “(θl, θu)-Parametric Multi-Task Optimization: Joint Search in Solution and Infinite Task Spaces” IEEE Transactions on Evolutionary Computation (2025).

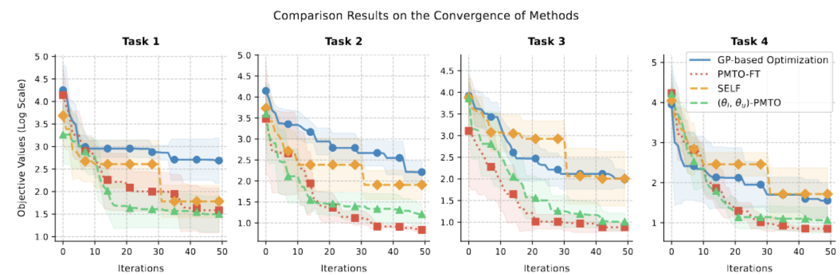
Parametric Multi-Task Optimization (Infinite Tasks)

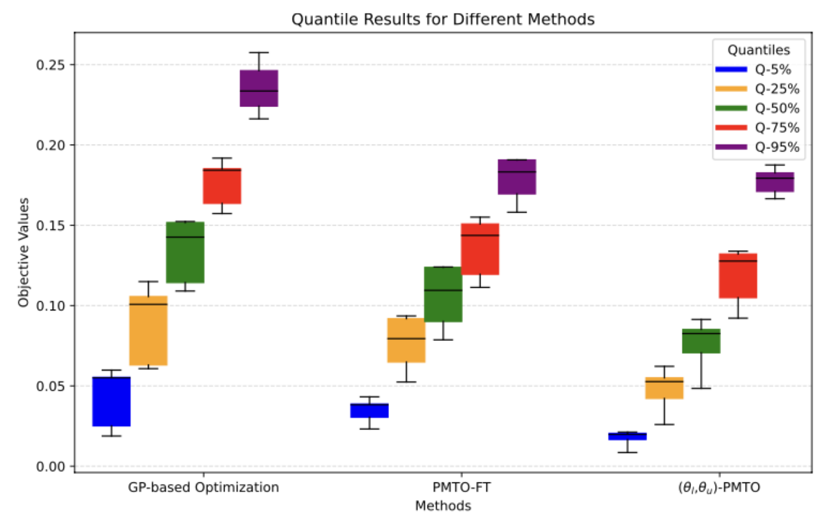
- $argmin_{x \in \mathcal{X}} f(x, \theta), \forall \theta \in \Theta$
- Good Optimization Results (Offline)
  - Sample multiple tasks
  - Check out the convergence
    - $f(\cdot, \theta_m), m \in [M]$
  - Theoretical Analysis of faster convergence

Theorem. *Let the kernel functions used in PMTO-FT and the independent strategy satisfy* $\kappa_{pmt}((\mathbf{x}, \boldsymbol{\theta}), (\mathbf{x}', \boldsymbol{\theta})) = \kappa_{ind}(\mathbf{x}, \mathbf{x}')$. *Then, the MIG in the independent strategy, denoted as* $\gamma_{T,m}^{ind}$, *and the MIG in the PMTO-FT, denoted as* $\gamma_{T,m}$, *satisfy* $\gamma_{T,m} \leq \gamma_{T,m}^{ind}, \forall m \in [M], \forall T \geq 1.$

- $\forall \theta \in \Theta,\ \mathcal{M}(\theta) = argmin_{x \in \mathcal{X}} f(x, \theta)$
- Good Inverse Model Predictive Results (Online)
  - Sample multiple (seen and unseen) tasks θ
  - Check out the solution quality of the model output
    - $f(x', \theta),\ x' = \mathcal{M}(\theta)$







Research Goal

- Few Shot Multi-Objective Optimization
  - $argmin_{x\in\mathcal{X}}\ F(x) = argmin_{x\in\mathcal{X}}(f_1(x), \dots, f_M(x))$

- How?

Our Prior Work

Multi-Objective Multi-Task Optimization[1]

- $argmin_{x_k\in\mathcal{X}_k} F_k(x_k), m \in [K]$

Why? ***Assuming*** *K multi-objective problems share certain similarities, we use their* ***Pareto-optimal solution distributions and surrogate models*** *as transferable knowledge to accelerate convergence.*

Our Current Work

Parametric Multi-Objective Multi-Task Optimization[2]

- $argmin_{x\in\mathcal{X}} F(x, \theta), \forall\theta \in \Theta$

We assume access to ***an external task space*** $\boldsymbol{\Theta}$ *and investigate how leveraging* ***task parameters*** $\boldsymbol{\theta}$ ***as side information*** *can accelerate optimization convergence and* ***return any preferred solution observing preference vector*** $\boldsymbol{\lambda}$ ***in any task*** $\boldsymbol{\theta}$***.***

[1]. Tingyang Wei, Jiao Liu, Abhishek Gupta, Puay Siew Tan, Yew-Soon Ong, “Bayesian Forward-Inverse Transfer for Multiobjective Optimization” International Conference on Parallel Problem Solving from Nature 2024.
[2]. Tingyang Wei, Jiao Liu, Abhishek Gupta, Chin Chun Ooi, Puay Siew Tan, Yew-Soon Ong, “Amortized Multi-Objective Optimization Across Tasks with Generative Solution Modeling”, IJCAI *2026.*

Why Parametric Multi-Objective Multi-Task Optimization?

- $argmin_{x \in \mathcal{X}} F(x, \theta) := (f_1(x, \theta), \ldots, f_M(x, \theta)), \forall \theta \in \Theta$

Opportunities:

- Explicit task parameters θ provide a structured way to **quantify task similarity**, contributing to **faster multi-objective optimization convergence** via efficient knowledge transfer.

Challenges:

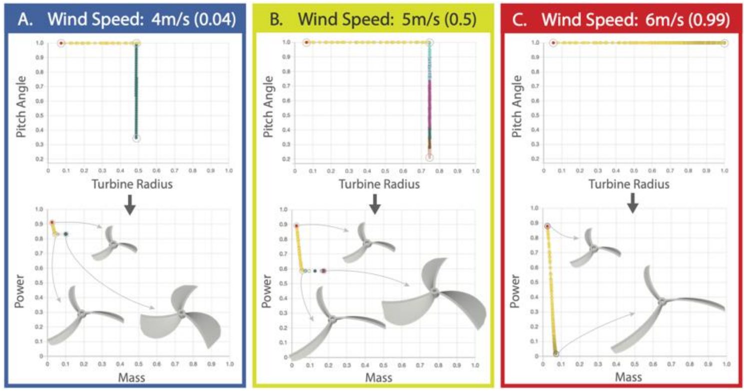
- The task parameters θ lie in a continuous space, meaning we have **infinitely many** multi-objective tasks to solve, and each multi-objective task require **potentially infinitely many solutions** to satisfy **arbitrary preference articulation** on Pareto Front for **any task**.

Application Example:

- In turbine design, θ can represent different **wind speeds**. Optimizing manufactured components x for diverse θ allows the parametric turbine to **adapt to unexpected environments**.

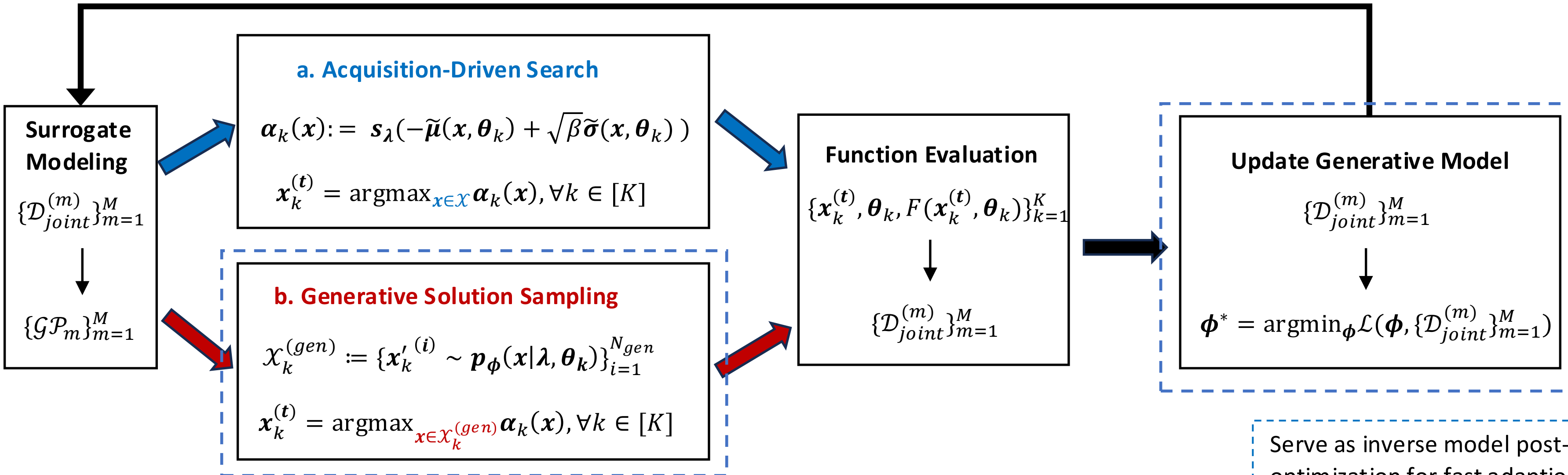
$$\forall \theta \in \Theta, \forall \lambda \in \Lambda$$
$$\mathcal{M}(\theta, \lambda) \approx argmin_{x \in \mathcal{X}} s_\lambda(F(x, \theta))$$

λ means preference vector to trade-off the objectives



Makatura, Liane, et al. "Pareto gamuts: exploring optimal designs across varying contexts." *ACM Transactions on Graphics (TOG)* 40.4 (2021): 1-17.

Methodology



Serve as inverse model post-optimization for fast adaption

Facilitate search to focus on the elite solutions in $\mathcal{X}$

We instantiate the Generative Model by either VAE or DDPM

- $\forall\theta \in \Theta, \forall\lambda \in \Lambda,\ \mathcal{M}(\theta,\lambda) \approx\ argmin_{x\in\mathcal{X}} s_\lambda(F(x,\theta))$

- Surrogate Models:
  - $\mathcal{GP}: \mathcal{X}\ \times\ \Theta\ \rightarrow \mathbb{R}$
  - (M models for M objectives)

Facilitate cross-task exploration in $\mathcal{X}\ \times\ \Theta$

- Cross-Task Sampling:
  - Forward Model contains cross-task information:
  - $\mathcal{GP}(\mu(\cdot,\theta), \sigma^2(\cdot,\theta))$

- Conditional Generative Model (Inverse Model):
  - $\mathcal{M}_{\boldsymbol{\phi}}: \Theta\ \times\ \Lambda \rightarrow \mathcal{X}$
  - (Identify the preferred solution by λ for task θ)

- Alternating Optimization Loop:
  - Per iteration, mode **a.** and **b.** alternates in a round robin way

Results

- $argmin_{x \in \mathcal{X}} F(x, \theta) \coloneqq (f_1(x, \theta), \dots, f_M(x, \theta)), \forall \theta \in \Theta$
- Good Optimization Results
  - Sample K multi-objective tasks
  - Check out the multi-objective convergence
    - $F(\cdot, \theta_k), k \in [K]$
  - Theoretical Analysis of faster convergence

Theorem 2 (Tightened Regret Bounds for PMT-MOBO). Under Assumption 1 and with the task-aware kernel that satisfies: $\widetilde{\kappa}_m((\mathbf{x}, \boldsymbol{\theta}_k), (\mathbf{x}', \boldsymbol{\theta}_k)) = \kappa_m(\mathbf{x}, \mathbf{x}'), \forall k \in [K]$, let $\widetilde{\gamma}_{T,k} = \max_m \widetilde{\gamma}_{T,m,k}$ be the largest MIG of PMT-MOBO across M objective functions over T scalarized evaluations for the k-th task, it holds that

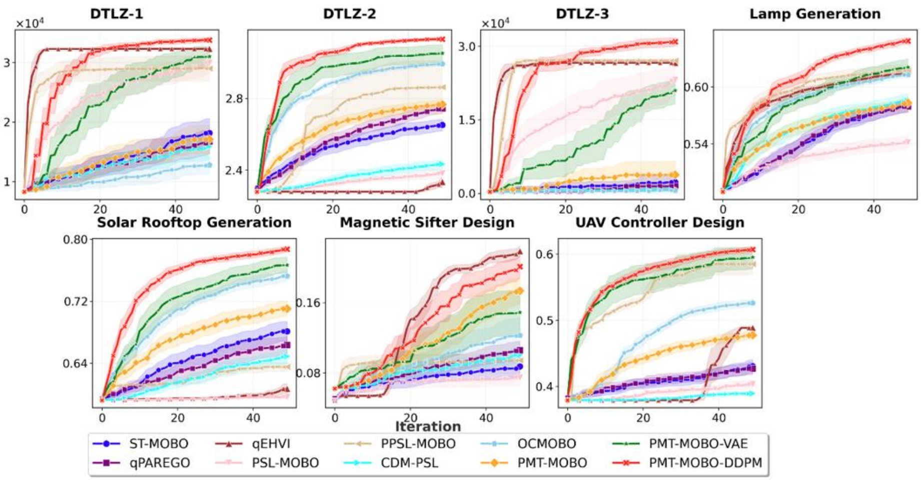
$$\widetilde{\gamma}_{T,m,k} \leq \gamma_{T,m,k}, \forall m \in [M], \forall k \in [K]. \quad (13)$$

Accordingly, it also holds that $\widetilde{\gamma}_{T,k} \leq \gamma_{T,k}, \forall k \in [K]$. Consider that both regret bounds from Theorem 1 still hold with $\gamma_{T,k}$ replaced by $\widetilde{\gamma}_{T,k}$, consequently, it holds that,

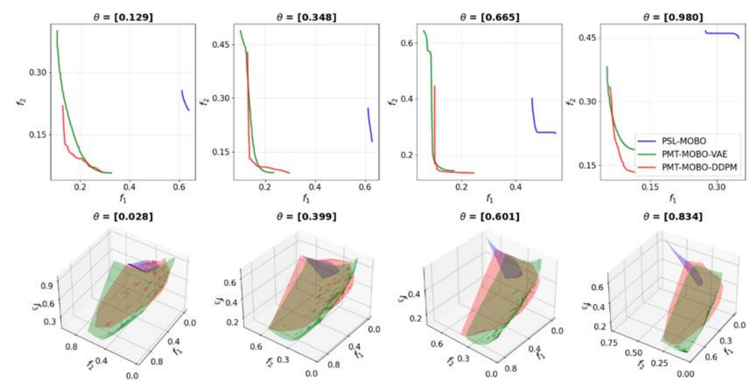
$$\mathbb{E}[\mathcal{R}_B^{\mathrm{PMT}}(T)] \leq \mathbb{E}[\mathcal{R}_B(T)], \quad (14)$$

i.e. PMT-MOBO achieves a tighter upper bound on Bayes regret than standard ST-MOBO.

- $\forall \theta \in \Theta, \forall \lambda \in \Lambda, \ \mathcal{M}(\theta, \lambda) \approx \ argmin_{x \in \mathcal{X}} s_\lambda(F(x, \theta))$
- Good Inverse Model Predictive Results
  - Sample multiple (unseen) multi-objective tasks θ
  - Sample diverse preference vectors $\lambda \in \Lambda$
  - Check out the solution quality of the model output
    - $F(x', \theta), \ x' = \mathcal{M}_{\boldsymbol{\phi}}(\theta, \lambda),$



Our PMT-MOBO-VAE/DDPM exhibit better hypervolume results



Our PMT-MOBO-VAE/DDPM generate better Pareto-optimal solutions

Detailed Comparisons

Method	DTLZ-1	DTLZ-2	DTLZ-3	LAMP	SOLAR	UAV
PSL-MOBO	2.82e+04 (1.78e+03)$^{\approx -}$	1.90e+00 (1.30e-01)$^{--}$	2.39e+04 (3.22e+03)$^{+-}$	5.01e-01 (7.32e-02)$^{--}$	3.54e-01 (4.06e-02)$^{--}$	4.12e-01 (7.30e-03)$^{--}$
CDM-PSL	1.92e+04 (3.85e+03)$^{--}$	2.52e+00 (8.49e-02)$^{--}$	1.83e+03 (1.31e+03)$^{--}$	6.19e-01 (4.45e-02)$^{\approx -}$	6.54e-01 (3.11e-02)$^{--}$	4.26e-01 (3.09e-02)$^{--}$
PPSL-MOBO	2.24e+04 (1.10e+02)$^{--}$	2.98e+00 (3.56e-02)$^{--}$	2.67e+04 (6.08e+02)$^{+-}$	6.12e-01 (1.05e-01)$^{\approx -}$	4.86e-01 (1.11e-01)$^{--}$	5.24e-01 (1.71e-01)$^{--}$
PMT-MOBO-VAE	2.88e+04 (2.57e+03)	3.10e+00 (2.99e-02)	1.78e+04 (7.28e+03)	6.31e-01 (7.33e-02)	**7.93e-01 (1.56e-02)**	**6.08e-01 (1.88e-02)**
PMT-MOBO-DDPM	**3.47e+04 (4.85e+02)**	**3.13e+00 (1.41e-02)**	**3.20e+04 (1.98e+03)**	**6.70e-01 (5.82e-02)**	7.65e-01 (1.30e-02)	6.05e-01 (2.44e-02)

Table 1: Inverse Model Generalization Performance on Unseen EMOP Tasks. Statistical significance is assessed using the Wilcoxon signed-rank test at the 95% confidence level. Superscripts $^{s_1 s_2}$ indicate that the competitor is significantly worse ($-$), significantly better ($+$), or statistically similar ($\approx$) relative to **PMT-MOBO-VAE** (s_1) and **PMT-MOBO-DDPM** (s_2).

Method	DTLZ-1	DTLZ-2	DTLZ-3	LAMP	SOLAR	MAGNETIC	UAV
ST-MOBO	1.82e+04 (4.80e+03)$^{--}$	2.65e+00 (6.30e-02)$^{--}$	2.24e+03 (3.03e+03)$^{--}$	5.80e-01 (1.20e-02)$^{--}$	6.81e-01 (2.51e-02)$^{--}$	8.72e-02 (2.15e-02)$^{--}$	4.31e-01 (1.99e-02)$^{--}$
qPAREGO	1.68e+04 (4.03e+03)$^{--}$	2.74e+00 (5.30e-02)$^{--}$	1.42e+03 (1.48e+03)$^{--}$	5.79e-01 (1.30e-02)$^{--}$	6.64e-01 (2.21e-02)$^{--}$	1.06e-01 (2.05e-02)$^{--}$	4.27e-01 (1.82e-02)$^{--}$
qEHVI	**3.23e+04 (5.06e+01)**$^{\approx -}$	2.33e+00 (6.20e-02)$^{--}$	2.66e+04 (1.85e-02)$^{+-}$	6.14e-01 (3.00e-03)$^{--}$	6.07e-01 (1.29e-02)$^{--}$	**2.18e-01 (1.15e-02)**$^{++}$	4.89e-01 (2.60e-03)$^{--}$
PSL-MOBO	3.00e+04 (3.52e+03)$^{\approx -}$	2.38e+00 (4.90e-02)$^{--}$	2.31e+04 (6.44e+03)$^{\approx -}$	5.42e-01 (1.40e-02)$^{--}$	5.96e-01 (3.60e-03)$^{--}$	7.42e-02 (1.79e-02)$^{--}$	4.03e-01 (1.46e-02)$^{--}$
CDM-PSL	1.59e+04 (4.15e+03)$^{--}$	2.43e+00 (5.04e-02)$^{--}$	7.76e+02 (9.41e+02)$^{--}$	5.85e-01 (1.39e-02)$^{--}$	6.49e-01 (1.97e-02)$^{--}$	1.00e-01 (1.82e-02)$^{--}$	3.89e-01 (5.70e-03)$^{--}$
PPSL-MOBO	3.20e+04 (1.72e+03)$^{\approx -}$	2.86e+00 (3.00e-01)$^{\approx -}$	**2.71e+04 (7.89e+02)**$^{+-}$	6.09e-01 (1.61e-02)$^{--}$	6.35e-01 (1.15e-02)$^{--}$	9.41e-02 (1.69e-02)$^{--}$	5.85e-01 (3.19e-02)$^{\approx -}$
OCMOBO	1.28e+04 (3.04e+03)$^{--}$	2.99e+00 (2.12e-02)$^{\approx -}$	6.20e+02 (6.29e+02)$^{--}$	6.13e-01 (6.60e-03)$^{--}$	7.53e-01 (1.36e-02)$^{--}$	1.23e-01 (3.95e-02)$^{--}$	5.26e-01 (1.09e-02)$^{--}$
PMT-MOBO	1.71e+04 (4.70e+03)$^{--}$	2.77e+00 (5.50e-02)$^{--}$	3.78e+03 (5.40e+03)$^{--}$	5.83e-01 (1.22e-02)$^{--}$	7.10e-01 (1.90e-02)$^{--}$	1.73e-01 (2.99e-02)$^{\approx -}$	4.78e-01 (1.27e-02)$^{-}$
PMT-MOBO-VAE	3.10e+04 (3.41e+03)	**3.05e+00 (9.58e-02)**	2.10e+04 (5.33e+03)	**6.21e-01 (1.73e-02)**	**7.67e-01 (2.01e-02)**	1.49e-01 (5.07e-02)	**5.95e-01 (3.33e-02)**
PMT-MOBO-DDPM	**3.37e+04 (8.35e+02)**	**3.13e+00 (1.62e-02)**	**3.09e+04 (1.67e+03)**	**6.48e-01 (5.10e-03)**	**7.87e-01 (7.80e-03)**	**2.01e-01 (2.50e-02)**	**6.07e-01 (7.50e-03)**

Table 5: Average Hypervolume results (mean $\pm$ std) across synthetic and real-world benchmarks. Statistical significance is assessed using the Wilcoxon signed-rank test at the 95% confidence level. Superscripts $^{s_1 s_2}$ indicate that the competitor is significantly worse ($-$), significantly better ($+$), or statistically similar ($\approx$) relative to **PMT-MOBO-VAE** (s_1) and **PMT-MOBO-DDPM** (s_2).

Conclusion

- We aim to provide a framework extending multitask optimization to infinite multi-task optimization.

- In TEVC2025, we present the first parametric multitask optimizer.

- In IJCAI2026, we present an alternation-based strategy to build a parametric multiobjective multitask optimizer based on generative models.
  - VAE/DDPM