CONSENSUS ACCELERATES EVOLUTION: BLACK-BOX MANY-TASK OPTIMIZATION VIA INFORMATION-GEOMETRIC CONSENSUS

Anonymous authors
Paper under double-blind review

ABSTRACT

Natural evolution strategies (NESs) provide a principled information-geometric framework for black-box optimization. However, scaling NESs to many concurrent tasks is hindered by hyperparameter sensitivity and high-variance natural-gradient estimates. With limited sampling per task, noisy distribution updates reduce useful progress and slow convergence. To address these limitations, we propose GemNES, an online meta-black-box optimizer using concurrent many-task optimization trajectories as supervision for data-efficient adaptation through dual-level information-geometric consensus (IGC). Specifically, global IGC pools normalized task-trajectory progress into a consensus signal for natural-gradient meta-policy updates and task-conditioned control. Meanwhile, local IGC aggregates aligned peer natural gradients in standardized tangent coordinates to improve each task update. Together, they encode cross-task agreement through natural-gradient updates on the meta-policy and task-distribution manifolds. Under alignment and regularity conditions, our analysis identifies a sufficient safe-transfer region, quantifies scale reduction relative to unsuppressed transfer, and derives tighter finite-horizon contraction for the evolving rank surrogate. Empirical results on synthetic many-task benchmarks and neural policy search demonstrate faster and more sustained convergence compared with representative baselines.

1 INTRODUCTION

Evolution strategies (ESs) (Hansen & Ostermeier, 2001; Beyer & Arnold, 2001; He et al., 2021) are black-box optimizers that update search distributions from sampled objective values rather than backpropagated gradients. This gradient-free distribution search preserves broad exploration on non-differentiable, deceptive, or noisy landscapes (Li et al., 2025a; Miikkulainen, 2025; Lange et al., 2023b), supporting applications from reinforcement learning (Salimans et al., 2017; Li et al., 2026b; 2024a) to language-model post-training (Qiu et al., 2026; Sarkar et al., 2026). Within this family, natural ESs (NESs) (Wierstra et al., 2014; Sun et al., 2009) optimize directly on the Riemannian manifold of probability distributions (Ollivier et al., 2017; Schaul, 2012). In high dimensions, limited Monte Carlo samples yield high-variance natural-gradient estimates that perturb both location and scale directions (Schaul et al., 2011), undermine reliable hyperparameter selection and informative updates, and thereby slow distributional convergence.

Real-world applications, such as cloud computing and large-scale automated simulations, frequently generate streams of related tasks. Evolutionary methods for many-task optimization (MaTO) (Gupta et al., 2016; Wei et al., 2022) exploit their shared structure (Bali et al., 2020; Bai et al., 2022; Li et al., 2025b). For NESs, MaTO creates an online meta-black-box optimization setting where concurrent trajectories provide richer online supervision and support adaptation across heterogeneous tasks. These signals directly address the two bottlenecks:

- *Global progress information* pools responses across tasks, landscapes, and search stages, providing richer online meta-data for adapting search controls to the evolving state.
- *Local geometric information* exposes aligned peer directions, providing structural priors that improve the effective update direction without requiring additional evaluations.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

These two information sources form a coupled feedback loop: controls learned through global IGC guide local IGC, whose transferred directions shape subsequent task progress. This coupling motivates dual-level information-geometric consensus (IGC) across the meta-policy and task-distribution levels as the organizing principle of GemNES (**Ge**ometric **m**eta-**NES**). GemNES realizes global IGC through online meta-policies and local IGC through affinity-guided natural-gradient aggregation, converting concurrent trajectories into continuously adapted controls and regularized location–scale distribution updates.

GemNES couples the learning of task-specific search controls with the construction of peer-informed update directions: global progress adapts learning rates, rank shaping, and transfer intensity, while geometric affinity and drift-coupled suppression shape location–scale consensus. Our analysis follows finite-sample rank utility along evolving search distributions to characterize how directional alignment, scale evolution, transfer intensity, curvature, and stochastic transfer cost jointly govern progress. It quantifies scale suppression and derives sufficient safe-rate, directional-gain, and contraction conditions for the refreshed rank surrogate. In summary, the main contributions are:

1. We propose GemNES, a black-box many-task optimizer via dual-level IGC: global IGC pools progress across task trajectories to learn policies for task-conditioned control, while local IGC regularizes task natural-gradient estimates through affinity-guided consensus.
2. Under explicit sufficient conditions, we theoretically establish a same-state utility advantage and a tighter contraction bound for the evolving frozen-rank surrogate.
3. Extensive experiments on synthetic benchmarks and high-dimensional policy search demonstrate faster and more sustained convergence with modest runtime overhead, and ablations characterize the contributions of global and local IGC.

2 Preliminaries & Related Work

Appendix Table A-I summarizes the mathematical notation used in this paper.

Black-Box Many-Task Optimization. Consider a black-box MaTO problem comprising K distinct tasks, aiming to find optimal solutions $\{\mathbf{x}_1^*, \ldots, \mathbf{x}_K^*\}$ that minimize respective objectives $f_k(\mathbf{x})$ across all tasks $k \in \{1, \ldots, K\}$. In deep neuroevolution, $\mathbf{x} \in \mathbb{R}^{D_k}$ typically represents the neural network weights, and f_k corresponds to the task-specific loss (*e.g.*, the negative cumulative reward). Traditionally, solving these tasks in isolation is sample-inefficient. Evolutionary multitasking (Gupta et al., 2016; Wei et al., 2022) optimizes tasks concurrently by exploiting latent structural priors via inter-task knowledge transfer. This paradigm has recently scaled to many concurrent tasks through several transfer mechanisms (Liaw & Ting, 2019; Jiang et al., 2023; Li et al., 2024b; 2025b). Beyond finite task sets, parametric multitask optimization extends joint search to solution and continuous, potentially infinite task spaces (Wei et al., 2026). For the mechanisms considered here, theoretical understanding of negative transfer and the interaction between transfer and distribution-scale evolution remains limited.

Natural Evolution Strategies. NESs (Wierstra et al., 2014; Sun et al., 2009; Ollivier et al., 2017) iteratively update a search distribution $P_{\boldsymbol{\theta}}$ to minimize $J(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}\sim P_{\boldsymbol{\theta}}}[f(\mathbf{x})]$. The information-geometric framework (Ollivier et al., 2017) estimates the natural gradient directly in the distribution's tangent space via score functions. For diagonal updates, GemNES uses a factorized Gaussian with location $\boldsymbol{\mu}$ and log-scale $\boldsymbol{\ell} = \log \boldsymbol{\sigma}$, corresponding to the separable NES parameterization (Schaul et al., 2011). Its score functions are $\mathbf{D}^{-1}\mathbf{z}$ and $\mathbf{z}^{\odot 2} - \mathbf{1}$ for $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ with $\mathbf{D} = \mathrm{diag}(\boldsymbol{\sigma})$. Monte Carlo estimates $\hat{\mathbf{g}}_{\boldsymbol{\mu}} = \sum_i w_i \mathbf{z}_i$ and $\hat{\mathbf{g}}_{\boldsymbol{\sigma}} = \sum_i w_i (\mathbf{z}_i^{\odot 2} - \mathbf{1})$ are obtained from λ fitness-ranked samples, where w_i are rank utilities. Despite favorable $\mathcal{O}(D)$ scaling, variance in these finite-sample natural-gradient estimates slows distributional progress and increases learning-rate sensitivity on complex landscapes. The proposed GemNES addresses this slow progress by adapting the rates online and supplementing each noisy direction with aligned peer evidence.

Meta-Black-Box Optimization. Meta-black-box optimization uses a hierarchical optimization structure to learn data-driven policies for base black-box optimizers. Offline methods learn update rules through self-attention (Lange et al., 2023b;a), symbolic distillation (Chen et al., 2024), or sequential decision models such as Q-Mamba (Ma et al., 2025a), while latent-space reverse engineering (Wang et al., 2026b) and neural landscape analysis (Ma et al., 2025b) improve general-

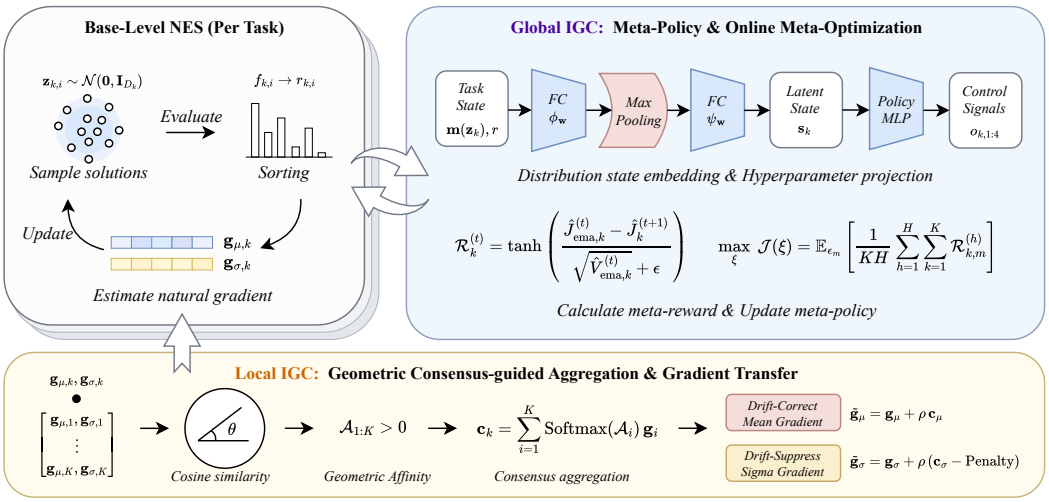


Figure 1: Overview of GemNES. **Top-Left:** Base-level NES estimates task-specific location and scale natural gradients from ranked samples. **Top-Right:** Global IGC embeds task states, produces task-conditioned controls through online meta-policies, and updates the policies from normalized progress. **Bottom:** Local IGC filters task gradients by geometric affinity, aggregates aligned peer directions, and applies drift correction and suppression.

ization. L2T learns state-conditioned transfer policies for subsequent multitask search (Wu et al., 2025). Online meta-learning instead updates policies *in situ* (Finn et al., 2019; Son et al., 2025). DLMCMA coordinates multiple inner operators and aggregates their evolution paths in distributed large-scale optimization (Duan et al., 2024), while task-free ABOM (Wang et al., 2026a) adapts evolutionary operators online from target populations. MaTO supplies multiple concurrent trajectories, providing richer supervision for efficient meta-policy adaptation without pre-training.

3 Information-Geometric Meta Natural Evolution Strategies

This section introduces GemNES in three parts: online meta-optimization for task-conditioned control, IGC-guided gradient transfer with theoretical analysis, and complexity analysis (Figure 1).

Information-Geometric Consensus (IGC) unifies the online meta-optimization and gradient-transfer modules by aligning cross-task coordination with the Fisher geometry of probability distributions. Global IGC aggregates normalized task-trajectory progress into a consensus signal that updates a shared meta-policy distribution via natural gradient, yielding task-conditioned controls. Local IGC aggregates peer natural gradients selected by geometric affinity to update each task distribution.

3.1 Global IGC: Online Meta-Optimization

GemNES realizes global IGC through an online embedding–mapping–meta-optimization loop: each evaluated sample batch is embedded into a task state, the meta-policy maps that state to task-specific controls, and cross-task progress updates its search distribution.

Dimension-Agnostic & Sample-Order-Invariant State Embedding. To operate across heterogeneous task dimensions and unordered sample sets, the meta-policy requires a common state representation. Given current samples $\mathcal{S}_{k,t} = \{(\mathbf{z}_{k,i}, r_{k,i})\}_{i=1}^{\lambda}$, a symmetric encoder produces:

$$\mathbf{s}_k^{(t)} = \psi_{\mathbf{w}}\left(\max_{i=1}^{\lambda} \phi_{\mathbf{w}}\left([\tfrac{1}{\sqrt{D_k}}\textstyle\sum_{d=1}^{D_k}[\mathbf{z}_{k,i}]_d;\ r_{k,i}]\right)\right) \in \mathbb{R}^{d_{\text{latent}}}. \tag{1}$$

Here $\mathbf{z}_{k,i} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the sampled perturbation, $r_{k,i} \in (0, 1]$ is its normalized ascending fitness rank, and $\phi_{\mathbf{w}}, \psi_{\mathbf{w}}$ are $\tanh$-activated fully connected layers. The normalized projection provides a perturbation–rank feature across dimensions, while max-pooling removes sample-order dependence.

State-to-Control Mapping. The policy passes $[\mathbf{s}_k^{(t)}; b_t; \iota_k]$, where b_t is the fraction of the evaluation budget consumed and $\iota_k \in (0, 1]$ is the task's normalized index, through two $\tanh$ hidden layers and applies a cubed $\tanh$ function to produce $\mathbf{o}_k \in (-1, 1)^4$. Its controls are

$$\begin{aligned} s_{\boldsymbol{\mu},k} &= \exp(o_{k,1}), \quad \eta_{\boldsymbol{\mu},k} = \eta_{\boldsymbol{\mu}}^{\text{base}} s_{\boldsymbol{\mu},k}, \quad \zeta_k = \exp(o_{k,3}), \\ s_{\boldsymbol{\sigma},k} &= \exp(o_{k,2}), \quad \eta_{\boldsymbol{\sigma},k} = \eta_{\boldsymbol{\sigma}}^{\text{base}} s_{\boldsymbol{\sigma},k}, \quad \rho_k = \rho_{\max} \max(0, o_{k,4}). \end{aligned} \tag{2}$$

The exponentials scale the base rates, while the rectified output bounds $0 \le \rho_k < \rho_{\max}$; Corollary 1 characterizes a sufficient state-dependent safe-rate region. The exponent ζ_k adapts the selection pressure of the shaped rank weights used to estimate the base natural gradients (Appendix B).

Natural-Gradient Meta-Optimization. The meta-policy is optimized online through $Q_{\boldsymbol{\xi}} = \mathcal{N}(\mathbf{w}, \operatorname{diag}(\boldsymbol{\sigma}_{\text{meta}}^2))$, with $\boldsymbol{\xi} = [\mathbf{w}; \boldsymbol{\sigma}_{\text{meta}}]$. For a sampled candidate m, let $\mathbf{w}_m = \mathbf{w} + \boldsymbol{\sigma}_{\text{meta}} \odot \boldsymbol{\epsilon}_m$, $\boldsymbol{\epsilon}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $\mathbf{ctrl}_{k,m}^{(t)} = \Pi_{\mathbf{w}_m}([\mathbf{s}_k^{(t)}; b_t; \iota_k])$:

$$\begin{aligned} &\textit{Meta-level:} \quad \max_{\boldsymbol{\xi}} \ \mathcal{J}(\boldsymbol{\xi}) = \mathbb{E}_{\mathbf{w}_m \sim Q_{\boldsymbol{\xi}}} \left[\mathbb{E}[\hat{\mathcal{J}}_m \mid \mathbf{w}_m] \right], \\ &\textit{Base-level:} \quad \min_{\boldsymbol{\theta}_k} \ J_k(\boldsymbol{\theta}_k) = \mathbb{E}_{\mathbf{x} \sim P_{\boldsymbol{\theta}_k}}[f_k(\mathbf{x})]. \end{aligned} \tag{3}$$

Here the inner expectation covers the evolving states visited by each candidate. The base-level objective is optimized by equations 8–9 with controls $\mathbf{ctrl}_{k,m}^{(t)} = (\eta_{\boldsymbol{\mu},k}, \eta_{\boldsymbol{\sigma},k}, \zeta_k, \rho_k)$. Candidates are evaluated on successive segments of a shared online trajectory, with H visits per candidate; normalized progress makes task contributions comparable. Appendix B details the evaluation procedure.

To compare progress across tasks with different objective scales, we define the task reward and candidate score as

$$\mathcal{R}_{k,m}^{(h)} = \tanh\left(\frac{\hat{J}_{\text{ema},k}^{(t)} - \hat{J}_k^{(t+1)}}{\sqrt{\hat{V}_{\text{ema},k}^{(t)}} + \epsilon} \right), \quad \hat{\mathcal{J}}_m = \frac{1}{KH} \sum_{h=1}^{H} \sum_{k=1}^{K} \mathcal{R}_{k,m}^{(h)}. \tag{4}$$

Here $\hat{J}_k^{(t+1)}$ is the mean objective of the new sample batch, while $\hat{J}_{\text{ema},k}^{(t)}$ and $\hat{V}_{\text{ema},k}^{(t)}$ provide a running task-specific baseline and scale. The same batch refreshes these statistics and becomes the next cached state, so the reward requires no extra evaluations.

Finally, the M trajectory-relative scores are converted to centered rank utilities v_m for an SNES-style update of the meta-search distribution. The mean update favors higher-scoring policies, while the scale adapts exploration in policy space:

$$\mathbf{w} \leftarrow \mathbf{w} + \eta_{\mathbf{w}} \boldsymbol{\sigma}_{\text{meta}} \odot \sum_{m=1}^{M} v_m \boldsymbol{\epsilon}_m, \quad \boldsymbol{\sigma}_{\text{meta}} \leftarrow \boldsymbol{\sigma}_{\text{meta}} \odot \exp\left(\frac{\eta_{\boldsymbol{\sigma},\text{meta}}}{2} \sum_{m=1}^{M} v_m (\boldsymbol{\epsilon}_m^{\odot 2} - \mathbf{1}) \right). \tag{5}$$

The pooled score $\hat{\mathcal{J}}_m$ forms a global consensus signal from normalized progress across task trajectories. Equation 5 uses this signal to update the shared meta-policy distribution via natural gradient and learn task-conditioned controls.

3.2 Local IGC: Natural-Gradient Transfer

Local IGC complements this global adaptation by regularizing high-variance task estimates with peer evidence in standardized natural-gradient coordinates.

Geometric Affinity. At generation t, let $\hat{\mathbf{g}}_k = [\hat{\mathbf{g}}_{\boldsymbol{\mu},k}; \hat{\mathbf{g}}_{\boldsymbol{\sigma},k}] \in \mathbb{R}^{2D_k}$ be task k's finite-sample rank-based natural-gradient estimate. For heterogeneous tasks, let π_{kj} map task-k coordinates in $\mathcal{I}_{kj}$ to prescribed task-j coordinates. Affinity is evaluated on this shared subspace:

$$\mathcal{A}_{k,j} = \frac{1}{2} \Big(\cos\left(\hat{\mathbf{g}}_{\boldsymbol{\mu},k}[\mathcal{I}_{kj}], \hat{\mathbf{g}}_{\boldsymbol{\mu},j}[\pi_{kj}(\mathcal{I}_{kj})]\right) + \cos\left(\hat{\mathbf{g}}_{\boldsymbol{\sigma},k}[\mathcal{I}_{kj}], \hat{\mathbf{g}}_{\boldsymbol{\sigma},j}[\pi_{kj}(\mathcal{I}_{kj})]\right) \Big). \tag{6}$$

For vector tasks, $\mathcal{I}_{kj} = \{1, \ldots, \min(D_k, D_j)\}$ and $\pi_{kj}(d) = d$; neural policies use the same hidden architecture and serialized parameter order, with π_{kj} defining a heuristic equal-index correspondence over the common prefix (Li et al., 2026b). The stabilized cosine is zero if either vector is zero, and $\mathcal{P}_k = \{j \neq k \mid \mathcal{A}_{k,j} > 0\}$ retains positively aligned peers.

Consensus Aggregation & Drift-Coupled Suppression. Softmax affinity weighting smoothly reweights positively aligned peers. For $q \in \{\mu, \sigma\}$, coordinate $d \le D_k$, and nonempty $\mathcal{P}_{k,d} = \{i \in \mathcal{P}_k : d \in \mathcal{I}_{ki}\}$,

$$[\mathbf{c}_{q,k}]_d = \sum_{i \in \mathcal{P}_{k,d}} \frac{\exp(\mathcal{A}_{k,i})}{\sum_{j \in \mathcal{P}_{k,d}} \exp(\mathcal{A}_{k,j})} [\hat{\mathbf{g}}_{q,i}]_{\pi_{ki}(d)}. \tag{7}$$

If $\mathcal{P}_k = \emptyset$, transfer is disabled. For any coordinate with $\mathcal{P}_{k,d} = \emptyset$, we set $[\mathbf{c}_{\boldsymbol{\mu},k}]_d = [\mathbf{c}_{\boldsymbol{\sigma},k}]_d = 0$, so the update on that coordinate remains the isolated task gradient.

Location consensus can be injected additively, but treating log-scale consensus in the same way can expand the search variance and destabilize exploration. We therefore apply drift-coupled suppression:

$$\tilde{\mathbf{g}}_{\boldsymbol{\mu},k} = \hat{\mathbf{g}}_{\boldsymbol{\mu},k} + \rho_k \mathbf{c}_{\boldsymbol{\mu},k}, \quad \tilde{\mathbf{g}}_{\boldsymbol{\sigma},k} = \hat{\mathbf{g}}_{\boldsymbol{\sigma},k} + \rho_k \Big(\mathbf{c}_{\boldsymbol{\sigma},k} - \alpha \frac{\|\mathbf{c}_{\boldsymbol{\sigma},k}\|_1}{\|\mathbf{c}_{\boldsymbol{\mu},k}\|_1 + \epsilon} |\mathbf{c}_{\boldsymbol{\mu},k}| \Big). \tag{8}$$

Here ρ_k is the meta-learned transfer rate, α controls suppression, and absolute values are elementwise. The L_1 ratio bounds the total correction by $\alpha \|\mathbf{c}_{\boldsymbol{\sigma},k}\|_1$ and distributes it according to $|\mathbf{c}_{\boldsymbol{\mu},k}|$; Proposition 1 quantifies the resulting scale reduction relative to unsuppressed transfer.

Finally, the task distribution uses the regularized directions together with the meta-learned learning rates:

$$\boldsymbol{\mu}_k \leftarrow \boldsymbol{\mu}_k + \eta_{\boldsymbol{\mu},k} \cdot \big(\boldsymbol{\sigma}_k \odot \tilde{\mathbf{g}}_{\boldsymbol{\mu},k} \big), \quad \boldsymbol{\sigma}_k \leftarrow \boldsymbol{\sigma}_k \odot \exp\Big(\frac{\eta_{\boldsymbol{\sigma},k}}{2} \tilde{\mathbf{g}}_{\boldsymbol{\sigma},k} \Big). \tag{9}$$

Theoretical Analysis. We analyze local IGC on the joint location–scale trajectory, deriving conditions under which aligned consensus improves same-state rank utility and finite-horizon contraction.

Why Can Local IGC Help? Finite-sample noise makes isolated natural-gradient updates unreliable, slowing distributional progress. Local IGC aggregates affinity-filtered peer directions, while learned transfer and scale suppression control risk. We quantify scale suppression and directional gains, then connect them to tighter contraction under alignment.

For readability, we fix a generic task and drop the subscript k. Let $\mathcal{F}_t$ be the pre-sampling history, r_i the randomized fitness rank of $\mathbf{X}_i$, and I an independent uniform sample index. Using the shaped weight $w_{\zeta,\mathsf{r}}$ defined in Appendix B, define the finite-sample rank utility and its frozen-rank surrogate (Ollivier et al., 2017) by

$$u_t(\mathbf{x}) = \lambda \mathbb{E}[w_{\zeta_t,\mathsf{r}_I} \mid \mathbf{X}_I = \mathbf{x}, \mathcal{F}_t], \quad \Phi_t(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = \mathbb{E}_{\mathbf{X} \sim P_{\boldsymbol{\theta}}}[u_t(\mathbf{X})]. \tag{10}$$

This surrogate freezes the current rank utility while the distribution moves, making candidate updates comparable at the same state. With $\boldsymbol{\ell}_t = \log \boldsymbol{\sigma}_t$ and $\mathbf{D}_t = \mathrm{diag}(\boldsymbol{\sigma}_t)$, conditioning on $\mathbf{X}_I$ at $\boldsymbol{\theta}^{(t)}$ gives $\mathbb{E}[\hat{\mathbf{g}}_{\boldsymbol{\mu}}^{(t)} \mid \mathcal{F}_t] = \mathbf{D}_t \nabla_{\boldsymbol{\mu}} \Phi_t$ and $\mathbb{E}[\hat{\mathbf{g}}_{\boldsymbol{\sigma}}^{(t)} \mid \mathcal{F}_t] = \nabla_{\boldsymbol{\ell}} \Phi_t$. Appendix A.1 gives the finite-λ identity. Let $\mathbf{y}_t = [\boldsymbol{\mu}_t; \boldsymbol{\ell}_t]$ and abbreviate $\Phi_t(\boldsymbol{\theta}(\mathbf{y}); \boldsymbol{\theta}^{(t)})$ as $\Phi_t(\mathbf{y})$, where $\boldsymbol{\theta}(\mathbf{y}) = [\boldsymbol{\mu}; \exp(\boldsymbol{\ell})]$. Given $\mathcal{F}_t$, the task distribution is fixed; sample-derived states, controls, and estimates are jointly random.

The transfer-induced scale direction is $\mathbf{b}_{\boldsymbol{\sigma}}^{(t)} = \mathbf{c}_{\boldsymbol{\sigma}}^{(t)} - \alpha \upsilon_t |\mathbf{c}_{\boldsymbol{\mu}}^{(t)}|$ with $\upsilon_t = \|\mathbf{c}_{\boldsymbol{\sigma}}^{(t)}\|_1 / (\|\mathbf{c}_{\boldsymbol{\mu}}^{(t)}\|_1 + \epsilon)$, where $\upsilon_t = 0$ if $\mathbf{c}_{\boldsymbol{\sigma}}^{(t)} = \mathbf{0}$.

Proposition 1 (Exact Scale Suppression)**.** *Let superscripts α and 0 denote the scale updates in equation 9 with and without suppression, all else fixed, and let $\boldsymbol{\Sigma}^{(\alpha)} = \mathrm{diag}((\boldsymbol{\sigma}^{(\alpha)})^{\odot 2})$. For $\alpha \ge 0$, $\eta_{\boldsymbol{\sigma},t} \ge 0$, and $\rho_t \ge 0$,*

$$\frac{[\boldsymbol{\sigma}^{(\alpha)}]_d}{[\boldsymbol{\sigma}^{(0)}]_d} = \exp\Big(-\frac{\eta_{\boldsymbol{\sigma},t} \rho_t \alpha \upsilon_t}{2} |[\mathbf{c}_{\boldsymbol{\mu}}^{(t)}]_d| \Big) \le 1, \quad \log \frac{\det \boldsymbol{\Sigma}^{(\alpha)}}{\det \boldsymbol{\Sigma}^{(0)}} = -\eta_{\boldsymbol{\sigma},t} \rho_t \alpha \upsilon_t \|\mathbf{c}_{\boldsymbol{\mu}}^{(t)}\|_1 \le 0. \tag{11}$$

Thus suppression cannot increase any marginal variance or the covariance determinant relative to unsuppressed transfer. (Proof and further bounds in Appendix A.3)

Factor $\eta_{\boldsymbol{\mu},t} = \tau_t \varphi_{\boldsymbol{\mu},t}$ and $\eta_{\boldsymbol{\sigma},t} = \tau_t \varphi_{\boldsymbol{\sigma},t}$ for deterministic $\tau_t > 0$. Define the isolated direction $\mathbf{p}_t$, unscaled consensus direction $\mathbf{d}_t$, and transferred direction $\mathbf{h}_t$ as

$$\mathbf{p}_t = \begin{bmatrix} \varphi_{\boldsymbol{\mu},t} \mathbf{D}_t \hat{\mathbf{g}}_{\boldsymbol{\mu}}^{(t)} \\ \frac{\varphi_{\boldsymbol{\sigma},t}}{2} \hat{\mathbf{g}}_{\boldsymbol{\sigma}}^{(t)} \end{bmatrix}, \quad \mathbf{d}_t = \begin{bmatrix} \varphi_{\boldsymbol{\mu},t} \mathbf{D}_t \mathbf{c}_{\boldsymbol{\mu}}^{(t)} \\ \frac{\varphi_{\boldsymbol{\sigma},t}}{2} \mathbf{b}_{\boldsymbol{\sigma}}^{(t)} \end{bmatrix}, \quad \mathbf{h}_t = \rho_t \mathbf{d}_t. \tag{12}$$

Thus equation 9 is $\mathbf{y}_{t+1} = \mathbf{y}_t + \tau_t(\mathbf{p}_t + \mathbf{h}_t)$, where $\mathbf{D}_t = \mathrm{diag}(\exp \boldsymbol{\ell}_t)$ evolves at each step and $\mathbf{h}_t$ isolates the IGC contribution.

Definition 1 (Positive Consensus Alignment). *For $\mathbf{R}_t = \nabla_{\mathbf{y}} \Phi_t(\mathbf{y}_t)$, positive consensus alignment means $s_t := \langle \mathbf{R}_t, \mathbb{E}[\mathbf{h}_t \mid \mathcal{F}_t] \rangle > 0$, i.e., the expected consensus direction has a positive projection onto the current surrogate-ascent direction.*

Geometric affinity in equation 6 provides an observable sample-level proxy for the distribution-level condition in Definition 1; the correspondence need not be exact. Appendix A.2 gives a sufficient finite-sample diagnostic for the aggregated IGC direction.

All smoothness conditions below concern convex regions containing the compared paths. Here $\mathbf{y}^{\mathrm{iso}}_{t+1}$ denotes the state produced by only the isolated update at step t.

Theorem 1 (Evolving-Distribution Utility Advantage). *Suppose Φ_t is L-smooth, $L > 0$, on such a region. For $\mathbf{y}^{\mathrm{iso}}_{t+1} = \mathbf{y}_t + \tau_t \mathbf{p}_t$ and $\mathcal{M}_t := \mathbb{E}[\|\mathbf{p}_t\|\|\mathbf{h}_t\| + \frac{1}{2}\|\mathbf{h}_t\|^2 \mid \mathcal{F}_t]$,*

$$\mathbb{E}[\Phi_t(\mathbf{y}_{t+1}) - \Phi_t(\mathbf{y}^{\mathrm{iso}}_{t+1}) \mid \mathcal{F}_t] \geq \tau_t(s_t - L\tau_t \mathcal{M}_t). \tag{13}$$

Hence a positively aligned step with $0 < \tau_t < s_t/(L\mathcal{M}_t)$ has a strict utility advantage. (Proof in Appendix A.4)

Corollary 1 (Safe Transfer-Rate Region). *Fix the realized directions $(\mathbf{p}_t, \mathbf{d}_t)$, and assume L-smoothness, with $L > 0$, on such a region for transfer multipliers in $[0, \rho_{\max}]$. Let $\Gamma_t := \langle \mathbf{R}_t, \mathbf{d}_t \rangle - L\tau_t\|\mathbf{p}_t\|\|\mathbf{d}_t\|$. If $\Gamma_t > 0$, $0 < \rho_t \leq \rho_{\max}$, and $\rho_t < 2\Gamma_t/(L\tau_t\|\mathbf{d}_t\|^2)$, then*

$$\Phi_t(\mathbf{y}_{t+1}) - \Phi_t(\mathbf{y}^{\mathrm{iso}}_{t+1}) \geq \tau_t \rho_t \left(\Gamma_t - \frac{L\tau_t \rho_t}{2} \|\mathbf{d}_t\|^2 \right) > 0, \tag{14}$$

and the lower bound is maximized at $\rho_t^{\mathrm{lb}} = \min\{\rho_{\max}, \Gamma_t/(L\tau_t\|\mathbf{d}_t\|^2)\}$. (Proof in Appendix A.5)

Whereas Corollary 1 bounds the transfer magnitude at a fixed step scale, the next result isolates the directional benefit at equal displacement.

Corollary 2 (Equal-Length Directional Advantage). *Suppose Φ_t is L-smooth on such a region and the relevant directions are nonzero. Let $\mathbf{u}_t^{\mathrm{iso}} = \mathbf{p}_t/\|\mathbf{p}_t\|$ and $\mathbf{u}_t^{\mathrm{igc}} = (\mathbf{p}_t + \mathbf{h}_t)/\|\mathbf{p}_t + \mathbf{h}_t\|$, and set $\omega_t = \tau_t\|\mathbf{p}_t + \mathbf{h}_t\|$. Define $\chi_t = \langle \mathbf{R}_t, \mathbf{u}_t^{\mathrm{igc}} - \mathbf{u}_t^{\mathrm{iso}} \rangle$ and $\mathbf{y}^{\mathrm{iso,eq}}_{t+1} = \mathbf{y}_t + \omega_t \mathbf{u}_t^{\mathrm{iso}}$. If $\chi_t > 0$ and $0 < \omega_t < \chi_t/L$, then*

$$\Phi_t(\mathbf{y}_{t+1}) - \Phi_t(\mathbf{y}^{\mathrm{iso,eq}}_{t+1}) \geq \omega_t(\chi_t - L\omega_t) > 0. \tag{15}$$

If $\mathbf{h}_t = \bar{\rho}\,\mathbf{p}_t$ for $\bar{\rho} \geq 0$, the two normalized directions coincide and $\chi_t = 0$, so the directional-gain term vanishes. (Proof and a finite-sample diagnostic are in Appendix A.6)

For $\mathbf{R}_t \neq \mathbf{0}$, define the normalized consensus contribution $\beta_t^{\mathrm{igc}} = s_t/\|\mathbf{R}_t\|^2$. Dividing equation 13 by $\|\mathbf{R}_t\|^2$ then gives the net consensus margin $\delta_t := \beta_t^{\mathrm{igc}} - L\tau_t\mathcal{M}_t/\|\mathbf{R}_t\|^2$. The condition $\delta_t > 0$ makes the expected same-state gain positive and supplies the extra contraction term in Theorem 2.

Theorem 2 (Conditional Joint-Trajectory Contraction). *Assume the local Polyak–Łojasiewicz (PL) gradient-dominance, base-moment, and surrogate-refresh conditions in Appendix A.7, with PL constant c and alignment/moment constants $\underline{h}$, $\bar{h}$, and ν. Let $G_t = \Phi_t^* - \Phi_t(\mathbf{y}_t) > 0$, where Φ_t^* is the local supremum of the current surrogate, and define*

$$\kappa = \operatorname*{ess\,inf}_{t<T} 2c\tau_t \left(\underline{h} - \frac{L\tau_t \bar{h}^2}{2} \right), \quad \gamma = \operatorname*{ess\,inf}_{t<T} 2c\tau_t \delta_t. \tag{16}$$

Here κ and γ are the isolated and IGC contraction contributions. If $0 < \kappa < \kappa + \gamma < 1$, let $\varpi = 1 - \kappa - \gamma$ and $e_t = L\tau_t^2\nu/2 + 2v_t$, where v_t bounds the expected rank-utility refresh variation. Then

$$\mathbb{E}[G_T] \leq \varpi^T \mathbb{E}[G_0] + \sum_{t<T} \varpi^{T-1-t} e_t. \tag{17}$$

Since $\varpi < 1 - \kappa$, IGC improves the corresponding isolated worst-case contraction factor. With $\Delta_t^{\mathrm{iso}} = \Phi_t(\mathbf{y}^{\mathrm{iso}}_{t+1}) - \Phi_t(\mathbf{y}_t)$, it also satisfies

$$\sum_{t<T} \mathbb{E}[\Phi_t(\mathbf{y}_{t+1}) - \Phi_t(\mathbf{y}_t)] \geq \sum_{t<T} \mathbb{E}[\Delta_t^{\mathrm{iso}} + \gamma G_t], \tag{18}$$

so IGC contributes at least $\sum_{t<T} \mathbb{E}[\gamma G_t] > 0$ beyond the corresponding isolated updates. (Proof in Appendix A.8)

3.3 Complete Algorithm & Complexity Analysis

Algorithm A-1 and Appendix B give the complete procedure and complexity. GemNES organizes the K tasks into $G = \lceil K/B \rceil$ groups of at most B tasks. Each group evolves an independent meta-policy and performs local IGC, while all groups are synchronized at every all-task update. Each update therefore costs $\mathcal{O}(K\lambda D + K\lambda \log \lambda + KBD + K|\mathbf{w}|)$, with MH updates per meta-cycle.

For extremely large task sets, fixing the large cap B makes the coordination term $\mathcal{O}(KBD)$ and hence linear in K. The task-scale study in Appendix E shows that bounded coordination retains performance comparable to full coordination, while runtime follows the same near-linear growth trend as an independently run single-task baseline.

4 Empirical Studies

Tasks. We employ two scenarios: (i) synthetic MaTO benchmarks (Feng et al., 2019; 2022) comprising 16 instances with 50 concurrent tasks each for continuous objective minimization; and (ii) the Brax-MuJoCo (Freeman et al., 2021) policy search suite involving 10 concurrent heterogeneous control tasks, where each task evolves a 4×32 MLP policy network to maximize cumulative reward. Experiments are performed in MToP (Li et al., 2026a). Task details are in Appendix C.1. Appendix E further evaluates task-scale efficiency with $K = 2^5$ to 2^{10} tasks.

Baselines. We first compare against ES baselines: single-task SNES (Schaul et al., 2011); offline meta LES (Lange et al., 2023b); many-task TNG-SNES (Li et al., 2025b), MES-RET (Li et al., 2026b), MTES-KG (Li et al., 2024b), and SBCMAES (Liaw & Ting, 2019). Additionally, for synthetic tasks, we include population-based BoKTDE (Jiang et al., 2023), MFEA (Gupta et al., 2016), and online meta ABOM (Wang et al., 2026a). For policy search tasks, we include OpenAI-ES (Salimans et al., 2017), PPO (Schulman et al., 2017), and SAC (Haarnoja et al., 2018). Within each scenario, all black-box methods use the same number of function evaluations as their stopping criterion. Detailed settings are in Appendix C.2.

Ablation Variants. We test **w/o-meta** (all meta-controlled settings frozen), **w/o-transfer** (meta-adapted controls without local IGC), and **w/o-similarity** (unfiltered local consensus). We also freeze individual controls at prescribed reference values on synthetic benchmarks. Appendix D further reports parameter sensitivity.

Performance Metrics. For synthetic tasks, we report UOV (Li et al., 2025b), a comparison-set-relative metric, over 30 independent runs. For policy search, we report the mean cumulative reward over 5 rollouts across 10 independent runs. We further report Wilcoxon rank-sum post-hoc comparisons, Friedman ranks, and the number of best results (#B). Appendix C.3 provides the formal metric definition, and Appendix F gives full result tables, convergence plots, and supplementary analyses.

4.1 Comparison Results

Table 1: Summary of comparative results.

Synthetic Benchmarks						Policy Search MuJoCo					
Algorithm	+	−	=	Rank	#B	Algorithm	+	−	=	Rank	#B
GemNES	↓	↑		**2.31**	**9**	**GemNES**	↓	↑		**2.25**	**6**
SNES	1	13	2	3.81	0	SNES	0	4	6	4.50	2
LES	0	16	0	6.56	0	LES	1	4	5	4.65	4
TNG-SNES	4	10	2	3.44	2	TNG-SNES	0	4	6	4.75	3
MES-RET	3	12	1	3.50	2	MES-RET	1	3	6	3.40	3
MTES-KG	3	13	0	4.44	2	MTES-KG	0	3	7	3.85	2
SBCMAES	3	11	2	5.88	0	SBCMAES	0	7	3	6.35	2
BoKTDE	1	15	0	8.00	0	OpenAI-ES	0	5	5	6.25	2
MFEA	1	15	0	8.88	0	PPO*	1	6	3	–	
ABOM	1	15	0	8.19	1	SAC*	1	8	1	–	

*PPO and SAC use different interaction budgets; their rank and #B are omitted.

Synthetic Benchmarks. We compare GemNES with isolated, meta, and many-task baselines. As summarized in Table 1 (left), GemNES achieves the best rank (2.31) and nine best mean results. The reported taskwise tests favor it over SNES, LES, and ABOM on 13–16 instances. GemNES also records 10–13 taskwise wins against TNG-SNES, MES-RET, MTES-KG, and SBCMAES, and 15 each against BoKTDE and MFEA. This broad advantage is consistent with online task-conditioned control and affinity-filtered consensus reusing aligned task information while suppressing conflict-

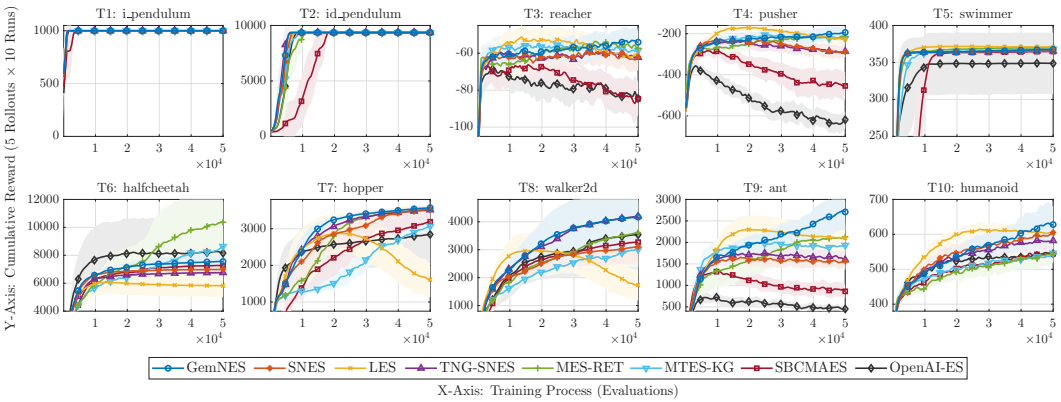


Figure 2: Convergence curves of cumulative reward ↑ on policy search tasks. Shaded areas show the mean ±1.96 standard errors across 10 independent runs (each averaged over 5 rollouts). GemNES demonstrates sustained convergence, especially on *walker2d* and *ant*.

ing transfer. GemNES requires an average of 0.51 hours per run, remaining within the runtime range of the ES baselines while requiring substantially less time than BoKTDE and ABOM.

Figure A-3 illustrates the convergence behavior. GemNES maintains sustained improvement across most instances, while LES and ABOM often plateau earlier. Its gains on the low-similarity WCCI22-MaTSO9–10 instances (Feng et al., 2022) are consistent with online meta-adaptation refining task-conditioned controls and affinity filtering suppressing harmful transfer.

Policy Search Tasks. To evaluate GemNES in high-dimensional neuroevolution policy search, we benchmark it against ESs and use gradient-based RL methods as cross-paradigm references on the Brax-MuJoCo suite. As detailed in Table 1 (right), GemNES achieves the best Friedman rank (2.25) and six best mean results, including ties. Together with its synthetic results, this demonstrates consistent effectiveness in high-dimensional neural policy search. It records 3–7 taskwise wins against every ES baseline by the reported test; only LES and MES-RET are better on *swimmer*. GemNES takes 8.04 hours per run, nearly identical to SNES (8.01 hours) and comparable to the other ESs (Figure A-1). Under their respective budgets, GemNES attains significantly higher rewards than PPO and SAC on six and eight tasks, with three and one non-significant differences, respectively.

Figure 2 and Figure A-4 further illustrate this optimization behavior. While most ESs converge rapidly on simpler tasks, several baselines plateau at lower rewards on *pusher* and *ant*. LES also declines late on *pusher*, *hopper*, and *walker2d*, whereas GemNES sustains progress on *walker2d* and *ant*. Global IGC adapts task-conditioned controls, while local IGC filters misaligned peers as task trajectories diverge. Under their respective budgets, PPO and SAC finish lower on several simpler tasks and achieve higher rewards on *humanoid*.

4.2 Ablation Study

Table 2: Summary of ablation study results.

Algorithm	Synthetic Benchmarks					Policy Search MuJoCo				
	+	−	=	Rank	#B	+	−	=	Rank	#B
GemNES	↓	↑		**3.00**	**4**	↓	↑		**1.75**	**6**
w/o-meta	0	14	2	6.38	0	0	6	4	3.35	2
w/o-transfer	0	5	11	4.25	2	0	3	7	2.45	5
w/o-similarity	0	2	14	3.31	1	0	1	9	2.45	3
w/o-ada-stepsize	0	8	8	3.88	3			–		
w/o-ada-shaping	0	2	14	3.31	3			–		
w/o-ada-transfer	0	3	13	3.88	3			–		

Impact of Global IGC. w/o-meta freezes all meta-controlled settings, yielding average ranks of 6.38 on synthetic tasks and 3.35 on MuJoCo. Among the single-control variants, freezing the two step-size scales causes eight taskwise losses, compared with two for rank shaping and three for the transfer rate. These results highlight the step-size adaptation driven by global IGC, with the best overall rank attained by adapting all controls jointly.

Impact of Local IGC. w/o-transfer incurs five taskwise losses on synthetic tasks and three on MuJoCo. GemNES sustains improvement on *halfcheetah* and *ant* after this variant begins to plateau,

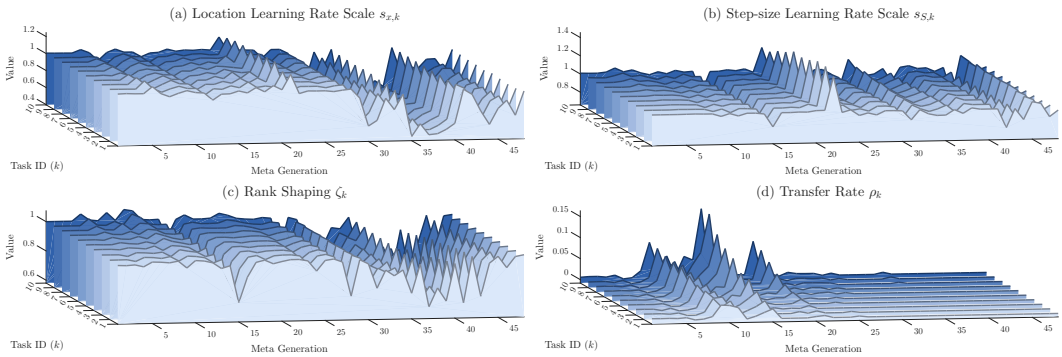


Figure 3: Meta-learned dynamics of four controls across Brax-MuJoCo tasks over meta-generations. The stacked task trajectories show shared adaptation with task-specific deviations: the location-rate scale and rank-shaping exponent generally decline, the scale-rate factor remains comparatively stable, and transfer is concentrated in the early stages.

achieving 11.4% and 30.3% higher terminal mean rewards, respectively. These gains highlight how peer-informed updates complement global control under limited per-task sampling.

Impact of Geometric Affinity. GemNES outperforms **w/o-similarity** on two synthetic instances and on *ant*, supporting the role of affinity filtering in mitigating negative transfer. These results suggest that peer aggregation exploits shared search structure, with affinity filtering providing selective gains by excluding less aligned peer directions.

4.3 Online Meta-Learned Hyperparameter Dynamics

Figure 3 shows shared trends and task-specific deviations in the learned controls. The policy combines state features, budget progress, and task identity to adjust search behavior across tasks and optimization stages.

Learning-Rate Scales. The location-rate scale $s_{\boldsymbol{\mu},k}$ generally decreases, indicating increasingly conservative location updates; this qualitative trend agrees with the small-step sufficient condition in Theorem 1. The scale-rate factor $s_{\boldsymbol{\sigma},k}$ varies less over optimization. Proposition 1 separately establishes suppression relative to unsuppressed transfer.

Rank-Shaping Exponent. ζ_k initially undergoes a steady decline, followed by stronger fluctuations in later stages, indicating task- and stage-dependent adaptation of selection pressure. A smaller ζ_k flattens the normalized rank weights, preventing a few top samples from dominating early updates, while later fluctuations adjust their concentration to task-specific progress.

Transfer Rate. ρ_k is initially high, placing greater weight on the consensus update analyzed in Theorems 1 and 2. Tasks with longer optimization plateaus tend to exhibit larger ρ_k values, suggesting greater reliance on gradient transfer. As optimization progresses and task trajectories diverge, ρ_k gradually decreases, indicating more conservative cross-task transfer. This schedule is qualitatively consistent with the finite safe-rate region in Corollary 1.

5 Conclusion & Discussion

Summary. This work proposes GemNES, an online meta-black-box optimizer for MaTO. GemNES mitigates slow distributional convergence from noisy natural-gradient estimates under limited sampling through dual-level IGC: global IGC adapts task-conditioned controls via online meta-policy distributions, while local IGC aggregates aligned peer directions. Theory gives conditional scale, transfer, and contraction results; experiments show faster and more sustained convergence on many-task benchmarks and high-dimensional neural policy search, together with task-count scalability.

Limitations & Future Work. The current formulation uses continuous search spaces, diagonal-Gaussian distributions, and prescribed cross-task coordinate correspondences. Future work can explore learned coordinate alignment for neural policies, richer distributions, structured or discrete variables, and consensus efficiency on nonstationary tasks.

AI USE STATEMENT

In this work, we used generative AI tools to edit the paper for readability, including grammar, clarity, and concision, and to assist in checking the mathematical proofs by reviewing assumptions and algebraic steps. The authors reviewed all AI-assisted suggestions and verified the final proofs. We take full responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

Reproducibility is supported by the complete update equations and pseudocode in Algorithm A-1 and Appendix B, including the execution order and complexity analysis. The assumptions and complete proofs of the theoretical results are provided in Appendix A. Appendix Sections C.1–C.3 document the task settings, evaluation budgets, baselines, hyperparameters, and performance metrics. Parameter sweeps and full numerical results with convergence curves are reported in Appendices D and F. The source code is included in the supplementary materials and will be released publicly.

REFERENCES

Lu Bai, Wu Lin, Abhishek Gupta, and Yew-Soon Ong. From multitask gradient descent to gradient-free evolutionary multitasking: A proof of faster convergence. *IEEE Transactions on Cybernetics*, 52(8):8561–8573, 2022. doi: 10.1109/TCYB.2021.3052509.

Kavitesh Kumar Bali, Yew-Soon Ong, Abhishek Gupta, and Puay Siew Tan. Multifactorial evolutionary algorithm with online transfer parameter estimation: MFEA-II. *IEEE Transactions on Evolutionary Computation*, 24(1):69–83, 2020. doi: 10.1109/TEVC.2019.2906927.

H-G Beyer and Dirk V Arnold. Theory of evolution strategies—a tutorial. *Theoretical aspects of evolutionary computing*, pp. 109–133, 2001.

Jiacheng Chen, Zeyuan Ma, Hongshu Guo, Yining Ma, Jie Zhang, and Yue-jiao Gong. Symbol: Generating flexible black-box optimizers through symbolic equation learning. In *The Twelfth International Conference on Learning Representations*, 2024.

Qiqi Duan, Chang Shao, Guochen Zhou, Minghan Zhang, Qi Zhao, and Yuhui Shi. Distributed evolution strategies with multi-level learning for large-scale black-box optimization. *IEEE Transactions on Parallel and Distributed Systems*, 35(11):2087–2101, 2024. doi: 10.1109/TPDS.2024.3437688.

Liang Feng, Kai Qin, Abhishek Gupta, Yuan Yuan, Yew-Soon Ong, and Xu Chi. CEC 2019 competition on evolutionary multi-task optimization. *2019 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–1, 2019. URL http://www.bdsc.site/websites/MTO_competiton_2019/MTO_Competition_CEC_2019.html.

Liang Feng, Kai Qin, Abhishek Gupta, Yuan Yuan, Yew-Soon Ong, and Xu Chi. WCCI 2022 competition on evolutionary multi-task optimization. *2022 IEEE World Congress on Computational Intelligence (WCCI)*, pp. 1–1, 2022. URL http://www.bdsc.site/websites/MTO_competition_2021/MTO_Competition_WCCI_2022.html.

Chelsea Finn, Aravind Rajeswaran, Sham Kakade, and Sergey Levine. Online meta-learning. In *International Conference on Machine Learning*, pp. 1920–1930. PMLR, 2019.

C. Daniel Freeman, Erik Frey, Anton Raichuk, Sertan Girgin, Igor Mordatch, and Olivier Bachem. Brax - a differentiable physics engine for large scale rigid body simulation. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021. URL http://github.com/google/brax.

Abhishek Gupta, Yew-Soon Ong, and Liang Feng. Multifactorial evolution: Toward evolutionary multitasking. *IEEE Transactions on Evolutionary Computation*, 20(3):343–357, 2016.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International Conference on Machine Learning*, pp. 1861–1870. PMLR, 2018.

Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001. doi: 10.1162/106365601750190398.

Xiaoyu He, Zibin Zheng, and Yuren Zhou. Mmes: Mixture model-based evolution strategy for large-scale optimization. *IEEE Transactions on Evolutionary Computation*, 25(2):320–333, 2021. doi: 10.1109/TEVC.2020.3034769.

Yi Jiang, Zhi-Hui Zhan, Kay Chen Tan, and Jun Zhang. A bi-objective knowledge transfer framework for evolutionary many-task optimization. *IEEE Transactions on Evolutionary Computation*, 27(5):1514–1528, 2023. doi: 10.1109/TEVC.2022.3210783.

Robert Lange, Tom Schaul, Yutian Chen, Chris Lu, Tom Zahavy, Valentin Dalibard, and Sebastian Flennerhag. Discovering attention-based genetic algorithms via meta-black-box optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '23, pp. 929–937, New York, NY, USA, 2023a. Association for Computing Machinery. ISBN 9798400701191. doi: 10.1145/3583131.3590496.

Robert Tjarko Lange, Tom Schaul, Yutian Chen, Tom Zahavy, Valentin Dalibard, Chris Lu, Satinder Singh, and Sebastian Flennerhag. Discovering evolution strategies via meta-black-box optimization. In *The Eleventh International Conference on Learning Representations*, 2023b.

Pengyi Li, Yan Zheng, Hongyao Tang, Xian Fu, and Jianye Hao. EvoRainbow: Combining improvements in evolutionary reinforcement learning for policy search. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 29427–29447. PMLR, 21–27 Jul 2024a.

Pengyi Li, Jianye Hao, Hongyao Tang, Xian Fu, Yan Zhen, and Ke Tang. Bridging evolutionary algorithms and reinforcement learning: A comprehensive survey on hybrid algorithms. *IEEE Transactions on Evolutionary Computation*, 29(5):1707–1728, 2025a. doi: 10.1109/TEVC.2024.3443913.

Yanchi Li, Wenyin Gong, and Shuijia Li. Multitask evolution strategy with knowledge-guided external sampling. *IEEE Transactions on Evolutionary Computation*, 28(6):1733–1745, 2024b. doi: 10.1109/TEVC.2023.3330265.

Yanchi Li, Wenyin Gong, and Qiong Gu. Transfer task-averaged natural gradient for efficient many-task optimization. *IEEE Transactions on Evolutionary Computation*, 29(5):1952–1965, 2025b. doi: 10.1109/TEVC.2024.3459862.

Yanchi Li, Wenyin Gong, Tingyu Zhang, Fei Ming, Shuijia Li, Qiong Gu, and Yew-Soon Ong. MToP: A MATLAB benchmarking platform for evolutionary multitasking. *ACM Transactions on Evolutionary Learning and Optimization*, 2026a. ISSN 2688-299X. doi: 10.1145/3812535.

Yanchi Li, Jiao Liu, Wenyin Gong, Qiong Gu, Yue Zhao, and Yew-Soon Ong. Breaking multitask curse: Reward-weighted evolution for black-box many-task optimization. In *Forty-third International Conference on Machine Learning*, 2026b. URL https://openreview.net/forum?id=lkGnJhXUNu.

Rung-Tzuo Liaw and Chuan-Kang Ting. Evolutionary manytasking optimization based on symbiosis in biocoenosis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01): 4295–4303, Jul. 2019. doi: 10.1609/aaai.v33i01.33014295.

Zeyuan Ma, Zhiguang Cao, Zhou Jiang, Hongshu Guo, and Yue-Jiao Gong. Meta-black-box-optimization through offline q-function learning. In *Forty-second International Conference on Machine Learning*, 2025a. URL https://openreview.net/forum?id=7qCkf5OT3x.

Zeyuan Ma, Jiacheng Chen, Hongshu Guo, and Yue-Jiao Gong. Neural exploratory landscape analysis for meta-black-box-optimization. In *The Thirteenth International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=EEI5R89Cmv.

Risto Miikkulainen. Neuroevolution insights into biological neural computation. *Science*, 387 (6735):eadp7478, 2025. doi: 10.1126/science.adp7478.

Yann Ollivier, Ludovic Arnold, Anne Auger, and Nikolaus Hansen. Information-geometric optimization algorithms: A unifying picture via invariance principles. *Journal of Machine Learning Research*, 18(1):564–628, 2017.

Xin Qiu, Yulu Gan, Conor F. Hayes, Qiyao Liang, Yinggan XU, Roberto Dailey, Elliot Meyerson, Babak Hodjat, and Risto Miikkulainen. Evolution strategies at scale: LLM fine-tuning beyond reinforcement learning. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=i0P4ew9GpS.

Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, and Ilya Sutskever. Evolution strategies as a scalable alternative to reinforcement learning, 2017. arXiv:1703.03864.

Bidipta Sarkar, Mattie Fellows, Juan Agustin Duque, Alistair Letcher, Antonio León Villares, Anya Sims, Clarisse Wibault, Dmitry Samsonov, Dylan Cope, Jarek Luca Liesen, Kang Li, Lukas Seier, Theo Wolf, Uljad Berdica, Valentin Mohl, Alexander David Goldie, Aaron Courville, Karin Sevegnani, Shimon Whiteson, and Jakob Nicolaus Foerster. Evolution strategies at the hyperscale. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=bfVJ4GsHrO.

Tom Schaul. Natural evolution strategies converge on sphere functions. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, GECCO '12, pp. 329–336, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450311779. doi: 10.1145/2330163.2330211.

Tom Schaul, Tobias Glasmachers, and Jürgen Schmidhuber. High dimensions and heavy tails for natural evolution strategies. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11, pp. 845–852, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450305570. doi: 10.1145/2001576.2001692.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017. arXiv:1707.06347.

Jaehyeon Son, Soochan Lee, and Gunhee Kim. When meta-learning meets online and continual learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(1): 413–432, 2025. doi: 10.1109/TPAMI.2024.3463709.

Yi Sun, Daan Wierstra, Tom Schaul, and Juergen Schmidhuber. Efficient natural evolution strategies. In *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, GECCO '09, pp. 539–546, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605583259. doi: 10.1145/1569901.1569976.

Chao Wang, Licheng Jiao, Lingling Li, Jiaxuan Zhao, Guanchun Wang, Fang Liu, and Shuyuan Yang. Task-free adaptive meta black-box optimization. In *The Fourteenth International Conference on Learning Representations*, 2026a. URL https://openreview.net/forum?id=AufVSUgMUo.

Chen Wang, Yue-Jiao Gong, Zhiguang Cao, and Zeyuan Ma. Instance generation for meta-black-box optimization through latent space reverse engineering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 37178–37186, 2026b.

Tingyang Wei, Shibin Wang, Jinghui Zhong, Dong Liu, and Jun Zhang. A review on evolutionary multi-task optimization: Trends and challenges. *IEEE Transactions on Evolutionary Computation*, 26(5):941–960, 2022. doi: 10.1109/TEVC.2021.3139437.

Tingyang Wei, Jiao Liu, Abhishek Gupta, Puay Siew Tan, and Yew-Soon Ong. (θ_l, θ_u)-parametric multitask optimization: Joint search in solution and infinite task spaces. *IEEE Transactions on Evolutionary Computation*, 30(3):1270–1283, 2026. doi: 10.1109/TEVC.2025.3598333.

Daan Wierstra, Tom Schaul, Tobias Glasmachers, Yi Sun, Jan Peters, and Jürgen Schmidhuber. Natural evolution strategies. *Journal of Machine Learning Research*, 15(27):949–980, 2014.

Sheng-Hao Wu, Yuxiao Huang, Xingyu Wu, Liang Feng, Zhi-Hui Zhan, and Kay Chen Tan. Learning to transfer for evolutionary multitasking. *IEEE Transactions on Cybernetics*, 55(7):3342–3355, 2025. URL https://arxiv.org/abs/2406.14359.

Appendix

TABLE OF CONTENTS

A Proofs & Additional Theoretical Details **13**
A.1 Finite-Sample Rank Identity . . . 15
A.2 Finite-Sample Diagnostic for Aggregated Alignment . . . 15
A.3 Proof of Proposition 1 . . . 15
A.4 Proof of Theorem 1 . . . 16
A.5 Proof of Corollary 1 . . . 16
A.6 Proof & Diagnostic for Corollary 2 . . . 16
A.7 Conditions for Theorem 2 . . . 17
A.8 Proof of Theorem 2 . . . 17

B Complete Algorithm **18**
B.1 Procedure & Update Details . . . 18
B.2 Complexity Analysis . . . 20

C Experimental Details **21**
C.1 Task Settings . . . 21
C.2 Baselines . . . 21
C.3 Performance Metrics . . . 23

D Parameter Sensitivity Analysis **24**

E Task-Scale Efficiency Analysis **25**

F Detailed Results & Supplementary Analysis **26**
F.1 Synthetic Benchmarks . . . 26
F.2 Policy Search: Comparison with ES Baselines . . . 26
F.3 Policy Search: Comparison with RL Baselines . . . 28
F.4 Ablation Studies . . . 28
F.5 Parameter Sensitivity Details . . . 31

A PROOFS & ADDITIONAL THEORETICAL DETAILS

We first establish the finite-sample rank identity and an alignment diagnostic, then prove scale suppression, the local utility advantages, and joint-trajectory contraction. The results concern the evolving frozen-rank surrogate and give sufficient conditions: affinity is an observable proxy for expected alignment, the safe-rate bound is state dependent, the isolated comparison is evaluated at the same state, and Proposition 1 controls scale relative to unsuppressed transfer. Table A-I summarizes the notation used throughout. As in Section 3.2, we suppress the task index k and retain the generation index t throughout the proofs.

Table A-I: Summary of notation. Bold symbols denote vectors or matrices, hats denote finite-sample estimates, and tildes denote consensus-regularized quantities.

Symbol	Description
General MaTO	
K	Number of concurrent tasks.
$B,\ G$	Maximum coordination-group size $1 \le B \le K$ and number of groups $G = \lceil K/B \rceil$.
$\mathcal{G}_g,\ K_g$	Task-index set of group g and its size $K_g = \lvert\mathcal{G}_g\rvert \le B$.
$D_k,\ D$	Task-k dimensionality and maximum dimension $D = \max_k D_k$.
$f_k(\cdot)$	Objective function for task k.
λ	Number of samples drawn per task update.
ϵ	Numerical stabilizer (10^{-8}) used in reward normalization and scale suppression.
Base-Level NES	
$P_{\boldsymbol{\theta}_k},\ \boldsymbol{\theta}_k = [\boldsymbol{\mu}_k; \boldsymbol{\sigma}_k]$	Gaussian search distribution for task k, with location $\boldsymbol{\mu}_k$ and scale $\boldsymbol{\sigma}_k$.
$J_k(\boldsymbol{\theta}_k)$	Expected objective value for task k.
$\mathbf{X},\ \mathbf{X}_i,\ \mathbf{x}$	Random samples and a realized search point.
$\mathbf{z}_{k,i}$	Isotropic perturbation vector for sample i of task k.
$\mathcal{S}_{k,t}$	Evaluated sample batch cached for task k at generation t.
$w_{\zeta_k,i},\ r_{k,i}$	Shaped utility at rank position i and normalized fitness rank of sample i.
$\hat{\mathbf{g}}_{\boldsymbol{\mu},k}, \hat{\mathbf{g}}_{\boldsymbol{\sigma},k}$	Finite-sample natural-gradient estimates for location and scale.
$\eta_{\boldsymbol{\mu},k}, \eta_{\boldsymbol{\sigma},k}$	Learning rates for location and scale updates.
Meta-Level Optimization	
$\Pi_{\mathbf{w}},\ \lvert\mathbf{w}\rvert$	Neural meta-policy parameterized by $\mathbf{w}$ and its parameter count.
$Q_{\boldsymbol{\xi}},\ \boldsymbol{\xi} = [\mathbf{w}; \boldsymbol{\sigma}_{\text{meta}}]$	Diagonal meta-policy search distribution and its location–scale state.
$\mathcal{J}(\boldsymbol{\xi}),\ \hat{\mathcal{J}}_m$	Expected meta-objective and candidate m's trajectory-relative progress score.
$\mathbf{ctrl}_{k,m}^{(t)}$	Candidate-conditioned controls (learning rates, rank-shaping exponent, and transfer rate).
$\mathcal{R}_{k,m}^{(h)}$	Normalized meta-reward of task k under candidate m at pass h.
$\hat{J}_{\text{ema},k}^{(t)},\ \hat{V}_{\text{ema},k}^{(t)},\ \beta_{\text{ema}}$	EMA estimates of the objective mean and variance, and the EMA coefficient.
$M,\ H$	Numbers of candidate policies and interleaved visits per candidate.
$\mathbf{w}_m,\ \boldsymbol{\epsilon}_m,\ v_m$	Candidate-policy weights, perturbation, and centered rank utility.
$\mathbf{s}_k^{(t)},\ b_t,\ \iota_k$	Task embedding, normalized evaluation-budget progress, and normalized within-group index.
$\eta_{\boldsymbol{\mu}}^{\text{base}}, \eta_{\boldsymbol{\sigma}}^{\text{base}}$	Base learning rates prior to meta-modulation.
$\eta_{\mathbf{w}}, \eta_{\boldsymbol{\sigma},\text{meta}}$	Meta location and scale learning rates.
ζ_k	Exponent controlling rank-based selection pressure.
$\rho_k, \rho_{\max}$	Gradient transfer rate and its maximum value.
$s_{\boldsymbol{\mu},k}, s_{\boldsymbol{\sigma},k}$	Meta-learned scaling factors for learning rates.
Information-Geometric Consensus Mechanism	
$\mathcal{A}_{k,j}$	Geometric affinity between tasks k and j.
$\mathcal{I}_{kj},\ \pi_{kj}$	Selected task-k coordinates and their task-j mapping.
$\mathcal{P}_k$	Set of affinity-selected peer tasks for task k.
$\mathbf{c}_{\boldsymbol{\mu},k}, \mathbf{c}_{\boldsymbol{\sigma},k}$	Consensus directions for location and scale.
$\tilde{\mathbf{g}}_{\boldsymbol{\mu},k}, \tilde{\mathbf{g}}_{\boldsymbol{\sigma},k}$	Final regularized gradients after consensus injection.
α	Coefficient of drift-coupled scale suppression.
Theoretical Analysis	
$\mathcal{F}_t,\ \mathsf{r}_i,\ w_{\zeta_t,\mathsf{r}_i},\ q_t$	Pre-sampling history, randomized rank, rank-shaped weight, and comparison probability.
$u_t,\ \Phi_t$	Finite-sample rank utility and its frozen-rank surrogate.
$\boldsymbol{\ell}_t,\ \mathbf{D}_t$	Log-scale vector and $\text{diag}(\exp \boldsymbol{\ell}_t)$.
$\mathbf{y}_t = [\boldsymbol{\mu}_t; \boldsymbol{\ell}_t]$	Joint location–log-scale state of the evolving search distribution.
$\tau_t,\ \varphi_{\boldsymbol{\mu},t},\ \varphi_{\boldsymbol{\sigma},t}$	Deterministic step scale and state-dependent location/scale multipliers.
$\mathbf{p}_t,\ \mathbf{d}_t,\ \mathbf{h}_t$	Isolated, unscaled IGC, and transfer-scaled IGC directions.
$\upsilon_t,\ \mathbf{b}_{\boldsymbol{\sigma}}^{(t)}$	Suppression ratio and resulting transfer-induced scale direction.
$\mathbf{R}_t,\ \hat{\mathbf{R}}_t$	Joint surrogate gradient and its empirical estimate.
$s_t,\ \beta_t^{\text{igc}}$	Expected consensus alignment and its normalized contribution.
$\mathcal{M}_t,\ \delta_t$	Expected consensus cost and net margin.
$\Gamma_t,\ \rho_t^{\text{lb}}$	Realized margin and transfer rate maximizing its safe lower bound.
$\mathcal{U},\ T,\ \Phi_t^*,\ G_t$	Local neighborhood, update horizon, local surrogate supremum, and optimality gap.
$c,\ L$	PL and smoothness constants of the frozen-rank surrogate on the joint state.
$\kappa,\ \gamma,\ \varpi$	Isolated and IGC contraction contributions and their joint contraction factor.
$\mathbf{u}_t^{\text{iso}},\ \mathbf{u}_t^{\text{igc}},\ \omega_t,\ \chi_t$	Equal-length isolated/IGC directions, common displacement, and directional gain.
$\underline{h},\ \bar{h},\ \nu$	Lower alignment and multiplicative/additive second-moment constants.
$v_t,\ e_t$	Surrogate-refresh bound and total recurrence error.
Δ_t^{iso}	Isolated-update improvement under the current frozen-rank surrogate.

A.1 Finite-Sample Rank Identity

Fix time t. The history $\mathcal{F}_t$ precedes generation of the cached batch and determines $(\boldsymbol{\mu}_t, \mathbf{D}_t)$. Each sample has conditional marginal $\mathbf{X}_i = \boldsymbol{\mu}_t + \mathbf{D}_t \mathbf{z}_i$, with $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The state encoder and controls, including ζ_t, may depend on the entire batch. Let I be uniform on $\{1, \dots, \lambda\}$ and independent of all sampling and algorithmic randomness. Suppressing time indices, the tower property gives

$$\mathbb{E}\left[\sum_{i=1}^{\lambda} w_{\zeta,\mathrm{r}_i} \mathbf{z}_i \mid \mathcal{F}_t\right] = \lambda \mathbb{E}[w_{\zeta,\mathrm{r}_I} \mathbf{z}_I \mid \mathcal{F}_t] = \mathbb{E}[u_t(\mathbf{X}_I)\mathbf{z}_I \mid \mathcal{F}_t].$$

Replacing each $\mathbf{z}_i$ by $\mathbf{z}_i^{\odot 2} - \mathbf{1}$ gives the scale identity. This argument applies to both independent and antithetic Gaussian batches. In the independent-sampling special case with ζ_t fixed before sampling and tie-free objective values, conditional on $\mathbf{X}_1 = \mathbf{x}$, the number of other samples whose objective values are no larger is $B \sim \mathrm{Bin}(\lambda - 1, q_t(\mathbf{x}))$, and hence

$$u_t(\mathbf{x}) = \lambda \mathbb{E}_{B \sim \mathrm{Bin}(\lambda-1, q_t(\mathbf{x}))}[w_{\zeta_t, B+1}], \quad q_t(\mathbf{x}) = \Pr(f(\mathbf{X}) \le f(\mathbf{x}) \mid \mathcal{F}_t).$$

Freezing u_t and differentiating its Gaussian expectation gives the two surrogate-gradient relations stated after equation 10. Randomized tie-breaking is included in the conditional expectation defining u_t.

A.2 Finite-Sample Diagnostic for Aggregated Alignment

This diagnostic evaluates the complete transfer direction after affinity filtering and aggregation. Let $\hat{\mathbf{R}}_t = [\mathbf{D}_t^{-1} \hat{\mathbf{g}}_{\boldsymbol{\mu}}^{(t)}; \hat{\mathbf{g}}_{\boldsymbol{\sigma}}^{(t)}]$, with error radii $\|\hat{\mathbf{R}}_t - \mathbf{R}_t\| \le e_{r,t}$ and $\|\mathbf{h}_t - \mathbb{E}[\mathbf{h}_t \mid \mathcal{F}_t]\| \le e_{h,t}$. Positive consensus alignment is certified whenever

$$\langle \hat{\mathbf{R}}_t, \mathbf{h}_t \rangle > e_{r,t} \|\mathbf{h}_t\| + (\|\hat{\mathbf{R}}_t\| + e_{r,t}) e_{h,t}. \tag{19}$$

With the fixed generation index suppressed, write $\hat{\mathbf{R}} = \mathbf{R} + \mathbf{e}_r$ and $\mathbf{h} = \mathbb{E}[\mathbf{h}] + \mathbf{e}_h$, where $\|\mathbf{e}_r\| \le e_r$ and $\|\mathbf{e}_h\| \le e_h$. Then

$$\langle \mathbf{R}, \mathbb{E}[\mathbf{h}] \rangle = \langle \hat{\mathbf{R}}, \mathbf{h} \rangle + \langle \mathbf{R} - \hat{\mathbf{R}}, \mathbf{h} \rangle + \langle \mathbf{R}, \mathbb{E}[\mathbf{h}] - \mathbf{h} \rangle.$$

Cauchy–Schwarz and $\|\mathbf{R}\| \le \|\hat{\mathbf{R}}\| + e_r$ give

$$\langle \mathbf{R}, \mathbb{E}[\mathbf{h}] \rangle \ge \langle \hat{\mathbf{R}}, \mathbf{h} \rangle - e_r \|\mathbf{h}\| - (\|\hat{\mathbf{R}}\| + e_r) e_h,$$

which is positive under equation 19. Thus equation 19 is a sufficient finite-sample diagnostic for the aggregated IGC direction at the specified error radii; geometric affinity is the sample-level proxy used by the algorithm.

A.3 Proof of Proposition 1

Proof. The two scale updates differ only by the suppression term, so equation 9 gives, coordinate-wise,

$$\frac{[\boldsymbol{\sigma}^{(\alpha)}]_d}{[\boldsymbol{\sigma}^{(0)}]_d} = \exp\left(-\frac{\eta_{\boldsymbol{\sigma},t} \rho_t \alpha \upsilon_t}{2} |[\mathbf{c}_{\boldsymbol{\mu}}^{(t)}]_d|\right).$$

This ratio is at most one because all scalar factors in its exponent are nonnegative. Squaring the coordinate ratios and multiplying them yields

$$\frac{\det \boldsymbol{\Sigma}^{(\alpha)}}{\det \boldsymbol{\Sigma}^{(0)}} = \exp\left(-\eta_{\boldsymbol{\sigma},t} \rho_t \alpha \upsilon_t \|\mathbf{c}_{\boldsymbol{\mu}}^{(t)}\|_1\right),$$

which proves equation 11.

Moreover,

$$\upsilon_t \|\mathbf{c}_{\boldsymbol{\mu}}^{(t)}\|_1 = \|\mathbf{c}_{\boldsymbol{\sigma}}^{(t)}\|_1 \frac{\|\mathbf{c}_{\boldsymbol{\mu}}^{(t)}\|_1}{\|\mathbf{c}_{\boldsymbol{\mu}}^{(t)}\|_1 + \epsilon} \le \|\mathbf{c}_{\boldsymbol{\sigma}}^{(t)}\|_1.$$

Since $\mathbf{b}_{\boldsymbol{\sigma}}^{(t)} - \mathbf{c}_{\boldsymbol{\sigma}}^{(t)} = -\alpha \upsilon_t |\mathbf{c}_{\boldsymbol{\mu}}^{(t)}| \preceq \mathbf{0}$,

$$\|\mathbf{b}_{\boldsymbol{\sigma}}^{(t)} - \mathbf{c}_{\boldsymbol{\sigma}}^{(t)}\|_1 = \alpha \upsilon_t \|\mathbf{c}_{\boldsymbol{\mu}}^{(t)}\|_1 \le \alpha \|\mathbf{c}_{\boldsymbol{\sigma}}^{(t)}\|_1.$$

□

A.4 Proof of Theorem 1

Proof. Suppress the generation index and write $\mathbf{y} = \mathbf{y}_t$, $\mathbf{p} = \mathbf{p}_t$, $\mathbf{h} = \mathbf{h}_t$, and $\tau = \tau_t$. Let

$$A = \Phi(\mathbf{y} + \tau\mathbf{p} + \tau\mathbf{h}) - \Phi(\mathbf{y} + \tau\mathbf{p}).$$

The fundamental theorem of calculus along the consensus segment gives

$$A = \tau \int_0^1 \nabla\Phi(\mathbf{y} + \tau\mathbf{p} + a\tau\mathbf{h})^\top \mathbf{h}\, da.$$

Add and subtract $\tau\nabla\Phi(\mathbf{y})^\top\mathbf{h}$. The remainder is

$$B = \tau \int_0^1 [\nabla\Phi(\mathbf{y} + \tau\mathbf{p} + a\tau\mathbf{h}) - \nabla\Phi(\mathbf{y})]^\top \mathbf{h}\, da.$$

By Cauchy–Schwarz and L-smoothness,

$$\begin{aligned} |B| &\le \tau L \int_0^1 \|\tau\mathbf{p} + a\tau\mathbf{h}\|\|\mathbf{h}\|\, da \\ &\le L\tau^2 \left(\|\mathbf{p}\|\|\mathbf{h}\| + \frac{1}{2}\|\mathbf{h}\|^2\right). \end{aligned}$$

Therefore

$$A \ge \tau\nabla\Phi(\mathbf{y})^\top\mathbf{h} - L\tau^2 \left(\|\mathbf{p}\|\|\mathbf{h}\| + \frac{1}{2}\|\mathbf{h}\|^2\right).$$

Taking conditional expectations gives $\mathbb{E}[\nabla\Phi(\mathbf{y})^\top\mathbf{h} \mid \mathcal{F}_t] = s$ and the second term equals $L\tau^2\mathcal{M}$. This proves equation 13; its right-hand side is positive when $s > L\tau\mathcal{M}$. □

A.5 Proof of Corollary 1

Proof. Suppress t on all quantities. The pathwise inequality preceding the conditional expectation in Theorem 1, applied with $\mathbf{h} = \bar{\rho}\,\mathbf{d}$, gives

$$\begin{aligned} &\Phi(\mathbf{y} + \tau\mathbf{p} + \tau\bar{\rho}\,\mathbf{d}) - \Phi(\mathbf{y} + \tau\mathbf{p}) \\ &\quad \ge \tau\bar{\rho}\langle\mathbf{R}, \mathbf{d}\rangle - L\tau^2 \left(\bar{\rho}\|\mathbf{p}\|\|\mathbf{d}\| + \frac{\bar{\rho}^2}{2}\|\mathbf{d}\|^2\right) \\ &\quad = \tau\bar{\rho} \left(\Gamma - \frac{L\tau\bar{\rho}}{2}\|\mathbf{d}\|^2\right). \end{aligned}$$

Setting $\bar{\rho} = \rho$ gives equation 14, which is positive under the stated rate condition. As a function of $\bar{\rho}$, the lower bound is a concave quadratic whose unconstrained maximizer is $\Gamma/(L\tau\|\mathbf{d}\|^2)$. Projection onto $[0, \rho_{\max}]$ gives the stated ρ^{lb}. □

A.6 Proof & Diagnostic for Corollary 2

Proof. Suppress t on all quantities and let $\mathbf{y} = \mathbf{y}_t$, $\mathbf{u}_I = \mathbf{u}^{\mathrm{igc}}$, $\mathbf{u}_0 = \mathbf{u}^{\mathrm{iso}}$, and $\omega = \omega_t$. By construction,

$$\mathbf{y}_{t+1} = \mathbf{y} + \omega\mathbf{u}_I, \quad \mathbf{y}_{t+1}^{\mathrm{iso,eq}} = \mathbf{y} + \omega\mathbf{u}_0, \quad \|\omega\mathbf{u}_I\| = \|\omega\mathbf{u}_0\| = \omega.$$

Thus, the comparison holds the total joint-state displacement fixed. The lower and upper forms of the descent lemma give

$$\Phi(\mathbf{y} + \omega\mathbf{u}_I) \ge \Phi(\mathbf{y}) + \omega\langle\mathbf{R}, \mathbf{u}_I\rangle - \frac{L\omega^2}{2},$$

and

$$\Phi(\mathbf{y} + \omega\mathbf{u}_0) \le \Phi(\mathbf{y}) + \omega\langle\mathbf{R}, \mathbf{u}_0\rangle + \frac{L\omega^2}{2}.$$

Subtracting gives

$$\Phi(\mathbf{y} + \omega\mathbf{u}_I) - \Phi(\mathbf{y} + \omega\mathbf{u}_0) \ge \omega\langle\mathbf{R}, \mathbf{u}_I - \mathbf{u}_0\rangle - L\omega^2 = \omega(\chi - L\omega),$$

which is strictly positive when $\chi > 0$ and $0 < \omega < \chi/L$.

If $\mathbf{h} = \bar{\rho}\,\mathbf{p}$ with $\bar{\rho} \ge 0$, then $\mathbf{p} + \mathbf{h} = (1 + \bar{\rho})\mathbf{p}$ and hence $\mathbf{u}_I = \mathbf{u}_0$, so $\chi = 0$ and the directional comparison is unchanged by pure rescaling. □

The condition $\chi_t > 0$ has a finite-sample diagnostic using the surrogate-gradient estimator from Appendix A.2. Let $\mathbf{v}_t = \mathbf{u}_t^{\mathrm{igc}} - \mathbf{u}_t^{\mathrm{iso}}$ and suppose $\|\hat{\mathbf{R}}_t - \mathbf{R}_t\| \le e_{r,t}$. Then

$$\chi_t = \langle \mathbf{R}_t, \mathbf{v}_t \rangle \ge \langle \hat{\mathbf{R}}_t, \mathbf{v}_t \rangle - e_{r,t}\|\mathbf{v}_t\|.$$

Consequently, $\langle \hat{\mathbf{R}}_t, \mathbf{v}_t \rangle > e_{r,t}\|\mathbf{v}_t\|$ certifies the required positive directional improvement. This diagnostic is defined relative to the realized finite-sample isolated direction: if $\mathbf{u}_t^{\mathrm{iso}} = \mathbf{R}_t/\|\mathbf{R}_t\|$, then Cauchy–Schwarz gives $\chi_t \le 0$ for every unit IGC direction.

A.7 Conditions for Theorem 2

Let a convex neighborhood $\mathcal{U}$ contain all iterates over the T updates. For $L, c, \underline{h} > 0$ and $\bar{h}, \nu \ge 0$, let Φ_t be L-smooth and $\Phi_t^* = \sup_{\mathbf{y}\in\mathcal{U}} \Phi_t(\mathbf{y})$. The local PL condition is $\|\mathbf{R}_t\|^2 \ge 2cG_t$. The full isolated direction, including its sample-dependent controls, obeys

$$\langle \mathbf{R}_t, \mathbb{E}[\mathbf{p}_t \mid \mathcal{F}_t] \rangle \ge \underline{h}\|\mathbf{R}_t\|^2, \quad \mathbb{E}[\|\mathbf{p}_t\|^2 \mid \mathcal{F}_t] \le \bar{h}^2\|\mathbf{R}_t\|^2 + \nu,$$

and the net consensus margin satisfies $\delta_t > 0$. Finally, the refreshed rank utility satisfies $\mathbb{E}[\|u_{t+1} - u_t\|_\infty \mid \mathcal{F}_t] \le v_t$ for deterministic $v_t \ge 0$.

The constants are uniform over the analyzed finite horizon and region, including the evolving distribution scales. The PL condition converts gradient progress into gap contraction, the two moment bounds characterize alignment and finite-sample variability of the isolated update, $\delta_t > 0$ supplies the additional IGC gain, and v_t controls variation when the rank surrogate is refreshed.

A.8 Proof of Theorem 2

Proof. Write Φ_t for the current frozen-rank surrogate. Applying L-smoothness to the isolated counterfactual gives

$$\Phi_t(\mathbf{y}_t + \tau_t\mathbf{p}_t) - \Phi_t(\mathbf{y}_t) \ge \tau_t\mathbf{R}_t^\top\mathbf{p}_t - \frac{L\tau_t^2}{2}\|\mathbf{p}_t\|^2.$$

Taking conditional expectations and applying the two base-direction moment bounds gives

$$\mathbb{E}[\Delta_t^{\mathrm{iso}} \mid \mathcal{F}_t] \ge \tau_t\left(\underline{h} - \frac{L\tau_t\bar{h}^2}{2}\right)\|\mathbf{R}_t\|^2 - \frac{L\tau_t^2\nu}{2}.$$

The PL condition $\|\mathbf{R}_t\|^2 \ge 2cG_t$ and the definition of κ imply

$$\mathbb{E}[\Delta_t^{\mathrm{iso}} \mid \mathcal{F}_t] \ge \kappa G_t - \frac{L\tau_t^2\nu}{2}.$$

Theorem 1 and the definition of δ_t give the same-state consensus gain

$$\begin{aligned}&\mathbb{E}[\Phi_t(\mathbf{y}_{t+1}) - \Phi_t(\mathbf{y}_{t+1}^{\mathrm{iso}}) \mid \mathcal{F}_t] \\ &\quad \ge \tau_t\delta_t\|\mathbf{R}_t\|^2 \ge 2c\tau_t\delta_t G_t \ge \gamma G_t.\end{aligned} \tag{20}$$

Adding the isolated and consensus gains yields a bound on the post-update gap measured by Φ_t:

$$\mathbb{E}[\Phi_t^* - \Phi_t(\mathbf{y}_{t+1}) \mid \mathcal{F}_t] \le (1 - \kappa - \gamma)G_t + \frac{L\tau_t^2\nu}{2}.$$

Define $V_t = \sup_{\mathbf{y}\in\mathcal{U}} |\Phi_{t+1}(\mathbf{y}) - \Phi_t(\mathbf{y})|$. For each fixed $\mathbf{y}$, both surrogates integrate over the same distribution $P_{\boldsymbol{\theta}(\mathbf{y})}$, so

$$V_t \le \sup_{\mathbf{x}} |u_{t+1}(\mathbf{x}) - u_t(\mathbf{x})| = \|u_{t+1} - u_t\|_\infty.$$

This variation controls both the optimal value and the value at the new iterate:

$$\Phi_{t+1}^* \le \Phi_t^* + V_t, \quad -\Phi_{t+1}(\mathbf{y}_{t+1}) \le -\Phi_t(\mathbf{y}_{t+1}) + V_t.$$

Consequently, $G_{t+1} \le \Phi_t^* - \Phi_t(\mathbf{y}_{t+1}) + 2V_t$. Taking conditional expectations and using $\mathbb{E}[V_t \mid \mathcal{F}_t] \le \mathbb{E}[\|u_{t+1} - u_t\|_\infty \mid \mathcal{F}_t] \le v_t$ gives

$$\mathbb{E}[G_{t+1} \mid \mathcal{F}_t] \le (1 - \kappa - \gamma)G_t + \frac{L\tau_t^2\nu}{2} + 2v_t = \varpi G_t + e_t. \tag{21}$$

This immediately gives strict one-step contraction when $e_t < (\kappa + \gamma)G_t$. Taking total expectations in equation 21 gives $\mathbb{E}[G_{t+1}] \le \varpi\mathbb{E}[G_t] + e_t$. Repeated substitution yields

$$\mathbb{E}[G_1] \le \varpi\mathbb{E}[G_0] + e_0,$$
$$\mathbb{E}[G_2] \le \varpi^2\mathbb{E}[G_0] + \varpi e_0 + e_1,$$
$$\mathbb{E}[G_T] \le \varpi^T\mathbb{E}[G_0] + \sum_{t=0}^{T-1} \varpi^{T-1-t} e_t,$$

which is equation 17. Applying the same calculation to the isolated update omits the consensus-gain term in equation 20, yielding the worst-case factor $1 - \kappa$.

Finally, for the current surrogate at each actual iterate,

$$\Phi_t(\mathbf{y}_{t+1}) - \Phi_t(\mathbf{y}_t) = \Delta_t^{\text{iso}} + \Phi_t(\mathbf{y}_{t+1}) - \Phi_t(\mathbf{y}_{t+1}^{\text{iso}}).$$

Conditional expectation and equation 20 give

$$\mathbb{E}[\Phi_t(\mathbf{y}_{t+1}) - \Phi_t(\mathbf{y}_t) \mid \mathcal{F}_t] \ge \mathbb{E}[\Delta_t^{\text{iso}} \mid \mathcal{F}_t] + \gamma G_t.$$

Taking total expectations and summing this inequality over $t = 0, \dots, T-1$ proves equation 18. □

B Complete Algorithm

B.1 Procedure & Update Details

Algorithm A-1 details the iterative procedure. Candidates are interleaved across H passes: each sampled policy controls one all-task update per pass, and every update continues from the state left by the preceding candidate. The sample batch drawn after an update determines that transition's reward and becomes the next cache. Thus $\hat{\mathcal{J}}_m$ measures progress at the states reached during candidate m's visits, including the effects of preceding updates; equation 5 uses their trajectory-relative ranking for online adaptation.

State features. The scalar $q = \mathbf{a}^\top \mathbf{z}$, with $\mathbf{a} = \mathbf{1}/\sqrt{D}$, has a standard Gaussian distribution at every dimension. Paired with fitness rank, it supplies a compact directional feature: for a local linear objective with perturbation-space gradient $\mathbf{g}$, $\mathrm{Cov}(q, \mathbf{g}^\top \mathbf{z}) = \mathbf{a}^\top \mathbf{g}$. The encoder therefore captures projected rank structure, with directional resolution limited to $\mathbf{a}$. Evaluation progress is $b_t = \mathrm{FE}_t/\mathrm{FE}_{\max}$, using cumulative objective-evaluation counts and the total budget.

Coordination-group construction. Let $1 \le B \le K$ denote the maximum coordination-group size; in our experiments, $B = \min\{100, K\}$. We choose B as a large scalability cap rather than a small group-size hyperparameter: $B = K$ at ordinary task scales, and $B < K$ only when the task count exceeds the cap. GemNES partitions the K tasks according to their arrival order into $G = \lceil K/B \rceil$ balanced, disjoint groups $\{\mathcal{G}_g\}_{g=1}^G$ whose sizes differ by at most one. Each group maintains a meta-policy distribution $Q_{\boldsymbol{\xi}_g} = \mathcal{N}(\mathbf{w}_g, \mathrm{diag}(\boldsymbol{\sigma}^2_{\text{meta},g}))$ and performs affinity filtering and IGC among its members. For task $k \in \mathcal{G}_g$, let $\iota_k = \ell_g(k)/|\mathcal{G}_g|$, where $\ell_g(k)$ is its local index in group g. For candidate m, group g uses

$$\hat{\mathcal{J}}_{g,m} = \frac{1}{|\mathcal{G}_g| H} \sum_{h=1}^{H} \sum_{k \in \mathcal{G}_g} \mathcal{R}_{k,m}^{(h)}, \tag{22}$$

and applies equation 5 to $(\mathbf{w}_g, \boldsymbol{\sigma}_{\text{meta},g})$. The candidate index m synchronizes trajectory segments across groups, whereas policy candidates and rewards remain group-specific. Within every (h, m) step, all groups finish their task updates before the generation advances, so each task receives one update per generation and each affinity computation involves at most B tasks. This cap restricts transfer to within-group peers; the synthetic and policy-search comparisons use full coordination with $K \le 100$.

For rank position i, let $a_i = \log(\lambda + 1) - \log i$, $\bar{a}_i = a_i^{\zeta_k} / \sum_j a_j^{\zeta_k}$, and $\bar{a}_{\text{med}} = \mathrm{med}_j(\bar{a}_j)$. The shaped rank weight used to estimate the base natural gradients is

$$w_{\zeta_k, i} = \frac{\max\{0, \bar{a}_i - \bar{a}_{\text{med}}\}}{\sum_l \max\{0, \bar{a}_l - \bar{a}_{\text{med}}\}} - \frac{1}{\lambda}.$$

Let $\hat{J}_k$ and $\hat{\sigma}_{J,k}$ be the objective mean and standard deviation of the new sample batch. After equation 4, the implementation updates

$$(\hat{J}_{\text{ema},k}^{(t+1)}, \hat{V}_{\text{ema},k}^{(t+1)}) = (1-\beta_{\text{ema}})(\hat{J}_{\text{ema},k}^{(t)}, \hat{V}_{\text{ema},k}^{(t)}) + \beta_{\text{ema}}(\hat{J}_k^{(t+1)}, (\hat{\sigma}_{J,k}^{(t+1)})^2 + \epsilon).$$

For descending candidate rank $\mathrm{r}_m^{\text{meta}}$, it uses $\bar{v}_m = \max(0, \log(M/2+1) - \log \mathrm{r}_m^{\text{meta}})$ and $v_m = \bar{v}_m / \sum_n \bar{v}_n - 1/M$ in equation 5.

Algorithm A-1: Geometric Meta-NES (GemNES)

Input: Number of tasks K, samples per task update λ, number of meta-candidates M, meta-horizon H, maximum coordination-group size $1 \le B \le K$.
Output: Optimized task parameters $\{\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k\}_{k=1}^K$.

Partition $\{1,\dots,K\}$ into $G = \lceil K/B \rceil$ balanced groups $\{\mathcal{G}_g\}_{g=1}^G$;
Initialize base-task search distributions $\boldsymbol{\theta}_k = [\boldsymbol{\mu}_k; \boldsymbol{\sigma}_k]$ for all $k \in \{1,\dots,K\}$;
Initialize meta-policy location $\mathbf{w}_g$ and scale $\boldsymbol{\sigma}_{\text{meta},g}$ independently for every group g;
Antithetically sample and evaluate $\mathcal{S}_{k,0}$ for every task; set $t \leftarrow 0$;
Initialize $\hat{J}_{\text{ema},k}^{(0)} = \hat{J}_k^{(0)}$ and $\hat{V}_{\text{ema},k}^{(0)} = (\hat{\sigma}_{J,k}^{(0)})^2 + \epsilon$;
while *optimization budget is not exhausted* **do**
  Phase 1: Interleaved Meta-Level Evaluation
  for *each group* $g \in \{1,\dots,G\}$ **do**
    Sample antithetic $\{\boldsymbol{\epsilon}_{g,m}\}_{m=1}^M$ and set $\mathbf{w}_{g,m} = \mathbf{w}_g + \boldsymbol{\sigma}_{\text{meta},g} \odot \boldsymbol{\epsilon}_{g,m}$;
    Initialize every $\hat{\mathcal{J}}_{g,m} = 0$;
  end
  for *pass* $h = 1$ **to** H **do**
    for *each sampled meta-policy* m **do**
      for *each group* $g \in \{1,\dots,G\}$ **do**
        for *each task* $k \in \mathcal{G}_g$ **do**
          Extract $\mathbf{s}_k^{(t)}$ from cached $\mathcal{S}_{k,t}$;
          Decode $\eta_{\boldsymbol{\mu},k}, \eta_{\boldsymbol{\sigma},k}, \zeta_k, \rho_k = \Pi_{\mathbf{w}_{g,m}}([\mathbf{s}_k^{(t)}; b_t; \ell_g(k)/|\mathcal{G}_g|])$;
          Estimate the shaped local gradients;
        end
        Calculate group-local affinities and consensus via equation 6–equation 7;
      end
      for *each group* $g \in \{1,\dots,G\}$ **do**
        Advance every $\boldsymbol{\theta}_k$, $k \in \mathcal{G}_g$, via equation 8–equation 9;
        for *each task* $k \in \mathcal{G}_g$ **do**
          Antithetically sample and evaluate $\mathcal{S}_{k,t+1}$;
          Compute $\mathcal{R}_{k,m}^{(h)}$ via equation 4 and update the EMA moments above;
          Replace the cache by $\mathcal{S}_{k,t+1}$;
        end
        $\hat{\mathcal{J}}_{g,m} \leftarrow \hat{\mathcal{J}}_{g,m} + \frac{1}{|\mathcal{G}_g| H} \sum_{k \in \mathcal{G}_g} \mathcal{R}_{k,m}^{(h)}$;
      end
      Complete the synchronized all-task generation; set $t \leftarrow t+1$;
    end
  end
  Phase 2: Meta-Policy Update
  for *each group* $g \in \{1,\dots,G\}$ **do**
    Compute centered rank utilities $v_{g,m}$ from $\{\hat{\mathcal{J}}_{g,m}\}_{m=1}^M$;
    Update $(\mathbf{w}_g, \boldsymbol{\sigma}_{\text{meta},g})$ via the group-indexed form of equation 5;
  end
end

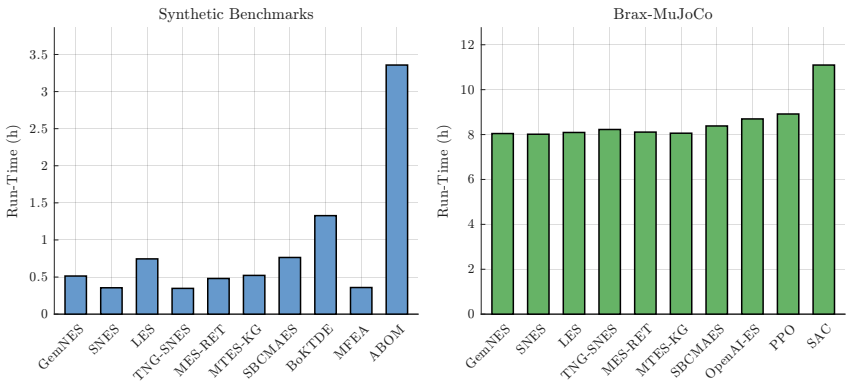


Figure A-1: Mean end-to-end runtime over repeated independent runs. (a) Synthetic many-task benchmarks. (b) Brax-MuJoCo policy search. Both are measured in hours.

B.2 Complexity Analysis

Let K denote the number of concurrent tasks, λ be the sample size per task, $D = \max_{1 \le k \le K} D_k$ represent the maximum search space dimensionality, M indicate the number of perturbed candidate policies, $1 \le B \le K$ be the maximum coordination-group size, $G = \lceil K/B \rceil$, and $|\mathbf{w}|$ be the parameter count of one meta-policy. Write $K_g = |\mathcal{G}_g|$, so $\sum_{g=1}^{G} K_g = K$ and $K_g \le B$.

Excluding the problem-dependent cost of black-box objective evaluation, one fixed (h, m) update requires $\mathcal{O}(K\lambda D + K\lambda \log \lambda)$ for sampling, sorting, state embedding, and natural-gradient estimation, while executing the selected group policies costs $\mathcal{O}(K|\mathbf{w}|)$. Pairwise affinity evaluation and IGC cost

$$\mathcal{O}\left(D \sum_{g=1}^{G} K_g^2\right),$$

because transfer is restricted to each coordination group. The reward uses the same newly evaluated sample batches, so it adds no objective evaluations.

Thus one synchronized all-task base-update step costs

$$\mathcal{O}\left(K\lambda D + K\lambda \log \lambda + D \sum_{g=1}^{G} K_g^2 + K|\mathbf{w}|\right).$$

Using $\sum_g K_g^2 \le B \sum_g K_g = BK$, this is bounded by

$$\mathcal{O}(K\lambda D + K\lambda \log \lambda + BKD + K|\mathbf{w}|),$$

and a complete outer meta-cycle costs

$$\mathcal{O}(MH(K\lambda D + K\lambda \log \lambda + BKD + K|\mathbf{w}|) + GM|\mathbf{w}|),$$

where the final term covers sampling and mean–scale aggregation for the G meta-policy distributions. With $G = 1$, the coordination term is $\mathcal{O}(K^2 D)$; when a very large task set exceeds the cap and $B < K$, it becomes $\mathcal{O}(BKD)$ and is linear in the total number of tasks for fixed B.

The mean end-to-end runtimes in Figure A-1 complement these arithmetic counts, while Section E examines runtime as the number of tasks increases. The synthetic and Brax-MuJoCo experiments instantiate $G = 1$ with $K = 50$ and $K = 10$ tasks, respectively. On the synthetic benchmarks (Figure A-1(a)), GemNES requires an average of 0.51 hours (1,850 s) per run, within the ES-baseline range of 0.35–0.76 hours. Its runtime is substantially lower than those of BoKTDE (1.33 hours; 4,780 s) and ABOM (3.36 hours; 12,088 s). On Brax-MuJoCo (Figure A-1(b)), GemNES takes 8.04 hours (28,963 s) per run, close to SNES at 8.01 hours (28,853 s) and LES at 8.09 hours (29,135 s). These measurements show that GemNES maintains a runtime comparable to the ES baselines in both scenarios while achieving the best overall optimization rank.

C Experimental Details

C.1 Task Settings

We employ two distinct experimental scenarios to evaluate the performance of GemNES:

Synthetic Benchmarks (CEC19 & WCCI22 MaTSO). We utilize 16 problem instances from the CEC19 and WCCI22 Many-Task Single-Objective Optimization (MaTSO) competitions (Feng et al., 2019; 2022) in MToP (Li et al., 2026a). Each problem consists of $K = 50$ concurrent tasks. The search space dimensionality D_k for each task is 50. For these benchmarks, we draw $\lambda = 50$ samples per task update. The optimization budget is set to $50,000$ function evaluations per task. The shifted, rotated, and multimodal instances cover diverse global-search and inter-task-transfer settings. Table A-II summarizes the landscape composition of these instances.

Table A-II: Landscape composition of the synthetic MaTSO benchmarks. Each instance contains 50 shifted and rotated 50-dimensional tasks; mixed WCCI22 instances assign the listed landscapes cyclically across tasks. CEC19 sets inter-task similarity according to the landscape function, whereas increasing WCCI22 instance indices correspond to greater shift separation and lower inter-task landscape similarity. The function-specific search domains are $[-100, 100]^{50}$ for Sphere and Griewank, $[-50, 50]^{50}$ for Rosenbrock, Ackley, and Rastrigin, $[-0.5, 0.5]^{50}$ for Weierstrass, and $[-500, 500]^{50}$ for Schwefel.

Suite	Instance	Landscape function(s)	Composition
CEC19	MaTSO1	Rosenbrock	Single
	MaTSO2	Ackley	Single
	MaTSO3	Rastrigin	Single
	MaTSO4	Griewank	Single
	MaTSO5	Weierstrass	Single
	MaTSO6	Schwefel	Single
WCCI22	MaTSO1	Sphere	Single
	MaTSO2	Rosenbrock	Single
	MaTSO3	Rastrigin	Single
	MaTSO4	Sphere, Rosenbrock, Ackley	Cyclic mixture
	MaTSO5	Rastrigin, Griewank, Weierstrass	Cyclic mixture
	MaTSO6	Rosenbrock, Griewank, Schwefel	Cyclic mixture
	MaTSO7	Ackley, Rastrigin, Weierstrass	Cyclic mixture
	MaTSO8	Rosenbrock, Ackley, Rastrigin, Griewank, Weierstrass	Cyclic mixture
	MaTSO9	Rosenbrock, Ackley, Rastrigin, Griewank, Weierstrass, Schwefel	Cyclic mixture
	MaTSO10	Ackley, Rastrigin, Griewank, Weierstrass, Schwefel	Cyclic mixture

Neuroevolution Policy Search (Brax-MuJoCo). We further evaluate GemNES on 10 heterogeneous continuous control tasks from the MuJoCo physics engine, integrated via the Brax framework (Freeman et al., 2021). As detailed in Table A-III, each task requires evolving the weights of a policy network (an MLP with 4 hidden layers of 32 units each, using $\tanh$ activations). Parameters are serialized as the first hidden bias, hidden-to-hidden weights and biases, output weights and bias, and input weights. Equal indices over the common prefix define the cross-task mapping, with no transfer to unmatched suffixes. The fixed hidden blocks share layer positions; the variable-sized input and output blocks use a heuristic index correspondence. The dimensionality D_k varies significantly across tasks. For ES-based methods, one function evaluation corresponds to the mean return over 5 random rollouts, with a maximum horizon of 1,000 steps per episode. We draw $\lambda = 50$ samples per task update and allow $50,000$ evaluations per task for all ES baselines. PPO and SAC use their standard environment-timestep budgets of 5×10^7 and 10^7 steps, respectively, following the Brax configurations. Comparisons with PPO and SAC report terminal performance under these respective budgets.

C.2 Baselines

We compare GemNES against three categories of baselines:

Table A-III: Overview of the 10 Brax-MuJoCo reinforcement learning tasks. All tasks use 4×32 MLP policies with tanh activations, while their different observation and action spaces yield parameter dimensions (D_k) from 3,361 to 11,569, spanning compact control problems and high-dimensional humanoid locomotion.

Task	Environment	Obs Dim	Act Dim	Parameter Dim (D_k)
T1	InvertedPendulum	4	1	3,361
T2	InvertedDoublePendulum	8	1	3,489
T3	Reacher	11	2	3,618
T4	Pusher	23	7	4,167
T5	Swimmer	8	2	3,522
T6	HalfCheetah	17	6	3,942
T7	Hopper	11	3	3,651
T8	Walker2d	17	6	3,942
T9	Ant	27	8	4,328
T10	Humanoid	244	17	11,569

Evolution Strategies. **SNES** (Schaul et al., 2011) is a standard single-task NES using a separable Gaussian distribution. **LES** (Lange et al., 2023b) is an offline meta-learned, single-task ES. It uses sample-order-invariant self-attention to produce recombination weights and an evolution-path-conditioned network to modulate per-dimension location and scale learning rates. **TNG-SNES** (Li et al., 2025b) is a many-task SNES that directly transfers averaged natural gradients among tasks. **MES-RET** (Li et al., 2026b) is a recently proposed many-task CMA-ES using reward-weighted evaluation and transfer to reduce negative transfer as the number of tasks grows. **MTES-KG** (Li et al., 2024b) is a many-task CMA-ES with knowledge-guided external sampling. **SBCMAES** (Liaw & Ting, 2019) is a symbiosis-inspired many-task CMA-ES. Following their original settings, these three methods use separable CMA-ES with diagonal covariance for high-dimensional policy search ($D_k > 3000$), while retaining full CMA-ES on the synthetic benchmarks ($D_k = 50$). **OpenAI-ES** (Salimans et al., 2017) is a high-performance ES variant widely used in reinforcement learning, utilizing antithetic sampling and virtual batch normalization.

Population-based Evolutionary Algorithms. **BoKTDE** (Jiang et al., 2023) is a many-task evolutionary algorithm using bi-objective similarity measurement. **MFEA** (Gupta et al., 2016) is the first evolutionary multitasking algorithm. It transfers genetic material through cross-task mating and inherits task assignments from parents. **ABOM** (Wang et al., 2026a) is a recently proposed online meta-optimizer that adapts evolutionary operators from target-task populations. We do not evaluate these non-distributional population-based methods on policy search because adapting their representations and evolutionary operators to neural policies requires additional method-specific design not provided by their original implementations. They are therefore evaluated only on the synthetic benchmarks.

Gradient-Based Reinforcement Learning Methods. **PPO** (Schulman et al., 2017) is an on-policy method using objective clipping, whereas **SAC** (Haarnoja et al., 2018) is an off-policy maximum-entropy method. Both implementations are provided by the Brax library (Freeman et al., 2021).

The hyperparameters of GemNES and baselines are summarized in Table A-IV.

Ablation Controls. **w/o-meta** freezes all meta-controlled settings throughout optimization. Each task retains one continuous search trajectory under the same evaluation budget. **w/o-transfer** retains meta-adaptation of the other controls but sets $\rho = 0$ to remove local IGC, and **w/o-similarity** aggregates all peers without the positive-affinity filter. The single-control variants use $(s_{\boldsymbol{\mu}}, s_{\boldsymbol{\sigma}}) = (1, 1)$ for **w/o-ada-stepsize**, $\zeta = 1$ for **w/o-ada-shaping**, and the fixed reference $\rho = 0.25$ for **w/o-ada-transfer**; all remaining controls continue to be produced by the meta-policy.

Table A-IV: Hyperparameter settings for GemNES and baselines.

Algorithm	Hyperparameter	Value
GemNES	Meta-optimizer	SNES ($M = 10, \boldsymbol{\sigma}^{(0)}_{\text{meta}} = 0.1 \cdot \mathbf{1}, \eta_{\mathbf{w}} = 1.0, \eta_{\boldsymbol{\sigma},\text{meta}} = 0.1, H = 2, \beta_{\text{ema}} = 0.2$)
	Meta-policy architecture	Encoder (2-8-4, max pooling); policy (6-8-8-4); tanh activations and cubed tanh output
	Transfer parameters	$\rho_{\max} = 0.5, \alpha = 0.2$ (Synthetic), $\alpha = 1.0$ (Policy Search)
	Base-level optimizer	SNES ($\eta^{\text{base}}_{\boldsymbol{\mu}} = 1.0, \eta^{\text{base}}_{\boldsymbol{\sigma}} = \frac{3+\log(D)}{5\sqrt{D}}$)
SNES	Learning rates	$\eta_{\boldsymbol{\mu}} = 1.0, \eta_{\boldsymbol{\sigma}} = \frac{3+\log(D)}{5\sqrt{D}}$
LES	Offline learned model	checkpoint `2023_03_les_v1.pkl`
TNG-SNES	Transfer parameters	$\rho_0 = 0.1, \alpha_0 = 0.7$, gap $= 100$
	Optimizer	SNES ($\eta_{\boldsymbol{\mu}} = 1.0, \eta_{\boldsymbol{\sigma}} = \frac{3+\log(D)}{5\sqrt{D}}$)
MES-RET	Transfer parameters	$\tau = 2$
	Optimizer	CMA-ES (Synthetic), separable CMA-ES (Policy Search)
MTES-KG	Transfer parameters	$\tau_0 = 2, \alpha = 0.5$, adjGap $= 50$
	Optimizer	CMA-ES (Synthetic), separable CMA-ES (Policy Search)
SBCMAES	Transfer parameters	Benefit $= 0.25$, Harm $= 0.5$
	Optimizer	CMA-ES (Synthetic), separable CMA-ES (Policy Search)
OpenAI-ES	Optimizer	Adam ($lr = 0.01, \beta_1 = 0.9, \beta_2 = 0.999$)
BoKTDE	Transfer parameters	$\Sigma = 0.9$
	Optimizer	Differential evolution ($F = 0.5, Cr = 0.7$)
MFEA	Transfer parameters	$RMP = 0.3$
	Optimizer	Simulated binary crossover and polynomial mutation ($\mu_C = 2, \mu_M = 5$)
ABOM	Online meta-optimizer	$lr = 10^{-3}, wd = 10^{-5}, p_C = p_M = 0.95, d_A = D, d_M = 2^{\lfloor \log_2 D \rfloor}$
PPO	Optimizer	envs $= 2048$, batch $= 1024$, minibatch $= 32$, $lr = 3 \times 10^{-4}, \gamma = 0.99$, ent $= 10^{-3}$
SAC	Optimizer	envs $= 128$, batch $= 256$, min_replay $= 8192$, $lr = 3 \times 10^{-4}, \gamma = 0.99$, scale $= 0.1$
Common	ES initial step size	$\sigma_0 = 0.3(U - L)$; policy weights use nominal $(L, U) = (-1, 1)$ only for initialization
	CMA learning rates	$c_1 = \frac{2}{(D+1.3)^2+\mu_{\text{eff}}}, c_\mu = \min(1 - c_1,\ 2 \cdot \frac{\mu_{\text{eff}}-2+1/\mu_{\text{eff}}}{(D+2)^2+\mu_{\text{eff}}})$
	Separable CMA learning rates	$c_{\text{cov}} = \frac{1}{\mu_{\text{eff}}} \cdot \frac{2}{(D+\sqrt{2})^2} + \left(1 - \frac{1}{\mu_{\text{eff}}}\right) \cdot \min\left(1,\ \frac{2\mu_{\text{eff}}-1}{(D+2)^2+\mu_{\text{eff}}}\right)$
	CMA parameters	$c_c = \frac{4+\mu_{\text{eff}}/D}{4+D+2\mu_{\text{eff}}/D}, c_\sigma = \frac{\mu_{\text{eff}}+2}{D+\mu_{\text{eff}}+5}, d_\sigma = 1 + c_\sigma + 2 \cdot \max\left(\sqrt{\frac{\mu_{\text{eff}}-1}{D+1}},\ 0\right)$.

C.3 Performance Metrics

Synthetic benchmark results use **30 independent runs**. Following Li et al. (2025b), the unified objective value for an algorithm is

$$\text{UOV} = \frac{1}{K} \sum_{i=1}^{K} \frac{f_i^* - f_i^{\min}}{f_i^{\max} - f_i^{\min}},$$

where f_i^* is the best objective value achieved by the current algorithm on task i, and $f_i^{\max}$ and $f_i^{\min}$ are the maximum and minimum objective values attained by the comparison set over all generations on that task. Thus UOV is relative to the algorithms included in each experiment. The baseline, ablation, and sensitivity experiments construct their normalizers from their respective comparison sets and use identical seed indices across algorithms.

The Brax-MuJoCo policy search tasks use **10 independent runs**, and each candidate's fitness is the mean cumulative reward over **5 independent rollouts**. To account for noisy returns, the incumbent policy is reevaluated with five fresh rollouts at each generation; tables report its terminal reevaluation. Shaded bands show the run mean ± 1.96 standard errors.

We employ two non-parametric statistical tests: the two-sided **Wilcoxon rank-sum post-hoc test** ($\alpha = 0.05$) compares GemNES with each baseline per task, while the **Friedman test** gives the overall ranking across tasks. The symbol $=$ denotes a non-significant difference, not statistical equivalence. We also report the **number of best results** (#Best or #B).

Pairwise reward comparisons cover all methods. Friedman rankings and #B statistics cover the ES methods evaluated under a common rollout budget (50,000 evaluations $\times$ 5 rollouts per task); PPO and SAC results use environment-timestep budgets.

All experiments are conducted on a Mac mini equipped with an Apple M4 Pro chip and 64 GB of RAM. The framework is implemented in MATLAB and Python, with rollout evaluations parallelized across CPU cores.

D Parameter Sensitivity Analysis

We analyze the sensitivity of GemNES to the number of meta-candidates M, meta-horizon H, meta-policy architecture (hidden width d_{hidden} and latent dimension d_{latent}), maximum transfer rate $\rho_{\max}$, and suppression coefficient α. These sweeps characterize parameter sensitivity on the controlled synthetic landscapes of CEC19 and WCCI22. Policy search provides a separate high-dimensional application setting, with component ablations reported in Section 4.2.

Table A-V summarizes Friedman rankings and two-sided Wilcoxon rank-sum *post-hoc* p-values across the 16 instances against the marked reference. Detailed results appear in Tables A-XII–A-XVI; all post-hoc p-values exceed 0.05, indicating limited detected sensitivity across the tested configurations.

Table A-V: Friedman rankings and two-sided Wilcoxon rank-sum *post-hoc* p-values for hyperparameter configurations on the synthetic benchmarks; Ref. marks the significance-test reference.

Hyperparameter	Value	Ranking	p-value
Number of meta-candidates M	6	2.88	0.9110
	8	2.88	0.9110
	10	**2.81**	**Ref.**
	20	3.12	0.5762
	40	3.31	0.3711
Meta-horizon H	1	**2.62**	0.2636
	2	3.12	0.8231
	3	3.25	**Ref.**
	4	3.12	0.8231
	5	2.88	0.5023
Meta-policy architecture $(d_{\text{hidden}}, d_{\text{latent}})$	(2, 1)	3.19	0.3143
	(4, 2)	2.75	0.8231
	(8, 4)	**2.62**	**Ref.**
	(16, 8)	3.12	0.3711
	(32, 16)	3.31	0.2188
Maximum transfer rate $\rho_{\max}$	0.1	2.62	0.7373
	0.3	**2.50**	0.5762
	0.5	2.81	**Ref.**
	0.7	3.44	0.2636
	0.9	3.62	0.1461
Penalty coefficient α	0.0	3.62	0.0573
	0.1	3.31	0.1797
	0.2	**2.56**	**Ref.**
	0.5	2.94	0.5023
	1.0	**2.56**	1.0000

- **Number of meta-candidates** M: This parameter controls how many perturbed policies are sampled for each joint mean–scale natural-gradient update. The default $M = 10$ achieves the lowest Friedman rank (2.81).
- **Meta-horizon** H: This parameter is the number of interleaved visits per candidate within a meta-cycle, yielding MH all-task updates before each meta-policy update. $H = 1$ yields the lowest rank (2.62), and the tested values show no significant difference from the reference. The main experiments use $H = 2$; $H = 3$ is the significance-test reference for this sweep.
- **Meta-policy architecture** $(d_{\text{hidden}}, d_{\text{latent}})$: The hidden width d_{hidden} and latent dimension d_{latent} govern the representational capacity of the meta-policy. The default $(8, 4)$ achieves the lowest rank (2.62).

- **Maximum gradient transfer rate** $\rho_{\max}$: This hyperparameter bounds the intensity of IGC injection. Smaller values have better ranks, with $\rho_{\max} = 0.3$ attaining the lowest rank (2.50). The deterioration at 0.7 and 0.9 is qualitatively consistent with Corollary 1.
- **Penalty coefficient** α: This parameter controls drift-coupled scale suppression. On synthetic benchmarks, $\alpha = 0.2$ and 1.0 share the lowest rank (2.56), with no significant difference detected ($p = 1.0000$). We therefore retain $\alpha = 0.2$ as the reference setting for synthetic tasks and use $\alpha = 1.0$ to strengthen scale suppression in high-dimensional policy search. Removing suppression ($\alpha = 0$) yields a worse rank (3.62); its comparison with the $\alpha = 0.2$ reference gives $p = 0.0573$, close to but above the 0.05 significance threshold.

E Task-Scale Efficiency Analysis

In the planar kinematic arm control problem (PKACP) (Jiang et al., 2023), task k uses a joint-angle vector $\mathbf{x} \in [0, 1]^D$. To generate K tasks, $50K$ points sampled uniformly from $[0, 1]^2$ are partitioned by K-means; centroid (A_k, L_k) controls the angular range and total arm length of task k. With cumulative heading $\Theta_{k,d}$ and equal link length L_k/D, the distance from the end effector to the fixed target $[0.5; 0.5]$ is

$$\Theta_{k,d} = \frac{2\pi A_k}{D} \sum_{j=1}^{d} (x_j - \tfrac{1}{2}), \qquad f_k(\mathbf{x}) = \left\| \frac{L_k}{D} \sum_{d=1}^{D} \begin{bmatrix} \cos \Theta_{k,d} \\ \sin \Theta_{k,d} \end{bmatrix} - [0.5; 0.5] \right\|_2 . \tag{23}$$

We fix $D = 20$. Evaluating a candidate requires D forward-kinematics updates and therefore costs $\mathcal{O}(D)$ time. Increasing K only generates additional centroids and leaves both dimension and per-task evaluation cost unchanged. PKACP is thus a low-dimensional and inexpensive setting in which the observed runtime growth primarily reflects optimizer-side task coordination. We test $K = 2^5, \ldots, 2^{10}$ using $\lambda = 20$ and $10,000$ evaluations per task, with shared seeds and hardware.

We compare GemNES with **w/o-group**, which coordinates all K tasks jointly, and isolated **SNES**. GemNES uses $B = \min\{100, K\}$ and $G = \lceil K/B \rceil$ balanced groups. Mean runtime and UOV over tasks and 30 independent runs measure computational scaling and the optimization effect of bounded coordination, respectively.

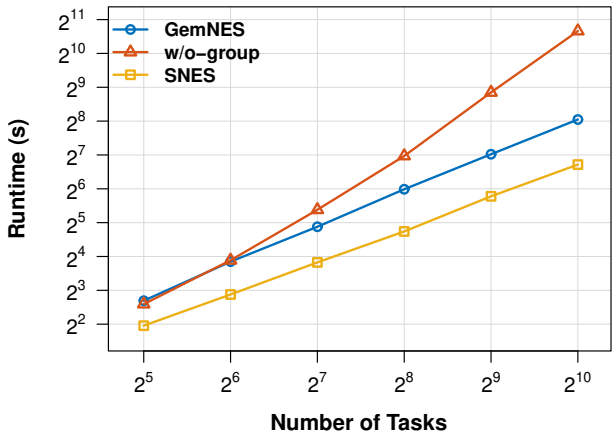


Figure A-2: Mean end-to-end runtime across 30 independent runs on the $D = 20$ PKACP as the number of tasks increases from 2^5 to 2^{10}. Both axes use logarithmic scales.

Runtime Scaling. Figure A-2 shows that GemNES and SNES follow a similar near-linear growth trend as K doubles, whereas the runtime of w/o-group grows increasingly rapidly once $K > B$. At $K = 1024$, grouping reduces the GemNES runtime from 1618.68 s to 264.58 s, a 6.12× speedup. SNES remains faster in absolute terms because it performs no cross-task coordination.

Performance Retention. Table A-VI gives GemNES and w/o-group the same Friedman rank of 1.50. Full coordination is significantly better only at $K = 128$ and shows no significant difference at

Table A-VI: Task-scale performance on PKACP. The table reports the mean and standard deviation of UOV across 30 independent runs. The signs (+ / − / =) indicate whether the compared variant is statistically better, worse, or not significantly different from GemNES according to the two-sided Wilcoxon rank-sum post-hoc test. Ranking is obtained by the Friedman test.

	GemNES	w/o-group	SNES
$K = 2^5$	**2.0277e-06 (1.31e-06)**	**2.0277e-06 (1.31e-06)** =	2.2449e-06 (9.43e-07) =
$K = 2^6$	**2.1499e-05 (1.71e-05)**	**2.1499e-05 (1.71e-05)** =	2.4729e-05 (1.28e-05) =
$K = 2^7$	3.5018e-06 (2.83e-06)	**2.3738e-06 (1.00e-06)** +	4.3147e-06 (1.26e-06) −
$K = 2^8$	**5.7215e-06 (3.05e-06)**	6.0115e-06 (3.63e-06) =	1.0206e-05 (4.00e-06) −
$K = 2^9$	**6.4833e-06 (3.22e-06)**	7.2212e-06 (4.14e-06) =	9.2584e-06 (3.33e-06) −
$K = 2^{10}$	5.7956e-06 (2.53e-06)	**5.6678e-06 (2.32e-06)** =	9.1037e-06 (2.72e-06) −
+ / − / =	Ours	1 / 0 / 5	0 / 4 / 2
Ranking	**1.50**	**1.50**	3.00

the other five scales, while SNES is significantly worse at the four largest scales. Bounded grouping therefore retains the optimization benefit of cross-task coordination while substantially improving efficiency for large task sets.

F Detailed Results & Supplementary Analysis

This section provides the full convergence curves and numerical tables with statistical comparisons, together with supplementary task-level analysis. We organize the material by scenario: synthetic benchmarks (Section F.1), policy search against ES baselines (Section F.2) and RL baselines (Section F.3), ablation studies (Section F.4), and parameter sensitivity details (Section F.5).

F.1 Synthetic Benchmarks

Figure A-3 illustrates the unified objective value convergence trajectories across all 16 synthetic MaTO benchmark instances, with detailed performance metrics summarized in Table A-VII.

Performance differences track landscape structure. Many-task ESs are strong on the smooth single-function WCCI22-MaTSO1–2, where task directions remain compatible and direct reuse is effective. The results on WCCI22-MaTSO4 and 8 also expose different inductive biases: full-covariance CMA-ES can model cross-variable dependence on the synthetic tasks, while direct-gradient transfer remains effective when alignment persists. GemNES is more reliable on most multimodal cyclic mixtures, where useful relationships vary with the search stage: global IGC adapts step sizes and selection pressure from online progress, while local IGC admits only aligned peers. CEC19-MaTSO6 is another informative exception: MFEA and ABOM achieve lower mean UOVs (0.0749 and 0.0390) than GemNES (0.103), with both differences significant. This suggests a benefit from their population-based operators on Schwefel. Across the full suite, GemNES outperforms MFEA on the remaining 15 instances and attains nine best mean results, with an overall rank of 2.31 versus 8.88 for MFEA.

F.2 Policy Search: Comparison with ES Baselines

Table A-VIII compares the mean cumulative reward of GemNES with the ES baselines across the 10 Brax-MuJoCo tasks, together with per-task Wilcoxon rank-sum post-hoc tests, the Friedman ranking, and the number of best results.

The task pattern distinguishes both outcomes and transfer mechanisms. After all ESs saturate the two pendulum tasks, GemNES leads *pusher*, *hopper*, *ant*, and *humanoid*; LES leads *reacher* and *swimmer*, MES-RET's reward weighting leads *halfcheetah*, and TNG-SNES leads *walker2d*. Only LES and MES-RET are better by the reported taskwise test, both on *swimmer*. TNG-SNES's direct averaging is vulnerable to conflicting directions, consistent with its lower results on *pusher* and *ant*, while the separable CMA variants cannot model cross-coordinate dependence. GemNES is also diagonal, but task-conditioned rates and affinity gating control when and from whom to borrow.

Table A-VII: Performance comparison on CEC19 and WCCI22 synthetic MaTO benchmark suites. The table reports the mean and standard deviation of the unified objective value across 30 independent runs. The signs (+ / − / =) indicate whether the compared algorithm is statistically better, worse, or not significantly different from GemNES according to the two-sided Wilcoxon rank-sum post-hoc test. Ranking is obtained by the Friedman test, and #Best denotes the number of best results achieved by each algorithm.

	GemNES	SNES	LES	TNG-SNES	MES-RET	MTES-KG	SBCMAES	BoKTDE	MFEA	ABOM
CEC19-MaTSO1	8.48e-10 (8.7e-11)	1.26e-09 (8.6e-11) −	2.68e-09 (4.9e-10) −	1.07e-09 (9.2e-11) −	**7.56e-10 (2.7e-10)** =	9.32e-10 (2.2e-10) −	1.06e-09 (5.0e-10) =	1.16e-08 (3.5e-09) −	3.57e-07 (2.3e-08) −	4.47e-08 (1.0e-08) −
CEC19-MaTSO2	**2.36e-15 (9.7e-15)**	5.89e-09 (8.6e-11) −	2.86e-03 (1.7e-05) −	3.80e-14 (7.5e-16) −	3.15e-15 (3.5e-17) −	2.68e-15 (3.5e-17) −	5.96e-01 (4.6e-01) −	9.41e-01 (4.1e-04) −	9.89e-03 (3.2e-04) −	3.45e-01 (6.2e-02) −
CEC19-MaTSO3	**2.33e-04 (7.7e-05)**	3.68e-04 (7.7e-05) −	7.21e-04 (3.3e-05) −	2.72e-04 (1.4e-05) −	4.58e-04 (2.3e-05) −	4.29e-04 (1.7e-05) −	4.04e-04 (1.9e-05) −	6.24e-03 (3.3e-05) −	1.36e-02 (3.5e-04) −	6.70e-03 (6.2e-05) −
CEC19-MaTSO4	2.77e-15 (1.5e-14)	5.07e-12 (1.8e-13) −	2.87e-05 (1.9e-06) −	**1.59e-19 (8.3e-19)** +	8.72e-08 (4.8e-07) −	1.91e-07 (7.3e-07) −	3.59e-07 (9.3e-07) =	9.21e-06 (5.2e-06) −	1.36e-02 (1.8e-04) −	1.34e-03 (6.2e-05) −
CEC19-MaTSO5	**4.25e-09 (1.7e-08)**	1.84e-05 (1.8e-07) −	9.03e-03 (3.4e-04) −	6.50e-09 (1.2e-08) −	1.06e-04 (1.1e-04) −	3.88e-04 (1.8e-04) −	5.53e-04 (2.7e-04) −	2.41e-03 (4.8e-04) −	4.76e-02 (5.1e-04) −	1.39e-02 (4.0e-04) −
CEC19-MaTSO6	1.03e-01 (1.4e-02)	2.20e-01 (3.4e-03) −	2.43e-01 (4.6e-03) −	2.52e-01 (5.0e-03) −	4.29e-01 (7.1e-03) −	5.30e-01 (7.7e-03) −	5.84e-01 (2.1e-03) −	6.10e-01 (1.9e-02) −	7.49e-02 (6.5e-03) +	**3.90e-02 (3.2e-03)** +
WCCI22-MaTSO1	7.22e-27 (3.8e-26)	5.89e-19 (1.5e-20) −	1.67e-07 (1.8e-09) −	2.52e-29 (9.8e-31) +	4.00e-31 (7.8e-33) +	**3.19e-31 (6.9e-33)** +	6.62e-31 (1.8e-32) +	6.50e-09 (7.6e-09) −	1.20e-04 (3.6e-06) −	1.61e-06 (9.3e-08) −
WCCI22-MaTSO2	2.07e-08 (1.2e-08)	8.73e-09 (4.1e-09) +	2.68e-08 (1.3e-08) −	1.18e-08 (4.2e-09) +	5.65e-10 (2.9e-10) +	**4.65e-10 (1.2e-10)** +	6.90e-10 (5.4e-10) +	9.27e-09 (4.1e-09) +	4.63e-07 (3.1e-07) −	4.13e-08 (1.4e-08) −
WCCI22-MaTSO3	**2.37e-04 (1.5e-04)**	3.75e-04 (6.5e-05) −	7.83e-04 (2.3e-05) −	2.97e-04 (1.9e-05) −	4.77e-04 (2.0e-05) −	4.41e-04 (1.9e-05) −	4.43e-04 (1.6e-05) −	6.53e-03 (4.1e-05) −	9.29e-03 (1.9e-04) −	6.64e-03 (1.7e-04) −
WCCI22-MaTSO4	7.27e-09 (6.0e-09)	4.91e-09 (1.7e-09) =	9.13e-04 (1.0e-05) −	4.63e-09 (2.9e-09) =	**3.26e-10 (7.3e-11)** +	6.84e-10 (1.2e-10) +	7.23e-10 (1.3e-10) +	1.15e-03 (8.6e-04) −	9.53e-02 (4.0e-03) −	1.86e-02 (1.3e-03) −
WCCI22-MaTSO5	**7.90e-05 (2.9e-05)**	1.18e-04 (3.8e-05) −	7.20e-03 (1.3e-03) −	8.16e-05 (6.7e-06) =	7.69e-04 (4.6e-04) −	1.08e-03 (4.8e-04) −	1.68e-03 (9.0e-04) −	1.69e-02 (1.6e-03) −	1.63e-01 (7.9e-03) −	7.01e-02 (4.1e-03) −
WCCI22-MaTSO6	**4.59e-02 (8.1e-03)**	9.84e-02 (4.6e-03) −	6.18e-02 (5.4e-03) −	1.04e-01 (3.0e-03) −	9.09e-02 (4.5e-03) −	1.33e-01 (4.2e-03) −	1.93e-01 (2.1e-02) −	1.89e-01 (1.0e-02) −	7.52e-02 (6.0e-03) −	1.14e-01 (1.7e-02) −
WCCI22-MaTSO7	**7.45e-05 (2.8e-05)**	1.19e-04 (3.5e-05) −	8.34e-03 (9.6e-04) −	8.79e-05 (7.6e-06) −	8.03e-04 (3.8e-04) −	1.18e-03 (6.5e-04) −	1.36e-03 (6.2e-04) −	1.80e-02 (2.3e-03) −	2.71e-01 (6.9e-03) −	9.07e-02 (3.9e-03) −
WCCI22-MaTSO8	6.10e-04 (7.9e-04)	1.64e-04 (1.5e-04) =	2.90e-02 (1.9e-02) −	**1.21e-04 (1.7e-04)** +	1.82e-03 (9.9e-04) −	3.15e-03 (1.0e-03) −	1.21e-02 (8.0e-03) −	2.54e-02 (5.8e-03) −	3.09e-01 (5.7e-03) −	6.33e-02 (5.5e-03) −
WCCI22-MaTSO9	**2.36e-02 (4.9e-03)**	4.67e-02 (3.2e-03) −	5.11e-02 (1.7e-02) −	4.91e-02 (3.2e-03) −	4.46e-02 (4.7e-03) −	6.40e-02 (3.7e-03) −	8.15e-02 (6.7e-03) −	1.10e-01 (8.5e-03) −	3.02e-01 (5.2e-03) −	9.82e-02 (1.4e-02) −
WCCI22-MaTSO10	**2.90e-02 (4.5e-03)**	5.71e-02 (2.8e-03) −	6.48e-02 (2.2e-02) −	6.07e-02 (3.2e-03) −	5.49e-02 (5.6e-03) −	7.94e-02 (3.9e-03) −	1.04e-01 (9.5e-03) −	1.40e-01 (1.4e-02) −	3.54e-01 (6.4e-03) −	1.24e-01 (1.7e-02) −
+ / − / =	Ours	1 / 13 / 2	0 / 16 / 0	4 / 10 / 2	3 / 12 / 1	3 / 13 / 0	3 / 11 / 2	1 / 15 / 0	1 / 15 / 0	1 / 15 / 0
Ranking	**2.31**	3.81	6.56	3.44	3.50	4.44	5.88	8.00	8.88	8.19
#Best	**9**	0	0	2	2	2	0	0	0	1

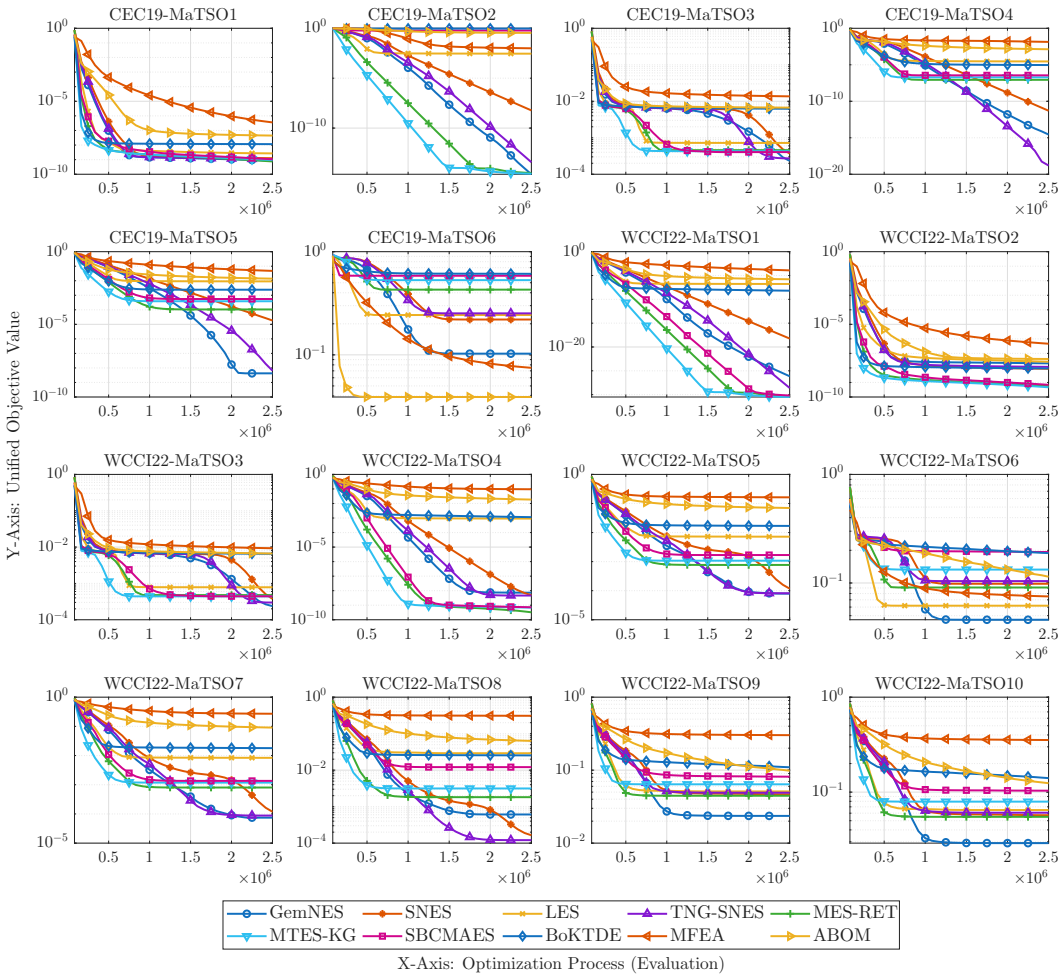


Figure A-3: Convergence curves of unified objective value ↓ on synthetic MaTO benchmark suites across 30 independent runs. Each problem instance contains 50 concurrent tasks.

LES progresses rapidly but declines late on *pusher*, *hopper*, and *walker2d*. GemNES sustains progress on *walker2d* and *ant*, while adapting update aggressiveness online and filtering diverging peers.

F.3 Policy Search: Comparison with RL Baselines

Table A-IX and Figure A-4 provide the detailed comparison with the gradient-based RL references PPO and SAC on the 10 Brax-MuJoCo tasks.

On the easier tasks, large return differences provide reliable rankings of sampled candidates, allowing GemNES to discover stable policies from episodic returns alone. Global IGC adapts exploration and selection pressure from online progress, while local IGC pools aligned task evidence. On *humanoid*, PPO and SAC attain higher rewards than GemNES. This comparison contrasts GemNES's rank-based episodic feedback with the transition-level actor–critic objectives used by PPO and SAC on long-horizon, high-dimensional control.

F.4 Ablation Studies

Detailed ablation results and trajectories are reported in Tables A-X and A-XI and Figure A-5.

The variants clarify the division of labor. **w/o-meta** freezes all meta-controlled settings. In **w/o-transfer**, each task estimates its direction from only 50 samples, leaving noisy long-horizon returns

Table A-VIII: Performance comparison on Brax-MuJoCo policy search tasks. The table reports the mean and standard deviation of the cumulative reward (averaged over 5 rollouts) across 10 independent runs. The signs (+ / − / =) indicate whether the compared algorithm is statistically better, worse, or not significantly different from GemNES according to the two-sided Wilcoxon rank-sum post-hoc test. Ranking is obtained by the Friedman test, and #Best denotes the number of best results achieved by each algorithm.

		GemNES	SNES	LES	TNG-SNES	MES-RET	MTES-KG	SBCMAES	OpenAI-ES
T1	i_pendulum	**1.0000e+03 (0.00e+00)**	**1.0000e+03 (0.00e+00)** =	**1.0000e+03 (0.00e+00)** =	**1.0000e+03 (0.00e+00)** =	**1.0000e+03 (0.00e+00)** =	**1.0000e+03 (0.00e+00)** =	**1.0000e+03 (0.00e+00)** =	**1.0000e+03 (0.00e+00)** =
T2	id_pendulum	**9.3600e+03 (8.04e-04)**	**9.3600e+03 (8.55e-04)** =	**9.3600e+03 (6.53e-03)** =	**9.3600e+03 (2.17e-03)** =	**9.3600e+03 (1.26e-03)** =	**9.3600e+03 (8.97e-04)** =	**9.3600e+03 (1.41e-02)** =	**9.3600e+03 (1.51e-03)** =
T3	reacher	-5.6105e+01 (8.45e+00)	-6.2883e+01 (8.45e+00) =	**-5.5471e+01 (1.42e+01)** =	-6.7079e+01 (1.41e+01) =	-6.2628e+01 (8.36e+00) −	-5.6314e+01 (1.03e+01) =	-8.0999e+01 (8.30e+00) −	-8.4997e+01 (1.43e+01) −
T4	pusher	**-1.8858e+02 (4.66e+01)**	-2.9830e+02 (3.52e+01) −	-2.3400e+02 (4.51e+01) −	-2.8927e+02 (4.19e+01) −	-2.1708e+02 (4.49e+01) =	-2.1447e+02 (3.16e+01) =	-4.6540e+02 (1.02e+02) −	-6.2689e+02 (4.40e+01) −
T5	swimmer	3.6774e+02 (1.39e+00)	3.6540e+02 (8.87e-01) −	**3.7097e+02 (7.69e-01)** +	3.6517e+02 (1.12e+00) −	3.6987e+02 (8.76e-01) +	3.6904e+02 (1.37e+00) =	3.6493e+02 (1.18e+00) −	3.4829e+02 (6.77e+01) −
T6	halfcheetah	7.5510e+03 (7.93e+02)	6.9914e+03 (1.47e+02) −	5.8308e+03 (1.30e+03) −	6.7601e+03 (7.89e+02) −	**1.0468e+04 (4.53e+03)** =	8.7175e+03 (2.64e+03) =	7.2762e+03 (1.29e+03) −	8.1779e+03 (4.14e+03) =
T7	hopper	**3.5743e+03 (7.36e+01)**	3.5246e+03 (1.78e+02) =	1.6031e+03 (7.77e+02) −	3.5099e+03 (1.75e+02) =	3.4985e+03 (1.34e+02) =	3.0522e+03 (2.01e+02) −	3.1911e+03 (4.65e+02) −	2.8401e+03 (1.21e+03) =
T8	walker2d	4.1671e+03 (1.83e+03)	3.1060e+03 (9.29e+02) =	1.7547e+03 (9.24e+02) −	**4.2010e+03 (1.84e+03)** =	3.6157e+03 (1.91e+03) =	3.0077e+03 (9.70e+02) =	3.2795e+03 (1.58e+03) =	3.5336e+03 (2.01e+03) =
T9	ant	**2.6583e+03 (6.91e+02)**	1.5115e+03 (3.28e+02) −	2.1249e+03 (5.26e+02) =	1.4958e+03 (3.73e+02) −	2.1522e+03 (7.13e+02) −	1.9647e+03 (2.97e+02) −	8.4846e+02 (2.62e+02) −	4.8140e+02 (1.40e+02) −
T10	humanoid	**6.1568e+02 (6.64e+01)**	6.0447e+02 (5.55e+01) =	5.9977e+02 (3.13e+01) =	5.8888e+02 (5.11e+01) =	5.4056e+02 (2.58e+01) −	5.4246e+02 (4.95e+01) −	5.3114e+02 (4.01e+01) −	5.5320e+02 (2.38e+01) −
	+ / − / =	Ours	0 / 4 / 6	1 / 4 / 5	0 / 4 / 6	1 / 3 / 6	0 / 3 / 7	0 / 7 / 3	0 / 5 / 5
	Ranking	**2.25**	4.50	4.65	4.75	3.40	3.85	6.35	6.25
	#Best	**6**	2	4	3	3	2	2	2

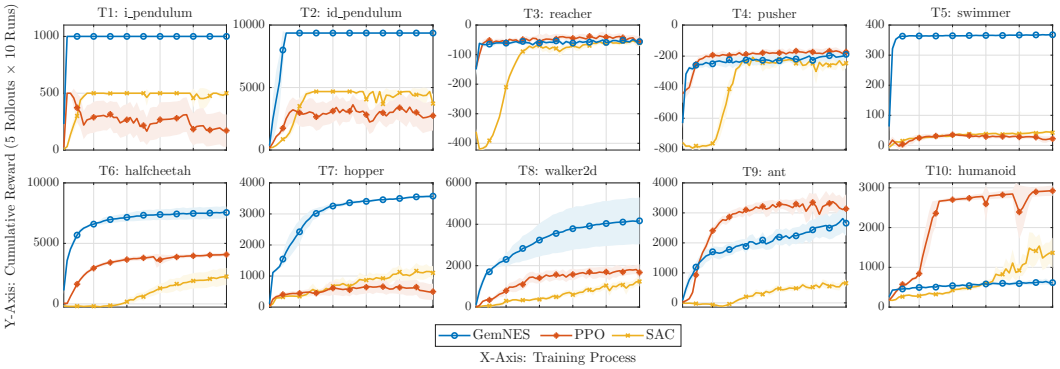


Figure A-4: Comparison of cumulative reward ↑ with gradient-based RL references under method-specific interaction budgets. Curves show each method's native training progress and are not a matched sample-efficiency comparison; shaded areas show the mean ±1.96 standard errors across 10 independent runs.

Table A-IX: Comparison with gradient-based RL methods on Brax-MuJoCo under method-specific interaction budgets. We report the mean and standard deviation of cumulative reward across 10 runs; GemNES evaluations average 5 rollouts, and signs reflect the two-sided Wilcoxon rank-sum post-hoc test against GemNES.

		GemNES	PPO	SAC
T1	i_pendulum	**1.0000e+03 (0.00e+00)**	1.7082e+02 (2.29e+02) –	5.0000e+02 (0.00e+00) –
T2	id_pendulum	**9.3600e+03 (8.04e-04)**	2.7451e+03 (1.91e+03) –	3.7172e+03 (1.25e+03) –
T3	reacher	-5.6105e+01 (8.45e+00)	**-5.3715e+01 (2.84e+01)** =	-6.0546e+01 (1.74e+01) =
T4	pusher	-1.8858e+02 (4.66e+01)	**-1.7306e+02 (2.57e+01)** =	-2.4553e+02 (5.12e+01) –
T5	swimmer	**3.6774e+02 (1.39e+00)**	2.2279e+01 (1.53e+01) –	4.2430e+01 (5.91e+00) –
T6	halfcheetah	**7.5510e+03 (7.93e+02)**	4.0791e+03 (2.57e+02) –	2.2579e+03 (1.15e+03) –
T7	hopper	**3.5743e+03 (7.36e+01)**	4.9617e+02 (4.19e+02) –	1.1021e+03 (3.56e+02) –
T8	walker2d	**4.1671e+03 (1.83e+03)**	1.6714e+03 (6.44e+02) –	1.2360e+03 (3.87e+02) –
T9	ant	2.6583e+03 (6.91e+02)	**3.1359e+03 (7.87e+02)** =	6.5572e+02 (1.88e+02) –
T10	humanoid	6.1568e+02 (6.64e+01)	**2.9257e+03 (2.12e+02)** +	1.3639e+03 (6.20e+02) +
	+/−/=	Ours	1/6/3	1/8/1

unregularized. Removing similarity filtering pools all peer directions. The resulting performance loss is significant on *ant*, a challenging 3D locomotion task whose quadrupedal morphology differs from the other agents in the suite. GemNES achieves a mean return of 2658.3 versus 2250.2 for **w/o-similarity** (Table A-XI). This gain is consistent with affinity filtering reducing negative transfer from less compatible peer directions during heterogeneous policy search.

Table A-X: Ablation study on CEC19 and WCCI22 synthetic MaTO benchmark suites. The table reports the mean and standard deviation of the unified objective value across 30 independent runs. The signs (+ / − / =) indicate whether the compared variant is statistically better, worse, or not significantly different from GemNES according to the two-sided Wilcoxon rank-sum post-hoc test. Ranking is obtained by the Friedman test, and #Best denotes the number of best results achieved by each algorithm.

	GemNES	w/o-meta	w/o-transfer	w/o-similarity	w/o-ada-stepsize	w/o-ada-shaping	w/o-ada-transfer
CEC19-MaTSO1	2.76e-10 (9.1e-11)	7.67e-10 (1.9e-10) –	3.69e-10 (8.9e-11) –	**2.65e-10 (1.1e-10)** =	4.85e-10 (6.8e-11) –	2.98e-10 (5.9e-11) =	2.93e-10 (8.1e-11) =
CEC19-MaTSO2	2.41e-15 (9.8e-15)	7.39e-12 (6.3e-13) –	1.09e-13 (5.9e-13) –	3.14e-15 (1.1e-14) =	1.10e-11 (3.8e-11) –	**2.89e-16 (6.2e-16)** =	7.07e-15 (2.7e-14) =
CEC19-MaTSO3	2.96e-04 (8.0e-05)	6.99e-04 (2.7e-05) –	3.28e-04 (2.4e-04) =	3.20e-04 (2.4e-04) =	3.27e-04 (1.9e-05) –	2.96e-04 (1.8e-04) =	**2.71e-04 (7.4e-05)** =
CEC19-MaTSO4	2.85e-15 (1.6e-14)	4.78e-07 (1.1e-06) –	9.25e-14 (4.3e-13) –	2.58e-15 (1.4e-14) =	5.47e-14 (1.3e-13) –	**1.42e-17 (5.4e-17)** =	3.82e-15 (1.7e-14) =
CEC19-MaTSO5	4.29e-09 (1.8e-08)	2.07e-05 (7.4e-05) –	1.67e-08 (8.7e-08) –	7.53e-09 (2.9e-08) –	5.69e-07 (1.6e-06) –	1.05e-08 (4.7e-08) –	**9.15e-11 (5.0e-10)** =
CEC19-MaTSO6	7.28e-02 (1.4e-02)	1.06e-01 (4.0e-03) –	**6.60e-02 (7.7e-03)** =	7.35e-02 (1.4e-02) =	7.05e-02 (7.5e-03) =	7.81e-02 (1.6e-02) =	8.50e-02 (1.1e-02) –
WCCI22-MaTSO1	**5.07e-27 (1.7e-26)**	5.94e-25 (1.3e-25) –	2.71e-24 (1.5e-23) –	6.00e-25 (3.3e-24) –	5.95e-22 (1.9e-21) –	1.41e-25 (7.7e-25) –	7.44e-27 (3.9e-26) =
WCCI22-MaTSO2	2.10e-08 (1.2e-08)	3.65e-08 (1.8e-08) –	1.94e-08 (1.1e-08) =	1.95e-08 (1.0e-08) =	**1.83e-08 (7.8e-09)** =	1.98e-08 (1.0e-08) =	2.04e-08 (1.2e-08) =
WCCI22-MaTSO3	3.01e-04 (1.6e-04)	7.47e-04 (2.7e-05) –	4.80e-04 (8.7e-04) =	2.90e-04 (8.5e-05) =	3.58e-04 (3.6e-05) –	3.00e-04 (7.9e-05) =	**2.85e-04 (9.1e-05)** =
WCCI22-MaTSO4	7.16e-09 (6.8e-09)	1.34e-08 (1.0e-08) –	7.05e-09 (5.1e-09) =	6.92e-09 (5.4e-09) =	5.32e-09 (4.7e-09) =	**4.47e-09 (3.3e-09)** =	5.69e-09 (5.6e-09) =
WCCI22-MaTSO5	**9.63e-05 (2.7e-05)**	7.95e-04 (6.2e-04) –	1.31e-04 (2.3e-04) =	1.02e-04 (5.7e-05) =	1.01e-04 (1.2e-05) =	1.00e-04 (3.0e-05) =	9.82e-05 (2.8e-05) =
WCCI22-MaTSO6	5.57e-02 (8.1e-03)	5.48e-02 (5.2e-03) =	5.58e-02 (7.7e-03) =	5.84e-02 (7.9e-03) =	**5.45e-02 (5.7e-03)** =	5.93e-02 (1.0e-02) =	5.98e-02 (5.8e-03) –
WCCI22-MaTSO7	**9.15e-05 (3.2e-05)**	7.41e-04 (4.7e-04) –	9.34e-05 (2.9e-05) =	9.27e-05 (2.7e-05) =	1.05e-04 (1.2e-05) –	9.47e-05 (2.8e-05) =	9.53e-05 (2.3e-05) =
WCCI22-MaTSO8	**6.29e-04 (8.0e-04)**	2.44e-03 (1.1e-03) –	7.52e-04 (8.1e-04) =	7.46e-04 (7.4e-04) =	7.25e-04 (6.3e-04) =	7.43e-04 (9.0e-04) =	7.75e-04 (8.4e-04) =
WCCI22-MaTSO9	2.84e-02 (5.0e-03)	3.14e-02 (3.8e-03) –	2.88e-02 (4.0e-03) =	2.78e-02 (4.8e-03) =	**2.72e-02 (4.2e-03)** =	3.09e-02 (4.9e-03) =	3.10e-02 (4.5e-03) –
WCCI22-MaTSO10	3.74e-02 (4.5e-03)	3.90e-02 (4.0e-03) =	**3.57e-02 (5.2e-03)** =	3.77e-02 (5.2e-03) =	3.60e-02 (5.0e-03) =	3.85e-02 (5.1e-03) =	4.01e-02 (4.3e-03) =
+/−/=	Ours	0/14/2	0/5/11	0/2/14	0/8/8	0/2/14	0/3/13
Ranking	**3.00**	6.38	4.25	3.31	3.88	3.31	3.88
#Best	**4**	0	2	1	3	3	3

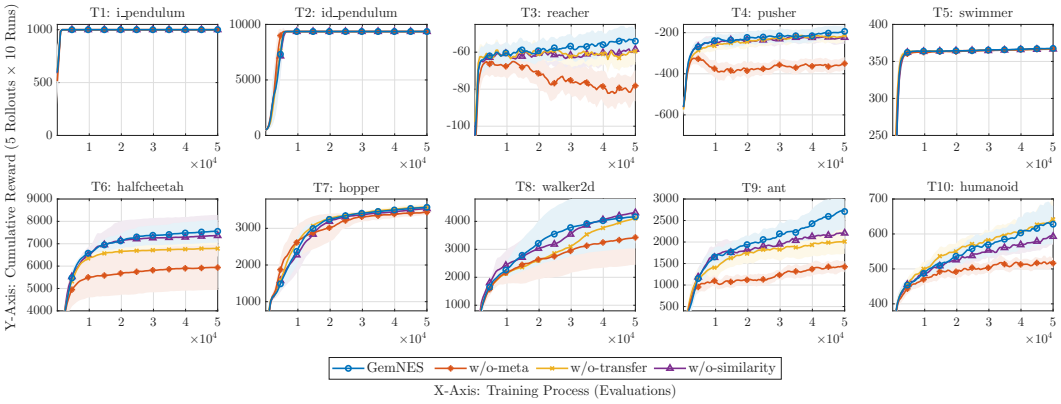


Figure A-5: Convergence curves of cumulative reward ↑ for the ablation study on policy search tasks. Shaded areas show the mean ±1.96 standard errors across 10 independent runs (each averaged over 5 rollouts).

Table A-XI: Ablation study on Brax-MuJoCo policy search tasks to evaluate the structural components of GemNES. The table reports the mean and standard deviation of the cumulative reward (averaged over 5 rollouts) across 10 independent runs. The signs (+ / − / =) indicate whether the compared variant is statistically better, worse, or not significantly different from GemNES according to the two-sided Wilcoxon rank-sum post-hoc test. Ranking is obtained by the Friedman test, and #Best denotes the number of best results achieved by each algorithm.

		GemNES	w/o-meta	w/o-transfer	w/o-similarity
T1	i_pendulum	**1.0000e+03 (0.00e+00)**	**1.0000e+03 (0.00e+00)** =	**1.0000e+03 (0.00e+00)** =	**1.0000e+03 (0.00e+00)** =
T2	id_pendulum	**9.3600e+03 (8.04e-04)**	**9.3600e+03 (7.21e-04)** =	**9.3600e+03 (1.11e-03)** =	**9.3600e+03 (5.15e-04)** =
T3	reacher	-5.6105e+01 (8.45e+00)	-8.2096e+01 (1.36e+01) −	**-5.5639e+01 (1.12e+01)** =	-5.5856e+01 (6.48e+00) =
T4	pusher	**-1.8858e+02 (4.66e+01)**	-3.5358e+02 (5.29e+01) −	-2.1148e+02 (4.05e+01) =	-2.1793e+02 (5.07e+01) =
T5	swimmer	**3.6774e+02 (1.39e+00)**	3.6715e+02 (1.09e+00) =	3.6644e+02 (7.92e-01) −	3.6688e+02 (1.41e+00) =
T6	halfcheetah	**7.5510e+03 (7.93e+02)**	5.9531e+03 (1.63e+03) −	6.7794e+03 (5.59e+02) −	7.3666e+03 (1.49e+03) =
T7	hopper	3.5743e+03 (7.36e+01)	3.4497e+03 (1.08e+02) −	**3.5850e+03 (1.25e+02)** =	3.5338e+03 (9.03e+01) =
T8	walker2d	4.1671e+03 (1.83e+03)	3.4257e+03 (1.57e+03) =	4.1046e+03 (1.28e+03) =	**4.3127e+03 (1.77e+03)** =
T9	ant	**2.6583e+03 (6.91e+02)**	1.4194e+03 (2.50e+02) −	2.0409e+03 (4.77e+02) −	2.2502e+03 (6.04e+02) −
T10	humanoid	6.1568e+02 (6.64e+01)	5.0813e+02 (2.08e+01) −	**6.3785e+02 (7.30e+01)** =	5.9621e+02 (3.89e+01) =
	+ / − / =	Ours	0 / 6 / 4	0 / 3 / 7	0 / 1 / 9
	Ranking	**1.75**	3.35	2.45	2.45
	#Best	**6**	2	5	3

F.5 Parameter Sensitivity Details

Tables A-XII–A-XVI report the instance-wise performance metrics (mean and standard deviation of the unified objective value across 30 runs) for each hyperparameter sweep on the synthetic benchmarks, together with the Friedman rankings and *post-hoc* p-values summarized in Table A-V.

Table A-XII: Parameter sensitivity analysis of the number of meta-candidates M on the CEC19 and WCCI22 synthetic benchmarks.

	$M = 6$	$M = 8$	$M = 10$	$M = 20$	$M = 40$
CEC19-MaTSO1	**2.8100e-10 (1.20e-10)**	3.3071e-10 (1.30e-10)	3.5718e-10 (9.07e-11)	5.3702e-10 (9.06e-11)	6.5256e-10 (5.08e-11)
CEC19-MaTSO2	**2.2665e-16 (1.17e-16)**	2.5796e-16 (2.34e-16)	2.4031e-15 (9.82e-15)	4.5459e-11 (2.48e-10)	1.4076e-10 (3.34e-10)
CEC19-MaTSO3	3.7515e-04 (6.53e-04)	**2.9515e-04 (8.60e-05)**	3.2361e-04 (7.99e-05)	4.3544e-04 (4.58e-04)	3.7369e-04 (7.04e-05)
CEC19-MaTSO4	**8.4411e-19 (3.34e-18)**	3.7493e-18 (1.89e-17)	2.8523e-15 (1.56e-14)	1.6148e-13 (4.45e-13)	3.2447e-13 (8.07e-13)
CEC19-MaTSO5	1.0693e-07 (3.69e-07)	3.1695e-08 (1.06e-07)	**4.2925e-09 (1.76e-08)**	1.5814e-07 (8.58e-07)	1.2931e-06 (3.06e-06)
CEC19-MaTSO6	8.1016e-02 (1.69e-02)	7.3219e-02 (1.54e-02)	7.0283e-02 (1.45e-02)	6.2764e-02 (7.28e-03)	**5.9473e-02 (6.47e-03)**
WCCI22-MaTSO1	**5.5552e-30 (2.31e-29)**	9.1166e-25 (4.99e-24)	7.4447e-27 (3.94e-26)	5.4499e-22 (1.98e-21)	4.2709e-21 (1.28e-20)
WCCI22-MaTSO2	1.8651e-08 (9.45e-09)	1.9443e-08 (1.07e-08)	2.0990e-08 (1.22e-08)	**1.6840e-08 (1.15e-08)**	2.0060e-08 (1.24e-08)
WCCI22-MaTSO3	3.1313e-04 (4.44e-04)	3.1066e-04 (1.91e-04)	**3.0419e-04 (1.59e-04)**	4.4999e-04 (4.87e-04)	3.9089e-04 (9.22e-05)
WCCI22-MaTSO4	6.5822e-09 (5.96e-09)	6.1726e-09 (5.51e-09)	7.1545e-09 (6.76e-09)	5.6952e-09 (4.95e-09)	**5.1495e-09 (4.25e-09)**
WCCI22-MaTSO5	**7.9338e-05 (2.99e-05)**	9.3989e-05 (3.63e-05)	9.8357e-05 (2.99e-05)	9.6889e-05 (1.72e-05)	1.0146e-04 (1.71e-05)
WCCI22-MaTSO6	6.4863e-02 (1.22e-02)	5.7839e-02 (9.30e-03)	5.5165e-02 (8.19e-03)	**5.1370e-02 (5.08e-03)**	5.1502e-02 (6.89e-03)
WCCI22-MaTSO7	**8.2549e-05 (2.74e-05)**	9.3302e-05 (3.38e-05)	9.3112e-05 (2.85e-05)	1.0508e-04 (2.26e-05)	1.0853e-04 (1.61e-05)
WCCI22-MaTSO8	6.5506e-04 (8.32e-04)	6.5512e-04 (7.68e-04)	6.2454e-04 (7.96e-04)	6.3309e-04 (7.56e-04)	**6.0281e-04 (7.52e-04)**
WCCI22-MaTSO9	3.4126e-02 (5.99e-03)	3.0618e-02 (5.20e-03)	2.9995e-02 (4.95e-03)	2.7705e-02 (5.26e-03)	**2.6935e-02 (3.51e-03)**
WCCI22-MaTSO10	3.8227e-02 (7.65e-03)	3.6383e-02 (6.48e-03)	3.4451e-02 (4.58e-03)	3.1986e-02 (4.60e-03)	**3.0026e-02 (3.64e-03)**
Ranking	2.88	2.88	**2.81**	3.12	3.31
p-value	0.9110	0.9110	Ref.	0.5762	0.3711

Table A-XIII: Parameter sensitivity analysis of the meta-horizon H on the CEC19 and WCCI22 synthetic benchmarks.

	$H = 1$	$H = 2$	$H = 3$	$H = 4$	$H = 5$
CEC19-MaTSO1	**2.5927e-10 (9.41e-11)**	3.6050e-10 (9.07e-11)	4.5518e-10 (8.89e-11)	5.3113e-10 (8.58e-11)	6.1509e-10 (9.16e-11)
CEC19-MaTSO2	**1.7721e-16 (7.73e-17)**	2.3931e-15 (9.82e-15)	8.3989e-12 (4.38e-11)	5.6861e-11 (1.99e-10)	9.0216e-12 (3.53e-11)
CEC19-MaTSO3	**2.2093e-04 (8.77e-05)**	2.9300e-04 (7.99e-05)	4.8614e-04 (9.67e-04)	4.6165e-04 (4.66e-04)	3.1040e-04 (1.47e-04)
CEC19-MaTSO4	**4.5968e-19 (1.38e-18)**	2.8523e-15 (1.56e-14)	1.7170e-14 (6.14e-14)	1.6746e-13 (4.45e-13)	2.3241e-13 (6.51e-13)
CEC19-MaTSO5	1.1718e-07 (2.85e-07)	**4.2925e-09 (1.76e-08)**	3.5195e-07 (1.92e-06)	2.2818e-07 (8.07e-07)	3.8251e-07 (1.43e-06)
CEC19-MaTSO6	7.9735e-02 (1.74e-02)	7.1775e-02 (1.44e-02)	6.6705e-02 (1.00e-02)	6.5075e-02 (6.93e-03)	**6.3432e-02 (8.10e-03)**
WCCI22-MaTSO1	**1.9207e-32 (5.56e-32)**	7.4447e-27 (3.94e-26)	1.0508e-24 (4.48e-24)	1.3860e-21 (5.82e-21)	2.6303e-21 (1.22e-20)
WCCI22-MaTSO2	1.8196e-08 (7.40e-09)	2.0972e-08 (1.22e-08)	1.8347e-08 (1.14e-08)	1.8105e-08 (1.01e-08)	**1.5534e-08 (1.09e-08)**
WCCI22-MaTSO3	**2.7974e-04 (1.82e-04)**	2.9794e-04 (1.59e-04)	4.2816e-04 (6.75e-04)	4.8813e-04 (8.88e-04)	3.2174e-04 (1.44e-04)
WCCI22-MaTSO4	6.8988e-09 (6.32e-09)	7.1545e-09 (6.76e-09)	6.2271e-09 (4.89e-09)	5.8267e-09 (5.62e-09)	**4.8138e-09 (3.44e-09)**
WCCI22-MaTSO5	**8.0376e-05 (2.71e-05)**	9.7181e-05 (2.99e-05)	9.0646e-05 (2.61e-05)	9.3083e-05 (2.23e-05)	9.2035e-05 (3.37e-05)
WCCI22-MaTSO6	5.9991e-02 (9.42e-03)	5.2171e-02 (8.27e-03)	5.0133e-02 (6.58e-03)	**4.8948e-02 (4.78e-03)**	4.9544e-02 (5.52e-03)
WCCI22-MaTSO7	**9.1194e-05 (3.42e-05)**	9.4806e-05 (2.85e-05)	9.2738e-05 (2.90e-05)	1.0312e-04 (2.23e-05)	9.9858e-05 (3.25e-05)
WCCI22-MaTSO8	6.4891e-04 (8.19e-04)	6.2576e-04 (7.96e-04)	6.3244e-04 (7.97e-04)	6.2474e-04 (7.79e-04)	**5.8506e-04 (7.28e-04)**
WCCI22-MaTSO9	3.3058e-02 (6.73e-03)	2.8588e-02 (5.00e-03)	2.7283e-02 (5.05e-03)	2.6246e-02 (3.12e-03)	**2.5070e-02 (4.33e-03)**
WCCI22-MaTSO10	3.9248e-02 (6.00e-03)	3.4545e-02 (4.56e-03)	3.3296e-02 (4.89e-03)	**2.9821e-02 (5.36e-03)**	3.1426e-02 (3.98e-03)
Ranking	**2.62**	3.12	3.25	3.12	2.88
p-value	0.2636	0.8231	Ref.	0.8231	0.5023

Table A-XIV: Parameter sensitivity analysis of the meta-policy architecture (hidden width d_{hidden} and latent dimension d_{latent}) on the CEC19 and WCCI22 synthetic benchmarks.

	$d_{\text{hidden}} = 2$ $d_{\text{latent}} = 1$	$d_{\text{hidden}} = 4$ $d_{\text{latent}} = 2$	$d_{\text{hidden}} = 8$ $d_{\text{latent}} = 4$	$d_{\text{hidden}} = 16$ $d_{\text{latent}} = 8$	$d_{\text{hidden}} = 32$ $d_{\text{latent}} = 16$
CEC19-MaTSO1	5.6270e-10 (7.61e-11)	4.9829e-10 (1.20e-10)	3.9035e-10 (8.43e-11)	3.3041e-10 (1.10e-10)	**3.1135e-10 (8.01e-11)**
CEC19-MaTSO2	8.0297e-11 (3.92e-10)	1.5757e-11 (8.03e-11)	2.3903e-15 (9.81e-15)	**1.8954e-16 (9.36e-17)**	2.5661e-16 (1.14e-16)
CEC19-MaTSO3	3.4104e-04 (1.04e-04)	3.8216e-04 (4.69e-04)	3.0355e-04 (7.68e-05)	**2.3327e-04 (1.11e-04)**	2.3944e-04 (8.77e-05)
CEC19-MaTSO4	3.3447e-13 (7.26e-13)	1.4214e-13 (3.20e-13)	2.7175e-15 (1.48e-14)	7.8601e-19 (3.53e-18)	**6.3579e-19 (1.38e-18)**
CEC19-MaTSO5	3.7739e-07 (1.00e-06)	**2.5869e-10 (9.66e-10)**	4.2271e-09 (1.73e-08)	4.0378e-08 (1.81e-07)	6.2555e-08 (1.01e-07)
CEC19-MaTSO6	**6.0192e-02 (7.16e-03)**	6.5068e-02 (9.06e-03)	6.9143e-02 (1.43e-02)	7.5669e-02 (1.05e-02)	8.5609e-02 (1.87e-02)
WCCI22-MaTSO1	8.3395e-21 (2.85e-20)	3.9025e-21 (2.13e-20)	7.1454e-27 (3.78e-26)	1.2988e-25 (7.03e-25)	**4.9911e-33 (4.27e-33)**
WCCI22-MaTSO2	1.5803e-08 (8.90e-09)	**1.4506e-08 (8.78e-09)**	1.9316e-08 (1.15e-08)	1.6755e-08 (8.68e-09)	1.8421e-08 (9.57e-09)
WCCI22-MaTSO3	4.2142e-04 (4.84e-04)	3.5448e-04 (2.65e-04)	2.8514e-04 (1.52e-04)	3.7161e-04 (3.82e-04)	**2.3200e-04 (7.79e-05)**
WCCI22-MaTSO4	**5.2369e-09 (4.07e-09)**	5.3147e-09 (4.19e-09)	6.1599e-09 (5.64e-09)	6.2363e-09 (4.80e-09)	7.3872e-09 (5.04e-09)
WCCI22-MaTSO5	9.4294e-05 (4.16e-05)	9.3171e-05 (7.78e-05)	**9.1865e-05 (2.80e-05)**	1.6964e-04 (2.81e-04)	1.1115e-04 (1.60e-04)
WCCI22-MaTSO6	**4.8983e-02 (6.80e-03)**	5.3905e-02 (6.65e-03)	5.3041e-02 (8.08e-03)	6.0211e-02 (8.49e-03)	6.5635e-02 (9.79e-03)
WCCI22-MaTSO7	1.1709e-04 (8.84e-05)	1.0695e-04 (1.37e-04)	**8.5523e-05 (2.71e-05)**	9.2115e-05 (9.81e-05)	9.5995e-05 (2.53e-05)
WCCI22-MaTSO8	5.3504e-04 (5.46e-04)	**5.3445e-04 (5.89e-04)**	6.0820e-04 (7.81e-04)	6.1228e-04 (8.45e-04)	9.6241e-04 (9.06e-04)
WCCI22-MaTSO9	**2.4682e-02 (4.37e-03)**	2.7057e-02 (5.62e-03)	2.8340e-02 (4.81e-03)	3.1562e-02 (4.84e-03)	3.4167e-02 (4.57e-03)
WCCI22-MaTSO10	**2.9610e-02 (4.00e-03)**	3.1584e-02 (5.51e-03)	3.3727e-02 (4.50e-03)	3.4900e-02 (3.87e-03)	3.9708e-02 (5.55e-03)
Ranking	3.19	2.75	**2.62**	3.12	3.31
p-value	0.3143	0.8231	Ref.	0.3711	0.2188

Table A-XV: Parameter sensitivity analysis of the maximum gradient transfer rate $\rho_{\max}$ on the CEC19 and WCCI22 synthetic benchmarks.

	$\rho_{\max} = 0.1$	$\rho_{\max} = 0.3$	$\rho_{\max} = 0.5$	$\rho_{\max} = 0.7$	$\rho_{\max} = 0.9$
CEC19-MaTSO1	3.3430e-10 (9.05e-11)	3.0819e-10 (9.07e-11)	2.6002e-10 (9.07e-11)	**2.3583e-10 (8.95e-11)**	2.5333e-10 (9.22e-11)
CEC19-MaTSO2	1.3411e-14 (5.71e-14)	7.4102e-15 (3.82e-14)	2.4198e-15 (9.82e-15)	**7.7045e-16 (3.01e-15)**	1.0482e-15 (3.68e-15)
CEC19-MaTSO3	2.9648e-04 (1.49e-04)	**2.7136e-04 (1.81e-04)**	3.0239e-04 (7.99e-05)	4.0582e-04 (6.23e-04)	2.7897e-04 (7.89e-05)
CEC19-MaTSO4	7.0263e-14 (3.04e-13)	3.2048e-14 (1.75e-13)	2.8523e-15 (1.56e-14)	1.4188e-15 (7.74e-15)	**3.1295e-18 (8.18e-18)**
CEC19-MaTSO5	1.3366e-07 (7.32e-07)	**3.0842e-09 (1.61e-08)**	4.2925e-09 (1.76e-08)	1.8739e-08 (6.54e-08)	4.6002e-08 (1.20e-07)
CEC19-MaTSO6	**6.5116e-02 (8.57e-03)**	6.7409e-02 (9.28e-03)	6.9959e-02 (1.45e-02)	7.0100e-02 (1.03e-02)	7.1130e-02 (1.65e-02)
WCCI22-MaTSO1	6.1354e-25 (3.29e-24)	1.8461e-26 (6.87e-26)	**7.4447e-27 (3.94e-26)**	3.3669e-25 (1.84e-24)	2.4623e-26 (1.19e-25)
WCCI22-MaTSO2	**1.8575e-08 (1.00e-08)**	2.0106e-08 (1.35e-08)	2.0968e-08 (1.22e-08)	2.0863e-08 (1.10e-08)	2.1160e-08 (1.15e-08)
WCCI22-MaTSO3	**2.7353e-04 (1.13e-04)**	5.3665e-04 (8.88e-04)	2.7986e-04 (1.59e-04)	4.2315e-04 (5.09e-04)	2.8945e-04 (9.89e-05)
WCCI22-MaTSO4	5.9237e-09 (5.17e-09)	**5.6051e-09 (4.53e-09)**	7.1472e-09 (6.76e-09)	6.4976e-09 (5.10e-09)	6.8298e-09 (5.91e-09)
WCCI22-MaTSO5	**8.2526e-05 (2.72e-05)**	9.0668e-05 (2.89e-05)	9.7323e-05 (2.99e-05)	9.2594e-05 (3.24e-05)	1.0308e-04 (7.29e-05)
WCCI22-MaTSO6	5.5333e-02 (6.03e-03)	5.5309e-02 (7.97e-03)	**5.5295e-02 (8.17e-03)**	5.7630e-02 (7.96e-03)	5.9360e-02 (9.50e-03)
WCCI22-MaTSO7	9.2580e-05 (2.97e-05)	**9.1481e-05 (3.22e-05)**	9.4890e-05 (2.85e-05)	9.9137e-05 (2.52e-05)	1.2751e-04 (1.02e-04)
WCCI22-MaTSO8	**6.0117e-04 (6.93e-04)**	7.1770e-04 (8.36e-04)	6.2604e-04 (7.96e-04)	7.3208e-04 (9.33e-04)	6.7803e-04 (7.68e-04)
WCCI22-MaTSO9	**2.8389e-02 (4.85e-03)**	2.8703e-02 (3.12e-03)	2.8933e-02 (5.02e-03)	2.9373e-02 (3.87e-03)	2.9688e-02 (4.85e-03)
WCCI22-MaTSO10	**3.5800e-02 (4.54e-03)**	3.6332e-02 (5.64e-03)	3.6317e-02 (4.53e-03)	3.7398e-02 (5.65e-03)	3.7167e-02 (5.77e-03)
Ranking	2.62	**2.50**	2.81	3.44	3.62
p-value	0.7373	0.5762	Ref.	0.2636	0.1461

Table A-XVI: Parameter sensitivity analysis of the penalty coefficient α on the CEC19 and WCCI22 synthetic benchmarks.

	$\alpha = 0.0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 1.0$
CEC19-MaTSO1	6.2859e-10 (2.40e-10)	3.1279e-10 (1.21e-10)	3.3052e-10 (9.07e-11)	3.7862e-10 (9.70e-11)	**2.8708e-10 (1.20e-10)**
CEC19-MaTSO2	1.5981e-15 (5.25e-15)	5.0156e-14 (2.73e-13)	2.4132e-15 (9.82e-15)	**3.6268e-16 (9.06e-16)**	7.4079e-16 (2.70e-15)
CEC19-MaTSO3	3.0233e-04 (2.00e-04)	3.9276e-04 (6.28e-04)	2.9038e-04 (7.99e-05)	**2.4426e-04 (7.08e-05)**	4.4919e-04 (9.81e-04)
CEC19-MaTSO4	**1.6367e-17 (7.56e-17)**	8.7543e-15 (4.78e-14)	2.8523e-15 (1.56e-14)	2.9637e-16 (1.53e-15)	1.0933e-14 (5.95e-14)
CEC19-MaTSO5	8.1752e-08 (2.61e-07)	1.4494e-08 (6.50e-08)	**4.2925e-09 (1.76e-08)**	2.4022e-08 (6.12e-08)	4.8881e-09 (1.96e-08)
CEC19-MaTSO6	7.9483e-02 (2.68e-02)	7.1612e-02 (1.18e-02)	7.0903e-02 (1.45e-02)	7.3801e-02 (1.38e-02)	**6.9287e-02 (1.12e-02)**
WCCI22-MaTSO1	1.0192e-26 (4.68e-26)	1.0978e-26 (4.97e-26)	**7.4447e-27 (3.94e-26)**	6.3058e-26 (3.43e-25)	4.5713e-26 (2.46e-25)
WCCI22-MaTSO2	2.4117e-08 (1.23e-08)	2.1005e-08 (1.21e-08)	2.0974e-08 (1.22e-08)	1.9376e-08 (1.32e-08)	**1.8734e-08 (1.15e-08)**
WCCI22-MaTSO3	**2.7345e-04 (8.92e-05)**	5.0141e-04 (8.93e-04)	2.8564e-04 (1.59e-04)	2.7792e-04 (1.01e-04)	5.2917e-04 (1.02e-03)
WCCI22-MaTSO4	6.2717e-09 (5.93e-09)	**5.3025e-09 (5.35e-09)**	7.1476e-09 (6.76e-09)	7.1466e-09 (5.31e-09)	6.0496e-09 (5.50e-09)
WCCI22-MaTSO5	9.2108e-05 (2.95e-05)	9.8098e-05 (2.82e-05)	9.5420e-05 (2.99e-05)	**9.1773e-05 (2.89e-05)**	9.7277e-05 (2.79e-05)
WCCI22-MaTSO6	5.6924e-02 (7.94e-03)	5.6500e-02 (7.19e-03)	**5.4929e-02 (8.21e-03)**	5.6877e-02 (7.51e-03)	5.6460e-02 (7.94e-03)
WCCI22-MaTSO7	9.3744e-05 (2.41e-05)	9.0480e-05 (2.76e-05)	8.8555e-05 (2.85e-05)	9.5336e-05 (2.63e-05)	**8.5682e-05 (2.64e-05)**
WCCI22-MaTSO8	8.0300e-04 (8.73e-04)	7.3477e-04 (8.78e-04)	6.2503e-04 (7.96e-04)	**5.7935e-04 (7.00e-04)**	7.2325e-04 (1.12e-03)
WCCI22-MaTSO9	3.1210e-02 (5.78e-03)	3.0826e-02 (3.77e-03)	3.0488e-02 (4.95e-03)	2.9601e-02 (4.08e-03)	**2.8832e-02 (4.82e-03)**
WCCI22-MaTSO10	3.4793e-02 (5.33e-03)	**3.3558e-02 (5.76e-03)**	3.4532e-02 (4.59e-03)	3.5292e-02 (5.41e-03)	3.3614e-02 (5.33e-03)
Ranking	3.62	3.31	**2.56**	2.94	**2.56**
p-value	0.0573	0.1797	Ref.	0.5023	1.0000