CLOSING THE PARETO-RESPONSE LOOP: PREFERENCE-AWARE OPTIMISTIC SELECTION FOR MULTI-OBJECTIVE BILEVEL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-objective bilevel learning (MOBL) coordinates conflicting upper-level objectives through lower-level responses. Yet prevailing methods typically use a *response-first organization*: they form the lower-level response first and only afterward reconcile its induced objectives or hypergradients according to user preferences, without directly feeding the Pareto mixture into the lower objective. This response-formation channel remains *open-loop*, although under non-convex or finite-step optimization, the reached response is *highly path-dependent* and fundamentally shapes the ensuing Pareto geometry. We introduce **Preference-Aware Optimistic Selection (PAOS)**, which uses a selector-controlled lower objective for optimistic branch selection among non-unique lower solutions and preference-guided response shaping under finite optimization budgets. Specifically, a stateful response-selection variable integrates the preference-weighted Pareto mixture produced by the current subproblem and guides the subsequent lower-level response, thereby closing the Pareto-response loop. On an analytically tractable double-well problem, PAOS systematically selects the preference-compatible branch from near-boundary starts. Across data hyper-cleaning, few-shot bilevel learning, physics-informed neural networks, and medical image segmentation, PAOS achieves the strongest overall performance among the evaluated MOBL methods. These results establish preference-aware optimistic response selection as an actionable degree of freedom in multi-objective bilevel decision-making. Our code is available at Repository.

1 INTRODUCTION

Modern learning systems rarely optimize a single criterion in isolation. Data quality, predictive accuracy, physical fidelity, and deployment reliability are often coupled through a model produced by training. Bilevel learning provides a natural abstraction for this hierarchy and underpins applications including data reweighting, meta-learning, and scientific machine learning (Franceschi et al., 2018; Dagréou et al., 2022). When multiple upper-level criteria conflict, this structure gives rise to multi-objective bilevel learning (MOBL) (Zhang et al., 2026a). Because these criteria are evaluated through lower-level optimization, the lower-level response is the interface through which upper-level trade-offs are ultimately expressed.

Recent work has substantially advanced the coordination of conflicting objectives. Preference-guided multi-objective learning targets user-specified regions of the Pareto set (Mahapatra & Rajan, 2020; Navon et al., 2021; Chen et al., 2024; 2025), while MOBL methods combine multi-gradient aggregation, robust formulations, first-order reformulations, Pareto stationarity, and preference-aware subproblems (Ye et al., 2021; Gu et al., 2023; Ye et al., 2024; Yang et al., 2024; Zhang et al., 2026a). Most existing MOBL pipelines, however, share a *response-first organization*: they first form a lower-level response and only afterward coordinate its induced objectives or hypergradients according to user preferences. Preference influences subsequent responses through upper-variable updates, but the Pareto mixture does not directly enter the lower training objective. We call this missing direct feedback channel *open-loop response formation*.

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.

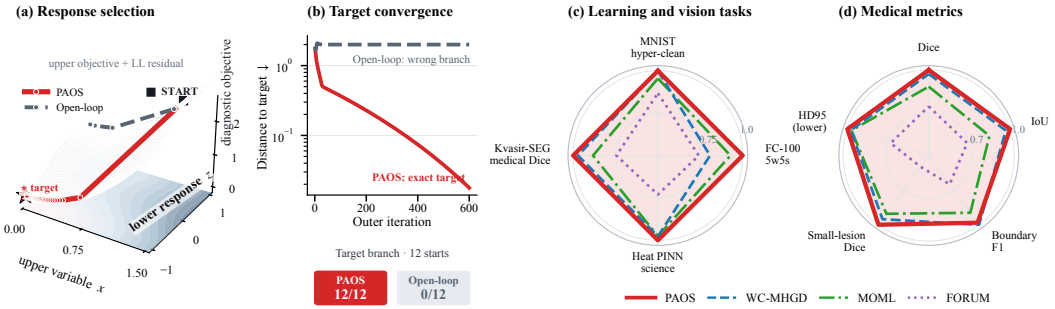


Figure 1: **(a–b)** Open-loop optimization follows a preference-incompatible branch, whereas Pareto feedback enables PAOS to select the preference-compatible branch across all 12 near-boundary hard starts of the aligned-objective control (Appendix B). **(c–d)** This active response selection yields leading performance across learning and medical-vision benchmarks.

This separation becomes especially consequential under non-convex or finite-step lower-level optimization. The reached response can be *highly path-dependent*, varying with initialization, optimization trajectory, and transient training dynamics (Liu et al., 2020; Vicol et al., 2022; Arbel & Mairal, 2022; Liu et al., 2023). Different responses induce different upper-level objectives and hypergradients, thereby fundamentally shaping the Pareto geometry encountered above. A preference introduced only after response formation may therefore arrive too late to influence which response is reached. This raises the central question: **Can the upper-level preference actively select the lower-level response from which the next Pareto geometry emerges?** We answer this question with **Preference-Aware Optimistic Selection (PAOS)**. PAOS operationalizes the optimistic bilevel principle in a preference-aware form: the evolving Pareto mixture favors lower-level responses compatible with the desired upper-level trade-off. By carrying this selection signal into subsequent lower-level optimization, PAOS actively closes the Pareto-response loop.

PAOS realizes active response selection through a stateful selector and a selector-controlled lower-level objective, supported by response-conditioned implicit differentiation and a conditional feedback-variation bound. Figure 1 further previews its strong performance across data hypercleaning (Ren et al., 2018; Franceschi et al., 2018), few-shot bilevel learning (Oreshkin et al., 2018; Rajeswaran et al., 2019), physics-informed neural networks (Raissi et al., 2019), and medical image segmentation (Jha et al., 2020). Controlled ablations show cumulative gains from response selection, preference conditioning, and dynamic Pareto feedback. Our contributions are threefold:

- **A closed-loop perspective on MOBL.** We identify a missing direct feedback channel between Pareto coordination and response formation, and establish the latter as an actionable degree of freedom in multi-objective decision-making.
- **Preference-Aware Optimistic Selection.** PAOS transforms preference-weighted Pareto mixtures into a stateful selector and a selector-controlled lower-level objective, thereby enabling optimistic branch selection and finite-budget response shaping.
- **Mechanistic and cross-domain validation.** We demonstrate preference-consistent branch selection on an analytic problem and leading performance across four domains; controlled ablations isolate dynamic-feedback gains beyond static preference selection.

2 Related Work

Bilevel Optimization and Learning. Bilevel optimization models learning systems in which one decision is defined through another optimization process, providing a common foundation for hyperparameter optimization and meta-learning (Maclaurin et al., 2015; Franceschi et al., 2018). Its algorithmic development spans approximate implicit differentiation, truncated or iterative differentiation, scalable hypergradient computation, and stochastic variance-reduced optimization (Pedregosa, 2016; Shaban et al., 2019; Lorraine et al., 2020; Grazzi et al., 2020; Dagréou et al., 2022). In overparameterized regimes, the attained response further reflects initialization, optimization dynamics, and selection among non-unique lower-level solutions (Liu et al., 2020; Arbel & Mairal, 2022; Vicol

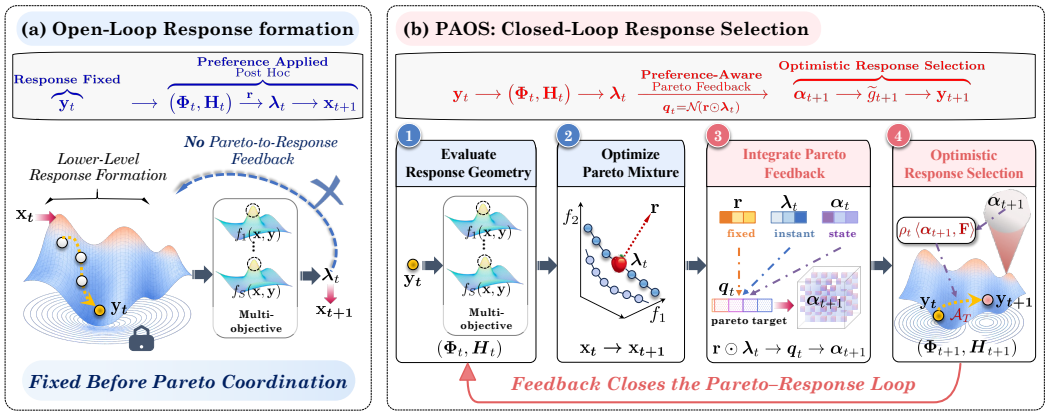


Figure 2: (a) Open-loop MOBL coordinates objectives after response formation. (b) PAOS feeds the preference-conditioned Pareto mixture into a stateful selector that shapes the next response.

et al., 2022; Masiha et al., 2026). These works study lower-solution selection under a single upper-level objective. PAOS instead derives the selection signal from the current multi-objective Pareto subproblem and feeds it back to recursively reshape the next response geometry.

Multi-Objective and Preference-Aware Learning. Gradient-based multi-objective optimization constructs descent directions that reconcile conflicting criteria (Désidéri, 2012; Sener & Koltun, 2018). Related multi-task learning methods develop projection-, conflict-, and bargaining-based mechanisms for coordinating task gradients (Yu et al., 2020; Liu et al., 2021; Navon et al., 2022). Preference-aware formulations further target designated Pareto regions or learn an entire preference-conditioned solution family (Lin et al., 2019; Mahapatra & Rajan, 2020; Navon et al., 2021), with stochastic extensions addressing biased multi-objective gradients (Fernando et al., 2023). These works establish preference-guided coordination in objective space. In a bilevel pipeline, objective-space coordination leaves the underlying lower-level response formation unchanged. PAOS extends preference guidance to this, allowing each Pareto decision to shape its induced objective geometry.

Multi-Objective Bilevel Learning. Multi-objective bilevel learning combines nested response generation with Pareto coordination. Existing formulations cover multi-objective meta-learning and robust min-max learning (Ye et al., 2021; Gu et al., 2023), gradient-based Pareto-stationary algorithms and value-function-based first-order optimization (Yang et al., 2024; Ye et al., 2024), and preference-guided multi-hypergradient descent (Zhang et al., 2026a). Recent developments further address non-convex lower levels through Moreau-envelope relaxations and general-convexity settings through equality-constrained multi-objective reformulations (Jiang & Huang, 2026; Zhang et al., 2026b). These methods nevertheless remain *response first*: Pareto coordination follows an already-formed lower-level response. PAOS instead feeds the current Pareto mixture back through a stateful selector, guiding the next response and closing the Pareto-response loop.

3 Method

Figure 2 summarizes the PAOS workflow. Section 3.1 introduces the MOBL formulation and its conventional open-loop organization. Section 3.2 formulates preference-aware optimistic selection, Section 3.3 develops the closed-loop feedback update, and Section 3.4 presents its theoretical properties. Algorithm 1 states the complete procedure.

3.1 Preliminaries

Multi-objective bilevel learning. Let $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^p$ and $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^q$ denote the upper- and lower-level variables. We consider S upper-level objectives collected by $\mathbf{F}(\mathbf{x}, \mathbf{y}) := (f_1(\mathbf{x}, \mathbf{y}), \ldots, f_S(\mathbf{x}, \mathbf{y}))^\top$, coupled through a scalar lower-level objective g. Relative to a selected

response branch, MOBL seeks a Pareto solution of

$$\min_{\mathbf{x}\in\mathcal{X}} \quad \mathbf{\Phi}(\mathbf{x}) := \mathbf{F}\big(\mathbf{x}, \mathbf{y}^\star(\mathbf{x})\big), \quad \mathbf{y}^\star(\mathbf{x}) \in \mathcal{Y}^\star(\mathbf{x}) := \arg\min_{\mathbf{y}\in\mathcal{Y}} g(\mathbf{x}, \mathbf{y}). \tag{1}$$

Here $\mathbf{\Phi} : \mathcal{X} \to \mathbb{R}^S$ is the induced upper-level objective map associated with the selected branch. A preference $\mathbf{r} \in \mathrm{int}(\Delta_S)$ specifies the desired trade-off, with $\Delta_S := \{\mathbf{w} \in \mathbb{R}^S_+ : \mathbf{1}^\top \mathbf{w} = 1\}$.

Open-loop response formation. In learning applications, the exact response is commonly replaced by a finite-step solver, initialized from the preceding iterate. The attained response can therefore depend on initialization, local geometry, and the optimization path. Conventional MOBL first forms this response and only afterward coordinates its induced upper-level criteria according to the preference. Thus, the preference determines how an obtained response is assessed, while the current Pareto decision is not directly fed into the lower-level objective that forms the response. We refer to this response-first organization as *open-loop response formation*; see Figure 2(a).

3.2 Preference-Aware Optimistic Selection

Optimistic response selection. When the lower problem admits multiple valid responses, the choice among them can materially change the induced upper-level trade-off. For a mixture $\boldsymbol{\alpha} \in \Delta_S$, let $\psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}) := \langle \boldsymbol{\alpha}, \mathbf{F}(\mathbf{x}, \mathbf{y})\rangle$ denote its scalarized upper criterion. We define the ideal optimistic response set as

$$\mathcal{Y}^{\mathrm{opt}}_{\boldsymbol{\alpha}}(\mathbf{x}) := \arg\min_{\mathbf{y}\in\mathcal{Y}^\star(\mathbf{x})} \psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}). \tag{2}$$

Here, g enforces lower-level validity, while $\psi_{\boldsymbol{\alpha}}$ selects the response best aligned with the current upper-level trade-off, extending optimistic bilevel selection (Liu et al., 2020) to multiple objectives.

Stateful selection surrogate. PAOS makes the upper-level trade-off operative during response formation. At iteration t, it combines the original lower objective with the upper-objective mixture specified by its selector state, then warm-starts a T-step lower solve (Step 4 of Figure 2(b)):

$$\widetilde{g}_t(\mathbf{x}, \mathbf{y}) := g(\mathbf{x}, \mathbf{y}) + \rho_t \langle \boldsymbol{\alpha}_t, \mathbf{F}(\mathbf{x}, \mathbf{y})\rangle = g(\mathbf{x}, \mathbf{y}) + \rho_t \sum_{s=1}^{S} \alpha_{t,s} f_s(\mathbf{x}, \mathbf{y}), \quad \mathbf{y}_t := \mathcal{A}_T(\mathbf{y}_{t-1}; \mathbf{x}_t, \widetilde{g}_t). \tag{3}$$

Here $\rho_t \geq 0$ sets the selection strength, $\boldsymbol{\alpha}_t \in \Delta_S$ is the stateful response selector, and $\mathcal{A}_T$ denotes T lower-level optimization steps initialized at $\mathbf{y}_{t-1}$. The upper state $\mathbf{x}_t$ and selector $\boldsymbol{\alpha}_t$ remain fixed during these steps. We initialize $\boldsymbol{\alpha}_0 = \mathbf{r}$: the user preference fixes the target trade-off, while later Pareto mixtures update $\boldsymbol{\alpha}_t$ as the response-conditioned geometry evolves. The resulting selected response defines the next geometry through which the upper-level objectives are coordinated. The small-ρ connection of Eq. (3) to the optimistic set in Eq. (2) is established in Proposition 1. PAOS realizes optimistic branch selection when the lower solution set is non-singleton, and preference-guided finite-budget response shaping when the lower response is unique. In the latter case, the selector shapes the finite-step trajectory through the perturbed objective rather than choosing among multiple exact minimizers of g.

Appendix A.2 proves the small-ρ selection limit, while Appendix A.3 quantifies the induced response and Pareto-geometry shifts.

3.3 Closed-Loop Pareto-Response Optimization

Response-conditioned Pareto geometry. At the selected response $(\mathbf{x}_t, \mathbf{y}_t)$, PAOS evaluates $\mathbf{\Phi}_t := \mathbf{F}(\mathbf{x}_t, \mathbf{y}_t)$ and differentiates each objective through the selected lower problem (Step 1 of Figure 2(b)). All derivatives below are evaluated at $(\mathbf{x}_t, \mathbf{y}_t)$ while $\boldsymbol{\alpha}_t$ is held fixed. Define the damped selected-response Hessian $A_t := \nabla^2_{\mathbf{yy}} \widetilde{g}_t + \delta I$ and mixed Hessian $B_t := \nabla^2_{\mathbf{xy}} \widetilde{g}_t$, where $\delta > 0$. For each objective $s \in [S]$, warm-started conjugate gradients produce $\mathbf{v}_{t,s}$ satisfying $A_t \mathbf{v}_{t,s} \approx \nabla_{\mathbf{y}} f_s$. The resulting implicit hypergradient and hypergradient matrix are

$$\widehat{\mathbf{h}}_{t,s} := \nabla_{\mathbf{x}} f_s - B_t \mathbf{v}_{t,s}, \qquad \mathbf{H}_t := [\widehat{\mathbf{h}}_{t,1}, \ldots, \widehat{\mathbf{h}}_{t,S}] \in \mathbb{R}^{p\times S}. \tag{4}$$

Finite-response, CG, and damping errors are analyzed in Appendix A.4; warm-start and finite-step solver details appear in Appendix C.2.

Preference-aware Pareto feedback. Stages 2-3 of Figure 2(b) convert the response-conditioned geometry into the control signal that forms the next lower-level response. Given $(\mathbf{\Phi}_t, \mathbf{H}_t)$, PAOS first computes an instantaneous Pareto mixture through the preference-aware weighted-Chebyshev multi-hypergradient subproblem (Zhang et al., 2026a):

$$\boldsymbol{\lambda}_t \in \arg\min_{\boldsymbol{\lambda}\in\Delta_S} \|\mathbf{H}_t(\mathbf{r}\odot\boldsymbol{\lambda})\|_2^2 - u\,\boldsymbol{\lambda}^\top(\mathbf{r}\odot\mathbf{\Phi}_t). \tag{5}$$

Here $\boldsymbol{\lambda}_t \in \Delta_S$ reconciles the current objective conflicts, $\mathbf{r} \in \mathrm{int}(\Delta_S)$ is the fixed user preference, $\odot$ denotes element-wise multiplication, and $u \geq 0$ weights the preference-aware objective-space term. The effective mixture $\mathbf{r}\odot\boldsymbol{\lambda}_t$ determines the upper-level direction and update:

$$\mathbf{d}_t := \mathbf{H}_t(\mathbf{r}\odot\boldsymbol{\lambda}_t), \qquad \mathbf{x}_{t+1} := \Pi_{\mathcal{X}}(\mathbf{x}_t - \eta_x \mathbf{d}_t). \tag{6}$$

Here $\eta_x > 0$ is the upper-level step size, $\Pi_{\mathcal{X}}$ is the projection onto $\mathcal{X}$ (the identity when $\mathcal{X} = \mathbb{R}^p$).

The defining PAOS operation is to reuse the same Pareto decision as feedback for subsequent response formation. Let $\mathcal{N}(\mathbf{a}) := \mathbf{a}/(\mathbf{1}^\top\mathbf{a})$ denote simplex normalization. PAOS forms the instantaneous Pareto target and updates its response-selector state as

$$\mathbf{q}_t := \mathcal{N}(\mathbf{r}\odot\boldsymbol{\lambda}_t), \qquad \boldsymbol{\alpha}_{t+1} := (1-\beta)\boldsymbol{\alpha}_t + \beta\mathbf{q}_t, \quad \beta\in[0,1]. \tag{7}$$

The four quantities have distinct roles: $\mathbf{r}$ specifies the global intent, $\boldsymbol{\lambda}_t$ resolves the current response-conditioned geometry, $\mathbf{q}_t$ is the resulting instantaneous target, and $\boldsymbol{\alpha}_{t+1}$ carries that target into the next selected lower objective. The feedback rate β controls how rapidly current Pareto decisions enter response selection: $\beta = 0$ recovers static preference selection when $\boldsymbol{\alpha}_0 = \mathbf{r}$, whereas $\beta > 0$ closes the feedback loop. Since $\mathbf{r} \in \mathrm{int}(\Delta_S)$ and $\boldsymbol{\lambda}_t \in \Delta_S$, the normalization is well-defined; because $\boldsymbol{\alpha}_t, \mathbf{q}_t \in \Delta_S$, the state update preserves simplex membership. The feedback construction requires only a simplex-valued Pareto mixture.

Simplex invariance and feedback tracking are established in Appendix A.5; selector schedules are specified in Appendix C.1.

Open- versus closed-loop recursion. Figure 2 contrasts when preference enters response formation. In the open-loop path, the response is formed before preference enters Pareto coordination; the resulting mixture influences the next response through the update of $\mathbf{x}$. PAOS also updates $\mathbf{x}$ via Eq. (6), while routing the mixture through $\mathbf{q}_t$ and $\boldsymbol{\alpha}_{t+1}$ to shape the next lower-level response:

$$\textbf{Open loop:}\ \overbrace{\mathbf{y}_t}^{\substack{\text{Response}\\ \text{Fixed}}} \overbrace{\rightarrow (\mathbf{\Phi}_t, \mathbf{H}_t) \xrightarrow{\mathbf{r}} \boldsymbol{\lambda}_t \rightarrow \mathbf{x}_{t+1}}^{\text{Preference Applied Post Hoc}} \quad\Big|\quad \textbf{PAOS:}\ \mathbf{y}_t \rightarrow (\mathbf{\Phi}_t, \mathbf{H}_t) \rightarrow \boldsymbol{\lambda}_t \xrightarrow[\mathbf{q}_t=\mathcal{N}(\mathbf{r}\odot\boldsymbol{\lambda}_t)]{\substack{\textbf{Preference-Aware}\\ \text{Pareto Feedback}}} \overbrace{\boldsymbol{\alpha}_{t+1} \rightarrow \widetilde{g}_{t+1} \rightarrow \mathbf{y}_{t+1}}^{\text{Optimistic Response Selection}} \tag{8}$$

Thus, PAOS turns the current response into a source of Pareto feedback that actively shapes the response formed at the next iteration.

Appendix C.1 specifies the selection schedule; Appendices C.2-C.4 provide finite-step solver details, computational complexity, and the ablation mapping.

3.4 Theoretical Analysis

We characterize optimistic selection and local response shaping, and derive a conditional bound on the effect of selector variation. Proofs and assumptions are deferred to Appendix A. For fixed $(\mathbf{x}, \boldsymbol{\alpha})$ and $\rho > 0$, let $\widetilde{g}_\rho(\mathbf{x}, \mathbf{y}; \boldsymbol{\alpha}) := g(\mathbf{x}, \mathbf{y}) + \rho\psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y})$ and define $\mathbf{y}_{\rho,\boldsymbol{\alpha}}(\mathbf{x}) \in \arg\min_{\mathbf{y}\in\mathcal{Y}} \widetilde{g}_\rho(\mathbf{x}, \mathbf{y}; \boldsymbol{\alpha})$. Proposition 1 connects the selected objective to the preference-compatible refinement of the original response set.

Proposition 1 (Optimistic-selection consistency)**.** *Under Assumption 1, as $\rho \downarrow 0$, every accumulation point of $\mathbf{y}_{\rho,\boldsymbol{\alpha}}$ belongs to* $\arg\min_{\mathbf{y}\in\arg\min_{\mathbf{z}} g(\mathbf{x},\mathbf{z})} \psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y})$.

Proposition 2 (First-order preference gain)**.** *Let $\mathbf{y}^0$ be a strict local minimizer of $g(\mathbf{x}, \cdot)$ with $G := \nabla^2_{\mathbf{yy}} g(\mathbf{x}, \mathbf{y}^0) \succ 0$. If g and f_s are three-times continuously differentiable near $(\mathbf{x}, \mathbf{y}^0)$, then, for sufficiently small ρ, there exists a differentiable selected-response branch satisfying*

$$\mathbf{y}_{\rho,\boldsymbol{\alpha}} = \mathbf{y}^0 - \rho G^{-1}\mathbf{a}_{\boldsymbol{\alpha}} + \mathcal{O}(\rho^2), \qquad \mathbf{a}_{\boldsymbol{\alpha}} := \nabla_{\mathbf{y}}\psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}^0). \tag{9}$$

$$\psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}_{\rho,\boldsymbol{\alpha}}) - \psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}^0) = -\rho\mathbf{a}_{\boldsymbol{\alpha}}^\top G^{-1}\mathbf{a}_{\boldsymbol{\alpha}} + \mathcal{O}(\rho^2),\ g(\mathbf{x}, \mathbf{y}_{\rho,\boldsymbol{\alpha}}) - g(\mathbf{x}, \mathbf{y}^0) = \frac{\rho^2}{2}\mathbf{a}_{\boldsymbol{\alpha}}^\top G^{-1}\mathbf{a}_{\boldsymbol{\alpha}} + \mathcal{O}(\rho^3). \tag{10}$$

Algorithm 1 Preference-Aware Optimistic Selection (PAOS)

Require: Initial states $\mathbf{x}_0, \mathbf{y}_{-1}$; preference $\mathbf{r}$; outer, lower, and CG steps N, T, K
Require: Learning rates η_x, η_y; feedback rate β; damping δ; selection schedule $\{\rho_t\}_{t=0}^{N-1}$

$\boldsymbol{\alpha}_0 \leftarrow \mathbf{r}$; initialize CG states $\{\mathbf{v}_{-1,s}\}_{s=1}^{S}$
for $t = 0, \dots, N-1$ **do**
  ▷ **Selected-response formation**
  $\widetilde{g}_t \leftarrow g + \rho_t \langle \boldsymbol{\alpha}_t, \mathbf{F} \rangle$; $\mathbf{y}_t^{(0)} \leftarrow \mathbf{y}_{t-1}$
  for $j = 0, \dots, T-1$ **do**
    $\mathbf{y}_t^{(j+1)} \leftarrow \mathbf{y}_t^{(j)} - \eta_y \nabla_{\mathbf{y}} \widetilde{g}_t(\mathbf{x}_t, \mathbf{y}_t^{(j)})$
  end for
  $\mathbf{y}_t \leftarrow \mathbf{y}_t^{(T)}$; evaluate $\boldsymbol{\Phi}_t$
  ▷ **Response-conditioned geometry**
  for $s = 1, \dots, S$ **do**
    Approximately solve $(\nabla^2_{\mathbf{yy}} \widetilde{g}_t + \delta I)\mathbf{v}_{t,s} = \nabla_{\mathbf{y}} f_s$ using K warm-started CG steps
    $\widehat{\mathbf{h}}_{t,s} \leftarrow \nabla_{\mathbf{x}} f_s - \nabla^2_{\mathbf{xy}} \widetilde{g}_t \mathbf{v}_{t,s}$
  end for
  $\mathbf{H}_t \leftarrow [\widehat{\mathbf{h}}_{t,1}, \dots, \widehat{\mathbf{h}}_{t,S}]$; solve Eq. (5) for $\boldsymbol{\lambda}_t$
  $\mathbf{d}_t \leftarrow \mathbf{H}_t(\mathbf{r} \odot \boldsymbol{\lambda}_t)$; $\mathbf{x}_{t+1} \leftarrow \Pi_{\mathcal{X}}(\mathbf{x}_t - \eta_x \mathbf{d}_t)$
  ▷ **Preference-aware Pareto feedback**
  $\mathbf{q}_t \leftarrow \mathcal{N}(\mathbf{r} \odot \boldsymbol{\lambda}_t)$, $\boldsymbol{\alpha}_{t+1} \leftarrow (1-\beta)\boldsymbol{\alpha}_t + \beta \mathbf{q}_t$
end for
return the evaluated trajectory and selector state $\{(\mathbf{x}_t, \mathbf{y}_t)\}_{t=0}^{N-1}, \boldsymbol{\alpha}_N$

For the following bound, consider $\mathcal{X} = \mathbb{R}^p$ and a fixed ρ after warm-up, reindexing this phase from $t = 0$. Let $\bar{\mathbf{d}}_t$ be the exact direction on a smooth selected-response branch and let $\mathbf{e}_t := \mathbb{E}_t[\mathbf{d}_t] - \bar{\mathbf{d}}_t$ be the total conditional direction bias, including response, hypergradient, and Pareto-mixture approximation. Define $V_N := \sum_{t=0}^{N-1} \mathbb{E}\|\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t\|$ and $\bar{e}_N^2 := N^{-1} \sum_{t=0}^{N-1} \mathbb{E}\|\mathbf{e}_t\|^2$.

Theorem 1 (Conditional feedback-variation bound)**.** *Suppose the fixed-selector coordinator admits a lower-bounded merit function satisfying sufficient-descent and selector-Lipschitz properties in Assumption 4. For a step size satisfying the assumption, the feedback-coupled iterates obey*

$$\frac{1}{N} \sum_{t=0}^{N-1} \mathbb{E}\|\bar{\mathbf{d}}_t\|^2 \le \frac{C_0}{\eta_x N} + \frac{C_1 V_N}{\eta_x N} + C_2 \bar{e}_N^2 + C_3 \eta_x \sigma^2. \tag{11}$$

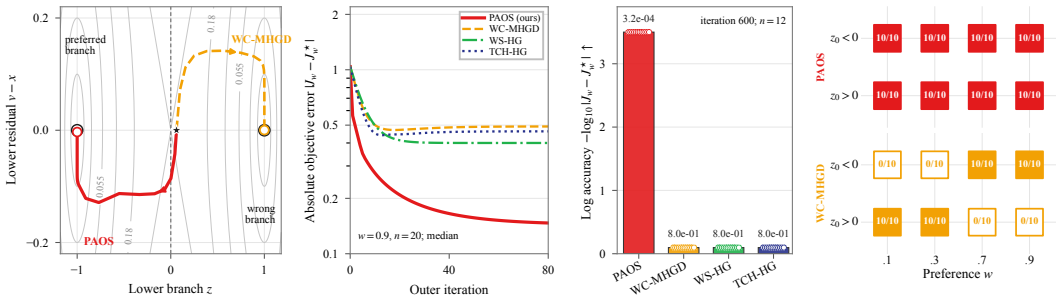
Here $C_0, \dots, C_3$ are independent of N, and σ^2 bounds the conditional direction-noise variance. If all terms on the right vanish, the averaged squared exact direction tends to zero.

The bound isolates the cost of evolving feedback, conditional on descent of the base coordinator; it does not establish that property for Eq. (5). Appendix A.4 separates hypergradient error from Pareto-mixture error, and Appendix A.6 gives the telescoping argument.

4 Experiments

Benchmarks. We evaluate PAOS on an analytical double-well problem and four multi-objective bilevel learning benchmarks. *Data hyper-cleaning* uses MNIST to learn a shared sample-weighting policy across five label-corruption levels. *Few-shot bilevel learning* uses a fixed 5-way 5-shot support-query task from FC-100 (Oreshkin et al., 2018), coordinating five class-wise losses on the query set used for upper-level optimization. *Physics-informed learning* reweights residual, initial-condition, boundary, and observation points for a neural solver of the one-dimensional heat equation (Raissi et al., 2019). *Medical segmentation* learns the training-loss composition of a compact U-Net on Kvasir-SEG (Jha et al., 2020), balancing regional overlap, boundary fidelity, and small-lesion quality. These tasks span sample weights, shared representations, physical training points, and loss mixtures as upper-level decisions. Complete task formulations, data splits, models, and metrics are given in Appendix D.

Baselines and implementation. We compare PAOS with MOML (Ye et al., 2021), FORUM (Ye et al., 2024), gMOBA (Yang et al., 2024), MGDA-HG (Sener & Koltun, 2018; Zhang et al., 2026a),



(a) Response Mechanism **(b)** Error Convergence **(c)** Final Robustness **(d)** Selection Consistency

Figure 3: **Pareto feedback shapes response attraction basins.** (a,b,d) Near-boundary branch selection on the 80-step problem; (b) uses weighted-sum reference regret. (c) Long-horizon refinement on a separate aligned-objective control (Appendix B).

Table 1: MNIST test accuracy. Best and second-best values in each row are **bold** and underlined.

Corruption	MOML	FORUM	gMOBA	MGDA-HG	WS-HG	TCH-HG	WC-MHGD	**PAOS (ours)**
0% ↑	81.36±0.08	73.21±1.09	84.14±0.08	84.04±0.00	84.04±0.00	84.04±0.00	86.98±0.03	**87.14±0.00**
10% ↑	81.20±0.28	73.20±0.28	83.69±0.38	83.67±0.30	83.59±0.01	83.59±0.01	86.05±0.13	**86.20±0.03**
20% ↑	80.99±0.47	73.45±0.21	83.45±0.58	83.35±0.44	83.20±0.06	83.20±0.06	84.98±0.06	**85.22±0.11**
30% ↑	79.79±0.52	73.25±0.16	82.29±0.18	82.20±0.08	82.56±0.14	82.56±0.14	83.07±0.13	**83.40±0.11**
40% ↑	79.73±0.10	73.06±0.00	**81.77±0.16**	81.53±0.16	81.13±0.58	81.13±0.58	80.89±0.18	81.68±0.17
Avg. ↑	80.61±0.03	73.23±0.09	83.07±0.28	82.96±0.20	82.90±0.13	82.90±0.13	84.39±0.02	**84.73±0.07**

and WC-MHGD (Zhang et al., 2026a); weighted-sum and Chebyshev hypergradient controls are denoted WS-HG and TCH-HG. Experiments are implemented in PyTorch and run on NVIDIA RTX A6000 GPUs. Task formulations and implementation details are provided in Appendix D.

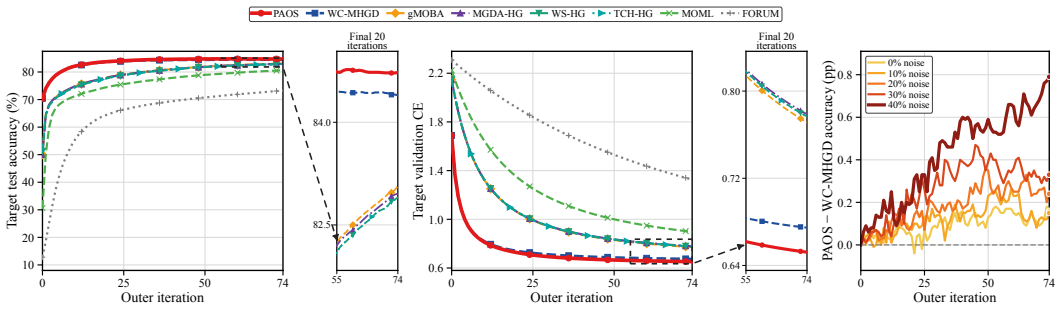
Mechanism analysis: Pareto-guided response formation. We isolate response selection on an analytical double-well problem with two equally valid lower-level branches, while the upper preference w selects one:

$$\mathcal{S}(x) = \{(-1, x), (+1, x)\}, \qquad z_w^\star = -\operatorname{sign}(w - \tfrac{1}{2}), \qquad z_{\text{sep}}(\alpha) = \tfrac{3\rho}{4}(2\alpha_1 - 1). \tag{12}$$

PAOS feeds the evolving Pareto mixture into α_t, moving this attraction boundary before the response settles. Across preferences and paired near-boundary starts, only PAOS consistently reaches the valid preferred branch and attains the lowest projected regret. Long-horizon and initialization-distance controls further confirm robust basin steering (Figure 3; Appendix B). Together, these results show that Pareto feedback changes which valid response is formed, rather than merely improving optimization within an inherited basin.

Data hyper-cleaning results. On MNIST, a shared sample-weighting policy coordinates five corruption-dependent validation objectives. Table 1 reports target-aligned test accuracy: PAOS leads on four corruption levels and attains the best average, with its gain over WC-MHGD growing toward the most corrupted target. Figure 4 shows that PAOS sustains higher target accuracy and lower validation cross-entropy during optimization, while its paired preference gain increases across corruption levels. Appendix D.2 gives the task formulation and evaluation protocol.

Ablation studies. Table 2 isolates three stages of preference-aware response formation. Uniform selection improves on the open-loop baseline; static preference guidance produces a larger gain, and feeding back the evolving Pareto mixture improves every corruption target further. Full PAOS achieves the highest average accuracy and lowest target validation CE. Across the ten preference–seed pairs, full PAOS improves accuracy in nine and ties in one, while reducing target validation CE in all ten. This consistency shows how response selection and dynamic Pareto feedback jointly realize the target preference in lower-level training. Complementary feedback-rate sweeps confirm this effect beyond MNIST: dynamic feedback improves valid branch selection in the double well and lowers PINN prediction error relative to static selection (Appendix Tables 7 and 9).

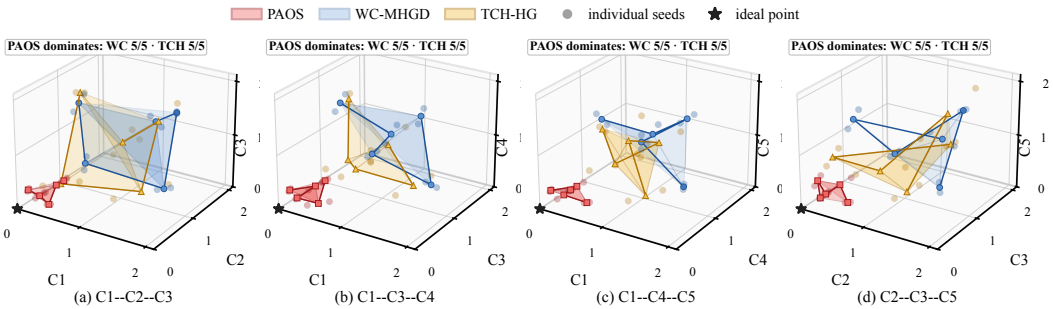


(a) Target-accuracy convergence **(b)** Upper-objective convergence **(c)** Paired preference gain

Figure 4: **MNIST hyper-cleaning trajectories.** **(a)** Target-aligned accuracy; **(b)** target validation cross-entropy; **(c)** paired accuracy gains over WC-MHGD across corruption preferences.

Table 2: **Component ablation on MNIST hyper-cleaning.** Reports target-aligned test accuracy (%) and target validation CE across 5 corruption levels. Best in **bold**, second in underlined.

Variant	0% ↑	10% ↑	20% ↑	30% ↑	40% ↑	Avg. Acc. ↑	Target Val. CE ↓
Open loop ($\rho_t = 0$)	86.98±0.03	86.05±0.13	84.98±0.06	83.07±0.13	80.89±0.18	84.39±0.02	0.6746±0.0042
Uniform selection ($\alpha_t = \mathbf{1}/S$)	87.01±0.04	86.08±0.14	85.04±0.06	83.11±0.16	81.13±0.13	84.47±0.05	0.6699±0.0042
Static preference ($\alpha_t = \mathbf{r}$)	87.12±0.00	86.17±0.01	85.13±0.04	83.32±0.14	81.54±0.00	84.66±0.04	0.6565±0.0041
Full PAOS	**87.14±0.00**	**86.20±0.03**	**85.22±0.11**	**83.40±0.11**	**81.68±0.17**	**84.73±0.07**	**0.6522±0.0041**



(a) C1--C2--C3 (b) C1--C3--C4 (c) C1--C4--C5 (d) C2--C3--C5

Figure 5: **FC-100 preference-conditioned trade-off projections.** Four three-objective views show five class-wise query preferences. The other six views appear in Appendix Figure 7.

Table 3: FC-100: accuracy on the optimization query set. Best in **bold**, second in underlined.

Metric	MOML	FORUM	gMOBA	MGDA-HG	WS-HG	TCH-HG	WC-MHGD	**PAOS (ours)**
Final Acc. ↑	64.67 ± 12.26	30.00 ± 10.37	86.67 ± 3.77	92.00 ± 3.77	74.67 ± 2.64	74.00 ± 7.35	95.47 ± 0.00	**96.27 ± 1.13**
Last-100 Acc. ↑	64.92 ± 12.61	30.29 ± 6.97	84.23 ± 4.64	92.67 ± 0.20	72.30 ± 2.42	67.35 ± 4.14	94.30 ± 0.37	**94.85 ± 0.67**
Acc. AUC ↑	64.43 ± 12.38	32.79 ± 4.94	70.53 ± 1.81	78.09 ± 1.58	56.64 ± 0.30	51.55 ± 1.20	80.30 ± 0.60	**80.52 ± 0.80**

Few-shot bilevel learning results. On the fixed 5-way 5-shot task, PAOS shapes finite-budget classifier responses to improve the five objectives on the optimization query set. PAOS dominates all five preference means of both WC-MHGD and TCH-HG in each of the ten three-objective projections (Figure 5 and Appendix Figure 7). This trade-off advantage is complemented by the best final, last-100, and trajectory-average accuracies on that same query set (Table 3). These results characterize within-task optimization and its preference-conditioned trade-off geometry.

Physics-informed learning results. We study point reweighting for a heat-equation PINN with residual, initial-condition, boundary-condition, and observation objectives. Across terminal, trajectory-average, and late-stage relative L_2 errors, PAOS achieves the best results (Table 4). Holding all other settings fixed, dynamic feedback at $\beta = 0.25$ reduces relative L_2 error by 23.6% and test MSE by 49.4% over static selection ($\beta = 0$), with paired improvements in $6/8$ test runs and $7/8$ validation runs. The complete feedback-rate sweep appears in Appendix D.4. Together, these

Table 4: Heat-equation PINN point-reweighting: test relative L_2 error. AUC averages the trajectory; Final-20 averages the last 20 iterations; PAOS uses $\beta = 0.25$. Best and second-best values are **bold** and underlined.

Metric	MOML	FORUM	gMOBA	MGDA-HG	WS-HG	TCH-HG	WC-MHGD	**PAOS (ours)**
Test rel. $L_2 \downarrow$	0.055 ± 0.008	0.411 ± 0.050	0.032 ± 0.001	0.039 ± 0.009	0.032 ± 0.001	0.032 ± 0.000	0.055 ± 0.034	$\mathbf{0.028 \pm 0.000}$
Rel.-L_2 AUC $\downarrow$	0.128 ± 0.000	0.484 ± 0.060	0.087 ± 0.006	0.087 ± 0.006	0.087 ± 0.007	0.087 ± 0.007	0.088 ± 0.008	$\mathbf{0.082 \pm 0.005}$
Final-20 rel. $L_2 \downarrow$	0.062 ± 0.009	0.423 ± 0.051	0.051 ± 0.021	0.052 ± 0.022	0.051 ± 0.022	0.051 ± 0.022	0.054 ± 0.026	$\mathbf{0.040 \pm 0.003}$

Table 5: Kvasir-SEG uniform-preference results with a compact U-Net.

Method	Dice ↑	IoU ↑	Boundary F1 ↑	Small Dice ↑	HD95 ↓
MOML	0.4281 ± 0.0560	0.2996 ± 0.0493	0.3294 ± 0.0257	0.2374 ± 0.0415	32.6203 ± 2.6727
FORUM	0.3612 ± 0.0168	0.2416 ± 0.0150	0.2386 ± 0.0283	0.1305 ± 0.0033	46.4731 ± 4.2473
WC-MHGD	0.4711 ± 0.0314	0.3422 ± 0.0266	$\mathbf{0.3676 \pm 0.0135}$	0.2496 ± 0.0199	32.4722 ± 1.2692
PAOS	$\mathbf{0.4849 \pm 0.0231}$	$\mathbf{0.3535 \pm 0.0189}$	0.3611 ± 0.0239	$\mathbf{0.2623 \pm 0.0278}$	$\mathbf{31.6363 \pm 1.2706}$

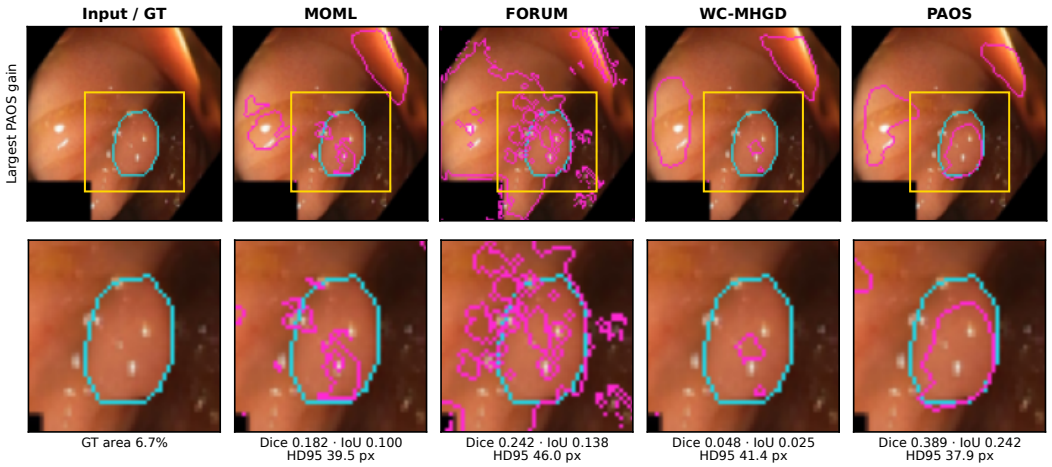


Figure 6: Qualitative Kvasir-SEG comparison on the mid-size test case with the largest PAOS gain.

ablations separate two roles: static preference initializes response selection, while dynamic Pareto feedback adapts it as response-conditioned trade-offs evolve.

Medical segmentation results. On Kvasir-SEG, PAOS yields the strongest overall uniform-preference segmentation profile (Table 5), improving regional overlap, small-lesion recovery, and boundary distance over WC-MHGD. The qualitative comparison in Figure 6 illustrates the corresponding mask fidelity; further lesion-size comparisons appear in Appendix E.2.

Quality-efficiency analysis. Table 10 shows that PAOS attains the best target accuracy, final CE, and CE AUC. It reaches CE ≤ 0.8 in 11 iterations-about $2.8\times$ faster than gMOBA and MGDA-HG at this threshold-while using over $10\times$ less allocated memory than unrolled MOML.

5 Conclusion

PAOS closes the Pareto-response loop by turning each preference-conditioned mixture into a stateful selector that shapes the next response and its objective geometry. Double-well and PINN ablations isolate feedback gains beyond static selection, while four application domains demonstrate consistent gains from feedback-guided response formation. These results establish response formation as an actionable decision variable in multi-objective bilevel learning. **Limitation and future work.** PAOS is developed for differentiable MOBL with a scalar lower-level objective and a simplex-valued Pareto coordinator. Extending closed-loop response selection to genuinely multi-objective lower levels and black-box evolutionary settings is a natural next step (Chauhan et al., 2026).

AI USE STATEMENT

Generative AI tools were used to assist with code development and debugging, theoretical and experimental analysis, literature organization, language editing, and manuscript preparation. The authors verified and take responsibility for all technical content.

REPRODUCIBILITY STATEMENT

The paper and appendix specify the update equations, task definitions, optimization settings, preference protocols, evaluation metrics, and seed aggregation procedures. The anonymous code artifact referenced in the abstract provides the corresponding implementation and experiment configurations. Our code is available at Repository.

REFERENCES

Michael Arbel and Julien Mairal. Non-convex bilevel games with critical point selection maps. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Dikshit Chauhan, Hisao Ishibuchi, Kalyanmoy Deb, Anupam Trivedi, and Dipti Srinivasan. A review on bilevel optimization using evolutionary algorithms and machine learning approaches. *IEEE Transactions on Evolutionary Computation*, 2026.

Lisha Chen, A. F. M. Saif, Yanning Shen, and Tianyi Chen. FERERO: A flexible framework for preference-guided multi-objective learning. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Lisha Chen, Quan Xiao, Ellen Hidemi Fukuda, Xinyi Chen, Kun Yuan, and Tianyi Chen. Efficient first-order optimization on the pareto set for multi-objective learning under preference guidance. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 9443–9486. PMLR, 2025.

Mathieu Dagréou, Pierre Ablin, Samuel Vaiter, and Thomas Moreau. A framework for bilevel optimization that enables stochastic and global variance reduction algorithms. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26698–26710, 2022.

Jean-Antoine Désidéri. Multiple-gradient descent algorithm (MGDA) for multiobjective optimization. *Comptes Rendus. Mathématique*, 350(5–6):313–318, 2012.

Heshan Devaka Fernando, Han Shen, Miao Liu, Subhajit Chaudhury, Keerthiram Murugesan, and Tianyi Chen. Mitigating gradient bias in multi-objective learning: A provably convergent approach. In *The Eleventh International Conference on Learning Representations*, 2023.

Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimiliano Pontil. Bilevel programming for hyperparameter optimization and meta-learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1568–1577. PMLR, 2018.

Riccardo Grazzi, Luca Franceschi, Massimiliano Pontil, and Saverio Salzo. On the iteration complexity of hypergradient computation. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3748–3758. PMLR, 2020.

Alex Gu, Songtao Lu, Parikshit Ram, and Tsui-Wei Weng. Min-max multi-objective bilevel optimization with applications in robust machine learning. In *The Eleventh International Conference on Learning Representations*, 2023.

Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D. Johansen. Kvasir-SEG: A segmented polyp dataset. In *MultiMedia Modeling*, pp. 451–462. Springer, 2020.

Yicong Jiang and Feihu Huang. Efficient hessian-free methods for multi-objective bilevel optimization with nonconvex lower level. *arXiv preprint arXiv:2608.12704*, 2026.

Xi Lin, Hui-Ling Zhen, Zhenhua Li, Qing-Fu Zhang, and Sam Kwong. Pareto multi-task learning. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. In *Advances in Neural Information Processing Systems*, volume 34, pp. 18878–18890, 2021.

Risheng Liu, Pan Mu, Xiaoming Yuan, Shangzhi Zeng, and Jin Zhang. A generic first-order algorithmic framework for bi-level programming beyond lower-level singleton. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6305–6315. PMLR, 2020.

Risheng Liu, Yaohua Liu, Wei Yao, Shangzhi Zeng, and Jin Zhang. Averaged method of multipliers for bi-level optimization without lower-level strong convexity. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 21839–21866. PMLR, 2023.

Jonathan Lorraine, Paul Vicol, and David Duvenaud. Optimizing millions of hyperparameters by implicit differentiation. In *Proceedings of the Twenty-Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 1540–1552. PMLR, 2020.

Dougal Maclaurin, David Duvenaud, and Ryan Adams. Gradient-based hyperparameter optimization through reversible learning. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2113–2122. PMLR, 2015.

Debabrata Mahapatra and Vaibhav Rajan. Multi-task learning with user preferences: Gradient descent with controlled ascent in pareto optimization. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6597–6607. PMLR, 2020.

Saeed Masiha, Zebang Shen, Negar Kiyavash, and Niao He. Select-then-differentiate: Solving bilevel optimization with manifold lower-level solution sets. *arXiv preprint arXiv:2605.09209*, 2026.

Aviv Navon, Aviv Shamsian, Ethan Fetaya, and Gal Chechik. Learning the pareto front with hypernetworks. In *International Conference on Learning Representations*, 2021.

Aviv Navon, Aviv Shamsian, Idan Achituve, Haggai Maron, Kenji Kawaguchi, Gal Chechik, and Ethan Fetaya. Multi-task learning as a bargaining game. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 16428–16446. PMLR, 2022.

Boris N. Oreshkin, Pau Rodriguez, and Alexandre Lacoste. TADAM: task dependent adaptive metric for improved few-shot learning. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 719–729, Red Hook, NY, USA, 2018. Curran Associates Inc.

Fabian Pedregosa. Hyperparameter optimization with approximate gradient. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 737–746. PMLR, 2016.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Aravind Rajeswaran, Chelsea Finn, Sham M. Kakade, and Sergey Levine. Meta-learning with implicit gradients. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4334–4343. PMLR, 2018.

Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Amirreza Shaban, Ching-An Cheng, Nathan Hatch, and Byron Boots. Truncated back-propagation for bilevel optimization. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1723–1732. PMLR, 2019.

Paul Vicol, Jonathan P. Lorraine, Fabian Pedregosa, David Duvenaud, and Roger B. Grosse. On implicit bias in overparameterized bilevel optimization. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 22234–22259. PMLR, 2022.

Xinmin Yang, Wei Yao, Haian Yin, Shangzhi Zeng, and Jin Zhang. Gradient-based algorithms for multi-objective bi-level optimization. *Science China Mathematics*, 67:1419–1438, 2024.

Feiyang Ye, Baijiong Lin, Zhixiong Yue, Pengxin Guo, Qiao Xiao, and Yu Zhang. Multi-objective meta learning. In *Advances in Neural Information Processing Systems*, volume 34, pp. 21338–21351, 2021.

Feiyang Ye, Baijiong Lin, Xiaofeng Cao, Yu Zhang, and Ivor W. Tsang. A first-order multi-gradient algorithm for multi-objective bi-level optimization. In *ECAI 2024: 27th European Conference on Artificial Intelligence*, pp. 2621–2628, 2024.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Zhiyao Zhang, Zhuqing Liu, Xin Zhang, Wen-Yen Chen, Jiyan Yang, and Jia Liu. Multi-objective bilevel learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(34):28653–28661, 2026a.

Zhiyao Zhang, Myeung Suk Oh, Zhen Qin, Jiaxiang Li, Xin Zhang, and Jia Liu. A tale of two problems: Multi-task bilevel learning meets equality constrained multi-objective optimization. *arXiv preprint arXiv:2605.09094*, 2026b.

APPENDIX OUTLINE

A Theoretical Analysis of PAOS **14**

A.1 Definitions and Assumptions 14

A.2 Proof of Optimistic-Selection Consistency 15

A.3 Local Response and Geometry Perturbation 15

A.4 Hypergradient Approximation Error 16

A.5 Feedback Tracking 17

A.6 Proof of the Conditional Feedback-Variation Bound 17

B Details of the Analytical Mechanism Study **18**

C Additional Algorithmic and Complexity Details **20**

C.1 Selector Initialization and Selection Schedule 20

C.2 Warm Starts and Finite-Step Solvers 20

C.3 Oracle Accounting 21

C.4 Ablation Configurations 21

D Experimental Details **21**

D.1 Preferences and Implementation 21

D.2 MNIST Data Hyper-Cleaning 21

D.3 Few-shot Bilevel Learning on FC-100 22

D.4 Physics-Informed Learning on the Heat Equation 23

D.5 Kvasir-SEG Uniform-Preference Loss Design 25

E Additional Experimental Results **26**

E.1 Quality-to-Target Efficiency 26

E.2 Additional Kvasir-SEG Results 26

A THEORETICAL ANALYSIS OF PAOS

This appendix provides the assumptions and proofs supporting Section 3.4. We separate the set-valued optimistic-selection result, which permits non-convex and non-unique lower-level solutions, from the local smooth analysis used for hypergradient computation and the conditional feedback-variation bound.

A.1 DEFINITIONS AND ASSUMPTIONS

For the scalar lower-level objective in Eq. (1), fix $\boldsymbol{\alpha} \in \Delta_S$ and define

$$\begin{aligned} \widetilde{g}_\rho(\mathbf{x}, \mathbf{y}; \boldsymbol{\alpha}) &:= g(\mathbf{x}, \mathbf{y}) + \rho \psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}) \\ &= g(\mathbf{x}, \mathbf{y}) + \rho \sum_{s=1}^{S} \alpha_s f_s(\mathbf{x}, \mathbf{y}), \\ \psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}) &:= \langle \boldsymbol{\alpha}, \mathbf{F}(\mathbf{x}, \mathbf{y}) \rangle. \end{aligned} \tag{13}$$

The main-text objective $\widetilde{g}_t$ in Eq. (3) is the time-varying specialization $\widetilde{g}_{\rho_t}(\cdot, \cdot; \boldsymbol{\alpha}_t)$. Let

$$\begin{aligned} \mathcal{Y}^\star(\mathbf{x}) &:= \arg\min_{\mathbf{y} \in \mathcal{Y}} g(\mathbf{x}, \mathbf{y}), \\ \mathcal{Y}^{\mathrm{opt}}_{\boldsymbol{\alpha}}(\mathbf{x}) &:= \arg\min_{\mathbf{y} \in \mathcal{Y}^\star(\mathbf{x})} \psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}). \end{aligned} \tag{14}$$

For the local smooth analysis, select an interior local-minimizer branch $\mathbf{y}_\rho(\mathbf{x}, \boldsymbol{\alpha})$ satisfying $\nabla_{\mathbf{y}} \widetilde{g}_\rho(\mathbf{x}, \mathbf{y}_\rho; \boldsymbol{\alpha}) = 0$ and the positive Hessian condition below. This branch need not be a global minimizer, unlike the responses used in Proposition 1. Define

$$\boldsymbol{\Phi}_\rho(\mathbf{x}, \boldsymbol{\alpha}) := \mathbf{F}\big(\mathbf{x}, \mathbf{y}_\rho(\mathbf{x}, \boldsymbol{\alpha})\big), \tag{15}$$

$$A_\rho(\mathbf{x}, \boldsymbol{\alpha}) := \nabla^2_{\mathbf{yy}} \widetilde{g}_\rho\big(\mathbf{x}, \mathbf{y}_\rho(\mathbf{x}, \boldsymbol{\alpha}); \boldsymbol{\alpha}\big),$$

$$B_\rho(\mathbf{x}, \boldsymbol{\alpha}) := \nabla^2_{\mathbf{xy}} \widetilde{g}_\rho\big(\mathbf{x}, \mathbf{y}_\rho(\mathbf{x}, \boldsymbol{\alpha}); \boldsymbol{\alpha}\big), \tag{16}$$

$$\mathbf{h}_{\rho,s}(\mathbf{x}, \boldsymbol{\alpha}) := \nabla_{\mathbf{x}} f_s - B_\rho A_\rho^{-1} \nabla_{\mathbf{y}} f_s, \tag{17}$$

$$\mathbf{H}_\rho(\mathbf{x}, \boldsymbol{\alpha}) := [\mathbf{h}_{\rho,1}(\mathbf{x}, \boldsymbol{\alpha}), \ldots, \mathbf{h}_{\rho,S}(\mathbf{x}, \boldsymbol{\alpha})].$$

where the derivatives of f_s in Eq. (17) are evaluated at $(\mathbf{x}, \mathbf{y}_\rho(\mathbf{x}, \boldsymbol{\alpha}))$. Let $\boldsymbol{\lambda}(\mathbf{x}, \boldsymbol{\alpha}) \in \Delta_S$ denote the solution selected by the Pareto subproblem in Eq. (5), and define its exact direction by $\bar{\mathbf{d}}(\mathbf{x}, \boldsymbol{\alpha}) := \mathbf{H}_\rho(\mathbf{x}, \boldsymbol{\alpha})(\mathbf{r} \odot \boldsymbol{\lambda}(\mathbf{x}, \boldsymbol{\alpha}))$.

The following assumptions are used selectively; Proposition 1 requires only Assumption 1, whereas the local implicit analysis uses Assumptions 2–3.

Assumption 1 (Existence and compactness)**.** *For every $\mathbf{x}$ in the region of interest, $g(\mathbf{x}, \cdot)$ and $\psi_{\boldsymbol{\alpha}}(\mathbf{x}, \cdot)$ are continuous. The relevant lower-level sublevel set is compact, and both $\mathcal{Y}^\star(\mathbf{x})$ and the minimizer set of $\widetilde{g}_\rho(\mathbf{x}, \cdot; \boldsymbol{\alpha})$ are nonempty for sufficiently small $\rho > 0$.*

Assumption 2 (Local response regularity)**.** *Along the selected branch visited by PAOS, $\widetilde{g}_\rho(\mathbf{x}, \cdot; \boldsymbol{\alpha})$ is locally μ-strongly convex for some $\mu > 0$. Equivalently, its Hessian satisfies $\nabla^2_{\mathbf{yy}} \widetilde{g}_\rho \succeq \mu I$ in a neighborhood containing the iterates and the exact selected response.*

Assumption 3 (Smoothness and bounded derivatives)**.** *The functions g and f_s are three-times continuously differentiable in the local response region. Their gradients, mixed Hessians, and lower-level Hessians are Lipschitz, and the required derivatives are uniformly bounded. The upper iterates remain in a compact subset of $\mathcal{X}$.*

Assumption 4 (Conditional base-coordinator descent)**.** *On an unconstrained upper domain $\mathcal{X} = \mathbb{R}^p$, fix ρ and a smooth selected-response branch. Assume that there exists a merit function $\mathcal{M}(\mathbf{x}, \boldsymbol{\alpha}) \geq \mathcal{M}_{\inf} > -\infty$ uniformly over the visited region. Let $\mathbb{E}_t$ condition on the history before computing the iteration-t direction. Write the implemented direction as $\mathbf{d}_t = \bar{\mathbf{d}}_t + \mathbf{e}_t + \boldsymbol{\xi}_t$, where $\mathbf{e}_t := \mathbb{E}_t[\mathbf{d}_t] - \bar{\mathbf{d}}_t$ is the total conditional bias and $\mathbb{E}_t[\boldsymbol{\xi}_t] = 0$, $\mathbb{E}_t\|\boldsymbol{\xi}_t\|^2 \leq \sigma^2$. There exist constants $c_d, c_e, c_\sigma, L_\alpha > 0$ such that, for a sufficiently small η_x,*

$$\begin{aligned} \mathbb{E}_t[\mathcal{M}(\mathbf{x}_{t+1}, \boldsymbol{\alpha}_t)] \leq \mathcal{M}(\mathbf{x}_t, \boldsymbol{\alpha}_t) &- c_d \eta_x \|\bar{\mathbf{d}}_t\|^2 \\ &+ c_e \eta_x \|\mathbf{e}_t\|^2 + c_\sigma \eta_x^2 \sigma^2, \end{aligned} \tag{18}$$

$$|\mathcal{M}(\mathbf{x}, \boldsymbol{\alpha}) - \mathcal{M}(\mathbf{x}, \boldsymbol{\alpha}')| \leq L_\alpha \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|. \tag{19}$$

Here $\mathbf{x}_{t+1} = \mathbf{x}_t - \eta_x \mathbf{d}_t$. The first inequality is an explicit condition on the base coordinator, not a consequence proved here for the weighted-Chebyshev subproblem. The second controls selector-induced drift. The theorem applies to a fixed-ρ phase; any finite warm-up precedes its initial state.

Assumption 2 supports implicit differentiation along the selected branch while allowing multiple response branches globally.

A.2 Proof of Optimistic-Selection Consistency

Proof of Proposition 1. Fix $\mathbf{x}$ and $\boldsymbol{\alpha}$, and abbreviate $g(\mathbf{y}) := g(\mathbf{x}, \mathbf{y})$ and $\psi(\mathbf{y}) := \psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y})$. Let $g^\star := \min_{\mathbf{y}} g(\mathbf{y})$ and choose any $\mathbf{y}^\star \in \mathcal{Y}^\star(\mathbf{x})$. By optimality of $\mathbf{y}_{\rho,\boldsymbol{\alpha}}$,

$$g(\mathbf{y}_{\rho,\boldsymbol{\alpha}}) + \rho\psi(\mathbf{y}_{\rho,\boldsymbol{\alpha}}) \le g^\star + \rho\psi(\mathbf{y}^\star). \tag{20}$$

Because the relevant sublevel set is compact, ψ is bounded below there by some $m_\psi > -\infty$. Hence

$$0 \le g(\mathbf{y}_{\rho,\boldsymbol{\alpha}}) - g^\star \le \rho\big(\psi(\mathbf{y}^\star) - m_\psi\big), \tag{21}$$

which converges to zero as $\rho \downarrow 0$. Every accumulation point $\bar{\mathbf{y}}$ of $\mathbf{y}_{\rho,\boldsymbol{\alpha}}$ therefore lies in $\mathcal{Y}^\star(\mathbf{x})$ by continuity of g.

Furthermore, since $g(\mathbf{y}_{\rho,\boldsymbol{\alpha}}) \ge g^\star$, Eq. (20) implies

$$\psi(\mathbf{y}_{\rho,\boldsymbol{\alpha}}) \le \psi(\mathbf{y}^\star) - \frac{g(\mathbf{y}_{\rho,\boldsymbol{\alpha}}) - g^\star}{\rho} \le \psi(\mathbf{y}^\star). \tag{22}$$

Since $\mathbf{y}^\star$ was arbitrary in $\mathcal{Y}^\star(\mathbf{x})$, taking a convergent subsequence and using continuity of ψ yields

$$\psi(\bar{\mathbf{y}}) \le \psi(\mathbf{y}), \qquad \forall \mathbf{y} \in \mathcal{Y}^\star(\mathbf{x}). \tag{23}$$

Thus $\bar{\mathbf{y}} \in \mathcal{Y}^{\mathrm{opt}}_{\boldsymbol{\alpha}}(\mathbf{x})$. □

A.3 Local Response and Geometry Perturbation

Proof of Proposition 2. Fix $\mathbf{x}$ and define

$$\mathcal{G}(\mathbf{y}, \rho) := \nabla_{\mathbf{y}} g(\mathbf{x}, \mathbf{y}) + \rho \nabla_{\mathbf{y}} \psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}). \tag{24}$$

At $(\mathbf{y}^0, 0)$, we have $\mathcal{G}(\mathbf{y}^0, 0) = 0$ and $\nabla_{\mathbf{y}} \mathcal{G}(\mathbf{y}^0, 0) = G \succ 0$. The implicit function theorem therefore gives a differentiable branch $\mathbf{y}_{\rho,\boldsymbol{\alpha}}$ satisfying $\mathcal{G}(\mathbf{y}_{\rho,\boldsymbol{\alpha}}, \rho) = 0$ for sufficiently small ρ. Differentiating at $\rho = 0$ gives

$$G \left.\frac{\mathrm{d}\mathbf{y}_{\rho,\boldsymbol{\alpha}}}{\mathrm{d}\rho}\right|_{\rho=0} + \mathbf{a}_{\boldsymbol{\alpha}} = 0, \tag{25}$$

and hence

$$\left.\frac{\mathrm{d}\mathbf{y}_{\rho,\boldsymbol{\alpha}}}{\mathrm{d}\rho}\right|_{\rho=0} = -G^{-1}\mathbf{a}_{\boldsymbol{\alpha}}. \tag{26}$$

Taylor expansion proves Eq. (9).

Using $\Delta\mathbf{y} := \mathbf{y}_{\rho,\boldsymbol{\alpha}} - \mathbf{y}^0 = -\rho G^{-1}\mathbf{a}_{\boldsymbol{\alpha}} + \mathcal{O}(\rho^2)$, we obtain

$$\psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}_{\rho,\boldsymbol{\alpha}}) - \psi_{\boldsymbol{\alpha}}(\mathbf{x}, \mathbf{y}^0) = \mathbf{a}_{\boldsymbol{\alpha}}^\top \Delta\mathbf{y} + \mathcal{O}(\|\Delta\mathbf{y}\|^2) \tag{27}$$

$$= -\rho \mathbf{a}_{\boldsymbol{\alpha}}^\top G^{-1}\mathbf{a}_{\boldsymbol{\alpha}} + \mathcal{O}(\rho^2), \tag{28}$$

which proves the first objective expansion in Proposition 2. Since $\nabla_{\mathbf{y}} g(\mathbf{x}, \mathbf{y}^0) = 0$,

$$g(\mathbf{x}, \mathbf{y}_{\rho,\boldsymbol{\alpha}}) - g(\mathbf{x}, \mathbf{y}^0) = \frac{1}{2}\Delta\mathbf{y}^\top G \Delta\mathbf{y} + \mathcal{O}(\|\Delta\mathbf{y}\|^3) \tag{29}$$

$$= \frac{\rho^2}{2}\mathbf{a}_{\boldsymbol{\alpha}}^\top G^{-1}\mathbf{a}_{\boldsymbol{\alpha}} + \mathcal{O}(\rho^3), \tag{30}$$

proving the corresponding lower-objective expansion. □

The same argument yields an objective-wise expansion.

Corollary 1 (Objective and geometry shift)**.** *Under the conditions of Proposition 2, for each $j \in [S]$,*

$$f_j(\mathbf{x}, \mathbf{y}_{\rho,\boldsymbol{\alpha}}) - f_j(\mathbf{x}, \mathbf{y}^0) = -\rho \, \nabla_{\mathbf{y}} f_j(\mathbf{x}, \mathbf{y}^0)^\top G^{-1} \mathbf{a}_{\boldsymbol{\alpha}} + \mathcal{O}(\rho^2). \tag{31}$$

If the derivatives entering the implicit hypergradient are locally Lipschitz, then

$$\|\mathbf{H}_\rho(\mathbf{x}, \boldsymbol{\alpha}) - \mathbf{H}_0(\mathbf{x})\|_{\mathrm{op}} \le C_H \rho \tag{32}$$

for some local constant $C_H > 0$.

Proof. Equation (31) follows from a first-order Taylor expansion of f_j and Eq. (9). The hypergradient bound follows by applying the local Lipschitz assumptions to the gradient, mixed-Hessian, and inverse-Hessian factors in the implicit-gradient formula. □

Define the standard Pareto-stationarity residual of a hypergradient matrix H by

$$\mathfrak{s}(H) := \min_{\mathbf{w} \in \Delta_S} \|H\mathbf{w}\|. \tag{33}$$

Since $\|\mathbf{w}\|_2 \le 1$ for $\mathbf{w} \in \Delta_S$, Corollary 1 also gives

$$\mathfrak{s}(\mathbf{H}_0(\mathbf{x})) \le \mathfrak{s}(\mathbf{H}_\rho(\mathbf{x}, \boldsymbol{\alpha})) + C_H \rho. \tag{34}$$

Thus an ϵ-stationary selected response is locally an $(\epsilon + \mathcal{O}(\rho))$-stationary response on the corresponding unperturbed branch.

A.4 Hypergradient Approximation Error

Let $\mathbf{y}^\star := \mathbf{y}_\rho(\mathbf{x}, \boldsymbol{\alpha})$ be an exact selected response. Define

$$A^\star := \nabla^2_{\mathbf{yy}} \widetilde{g}_\rho(\mathbf{x}, \mathbf{y}^\star; \boldsymbol{\alpha}), \tag{35}$$

$$B^\star := \nabla^2_{\mathbf{xy}} \widetilde{g}_\rho(\mathbf{x}, \mathbf{y}^\star; \boldsymbol{\alpha}), \tag{36}$$

$$\mathbf{h}^\star_s := \nabla_{\mathbf{x}} f_s(\mathbf{x}, \mathbf{y}^\star) - B^\star (A^\star)^{-1} \nabla_{\mathbf{y}} f_s(\mathbf{x}, \mathbf{y}^\star). \tag{37}$$

At an approximate response $\widehat{\mathbf{y}}$, let

$$\widehat{A} := \nabla^2_{\mathbf{yy}} \widetilde{g}_\rho(\mathbf{x}, \widehat{\mathbf{y}}; \boldsymbol{\alpha}), \qquad \widehat{B} := \nabla^2_{\mathbf{xy}} \widetilde{g}_\rho(\mathbf{x}, \widehat{\mathbf{y}}; \boldsymbol{\alpha}), \tag{38}$$

and suppose $\widehat{\mathbf{v}}_s$ has damped linear residual

$$\mathbf{e}_{v,s} := (\widehat{A} + \delta I)\widehat{\mathbf{v}}_s - \nabla_{\mathbf{y}} f_s(\mathbf{x}, \widehat{\mathbf{y}}). \tag{39}$$

Lemma 1 (Hypergradient error decomposition)**.** *Under Assumptions 2 and 3, let $\epsilon_y := \|\nabla_{\mathbf{y}} \widetilde{g}_\rho(\mathbf{x}, \widehat{\mathbf{y}}; \boldsymbol{\alpha})\|$. There exist constants $C_y, C_v, C_\delta > 0$ such that*

$$\|\widehat{\mathbf{h}}_s - \mathbf{h}^\star_s\| \le C_y \epsilon_y + C_v \|\mathbf{e}_{v,s}\| + C_\delta \delta. \tag{40}$$

Proof. Local μ-strong convexity implies

$$\|\widehat{\mathbf{y}} - \mathbf{y}^\star\| \le \mu^{-1} \epsilon_y. \tag{41}$$

The inverse perturbation identity and the derivative Lipschitz bounds therefore control the error caused by replacing $(A^\star, B^\star, \mathbf{y}^\star)$ with $(\widehat{A}, \widehat{B}, \widehat{\mathbf{y}})$ by $\mathcal{O}(\epsilon_y)$. Next,

$$\left\| (\widehat{A} + \delta I)^{-1} - (\widehat{A})^{-1} \right\|_{\mathrm{op}} \le \frac{\delta}{\mu(\mu + \delta)} \le \frac{\delta}{\mu^2}, \tag{42}$$

which gives the damping term. Finally,

$$\left\| \widehat{\mathbf{v}}_s - (\widehat{A} + \delta I)^{-1} \nabla_{\mathbf{y}} f_s(\mathbf{x}, \widehat{\mathbf{y}}) \right\| \le \frac{1}{\mu + \delta} \|\mathbf{e}_{v,s}\|. \tag{43}$$

Multiplying by the uniformly bounded mixed Hessian and combining the three terms proves Eq. (40). □

At iteration t, define the lower-solver residual and aggregate CG residual by

$$\begin{aligned}\epsilon_{\mathrm{in},t} &:= \|\nabla_{\mathbf{y}}\widetilde{g}_t(\mathbf{x}_t,\mathbf{y}_t)\|,\\ \epsilon^2_{\mathrm{CG},t} &:= \sum_{s=1}^{S}\|A_t\mathbf{v}_{t,s}-\nabla_{\mathbf{y}}f_s(\mathbf{x}_t,\mathbf{y}_t)\|^2.\end{aligned} \tag{44}$$

Hypergradient error and Pareto-mixture error enter the direction separately. For $H^\star := \mathbf{H}_\rho(\mathbf{x}_t,\boldsymbol{\alpha}_t)$, let $\boldsymbol{\lambda}^\star$ be the chosen exact subproblem solution and $\widehat{\boldsymbol{\lambda}}$ its computed counterpart. A fixed selection rule is used if the exact solution is non-unique. Then

$$\mathbf{d}_t-\bar{\mathbf{d}}_t=(\mathbf{H}_t-H^\star)(\mathbf{r}\odot\widehat{\boldsymbol{\lambda}})+H^\star\big(\mathbf{r}\odot(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^\star)\big). \tag{45}$$

For deterministic evaluations satisfying Lemma 1, bounded preference weights and $H^\star$ therefore yield

$$\|\mathbf{d}_t-\bar{\mathbf{d}}_t\|^2\le C_{\mathrm{in}}\epsilon^2_{\mathrm{in},t}+C_{\mathrm{CG}}\epsilon^2_{\mathrm{CG},t}+C_{\mathrm{damp}}\delta^2+C_\lambda\|\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}^\star\|^2. \tag{46}$$

Absorbing the final term into response and hypergradient errors requires an additional locally Lipschitz solution selection for the Pareto subproblem with respect to both $(H,\boldsymbol{\Phi})$, and control of numerical QP error. Theorem 1 instead retains the total bias $\mathbf{e}_t=\mathbb{E}_t[\mathbf{d}_t]-\bar{\mathbf{d}}_t$ explicitly, including the mixture error and any bias induced by stochastic subproblem inputs.

A.5 Feedback Tracking

The selector update is a convex combination of the previous state and the current Pareto target:

$$\boldsymbol{\alpha}_{t+1}=(1-\beta)\boldsymbol{\alpha}_t+\beta\mathbf{q}_t. \tag{47}$$

Lemma 2 (Simplex invariance and target tracking)**.** *If $\boldsymbol{\alpha}_0\in\Delta_S$ and $\beta\in[0,1]$, then $\boldsymbol{\alpha}_t\in\Delta_S$ for every t. Moreover,*

$$\|\boldsymbol{\alpha}_{t+1}-\mathbf{q}_{t+1}\|\le(1-\beta)\|\boldsymbol{\alpha}_t-\mathbf{q}_t\|+\|\mathbf{q}_{t+1}-\mathbf{q}_t\|. \tag{48}$$

Consequently, if $\beta\in(0,1]$ and $\mathbf{q}_t\to\mathbf{q}^\star$, then $\boldsymbol{\alpha}_t\to\mathbf{q}^\star$.

Proof. The simplex is convex, so the update preserves simplex membership. Subtracting $\mathbf{q}_{t+1}$ from the update gives

$$\boldsymbol{\alpha}_{t+1}-\mathbf{q}_{t+1}=(1-\beta)(\boldsymbol{\alpha}_t-\mathbf{q}_t)+(\mathbf{q}_t-\mathbf{q}_{t+1}), \tag{49}$$

and the triangle inequality proves Eq. (48). Iterating the recursion proves the convergence statement. □

A.6 Proof of the Conditional Feedback-Variation Bound

Proof of Theorem 1. Condition on the history at t, apply Eq. (18), and account for the selector change using Eq. (19). Taking total expectations gives

$$\begin{aligned}\mathbb{E}\mathcal{M}(\mathbf{x}_{t+1},\boldsymbol{\alpha}_{t+1}) &\le \mathbb{E}\mathcal{M}(\mathbf{x}_{t+1},\boldsymbol{\alpha}_t)+L_\alpha\mathbb{E}\|\boldsymbol{\alpha}_{t+1}-\boldsymbol{\alpha}_t\| &(50)\\ &\le \mathbb{E}\mathcal{M}(\mathbf{x}_t,\boldsymbol{\alpha}_t)-c_d\eta_x\mathbb{E}\|\bar{\mathbf{d}}_t\|^2\\ &\quad+c_e\eta_x\mathbb{E}\|\mathbf{e}_t\|^2+c_\sigma\eta_x^2\sigma^2\\ &\quad+L_\alpha\mathbb{E}\|\boldsymbol{\alpha}_{t+1}-\boldsymbol{\alpha}_t\|. &(51)\end{aligned}$$

Summing from $t=0$ to $N-1$ and using the lower bound $\mathcal{M}_{\mathrm{inf}}$ yields

$$\begin{aligned}c_d\eta_x\sum_{t=0}^{N-1}\mathbb{E}\|\bar{\mathbf{d}}_t\|^2 &\le \mathbb{E}\mathcal{M}(\mathbf{x}_0,\boldsymbol{\alpha}_0)-\mathcal{M}_{\mathrm{inf}}+L_\alpha V_N\\ &\quad+c_e\eta_x\sum_{t=0}^{N-1}\mathbb{E}\|\mathbf{e}_t\|^2+c_\sigma N\eta_x^2\sigma^2.\end{aligned} \tag{52}$$

Dividing by $c_d\eta_x N$ proves Eq. (11), with $C_0=(\mathbb{E}\mathcal{M}(\mathbf{x}_0,\boldsymbol{\alpha}_0)-\mathcal{M}_{\mathrm{inf}})/c_d$, $C_1=L_\alpha/c_d$, $C_2=c_e/c_d$, and $C_3=c_\sigma/c_d$. The asymptotic implication follows when $1/(\eta_x N)$, $V_N/(\eta_x N)$, $\bar{e}^2_N$, and $\eta_x\sigma^2$ all tend to zero. □

The term $V_N/(\eta_x N)$ isolates selector drift. A constant feedback rate alone does not imply sublinear V_N. For $r_{\min} := \min_s r_s > 0$, $\mathfrak{s}(H^\star) \le \|\bar{\mathbf{d}}_t\|/r_{\min}$ by normalizing $\mathbf{r} \odot \boldsymbol{\lambda}^\star$. Thus a vanishing exact direction implies stationarity of the smooth selected-response branch under the stated conditions, rather than a global convergence guarantee across response basins.

B Details of the Analytical Mechanism Study

Problem formulation. As introduced in Section 4, we construct an analytical double-well problem to isolate whether Pareto feedback can actively shape lower-level response formation. Let the upper variable be $x \in \mathbb{R}$ and the lower variable be $(z, v) \in \mathbb{R}^2$. The lower-level objective and the two upper-level objectives are defined as

$$g(x, z, v) = \frac{1}{4}(z^2 - 1)^2 + \frac{1}{2}(v - x)^2, \tag{53}$$

$$b(z) = \frac{2 + 3z - z^3}{4}, \tag{54}$$

$$f_1(x, z, v) = (v + 1)^2 + b(z), \tag{55}$$

$$f_2(x, z, v) = (v - 1)^2 + 1 - b(z). \tag{56}$$

For every $x \in \mathbb{R}$, the original lower objective satisfies $g^\star(x) = 0$, and its optimal response set consists of two equally valid branches:

$$\mathcal{S}(x) := \arg\min_{z,v} g(x, z, v) = \{(-1, x), (+1, x)\}. \tag{57}$$

Attraction-basin mechanism. For a frozen selector $\boldsymbol{\alpha} = (\alpha_1, 1 - \alpha_1)$, the branch derivative of the selected objective factorizes as

$$\partial_z \widetilde{g} = (z^2 - 1)\left[z - \frac{3\rho}{4}(2\alpha_1 - 1)\right], \qquad z_{\text{sep}} = \frac{3\rho}{4}(2\alpha_1 - 1). \tag{58}$$

When $|z_{\text{sep}}| < 1$, $z = \pm 1$ are stable equilibria and z_{sep} is the intervening unstable equilibrium of the frozen gradient flow. Pareto feedback moves this attraction boundary during response formation, allowing a near-boundary initialization to enter the preferred basin. The endpoints $z = \pm 1$ remain stationary for every selector: the mechanism concerns basin allocation before convergence, not escape from an already attained exact branch. For $\rho = 0.15$, the boundary satisfies $|z_{\text{sep}}| \le 0.1125$.

Weighted-sum reference solution. Restricting the preference-weighted upper objective $J_w = wf_1 + (1 - w)f_2$ to the two exact response branches gives

$$J_w(x, -1, x) = \big(x - (1 - 2w)\big)^2 + 4w(1 - w) + (1 - w), \tag{59}$$

$$J_w(x, +1, x) = \big(x - (1 - 2w)\big)^2 + 4w(1 - w) + w. \tag{60}$$

Both branches share the same optimal upper variable $x_w^\star = 1 - 2w$. Comparing their remaining branch-dependent terms yields $z_w^\star = -1$ for $w > 1/2$ and $z_w^\star = +1$ for $w < 1/2$, with

$$J_w^\star = 4w(1 - w) + \min\{w, 1 - w\}. \tag{61}$$

At $w = 1/2$, the two branches tie; we therefore use off-center preferences to obtain an identifiable selection target. This weighted-sum oracle supplies a common evaluation reference. The weighted-Chebyshev coordinator in Eq. (5) is not an optimizer of J_w; its upper iterates need not converge to $x_w^\star$.

Experimental protocol. We consider $w \in \{0.1, 0.3, 0.7, 0.9\}$. For each preference, the 20 paired initializations have balanced response signs, with ten starts on either side of the branch boundary and

$$x_0 = v_0 = 0, \qquad |z_0| \in [0.01, 0.08]. \tag{62}$$

The same initial state is paired across all methods.

Each run uses 80 outer iterations and five lower-level gradient steps per outer iteration. The lower and upper learning rates are 0.1 and 0.04, respectively, and implicit-gradient methods use five

conjugate-gradient iterations with damping 10^{-3}. PAOS uses $\rho = 0.15$, feedback rate $\beta = 1$, and $u = 10$ in the weighted-Chebyshev subproblem. The balanced study uses seeds 20–39. WC-MHGD serves as the matched open-loop counterpart of PAOS. WS-HG and TCH-HG are weighted-sum and Chebyshev hypergradient controls that receive the same preference but do not return the Pareto decision to lower-level response formation.

Evaluation metrics. A terminal response is counted as preference-consistent when

$$\begin{aligned} \operatorname{sign}(z_T) &= -\operatorname{sign}\left(w - \frac{1}{2}\right), \\ |z_T| \geq 0.9, &\qquad g(x_T, z_T, v_T) \leq 10^{-3}. \end{aligned} \tag{63}$$

For upper-level quality, we project the current response to its nearest exact lower branch:

$$\bar{z}_t \in \operatorname*{arg\,min}_{s \in \{-1,+1\}} |z_t - s|, \qquad \bar{v}_t = x_t. \tag{64}$$

The projected weighted-oracle regret is then

$$\mathcal{R}_w^{(t)} = J_w(x_t, \bar{z}_t, \bar{v}_t) - J_w^\star. \tag{65}$$

This projection defines the evaluation metric; optimization proceeds on the original unprojected trajectory.

Preference-wise branch selection. Table 6 reports the preference-consistent branch count for every tested preference. PAOS selects the requested response in all paired runs, whereas the three open-loop controls preserve the balanced initialization split.

Table 6: Preference-consistent response counts over 20 paired initializations per preference. Bold denotes the best result.

Method	$w = 0.1$	$w = 0.3$	$w = 0.7$	$w = 0.9$
WS-HG	$10/20$	$10/20$	$10/20$	$10/20$
TCH-HG	$10/20$	$10/20$	$10/20$	$10/20$
WC-MHGD	$10/20$	$10/20$	$10/20$	$10/20$
PAOS	$\mathbf{20/20}$	$\mathbf{20/20}$	$\mathbf{20/20}$	$\mathbf{20/20}$

At the default selection strength $\rho = 0.15$, PAOS selects the preference-compatible branch in all $80/80$ runs across the four preferences, with median terminal lower-level gap 2.32×10^{-6}. In the separate selection-strength study (seeds 0–19, 20 near-boundary starts per preference, otherwise the same 80-step configuration), $\rho = 0.15$ likewise yields $80/80$ valid preference-compatible responses with median lower-level gap 2.4×10^{-6}. These results jointly demonstrate branch selection and approximate validity under the original lower objective.

For the representative $w = 0.9$ setting, the median projected oracle regrets over all 20 balanced starts are 0.147, 0.492, 0.400, and 0.463 for PAOS, WC-MHGD, WS-HG, and TCH-HG, respectively. These values compare attained responses against the weighted-sum reference; branch counts directly measure the selection mechanism.

Feedback-strength ablation. Table 7 isolates dynamic feedback by varying only β. Static selection often attains the preferred sign but no valid preferred response, whereas dynamic feedback raises valid branch selection from $0/80$ to $80/80$ and reduces median regret from 0.243 to 0.134.

Table 7: Double-well feedback-rate ablation over 80 paired near-boundary runs; valid branch additionally enforces Eq. (63).

Feedback rate	Preferred sign ↑	Valid branch ↑	Median regret ↓
$\beta = 0$ (static)	$67/80$	$0/80$	0.243
$\beta = 0.25$	$76/80$	$76/80$	0.148
$\beta = 0.50$	$78/80$	$78/80$	0.147
$\beta = 0.70$	$79/80$	$79/80$	0.147
$\beta = 1.00$ (PAOS)	$\mathbf{80/80}$	$\mathbf{80/80}$	$\mathbf{0.134}$

Initialization-distance sweep. An additional paired sweep at $w = 0.9$ varies positive initial z_0 while keeping $x_0 = v_0 = 0$ and the 80-step configuration above. Twenty starts are sampled uniformly in each interval using random seed 2027. Table 8 reports all intervals from the existing sweep. The transition across intervals is consistent with the bounded displacement of the attraction boundary in Eq. (58).

Table 8: Valid preferred-branch counts versus initial distance at $w = 0.9$ (20 paired starts per interval, $\rho = 0.15$).

Initial z_0 interval	PAOS	WC-MHGD
$[0.002, 0.015]$	$20/20$	$0/20$
$[0.015, 0.030]$	$20/20$	$0/20$
$[0.030, 0.045]$	$20/20$	$0/20$
$[0.045, 0.060]$	$20/20$	$0/20$
$[0.060, 0.075]$	$20/20$	$0/20$
$[0.075, 0.090]$	$20/20$	$0/20$
$[0.090, 0.120]$	$9/20$	$0/20$
$[0.120, 0.180]$	$0/20$	$0/20$
$[0.180, 0.350]$	$0/20$	$0/20$
$[0.350, 0.600]$	$0/20$	$0/20$
$[0.600, 0.850]$	$0/20$	$0/20$

Separate aligned-objective control. Figure 3(c) and Figure 1(b) use a distinct control problem that retains g and b but aligns the continuous upper-objective terms:

$$f_1^{\text{align}} = v^2 + b(z), \qquad f_2^{\text{align}} = v^2 + 1 - b(z). \tag{66}$$

Its exact preferred response has $x_w^\star = v_w^\star = 0$, branch $z_w^\star$ as above, and weighted-sum optimum $\min\{w, 1-w\}$. Projected regret is $x_t^2 + |2w-1|\,\mathbf{1}\{\text{sign}(z_t) \neq z_w^\star\}$. We use $w \in \{0.1, 0.9\}$, seeds 0–11 for each preference, $x_0 = v_0 = 1.5$, and balanced signs with $|z_0| \in [0.01, 0.08]$. The displayed hard-start subset contains the six initially incompatible signs per preference, pooled into 12 paired starts. Runs last 600 outer iterations, with five lower steps, lower/upper learning rates $0.1/0.04$, five CG steps, damping 10^{-3}, $u = 10$, and $\beta = 1$. PAOS anneals $\rho_t = 0.2[1 - (t+1)/600]$ for $t = 0, \ldots, 599$. The final median projected regret is 3.16×10^{-4} for PAOS and 0.8 for each open-loop control. This experiment evaluates long-horizon refinement after branch selection on the aligned control, separately from the conflicting continuous objectives in the 80-step basin study.

C Additional Algorithmic and Complexity Details

C.1 Selector Initialization and Selection Schedule

We initialize the response-selector state with the user preference, $\boldsymbol{\alpha}_0 = \mathbf{r}$. The response-selection coefficient is increased smoothly during an initial warm-up period:

$$\rho_t = \rho_0 + (\rho_{\max} - \rho_0) \min\left\{1, \frac{t+1}{W}\right\}^{p_\rho}, \tag{67}$$

where W is the warm-up length and p_ρ controls the schedule shape. This schedule progressively activates preference-guided response shaping while leaving the subsequent PAOS recursion unchanged. The 80-step double-well study instead uses constant ρ, and the separate aligned-objective control uses the annealing schedule in Appendix B.

C.2 Warm Starts and Finite-Step Solvers

PAOS warm-starts the lower response from the previous outer iterate and stores one CG state per upper objective. The lower optimization steps are detached between iterations in the implementation; hence PAOS uses the implicit hypergradient in Eq. (4) rather than differentiating through the complete T-step trajectory. This keeps memory usage independent of the unrolling depth T, up to the model state and the S CG vectors.

C.3 Oracle Accounting

Let C_F denote the cost of evaluating the selector-weighted derivatives of the upper objectives, C_g the cost of a lower-level gradient, and C_{HVP} the cost of one Hessian-vector product. One deterministic outer iteration uses approximately

$$T(C_g + C_F) + SKC_{\mathrm{HVP}} + C_{\mathrm{QP}}(S). \tag{68}$$

The Pareto subproblem depends on the number of objectives, while the response feedback itself requires $\mathcal{O}(S)$ arithmetic and memory.

C.4 Ablation Configurations

The following configurations isolate the components of PAOS:

Configuration	Setting	Interpretation
Open-loop baseline	$\rho_t = 0$	Original lower response
Uniform selection	$\boldsymbol{\alpha}_t = \mathbf{1}/S$	Preference-agnostic response selection
Static preference	$\beta = 0,\ \boldsymbol{\alpha}_0 = \mathbf{r}$	No dynamic Pareto feedback
Full PAOS	$\beta > 0$, dynamic $\boldsymbol{\alpha}_t$	Preference-aware Pareto feedback

The main ablation compares these cases under matched lower-step and hypergradient budgets, isolating response selection, preference conditioning, and dynamic Pareto feedback.

D Experimental Details

This appendix specifies the task-level bilevel formulations, datasets, models, optimization settings, and evaluation metrics used in Section 4.

D.1 Preferences and Implementation

Baseline implementation. We implement MOML, FORUM, gMOBA, MGDA-HG, WS-HG, TCH-HG, WC-MHGD, and PAOS using their published update rules with separately selected task-specific learning rates; complete configurations are provided in the accompanying code. The gMOBA row uses its single-lower-objective specialization.

Computing environment. Learning experiments run on Linux with two NVIDIA RTX A6000 GPUs, Python 3.10.20, PyTorch 2.5.1, CUDA 12.1, and cuDNN 9.1.

D.2 MNIST Data Hyper-Cleaning

Task construction. We partition MNIST into 5,000 lower-level training images, 5,000 clean validation images, and 5,000 test images. Each image is flattened into a 784-dimensional vector. We construct five training environments with label corruption rates

$$c_s \in \{0, 0.1, 0.2, 0.3, 0.4\}, \qquad s \in \{1, \dots, 5\}. \tag{69}$$

Let $\mathcal{D}_{\mathrm{tr}}^{(s)} = \{(a_i, \widetilde{b}_i^{(s)})\}_{i=1}^n$ denote the training set whose labels are corrupted at rate c_s.

The upper variable $x \in \mathbb{R}^n$ contains one cleaning logit for each training example, and $\sigma(x_i)$ determines its contribution to lower-level training. The lower response $y = (y_1, \dots, y_5)$ consists of five linear softmax classifiers, one for each corruption environment. The original lower-level objective is

$$g(x, y) = \frac{1}{5}\sum_{s=1}^{5}\frac{1}{n}\sum_{i=1}^{n}\sigma(x_i)\ell_{\mathrm{ce}}\big(h_{y_s}(a_i), \widetilde{b}_i^{(s)}\big) + \frac{\lambda_y}{10}\sum_{s=1}^{5}\|y_s\|_2^2 + \frac{\lambda_x}{2}\|x\|_2^2.$$

Each upper-level objective evaluates one corruption-specific classifier on the shared clean validation set:

$$f_s(x, y) = \frac{1}{|\mathcal{D}_{\mathrm{val}}|}\sum_{(a,b)\in\mathcal{D}_{\mathrm{val}}}\ell_{\mathrm{ce}}\big(h_{y_s}(a), b\big),\ s \in \{1, \dots, 5\}. \tag{70}$$

The resulting MOBL problem learns a shared cleaning policy while coordinating the validation objectives induced by five differently corrupted lower-level responses.

For fixed sample weights, positive quadratic regularization makes the original linear-classifier lower problem strongly convex. PAOS therefore shapes finite-budget responses through the selected training objective here; the exact original lower response is unique.

Preference-conditioned evaluation. For target corruption environment s, we use

$$r_j^{(s)} = \begin{cases} 0.8, & j = s, \\ 0.05, & j \neq s, \end{cases} \qquad \sum_{j=1}^{5} r_j^{(s)} = 1. \tag{71}$$

For a preference-conditioned method, the target-aligned accuracy at corruption level c_s is the test accuracy of classifier h_{y_s} from the run using $r^{(s)}$. For a preference-independent method, we evaluate the corresponding classifier from its shared multi-objective run. The average reported in Table 1 is

$$\mathrm{Acc}_{\mathrm{avg}} = \frac{1}{5} \sum_{s=1}^{5} \mathrm{Acc}_{\mathrm{test}}\big(h_{y_s}^{(r^{(s)})}\big). \tag{72}$$

Optimization details. PAOS uses 75 outer iterations and 200 lower-level gradient steps per outer iteration, with lower and upper learning rates of 0.05 and 0.1, respectively, ten conjugate-gradient iterations, and damping 10^{-4}. Classifier weight decay is 10^{-3} and upper-logit regularization is 10^{-5}. For PAOS, the response-selection strength is linearly increased from zero to 0.005 during the first 25 outer iterations, and the selection-state momentum is 0.5. Results are reported at $T = 75$.

D.3 Few-shot Bilevel Learning on FC-100

Dataset and task construction. FC-100 (Oreshkin et al., 2018) partitions the CIFAR-100 classes into disjoint training, validation, and test subsets. We construct a 5-way 5-shot support–query task from the training classes. For every selected class, the support set contains five labeled images and the query set contains 15 images. All images are resized to 32×32 pixels and processed by a shared augmentation and normalization pipeline. Each optimization run performs 500 outer iterations on one sampled support–query task. The same query set supplies the upper-level optimization losses and reported accuracies; this protocol evaluates within-task optimization, not generalization to unseen examples or tasks.

Bilevel formulation. Let ϕ_x denote a shared feature encoder parameterized by the upper-level variable x, and let h_y denote a task-specific linear classifier parameterized by the lower-level variable y. Given a task with support set $\mathcal{D}_{\mathrm{sup}}$, the lower-level classifier is adapted by minimizing

$$g(x, y) = \frac{1}{|\mathcal{D}_{\mathrm{sup}}|} \sum_{(a,b) \in \mathcal{D}_{\mathrm{sup}}} \ell_{\mathrm{ce}}(h_y(\phi_x(a)), b) + \frac{\lambda_y}{2} \|y\|_2^2, \tag{73}$$

where ℓ_{ce} is the cross-entropy loss and λ_y controls classifier regularization. With a fixed encoder and positive quadratic regularization on the classifier parameters, this lower problem has a unique minimizer. PAOS therefore performs finite-budget response shaping here, rather than selection among multiple exact lower minimizers. The upper level contains one query objective for each of the five task classes:

$$f_c(x, y) = \frac{1}{|\mathcal{D}_{\mathrm{qry}}^c|} \sum_{(a,b) \in \mathcal{D}_{\mathrm{qry}}^c} \ell_{\mathrm{ce}}(h_y(\phi_x(a)), b), \qquad c \in \{1, \ldots, 5\}. \tag{74}$$

where $\mathcal{D}_{\mathrm{qry}}^c$ is the class-c subset of the query set. These labeled query objectives enter both upper-level optimization and PAOS's selected lower objective. This construction yields a five-objective bilevel problem in which improving one class can alter the representation and adaptation trajectory shared by the others.

For each task, we construct five preference-conditioned runs by selecting each class once as the preferred objective. The corresponding preference vector assigns weight 0.8 to the selected class

and distributes the remaining mass uniformly across the other four objectives:

$$r_j^{(c)} = \begin{cases} 0.8, & j = c, \\ 0.05, & j \neq c. \end{cases} \tag{75}$$

PAOS uses the Pareto mixture induced by this preference to update its stateful selection variable and construct the lower-level objective used in the subsequent adaptation stage.

Model and optimization settings. The feature extractor is a four-layer convolutional network with 32 channels per layer, followed by a task-specific linear classifier. In the PAOS configuration, each outer iteration performs 32 lower-level adaptation steps with learning rate 0.2, followed by an upper-level update with learning rate 0.01. Hypergradient computation uses five conjugate-gradient iterations with damping coefficient 0.01. We apply gradient-clipping thresholds of 10 and 2 to the lower- and upper-level updates, respectively, and set the classifier weight decay to 10^{-3}.

For PAOS, the selection strength is linearly warmed from zero to 2×10^{-3} during the first 100 outer iterations. The response-selection state is updated with momentum coefficient 0.5.

Evaluation metrics. We report three accuracy summaries on the optimization query set. Final accuracy is measured after the last outer iteration. Last-100 accuracy averages the final 100 iterations and captures late-stage optimization quality. Accuracy AUC is the normalized discrete average over all 500 outer iterations:

$$\mathrm{AUC}_{\mathrm{acc}} = \frac{1}{T} \sum_{t=1}^{T} \mathrm{Acc}_{\mathrm{qry}}^{(t)}, \qquad T = 500. \tag{76}$$

Table 3 reports the mean and standard deviation across five independent seeds.

Additional preference-conditioned trade-off projections. The four projections in Figure 5 jointly involve all five class-wise query losses. Figure 7 supplies the six remaining three-objective views, completing the ten possible visualizations of the optimized query-loss geometry. Within every displayed projection, the five PAOS preference solutions dominate all five preference means of both WC-MHGD and TCH-HG.

D.4 Physics-Informed Learning on the Heat Equation

Problem construction. We consider the one-dimensional heat equation

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial \xi^2}, \qquad (\xi, t) \in [-1, 1] \times [0, 1], \qquad \nu = 0.1, \tag{77}$$

with initial and boundary conditions

$$u(\xi, 0) = \sin(\pi \xi), \qquad u(-1, t) = u(1, t) = 0. \tag{78}$$

Its analytical solution is

$$u^\star(\xi, t) = \exp(-\pi^2 \nu t) \sin(\pi \xi). \tag{79}$$

The lower-level variable y parameterizes a three-layer tanh multilayer perceptron $u_y(\xi, t)$ with hidden width 32. Its physics residual is

$$\mathcal{R}_y(\xi, t) = \frac{\partial u_y(\xi, t)}{\partial t} - \nu \frac{\partial^2 u_y(\xi, t)}{\partial \xi^2}. \tag{80}$$

We construct four groups of pointwise training losses:

$$\ell_{\mathrm{res},i}(y) = \left| \mathcal{R}_y(\xi_i, t_i) \right|^2, \tag{81}$$

$$\ell_{\mathrm{IC},i}(y) = \left| u_y(\xi_i, 0) - \sin(\pi \xi_i) \right|^2, \tag{82}$$

$$\ell_{\mathrm{BC},i}(y) = \left| u_y(\xi_i^{\mathrm{bd}}, t_i) \right|^2, \qquad \xi_i^{\mathrm{bd}} \in \{-1, 1\}, \tag{83}$$

$$\ell_{\mathrm{obs},i}(y) = \left| u_y(\xi_i, t_i) - \widetilde{u}_i \right|^2, \tag{84}$$

where

$$\widetilde{u}_i = u^\star(\xi_i, t_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, 0.02^2), \tag{85}$$

denotes a noisy observation of the analytical solution.

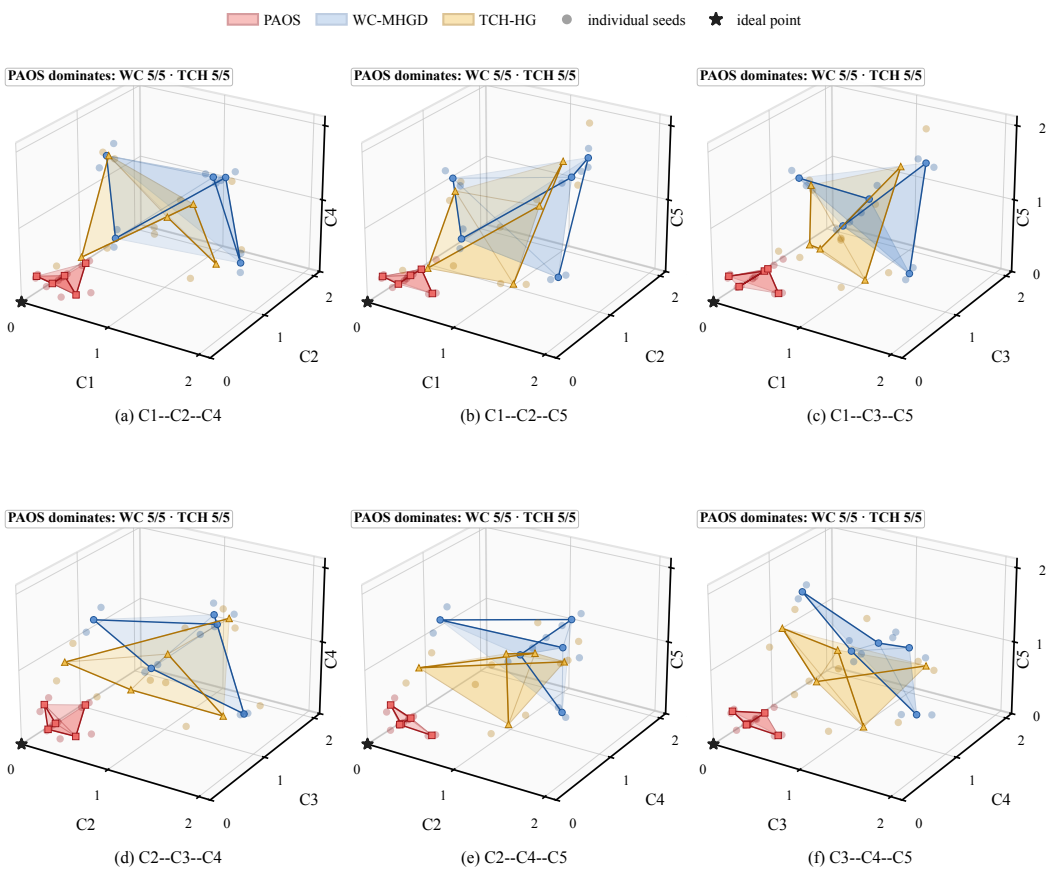


Figure 7: **Additional FC-100 preference-conditioned trade-off projections.** These six views complete all ten three-objective projections. In every view, PAOS dominates all five preference means of both WC-MHGD and TCH-HG (lower is better); translucent markers show individual seeds.

Point-reweighting bilevel formulation. The upper-level variable $x = \{x_{k,i} : k = 1, \dots, 4,\ i = 1, \dots, n_k\}$ contains one logit for every training point. For constraint group k, we convert these logits into positive mean-one weights:

$$w_{k,i}(x) = n_k \frac{\exp(x_{k,i})}{\sum_{j=1}^{n_k} \exp(x_{k,j})}, \qquad \frac{1}{n_k} \sum_{i=1}^{n_k} w_{k,i}(x) = 1. \tag{86}$$

The resulting lower-level objective is

$$g(x, y) = \sum_{k=1}^{4} \frac{1}{n_k} \sum_{i=1}^{n_k} w_{k,i}(x) \ell_{k,i}(y) + \frac{\lambda_y}{2} \|y\|_2^2 + \frac{\lambda_x}{2} \|x\|_2^2. \tag{87}$$

The four upper-level objectives evaluate the corresponding constraints on independently sampled, noise-free validation points:

$$\mathbf{F}(x, y) = \begin{bmatrix} F_{\text{res}}(x, y) \\ F_{\text{IC}}(x, y) \\ F_{\text{BC}}(x, y) \\ F_{\text{obs}}(x, y) \end{bmatrix} = \begin{bmatrix} \text{MSE}_{\text{res}}^{\text{val}} \\ \text{MSE}_{\text{IC}}^{\text{val}} \\ \text{MSE}_{\text{BC}}^{\text{val}} \\ \text{MSE}_{\text{obs}}^{\text{val}} \end{bmatrix}. \tag{88}$$

Thus, the upper level learns how the training points should shape the lower-level response, while its four objectives coordinate the physical, initial, boundary, and observational properties of the resulting solution.

Table 9: PINN feedback-rate ablation over four preference directions and two seeds. Metrics are macro-averaged over preferences within each seed; only β varies.

Feedback rate	Science score ↑	Relative L_2 ↓	Test MSE $\times 10^{-4}$ ↓	Selected val. $\times 10^{-4}$ ↓
$\beta = 0$ (static)	0.96480 ± 0.00105	0.03674 ± 0.00142	3.52 ± 0.91	8.46 ± 2.74
$\beta = 0.25$ (PAOS)	$\mathbf{0.97274 \pm 0.00009}$	$\mathbf{0.02805 \pm 0.00012}$	$\mathbf{1.78 \pm 0.08}$	$\mathbf{6.52 \pm 0.98}$
$\beta = 0.50$	0.96570 ± 0.00266	0.03558 ± 0.00290	2.90 ± 0.55	11.01 ± 0.53
$\beta = 0.70$	0.97087 ± 0.00456	0.03005 ± 0.00488	2.07 ± 0.72	7.76 ± 1.66
$\beta = 1.00$	0.94370 ± 0.03954	0.06395 ± 0.04912	20.10 ± 25.66	30.38 ± 32.76

Data and preference protocol. The training set contains 256 PDE-residual points, 96 initial-condition points, 96 boundary-condition points, and 96 observation points, yielding 544 point-weighting variables in total. The validation sets contain independently sampled points for the same four constraint groups and use noise-free targets. Testing is conducted on a dense 48×96 space–time grid.

We evaluate four preference directions, with each upper objective selected once as the preferred objective. For preference $p \in \{1, \dots, 4\}$, the simplex-normalized preference is

$$r_k^{(p)} = \begin{cases} 16/19, & k = p, \\ 1/19, & k \neq p. \end{cases} \tag{89}$$

This corresponds to assigning unnormalized weights 0.8 and 0.05 to the preferred and remaining objectives, respectively. Results are reported over two independent seeds per preference direction.

Optimization settings. PAOS uses 80 outer iterations and 60 lower-level steps per outer iteration. The lower- and upper-level learning rates are 0.01 and 0.02; hypergradient estimation uses eight conjugate-gradient steps with damping coefficient 10^{-3}. The lower gradient and upper update direction are clipped at norms 10 and 3, respectively.

For PAOS, the selection strength is linearly warmed from zero to 0.005 during the first 50 outer iterations, and the feedback rate is set to $\beta = 0.25$. We select β from $\{0, 0.25, 0.5, 0.7, 1\}$ using the target-aligned validation objective and reserve test metrics for final evaluation.

Evaluation metrics. Let $\mathcal{G}_{\text{test}}$ denote the dense space–time test grid. The primary evaluation metric is the relative L_2 error

$$E_{\text{rel}} = \frac{\|u_y - u^\star\|_{2,\mathcal{G}_{\text{test}}}}{\|u^\star\|_{2,\mathcal{G}_{\text{test}}}}. \tag{90}$$

We additionally report the iteration-normalized relative-L_2 AUC,

$$E_{\text{AUC}} = \frac{1}{T} \sum_{s=1}^{T} E_{\text{rel}}^{(s)}, \qquad T = 80, \tag{91}$$

and the average relative L_2 error over the final 20 iterations.

Feedback-rate ablation. Table 9 varies only β. Relative to static selection, $\beta = 0.25$ reduces relative L_2 error by 23.6%, test MSE by 49.4%, and the selected validation objective by 22.9%, improving $6/8$ paired test runs and $7/8$ validation runs. The sweep favors moderate, stateful feedback over immediate target replacement. Science score is $1/(1 + E_{\text{rel}})$; selected validation denotes the preference-indexed upper objective.

D.5 Kvasir-SEG Uniform-Preference Loss Design

Kvasir-SEG contains 1,000 endoscopic images and corresponding polyp masks. We stratify by lesion area and use a fixed $700/150/150$ train/validation/test split with split seed 2027. Images and masks are resized to 96×96 using bilinear and nearest-neighbor interpolation, respectively. Training augmentation consists of random horizontal flips and lower-probability vertical flips.

Table 10: MNIST hyper-cleaning quality-to-target efficiency over 75 outer iterations (mean ± std. over three seeds, RTX A6000). "–" means target CE ≤ 0.8 was not reached. GPU memory is allocated memory in MiB.

	Solution quality			Reach CE ≤ 0.8		GPU memory	Oracle calls / outer	
Method	Target Acc. ↑	Final CE ↓	CE AUC ↓	Iter. ↓	Time (s) ↓	Alloc. ↓	LL grad.	HVP
MOML	80.62±0.23	0.8752±0.0078	1.2099±0.0060	–	–	3133.4±0.4	20	0
FORUM	72.89±0.08	1.3322±0.0011	1.7275±0.0015	–	–	188.0±0.4	64	0
gMOBA	83.39±0.08	0.7490±0.0086	0.9932±0.0072	53.7±2.3	207.5±9.8	162.3±0.5	200	50
MGDA-HG	83.24±0.08	0.7561±0.0087	0.9969±0.0072	55.3±2.9	213.7±10.8	162.1±0.4	200	50
WC-MHGD	83.61±0.12	0.6923±0.0094	0.7846±0.0089	11.7±0.6	**44.9±2.0**	162.3±0.4	200	50
PAOS	**85.17±0.11**	**0.6417±0.0089**	**0.7200±0.0085**	**11.0±0.0**	75.2±0.2	281.3±0.4	200	50

The upper variable $x \in \mathbb{R}^3$ contains training-loss logits, with $\omega(x) = 3\,\mathrm{softmax}(x)$. The lower response y comprises the 7,697 parameters of a compact U-Net with base width four. The lower objective is

$$g(x, y) = \omega_1(x)\mathcal{L}_{\mathrm{BCE}}(y) + \omega_2(x)\mathcal{L}_{\mathrm{region}}(y) + \omega_3(x)\mathcal{L}_{\mathrm{boundary}}(y) + \frac{\lambda_y}{2}\|y\|^2, \tag{92}$$

where the latter two terms are soft regional-Dice and boundary-Dice losses. On the validation split, we optimize

$$f_1 = 1 - \mathrm{Dice}_{\mathrm{region}}, \tag{93}$$

$$f_2 = 1 - \mathrm{Dice}_{\mathrm{boundary}}, \tag{94}$$

$$f_3 = 1 - \mathrm{Dice}_{\mathrm{small/hard}}\,. \tag{95}$$

The third objective is evaluated on small-lesion validation examples. All methods use the same uniform user preference.

PAOS uses 180 outer iterations, six lower steps, two CG steps, damping 0.05, and lower and upper learning rates of 0.03. The selection strength is warmed to 0.1 over 20 iterations with selector momentum $\beta = 0.7$. Training and validation batch sizes are six, and the test batch size is 16.

Test metrics are Dice, IoU, Boundary F1 with a two-pixel tolerance, Dice on the lowest lesion-area quartile, and the 95th-percentile Hausdorff distance (HD95). Dice, IoU, Boundary F1, and small-lesion Dice are higher-is-better; HD95 is lower-is-better and is measured in resized-image pixels.

E Additional Experimental Results

E.1 Quality-to-Target Efficiency

E.2 Additional Kvasir-SEG Results

The supplementary qualitative comparison presents representative test cases across ground-truth lesion-area quartiles. Each row shows the full image, a shared region-of-interest crop, and per-case Dice, IoU, and HD95.

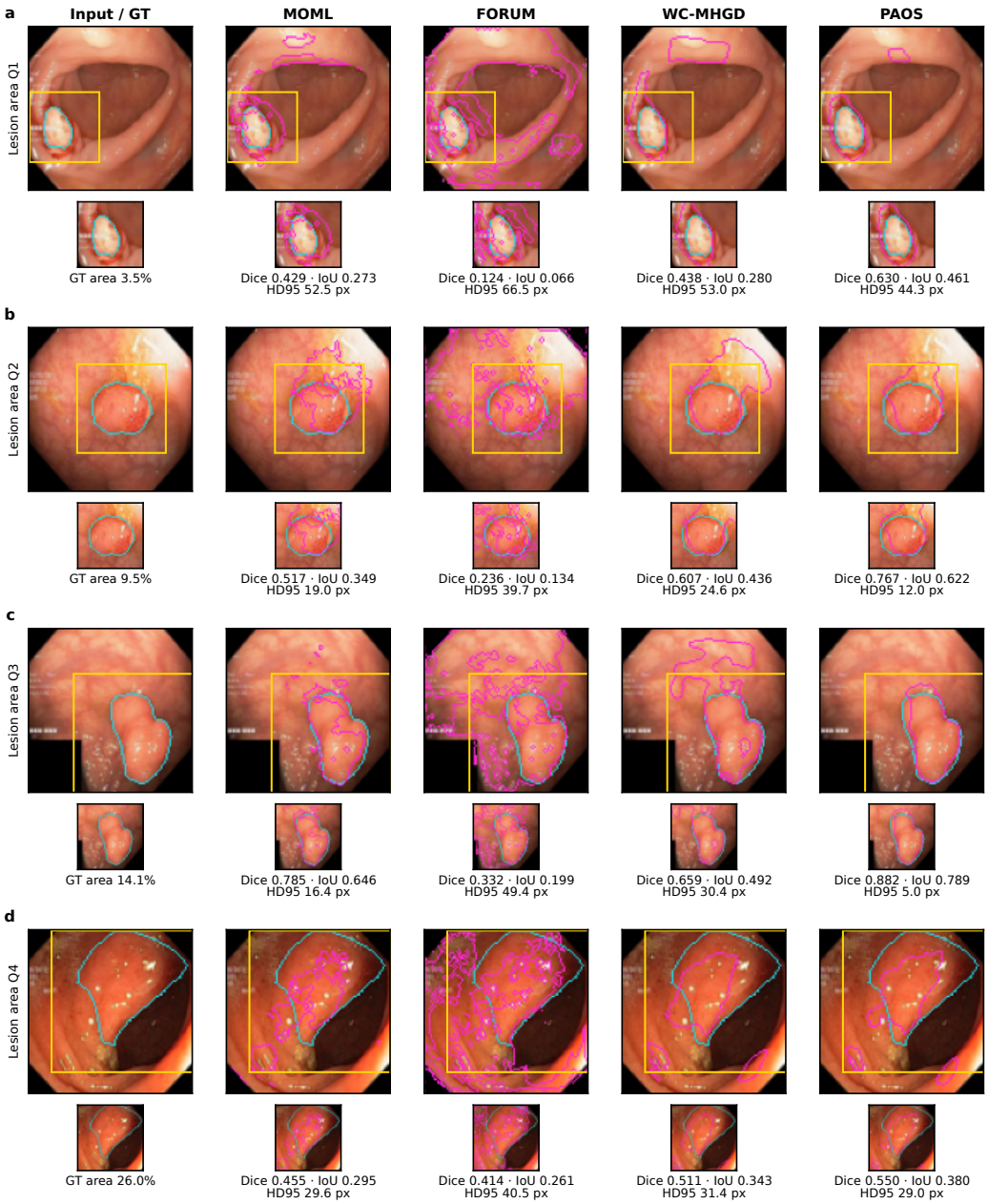


Figure 8: Kvasir-SEG qualitative comparisons across four lesion-area quartiles. All methods use the same image, region of interest, and display scale within each row.