A FULLY FIRST-ORDER SINGLE-LOOP ALGORITHM FOR STOCHASTIC MULTI-OBJECTIVE BILEVEL OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study stochastic multi-objective bilevel optimization, where each upper-level objective depends on the solution of an objective-specific lower-level problem. Existing methods either require computationally expensive Hessian information, rely on accurate inner-loop solutions, or lack rigorous non-asymptotic convergence guarantees. To this end, we propose `FOSMO`, a fully first-order single-loop algorithm for stochastic multi-objective bilevel optimization that avoids second-order computations, and establish a non-asymptotic convergence rate of $\widetilde{\mathcal{O}}(T^{-2/5})$, yielding an $\widetilde{\mathcal{O}}(\epsilon^{-5/2})$ oracle complexity for a fixed number of objectives. Furthermore, we adopt smooth Tchebycheff aggregation for preference-guided trade-offs among different objectives, enabling broader exploration of the Pareto front. Experiments on multi-task representation learning and hyperparameter optimization demonstrate the effectiveness of `FOSMO`.

1 INTRODUCTION

Bilevel optimization has been widely used to model hierarchical learning problems in which the upper-level objective depends on the optimal solution of a lower-level problem. Representative applications include hyperparameter optimization (Alcantara et al., 2025; MacKay et al., 2019; Salehi et al., 2025; Li et al., 2021b), meta-learning (Qin et al., 2023; Rajeswaran et al., 2019), model selection (Bennett et al., 2006), and data reweighting (Pan et al., 2025). Multi-objective bilevel optimization extends this framework to settings with multiple conflicting upper-level objectives. Such problems naturally arise in many applications, for example, in large language model alignment, where different objectives such as safety and helpfulness need to be balanced simultaneously (Shi et al., 2024; Yang et al., 2024a; Wang et al., 2024; Lin et al., 2025; Guo et al., 2024). The general formulation of multi-objective bilevel optimization is given as follows:

$$\min_{x\in\mathcal{X}} \boldsymbol{\phi}(x) := (\phi_1(x), \cdots, \phi_m(x)) \quad \text{s.t. } y_i^* \in \arg\min_{y_i\in\mathbb{R}^{d_y}} g_i(x, y_i), \quad i \in [m], \tag{P}$$

where the upper-level objective functions $\phi_i(x) = f_i(x, y_i^*(x))$ are implicitly determined by the optimal solution mapping $y_i^*(x)$ of the lower-level problems. Throughout this paper, we assume that $\mathcal{X} = \mathbb{R}^{d_x}$ and g_i is strongly convex in y, so that $y^*(x)$ is uniquely well-defined for all $x \in \mathcal{X}$. For bilevel optimization, iterative methods have been widely adopted due to their simplicity and scalability. Initiated by Ghadimi & Wang (2018), a growing body of literature has focused on efficient iterative algorithms for bilevel problems, with particular emphasis on their non-asymptotic convergence guarantees. For problem (**P**), the hypergradient of the upper-level function $\phi_i(x)$ is given by:

$$\nabla\phi_i(x) = \nabla_x f_i(x, y_i^*(x)) - \nabla^2_{xy_i} g_i(x, y_i^*(x)) \times \left[\nabla^2_{y_i y_i} g_i(x, y_i^*(x))\right]^{-1} \nabla_{y_i} f_i(x, y_i^*(x)). \tag{1}$$

There are two major challenges in computing 1. **a) Inner-loop optimization.** Computing $y_i^*(x)$ requires multiple iterations to approximately solve the lower-level problem at each outer iteration. To alleviate this issue, existing works (Kwon et al., 2023; Liu et al., 2024; Cao et al., 2025; Gu et al., 2026; Shen et al., 2025; Gu et al., 2023; Li et al., 2021a) adopt a two-timescale update scheme, allowing the lower-level iterates to effectively track the lower-level optimum with only a few inner-loop updates. Inspired by this, we further show that a single-loop style can still be achieved using

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.

only first-order information. **b) Dependence on second-order information.** Equation1 involves second-order derivatives of the lower-level objective g_i. Computing such Hessian-related information incurs additional computational and memory costs (Gu et al., 2023; Chen et al., 2021b;a; Yang et al., 2021; Zhang et al., 2025). Moreover, in the stochastic setting, the estimation of second-order information is more sensitive to the batch size than first-order, leading to higher estimation variance and noisier hypergradient estimates.

Some existing works (Giovannelli et al., 2023; Sow et al., 2022; Liu et al., 2021; 2022; Kwon et al., 2023; Yang et al., 2024b) avoid computing second-order information and only utilize first-order information of the upper-level and lower-level functions. However, these methods either lack a non-asymptotic convergence analysis (Giovannelli et al., 2023; Liu et al., 2021) or do not consider the stochastic setting (Sow et al., 2022; Liu et al., 2022; Kwon et al., 2023). Moreover, first-order methods for multi-objective bilevel optimization remain largely underexplored. The most closely related work is Ye et al. (2024), which proposes a first-order method for multi-objective bilevel optimization. However, it neither considers the stochastic setting nor adopts a single-loop scheme.

Our goal in this work is to develop a fully first-order single-loop algorithm for stochastic multi-objective bilevel optimization. We employ a value-function reformulation to avoid Hessian computations and thereby enable a fully first-order optimization framework. A smooth Tchebycheff aggregation is further adopted to achieve preference-guided optimization, while a recursive momentum scheme is employed to reduce estimation errors and improve the stability of stochastic updates.

The main contributions of this work are as follows.

- We propose `FOSMO`, a novel fully first-order single-loop algorithm for stochastic multi-objective bilevel optimization. `FOSMO` requires no inner-loop iterations and avoids Hessian computations, while also supporting preference-guided pareto exploration via user-specified preference vectors.
- We establish non-asymptotic convergence guarantees for `FOSMO`. Using recursive variance reduction and smooth Tchebycheff aggregation, we derive an $\widetilde{\mathcal{O}}(T^{-2/5})$ convergence rate to Pareto-stationary points, with a first-order oracle complexity of $\widetilde{\mathcal{O}}(\epsilon^{-5/2}m^6)$.
- We conduct experiments on two representative applications, including multi-task representation learning and hyperparameter optimization. The experimental results demonstrate the superior robustness of `FOSMO` and its effectiveness in preference-guided multi-objective optimization.

2 Related Work

Single-Objective Bilevel Optimization. Bilevel optimization (BLO) models hierarchical optimization problems in which the upper-level objective depends on the optimal solution of a lower-level problem. When the lower-level problem is strongly convex, Ghadimi & Wang (2018) established non-asymptotic convergence using implicit differentiation, but such methods typically require Hessian inverse. To avoid Hessian computation, Kwon et al. (2023) proposed a fully first-order stochastic approximation method with convergence guarantees. For nonconvex lower-level problems, sequential averaging methods (Liu et al., 2021) and value-function-based approaches (Gao et al., 2022) have been developed, yet they often lack finite-time guarantees. Recent efforts focus on single-loop first-order algorithms: Shen et al. (2025) proposed a single-loop method for linearly constrained BLO with an $\mathcal{O}(\epsilon^{-3}\log(\epsilon^{-1}))$ rate, and Cao et al. (2025) developed single-loop algorithms for deterministic and stochastic pessimistic BLO. In the stochastic setting, Phan & Wang (2025) provided finite-time convergence for linearly constrained stochastic BLO using only first-order information.

Multi-Objective Bilevel Optimization. MORBiT (Gu et al., 2023) formulates multi-objective bilevel optimization as a min–max problem and uses implicit differentiation under strongly convex lower-level problems, establishing a non-asymptotic convergence rate. MOML (Ye et al., 2021) formulates multi-objective meta learning as a bilevel optimization problem and solves it via gradient-based multi-objective optimization. Classical multicriteria descent characterizes Pareto stationarity via convex combinations of objective gradients (Fliege & Svaiter, 2000). WC-MHGD (Zhang et al., 2025) applies smooth weighted Tchebycheff scalarization directly to objective-wise hypergradients, with finite-time Pareto-stationarity guarantees in both deterministic and stochastic settings. More recently, Ye et al. (2024) proposed a first-order multi-gradient aggregation method based on the

value-function approach, and Jiang & Huang (2026b) developed Hessian-free methods for MOBLO with nonconvex lower level.

First-Order Bilevel Optimization. Kwon et al. (2023) proposed F2SA, a value-function-based penalty method with non-asymptotic convergence guarantees using only first-order oracles. Building on this framework, Chen et al. (2025) derived faster convergence rates and adopted a two-timescale update scheme with improved first-order oracle complexity. Liu et al. (2024) developed a single-loop Hessian-free algorithm via Moreau envelope for nonconvex bilevel problems. Shen et al. (2025) proposed SFLCB for linearly constrained BLO, eliminating Hessian via penalty and augmented Lagrangian with $\mathcal{O}(\epsilon^{-3})$ complexity. For stochastic settings, Phan & Wang (2025) provided first finite-time guarantees for linearly constrained stochastic BLO using first-order methods, and Zhang et al. (2026a) developed a decentralized stochastic first-order algorithm for nonconvex lower-level problems satisfying the Polyak–Łojasiewicz (PL) condition, without requiring Hessian or Jacobian information. More recently, Gu et al. (2026) proposed SGHA for nonconvex-strongly-convex BLO, and Cao et al. (2025) developed two single-loop algorithms for deterministic and stochastic pessimistic BLO using only first-order information.

3 PRELIMINARIES

In MOBO problems, the upper-level objectives are generally conflicting, and hence there may not exist a single solution that simultaneously minimizes all objectives. Instead of seeking a common minimizer, one typically considers solutions that achieve a desirable trade-off among the objectives. In multi-objective optimization, Pareto optimality is commonly considered.

Definition 1 (Pareto optimality)**.** *We say that* x_1 dominates x_2*, denoted by* $x_1 \prec x_2$*, if* $\phi_i(x_1) \leq \phi_i(x_2)$ *for all* $i \in [m]$ *and* $\phi_j(x_1) < \phi_j(x_2)$ *for at least one* $j \in [m]$*. A point* x *is called* Pareto optimal *if there does not exist any* $x \in \mathcal{X}$ *such that* $x' \prec x$*.*

Similar to non-convex multi-objective optimization or BLO problems, finding a Pareto-optimal solution in non-convex MOBO is generally challenging. Therefore, it is often more practical to consider the following Pareto stationarity as a first-order notion of optimality for characterizing convergence. In the following analysis, we define the simplex as $\Delta_m := \left\{\omega \in \mathbb{R}_+^m : \sum_{i=1}^m \omega_i = 1\right\}$.

Definition 2 (Pareto stationarity)**.** *A point* x *is called* Pareto stationary *if there exists* $\boldsymbol{\omega} \in \Delta_m$ *such that* $\sum_{i=1}^m \omega_i \nabla \phi_i(x) = 0$*.*

To characterize the non-asymptotic convergence toward Pareto stationarity, we introduce the following notion of ϵ-Pareto stationarity.

Definition 3 (ϵ-Pareto stationarity)**.** *For any* $\epsilon > 0$*, a point* x *is called a* Pareto ϵ-stationary point *if there exists* $\boldsymbol{\omega} \in \Delta_m$ *such that* $\left\|\sum_{i=1}^m \omega_i \nabla \phi_i(x)\right\|^2 \leq \epsilon$*.*

To characterize the computational efficiency of an algorithm, we introduce the following notion of oracle complexity.

Definition 4 (Oracle complexity)**.** *For any* $\epsilon > 0$*, the oracle complexity is defined as the total number of oracle calls required by an algorithm to obtain an* ϵ*-Pareto stationary point.*

In multi-objective optimization, the relative importance of different objectives often depends on the specific application or preference. Therefore, beyond finding an arbitrary Pareto solution, it is also important to obtain solutions that reflect the preferences over the objectives. A preference vector provides a natural way to characterize such desired trade-offs and guide the optimization toward different regions of the Pareto front. Classical multi-objective optimization incorporates such preference information through reference-point and achievement-scalarization methods (Wierzbicki, 1980), while more recent gradient-based approaches use preference vectors to generate Pareto solutions with different trade-offs or to directly seek solutions consistent with user-specified preferences (Lin et al., 2019; Mahapatra & Rajan, 2020).

In the multi-objective bilevel setting, different formulations have been developed to handle trade-offs among upper-level objectives. MORBiT (Gu et al., 2023) adopts a min–max formulation that emphasizes the worst-performing objective and leads to a robustness-oriented solution. FORUM (Ye et al., 2024) instead constructs a common descent direction through multi-gradient aggregation and seeks Pareto-stationary solutions without explicitly incorporating a prescribed preference.

More recently, weighted Tchebycheff scalarization has been introduced into multi-objective bilevel learning to explicitly capture preference information and enable systematic exploration of different regions of the Pareto front (Zhang et al., 2025; 2026b; Jiang & Huang, 2026a). However, existing approaches either provide limited control over user-specified preferences or rely on hypergradient-based bilevel optimization procedures. Consequently, preference-guided stochastic multi-objective bilevel optimization with a fully first-order and single-loop scheme remains underexplored.

4 ALGORITHM AND ANALYSIS

In this section, we introduce a fully first-order single-loop algorithm FOSMO for the stochastic multi-objective bilevel problem **P**, and establish its rigorous convergence rate analysis and complexity guarantees. The proposed method employs a value-function-based reformulation to avoid second-order derivatives. For the theoretical analysis, we defer the complete proofs to Appendix B and C, and present the main theoretical results and key lemmas here.

4.1 ALGORITHM

We reformulate the bilevel problem equation **P** into a value-function-based constrained problem by introducing the optimal value function $\varphi_i(x)$ for each lower-level problem and enforcing $g_i(x, y_i) = \varphi_i(x)$. The reformulation is given by

$$\min_{x\in\mathcal{X}, y_i\in\mathbb{R}^{d_y}} \quad f_i(x, y_i) \quad \text{s.t.} \quad g_i(x, y_i) - \varphi_i(x) \le 0, \quad i = 1, \dots, m. \tag{2}$$

where $\varphi_i(x) = g_i(x, y_i^*(x)) = \min_{y_i} g_i(x, y_i)$ is the optimal value function of the i-th lower-level problem. The constraint $g_i(x, y_i) - \varphi_i(x) \le 0$ ensures that y_i reaches the lower-level optimal value, and since $g_i(x, y_i) \ge \varphi_i(x)$ always holds, it is equivalent to $g_i(x, y_i) = \varphi_i(x)$. This reformulation converts the bilevel problem into a single-level problem with value-function constraints. Based on the optimality of $\varphi_i(x)$, we can develop fully first-order algorithms. To handle the value-function constraint, we adopt a penalty-based approach and move the constraint into the objective as a penalty term, yielding the following penalized value function:

$$L_{i,\lambda}^*(x) = \min_{y_i}\{f_i(x, y_i) + \lambda\left(g_i(x, y_i) - \varphi_i(x)\right)\}, \tag{3}$$

where $\lambda > 0$ is the penalty parameter, and the penalty term ensures that the lower-level optimality condition is enforced. To derive the gradient of the penalized value function in equation 3 with respect to x, we define two optimal solutions: $y_{i,\lambda}^*(x)$ denotes the minimizer of the penalized objective over y_i, while $z_i^*(x)$ denotes the solution to the original lower-level problem. Specifically,

$$y_{i,\lambda}^*(x) = \arg\min_{y_i}\{f_i(x, y_i) + \lambda g_i(x, y_i)\}, \quad z_i^*(x) = \arg\min_{y_i} g_i(x, y_i) \tag{4}$$

Then we have $L_{i,\lambda}^*(x) = f_i(x, y_{i,\lambda}^*) + \lambda[g_i(x, y_{i,\lambda}^*) - g_i(x, z_i^*)]$. Due to the optimality of $y_{i,\lambda}^*(x)$ and $z_i^*(x)$, Danskin's theorem (Danskin, 1966) directly gives the gradient of $L_{i,\lambda}^*(x)$ with respect to x as

$$\nabla L_{i,\lambda}^*(x) = \nabla_x f_i(x, y_{i,\lambda}^*) + \lambda[\nabla_x g_i(x, y_{i,\lambda}^*) - \nabla_x g_i(x, z_i^*)] \tag{5}$$

Note that $\nabla L_{i,\lambda}^*(x)$ involves no second-order derivatives but only the first-order gradients of f_i and g_i. In addition, $\nabla L_{i,\lambda}^*(x)$ in equation 5 involves two optimal solutions, $y_{i,\lambda}^*(x)$ and $z_i^*(x)$, which are computationally intractable to compute during optimization. Following Kwon et al. (2023), we introduce auxiliary variables y_i and z_i to track $y_{i,\lambda}^*(x)$ and $z_i^*(x)$, respectively. Then we denote the approximation of $\nabla L_{i,\lambda_i}^*(x)$ as

$$q_{i,x} = \nabla_x L_{i,\lambda}(x) = \nabla_x f_i(x, y_i) + \lambda\left(\nabla_x g_i(x, y_i) - \nabla_x g_i(x, z_i)\right), \tag{6}$$

We denote the stochastic estimates of $\nabla_x L_{i,\lambda_i}(x)$ as $h_{i,x}$. Recall that variables y_i and z_i are intended to track the optimal solutions $y_{i,\lambda}^*(x)$ and $z_i^*(x)$ in equation 4. We define the following gradient blocks:

$$q_{i,y} = \nabla_{y_i} f_i(x, y_i) + \lambda\nabla_{y_i} g_i(x, y_i), \quad q_{i,z} = \nabla_{z_i} g_i(x, z_i), \quad i = 1, ..., m \tag{7}$$

We denote the stochastic estimates of $q_{i,y}$ and $q_{i,z}$ by $h_{i,y}$ and $h_{i,z}$, respectively. The stochastic estimators defined above can be significantly affected by stochastic noise, which can be further

exacerbated as the penalty parameter λ_t increases. Moreover, directly using the raw stochastic estimates may therefore lead to noisy gradient and function-value estimation. To mitigate this issue, we apply STORM-based recursive tracking to the stochastic estimators, which reduces estimation noise and improves the stability of the stochastic updates. The STORM-based recursive estimators are defined as follows:

$$\hat{h}_{i,z}^t = h_{i,z}^t(\zeta_t) + (1-\alpha_i^t)\left[\hat{h}_{i,z}^{t-1} - h_{i,z}^{t-1}(\zeta_t)\right], \tag{8}$$

where ζ_t denotes the fresh stochastic sample drawn at iteration t. The remaining estimators $\hat{h}_{i,y}^t$, $\hat{h}_{i,x}^t$, $\hat{F}_i^t$, $\hat{G}_{i,y}^t$, and $\hat{G}_{i,z}^t$ are defined analogously using the same recursive estimator, where $\hat{F}_i^t$, $\hat{G}_{i,y}^t$, and $\hat{G}_{i,z}^t$ denote the recursive stochastic estimators of the function values $f_i(x_t, y_i^t)$, $g_i(x_t, y_i^t)$, and $g_i(x_t, z_i^t)$, respectively. Since the upper-level variable x is common to all objectives, the objective-specific upper-level directions $\{\hat{h}_{i,x}^t\}_{i=1}^m$ are required to be aggregated to obtain a unified descent direction for the update of x. To aggregate the objective-specific upper-level directions, we adopt a smooth Tchebycheff scalarization. Let $\boldsymbol{w} = (w_1, \ldots, w_m) \in \Delta_m$ denote a preference vector, and let $\boldsymbol{r} = (r_1, \ldots, r_m)$ be a reference point. Then the aggregation direction is determined by $\hat{h}_x^t = \sum_{i=1}^m \hat{p}_i^t w_i \hat{h}_{i,x}^t$, where $\hat{p}_i^t = \frac{\exp\left(w_i(\hat{L}_{i,\lambda}^t(x) - r_i)/\tau\right)}{\sum_{j=1}^m \exp\left(w_j(\hat{L}_{j,\lambda}^t(x) - r_j)/\tau\right)}$ and $\hat{L}_{i,\lambda}^t(x) = \hat{F}_i^t + \lambda_t(\hat{G}_{i,y}^t - \hat{G}_{i,z}^t)$ With the above stochastic gradients estimates at hand, our FOSMO is summarized in Algorithm 1.

Algorithm 1 The FOSMO Method

1: **Input:** The numbers of iteration T, initialization values x_0, y_i^0, z_i^0, preference vector $\boldsymbol{w}$, reference point $\boldsymbol{r}$, temperature coefficient τ, step-sizes $\beta, \eta, \alpha, \lambda$, step-size ratio ξ.
2: **for** $t = 0, 1, \ldots, T-1$ **do**
3: Compute $\hat{h}_{i,x}^t, \hat{h}_{i,z}^t, \hat{h}_{i,y}^t$ and $\hat{p}_i^t$
4: Aggregate $\hat{h}_x^t = \sum_{i=1}^m \hat{p}_i^t w_i \hat{h}_{i,x}^t$
5: **for** $i \in [m]$ **do**
6: $z_i^{t+1} \leftarrow z^t - \beta_t \hat{h}_{i,z}^t$
7: $y_i^{t+1} \leftarrow y^t - \eta_t \hat{h}_{i,y}^t$
8: **end for**
9: $x_{t+1} = x_t - \xi\eta_t \hat{h}_x^t$
10: **end for**
11: **return** x, y_i, z_i.

At the t-th iteration, $\hat{p}_i^t$ denotes the adaptive weight determined by the current function values $\hat{L}_{i,\lambda}^t(x)$. Therefore, the aggregated direction $\hat{h}_x^t$ depends not only on the preference vector but also on the current objective values, making the aggregation weights dynamically adjusted throughout the optimization process. Our smooth Tchebycheff weights admit a closed-form expression and dynamically adjust the relative contributions of different objectives according to the current optimization status while preserving the user-specified preference. In contrast, Zhang et al. (2025) determines the adaptive aggregation weights by solving an additional dynamic weighting subproblem based on the objective-wise hypergradients at each outer iteration, which requires an extra optimization procedure to determine the aggregation weights. Meanwhile, our method provides a more direct and lightweight mechanism for preference-guided aggregation.

4.2 Main Results

We next present the theoretical analysis of FOSMO. The Hessian-free design of our method introduces additional challenges in the convergence analysis. To establish our theoretical results, we first make the following standard assumptions.

Assumption 1 (Upper-level functions)**.** *For each $i \in [m]$, f_i is continuously differentiable and jointly $l_{f,1}$-smooth in (x, y_i). Moreover, for any x, $f_i(x, \cdot)$ is $l_{f,0}$-Lipschitz, and $\|\nabla\phi_i(x)\| \le C_\phi$.*

Assumption 2 (Lower-level functions)**.** *For each $i \in [m]$, g_i is continuously differentiable and jointly $l_{g,1}$-smooth in (x, y_i), and is μ_g-strongly convex in y_i for any fixed x. Moreover, g_i is twice continuously differentiable, and its Hessian is jointly $l_{g,2}$-Lipschitz continuous.*

Assumption 3 (Stochastic estimators)**.** *For each $i \in [m]$, the stochastic estimates of the function values and gradients of f_i and g_i are unbiased, and their variances are uniformly bounded as*

$$\mathbb{E}\|\nabla f_i(x, y_i; \xi) - \nabla f_i(x, y_i)\|^2 \le \delta_f^2, \qquad \mathbb{E}\|\nabla g_i(x, y_i; \zeta) - \nabla g_i(x, y_i)\|^2 \le \delta_g^2,$$
$$\mathbb{E}|f_i(x, y_i; \xi) - f_i(x, y_i)|^2 \le \nu_f^2, \qquad \mathbb{E}|g_i(x, y_i; \zeta) - g_i(x, y_i)|^2 \le \nu_g^2.$$

Moreover, the stochastic sample functions are smooth with uniformly Lipschitz continuous gradients.

The strong convexity condition in Assumption 2 guarantees the uniqueness of each lower-level solution. Assumption 3 imposes standard unbiasedness and bounded-variance conditions on the stochastic estimators, ensuring that the stochastic errors remain well controlled throughout the optimization process. Under these standard assumptions, we now establish the convergence guarantee of FOSMO. To characterize the convergence of the preference-guided aggregation, we define the smooth Tchebycheff scalarized objective $\Phi(x) := \tau \log\left(\sum_{i=1}^{m} \exp\left(\frac{w_i(\phi_i(x)-r_i)}{\tau}\right)\right)$. We characterize the convergence of FOSMO through the stationarity of $\Phi(x)$, as stated in the following theorem. The complete proof is provided in Appendix B.

Theorem 1. *Under Assumptions 1–3, suppose that Φ is bounded below by $\Phi_{\inf} > -\infty$. Let $\eta_t = \eta_0(t+t_0)^{-3/5}, \beta_t = \beta_0(t+t_0)^{-2/5}$, $\alpha_t = \alpha_0(t+t_0)^{-4/5}, \lambda_t = \lambda_0(t+t_0)^{1/5}$, where $\eta_0, \beta_0, \alpha_0, \lambda_0, \xi > 0$ are fixed constants satisfying $\frac{\eta_0^2\lambda_0^2}{\alpha_0} \le c_\eta, \frac{\beta_0^2}{\alpha_0} \le c_\beta, \frac{\eta_0^2\lambda_0^2}{\beta_0^2} \le c_{\mathrm{ts}}$, for sufficiently small constants $c_\eta, c_\beta, c_{\mathrm{ts}} > 0$, and $\xi > 0$ is sufficiently small. For sufficiently large $t_0 \ge 1$, the iterates generated by Algorithm 1 satisfy, for any $T \ge 1$, there holds:*

$$\sum_{t=0}^{T-1} \xi\eta_t \mathbb{E}\|\nabla\Phi(x_t)\|^2 \le \mathcal{O}(1)\cdot\sum_t \frac{\xi\eta_t}{\lambda_t^2} + \mathcal{O}(1)\cdot\sum_t \xi\eta_t\alpha_t\left[\delta_f^2 + \nu_f^2 + \lambda_t^2(\delta_g^2+\nu_g^2)\right] + \mathcal{O}(1)\cdot\sum_t \xi\frac{\eta_t}{\alpha_t}(\Delta\lambda_t)^2\delta_g^2 + \mathcal{O}(1)\cdot\sum_t \xi\frac{(\Delta\lambda_t)^2}{\eta_t\lambda_t^4} + \mathcal{O}(1),$$

where $\Delta\lambda_t := \lambda_{t+1} - \lambda_t$, and $\mathcal{O}(1)$ denotes an instance-dependent constant independent of T.

From Theorem 1, we can immediately obtain the following corollary. The proof is provided in Appendix B.

Corollary 1. *Under the conditions of Theorem 1, let $R \sim \mathrm{Unif}\{0,\dots,T-1\}$, then there exists a constant $C > 0$, independent of T, such that $\mathbb{E}\left[\|\nabla\Phi(x_R)\|^2\right] \le C(1+\log T)/T^{2/5} = \widetilde{\mathcal{O}}(T^{-2/5})$.*

Note that the gradient of $\Phi(x)$ is a weighted combination of the objective gradients $\phi(x)$, i.e., $\nabla\Phi(x) = \sum_{i=1}^{m} p_i(x)w_i\nabla\phi_i(x)$, where $p_i(x)$ denotes the smooth Tchebycheff aggregation weight. Since $p_i(x)w_i \ge 0$, these weights can be normalized to form a vector in the probability simplex. Therefore, by Definition 2, the convergence result in Corollary 1 also establishes convergence toward Pareto stationarity. To further characterize the computational efficiency of Algorithm 1, we consider its oracle complexity as defined in Definition 4.

Theorem 2. *Under the conditions of Corollary 1, suppose that $\boldsymbol{w} \in \Delta_m$ and there exists a constant $c_w > 0$, independent of m, such that $w_i \ge c_w/m$ for all $i \in [m]$. Then Algorithm 1 converges to an expected ϵ-Pareto stationary point with an oracle complexity of $\widetilde{\mathcal{O}}\left(\epsilon^{-5/2}m^6\right)$.*

Furthermore, if we relax the simplex constraint $\boldsymbol{w} \in \Delta_m$ to the box constraint, i.e., $0 < c_1 \le w_i \le c_2$ for all $i \in [m]$, we can obtain the following result. The proof is given in Appendix B

Corollary 2. *Under the conditions of Corollary 1, suppose that $0 < c_1 \le w_i \le c_2$ for all $i \in [m]$, where c_1 and c_2 are constants independent of m, then we have $\mathbb{E}\left[\|\nabla\Phi(x_R)\|^2\right] = \widetilde{\mathcal{O}}(mT^{-2/5})$. Consequently, Algorithm 1 converges to an expected ϵ-Pareto stationary point with an oracle complexity of $\widetilde{\mathcal{O}}(m^{7/2}\epsilon^{-5/2})$.*

Proof Sketch. The proofs of the main theorem and its corollaries are deferred to Appendix B. Here, we only present two key lemmas underlying the convergence analysis. The first lemma shows that the value-function-based penalty proxy $L_{i,\lambda}^*(x)$ approximates both the value and the gradient of the true upper-level objective $\phi_i(x)$. The detailed proof is provided in Appendix C.

Lemma 1. *Under Assumptions 1 and 2, with $\lambda\mu_g - l_{f,1} > 0$, We have the following result:*

$$\|L_{i,\lambda}^*(x) - \phi_i(x)\| \le \frac{C_1}{\lambda}, \quad \|\nabla L_{i,\lambda}^*(x) - \nabla\phi_i(x)\| \le \frac{C_2}{\lambda}. \tag{9}$$

where $C_1 := \frac{l_{f,0}^2}{2\mu_g}, C_2 = \frac{l_{f,0}}{\mu_g}(l_{f,1} + \frac{l_{g,1}l_{f,1}}{\mu_g} + \frac{l_{g,1}l_{g,2}l_{f,0}}{2\mu_g^2} + \frac{l_{g,2}l_{f,0}}{2\mu_g})$.

This lemma implies that, by choosing the penalty parameter λ sufficiently large, the proxy gradient can approximate the true hypergradient arbitrarily well. Therefore, the proxy gradient can be

treated as a controllable approximation of the exact hypergradient without requiring second-order information.

Building on Lemma 1, we next characterize the one-step progress of the smooth Tchebycheff objective $\Phi(x)$. At each iteration, FOSMO uses y_i^t to approximate $y_{i,\lambda_t}^*(x_t)$ and z_i^t to approximate $z_i^*(x_t)$, so the resulting first-order direction is generally different from the exact gradient of the penalized value function. In addition, both the gradient blocks and the function values used to construct the adaptive Tchebycheff weights are estimated stochastically. Consequently, the actual update direction deviates from $\nabla\Phi(x_t)$ due to the lower-level tracking errors, stochastic estimation errors, and the finite-penalty approximation error. The following lemma quantifies the effect of these errors and establishes a one-step descent inequality for $\Phi(x)$.

Lemma 2. *Under Assumptions 1–3. suppose that $\lambda_t \geq 1$ and $L_\Phi \xi\eta_t \leq \frac{1}{4}$, where L_Φ denotes the smoothness constant of $\Phi(x)$. Then the following inequality holds:*

$$\begin{aligned}\mathbb{E}[\Phi(x_{t+1}) - \Phi(x_t)] \leq -\frac{\xi\eta_t}{4}\mathbb{E}[\|\nabla\Phi(x_t)\|^2] - \frac{\xi\eta_t}{4}\mathbb{E}[\|q_x^t\|^2] + 2\xi\eta_t\mathbb{E}[\|e_x^t\|^2] \\ + C_1\xi\eta_t\left[\lambda_t^2\sum_{i=1}^m w_i\mathbb{E}[\|y_i^t - y_{i,\lambda_t}^*\|^2] + \lambda_t^2\sum_{i=1}^m w_i\mathbb{E}[\|z_i^t - z_i^*\|^2] + \frac{1}{\lambda_t^2}\right] \quad (10) \\ + C_2\xi\eta_t\left[\sum_{i=1}^m w_i\mathbb{E}|e_{i,fv}^t|^2 + \lambda_t^2\sum_{i=1}^m w_i\mathbb{E}|e_{i,gy}^t|^2 + \lambda_t^2\sum_{i=1}^m w_i\mathbb{E}|e_{i,gz}^t|^2\right]\end{aligned}$$

where $e_{i,fv}^t := \widehat{F}_i^t - F_i^t, e_{i,gy}^t := \widehat{G}_{i,y}^t - G_{i,y}^t, e_{i,gz}^t := \widehat{G}_{i,z}^t - G_{i,z}^t, e_x^t = \hat{h}_x^t - q_x^t$, and C_1, C_2 are instance-dependent constants.

The crucial term in this inequality is the negative term $-\frac{1}{4}\xi\eta_t\mathbb{E}\|\nabla\Phi(x_t)\|^2$, which represents the descent obtained at iteration t. The positive terms account for the STORM estimation error, the proxy value estimation error, the tracking errors of the lower-level variables y_i^t and z_i^t, and the bias introduced by the penalty parameter λ_t. Thus, controlling these error terms through the step sizes and momentum parameters ensures sufficient descent. The detailed proof is provided in Appendix C.

5 EXPERIMENTS

In this section, we present two multi-objective bilevel experiments involving multi-task representation learning and hyperparameter optimization. In both experiments, we obtain the reference point for each objective by setting the corresponding preference weight to one and all other weights to zero, i.e., $w_i = 1$ and $w_j = 0$ for $j \neq i$. The resulting objective value is used as the reference value r_i, and temperature coefficient $\tau = 0.05$. All reported results are averaged over five independent runs with different random seeds. We follow Gu et al. (2023) for the problems setup. In both experiments, we use the stochastic method without variance reduction.

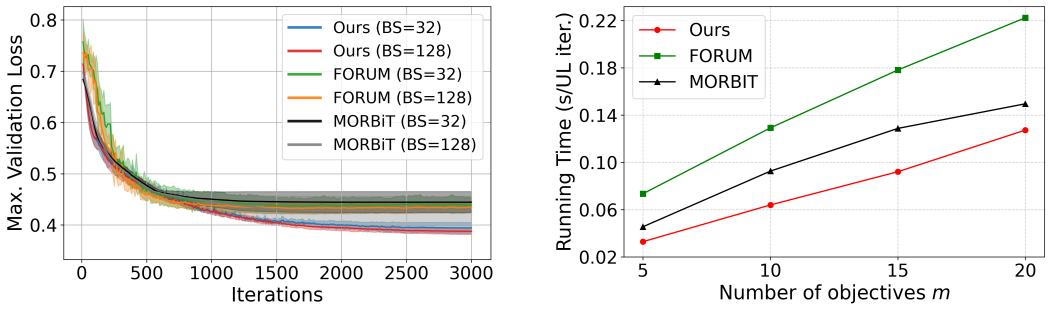
5.1 Multi-Task Representation Learning

Setup. In this experiment, we consider a multi-task representation learning problem, which can be naturally formulated as a bilevel optimization problem. The upper-level variable x parameterizes a shared representation network $\phi_x(\cdot)$, and each lower-task $i \in [m]$ has a task-specific liner model W_{y_i} parameterized by y_i. The data for task i are split into a training set $\mathcal{D}_i^{\text{tr}}$ and a validation set $\mathcal{D}_i^{\text{val}}$. Given the shared representation ϕ_x, the lower-level problem for each task i optimizes the task-specific parameters y_i by minimizing the training loss on $\mathcal{D}_i^{\text{tr}}$, yielding a task-specific model W_{y_i}, while the upper-level problem updates x according to the validation performance of W_{y_i} on $\mathcal{D}_i^{\text{val}}$. Let $\ell_i(f, \mathcal{D})$ denote the loss of model f on dataset $\mathcal{D}$. The resulting bilevel optimization problem is formulated as follows:

$$f_i(x, y_i) = \ell_i(W_{y_i} \circ \phi_x, \mathcal{D}_i^{\text{val}}), \qquad g_i(x, y_i) = \ell_i(W_{y_i} \circ \phi_x, \mathcal{D}_i^{\text{tr}}) + \rho||W_{y_i}||^2, \quad (11)$$

where $\rho > 0$ is the regularization parameter used to ensure the strong convexity of the lower-level problem, and we set $\rho = 0.001$ in all experiments.

We conduct this experiments on the CIFAR-100 (Krizhevsky, 2009) dataset and construct $n = 10$ binary classification tasks to form the multi-task learning setting. For each task i, we use the original



(a) Maximum validation loss over iterations.

(b) Running time with respect to m.

Figure 1: Experimental results on the CIFAR-100 dataset.

Table 1: Classification accuracy on the CIFAR-100 dataset, where o_i denotes the i-th objective.

Method	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}	Avg.
MORBiT	0.953	0.962	0.791	0.955	0.805	0.806	0.816	0.865	0.937	0.914	0.880
FORUM	**0.960**	0.977	0.829	0.967	0.816	0.833	0.841	0.930	0.958	0.937	0.905
Ours	**0.960**	**0.983**	**0.853**	**0.973**	**0.834**	**0.835**	**0.846**	**0.935**	**0.963**	**0.937**	**0.912**

training split as $\mathcal{D}_i^{\text{tr}}$. The original test split is randomly divided into two equally sized subsets, denoted by $\mathcal{D}_i^{\text{val}}$ and $\mathcal{D}_i^{\text{test}}$. The validation set $\mathcal{D}_i^{\text{val}}$ is used to evaluate the upper-level objective during bilevel optimization, while $\mathcal{D}_i^{\text{test}}$ is only used for the final evaluation. We compare FOSMO with two multi-objective bilevel optimization methods, FORUM (Ye et al., 2024) and MORBiT (Gu et al., 2023). FORUM is a fully first-order method but involves multiple nested loops. In our experiments, we set the number of inner iterations to 5. MORBiT adopts a single outer loop but requires hypergradient computation. By comparing against these two baselines, we aim to demonstrate the advantages of FOSMO in simultaneously achieving a single-loop structure and fully first-order updates.

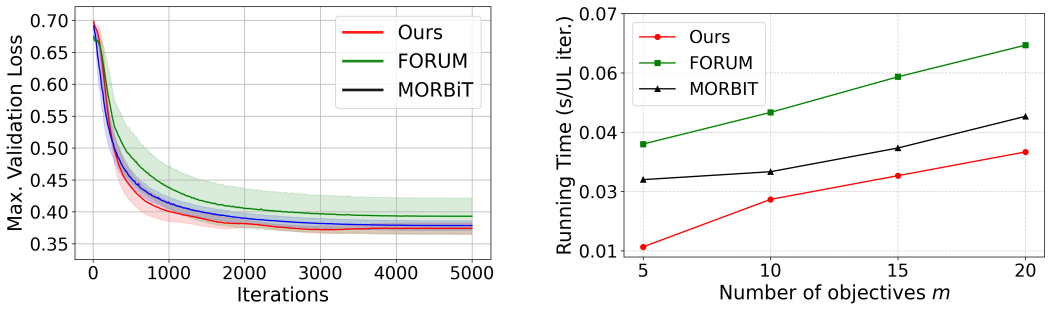
Figure 1a illustrates the evolution of the maximum validation loss for different methods over outer iterations. We observe that FOSMO consistently outperforms FORUM and MORBiT across different batch-size settings. Specifically, the performance of FORUM relies on sufficiently accurate solutions of the lower-level problems and additionally requires solving a quadratic programming subproblem; inaccurate lower-level solutions can therefore adversely affect its final performance. MORBiT employs iterative differentiation to approximate Hessian. Since the stochastic noise in Hessian estimates is typically larger than that in gradient estimates, the resulting noise may degrade its empirical performance. Figure 1b reports the per outer iteration runtime of different methods as the number of objectives varies. We observe that FOSMO consistently achieves the lowest runtime, benefiting from its single-loop structure and the absence of Hessian computations.

To further evaluate the ability to explore the Pareto front, we report the classification accuracy on different tasks in Table 1. When evaluating task i, FOSMO sets $w_i = 0.82$ and $w_j = 0.02$ for $j \neq i$. The results show that FOSMO achieves both a higher average accuracy and strong task-specific accuracy, indicating a broader ability to explore different trade-offs along the Pareto front.

5.2 Hyperparameter Optimization

Setup. Hyperparameter Optimization is a typical bilevel optimization problem. In this setting, we consider a hyperparameter optimization problem for kernel logistic regression with K random Fourier features. The upper-level variable is parameterized by $x = \{x_\sigma, x_\rho\}$, where x_σ represents the kernel bandwidth and x_ρ denotes the regularization parameter, with x_ϕ denoting the RFF[1]. Given the upper-level hyperparameters, each lower-level task i trains a model W_{y_i} parameterized by y_i. Similarly, $\mathcal{D}_i^{\text{tr}}$ is used for lower-level training, $\mathcal{D}_i^{\text{val}}$ for upper-level optimization. The resulting

[1]For a d-dimension point p, the RFF $\phi_{x_\sigma}(p) = \left[\sin(W(x_\sigma \odot p))^\top, \cos(W(x_\sigma \odot p))^\top\right]^\top \in \mathbb{R}^{2K}$, where $W \in \mathbb{R}^{K\times d}$ is a random normal matrix and the $\sin(\cdot)$ and $\cos(\cdot)$ are applied elementwise.



(a) Maximum validation loss over iterations.

(b) Running time with respect to m.

Figure 2: Experimental results on the Letter dataset.

Table 2: Classification accuracy on the Letter dataset, where o_i denotes the i-th objective.

Method	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}	Avg.
MORBiT	0.964	0.942	0.965	0.950	0.972	0.957	0.970	0.929	0.990	0.906	0.954
FORUM	0.940	0.850	0.935	0.961	0.923	0.900	0.938	0.924	0.951	0.922	0.925
Ours	**0.981**	**0.959**	**0.977**	**0.986**	**0.994**	**0.973**	**0.982**	**0.941**	**0.995**	**0.974**	**0.976**

bilevel optimization problem can be formulated as follows:

$$f_i(x, y_i) = \ell_i(W_{y_i} \circ \phi_{x_\sigma}, \mathcal{D}_i^{\text{val}}), \qquad g_i(x, y_i) = \ell_i(W_{y_i} \circ \phi_{x_\sigma}, \mathcal{D}_i^{\text{tr}}) + ||x_\rho \circ W_{y_i}||^2. \tag{12}$$

We conduct experiments on the Letter(Frey & Slate, 1991) dataset and construct $n = 10$ binary classification problems as the lower-level tasks. For each task, the data are randomly split into $\mathcal{D}_i^{\text{val}}$, $\mathcal{D}_i^{\text{tr}}$, and $\mathcal{D}_i^{\text{test}}$ with proportions of 20%, 60%, and 20%, respectively. We also compare FOSMO with FORUM and MORBiT.

Figure 2 presents the comparison of different methods in terms of validation loss and runtime per outer iteration. Similar to the previous experiment, FOSMO consistently achieves a lower validation loss than FORUM (Ye et al., 2024) and MORBiT (Gu et al., 2023), indicating better optimization performance. In addition, FOSMO requires less computational time for each outer iteration, benefiting from its single-loop structure and the avoidance of Hessian-related computations. To further evaluate the ability to explore the Pareto front, Table 2 reports the classification accuracy on the ten tasks under different preference settings. As in the previous experiment, when evaluating task i, we set $w_i = 0.82$ and $w_j = 0.02$ for $j \neq i$. FOSMO achieves competitive task-specific classification accuracy as well as a high average accuracy across the ten tasks, demonstrating its ability to explore diverse trade-offs among different objectives. Overall, these results are consistent with those in the multi-task representation learning experiment and further support the effectiveness, computational efficiency, and Pareto exploration capability of FOSMO.

6 Conclusion

In this paper, we propose FOSMO, a fully first-order single-loop approach for stochastic multi-objective bilevel optimization. FOSMO employs a value-function penalty reformulation to avoid second-order computations, together with recursive variance reduction to control the stochastic estimation errors. Separate auxiliary variables are updated once per iteration to track the lower-level and penalized optimal solutions, with their tracking errors shown to remain controllable throughout optimization, thereby enabling a single-loop scheme. We further use smooth Tchebycheff aggregation to enable preference-guided optimization and broader Pareto exploration. Under standard regularity assumptions, we establish a $\widetilde{\mathcal{O}}(T^{-2/5})$ convergence rate to a Pareto-stationary point, with a stochastic first-order oracle complexity of $\widetilde{\mathcal{O}}(m^6\epsilon^{-5/2})$. These results provide an efficient and theoretically guaranteed framework for stochastic multi-objective bilevel optimization using only first-order information.

AI USE STATEMENT

In this work, we used generative AI tools to help polish the language to improve readability. We have checked the LLM-generated content for potential errors. We take responsibility for the final content of this work.

REFERENCES

Jan Harold Alcantara, Chieu Thanh Nguyen, Takayuki Okuno, Akiko Takeda, and Jein-Shan Chen. Unified smoothing approach for best hyperparameter selection problem using a bilevel optimization strategy. *Mathematical Programming*, 212(1):479–518, 2025.

K.P. Bennett, Jing Hu, Xiaoyun Ji, G. Kunapuli, and Jong-Shi Pang. Model selection via bilevel optimization. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pp. 1922–1929, 2006. doi: 10.1109/IJCNN.2006.246935.

Qichao Cao, Shangzhi Zeng, and Jin Zhang. A single-loop gradient algorithm for pessimistic bilevel optimization via smooth approximation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=vKmWKHlQBe.

Lesi Chen, Yaohua Ma, and Jingzhao Zhang. Near-Optimal Nonconvex-Strongly-Convex Bilevel Optimization with Fully First-Order Oracles. *Journal of Machine Learning Research*, 26:1–56, 2025.

Tianyi Chen, Yuejiao Sun, and Wotao Yin. A single-timescale stochastic bilevel optimization method. *ArXiv*, abs/2102.04671, 2021a. URL https://api.semanticscholar.org/CorpusID:231855754.

Tianyi Chen, Yuejiao Sun, and Wotao Yin. Closing the gap: Tighter analysis of alternating stochastic gradient methods for bilevel problems. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021b.

John M. Danskin. The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, 14(4):641–664, 1966. doi: 10.1137/0114053.

Jörg Fliege and Benar Fux Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51:479–494, 2000.

Peter W. Frey and David J. Slate. Letter recognition using holland-style adaptive classifiers. 6(2), 1991. ISSN 0885-6125. doi: 10.1023/A:1022606404104. URL https://doi.org/10.1023/A:1022606404104.

Lucy L Gao, Jane Ye, Haian Yin, Shangzhi Zeng, and Jin Zhang. Value function based difference-of-convex algorithm for bilevel hyperparameter selection problems. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 7164–7182. PMLR, 17–23 Jul 2022.

Saeed Ghadimi and Mengdi Wang. Approximation methods for bilevel programming, 2018. URL https://arxiv.org/abs/1802.02246.

Tommaso Giovannelli, Griffin Dean Kent, and Luis Nunes Vicente. Inexact bilevel stochastic gradient methods for constrained and unconstrained lower-level problems, 2023. URL https://arxiv.org/abs/2110.00604.

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

Zhihao Gu, Qilong Wu, and Junchi Yang. Sgha: A single-loop fully first-order algorithm for nonconvex-strongly-convex bilevel optimization, 2026. URL https://arxiv.org/abs/2608.23211.

Yiju Guo, Ganqu Cui, Lifan Yuan, Ning Ding, Jiexin Wang, Huimin Chen, Bowen Sun, Ruobing Xie, Jie Zhou, Yankai Lin, Zhiyuan Liu, and Maosong Sun. Controllable preference optimization: Toward controllable multi-objective alignment. In *Conference on Empirical Methods in Natural Language Processing*, 2024. URL https://api.semanticscholar.org/CorpusID:268063181.

Yicong Jiang and Feihu Huang. Efficient hessian-free methods for multi-objective bilevel optimization with nonconvex lower level, 2026a. URL https://arxiv.org/abs/2608.12704.

Yicong Jiang and Feihu Huang. Efficient hessian-free methods for multi-objective bilevel optimization with nonconvex lower level, 2026b. URL https://arxiv.org/abs/2608.12704.

Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009. URL https://api.semanticscholar.org/CorpusID:18268744.

Jeongyeol Kwon, Dohyun Kwon, Stephen Wright, and Robert D. Nowak. A fully first-order method for stochastic bilevel optimization. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 18083–18113, 2023.

Junyi Li, Bin Gu, and Heng Huang. A fully single loop algorithm for bilevel optimization without hessian inverse, 2021a. URL https://arxiv.org/abs/2112.04660.

Qingna Li, Zhen Li, and Alain Zemkoho. Bilevel hyperparameter optimization for support vector classification: theoretical analysis and a solution method. *arXiv preprint arXiv:2110.01697*, 2021b.

Baijiong Lin, Weisen Jiang, Yuancheng Xu, Hao Chen, and Ying-Cong Chen. PARM: Multi-objective test-time alignment via preference-aware autoregressive reward model. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=zm53HtGiXN.

Xi Lin, Hui-Ling Zhen, Zhenhua Li, Qing-Fu Zhang, and Sam Kwong. Pareto multi-task learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/685bfde03eb646c27ed565881917c71c-Paper.pdf.

Bo Liu, Mao Ye, Stephen Wright, Peter Stone, and Qiang Liu. Bome! bilevel optimization made easy: A simple first-order approach. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 17248–17262. Curran Associates, Inc., 2022. doi: 10.52202/068431-1254.

Risheng Liu, Yaohua Liu, Shangzhi Zeng, and Jin Zhang. Towards gradient-based bilevel optimization with non-convex followers and beyond. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 8662–8675. Curran Associates, Inc., 2021.

Risheng Liu, Zhu Liu, Wei Yao, Shang-Zhi Zeng, and Jin Zhang. Moreau envelope for nonconvex bi-level optimization: A single-loop and hessian-free solution strategy. *ArXiv*, abs/2405.09927, 2024. URL https://api.semanticscholar.org/CorpusID:269790867.

Matthew MacKay, Paul Vicol, Jon Lorraine, David Duvenaud, and Roger Grosse. Self-tuning networks: Bilevel optimization of hyperparameters using structured best-response functions. *arXiv preprint arXiv:1903.03088*, 2019.

Debabrata Mahapatra and Vaibhav Rajan. Multi-task learning with user preferences: Gradient descent with controlled ascent in pareto optimization. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6597–6607. PMLR, 13–18 Jul 2020.

Rui Pan, Dylan Zhang, Hanning Zhang, Xingyuan Pan, Minrui Xu, Jipeng Zhang, Renjie Pi, Xiaoyu Wang, and Tong Zhang. Scalebio: Scalable bilevel optimization for llm data reweighting. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 31959–31982, 2025.

Cac Phan and Kai Wang. Bridging constraints and stochasticity: A fully first-order method for stochastic bilevel optimization with linear constraints, 2025. URL https://arxiv.org/abs/2511.09845.

Xiaorong Qin, Xinhang Song, and Shuqiang Jiang. Bi-level meta-learning for few-shot domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15900–15910, June 2023.

Aravind Rajeswaran, Chelsea Finn, Sham M. Kakade, and Sergey Levine. Meta-learning with implicit gradients. In *Neural Information Processing Systems*, 2019. URL https://api.semanticscholar.org/CorpusID:202542766.

Mohammad Sadegh Salehi, Subhadip Mukherjee, Lindon Roberts, and Matthias J Ehrhardt. An adaptively inexact first-order method for bilevel optimization with application to hyperparameter learning. *SIAM Journal on Mathematics of Data Science*, 7(3):906–936, 2025.

Wei Shen, Jiawei zhang, Minhui Huang, and Cong Shen. A single-loop first-order algorithm for linearly constrained bilevel optimization. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 57432–57479. Curran Associates, Inc., 2025. doi: 10.52202/085713-1925.

Ruizhe Shi, Yifang Chen, Yushi Hu, Alisa Liu, Hannaneh Hajishirzi, Noah A. Smith, and Simon Shaolei Du. Decoding-time language model alignment with multiple objectives. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=3csuL7TVpV.

Daouda Sow, Kaiyi Ji, Ziwei Guan, and Yingbin Liang. A primal-dual approach to bilevel optimization with multiple inner minima, 2022. URL https://arxiv.org/abs/2203.01123.

Haoxiang Wang, Yong Lin, Wei Xiong, Rui Yang, Shizhe Diao, Shuang Qiu, Han Zhao, and Tong Zhang. Arithmetic control of LLMs for diverse user preferences: Directional preference alignment with multi-objective rewards. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8642–8655, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.468. URL https://aclanthology.org/2024.acl-long.468/.

Andrzej P. Wierzbicki. The use of reference objectives in multiobjective optimization. In Günter Fandel and Tomas Gal (eds.), *Multiple Criteria Decision Making Theory and Application*, pp. 468–486, Berlin, Heidelberg, 1980. Springer Berlin Heidelberg.

Junjie Yang, Kaiyi Ji, and Yingbin Liang. Provably faster algorithms for bilevel optimization. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021.

Rui Yang, Xiaoman Pan, Feng Luo, Shuang Qiu, Han Zhong, Dong Yu, and Jianshu Chen. Rewards-in-context: Multi-objective alignment of foundation models with dynamic preference adjustment. In *Forty-first International Conference on Machine Learning*, 2024a. URL https://openreview.net/forum?id=QLcBzRI3V3.

Yifan Yang, Zhaofeng Si, Siwei Lyu, and Kaiyi Ji. First-order minimax bilevel optimization. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 24990–25035. Curran Associates, Inc., 2024b. doi: 10.52202/079017-0788. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/2c8047bf3ed8ef6905351608d641f02f-Paper-Conference.pdf.

Feiyang Ye, Baijiong Lin, Zhixiong Yue, Pengxin Guo, Qiao Xiao, and Yu Zhang. Multi-objective meta learning. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 21338–21351. Curran Associates, Inc., 2021.

Feiyang Ye, Baijiong Lin, Xiaofeng Cao, Yu Zhang, and Ivor Tsang. A first-order multi-gradient algorithm for multi-objective bi-level optimization, 2024. URL https://arxiv.org/abs/2401.09257.

Yihan Zhang, Xinwen Zhang, My T. Thai, Jie Wu, and Hongchang Gao. Convergence analysis of decentralized hessian-/jacobian-free algorithm for nonconvex stochastic bilevel optimization. In *Forty-third International Conference on Machine Learning*, 2026a.

Zhiyao Zhang, Zhuqing Liu, Xin Zhang, Wen-Yen Chen, Jiyan Yang, and Jia Liu. Multi-objective bilevel learning, 2025. URL https://arxiv.org/abs/2511.07824.

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, 2026b. URL https://arxiv.org/abs/2605.09094.

A NOTATIONS

For ease of proof, we introduce the following notation. Denote

$$\begin{aligned} q_{i,z}^t &= \nabla_{z_i} g_i(x_t, z_i^t) \\ q_{i,y}^t &= \nabla_{y_i} f_i(x_t, y_i^t) + \lambda_t \nabla_{y_i} g_i(x_t, y_i^t) \\ q_{i,x}^t &= \nabla_x f_i(x_t, y_i^t) + \lambda_t \left(\nabla_x g_i(x_t, y_i^t) - \nabla_x g_i(x_t, z_i^t)\right) \\ q_x^t &= \sum_{i=1}^m \hat{p}_i^t w_i q_{i,x}^t, \end{aligned} \tag{13}$$

where $q_{i,z}^t$, $q_{i,y}^t$ and $q_{i,x}^t$ are exact gradients without stochastic noises. Now we make the following definitional decomposition of the stochastic gradient

$$e_{i,z}^t = \hat{h}_{i,z}^t - q_{i,z}^t, \quad e_{i,y}^t = \hat{h}_{i,y}^t - q_{i,y}^t, \quad e_x^t = \hat{h}_x^t - q_x^t, \quad e_{i,x}^t = \hat{h}_{i,x}^t - q_{i,x}^t \tag{14}$$

Let

$$e_{i,fv}^t := \widehat{F}_i^t - F_i^t, \qquad e_{i,gy}^t := \widehat{G}_{i,y}^t - G_{i,y}^t, \qquad e_{i,gz}^t := \widehat{G}_{i,z}^t - G_{i,z}^t,$$

and define the weighted mean-square value-estimation errors

$$E_{fv,t} := \sum_{i=1}^m w_i \mathbb{E}|e_{i,fv}^t|^2, \qquad E_{gy,t} := \sum_{i=1}^m w_i \mathbb{E}|e_{i,gy}^t|^2, \qquad E_{gz,t} := \sum_{i=1}^m w_i \mathbb{E}|e_{i,gz}^t|^2.$$

B MAIN THEORETICAL PROOF

We first present a series of auxiliary lemmas that are used in the proof of the Theorem 1, with their proofs deferred to Appendix C.

B.1 AUXILIARY LEMMA

Lemma 3. *Let q_t denote the exact quantity, with $h_t(\omega)$ being its unbiased stochastic estimate, and let $\hat{h}_t$ be the recursive estimator defined in equation 8. Define $\varepsilon_{t+1} := h_{t+1}(\omega_{t+1}) - q_{t+1}, \varepsilon_t := h_t(\omega_{t+1}) - q_t$, where the same fresh sample ω_{t+1} is used in both stochastic evaluations. Then*

$$\begin{aligned} \mathbb{E}\|\hat{h}_{t+1} - q_{t+1}\|^2 \le (1-\alpha_{t+1})^2 \mathbb{E}\|\hat{h}_t - q_t\|^2 \\ + 2\mathbb{E}\|\varepsilon_{t+1} - \varepsilon_t\|^2 + 2\alpha_{t+1}^2 \mathbb{E}\|\varepsilon_{t+1}\|^2. \end{aligned} \tag{15}$$

Lemma 4. *Fix any $i \in [m]$, then*

$$\mathbb{E}\|e_{i,z}^{t+1}\|^2 \leq \left[(1-\alpha_i^{t+1})^2 + 4l_{g,1}^2\beta_t^2\right]\mathbb{E}\|e_{i,z}^t\|^2 + 4l_{g,1}^2\xi^2\eta_t^2\mathbb{E}\|e_x^t\|^2 + 4l_{g,1}^2\beta_t^2\mathbb{E}\|q_{i,z}^t\|^2 + 4l_{g,1}^2\xi^2\eta_t^2\mathbb{E}\|q_x^t\|^2 + 2(\alpha_i^{t+1})^2\delta_g^2. \tag{16}$$

Lemma 5. *Fix any $i \in [m]$, then*

$$\mathbb{E}\|e_{i,y}^{t+1}\|^2 \leq \left[(1-\alpha_i^{t+1})^2 + 12\big(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\big)\eta_t^2\right]\mathbb{E}\|e_{i,y}^t\|^2 + 12\big(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\big)\xi^2\eta_t^2\mathbb{E}\|e_x^t\|^2 + 12\big(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\big)\eta_t^2\mathbb{E}\|q_{i,y}^t\|^2 + 12\big(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\big)\xi^2\eta_t^2\mathbb{E}\|q_x^t\|^2 + 4(\alpha_i^{t+1})^2\left(\delta_f^2 + \lambda_{t+1}^2\delta_g^2\right) + 6(\Delta\lambda_t)^2\delta_g^2. \tag{17}$$

Lemma 6. *For each $i \in [m]$, then*

$$\mathbb{E}\|e_{i,x}^{t+1}\|^2 \leq (1-\alpha_i^{t+1})^2\mathbb{E}\|e_{i,x}^t\|^2 + 16\left(l_{f,1}^2 + 2\lambda_{t+1}^2 l_{g,1}^2\right)\xi^2\eta_t^2\left(\mathbb{E}\|e_x^t\|^2 + \mathbb{E}\|q_x^t\|^2\right) + 16\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\eta_t^2\left(\mathbb{E}\|e_{i,y}^t\|^2 + \mathbb{E}\|q_{i,y}^t\|^2\right) + 16\lambda_{t+1}^2 l_{g,1}^2\beta_t^2\left(\mathbb{E}\|e_{i,z}^t\|^2 + \mathbb{E}\|q_{i,z}^t\|^2\right) + 6(\alpha_i^{t+1})^2\left(\delta_f^2 + 2\lambda_{t+1}^2\delta_g^2\right) + 32(\Delta\lambda_t)^2\delta_g^2. \tag{18}$$

Moreover,

$$\mathbb{E}\|e_x^{t+1}\|^2 \leq \sum_{i=1}^m w_i\mathbb{E}\|e_{i,x}^{t+1}\|^2. \tag{19}$$

Lemma 7. *For every $i \in [m]$, then*

$$\mathbb{E}|e_{i,fv}^{t+1}|^2 \leq (1-\alpha_{t+1})^2\mathbb{E}|e_{i,fv}^t|^2 + 4l_{f,1}^2\xi^2\eta_t^2\left(\mathbb{E}\|e_x^t\|^2 + \mathbb{E}\|q_x^t\|^2\right) + 4l_{f,1}^2\eta_t^2\left(\mathbb{E}\|e_{i,y}^t\|^2 + \mathbb{E}\|q_{i,y}^t\|^2\right) + 2\alpha_{t+1}^2\nu_f^2, \tag{20}$$

and

$$\mathbb{E}|e_{i,gy}^{t+1}|^2 \leq (1-\alpha_{t+1})^2\mathbb{E}|e_{i,gy}^t|^2 + 4l_{g,1}^2\xi^2\eta_t^2\left(\mathbb{E}\|e_x^t\|^2 + \mathbb{E}\|q_x^t\|^2\right) + 4l_{g,1}^2\eta_t^2\left(\mathbb{E}\|e_{i,y}^t\|^2 + \mathbb{E}\|q_{i,y}^t\|^2\right) + 2\alpha_{t+1}^2\nu_g^2. \tag{21}$$

Moreover,

$$\mathbb{E}|e_{i,gz}^{t+1}|^2 \leq (1-\alpha_{t+1})^2\mathbb{E}|e_{i,gz}^t|^2 + 4l_{g,1}^2\xi^2\eta_t^2\left(\mathbb{E}\|e_x^t\|^2 + \mathbb{E}\|q_x^t\|^2\right) + 4l_{g,1}^2\beta_t^2\left(\mathbb{E}\|e_{i,z}^t\|^2 + \mathbb{E}\|q_{i,z}^t\|^2\right) + 2\alpha_{t+1}^2\nu_g^2. \tag{22}$$

Lemma 8. *Let $\bar{\mu}_t := \lambda_t\mu_g - l_{f,1}, \bar{L}_t := l_{f,1} + \lambda_t l_{g,1}$, and assume $\bar{\mu}_t > 0$. Define $\kappa_{y,t} := \frac{\bar{L}_t}{\bar{\mu}_t}$. Suppose that $\eta_t \leq \min\left\{\frac{\bar{\mu}_t}{4\bar{L}_t^2}, \frac{1}{\bar{\mu}_t}\right\}$. Then*

$$\mathbb{E}\left[\left\|y_i^{t+1} - y_{i,\lambda_{t+1}}^*(x^{t+1})\right\|^2\right] \leq \left(1 - \frac{\bar{\mu}_t\eta_t}{8}\right)\mathbb{E}\left[\left\|y_i^t - y_{i,\lambda_t}^*(x^t)\right\|^2\right] + \frac{5\eta_t}{\bar{\mu}_t}\mathbb{E}\|e_{i,y}^t\|^2 + \frac{12\kappa_{y,t}^2\xi^2\eta_t}{\bar{\mu}_t}\mathbb{E}\left[\|e_x^t\|^2 + \|q_x^t\|^2\right] + \frac{9l_{f,0}^2(\Delta\lambda_t)^2}{\bar{\mu}_t\eta_t\lambda_t^2\bar{\mu}_{t+1}^2}, \tag{23}$$

where

$$\Delta\lambda_t := \lambda_{t+1} - \lambda_t.$$

Lemma 9. *Let $\kappa_z := \frac{l_{g,1}}{\mu_g}$. Suppose that $\beta_t \leq \min\left\{\frac{\mu_g}{4l_{g,1}^2}, \frac{1}{\mu_g}\right\}$. Then we have*

$$\mathbb{E}\left[\left\|z_i^{t+1} - y_i^*(x^{t+1})\right\|^2\right] \leq \left(1 - \frac{\mu_g\beta_t}{4}\right)\mathbb{E}\left[\left\|z_i^t - y_i^*(x^t)\right\|^2\right] + \frac{4\beta_t}{\mu_g}\mathbb{E}\|e_{i,z}^t\|^2 + \frac{10\kappa_z^2\xi^2\eta_t^2}{\mu_g\beta_t}\mathbb{E}\left[\|e_x^t\|^2 + \|q_x^t\|^2\right]. \tag{24}$$

B.2 Proof of Theorem 1

Proof. For convenience, define the weighted gradient-estimation and lower-level tracking errors

$$E_{x,t} := \sum_{i=1}^{m} w_i \mathbb{E}|e_{i,x}^t|^2, \quad E_{y,t} := \sum_{i=1}^{m} w_i \mathbb{E}|e_{i,y}^t|^2, \quad E_{z,t} := \sum_{i=1}^{m} w_i \mathbb{E}|e_{i,z}^t|^2,$$
$$R_{y,t} := \sum_{i=1}^{m} w_i \mathbb{E}|y_i^t - y_{i,\lambda_t}^*(x_t)|^2, \quad R_{z,t} := \sum_{i=1}^{m} w_i \mathbb{E}|z_i^t - y_i^*(x_t)|^2. \tag{25}$$

Since $\hat{p}^t, w \in \Delta_m$, weighted Cauchy–Schwarz, the optimality conditions, and smoothness give

$$\mathbb{E}|e_x^t|^2 \le E_{x,t}, \qquad \sum_{i=1}^{m} w_i \mathbb{E}|q_{i,z}^t|^2 \le l_{g,1}^2 R_{z,t}, \qquad \sum_{i=1}^{m} w_i \mathbb{E}|q_{i,y}^t|^2 \le \bar{L}_t^2 R_{y,t}, \tag{26}$$

where $\bar{L}_t := l_{f,1} + \lambda_t l_{g,1}$. We first verify that the parameter schedules satisfy all conditions required by Lemma 2 and Lemma4–9. Let $s_t := t + t_0$. By the prescribed schedules, we have

$$\frac{\eta_t^2 \lambda_t^2}{\alpha_t} = \frac{\eta_0^2 \lambda_0^2}{\alpha_0} \le c_\eta, \qquad \frac{\beta_t^2}{\alpha_t} = \frac{\beta_0^2}{\alpha_0} \le c_\beta, \qquad \frac{\eta_t^2 \lambda_t^2}{\beta_t^2} = \frac{\eta_0^2 \lambda_0^2}{\beta_0^2} \le c_{\mathrm{ts}}. \tag{27}$$

For sufficiently large t_0, there exist constants $c_\mu, C_L, C_\kappa > 0$, independent of t and T, such that

$$\lambda_t \ge 1, \qquad \bar{\mu}_t \ge c_\mu \lambda_t, \qquad \bar{L}_t \le C_L \lambda_t, \qquad \frac{\bar{L}_t}{\bar{\mu}_t} \le C_\kappa.$$

Since $\eta_t = \Theta(s_t^{-3/5})$, $\beta_t = \Theta(s_t^{-2/5})$, and $\lambda_t^{-1} = \Theta(s_t^{-1/5})$, increasing t_0 if necessary ensures that

$$\eta_t \le \min\left\{\frac{\bar{\mu}_t}{4\bar{L}_t^2}, \frac{1}{\bar{\mu}_t}\right\}, \qquad \beta_t \le \min\left\{\frac{\mu_g}{4l_{g,1}^2}, \frac{1}{\mu_g}\right\}, \qquad L_\Phi \xi \eta_t \le \frac{1}{4}.$$

Moreover, by equation 27 and sufficiently small c_η, c_β and ξ, it holds that

$$16\big(l_{f,1}^2 + 2\lambda_{t+1}^2 l_{g,1}^2\big)\xi^2 \eta_t^2 \le \frac{\alpha_t}{4}, \quad 12\big(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\big)\eta_t^2 \le \frac{\alpha_t}{4}, \quad 4l_{g,1}^2 \beta_t^2 \le \frac{\alpha_t}{4}.$$

The Lyapunov weights below are slowly varying. Indeed,

$$\frac{\eta_{t+1}/\alpha_{t+1}}{\eta_t/\alpha_t} \le 1 + \frac{\alpha_t}{8}, \ \frac{\eta_{t+1}\lambda_{t+1}^2/\alpha_{t+1}}{\eta_t \lambda_t^2/\alpha_t} \le 1 + \frac{\alpha_t}{8}, \ \frac{\lambda_{t+1}}{\lambda_t} \le 1 + \frac{\bar{\mu}_t \eta_t}{16}, \ \frac{\eta_{t+1}\lambda_{t+1}^2/\beta_{t+1}}{\eta_t \lambda_t^2/\beta_t} \le 1 + \frac{\mu_g \beta_t}{8}, \tag{28}$$

where we used

$$\frac{\eta_{t+1}/\alpha_{t+1}}{\eta_t/\alpha_t} = \left(1 + \frac{1}{s_t}\right)^{1/5}, \qquad \frac{\eta_{t+1}\lambda_{t+1}^2/\alpha_{t+1}}{\eta_t \lambda_t^2/\alpha_t} = \left(1 + \frac{1}{s_t}\right)^{3/5},$$

and $s_t^{-1} = o(\alpha_t)$, $s_t^{-1} = o(\bar{\mu}_t \eta_t)$, and $s_t^{-1} = o(\beta_t)$. We now construct the Lyapunov function

$$V_t := \mathbb{E}[\Phi(x_t)] + K_x \frac{\xi \eta_t}{\alpha_t} E_{x,t} + K_y \frac{\xi \eta_t}{\alpha_t} E_{y,t} + K_z \frac{\xi \eta_t \lambda_t^2}{\alpha_t} E_{z,t} + K_{fv} \frac{\xi \eta_t}{\alpha_t} E_{fv,t} + K_{gy} \frac{\xi \eta_t \lambda_t^2}{\alpha_t} E_{gy,t}$$
$$+ K_{gz} \frac{\xi \eta_t \lambda_t^2}{\alpha_t} E_{gz,t} + K_{ry} \xi \lambda_t R_{y,t} + K_{rz} \frac{\xi \eta_t \lambda_t^2}{\beta_t} R_{z,t}, \tag{29}$$

where all K are positive constants independent of t and T. Multiply the recursions in Lemmas 4–7 by their corresponding weights in equation 29, multiply Lemmas 8 and 9 by $K_{ry}\xi\lambda_{t+1}$and$K_{rz}\frac{\xi\eta_{t+1}\lambda_{t+1}^2}{\beta_{t+1}}$, respectively, and add the resulting inequalities to the descent inequality in Lemma 2. By equation 28, the self-contraction terms yield negative contributions of order $\xi\eta_t$ for $E_{x,t}, E_{y,t}, E_{fv,t}$ and of order $\xi\eta_t\lambda_t^2$ for $E_{z,t}, R_{y,t}, R_{z,t}, E_{gy,t}, E_{gz,t}$. The remaining coupling terms are controlled by the three ratios in equation 27. In particular, the $E_{y,t}$ and $R_{y,t}$ couplings are proportional to $\frac{\eta_t^2\lambda_t^2}{\alpha_t}$, the $E_{z,t}$ and $R_{z,t}$ couplings are proportional to $\frac{\beta_t^2}{\alpha_t}$, and the additional $E_{x,t}$

and $\mathbb{E}|q_x^t|^2$ coupling generated by the z-tracking recursion is proportional to $\frac{\eta_t^2\lambda_t^2}{\beta_t^2}$. Hence, by choosing the constants in equation 29 sufficiently large and then choosing $c_\eta, c_\beta, c_{\text{ts}}$ and ξ sufficiently small, all these terms are absorbed by the corresponding negative Lyapunov terms and by the term $-\frac{\xi\eta_t}{4}\mathbb{E}\|q_x^t\|^2$ in Lemma 2. Therefore, there exist constants $c_0, C_0, C_1, C_2, C_3 > 0$, independent of t and T, such that

$$\begin{aligned} V_{t+1} - V_t \leq & -c_0\xi\eta_t\mathbb{E}|\nabla\Phi(x_t)|^2 + C_0\frac{\xi\eta_t}{\lambda_t^2} + C_1\xi\eta_t\alpha_t\left[\delta_f^2 + \nu_f^2 + \lambda_t^2(\delta_g^2 + \nu_g^2)\right] \\ & + C_2\xi\frac{\eta_t}{\alpha_t}(\Delta\lambda_t)^2\delta_g^2 + C_3\xi\frac{(\Delta\lambda_t)^2}{\eta_t\lambda_t^4}. \end{aligned} \tag{30}$$

Here the last term follows from Lemma 8 together with $\bar{\mu}_t \geq c_\mu\lambda_t, \bar{\mu}_{t+1} \geq c_\mu\lambda_{t+1}$, and all fixed problem-dependent factors are absorbed into C_3. Summing equation 30 from $t = 0$ to $T - 1$ gives

$$\begin{aligned} c_0\sum_{t=0}^{T-1}\xi\eta_t\mathbb{E}|\nabla\Phi(x_t)|^2 \leq V_0 - V_T + C_0\sum_{t=0}^{T-1}\frac{\xi\eta_t}{\lambda_t^2} + C_1\sum_{t=0}^{T-1}\xi\eta_t\alpha_t\left[\delta_f^2 + \nu_f^2 + \lambda_t^2(\delta_g^2 + \nu_g^2)\right] \\ + C_2\sum_{t=0}^{T-1}\xi\frac{\eta_t}{\alpha_t}(\Delta\lambda_t)^2\delta_g^2 + C_3\sum_{t=0}^{T-1}\xi\frac{(\Delta\lambda_t)^2}{\eta_t\lambda_t^4}. \end{aligned} \tag{31}$$

All error terms in equation 29 are nonnegative, and therefore $V_T \geq \mathbb{E}[\Phi(x_T)] \geq \Phi_{\text{inf}}$. Since $V_0 - \Phi_{\text{inf}}$ is finite and independent of T, dividing equation 31 by c_0 and absorbing all fixed constants into $\mathcal{O}(1)$ yields

$$\begin{aligned} \sum_{t=0}^{T-1}\xi\eta_t\mathbb{E}|\nabla\Phi(x_t)|^2 \leq \mathcal{O}(1)\sum_{t=0}^{T-1}\frac{\xi\eta_t}{\lambda_t^2} + \mathcal{O}(1)\sum_{t=0}^{T-1}\xi\eta_t\alpha_t\left[\delta_f^2 + \nu_f^2 + \lambda_t^2(\delta_g^2 + \nu_g^2)\right] \\ + \mathcal{O}(1)\sum_{t=0}^{T-1}\xi\frac{\eta_t}{\alpha_t}(\Delta\lambda_t)^2\delta_g^2 + \mathcal{O}(1)\sum_{t=0}^{T-1}\xi\frac{(\Delta\lambda_t)^2}{\eta_t\lambda_t^4} + \mathcal{O}(1), \end{aligned}$$

which proves the theorem. □

B.3 Proof of Corollary 1

Proof. Let $s_t := t + t_0$. Under the parameter schedules in Theorem 1,

$$\frac{\eta_t}{\lambda_t^2} = \frac{\eta_0}{\lambda_0^2}s_t^{-1}, \qquad \eta_t\alpha_t = \eta_0\alpha_0 s_t^{-7/5}, \qquad \eta_t\alpha_t\lambda_t^2 = \eta_0\alpha_0\lambda_0^2 s_t^{-1}. \tag{32}$$

Moreover, by the mean-value theorem, it holds

$$|\Delta\lambda_t| = \lambda_0\left[(s_t+1)^{1/5} - s_t^{1/5}\right] \leq \frac{\lambda_0}{5}s_t^{-4/5}. \tag{33}$$

Hence,

$$\frac{\eta_t}{\alpha_t}(\Delta\lambda_t)^2 = \mathcal{O}(s_t^{-7/5}), \qquad \frac{(\Delta\lambda_t)^2}{\eta_t\lambda_t^4} = \mathcal{O}(s_t^{-9/5}). \tag{34}$$

Therefore,

$$\begin{aligned} \sum_{t=0}^{T-1}\frac{\eta_t}{\lambda_t^2} = \mathcal{O}(1+\log(T+t_0)), \sum_{t=0}^{T-1}\eta_t\alpha_t\left[\delta_f^2 + \nu_f^2 + \lambda_t^2(\delta_g^2 + \nu_g^2)\right] = \mathcal{O}(1+\log(T+t_0)), \\ \sum_{t=0}^{T-1}\frac{\eta_t}{\alpha_t}(\Delta\lambda_t)^2\delta_g^2 = \mathcal{O}(1), \sum_{t=0}^{T-1}\frac{(\Delta\lambda_t)^2}{\eta_t\lambda_t^4} = \mathcal{O}(1). \end{aligned} \tag{35}$$

Substituting equation 35 into Theorem 1 gives

$$\sum_{t=0}^{T-1}\xi\eta_t\mathbb{E}|\nabla\Phi(x_t)|^2 \leq C\big(1+\log(T+t_0)\big), \tag{36}$$

where $C > 0$ is independent of T. Let $R \sim \mathrm{Unif}\, 0, \ldots, T-1$. Since η_t is non-increasing,

$$\mathbb{E}|\nabla\Phi(x_R)|^2 = \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}|\nabla\Phi(x_t)|^2 \quad \le \frac{1}{T\xi\eta_{T-1}}\sum_{t=0}^{T-1}\xi\eta_t\mathbb{E}|\nabla\Phi(x_t)|^2 \le \frac{C\big(1+\log(T+t_0)\big)}{T}\xi\eta_{T-1}. \tag{37}$$

Since t_0 is fixed and $\eta_{T-1} = \eta_0(T+t_0-1)^{-3/5}$, we have

$$\frac{1}{T\eta_{T-1}} = \mathcal{O}(T^{-2/5}).$$

Consequently,

$$\mathbb{E}|\nabla\Phi(x_R)|^2 \le C'\frac{1+\log T}{T^{2/5}} = \widetilde{\mathcal{O}}(T^{-2/5}), \tag{38}$$

where $C' > 0$ is independent of T. □

B.4 Proof of Theorem 2

Proof. Recall that the gradient of the smooth Tchebycheff objective satisfies

$$\nabla\Phi(x) = \sum_{i=1}^{m} p_i(x)w_i\nabla\phi_i(x), \tag{39}$$

where $p(x) \in \Delta_m$. Define

$$a(x) := \sum_{i=1}^{m} p_i(x)w_i, \qquad \omega_i(x) := \frac{p_i(x)w_i}{a(x)}. \tag{40}$$

Since $p_i(x)w_i \ge 0$, we have $\omega(x) \in \Delta_m$. Moreover, since $w_i \ge c_w/m$ and $\sum_i p_i(x) = 1$,

$$a(x) = \sum_{i=1}^{m} p_i(x)w_i \ge \frac{c_w}{m}. \tag{41}$$

Combining equation 39–equation 41 gives

$$\left|\sum_{i=1}^{m}\omega_i(x)\nabla\phi_i(x)\right|^2 = \frac{|\nabla\Phi(x)|^2}{a(x)^2} \quad \le \frac{m^2}{c_w^2}|\nabla\Phi(x)|^2. \tag{42}$$

Applying Corollary 1 at the random output x_R yields

$$\mathbb{E}\left|\sum_{i=1}^{m}\omega_i(x_R)\nabla\phi_i(x_R)\right|^2 \le \widetilde{\mathcal{O}}\left(m^2T^{-2/5}\right). \tag{43}$$

Thus, to obtain an expected ϵ-Pareto stationary point, it is sufficient to choose T such that $m^2T^{-2/5} \le \widetilde{\mathcal{O}}(\epsilon)$, which gives

$$T = \widetilde{\mathcal{O}}\left(m^5\epsilon^{-5/2}\right). \tag{44}$$

At each iteration, Algorithm 1 evaluates a constant number of stochastic function-value and first-order gradient oracles for each of the m objectives. Hence the oracle cost per iteration is $\mathcal{O}(m)$. Combining this with equation 44, the total oracle complexity is

$$\widetilde{\mathcal{O}}\,(mT) = \widetilde{\mathcal{O}}\left(m^6\epsilon^{-5/2}\right). \tag{45}$$

This proves the result. □

B.5 Proof of Corollary 2

Proof. Since $c_1 \le w_i \le c_2$ for all $i \in [m]$, we have that

$$c_1 m \le \sum_{i=1}^{m} w_i \le c_2 m. \tag{46}$$

Therefore, compared with the case $w \in \Delta_m$, the weighted stochastic-estimation and tracking recursions used in the proof of Theorem 1 introduce at most an additional factor of $\mathcal{O}(m)$ through the objective-wise noise and initialization terms. Meanwhile, since $p(x) \in \Delta_m$ and the weights w_i are uniformly bounded above and below, all aggregation-related constants remain independent of m. Consequently, the same Lyapunov argument as in Theorem 1 gives

$$\mathbb{E}\|\nabla\Phi(x_R)\|^2 = \widetilde{\mathcal{O}}\left(mT^{-2/5}\right). \tag{47}$$

It remains to relate the stationarity of Φ to Pareto stationarity. Recall that

$$\nabla\Phi(x) = \sum_{i=1}^{m} p_i(x) w_i \nabla\phi_i(x).$$

Define

$$a(x) := \sum_{i=1}^{m} p_i(x) w_i, \qquad \omega_i(x) := \frac{p_i(x) w_i}{a(x)}.$$

Since $p(x) \in \Delta_m$ and $c_1 \le w_i \le c_2$, we have $c_1 \le a(x) \le c_2$, and $\omega(x) \in \Delta_m$. Hence

$$\left\| \sum_{i=1}^{m} \omega_i(x_R) \nabla\phi_i(x_R) \right\|^2 = \frac{\|\nabla\Phi(x_R)\|^2}{a(x_R)^2} \le \frac{1}{c_1^2} \|\nabla\Phi(x_R)\|^2. \tag{48}$$

Combining equation 47 and equation 48, Algorithm 1 obtains an expected ϵ-Pareto stationary point whenever $mT^{-2/5} \le \epsilon$. Therefore,

$$T = \widetilde{\mathcal{O}}\left(m^{5/2}\epsilon^{-5/2}\right).$$

Since each iteration requires a constant number of stochastic first-order oracle calls for each of the m objectives, the oracle cost per iteration is $\mathcal{O}(m)$. Thus, the total oracle complexity is

$$\widetilde{\mathcal{O}}\left(m^{7/2}\epsilon^{-5/2}\right).$$

This completes the proof. □

C Proof of the Auxiliary Lemma

We first present the proof of Lemma 1.

C.1 Proof of lemma 1

For convenience, we denote $y_i^*(x) = z_i^*(x)$.

Proof. We first establish the approximation error in the function value. By the definition of $L_{i,\lambda}^*(x)$, we have

$$\begin{aligned} L_{i,\lambda}^*(x) &\le f_i(x, y_i^*(x)) + \lambda \left[g_i(x, y_i^*(x)) - g_i(x, y_i^*(x)) \right] \\ &= f_i(x, y_i^*(x)) = \phi_i(x). \end{aligned} \tag{49}$$

Since $f_i(x, \cdot)$ is $l_{f,0}$-Lipschitz, we have

$$f_i(x, y_{i,\lambda}^*(x)) - f_i(x, y_i^*(x)) \ge -l_{f,0} \|y_{i,\lambda}^*(x) - y_i^*(x)\|.$$

Moreover, since $g_i(x,\cdot)$ is μ_g-strongly convex and $\nabla_y g_i(x, y_i^*(x)) = 0$,

$$g_i(x, y_{i,\lambda}^*(x)) - g_i(x, y_i^*(x)) \geq \frac{\mu_g}{2}\|y_{i,\lambda}^*(x) - y_i^*(x)\|^2.$$

Let $d_\lambda := \|y_{i,\lambda}^*(x) - y_i^*(x)\|$, then we have

$$\begin{aligned} L_{i,\lambda}^*(x) - \phi_i(x) &= f_i(x, y_{i,\lambda}^*(x)) - f_i(x, y_i^*(x)) + \lambda\left[g_i(x, y_{i,\lambda}^*(x)) - g_i(x, y_i^*(x))\right] \\ &\geq -l_{f,0}d_\lambda + \frac{\lambda\mu_g}{2}d_\lambda^2 \\ &= \frac{\lambda\mu_g}{2}\left(d_\lambda - \frac{l_{f,0}}{\lambda\mu_g}\right)^2 - \frac{l_{f,0}^2}{2\lambda\mu_g} \\ &\geq -\frac{l_{f,0}^2}{2\lambda\mu_g}. \end{aligned} \tag{50}$$

Combining equation 49 and equation 50 gives

$$0 \leq \phi_i(x) - L_{i,\lambda}^*(x) \leq \frac{l_{f,0}^2}{2\lambda\mu_g}.$$

Therefore,

$$\left|L_{i,\lambda}^*(x) - \phi_i(x)\right| \leq \frac{C_1}{\lambda}, \qquad C_1 := \frac{l_{f,0}^2}{2\mu_g}.$$

We next establish the gradient approximation error. By the optimality conditions,

$$\nabla_y g_i(x, y_i^*(x)) = 0, \qquad \nabla_y f_i(x, y_{i,\lambda}^*(x)) + \lambda\nabla_y g_i(x, y_{i,\lambda}^*(x)) = 0.$$

Subtracting and taking the inner product with $y_{i,\lambda}^*(x) - y_i^*(x)$ yields

$$\langle\nabla_y f_i(x, y_{i,\lambda}^*(x)),\, y_{i,\lambda}^*(x) - y_i^*(x)\rangle = -\lambda\langle\nabla_y g_i(x, y_{i,\lambda}^*(x)) - \nabla_y g_i(x, y_i^*(x)),\, y_{i,\lambda}^*(x) - y_i^*(x)\rangle.$$

By μ_g-strong convexity of $g_i(x,\cdot)$, its gradient is μ_g-strongly monotone, hence

$$\langle\nabla_y g_i(x, y_{i,\lambda}^*(x)) - \nabla_y g_i(x, y_i^*(x)),\, y_{i,\lambda}^*(x) - y_i^*(x)\rangle \geq \mu_g\|y_{i,\lambda}^*(x) - y_i^*(x)\|^2.$$

Therefore,

$$\begin{aligned} \lambda\mu_g\|y_{i,\lambda}^*(x) - y_i^*(x)\|^2 &\leq -\langle\nabla_y f_i(x, y_{i,\lambda}^*(x)),\, y_{i,\lambda}^*(x) - y_i^*(x)\rangle \\ &\leq \|\nabla_y f_i(x, y_{i,\lambda}^*(x))\|\,\|y_{i,\lambda}^*(x) - y_i^*(x)\|. \end{aligned}$$

Since $f_i(x,\cdot)$ is $l_{f,0}$-Lipschitz and continuously differentiable, Thus,

$$\|y_{i,\lambda}^*(x) - y_i^*(x)\| \leq \frac{l_{f,0}}{\mu_g\lambda}. \tag{51}$$

The hypergradient of $\phi_i(x)$ is given by

$$\nabla\phi_i(x) = \nabla_x f_i(x, y_i^*(x)) - \nabla_{xy_i}^2 g_i(x, y_i^*(x))\left[\nabla_{y_iy_i}^2 g_i(x, y_i^*(x))\right]^{-1}\nabla_{y_i} f_i(x, y_i^*(x)).$$

By Danskin's theorem, together with the optimality of $y_{i,\lambda}^*(x)$ and $y_i^*(x)$, we have

$$\nabla L_{i,\lambda}^*(x) = \nabla_x f_i(x, y_{i,\lambda}^*(x)) + \lambda\left[\nabla_x g_i(x, y_{i,\lambda}^*(x)) - \nabla_x g_i(x, y_i^*(x))\right].$$

Therefore,

$$\begin{aligned} \nabla L_{i,\lambda}^*(x) - \nabla\phi_i(x) = {} & \underbrace{\nabla_x f_i(x, y_{i,\lambda}^*(x)) - \nabla_x f_i(x, y_i^*(x))}_{(i)} + \underbrace{\lambda\left[\nabla_x g_i(x, y_{i,\lambda}^*(x)) - \nabla_x g_i(x, y_i^*(x))\right]}_{(ii)} \\ & + \underbrace{\nabla_{xy_i}^2 g_i(x, y_i^*(x))\left[\nabla_{y_iy_i}^2 g_i(x, y_i^*(x))\right]^{-1}\nabla_{y_i} f_i(x, y_i^*(x))}_{(iii)}. \end{aligned} \tag{52}$$

For term (i), the $l_{f,1}$-smoothness of f_i and equation 51 imply

$$\|(i)\| \leq l_{f,1}\|y^*_{i,\lambda}(x) - y^*_i(x)\| \leq \frac{l_{f,1}l_{f,0}}{\mu_g \lambda}. \tag{53}$$

We next show that the leading-order component of (ii) cancels with (iii). Let

$$d := y^*_{i,\lambda}(x) - y^*_i(x), \qquad H := \nabla^2_{y_i y_i} g_i(x, y^*_i(x)), \qquad B := \nabla^2_{x y_i} g_i(x, y^*_i(x)).$$

By $\nabla_{y_i} g_i(x, y^*_i(x)) = 0$ and the Lipschitz continuity of the Hessian of g_i, it follows that

$$\nabla_y g_i(x, y^*_{i,\lambda}(x)) = Hd + r_g, \qquad \|r_g\| \leq \frac{l_{g,2}}{2}\|d\|^2. \tag{54}$$

By the $l_{f,1}$-smoothness of f_i,

$$\nabla_y f_i(x, y^*_{i,\lambda}(x)) = \nabla_y f_i(x, y^*_i(x)) + r_f, \qquad \|r_f\| \leq l_{f,1}\|d\|. \tag{55}$$

Substituting these expansions into

$$\nabla_y f_i(x, y^*_{i,\lambda}(x)) + \lambda \nabla_y g_i(x, y^*_{i,\lambda}(x)) = 0$$

yields

$$\lambda H d = -\nabla_y f_i(x, y^*_i(x)) - r_f - \lambda r_g.$$

Since $g_i(x, \cdot)$ is μ_g-strongly convex,

$$H \succeq \mu_g I, \qquad \|H^{-1}\| \leq \frac{1}{\mu_g}.$$

Then we have

$$\begin{aligned}
\left\|\lambda d + H^{-1}\nabla_y f_i(x, y^*_i(x))\right\| &\leq \frac{1}{\mu_g}\left(\|r_f\| + \lambda\|r_g\|\right) \\
&\leq \frac{1}{\mu_g}\left(l_{f,1}\frac{l_{f,0}}{\lambda\mu_g} + \frac{l_{g,2}l^2_{f,0}}{2\lambda\mu^2_g}\right) \\
&= \frac{1}{\lambda}\left(\frac{l_{f,1}l_{f,0}}{\mu^2_g} + \frac{l_{g,2}l^2_{f,0}}{2\mu^3_g}\right).
\end{aligned} \tag{56}$$

Similarly, Taylor expansion of $\nabla_x g_i$ with respect to y gives

$$\nabla_x g_i(x, y^*_{i,\lambda}(x)) - \nabla_x g_i(x, y^*_i(x)) = Bd + r_x, \qquad \|r_x\| \leq \frac{l_{g,2}}{2}\|d\|^2 \leq \frac{l_{g,2}l^2_{f,0}}{2\lambda^2\mu^2_g}.$$

Therefore,

$$(ii) + (iii) = B\left(\lambda d + H^{-1}\nabla_y f_i(x, y^*_i(x))\right) + \lambda r_x.$$

Since g_i is $l_{g,1}$-smooth and twice continuously differentiable, it holds that $\|B\| \leq l_{g,1}$. Together with equation 56, we obtain

$$\|(ii) + (iii)\| \leq \frac{1}{\lambda}\left[l_{g,1}\left(\frac{l_{f,1}l_{f,0}}{\mu^2_g} + \frac{l_{g,2}l^2_{f,0}}{2\mu^3_g}\right) + \frac{l_{g,2}l^2_{f,0}}{2\mu^2_g}\right]. \tag{57}$$

Combining equation 53 and equation 57 yields

$$\left\|\nabla L^*_{i,\lambda}(x) - \nabla\phi_i(x)\right\| \leq \frac{C_2}{\lambda},$$

where

$$C_2 = \frac{l_{f,1}l_{f,0}}{\mu_g} + \frac{l_{g,1}l_{f,1}l_{f,0}}{\mu^2_g} + \frac{l_{g,1}l_{g,2}l^2_{f,0}}{2\mu^3_g} + \frac{l_{g,2}l^2_{f,0}}{2\mu^2_g}.$$

This completes the proof. □

C.2 Proof of Lemma 2

Proof. By the L_Φ-smoothness of Φ,

$$\Phi(x_{t+1}) \le \Phi(x_t) + \langle \nabla\Phi(x_t), x_{t+1} - x_t\rangle + \frac{L_\Phi}{2}\|x_{t+1} - x_t\|^2. \tag{58}$$

For convenience of the following derivation, let $g_t := \nabla\Phi(x_t)$. By the update formula $x_{t+1} = x_t - \xi\eta_t\,(q_x^t + e_x^t)$, we have

$$\begin{aligned}\Phi(x_{t+1}) - \Phi(x_t) \le & -\xi\eta_t\langle g_t, q_x^t + e_x^t\rangle + \frac{L_\Phi\xi^2\eta_t^2}{2}\|q_x^t + e_x^t\|^2 \\ \le & -\xi\eta_t\langle g_t, e_x^t\rangle - \frac{\xi\eta_t}{2}\left(\|g_t\|^2 + \|q_x^t\|^2 - \|g_t - q_x^t\|^2\right) \\ & + L_\Phi\xi^2\eta_t^2\left(\|e_x^t\|^2 + \|q_x^t\|^2\right).\end{aligned} \tag{59}$$

Using Young's inequality,

$$-\langle g_t, e_x^t\rangle \le \frac{1}{4}\|g_t\|^2 + \|e_x^t\|^2.$$

Hence, since $L_\Phi\xi\eta_t \le 1/4$, we have

$$\Phi(x_{t+1}) - \Phi(x_t) \le -\frac{\xi\eta_t}{4}\|g_t\|^2 - \frac{\xi\eta_t}{4}\|q_x^t\|^2 + 2\xi\eta_t\|e_x^t\|^2 + \frac{\xi\eta_t}{2}\|g_t - q_x^t\|^2. \tag{60}$$

It remains to control $\|g_t - q_x^t\|^2$. Define the exact current-iterate weights

$$p_i^t = \frac{\exp(w_i(L_i^t - r_i)/\tau)}{\sum_{j=1}^m \exp\big(w_j(L_j^t - r_j)/\tau\big)},$$

and the intermediate exact-penalty weights

$$\bar{p}_i^t = \frac{\exp\Big(w_i(L^*_{i,\lambda_t}(x_t) - r_i)/\tau\Big)}{\sum_{j=1}^m \exp\Big(w_j(L^*_{j,\lambda_t}(x_t) - r_j)/\tau\Big)}.$$

By the definition of q_x^t and triangle inequality, we have

$$\begin{aligned}\|g_t - q_x^t\| \le & \|\sum_{i=1}^m \hat{p}_i^t w_i[\nabla\phi_i(x_t) - q_{i,x}^t]\| + \|\sum_{i=1}^m (p_{i,*}^t - \hat{p}_i^t) w_i\nabla\phi_i(x_t)\| \\ \le & \underbrace{\|\sum_{i=1}^m \hat{p}_i^t w_i(\nabla\phi_i(x_t) - \nabla L^*_{i,\lambda}(x_t))\|}_{(i)} + \underbrace{\sum_{i=1}^m \|\hat{p}_i^t w_i(\nabla L^*_{i,\lambda}(x_t) - \nabla L_{i,\lambda}(x_t))\|}_{(ii)} \\ & + \underbrace{\|\sum_{i=1}^m (p_{i,*}^t - \hat{p}_i^t) w_i\nabla\phi_i(x_t)\|}_{(iii)}\end{aligned} \tag{61}$$

For (i), Lemma 1 gives

$$\left\|\nabla\phi_i(x_t) - \nabla L^*_{i,\lambda_t}(x_t)\right\| \le \frac{C_2}{\lambda_t}.$$

Since $\widehat{p}^t \in \Delta_m$ and $w \in \Delta_m$, then we have $\sum_{i=1}^m \widehat{p}_i^t w_i \le 1$. Therefore,

$$(i) \le \frac{C_2}{\lambda_t}. \tag{62}$$

For (ii), by the joint smoothness of f_i and g_i,

$$\left\|\nabla L^*_{i,\lambda_t}(x_t) - q_{i,x}^t\right\| \le \bar{L}_t\left\|y_i^t - y^*_{i,\lambda_t}(x_t)\right\| + \lambda_t l_{g,1}\left\|z_i^t - y_i^*(x_t)\right\|, \tag{63}$$

where $\bar{L}_t = l_{f,1} + \lambda_t l_{g,1}$. Using $\widehat{p}_i^t \le 1$, we obtain

$$(ii) \le \sum_{i=1}^{m} w_i \left[\bar{L}_t \left\| y_i^t - y_{i,\lambda_t}^*(x_t) \right\| + \lambda_t l_{g,1} \left\| z_i^t - y_i^*(x_t) \right\| \right].$$

Since $\sum_i w_i = 1$, weighted Cauchy–Schwarz gives

$$(ii)^2 \le 2\bar{L}_t^2 \sum_{i=1}^{m} w_i \left\| y_i^t - y_{i,\lambda_t}^*(x_t) \right\|^2 + 2\lambda_t^2 l_{g,1}^2 \sum_{i=1}^{m} w_i \left\| z_i^t - y_i^*(x_t) \right\|^2. \tag{64}$$

Since $\lambda_t \ge 1$, we have $\bar{L}_t = l_{f,1} + \lambda_t l_{g,1} \le (l_{f,1} + l_{g,1})\lambda_t$. Thus there exists a constant $C_q > 0$ such that

$$(ii)^2 \le C_q \lambda_t^2 \left[\sum_{i=1}^{m} w_i \left\| y_i^t - y_{i,\lambda_t}^*(x_t) \right\|^2 + \sum_{i=1}^{m} w_i \left\| z_i^t - y_i^*(x_t) \right\|^2 \right]. \tag{65}$$

We next bound (iii). By Assumption 1, $\|\nabla \phi_i(x_t)\| \le C_\phi$, and since $w_i \le 1$,

$$(iii) \le C_\phi \|p_*^t - \widehat{p}^t\|_1. \tag{66}$$

Using the intermediate weights,

$$p_*^t - \widehat{p}^t = (p_*^t - \bar{p}^t) + (\bar{p}^t - p^t) + (p^t - \widehat{p}^t),$$

and hence

$$\|p_*^t - \widehat{p}^t\|_1^2 \le 3\|p_*^t - \bar{p}^t\|_1^2 + 3\|\bar{p}^t - p^t\|_1^2 + 3\|p^t - \widehat{p}^t\|_1^2. \tag{67}$$

For the first term, the softmax map with temperature τ is $1/(2\tau)$-Lipschitz from ℓ_1 to ℓ_1. Therefore,

$$\|p_*^t - \bar{p}^t\|_1 \le \frac{1}{2\tau} \sum_{i=1}^{m} w_i \left| \phi_i(x_t) - L_{i,\lambda_t}^*(x_t) \right| \le \frac{C_1}{2\tau\lambda_t}, \tag{68}$$

where Lemma 1 and $\sum_i w_i = 1$ were used. Hence

$$\|p_*^t - \bar{p}^t\|_1^2 \le \frac{C_1^2}{4\tau^2\lambda_t^2}. \tag{69}$$

For the second term, recall

$$L_i^t - L_{i,\lambda_t}^*(x_t) = F_{i,t}(y_i^t) - F_{i,t}(y_{i,\lambda_t}^*(x_t)) - \lambda_t \left[g_i(x_t, z_i^t) - g_i(x_t, y_i^*(x_t)) \right],$$

where

$$F_{i,t}(y) = f_i(x_t, y) + \lambda_t g_i(x_t, y).$$

By smoothness and the optimality of $y_{i,\lambda_t}^*(x_t)$ and $y_i^*(x_t)$,

$$|L_i^t - L_{i,\lambda_t}^*(x_t)| \le \frac{\bar{L}_t}{2} \left\| y_i^t - y_{i,\lambda_t}^*(x_t) \right\|^2 + \frac{\lambda_t l_{g,1}}{2} \left\| z_i^t - y_i^*(x_t) \right\|^2. \tag{70}$$

Thus,

$$\|\bar{p}^t - p^t\|_1 \le \frac{\bar{L}_t}{4\tau} \sum_{i=1}^{m} w_i \left\| y_i^t - y_{i,\lambda_t}^*(x_t) \right\|^2 + \frac{\lambda_t l_{g,1}}{4\tau} \sum_{i=1}^{m} w_i \left\| z_i^t - y_i^*(x_t) \right\|^2. \tag{71}$$

Since both $\bar{p}^t$ and p^t belong to the probability simplex, we have $\|\bar{p}^t - p^t\|_1 \le 2$, and hence $\|\bar{p}^t - p^t\|_1^2 \le 2\|\bar{p}^t - p^t\|_1$. Using $\lambda_t \ge 1$, there exists $C_{\tau,1} > 0$ such that

$$\|\bar{p}^t - p^t\|_1^2 \le C_{\tau,1} \lambda_t^2 \left[\sum_{i=1}^{m} w_i \left\| y_i^t - y_{i,\lambda_t}^*(x_t) \right\|^2 + \sum_{i=1}^{m} w_i \left\| z_i^t - y_i^*(x_t) \right\|^2 \right]. \tag{72}$$

It remains to bound the new term $\|p^t - \widehat{p}^t\|_1^2$. Again by the Lipschitz continuity of softmax,

$$\|p^t - \widehat{p}^t\|_1 \le \frac{1}{2\tau} \sum_{i=1}^{m} w_i |\widehat{L}_i^t - L_i^t|. \tag{73}$$

Since $\sum_i w_i = 1$, we obtain

$$\|p^t - \widehat{p}^t\|_1^2 \le \frac{1}{4\tau^2} \sum_{i=1}^m w_i |\widehat{L}_i^t - L_i^t|^2. \tag{74}$$

By $\widehat{L}_i^t - L_i^t = e_{i,fv}^t + \lambda_t e_{i,gy}^t - \lambda_t e_{i,gz}^t$, we obtain, $|\widehat{L}_i^t - L_i^t|^2 \le 3|e_{i,fv}^t|^2 + 3\lambda_t^2 |e_{i,gy}^t|^2 + 3\lambda_t^2 |e_{i,gz}^t|^2$. Substituting into equation 74 and taking expectation gives

$$\mathbb{E}\|p^t - \widehat{p}^t\|_1^2 \le \frac{3}{4\tau^2} \left(E_{fv,t} + \lambda_t^2 E_{gy,t} + \lambda_t^2 E_{gz,t} \right). \tag{75}$$

Combining equation 67, equation 69, equation 72, and equation 75, and then using equation 66, yields

$$\begin{aligned} \mathbb{E}[(iii)^2] \le C_{\tau,2} \Bigg[& \lambda_t^2 \sum_{i=1}^m w_i \mathbb{E} \left\| y_i^t - y_{i,\lambda_t}^*(x_t) \right\|^2 \\ & + \lambda_t^2 \sum_{i=1}^m w_i \mathbb{E} \left\| z_i^t - y_i^*(x_t) \right\|^2 + \frac{1}{\lambda_t^2} + E_{fv,t} + \lambda_t^2 E_{gy,t} + \lambda_t^2 E_{gz,t} \Bigg]. \end{aligned} \tag{76}$$

Finally, from

$$\|g_t - q_x^t\|^2 \le 3I_t^2 + 3(ii)^2 + 3(iii)^2,$$

together with equation 62, equation 65, and equation 76, there exist constants $C_\Phi, C_v > 0$ such that

$$\begin{aligned} \mathbb{E}\|g_t - q_x^t\|^2 \le C_\Phi \Bigg[& \lambda_t^2 \sum_{i=1}^m w_i \mathbb{E} \left\| y_i^t - y_{i,\lambda_t}^*(x_t) \right\|^2 \\ & + \lambda_t^2 \sum_{i=1}^m w_i \mathbb{E} \left\| z_i^t - y_i^*(x_t) \right\|^2 + \frac{1}{\lambda_t^2} \Bigg] + C_v \left(E_{fv,t} + \lambda_t^2 E_{gy,t} + \lambda_t^2 E_{gz,t} \right). \end{aligned} \tag{77}$$

Substituting equation 77 into equation 60 proves equation 10. □

C.3 PROOF OF LEMMA 3

Proof. Let $e_t := \hat{h}_t - q_t$, and define the centered stochastic errors evaluated using the same fresh sample ω_{t+1} as

$$\varepsilon_{t+1} := h_{t+1}(\omega_{t+1}) - q_{t+1}, \qquad \varepsilon_t := h_t(\omega_{t+1}) - q_t.$$

By unbiasedness, we have $\mathbb{E}[\varepsilon_{t+1} \mid \mathcal{F}_t] = \mathbb{E}[\varepsilon_t \mid \mathcal{F}_t] = 0$, where $\mathcal{F}_t$ denotes the history up to iteration t. Subtracting q_{t+1} from the recursive update gives $e_{t+1} = (1 - \alpha_{t+1})e_t + \varepsilon_{t+1} - (1 - \alpha_{t+1})\varepsilon_t$. Therefore, conditioning on $\mathcal{F}_t$ and using the conditional zero-mean property,

$$\begin{aligned} \mathbb{E}[\|e_{t+1}\|^2 \mid \mathcal{F}_t] = & (1 - \alpha_{t+1})^2 \|e_t\|^2 \\ & + \mathbb{E}\left[\|\varepsilon_{t+1} - (1 - \alpha_{t+1})\varepsilon_t\|^2 \mid \mathcal{F}_t \right]. \end{aligned} \tag{78}$$

Next, observe that

$$\varepsilon_{t+1} - (1 - \alpha_{t+1})\varepsilon_t = (1 - \alpha_{t+1})(\varepsilon_{t+1} - \varepsilon_t) + \alpha_{t+1}\varepsilon_{t+1}.$$

Using $\|a + b\|^2 \le 2\|a\|^2 + 2\|b\|^2$ and $(1 - \alpha_{t+1})^2 \le 1$, we obtain

$$\begin{aligned} & \mathbb{E}\left[\|\varepsilon_{t+1} - (1 - \alpha_{t+1})\varepsilon_t\|^2 \mid \mathcal{F}_t \right] \\ & \quad \le 2\mathbb{E}\left[\|\varepsilon_{t+1} - \varepsilon_t\|^2 \mid \mathcal{F}_t \right] + 2\alpha_{t+1}^2 \mathbb{E}\left[\|\varepsilon_{t+1}\|^2 \mid \mathcal{F}_t \right] \\ & \quad \le 2\mathbb{E}\left[\left\| \big(h_{t+1}(\omega_{t+1}) - q_{t+1}\big) - \big(h_t(\omega_{t+1}) - q_t\big) \right\|^2 \mid \mathcal{F}_t \right] + 2\alpha_{t+1}^2 \delta^2. \end{aligned} \tag{79}$$

Substituting this bound into equation 78 and taking total expectation yields

$$\begin{aligned} \mathbb{E}\|\hat{h}_{t+1} - q_{t+1}\|^2 \le & (1 - \alpha_{t+1})^2 \mathbb{E}\|\hat{h}_t - q_t\|^2 \\ & + 2\mathbb{E} \left\| \big(h_{t+1}(\omega_{t+1}) - q_{t+1}\big) - \big(h_t(\omega_{t+1}) - q_t\big) \right\|^2 + 2\alpha_{t+1}^2 \delta^2. \end{aligned} \tag{80}$$

This completes the proof. □

C.4 Proof of Lemma 4

Proof. Let $q_t = q_{i,z}^t, h_t(\omega) = \nabla_z g_i(x_t, z_i^t; \omega)$. Since the stochastic gradient variance is bounded by δ_g^2, by Lemma 3 we have

$$\mathbb{E}\|e_{i,z}^{t+1}\|^2 \le (1-\alpha_i^{t+1})^2 \mathbb{E}\|e_{i,z}^t\|^2 + 2\mathbb{E}\left\|\varepsilon_{i,z}^{t+1} - \varepsilon_{i,z}^t\right\|^2 + 2(\alpha_i^{t+1})^2\delta_g^2, \tag{81}$$

where

$$\varepsilon_{i,z}^t := \nabla_z g_i(x_t, z_i^t; \omega_{t+1}) - q_{i,z}^t.$$

By the conditional variance inequality and the mean-square smoothness of the stochastic gradient,

$$\mathbb{E}\left\|\varepsilon_{i,z}^{t+1} - \varepsilon_{i,z}^t\right\|^2 \le l_{g,1}^2 \mathbb{E}\left[\|x_{t+1} - x_t\|^2 + \|z_i^{t+1} - z_i^t\|^2\right]. \tag{82}$$

Using

$$x_{t+1} - x_t = -\xi\eta_t(e_x^t + q_x^t), \qquad z_i^{t+1} - z_i^t = -\beta_t(e_{i,z}^t + q_{i,z}^t),$$

we obtain

$$\mathbb{E}\|x_{t+1} - x_t\|^2 \le 2\xi^2\eta_t^2\left(\mathbb{E}\|e_x^t\|^2 + \mathbb{E}\|q_x^t\|^2\right),$$

and

$$\mathbb{E}\|z_i^{t+1} - z_i^t\|^2 \le 2\beta_t^2\left(\mathbb{E}\|e_{i,z}^t\|^2 + \mathbb{E}\|q_{i,z}^t\|^2\right).$$

Substituting the above bounds into equation 81 yields

$$\begin{aligned}\mathbb{E}\|e_{i,z}^{t+1}\|^2 \le\ & \left[(1-\alpha_i^{t+1})^2 + 4l_{g,1}^2\beta_t^2\right]\mathbb{E}\|e_{i,z}^t\|^2 \\ & + 4l_{g,1}^2\xi^2\eta_t^2\mathbb{E}\|e_x^t\|^2 + 4l_{g,1}^2\beta_t^2\mathbb{E}\|q_{i,z}^t\|^2 \\ & + 4l_{g,1}^2\xi^2\eta_t^2\mathbb{E}\|q_x^t\|^2 + 2(\alpha_i^{t+1})^2\delta_g^2.\end{aligned} \tag{83}$$

This proves the result. □

C.5 Proof of Lemma 5

Proof. Let $q_t = q_{i,y}^t$, $h_t(\omega) = \nabla_y f_i(x_t, y_i^t; \omega) + \lambda_t \nabla_y g_i(x_t, y_i^t; \omega)$. , $\zeta_f^t := \nabla_y f_i(x_t, y_i^t; \omega_{t+1}) - \nabla_y f_i(x_t, y_i^t)$, , $\zeta_g^t := \nabla_y g_i(x_t, y_i^t; \omega_{t+1}) - \nabla_y g_i(x_t, y_i^t)$. Then the centered stochastic error is

$$\varepsilon_{i,y}^t = \zeta_f^t + \lambda_t\zeta_g^t.$$

Moreover,

$$\varepsilon_{i,y}^{t+1} - \varepsilon_{i,y}^t = (\zeta_f^{t+1} - \zeta_f^t) + \lambda_{t+1}(\zeta_g^{t+1} - \zeta_g^t) + \Delta\lambda_t\,\zeta_g^t.$$

Hence, by mean-square smoothness and $\|a+b+c\|^2 \le 3(\|a\|^2 + \|b\|^2 + \|c\|^2)$,

$$\begin{aligned}\mathbb{E}\|\varepsilon_{i,y}^{t+1} - \varepsilon_{i,y}^t\|^2 \le\ & 3\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\mathbb{E}\left[\|x_{t+1} - x_t\|^2 + \|y_i^{t+1} - y_i^t\|^2\right] \\ & + 3(\Delta\lambda_t)^2\delta_g^2.\end{aligned} \tag{84}$$

Furthermore,

$$\mathbb{E}\|\varepsilon_{i,y}^{t+1}\|^2 \le 2\delta_f^2 + 2\lambda_{t+1}^2\delta_g^2.$$

Therefore, Lemma 3 gives

$$\begin{aligned}\mathbb{E}\|e_{i,y}^{t+1}\|^2 \le\ & (1-\alpha_i^{t+1})^2\mathbb{E}\|e_{i,y}^t\|^2 \\ & + 6\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\mathbb{E}\left[\|x_{t+1} - x_t\|^2 + \|y_i^{t+1} - y_i^t\|^2\right] \\ & + 4(\alpha_i^{t+1})^2\left(\delta_f^2 + \lambda_{t+1}^2\delta_g^2\right) + 6(\Delta\lambda_t)^2\delta_g^2.\end{aligned} \tag{85}$$

Using

$$x_{t+1} - x_t = -\xi\eta_t(e_x^t + q_x^t), \qquad y_i^{t+1} - y_i^t = -\eta_t(e_{i,y}^t + q_{i,y}^t),$$

and $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$ in equation 85 gives

$$\begin{aligned}\mathbb{E}\|e_{i,y}^{t+1}\|^2 \le\ & \left[(1-\alpha_i^{t+1})^2 + 12\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\eta_t^2\right]\mathbb{E}\|e_{i,y}^t\|^2 \\ & + 12\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\xi^2\eta_t^2\mathbb{E}\|e_x^t\|^2 \\ & + 12\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\eta_t^2\mathbb{E}\|q_{i,y}^t\|^2 \\ & + 12\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\xi^2\eta_t^2\mathbb{E}\|q_x^t\|^2 \\ & + 4(\alpha_i^{t+1})^2\left(\delta_f^2 + \lambda_{t+1}^2\delta_g^2\right) + 6(\Delta\lambda_t)^2\delta_g^2.\end{aligned} \tag{86}$$

This proves the result. □

C.6 Proof of Lemma 6

Proof. Let $q_t = q_{i,x}^t$ and $h_t(\omega) = \nabla_x f_i(x_t, y_i^t; \omega) + \lambda_t \left[\nabla_x g_i(x_t, y_i^t; \omega) - \nabla_x g_i(x_t, z_i^t; \omega)\right]$. Define the centered noises

$$\zeta_f^t = \nabla_x f_i(x_t, y_i^t; \omega_{t+1}) - \nabla_x f_i(x_t, y_i^t), \qquad \zeta_{g,y}^t = \nabla_x g_i(x_t, y_i^t; \omega_{t+1}) - \nabla_x g_i(x_t, y_i^t),$$

and

$$\zeta_{g,z}^t = \nabla_x g_i(x_t, z_i^t; \omega_{t+1}) - \nabla_x g_i(x_t, z_i^t).$$

Then

$$\varepsilon_{i,x}^t = \zeta_f^t + \lambda_t \left(\zeta_{g,y}^t - \zeta_{g,z}^t\right),$$

and

$$\begin{aligned}\varepsilon_{i,x}^{t+1} - \varepsilon_{i,x}^t = (\zeta_f^{t+1} - \zeta_f^t) + \lambda_{t+1}(\zeta_{g,y}^{t+1} - \zeta_{g,y}^t) \\ - \lambda_{t+1}(\zeta_{g,z}^{t+1} - \zeta_{g,z}^t) + \Delta\lambda_t(\zeta_{g,y}^t - \zeta_{g,z}^t).\end{aligned} \tag{87}$$

By mean-square smoothness and $\|a+b+c+d\|^2 \le 4(\|a\|^2 + \|b\|^2 + \|c\|^2 + \|d\|^2)$,

$$\begin{aligned}\mathbb{E}\|\varepsilon_{i,x}^{t+1} - \varepsilon_{i,x}^t\|^2 \le 4\left(l_{f,1}^2 + 2\lambda_{t+1}^2 l_{g,1}^2\right)\mathbb{E}\|x_{t+1} - x_t\|^2 \\ + 4\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\mathbb{E}\|y_i^{t+1} - y_i^t\|^2 \\ + 4\lambda_{t+1}^2 l_{g,1}^2 \mathbb{E}\|z_i^{t+1} - z_i^t\|^2 + 16(\Delta\lambda_t)^2\delta_g^2.\end{aligned} \tag{88}$$

At iteration $t+1$,

$$\mathbb{E}\|\varepsilon_{i,x}^{t+1}\|^2 \le 3\delta_f^2 + 6\lambda_{t+1}^2\delta_g^2.$$

Applying Lemma 3 therefore gives

$$\begin{aligned}\mathbb{E}\|e_{i,x}^{t+1}\|^2 \le (1-\alpha_i^{t+1})^2\mathbb{E}\|e_{i,x}^t\|^2 \\ + 8\left(l_{f,1}^2 + 2\lambda_{t+1}^2 l_{g,1}^2\right)\mathbb{E}\|x_{t+1} - x_t\|^2 \\ + 8\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\mathbb{E}\|y_i^{t+1} - y_i^t\|^2 \\ + 8\lambda_{t+1}^2 l_{g,1}^2 \mathbb{E}\|z_i^{t+1} - z_i^t\|^2 \\ + 6(\alpha_i^{t+1})^2\left(\delta_f^2 + 2\lambda_{t+1}^2\delta_g^2\right) + 32(\Delta\lambda_t)^2\delta_g^2.\end{aligned} \tag{89}$$

Using

$$x_{t+1} - x_t = -\xi\eta_t(e_x^t + q_x^t), \qquad y_i^{t+1} - y_i^t = -\eta_t(e_{i,y}^t + q_{i,y}^t),$$

and

$$z_i^{t+1} - z_i^t = -\beta_t(e_{i,z}^t + q_{i,z}^t),$$

in equation 89, together with $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$, gives

$$\begin{aligned}\mathbb{E}\|e_{i,x}^{t+1}\|^2 \le (1-\alpha_i^{t+1})^2\mathbb{E}\|e_{i,x}^t\|^2 \\ + 16\left(l_{f,1}^2 + 2\lambda_{t+1}^2 l_{g,1}^2\right)\xi^2\eta_t^2\left(\mathbb{E}\|e_x^t\|^2 + \mathbb{E}\|q_x^t\|^2\right) \\ + 16\left(l_{f,1}^2 + \lambda_{t+1}^2 l_{g,1}^2\right)\eta_t^2\left(\mathbb{E}\|e_{i,y}^t\|^2 + \mathbb{E}\|q_{i,y}^t\|^2\right) \\ + 16\lambda_{t+1}^2 l_{g,1}^2\beta_t^2\left(\mathbb{E}\|e_{i,z}^t\|^2 + \mathbb{E}\|q_{i,z}^t\|^2\right) \\ + 6(\alpha_i^{t+1})^2\left(\delta_f^2 + 2\lambda_{t+1}^2\delta_g^2\right) + 32(\Delta\lambda_t)^2\delta_g^2.\end{aligned} \tag{90}$$

Finally, we have

$$e_x^{t+1} = \sum_{i=1}^m \widehat{p}_i^{t+1} w_i e_{i,x}^{t+1}.$$

Since $\widehat{p}^{t+1}, w \in \Delta_m$, weighted Cauchy–Schwarz yields pathwise

$$\|e_x^{t+1}\|^2 \le \left(\sum_{i=1}^m \widehat{p}_i^{t+1} w_i\right)\sum_{i=1}^m \widehat{p}_i^{t+1} w_i \|e_{i,x}^{t+1}\|^2 \le \sum_{i=1}^m w_i \|e_{i,x}^{t+1}\|^2.$$

Taking expectation proves the aggregate bound. □

C.7 Proof of Lemma 7

Proof. We apply Lemma 3 to the three function-value estimators separately. For $\widehat{F}_i^t$, Let $q_t = F_i^t = f_i(x_t, y_i^t), h_t(\omega) = f_i(x_t, y_i^t; \omega)$. By the uniform variance bound on the stochastic gradients in Assumption 3, the centered function-value noise is mean-square Lipschitz. Specifically, letting $\varepsilon_f(u;\omega) := f_i(u;\omega) - f_i(u)$, we have $\mathbb{E}\left|\varepsilon_f(u;\omega) - \varepsilon_f(v;\omega)\right|^2 \le \delta_f^2\|u-v\|^2$. Therefore,

$$\mathbb{E}\left|\left(h_{t+1}(\omega) - F_i^{t+1}\right) - \left(h_t(\omega) - F_i^t\right)\right|^2 \le \delta_f^2\mathbb{E}\left[\|x_{t+1}-x_t\|^2 + \|y_i^{t+1}-y_i^t\|^2\right].$$

Since the function-value variance is bounded by ν_f^2, Lemma 3 and the update formulas give

$$\begin{aligned}\mathbb{E}|e_{i,fv}^{t+1}|^2 \le (1-\alpha_{t+1})^2\mathbb{E}|e_{i,fv}^t|^2 + 4l_{f,1}^2\xi^2\eta_t^2\left(\mathbb{E}\|e_x^t\|^2 + \mathbb{E}\|q_x^t\|^2\right)\\ + 4l_{f,1}^2\eta_t^2\left(\mathbb{E}\|e_{i,y}^t\|^2 + \mathbb{E}\|q_{i,y}^t\|^2\right) + 2\alpha_{t+1}^2\nu_f^2. \end{aligned} \tag{91}$$

For $\widehat{G}_{i,y}^t$, let $q_t = G_{i,y}^t = g_i(x_t, y_i^t), h_t(\omega) = g_i(x_t, y_i^t; \omega)$. The same argument, using the constant $l_{g,1}$ and variance bound ν_g^2, yields

$$\begin{aligned}\mathbb{E}|e_{i,gy}^{t+1}|^2 \le (1-\alpha_{t+1})^2\mathbb{E}|e_{i,gy}^t|^2 + 4l_{g,1}^2\xi^2\eta_t^2\left(\mathbb{E}\|e_x^t\|^2 + \mathbb{E}\|q_x^t\|^2\right)\\ + 4l_{g,1}^2\eta_t^2\left(\mathbb{E}\|e_{i,y}^t\|^2 + \mathbb{E}\|q_{i,y}^t\|^2\right) + 2\alpha_{t+1}^2\nu_g^2. \end{aligned} \tag{92}$$

Finally, for $\widehat{G}_{i,z}^t$, let $q_t = G_{i,z}^t = g_i(x_t, z_i^t), h_t(\omega) = g_i(x_t, z_i^t; \omega)$. Then we have

$$\mathbb{E}\left|\left(h_{t+1}(\omega) - G_{i,z}^{t+1}\right) - \left(h_t(\omega) - G_{i,z}^t\right)\right|^2 \le l_{g,1}^2\mathbb{E}\left[\|x_{t+1}-x_t\|^2 + \|z_i^{t+1}-z_i^t\|^2\right].$$

Using the updates of x_t and z_i^t gives

$$\begin{aligned}\mathbb{E}|e_{i,gz}^{t+1}|^2 \le (1-\alpha_{t+1})^2\mathbb{E}|e_{i,gz}^t|^2 + 4l_{g,1}^2\xi^2\eta_t^2\left(\mathbb{E}\|e_x^t\|^2 + \mathbb{E}\|q_x^t\|^2\right)\\ + 4l_{g,1}^2\beta_t^2\left(\mathbb{E}\|e_{i,z}^t\|^2 + \mathbb{E}\|q_{i,z}^t\|^2\right) + 2\alpha_{t+1}^2\nu_g^2. \end{aligned} \tag{93}$$

This completes the proof. □

C.8 Proof of Lemma 8

Proof. Let $F_{i,t}(x,y) := f_i(x,y) + \lambda_t g_i(x,y)$. Since f_i is $l_{f,1}$-smooth, for any y, y',

$$\langle \nabla_y f_i(x,y) - \nabla_y f_i(x,y'), y - y'\rangle \ge -l_{f,1}\|y-y'\|^2.$$

Since $g_i(x,\cdot)$ is μ_g-strongly convex,

$$\langle \nabla_y g_i(x,y) - \nabla_y g_i(x,y'), y - y'\rangle \ge \mu_g\|y-y'\|^2.$$

Therefore,

$$\langle \nabla_y F_{i,t}(x,y) - \nabla_y F_{i,t}(x,y'), y - y'\rangle \ge (\lambda_t\mu_g - l_{f,1})\|y-y'\|^2 = \bar{\mu}_t\|y-y'\|^2. \tag{94}$$

Hence $F_{i,t}(x,\cdot)$ is $\bar{\mu}_t$-strongly convex. Moreover, by the smoothness of f_i and g_i,

$$\|\nabla_y F_{i,t}(x,y) - \nabla_y F_{i,t}(x,y')\| \le \bar{L}_t\|y-y'\|. \tag{95}$$

We first establish the Lipschitz continuity of $y_{i,\lambda_t}^*(x)$ with respect to x. Let $u = y_{i,\lambda_t}^*(x), v = y_{i,\lambda_t}^*(x')$. By optimality, we have $\nabla_y F_{i,t}(x,u) = 0, \nabla_y F_{i,t}(x',v) = 0$. Using equation 94,

$$\begin{aligned}\bar{\mu}_t\|u-v\|^2 &\le \langle \nabla_y F_{i,t}(x,u) - \nabla_y F_{i,t}(x,v), u - v\rangle\\ &= \langle \nabla_y F_{i,t}(x',v) - \nabla_y F_{i,t}(x,v), u - v\rangle\\ &\le \bar{L}_t\|x-x'\|\|u-v\|, \end{aligned} \tag{96}$$

where we use the joint smoothness of the gradients with respect to (x,y). Hence

$$\left\|y_{i,\lambda_t}^*(x) - y_{i,\lambda_t}^*(x')\right\| \le \kappa_{y,t}\|x-x'\|. \tag{97}$$

We next study one lower-level update while keeping the target $y_{i,\lambda_t}^*(x^t)$ fixed. Denote

$$y_{i,t}^* := y_{i,\lambda_t}^*(x^t), \qquad d_{i,y}^t := y_i^t - y_{i,t}^*.$$

By definition, we have that $q_{i,y}^t = \nabla_y F_{i,t}(x^t, y_i^t)$, while $\nabla_y F_{i,t}(x^t, y_{i,t}^*) = 0$. Thus, by equation 94, we obtian

$$\left\langle d_{i,y}^t, q_{i,y}^t \right\rangle \geq \bar{\mu}_t \|d_{i,y}^t\|^2, \tag{98}$$

and by equation 95,

$$\|q_{i,y}^t\| \leq \bar{L}_t \|d_{i,y}^t\|. \tag{99}$$

Since $\widehat{h}_{i,y}^t = q_{i,y}^t + e_{i,y}^t$, the update is $y_i^{t+1} = y_i^t - \eta_t \left(q_{i,y}^t + e_{i,y}^t\right)$. Therefore,

$$\begin{aligned}\left\|y_i^{t+1} - y_{i,t}^*\right\|^2 = \|d_{i,y}^t\|^2 - 2\eta_t \left\langle d_{i,y}^t, q_{i,y}^t \right\rangle - 2\eta_t \left\langle d_{i,y}^t, e_{i,y}^t \right\rangle \\ + \eta_t^2 \left\|q_{i,y}^t + e_{i,y}^t\right\|^2.\end{aligned} \tag{100}$$

Using $-2\langle a, b\rangle \leq \bar{\mu}_t \|a\|^2 + \frac{1}{\bar{\mu}_t}\|b\|^2$, together with $\|a+b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$, equation 98, and equation 99, we obtain

$$\begin{aligned}\left\|y_i^{t+1} - y_{i,t}^*\right\|^2 \leq \left(1 - \bar{\mu}_t\eta_t + 2\bar{L}_t^2\eta_t^2\right)\|d_{i,y}^t\|^2 \\ + \left(\frac{\eta_t}{\bar{\mu}_t} + 2\eta_t^2\right)\|e_{i,y}^t\|^2.\end{aligned} \tag{101}$$

By $\eta_t \leq \frac{\bar{\mu}_t}{4\bar{L}_t^2}$, we have $2\bar{L}_t^2\eta_t^2 \leq \frac{\bar{\mu}_t\eta_t}{2}$. Moreover, since $\eta_t \leq \frac{1}{\bar{\mu}_t}$, we have $\frac{\eta_t}{\bar{\mu}_t} + 2\eta_t^2 \leq \frac{3\eta_t}{\bar{\mu}_t}$. Therefore,

$$\left\|y_i^{t+1} - y_{i,\lambda_t}^*(x^t)\right\|^2 \leq \left(1 - \frac{\bar{\mu}_t\eta_t}{2}\right)\left\|y_i^t - y_{i,\lambda_t}^*(x^t)\right\|^2 + \frac{3\eta_t}{\bar{\mu}_t}\|e_{i,y}^t\|^2. \tag{102}$$

We now account for the movement of the upper-level variable. Using

$$\begin{aligned}y_i^{t+1} - y_{i,\lambda_t}^*(x^{t+1}) = \left[y_i^{t+1} - y_{i,\lambda_t}^*(x^t)\right] \\ + \left[y_{i,\lambda_t}^*(x^t) - y_{i,\lambda_t}^*(x^{t+1})\right],\end{aligned} \tag{103}$$

and $\|a+b\|^2 \leq (1+\rho)\|a\|^2 + \left(1 + \frac{1}{\rho}\right)\|b\|^2$ with $\rho = \frac{\bar{\mu}_t\eta_t}{4}$, we obtain from equation 97 and equation 102

$$\begin{aligned}&\left\|y_i^{t+1} - y_{i,\lambda_t}^*(x^{t+1})\right\|^2 \\ &\leq \left(1 - \frac{\bar{\mu}_t\eta_t}{4}\right)\left\|y_i^t - y_{i,\lambda_t}^*(x^t)\right\|^2 + \frac{4\eta_t}{\bar{\mu}_t}\|e_{i,y}^t\|^2 + \frac{10\kappa_{y,t}^2\xi^2\eta_t}{\bar{\mu}_t}\left(\|e_x^t\|^2 + \|q_x^t\|^2\right),\end{aligned} \tag{104}$$

where we used $x^{t+1} - x^t = -\xi\eta_t(e_x^t + q_x^t)$ and hence we obtain $\|x^{t+1} - x^t\|^2 \leq 2\xi^2\eta_t^2\left(\|e_x^t\|^2 + \|q_x^t\|^2\right)$. Finally, we account for the change of the penalty parameter. Let $u = y_{i,\lambda_t}^*(x), v = y_{i,\lambda_{t+1}}^*(x)$. The optimality condition at u gives $\nabla_y f_i(x,u) + \lambda_t \nabla_y g_i(x,u) = 0$. Therefore, by $l_{f,0}$-Lipschitz and continuously differentiable of f_i, we have

$$\|\nabla_y g_i(x,u)\| \leq \frac{l_{f,0}}{\lambda_t}.$$

Moreover,

$$\nabla_y F_{i,t+1}(x,u) = \Delta\lambda_t \nabla_y g_i(x,u), \qquad \nabla_y F_{i,t+1}(x,v) = 0.$$

Since $F_{i,t+1}(x,\cdot)$ is $\bar{\mu}_{t+1}$-strongly convex,

$$\left\|y_{i,\lambda_{t+1}}^*(x) - y_{i,\lambda_t}^*(x)\right\| \leq \frac{l_{f,0}|\Delta\lambda_t|}{\lambda_t\bar{\mu}_{t+1}}. \tag{105}$$

Now write

$$\begin{aligned}y_i^{t+1} - y_{i,\lambda_{t+1}}^*(x^{t+1}) = \left[y_i^{t+1} - y_{i,\lambda_t}^*(x^{t+1})\right] \\ + \left[y_{i,\lambda_t}^*(x^{t+1}) - y_{i,\lambda_{t+1}}^*(x^{t+1})\right].\end{aligned} \tag{106}$$

Applying Young's inequality with $\rho = \frac{\bar{\mu}_t\eta_t}{8}$, and using $\bar{\mu}_t\eta_t \leq 1$, equation 104, and equation 105, gives

$$\begin{aligned}&\left\|y_i^{t+1} - y_{i,\lambda_{t+1}}^*(x^{t+1})\right\|^2 \\ &\leq \left(1 - \frac{\bar{\mu}_t\eta_t}{8}\right)\left\|y_i^t - y_{i,\lambda_t}^*(x^t)\right\|^2 + \frac{5\eta_t}{\bar{\mu}_t}\|e_{i,y}^t\|^2 \\ &\quad + \frac{12\kappa_{y,t}^2\xi^2\eta_t}{\bar{\mu}_t}\left(\|e_x^t\|^2 + \|q_x^t\|^2\right) + \frac{9l_{f,0}^2(\Delta\lambda_t)^2}{\bar{\mu}_t\eta_t\lambda_t^2\bar{\mu}_{t+1}^2}.\end{aligned} \tag{107}$$

Taking total expectation proves equation 23. □

C.9 PROOF OF LEMMA 9

Proof. We first establish the Lipschitz continuity of the lower-level solution mapping. Let $u = y_i^*(x), v = y_i^*(x')$. The optimality conditions give $\nabla_y g_i(x,u) = 0, \nabla_y g_i(x',v) = 0$. Since $g_i(x,\cdot)$ is μ_g-strongly convex,

$$\begin{aligned}\mu_g\|u-v\|^2 &\le \langle \nabla_y g_i(x,u) - \nabla_y g_i(x,v), u-v\rangle \\ &= \langle \nabla_y g_i(x',v) - \nabla_y g_i(x,v), u-v\rangle \\ &\le l_{g,1}\|x-x'\|\|u-v\|. \end{aligned} \tag{108}$$

Therefore,

$$\|y_i^*(x) - y_i^*(x')\| \le \kappa_z\|x-x'\|, \qquad \kappa_z = \frac{l_{g,1}}{\mu_g}. \tag{109}$$

Denote $z_{i,t}^* := y_i^*(x^t), d_{i,z}^t := z_i^t - z_{i,t}^*$. By the optimality condition, we have $\nabla_z g_i(x^t, z_{i,t}^*) = 0$. Since $g_i(x^t,\cdot)$ is μ_g-strongly convex,

$$\left\langle d_{i,z}^t, q_{i,z}^t \right\rangle \ge \mu_g \|d_{i,z}^t\|^2, \tag{110}$$

where $q_{i,z}^t = \nabla_z g_i(x^t, z_i^t)$. Moreover, by $l_{g,1}$-smoothness,

$$\|q_{i,z}^t\| \le l_{g,1}\|d_{i,z}^t\|. \tag{111}$$

Using $\widehat{h}_{i,z}^t = q_{i,z}^t + e_{i,z}^t$ and $z_i^{t+1} = z_i^t - \beta_t\left(q_{i,z}^t + e_{i,z}^t\right)$, we obtain

$$\begin{aligned}\left\|z_i^{t+1} - z_{i,t}^*\right\|^2 = \|d_{i,z}^t\|^2 &- 2\beta_t\left\langle d_{i,z}^t, q_{i,z}^t\right\rangle - 2\beta_t\left\langle d_{i,z}^t, e_{i,z}^t\right\rangle \\ &+ \beta_t^2\left\|q_{i,z}^t + e_{i,z}^t\right\|^2. \end{aligned} \tag{112}$$

Using $-2\langle a,b\rangle \le \mu_g\|a\|^2 + \frac{1}{\mu_g}\|b\|^2$, together with $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$, equation 110, and equation 111, gives

$$\begin{aligned}\left\|z_i^{t+1} - z_{i,t}^*\right\|^2 \le\ & \left(1 - \mu_g\beta_t + 2l_{g,1}^2\beta_t^2\right)\|d_{i,z}^t\|^2 \\ &+ \left(\frac{\beta_t}{\mu_g} + 2\beta_t^2\right)\|e_{i,z}^t\|^2. \end{aligned} \tag{113}$$

Since $\beta_t \le \frac{\mu_g}{4l_{g,1}^2}$, we have $2l_{g,1}^2\beta_t^2 \le \frac{\mu_g\beta_t}{2}$. Since $\beta_t \le \frac{1}{\mu_g}$, we also have $\frac{\beta_t}{\mu_g} + 2\beta_t^2 \le \frac{3\beta_t}{\mu_g}$. Thus,

$$\left\|z_i^{t+1} - y_i^*(x^t)\right\|^2 \le \left(1 - \frac{\mu_g\beta_t}{2}\right)\left\|z_i^t - y_i^*(x^t)\right\|^2 + \frac{3\beta_t}{\mu_g}\|e_{i,z}^t\|^2. \tag{114}$$

We next account for the movement of the upper-level variable:

$$z_i^{t+1} - y_i^*(x^{t+1}) = \left[z_i^{t+1} - y_i^*(x^t)\right] + \left[y_i^*(x^t) - y_i^*(x^{t+1})\right].$$

Applying $\|a+b\|^2 \le (1+\rho)\|a\|^2 + \left(1+\frac{1}{\rho}\right)\|b\|^2$ with $\rho = \frac{\mu_g\beta_t}{4}$, and using equation 109, gives

$$\begin{aligned}&\left\|z_i^{t+1} - y_i^*(x^{t+1})\right\|^2 \\ &\le \left(1 - \frac{\mu_g\beta_t}{4}\right)\left\|z_i^t - y_i^*(x^t)\right\|^2 + \frac{4\beta_t}{\mu_g}\|e_{i,z}^t\|^2 \\ &\quad + \left(1 + \frac{4}{\mu_g\beta_t}\right)\kappa_z^2\|x^{t+1} - x^t\|^2. \end{aligned} \tag{115}$$

Since $x^{t+1} - x^t = -\xi\eta_t(e_x^t + q_x^t)$, we have $\|x^{t+1} - x^t\|^2 \le 2\xi^2\eta_t^2\left(\|e_x^t\|^2 + \|q_x^t\|^2\right)$. Moreover, since $\mu_g\beta_t \le 1$, we have

$$2\left(1 + \frac{4}{\mu_g\beta_t}\right)\xi^2\eta_t^2 \le \frac{10\xi^2\eta_t^2}{\mu_g\beta_t}.$$

Therefore,

$$\begin{aligned}
&\left\|z_i^{t+1} - y_i^*(x^{t+1})\right\|^2 \\
&\leq \left(1 - \frac{\mu_g \beta_t}{4}\right) \left\|z_i^t - y_i^*(x^t)\right\|^2 + \frac{4\beta_t}{\mu_g} \|e_{i,z}^t\|^2 \\
&\quad + \frac{10\kappa_z^2 \xi^2 \eta_t^2}{\mu_g \beta_t} \left(\|e_x^t\|^2 + \|q_x^t\|^2\right). \qquad (116)
\end{aligned}$$

Taking total expectation proves equation 24. □