STOCHASTIC ALIGNED MULTI-OBJECTIVE OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Aligned Multi-Objective Optimization (AMOO) studies how to minimize multiple objectives that share a common optimum. Existing deterministic guarantees do not directly cover noisy minibatch updates. Unrestricted weight scaling can amplify noise, while choosing weights from the update batch can bias the combined gradient. We propose delayed update normalized AMOO (DUNAMOO), which selects nonnegative weights that sum to one and applies them one iteration later. This controls weight scaling and separates the applied weights from fresh gradient noise, using one minibatch per iteration. We provide, to our knowledge, the first convergence analysis of stochastic AMOO that handles noisy Polyak weight selection and delayed updates. Under the stated curvature and noise conditions, we prove finite expected entry into a near region. After entry, the expected squared error contributed by paths with no earlier exit has a geometrically decaying bound down to a noise-dependent level. Experiments on tasks with loss-scale imbalance show favorable convergence trends for DUNAMOO.

1 INTRODUCTION

Multi-objective optimization (MOO) arises in multi-task learning (Sener & Koltun, 2018; Lin & Zhang, 2023; Liu et al., 2021; Navon et al., 2022), reinforcement learning with multiple rewards (Dann et al., 2023; Lee et al., 2024), and large language model post-training (Guo et al., 2025; Ramé et al., 2023; Dai et al., 2024). Conflicting objectives require trade-offs. Aligned multi-objective optimization (AMOO) instead studies objectives sharing a common minimizer (Efroni et al., 2025), modeling complementary feedback toward a shared goal.

Sharing a minimizer does not make every objective equally informative. One objective may be flat in a direction where another provides useful feedback, and equal weighting can dilute that information (Efroni et al., 2025). The question is how to exploit such differences when losses and gradients must be estimated from sampled data.

Efroni et al. (2025) establish deterministic guarantees for Curvature AMOO (CAMOO) and Polyak AMOO (PAMOO). PAMOO uses gradients and objective gaps (values minus their minima) in a quadratic program (QP), without Hessians. Its rate depends on the curvature of a weighted objective at the common optimum. We study whether this approach remains reliable when full-data quantities are replaced by minibatch estimates.

Minibatch feedback can invalidate PAMOO's deterministic guarantee. Its noisy weight score may have no finite maximum (Proposition B.2). Applying batch-selected weights immediately can yield a limit away from the common optimum even at arbitrarily small learning rates (Proposition B.1).

Can adaptive weighting make provable progress
toward a common optimum when both losses and gradients are noisy?

We propose delayed update normalized AMOO (DUNAMOO). Weights are nonnegative and sum to one, fixing their scale. Applying them one iteration later makes them known before the update batch is drawn. Each iteration uses only one minibatch.

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.

Normalization makes the weight problem well posed, and delay removes bias from the current update. Three questions remain. First, how much true objective gap do weights selected from noisy data retain? Second, how much of that guarantee survives the intervening update? Third, how can local curvature imply progress in expectation when individual updates may move away from the optimum? Our proof sketch answers these questions in this order (Section 5.2).

Contributions.

- **First convergence guarantees for stochastic AMOO.** To our knowledge, this is the first analysis jointly controlling noisy Polyak weight selection and delayed dependence through local mixture curvature. Under an explicit noise condition, we prove finite expected entry into a near region. After entry, the expected squared error contributed by paths with no earlier exit is bounded by a geometrically decaying initial-error term plus a noise-dependent residual (Theorem 1). A separate global averaged maximum-gap bound holds under smooth convexity alone (Corollary 1).
- **DUNAMOO, a new algorithm for stochastic AMOO.** Our Hessian-free method combines normalized weight selection with one-iteration delayed application. These make selection well posed and remove same-batch bias using one fresh minibatch per iteration.
- **Empirical evaluation under loss-scale imbalance.** Multi-MNIST experiments spanning multiple learning rates and loss-scale imbalance settings show broadly consistent performance gains for DUNAMOO-SGD across most tested configurations (Section 6, Appendix D.3.2).

2 RELATED WORK

Aligned optimization. Efroni et al. (2025) analyze deterministic CAMOO and PAMOO under smoothness and self-concordance, distinguishing global adaptive curvature from mixture curvature at the optimum. Kretzu et al. (2025) derive convex maximum-gap guarantees and MG-AMOO, which passes a maximum-gap objective to a single-objective optimizer. Our analysis also covers delayed maximum-gap selection, while certifying the actual noisy Polyak QP.

Multi-task learning and stochastic MOO. MGDA seeks common descent (Désidéri, 2012; Sener & Koltun, 2018), while gradient surgery, CAGrad, and Nash-MTL address conflict and task balance (Yu et al., 2020; Liu et al., 2021; Navon et al., 2022). Stochastic aggregation bias is studied by SMG and MoCo (Liu & Vicente, 2024; Fernando et al., 2023). Further analyses address gradient manipulation, independent sampling, direction tracking, periodic reuse, and improved Pareto-stationarity rates (Zhou et al., 2022; Chen et al., 2024; Xiao et al., 2023; Xu et al., 2025; Chen, 2026). We target simultaneous objective optimality under alignment, which Pareto stationarity alone does not ensure when component Hessians are singular.

Stochastic Polyak and fixed-objective methods. Stochastic Polyak methods study coupling between step selection and noise (Loizou et al., 2021; Orvieto et al., 2022). Making adaptive choices depend only on past information is a known device for removing this coupling (Li & Orabona, 2019). We apply it to multi-objective weights while certifying the noisy QP and controlling delayed dependence. Related stochastic and self-concordant analyses concern fixed objectives (Gower et al., 2019; Bach, 2010; Bach & Moulines, 2013; Marteau-Ferey et al., 2019).

Stochastic scalarization and excess risks. ExcessMTL uses estimated excess risks (He et al., 2024), and (Ada)OMD-TCH gives stochastic Tchebycheff scalarization rates, including convex global guarantees under bounded parameter regions and loss/gradient bounds (Liu et al., 2026). With known optimal values, minimax scalarization also targets the maximum gap. Our two-phase result instead concerns an unprojected adaptive mixture under uniform centered-noise bounds and local curvature at an aligned optimum, with unbounded population losses and gradients allowed.

3 PRELIMINARIES

Notation. Let n be the parameter dimension, m the number of objectives, and $[m] = \{1, \dots, m\}$. Write $\mathbb{R}^m_+$ for the nonnegative cone and $\Delta_m = \{w \in \mathbb{R}^m_+ : \mathbf{1}^\top w = 1\}$ for the unit simplex, where $\mathbf{1}$ is the all-ones vector. I denotes the identity matrix. We call weights in Δ_m *normalized*, meaning nonnegative with unit sum. Vector norms are Euclidean, and $\lambda_{\min}(A)$ is the smallest eigenvalue of a symmetric matrix A. Expectations and probabilities are denoted by $\mathbb{E}$ and $\mathbb{P}$.

3.1 ALIGNED MULTI-OBJECTIVE OPTIMIZATION (AMOO)

We optimize population objectives $f_i(x) = \mathbb{E}_{\mathsf{Z}}[\ell_i(x;\mathsf{Z})]$ on $\mathbb{R}^n$, where Z is a data sample and ℓ_i its loss for objective i. For a fixed $w \in \Delta_m$, the mixture f_w and gradient matrix J satisfy

$$f_w(x) = \sum_i w_i f_i(x), \qquad J(x) = [\nabla f_1(x), \dots, \nabla f_m(x)] \in \mathbb{R}^{n\times m},$$

with $\nabla f_w(x) = J(x)w$ when w is held fixed.

Definition 1 (Aligned multi-objective optimization)**.** The objectives are aligned if there exists a common minimizer $x_\star \in \bigcap_i \arg\min_x f_i(x)$, which AMOO seeks (Efroni et al., 2025).

This population condition does not require sample interpolation, in which sample gradients vanish at the common minimizer (Vaswani et al., 2019).

Write $f_i^\star = f_i(x_\star)$, $\Delta_i(x) = f_i(x) - f_i^\star$, and $\Delta(x) = (\Delta_i(x))_{i=1}^m$. The weighted and maximum gaps are $\Delta_w(x) = w^\top \Delta(x)$ and $M(x) = \max_i \Delta_i(x)$. Alignment makes all gaps nonnegative, and $M(x) = 0$ means simultaneous population optimality.

Pareto stationarity, $\min_{w\in\Delta_m} \|J(x)w\| = 0$, need not imply $M(x) = 0$. For $f_1(x) = x_1^2/2$, $f_2(x) = x_2^2/2$ and $x = (1, 0)$, the weight $(0, 1)$ gives stationarity although $M(x) = 1/2$.

3.2 OBJECTIVE GEOMETRY AND STOCHASTIC FEEDBACK

Write C^r for functions with r continuous derivatives.

Assumption 1 (Smooth convex objectives)**.** Each f_i is convex, C^2, and globally β-smooth for some $\beta > 0$, meaning that for all $x, y \in \mathbb{R}^n$,

$$f_i(y) \le f_i(x) + \langle \nabla f_i(x), y - x\rangle + \tfrac{\beta}{2}\|y - x\|^2.$$

The notation $\nabla^3 f_i(x)[d, d, d]$ denotes the third directional derivative.

Assumption 2 (Self-concordance)**.** Each f_i is C^3 and M_{f}-self-concordant for some $M_{\mathrm{f}} \ge 0$, meaning that for all $x, d \in \mathbb{R}^n$,

$$|\nabla^3 f_i(x)[d, d, d]| \le 2M_{\mathrm{f}}\big(d^\top \nabla^2 f_i(x) d\big)^{3/2}. \tag{1}$$

Self-concordance limits how rapidly curvature can change along a direction. The two-phase result requires positive curvature for some normalized mixture at the common minimizer, measured by $\mu_{\mathrm{L}} = \max_{w\in\Delta_m} \lambda_{\min}(\nabla^2 f_w(x_\star)) > 0$. These conditions are those of deterministic PAMOO (Efroni et al., 2025, Section 5), permit singular component Hessians while ensuring a unique common minimizer, and enter only the analysis, since DUNAMOO computes no Hessians.

Appendix A.4 gives a nonquadratic example with singular component Hessians, $\mu_{\mathrm{L}} = 1/m > 0$ and vanishing mixture curvature far from the optimum, so global strong convexity is unnecessary.

Let $\mathcal{F}_k$ be the history before minibatch $\mathcal{B}_k$, containing the fixed initial point and all past oracle and solver outputs. The iterate x_k, stored weight w_k and step (learning rate) $\eta_k > 0$ are $\mathcal{F}_k$-measurable, meaning determined by this history, and we call them predictable. The new batch $\mathcal{B}_k$ is independent of $\mathcal{F}_k$. Write $\mathbb{E}_k = \mathbb{E}[\,\cdot \mid \mathcal{F}_k]$, $V_k = \|x_k - x_\star\|^2$, and call $\mathbb{E}V_k$ the mean-square error (MSE).

For a stopping time T, $\mathcal{F}_T$ denotes the history available at that random iteration. Oracle and solver outputs are finite-valued and jointly measurable, with exact ties resolved by Lemma C.2.

Minibatch moments. For the minibatch implementation, sample gradients are unbiased and differentiation under expectation is valid. Define the centered loss-error vector $r(x;\mathsf{Z}) = (\ell_i(x;\mathsf{Z}) - f_i(x))_{i=1}^m$ and the projection $P = I - \mathbf{1}\mathbf{1}^\top/m$ that removes its common component. For finite constants $\nu_\Delta, \sigma_{\mathrm{g}} \ge 0$, require uniformly over x

$$\mathbb{E}\|Pr(x;\mathsf{Z})\|^2 \le \nu_\Delta^2, \qquad \max_i \mathbb{E}\|\nabla_x \ell_i(x;\mathsf{Z}) - \nabla f_i(x)\|^2 \le \sigma_{\mathrm{g}}^2. \tag{2}$$

Common loss fluctuations shift all normalized-weight scores equally. Only loss contrasts therefore need a moment bound, and the gradient bound controls update noise. Cross-objective independence and bounded population losses or gradients are not assumed (Proposition A.1).

4 DELAYED UPDATE NORMALIZED AMOO (DUNAMOO)

From PAMOO to normalized weight selection. PAMOO maximizes $2\Delta(x)^\top u - \|J(x)u\|^2$ over $u \in \mathbb{R}^m_+$ and uses $x^+ = x - J(x)u$ (Efroni et al., 2025, Algorithm 3). Its weights set both relative importance and update magnitude. DUNAMOO instead sets $u = \eta w$ with $w \in \Delta_m$. Dividing the score by $\eta > 0$ gives $2\Delta(x)^\top w - \eta\|J(x)w\|^2$, so the application step fixes its coefficient without a separate penalty. Convexity gives

$$\|x - \eta J(x)w - x_\star\|^2 \le \|x - x_\star\|^2 - \eta\{2\Delta(x)^\top w - \eta\|J(x)w\|^2\}.$$

The QP therefore optimizes a lower bound on exact-feedback squared-distance decrease.

Delayed weight application. One fresh minibatch updates x_k using stored w_k and selects w_{k+1} for the next update. Only weights are delayed, and gradients are current. The score uses the next step η_{k+1}, known before the batch is drawn.

Minibatch estimates. For deterministic $N_k \ge 1$, draw $\mathcal{B}_k = (\mathsf{Z}_{k,s})_{s=1}^{N_k}$ with independent and identically distributed (i.i.d.) observations, independently of $\mathcal{F}_k$. At x_k, using optimal values supplied up to a common offset, write

$$\widehat{\Delta}_{k,i} = \frac{1}{N_k}\sum_{s=1}^{N_k} \ell_i(x_k; \mathsf{Z}_{k,s}) - f_i^\star, \qquad \widehat{g}_{k,i} = \frac{1}{N_k}\sum_{s=1}^{N_k} \nabla_x \ell_i(x_k; \mathsf{Z}_{k,s}). \tag{3}$$

Stack $\widehat{\Delta}_k = (\widehat{\Delta}_{k,i})_{i=1}^m$ and $\widehat{J}_k = [\widehat{g}_{k,1}, \dots, \widehat{g}_{k,m}]$, whose Gram is $\widehat{J}_k^\top \widehat{J}_k$.

Algorithm 1 DUNAMOO (normalized weight selection with delayed application)

1. **Input:** $x_0 \in \mathbb{R}^n$, positive step schedule $(\eta_k)_{k\ge 0}$, batch sizes (N_k), optimal values $(f_i^\star)$ up to a common additive offset, updates $K \ge 1$.
2. **Initialize:** $w_0 = \mathbf{1}/m$.
3. **For** $k = 0, \dots, K-1$ **do**
4. Draw one fresh minibatch $\mathcal{B}_k$ and compute $\widehat{\Delta}_k, \widehat{J}_k$ by Eq. (3).
5. Form $g_k = \widehat{J}_k w_k$ using the *stored* weight.
6. Select and store the *next* weight

$$w_{k+1} \in \arg\max_{w \in \Delta_m} \{2\widehat{\Delta}_k^\top w - \eta_{k+1}\|\widehat{J}_k w\|^2\}. \tag{4}$$

7. Update $x_{k+1} = x_k - \eta_k g_k$.
8. **End for.** Return x_K.

Only the optimal values, not $x_\star$, are needed. Errors that differ across objectives must meet the conditions in Appendix A.5.2, and a common additive offset cancels.

Conditional unbiasedness. With the stored weight, $\mathbb{E}_k[\widehat{J}_k w_k] = J(x_k)w_k$, and the conditional mean-square gradient noise is at most σ_g^2/N_k. The next weight still shares data with the displacement (Figure 1), which Section 5 controls.

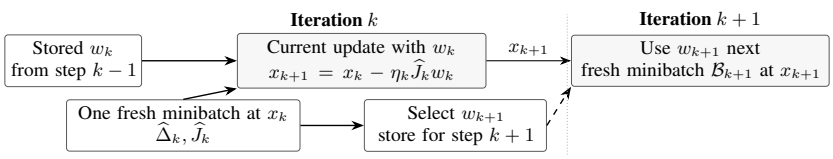


Figure 1: **Weight timing in DUNAMOO.** Only weights are delayed. Batch $\mathcal{B}_k$ updates x_k with stored w_k and selects w_{k+1} for the next step (dashed arrow). The next update uses fresh gradients at x_{k+1}, and both x_{k+1} and w_{k+1} depend on $\mathcal{B}_k$.

5 CONVERGENCE ANALYSIS

We first state the guarantees for a fixed batch size N and the basic step $\eta = 1/(2\beta)$, then explain how the noisy QP yields them. Appendix A gives the general-oracle and variable-batch extensions.

5.1 MAIN CONVERGENCE THEOREM

To account for delay, we track the current squared distance V_k together with the expected squared distance after the next update. Their weighted sum is the analysis-only *conditional energy*

$$\mathcal{L}_k = \tfrac{4}{3}\mathbb{E}_k V_{k+1} + \tfrac{2}{3}V_k, \qquad q = \frac{\mu_{\mathrm{L}}}{12\beta}, \tag{5}$$

where $\mathbb{E}_k = \mathbb{E}[\cdot \mid \mathcal{F}_k]$ and q determines geometric contraction. The energy controls V_k but is not a distance or algorithm input. The near-phase threshold R^2 and noise-dependent residual term are

$$\begin{aligned} R &= \frac{\sqrt{2}}{M_{\mathrm{f}}\sqrt{\beta}} \quad (M_{\mathrm{f}} > 0), \\ \mathcal{E}_\infty &= \left(\frac{34}{3\beta\mu_{\mathrm{L}}} + \frac{2}{3\beta^2}\right)\frac{\sigma_{\mathrm{g}}^2}{N} + \frac{32\nu_\Delta}{3\mu_{\mathrm{L}}\sqrt{N}}. \end{aligned} \tag{6}$$

Write $\mathbf{1}_A$ for an event indicator, $s \wedge t = \min\{s, t\}$ and $\inf \varnothing = +\infty$.

Theorem 1 (Two-phase convergence of DUNAMOO)**.** *Assume alignment, Assumptions 1–2, $\mu_{\mathrm{L}} > 0$, and the unbiased minibatch feedback in Eq.* (2)*. Run Algorithm 1 for all $k \geq 0$ with fixed x_0, $w_0 = \mathbf{1}/m$, $\eta = 1/(2\beta)$, fixed batch size $N \geq 1$, and fresh i.i.d. observations as in Eq.* (3)*. Optimal values are supplied up to a common additive offset. Then*

$$V_k \leq \mathcal{L}_k \leq 2V_k + \frac{3\sigma_{\mathrm{g}}^2}{8\beta^2 N}. \tag{7}$$

For $M_{\mathrm{f}} > 0$, let $T = \inf\{k \geq 0 : \mathcal{L}_k \leq R^2\}$ be the entry time, on $\{T < \infty\}$ let $T_{\mathrm{exit}} = \inf\{j \geq T : \mathcal{L}_j > R^2\}$ be the first exit time, and write $\gamma = q(R^2 - \mathcal{E}_\infty)/(2R)$ for the far-phase decrement. For all integers $k \geq 0$ and $s \geq 0$,

$$\begin{cases} \mathbb{E}\sqrt{\mathcal{L}_{k\wedge T}} \leq \sqrt{\mathcal{L}_0} - \gamma\,\mathbb{E}(k \wedge T), & \text{(i) far phase, before entry,} \\ \mathbb{E}\big[V_{T+s}\mathbf{1}_{\{T_{\mathrm{exit}} \geq T+s\}} \mid \mathcal{F}_T\big] \\ \quad \leq (1-q)^s\mathcal{L}_T + \mathcal{E}_\infty\{1-(1-q)^s\}, & \text{(ii) near phase, } s \text{ steps after entry.} \end{cases} \tag{8}$$

Branch (ii) holds on $\{T < \infty\}$ and includes a path at its first exit, but not at later iterates.

If $\mathcal{E}_\infty < R^2$, the decrement γ is positive, so $T < \infty$ almost surely and

$$\mathbb{E}T \leq \frac{\sqrt{\mathcal{L}_0}}{\gamma} = \frac{2R\sqrt{\mathcal{L}_0}}{q(R^2 - \mathcal{E}_\infty)}. \tag{9}$$

Branch (ii) does not need this small-error condition.

Quadratics. *If $M_{\mathrm{f}} = 0$, the ordinary MSE satisfies globally*

$$\mathbb{E}V_k \leq (1-2q)^k\mathcal{L}_0 + \tfrac{1}{2}\mathcal{E}_\infty\{1-(1-2q)^k\}. \tag{10}$$

What the two branches assert. Branch (i) replaces the fixed decrement of deterministic PAMOO's distance bound (Efroni et al., 2025) with expected progress of the stopped energy radius. Branch (ii) is a *localized second moment*. It averages the squared distance only over paths that have not left the near region before step $T + s$, so it is neither the ordinary MSE, which also counts paths after an exit, nor a bound on any single path. The quantity it leaves open is the probability of staying, which Corollary A.2 supplies. On entry and for every integer horizon $h \geq 0$,

$$\mathbb{P}(T_{\mathrm{exit}} > T + h \mid \mathcal{F}_T) \geq 1 - \alpha_h, \qquad \alpha_h = \frac{\mathcal{L}_T + q\mathcal{E}_\infty h}{R^2}, \tag{11}$$

and whenever $\alpha_h < 1$, dividing branch (ii) by $1 - \alpha_h$ bounds the MSE conditioned on no exit within h steps by $[(1-q)^h \mathcal{L}_T + \mathcal{E}_\infty\{1-(1-q)^h\}]/(1-\alpha_h)$. The near-phase guarantee is therefore a statement about a horizon of order $(R^2 - \mathcal{L}_T)/(q\mathcal{E}_\infty)$ steps after entry, which grows as the noise shrinks and is unbounded with exact feedback, where the near region is invariant (Eq. (16)). Paths that exit are excluded, not controlled. Proposition A.3 shows that their contribution can make the ordinary MSE diverge under finite variance even when $\mathcal{E}_\infty < R^2$, so the indicator in branch (ii) cannot be removed, and Corollary A.3 restores a margin below R^2 via entry into a smaller region.

5.2 Proof sketch

The three steps below answer the three questions in Section 1, in the same order.

1. Certify the quality of noisy selection. We need the selected mixture to retain a useful population gap, not its weights to approximate exact-data weights. We compare this gap with the largest population gap, whose index j is fixed before the current batch is drawn. The QP's first-order optimality condition (the Karush–Kuhn–Tucker condition) and square completion give

$$\widehat{\Delta}_{k,j} - \widehat{\Delta}_k^\top w_{k+1} \le \tfrac{\eta}{4}\|\widehat{g}_{k,j}\|^2.$$

The gradient estimate for j is unbiased, but the selected loss error need not average to zero. Bounding it by the contrast moment gives

$$\mathbb{E}_k \Delta_{w_{k+1}}(x_k) \ge \tfrac{3}{4} M(x_k) - \frac{\nu_\Delta}{\sqrt{N}} - \frac{\sigma_{\mathrm{g}}^2}{8\beta N}. \tag{12}$$

This lower bound on the retained population gap is the *selection certificate*. It needs no Gram inverse or bounded iterates and allows zero weights (Proposition A.1).

2. Transfer selection quality to the next iterate. The new weight and the intervening update use the same minibatch, so their dependence cannot be discarded. Smooth convexity bounds the loss of selection quality for every weight and pair of points:

$$\Delta_{w_{k+1}}(x_{k+1}) \ge \tfrac{2}{3}\Delta_{w_{k+1}}(x_k) - \beta\|x_{k+1} - x_k\|^2.$$

Conditional unbiasedness with the stored weight gives $2\eta\Delta_{w_k}(x_k) \le V_k - \mathbb{E}_k V_{k+1} + \eta^2\sigma_{\mathrm{g}}^2/N$, and the expected squared movement is at most $2\beta\eta^2\Delta_{w_k}(x_k) + \eta^2\sigma_{\mathrm{g}}^2/N$. Substitution combines two consecutive updates in a *one-lag recursion*. At $\eta = 1/(2\beta)$, it gives (Lemma A.5)

$$\mathbb{E}_k V_{k+2} \le \tfrac{1}{2}\mathbb{E}_k V_{k+1} + \tfrac{1}{2} V_k - \tfrac{2}{3\beta}\mathbb{E}_k \Delta_{w_{k+1}}(x_k) + \frac{5\sigma_{\mathrm{g}}^2}{8\beta^2 N}. \tag{13}$$

No selector continuity is needed, and the step uses only smooth convexity and the fact that w_{k+1} is fixed before batch $\mathcal{B}_{k+1}$ is drawn, so it applies to any adaptive weight or step selected from the previous batch. The propagated quantity $\mathbb{E}_k V_{k+1} + \frac{1}{2}V_k$ is proportional to $\mathcal{L}_k$, which controls V_k.

3. Turn local curvature into stochastic progress. Local curvature controls the realized iterate, not directly its mean squared distance. Integrating each objective's curvature *before* mixing avoids a minimum-weight restriction. A mixture attaining μ_{L} at $x_\star$ gives (Lemma A.2)

$$M(x) \ge \frac{\mu_{\mathrm{L}}\|x - x_\star\|^2}{2(1 + M_{\mathrm{f}}\sqrt{\beta}\,\|x - x_\star\|)}.$$

The nonlinear distance dependence prevents closing an ordinary-MSE recursion. Define $H(v) = v/(1 + M_{\mathrm{f}}\sqrt{\beta}\sqrt{v})$ for $v \ge 0$, so $M(x_k) \ge (\mu_{\mathrm{L}}/2)H(V_k)$. The function H is increasing, concave and 1-Lipschitz. Jensen gives $\mathbb{E}H(V_k) \le H(\mathbb{E}V_k)$, whereas Eq. (13) needs a lower bound on $\mathbb{E}H(V_k)$.

We instead retain the current history in $\mathcal{L}_k$, combining current and expected next-step squared distances. The pathwise energy comparison and the monotonicity and Lipschitz property of H transfer curvature from V_k to $\mathcal{L}_k$ at an additional noise cost (Lemma A.6):

$$\mathbb{E}_k \mathcal{L}_{k+1} \le \mathcal{L}_k - \frac{2q\mathcal{L}_k}{1 + \sqrt{\mathcal{L}_k}/R} + q\mathcal{E}_\infty \qquad (M_{\mathrm{f}} > 0). \tag{14}$$

When $\mathcal{L}_k > R^2$, conditional Jensen for the square root gives $\mathbb{E}_k\sqrt{\mathcal{L}_{k+1}} \le \sqrt{\mathcal{L}_k} - \gamma$. Stopping at entry and summing gives Eq. (8)(i). Inside the near region, the denominator is at most two, and retaining paths through their first exit gives the geometric bound in Eq. (8)(ii). For $M_{\mathrm{f}} = 0$, the denominator is one globally. The step uses only that H is increasing, concave and 1-Lipschitz, so any local curvature bound of this form can replace Lemma A.2. Appendix A.3 gives the full argument.

5.3 GLOBAL MAXIMUM-GAP GUARANTEE

The same one-lag recursion gives a global objective guarantee under smooth convexity alone.

Corollary 1 (Global convex maximum-gap bound)**.** *Assume alignment, Assumption 1, and unbiased feedback satisfying Eq.* (2)*. Run Algorithm 1 from fixed x_0 with $V_0 > 0$, $w_0 = \mathbf{1}/m$, fixed batch size N, and fresh i.i.d. observations as in Eq.* (3)*. Supply optimal values up to a common offset. For $K \geq 1$, report $\bar{x}_K = K^{-1}\sum_{k=0}^{K-1} x_k$. Use the constant step $\eta = \min\{1/(2\beta), \sqrt{NV_0}/(\sigma_{\mathrm{g}}\sqrt{K})\}$ for $\sigma_{\mathrm{g}} > 0$, and $\eta = 1/(2\beta)$ for $\sigma_{\mathrm{g}} = 0$. Then*

$$\mathbb{E}M(\bar{x}_K) = O\left(\frac{\beta V_0}{K} + \sigma_{\mathrm{g}}\sqrt{\frac{V_0}{NK}} + \frac{\nu_\Delta}{\sqrt{N}}\right). \tag{15}$$

Neither self-concordance nor positive mixture curvature is required.

A known positive upper bound may replace V_0 in this step choice and rate, and Appendix A.6 gives constants for every $0 < \eta \leq 1/\beta$.

5.4 TWO-PHASE GEOMETRY AND DETERMINISTIC LIMIT

Figure 2 summarizes the bounds, and Table A.1 lists their assumptions and error criteria.

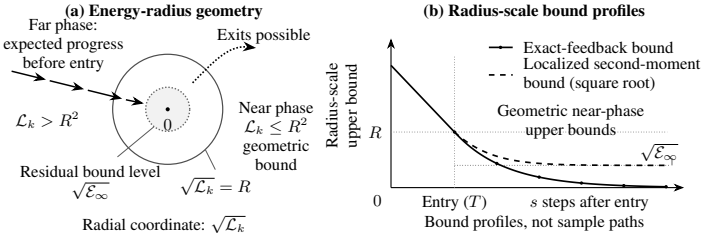


Figure 2: **Two-phase guarantees, schematically.** (a) The circles depict energy radii, not parameter-space level sets. Before entry, the stopped energy radius $\sqrt{\mathcal{L}_{k\wedge T}}$ makes progress in conditional expectation. The near-phase bound covers paths with no earlier exit, and exits remain possible. The inner circle is a residual bound level, not an invariant set. (b) Exact-feedback bound (solid) and square root of the localized second-moment bound in Eq. (8)(ii) (dashed), with $\mathcal{L}_T = R^2$ and $0 < \mathcal{E}_\infty < R^2$, as bounds rather than sample paths.

Exact-feedback specialization. With exact losses, gradients and QP solutions, correct optimal values up to a common offset, and deterministic ties, $\mathcal{E}_\infty = 0$. For $M_{\mathrm{f}} > 0$ and $T_0 = \inf\{k : \mathcal{L}_k \leq R^2\}$, Corollary A.4 gives

$$\sqrt{\mathcal{L}_k} \leq \begin{cases} \sqrt{\mathcal{L}_0} - kqR/2, & 0 \leq k < T_0, \\ \sqrt{\mathcal{L}_{T_0}}(1-q)^{(k-T_0)/2}, & k \geq T_0. \end{cases} \tag{16}$$

Since $\|x_k - x_\star\| \leq \sqrt{\mathcal{L}_k} \leq \sqrt{2}\|x_k - x_\star\|$, this envelope controls distance. Exact-feedback energy is nonincreasing, so the near region is invariant, which noise removes.

A larger certified entry region. With exact feedback, $\|x_k - x_\star\| \leq R/\sqrt{2} = 1/(M_{\mathrm{f}}\sqrt{\beta})$ suffices for entry, independently of m and μ_{L}. The original PAMOO proof uses $\sqrt{\mu_{\mathrm{L}}}/(3\sqrt{m}\,\beta M_{\mathrm{f}})$ (Efroni et al., 2025), smaller by $3\sqrt{m\beta/\mu_{\mathrm{L}}} \geq 3\sqrt{m}$. Componentwise integration also improves the original cone-weighted analysis, a proof improvement rather than universal acceleration (Appendix A.8).

What delay and larger batches resolve. Delay removes the step-independent same-batch bias (Proposition B.1), not the loss-selection error that can stall even exact-gradient updates (Proposition A.2). Larger batches reduce gradient and loss residuals at orders N^{-1} and $N^{-1/2}$, respectively, without removing localization (Corollary A.1). Appendix D.5 measures both effects.

Why use gradient-aware mixing? The delayed maximum-gap rule puts all weight on the largest estimated gap. Its generic guarantee retains a larger fraction of the largest population gap, but ignores gradient directions, whereas the QP optimizes the decrease bound in Section 4 at the *selection point*.

For $f_i(x) = x_i^2/2$, $i = 1, 2$, $x_0 = (\sqrt{5/4}, 1)$, $\eta = 1/2$ and $w_0 = (1/2, 1/2)$, two exact-feedback updates with the same delay give squared distance $181/256$ for the QP versus $189/256$ for maximum-gap selection (Appendix A.7). This proves a benefit on this instance, not a general rate or runtime advantage. Appendix D.4.2 compares both rules on Multi-MNIST, where the QP leads per update.

Comparison with equal weighting. The quadratic contraction certificate beats EW's only if mixture-curvature gains outweigh selection and delay costs. The condition is impossible for $m \leq 5$ and holds in the specified selection example for $m \geq 6$ (Appendix A.9). This orders upper-bound factors, not trajectories or training time.

5.5 Finite-accuracy weight selection

The analysis also covers a QP solved to certified finite accuracy. For a feasible weight $w \in \Delta_m$ with score gradient $p = 2\widehat{\Delta}_k - 2\eta_{k+1}\widehat{J}_k^\top \widehat{J}_k w$, the computable residual $\max_i p_i - p^\top w$ is nonnegative, vanishes at maximizers and bounds score suboptimality from above, and a deterministic residual budget ϵ adds at most $\epsilon/2$ to the selection allowance (Lemma A.7). Numerical accuracy thus enters as an explicit allowance, not an assumption that the computed weight is near an unknown optimizer.

6 Experiments

We evaluate the convergence speed of DunaMOO-SGD on a Multi-MNIST task with a shared CNN backbone and loss-scale imbalance. We compare against equal-weight SGD (EW-SGD), a regularized-cone PAMOO-SGD implementation (Efroni et al., 2025), and, in Appendix D.4, a same-batch simplex control and the delayed maximum-gap rule, with matched datasets, initializations and minibatch sequences (Appendix D).

Multi-MNIST dataset and methods. Each example places $m \in \{2, 4, 8\}$ non-overlapping 28×28 MNIST digits on a shared canvas, in the top-left and bottom-right quadrants of a 56×56 image for $m = 2$, in all four quadrants for $m = 4$, and in the outer cells of an 84×84 grid for $m = 8$ (Figure 3). The network classifies the digit at each position using a separate sample-weighted cross-entropy loss, one objective per position. EW-SGD uses uniform weights, PAMOO-SGD applies nonnegative cone weights, and DunaMOO-SGD applies normalized weights with a one-iteration delay.

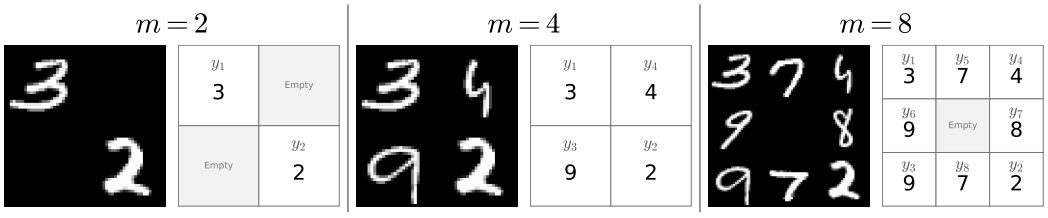


Figure 3: **Multi-MNIST examples** for $m = 2, 4$, and 8. Each panel shows an input image (left) and its labels $y_1, \ldots, y_m$ by position (right).

Sample-dependent loss multipliers. Sample-dependent loss multipliers create unequal objective scales and minibatch fluctuations. Let $\ell_i(x; \mathsf{Z})$ denote the cross-entropy loss of example Z for objective i, and let $a_{\mathsf{Z},i}$ denote its fixed loss multiplier. For a minibatch $\mathcal{B}_k$, the cross-entropy loss is $\widehat{f}_{k,i}(x) = \frac{1}{|\mathcal{B}_k|} \sum_{\mathsf{Z} \in \mathcal{B}_k} a_{\mathsf{Z},i} \ell_i(x; \mathsf{Z})$. Multipliers $(a_{\mathsf{Z},i})_{\mathsf{Z} \in \mathcal{B}_k}$ are assigned through an imbalanced mixture of regimes independently of digit labels, remain fixed throughout training, and have a constant sum across objectives for each example, so they change relative loss and gradient scales while preserving the targets. They are distinct from the adaptive weights applied to the objective gradients (construction in Appendix D.1).

Training protocol. All methods use SGD with uniform minibatch sampling with replacement, a common training budget, the same learning-rate decay schedule and the same grid of initial learning rates. The settings and learning-rate choices used in Figure 4 are given in Appendix D.3.1.

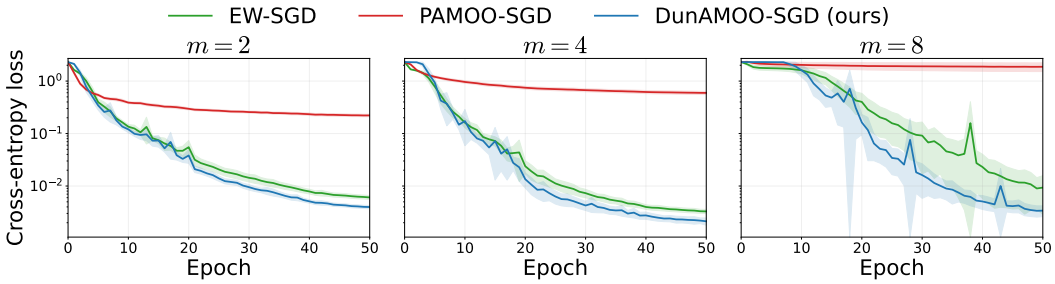


Figure 4: **Training loss of the final worst objective on Multi-MNIST.** Columns correspond to $m = 2, 4$, and 8 objectives over 50 epochs. For each method, solid curves show the seed-averaged loss trajectory of the objective with the largest seed-averaged loss at the final epoch, and shaded regions indicate pointwise approximate 95% confidence intervals.

Results. The best values in Tables D.2–D.7 improve over PAMOO-SGD in all six reported settings. Relative to EW-SGD, DUNAMOO-SGD improves in five settings, two of them by less than one standard deviation, and is worse in one. These are best-tuned comparisons, not uniform improvements across learning rates. Appendix D.3.2 gives the objective-wise plots and full learning-rate tables.

The learning-rate sweeps in Appendix D.3.1 show that the favorable performance of DUNAMOO-SGD is not confined to a single learning-rate choice or loss-multiplier configuration. Across multiple learning rates and different degrees of sample-dependent loss-scale imbalance, DUNAMOO-SGD generally achieves lower final maximum cross-entropy losses than the PAMOO-SGD baseline. PAMOO-SGD exhibits a rapid initial decrease in loss in the first five epochs, but its progress slows substantially as training proceeds, with only limited improvement during the later epochs. In contrast, DUNAMOO-SGD continues to reduce the losses over the remaining training budget, reaching substantially lower final losses. Together, these observations highlight not only the consistency of DUNAMOO-SGD's performance across multiple learning-rate and loss-multiplier settings, but also its sustained progress after PAMOO-SGD's initial improvement has largely subsided.

Synthetic checks. Appendix D.5 runs DUNAMOO on the analytic instances, where every constant of Theorem 1 is known. Every certified bound holds, the entry time is linear in $\sqrt{\mathcal{L}_0}$ as in Eq. (9), the residual follows the gradient-noise order N^{-1}, and the same-batch rule of Proposition B.1 converges to $\bar{x} \approx -0.4502$ at every step while DUNAMOO stays within its mean-square envelope.

7 CONCLUSION

Can adaptive weighting make provable progress toward a common optimum when both losses and gradients are noisy? For DUNAMOO the answer is yes, within limits stated as precisely as the guarantee. Normalized weights and one-iteration delay remove the two obstructions that minibatch noise creates for PAMOO, unrestricted weight scaling and a same-batch bias that no step size removes, and give, to our knowledge, the first convergence analysis of stochastic AMOO with noisy Polyak weight selection. Its one-lag recursion and conditional energy use only smooth convexity and a weight fixed before its batch is drawn, so they apply beyond the Polyak QP. Fixed batches leave a selection residual, exited paths are excluded rather than controlled, and the certified constants are conservative. Removing them and estimating the optimal values the method needs is the next step.

AI USE STATEMENT

Generative AI tools assisted with language and reference editing, manuscript organization and LaTeX figures, mathematical checking, and drafting the inner-entry corollary and the refined equal-weighting comparison. AI-generated scripts supported symbolic and finite-history checks. The authors remain responsible for the correctness, originality and attribution of all manuscript content.

REFERENCES

Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, 3rd edition, 2006. URL https://link.springer.com/book/10.1007/3-540-29587-9.

Francis Bach. Self-concordant analysis for logistic regression. *Electronic Journal of Statistics*, 4: 384–414, 2010. doi: 10.1214/09-EJS521.

Francis Bach and Eric Moulines. Non-strongly-convex smooth stochastic approximation with convergence rate O(1/n). In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. URL https://web.stanford.edu/~boyd/cvxbook/.

Lisha Chen. Improved convergence rates for stochastic multi-gradient descent which close the gap: A proof by AI. *arXiv preprint arXiv:2607.18174*, 2026.

Lisha Chen, Heshan Fernando, Yiming Ying, and Tianyi Chen. Three-way trade-off in multi-objective learning: Optimization, generalization and conflict-avoidance. *Journal of Machine Learning Research*, 25(193):1–53, 2024.

Juntao Dai, Xuehai Pan, Ruiyang Sun, Jiaming Ji, Xinbo Xu, Mickel Liu, Yizhou Wang, and Yaodong Yang. Safe RLHF: Safe reinforcement learning from human feedback. In *International Conference on Learning Representations (ICLR)*, 2024.

Christoph Dann, Yishay Mansour, and Mehryar Mohri. Reinforcement learning can be more efficient with multiple rewards. In *International Conference on Machine Learning (ICML)*, volume 202, pp. 6948–6967. PMLR, 2023.

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

Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 5th edition, 2019.

Yonathan Efroni, Ben Kretzu, Daniel R. Jiang, Jalaj Bhandari, Zheqing Zhu, and Karen Ullrich. Aligned multi objective optimization. In *International Conference on Machine Learning (ICML)*, volume 267, pp. 14989–15017. PMLR, 2025.

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

Robert Mansel Gower, Nicolas Loizou, Xun Qian, Alibek Sailanbayev, Egor Shulgin, and Peter Richtárik. SGD: General analysis and improved rates. In *International Conference on Machine Learning (ICML)*, volume 97, pp. 5200–5209. PMLR, 2019.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 645:633–638, 2025. doi: 10.1038/s41586-025-09422-z.

Yifei He, Shiji Zhou, Guojun Zhang, Hyokun Yun, Yi Xu, Belinda Zeng, Trishul Chilimbi, and Han Zhao. Robust multi-task learning with excess risks. In *International Conference on Machine Learning (ICML)*, volume 235, pp. 18094–18114. PMLR, 2024.

Ben Kretzu, Karen Ullrich, and Yonathan Efroni. Simple optimizers for convex aligned multi-objective optimization. *arXiv preprint arXiv:2509.05811*, 2025.

Seung Hyun Lee, Yinxiao Li, Junjie Ke, Innfarn Yoo, Han Zhang, Jiahui Yu, Qifei Wang, Fei Deng, Glenn Entis, Junfeng He, Gang Li, Sangpil Kim, Irfan Essa, and Feng Yang. Parrot: Pareto-optimal multi-reward reinforcement learning framework for text-to-image generation. In *European Conference on Computer Vision*, pp. 462–478. Springer, 2024.

Xiaoyu Li and Francesco Orabona. On the convergence of stochastic gradient descent with adaptive stepsizes. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89, pp. 983–992. PMLR, 2019.

Baijiong Lin and Yu Zhang. LibMTL: A Python library for deep multi-task learning. *Journal of Machine Learning Research*, 24(209):1–7, 2023.

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

Meitong Liu, Xiaoyuan Zhang, Chulin Xie, Kate Donahue, and Han Zhao. Adaptive online mirror descent for Tchebycheff scalarization in multi-objective learning. *Transactions on Machine Learning Research*, 2026. URL https://openreview.net/forum?id=MUTRffB3af.

Suyun Liu and Luís Nunes Vicente. The stochastic multi-gradient algorithm for multi-objective optimization and its application to supervised machine learning. *Annals of Operations Research*, 339:1119–1148, 2024.

Nicolas Loizou, Sharan Vaswani, Issam Hadj Laradji, and Simon Lacoste-Julien. Stochastic Polyak step-size for SGD: An adaptive learning rate for fast convergence. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130, pp. 1306–1314. PMLR, 2021.

Ulysse Marteau-Ferey, Dmitrii Ostrovskii, Francis Bach, and Alessandro Rudi. Beyond least-squares: Fast rates for regularized empirical risk minimization through self-concordance. In *Proceedings of the 32nd Conference on Learning Theory (COLT)*, volume 99, pp. 2294–2340. PMLR, 2019.

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

Yurii Nesterov. *Lectures on Convex Optimization*. Springer, 2nd edition, 2018. URL https://link.springer.com/book/10.1007/978-3-319-91578-4.

Antonio Orvieto, Simon Lacoste-Julien, and Nicolas Loizou. Dynamics of SGD with stochastic Polyak stepsizes: Truly adaptive variants and convergence to exact solution. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Alexandre Ramé, Guillaume Couairon, Corentin Dancette, Jean-Baptiste Gaya, Mustafa Shukor, Laure Soulier, and Matthieu Cord. Rewarded soups: towards Pareto-optimal alignment by interpolating weights fine-tuned on diverse rewards. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

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

Sharan Vaswani, Francis Bach, and Mark Schmidt. Fast and faster convergence of SGD for over-parameterized models and an accelerated perceptron. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89, pp. 1195–1204. PMLR, 2019.

Peiyao Xiao, Hao Ban, and Kaiyi Ji. Direction-oriented multi-objective learning: Simple and provable stochastic algorithms. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Mingjing Xu, Peizhong Ju, Jia Liu, and Haibo Yang. PSMGD: Periodic stochastic multi-gradient descent for fast multi-objective optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 21770–21778, 2025. doi: 10.1609/aaai.v39i20.35482. URL https://ojs.aaai.org/index.php/AAAI/article/view/35482.

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

Shiji Zhou, Wenpeng Zhang, Jiyan Jiang, Wenliang Zhong, Jinjie Gu, and Wenwu Zhu. On the convergence of stochastic multi-objective gradient manipulation and beyond. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

APPENDIX GUIDE

Main proof. Appendices A.1–A.3 give the general theorem, noisy selection certificate, dependent-update bound and conditional-energy proof. The last of these also proves finite-horizon exit control and the inner-entry guarantee. Shared geometry, measurable selection and random-time conditioning are recorded in Appendix C.

Applicability and comparisons. Appendices A.4–A.5 give an admissible nonquadratic model and the value/solver error conditions. Appendix A.6 proves the global convex guarantee; Appendices A.7–A.9 compare maximum-gap selection, original PAMOO and equal weighting.

Limits and experimental details. Appendix A.10 shows loss-selection stalling and why ordinary nonlinear MSE is not controlled. The motivating same-batch bias and noisy-cone counterexamples from the Introduction are in Appendix B. Appendix D contains the experimental details, and Appendix D.5 checks the certified constants and both counterexamples numerically.

A CONVERGENCE OF DUNAMOO

The proof follows the three questions in the main text: certify noisy selection, transfer it across the dependent update, and use local curvature to prove stochastic progress. Appendix A.2.2 establishes selection accuracy, and Lemma A.2 supplies the componentwise curvature bound; Appendix A.3 then proves the one-lag recursion, energy drift and stopping results. The general theorem below separates the update-noise and selection-error assumptions.

Population geometry	Basic DUNAMOO guarantee
Self-concordant, $\mu_{\mathrm{L}} > 0$	Expected entry if $\mathcal{E}_\infty < R^2$, localized second-moment bound (Theorem 1)
Quadratic, $\mu_{\mathrm{L}} > 0$	Ordinary global MSE to a residual (Theorem 1)
Smooth convex	Averaged maximum gap with selection residual (Corollary 1)

Table A.1: The basic guarantees under their stated oracle and step conditions.

A.1 GENERAL CONSTANT-STEP THEOREM

The general theorem permits any constant $0 < \eta \leq 1/\beta$ and any selector satisfying the population-gap bound below. Its energy $\mathcal{L}_k$, rate q, threshold R and residual $\mathcal{E}_\infty$ specialize to the main theorem at the basic step and batch size. Write e_i for a coordinate unit vector.

Update noise and selection accuracy. The general update condition bounds variance independently of selection accuracy. The history $\mathcal{F}_k$ determines x_k, w_k, and ξ denotes the fresh oracle randomness used to update them.

Assumption A.1 (Unbiased gradients with bounded variance)**.** The gradient oracles are measurable maps $g_i(x, \xi) \in \mathbb{R}^n$. For every x, they satisfy $\mathbb{E}_\xi g_i(x, \xi) = \nabla f_i(x)$, and for some $\sigma \geq 0$,

$$\sup_x \max_{i \in [m]} \mathbb{E}_\xi \|g_i(x, \xi) - \nabla f_i(x)\|^2 \leq \sigma^2.$$

The bound also holds for every history-fixed normalized mixture, by convexity of squared norm, and conversely simplex vertices recover the displayed condition. No cross-objective independence is required. Assumption A.1 applies uniformly to the update oracle used at each iteration. Write $\widehat{g}_{k,i} = g_i(x_k, \mathcal{B}_k)$ and $\widehat{J}_k = [\widehat{g}_{k,1}, \ldots, \widehat{g}_{k,m}]$. The parameter σ^2 bounds the variance of an update estimate, and the primitive sample-variance bound σ_{g}^2 yields $\sigma^2 = \sigma_{\mathrm{g}}^2/N$ for a fixed minibatch size N.

Selection accuracy is measured by a fixed retained fraction $0 < \chi \leq 1$ and a nonnegative, $\mathcal{F}_k$-measurable error allowance ε_k. The selector must satisfy

$$\mathbb{E}_k \Delta_{w_{k+1}}(x_k) \geq \chi M(x_k) - \varepsilon_k, \qquad 0 < \chi \leq 1, \quad \varepsilon_k \geq 0. \tag{A.1}$$

Thus the selected mixture retains a fraction of the largest population gap, up to additive error. For the basic QP Eq. (4), optimal values up to a common additive offset, fresh i.i.d. minibatches satisfying Eq. (2), and a constant step $0 < \eta\beta < 2$, Proposition A.1 gives

$$\chi = 1 - \frac{\eta\beta}{2}, \qquad \varepsilon_k = \frac{\nu_\Delta}{\sqrt{N_k}} + \frac{\eta\sigma_{\mathrm{g}}^2}{4N_k}. \tag{A.2}$$

The certificate permits a singular Gram and boundary solutions. Differential optimal-value errors and numerical solver residuals enter additional selection allowances in Appendix A.5.

General constant-step parameters. For $0 < \eta \le 1/\beta$ and the selector's certified fraction χ, define

$$\mathcal{L}_k = \frac{\mathbb{E}_k V_{k+1} + 2(\beta\eta)^2 V_k}{(1-\beta\eta)^2 + 2(\beta\eta)^2}, \qquad q = \frac{\eta\chi\mu_{\mathrm{L}}}{3\{1 + 2(\beta\eta)^2\}}. \tag{A.3}$$

For a uniform deterministic bound $0 \le \varepsilon_k \le \varepsilon < \infty$, set

$$\begin{aligned} R &= \frac{1}{M_{\mathrm{f}}\sqrt{\beta}} \sqrt{\frac{1 + 2(\beta\eta)^2}{(1-\beta\eta)^2 + 2(\beta\eta)^2}} \quad (M_{\mathrm{f}} > 0), \\ \mathcal{E}_\infty &= \frac{\eta^2\sigma^2\{1 + 2\beta\eta + 2(\beta\eta)^2 + 2q\} + \frac{4}{3}\eta\varepsilon}{\{(1-\beta\eta)^2 + 2(\beta\eta)^2\}q}. \end{aligned} \tag{A.4}$$

The residual is also defined when $M_{\mathrm{f}} = 0$, and only the threshold R is unnecessary. For the basic QP, $\eta = 1/(2\beta)$ gives $\chi = 3/4$ and the energy and rate in Eq. (5). Substituting its fixed-batch noise bounds gives Eq. (6). A different selector uses its own certified χ and allowance.

Theorem A.1 (General constant-step two-phase guarantee for certified selectors)**.** *Assume alignment, Assumptions 1, 2, A.1, and $\mu_{\mathrm{L}} > 0$. Run Algorithm 1 for all $k \ge 0$ with fixed x_0, $w_0 = \mathbf{1}/m$, constant $0 < \eta \le 1/\beta$, and any measurable next-weight selector on Δ_m satisfying Eq.* (A.1) *with fixed $0 < \chi \le 1$ and past-measurable $0 \le \varepsilon_k \le \varepsilon < \infty$. Use the parameters in Eq.* (A.3)–(A.4)*. For every $k \ge 0$,*

$$V_k \le \mathcal{L}_k \le 2V_k + \tfrac{3}{2}\eta^2\sigma^2. \tag{A.5}$$

For $M_{\mathrm{f}} > 0$, define $T = \inf\{k \ge 0 : \mathcal{L}_k \le R^2\}$ and $\gamma = q(R^2 - \mathcal{E}_\infty)/(2R)$. The two-branch inequalities in Eq. (8) *hold with the general energy and constants above, the far branch for every $k \ge 0$ whatever the sign of γ.*

(i) Finite expected entry. *If $\mathcal{E}_\infty < R^2$, then $T < \infty$ almost surely and*

$$\mathbb{E}T \le \frac{2R\sqrt{\mathcal{L}_0}}{q(R^2 - \mathcal{E}_\infty)}. \tag{A.6}$$

(ii) Localized second-moment bound up to and including the first exit. *On $\{T < \infty\}$, put $T_{\mathrm{exit}} = \inf\{j \ge T : \mathcal{L}_j > R^2\}$. For every integer $s \ge 0$,*

$$\mathbb{E}\big[V_{T+s}\mathbf{1}_{\{T_{\mathrm{exit}} \ge T+s\}} \mid \mathcal{F}_T\big] \le (1-q)^s\mathcal{L}_T + \mathcal{E}_\infty\{1 - (1-q)^s\}. \tag{A.7}$$

This conclusion holds whenever entry occurs. It does not require part (i)'s small-error condition and does not bound the contribution of trajectories that exited earlier.

Quadratics. *If $M_{\mathrm{f}} = 0$, the ordinary global MSE satisfies*

$$\mathbb{E}V_k \le (1-2q)^k\mathcal{L}_0 + \tfrac{1}{2}\mathcal{E}_\infty\{1 - (1-2q)^k\}. \tag{A.8}$$

Appendix A.3 proves this theorem and derives the main fixed-batch result.

Corollary A.1 (Variable minibatch sizes)**.** *Under the hypotheses of Theorem 1, replace the fixed batch size by deterministic sizes $N_k \ge N_{\min} \ge 1$, and write $\mathcal{E}_\infty(N)$ for the residual in Eq.* (6) *as a function of the batch size. With $\eta = 1/(2\beta)$, keep the energy, rate and threshold of that theorem.*

(a) All conclusions of Theorem 1 hold with the residual $\mathcal{E}_\infty(N_{\min})$. In its energy comparison, use $\sigma_{\mathrm{g}}^2/N_{\min}$ as the update-variance bound.

(b) If the sizes are nondecreasing, the residual sequence is nonincreasing, and it vanishes when $N_k \to \infty$. On $\{T < \infty\}$, for every integer $s \geq 0$,

$$\mathbb{E}\big[V_{T+s}\mathbf{1}_{\{T_{\mathrm{exit}} \geq T+s\}} \mid \mathcal{F}_T\big] \leq (1-q)^s \mathcal{L}_T + q\sum_{j<s}(1-q)^{s-1-j}\mathcal{E}_\infty(N_j),$$

and for $M_{\mathrm{f}} = 0$, $\mathbb{E}V_k \leq (1-2q)^k\mathcal{L}_0 + q\sum_{j<k}(1-2q)^{k-1-j}\mathcal{E}_\infty(N_j)$ for every $k \geq 0$. If $N_k \to \infty$, both right-hand sides tend to zero.

Proof. Proposition A.1 gives, for batch k, the update variance $\sigma_k^2 = \sigma_{\mathrm{g}}^2/N_k$, the fraction $\chi = 3/4$ and the allowance $\varepsilon_k = \nu_\Delta/\sqrt{N_k} + \sigma_{\mathrm{g}}^2/(8\beta N_k)$. Replacing N_k by $N_{\min}$ gives uniform bounds, substitution into Eq. (A.3)–(A.4) gives $\mathcal{E}_\infty(N_{\min})$, and Theorem A.1 proves (a).

For (b), retain the per-step variance and selection errors in the proof of Lemma A.6, keeping its abbreviations $t = \beta\eta$ and $c = 2t^2$. The additive term of the drift at index k collects $\eta^2\sigma_{k+1}^2$ from the descent at $k+1$, $\eta^2\sigma_k^2(2t+c+2q)$ from the movement, descent and comparison at k, and $\frac{4}{3}\eta\varepsilon_k$. Every term is nonincreasing in the batch size, so the additive term is at most $q\mathcal{E}_\infty(\min\{N_k, N_{k+1}\}) = q\mathcal{E}_\infty(N_k)$ for nondecreasing sizes.

Nondecreasing batch sizes also control errors after random entry. Since $T \geq 0$, $\mathcal{E}_\infty(N_{T+j}) \leq \mathcal{E}_\infty(N_j)$ for every j. The induction in the proof of part (ii) and the geometric summation in the quadratic clause then give the displayed bounds. A geometrically weighted sum of a bounded sequence tending to zero also tends to zero. □

Predictable update oracle. For the stored-weight update $g_k = \widehat{J}_k w_k$, predictability and Assumption A.1 give

$$\bar{g}_k = \nabla f_{w_k}(x_k), \quad \zeta_k = g_k - \bar{g}_k, \qquad \mathbb{E}_k\zeta_k = 0, \quad \mathbb{E}_k\|\zeta_k\|^2 \leq \sigma^2. \tag{A.9}$$

Current losses, gradients, Gram and next weight may be dependent. Fresh sampling restores this oracle property at the next step. The conditional energy is an analysis quantity that the algorithm does not compute.

Dependence on the number of objectives. The loss allowance admits $\nu_\Delta \leq \sqrt{m}\,\nu_f$ when ν_f bounds each centered sample loss in second moment, $\sup_{x,i}\mathbb{E}|\ell_i(x;\mathsf{Z}) - f_i(x)|^2 \leq \nu_f^2$ (Proposition A.1). Appendix A.9 gives the corresponding mixture-curvature comparison, and Appendix A.5 accounts for gradient, Gram and solver costs.

Probability of exit at a fixed lag. The localized second-moment bound controls the first-exit probability at a specified lag. The proof of part (ii) bounds $\mathbb{E}[\mathcal{L}_{T+s}\mathbf{1}_{\{T_{\mathrm{exit}} \geq T+s\}} \mid \mathcal{F}_T]$ by the right-hand side of Eq. (A.7), and $\{T_{\mathrm{exit}} = T+s\} \subseteq \{\mathcal{L}_{T+s} > R^2,\ T_{\mathrm{exit}} \geq T+s\}$, so Markov's inequality yields

$$\mathbb{P}(T_{\mathrm{exit}} = T+s \mid \mathcal{F}_T) \leq \frac{(1-q)^s\mathcal{L}_T + \mathcal{E}_\infty\{1-(1-q)^s\}}{R^2}.$$

This bounds a single exit time, not the probability of any exit before that time. Corollary A.2 supplies the finite-horizon bound, and Appendix A.10 explains why bounded variance alone does not guarantee permanent containment.

A.2 Selection guarantees and componentwise geometry

A.2.1 Smooth convex inequalities

Lemma A.1 (Smooth convex gap inequalities)**.** *Under alignment and Assumption 1, for every $w \in \Delta_m$ and $x \in \mathbb{R}^n$,*

$$\|\nabla f_w(x)\|^2 \leq 2\beta\Delta_w(x), \tag{A.10}$$

$$\langle \nabla f_w(x), x - x_\star\rangle \geq \Delta_w(x) + \frac{1}{2\beta}\|\nabla f_w(x)\|^2, \tag{A.11}$$

$$0 \leq \Delta_w(x) \leq \tfrac{\beta}{2}\|x - x_\star\|^2. \tag{A.12}$$

Proof. Lemma C.1 gives a convex, β-smooth mixture minimized at $x_\star$ with $\nabla f_w(x_\star) = 0$. The following descent and interpolation arguments are standard smooth-convex inequalities (Boyd & Vandenberghe, 2004, Chapters 3 and 9), proved here in the needed form. Applying its smoothness bound at $x - \nabla f_w(x)/\beta$ and using minimality of $x_\star$ gives (A.10). For (A.11), fix x and let $h(y) = f_w(y) - \langle \nabla f_w(x), y\rangle$. Since $\nabla h(x) = 0$, x minimizes h. A gradient step for h from $x_\star$ gives

$$h(x) \le h\big(x_\star - \nabla h(x_\star)/\beta\big) \le h(x_\star) - \frac{1}{2\beta}\|\nabla h(x_\star)\|^2.$$

Use $\nabla h(x_\star) = -\nabla f_w(x)$ and rearrange. Finally, alignment gives the lower bound in (A.12). Smoothness at $x_\star$ gives its upper bound. □

A.2.2 Selection accuracy without a bounded-iterate assumption

The primary moment condition Eq. (2) controls loss *contrasts*, because a perturbation shared by every loss adds a constant to the normalized-weight score and does not affect selection. The following proposition also covers a next update step chosen from the pre-batch history.

Proposition A.1 (Minibatch selection accuracy for DunAMOO)**.** *Assume alignment and Assumption 1. Use deterministic batch sizes N_k, i.i.d. joint observations fresh from $\mathcal{F}_k$, unbiased sample gradients, and the loss-contrast and gradient moment bounds in Eq.* (2)*. Supply the optimal values exactly or up to a common additive offset. In the exact normalized-weight QP of Algorithm 1, let η_{k+1} be $\mathcal{F}_k$-measurable and satisfy $0 < \eta_{k+1} \le \eta_{\max} < 2/\beta$ for a deterministic $\eta_{\max}$. Then*

$$\mathbb{E}_k \Delta_{w_{k+1}}(x_k) \ge \left(1 - \frac{\beta\eta_{k+1}}{2}\right) M(x_k) - \varepsilon_k, \qquad \varepsilon_k = \frac{\nu_\Delta}{\sqrt{N_k}} + \frac{\eta_{k+1}\sigma_{\mathrm{g}}^2}{4N_k}.$$

Consequently Eq. (A.1) *holds with the fixed fraction $\chi = 1 - \beta\eta_{\max}/2 > 0$ and the stated past-measurable ε_k. For constant steps $\eta_k = \eta$, taking $\eta_{\max} = \eta$ recovers exactly Eq.* (A.2)*, and at the basic step and fixed batch size it gives Eq.* (12)*. The stored-weight update variance is at most $\sigma_{\mathrm{g}}^2/N_k$. Hence $N_k \ge N_{\min} \ge 1$ supplies uniform bounds by replacing N_k with $N_{\min}$ and η_{k+1} with $\eta_{\max}$, without a trajectory-radius assumption. The stronger per-objective condition $\sup_{x,i} \mathbb{E}|\ell_i(x;\mathsf{Z}) - f_i(x)|^2 \le \nu_f^2$ is sufficient by taking $\nu_\Delta = \sqrt{m}\,\nu_f$.*

Proof. Condition on $\mathcal{F}_k$, write $x = x_k$ and $W = w_{k+1}$, and omit index k from hatted arrays within this proof. Choose the smallest index j maximizing $\Delta_i(x)$. Both this comparator and η_{k+1} are fixed under the conditioning. Let

$$\delta = \frac{1}{N_k}\sum_{s=1}^{N_k} (\ell_i(x;\mathsf{Z}_{k,s}) - f_i(x))_{i=1}^m, \qquad d = P\delta, \qquad P = I - \mathbf{1}\mathbf{1}^\top/m.$$

Unbiasedness and independent averaging give $\mathbb{E}_k d = 0$ and $\mathbb{E}_k\|d\|^2 \le \nu_\Delta^2/N_k$. The supplied gap vector is $\widehat{\Delta} = \Delta(x) + \delta + b\mathbf{1}$ for a common offset b, and this last term cancels from every feasible-direction comparison below, even if it depends on the batch.

Put $u = \widehat{J}W$ and $v = \widehat{g}_j$. First-order optimality of the differentiable concave score on the simplex gives

$$\langle 2\widehat{\Delta} - 2\eta_{k+1}\widehat{J}^\top\widehat{J}W, e_j - W\rangle \le 0.$$

Completing a square yields

$$\begin{aligned}\widehat{\Delta}_j - \widehat{\Delta}^\top W &\le \eta_{k+1}(\langle u, v\rangle - \|u\|^2) \\ &= \frac{\eta_{k+1}}{4}\|v\|^2 - \eta_{k+1}\|u - v/2\|^2 \le \frac{\eta_{k+1}}{4}\|v\|^2.\end{aligned} \tag{A.13}$$

Since $\mathbf{1}^\top(W - e_j) = 0$, the loss-error difference is $\delta^\top(W - e_j) = d^\top(W - e_j)$. The fixed comparator has $\mathbb{E}_k d_j = 0$, whereas the selected loss error must be retained. Consequently

$$\mathbb{E}_k \Delta_W(x) \ge M(x) - \frac{\eta_{k+1}}{4}\mathbb{E}_k\|\widehat{g}_j\|^2 - \mathbb{E}_k[d^\top W].$$

Smoothness and the sample-gradient moments give

$$\mathbb{E}_k\|\widehat{g}_j\|^2 \le \|\nabla f_j(x)\|^2 + \frac{\sigma_{\mathrm{g}}^2}{N_k} \le 2\beta M(x) + \frac{\sigma_{\mathrm{g}}^2}{N_k}.$$

Let $\|d\|_\infty = \max_i |d_i|$. Simplex feasibility and the contrast moment bound give, without independence between W and d,

$$\mathbb{E}_k[d^\top W] \le \mathbb{E}_k\|d\|_\infty \le \sqrt{\mathbb{E}_k\|d\|^2} \le \frac{\nu_\Delta}{\sqrt{N_k}}.$$

Substitution proves the stated conditional certificate. Since $M(x) \ge 0$ and $\eta_{k+1} \le \eta_{\max}$, replacing its fraction by $1 - \beta\eta_{\max}/2$ proves Eq. (A.1), and for constant steps this is Eq. (A.2). For the stored w_k, convexity of squared norm bounds the single-observation mixture variance by σ_{g}^2, and independent averaging divides it by N_k. Finally, P is an orthogonal projection, so the per-objective alternative implies

$$\mathbb{E}\|P(\ell_i(x;\mathsf{Z}) - f_i(x))_{i=1}^m\|^2 \le \sum_i \mathbb{E}|\ell_i(x;\mathsf{Z}) - f_i(x)|^2 \le m\nu_f^2.$$

This proves all claims. The proof permits dependent objectives, a singular Gram, boundary optimizers and discontinuous measurable selectors. It neither inverts the Gram nor assumes that the selected loss error is centered. Choosing the score coefficient after seeing the current batch would remove its measurability premise. □

What the loss-moment condition requires. Uniform individual loss variance is stronger than the contrast condition and is not implied by bounded gradient variance. For example, let Z take $+1$ and -1 with equal probabilities (a Rademacher variable). With $\ell(x;Z) = (x-Z)^2/2$, the centered gradient variance is one but the centered loss variance is x^2. A common centered perturbation across several objectives is removed by P, so it incurs no contrast error even when its individual variance grows with x. Objective-dependent loss contrasts still require their stated uniform moment bound. The per-objective ν_f condition remains a convenient sufficient condition for the explicit examples and batch budgets below.

Why entrywise Gram bounds are weaker. Let $e_{K,k}$ denote the largest absolute entry of the Gram error,

$$e_{K,k} = \max_{i,j} |(\widehat{J}_k^\top \widehat{J}_k - J(x_k)^\top J(x_k))_{ij}|.$$

Write $\Xi_k = \widehat{J}_k - J(x_k)$ with columns $\xi_{k,i}$, let $e = \max_i \|\xi_{k,i}\|$, and let $\|\cdot\|_F$ be the Frobenius norm. Each entry expands as $\langle \nabla f_i(x_k), \xi_{k,j}\rangle + \langle \xi_{k,i}, \nabla f_j(x_k)\rangle + \langle \xi_{k,i}, \xi_{k,j}\rangle$, and $\|\nabla f_i(x_k)\| \le \beta\|x_k - x_\star\|$, $\mathbb{E}_k e \le (\mathbb{E}_k\|\Xi_k\|_F^2)^{1/2} \le \sigma_{\mathrm{g}}\sqrt{m/N_k}$ and $\mathbb{E}_k e^2 \le m\sigma_{\mathrm{g}}^2/N_k$ give

$$\mathbb{E}_k e_{K,k} \le 2\beta\|x_k - x_\star\|\sigma_{\mathrm{g}}\sqrt{m/N_k} + m\sigma_{\mathrm{g}}^2/N_k.$$

This direct perturbation bound grows with the distance to the optimum. Proposition A.1 uses the fixed comparator in Eq. (A.13) to control the raw-Gram penalty through one gradient moment. This is the step that certifies the basic QP without a radius assumption. Appendix A.5 treats differential optimum-value errors and numerical solver residuals for the same QP.

A.2.3 Componentwise self-concordance and the local mixture curvature

Local curvature need not make every selected mixture strongly convex. We use a fixed comparison mixture at $x_\star$ only in the analysis. Integrating each component before mixing avoids the inverse minimum-weight factor that can arise from assigning a self-concordance constant to the weighted function itself.

Lemma A.2 (Local-curvature gap bound without minimum-weight dependence)**.** *Assume alignment, smoothness, self-concordance and $\mu_{\mathrm{L}} = \max_{v\in\Delta_m} \lambda_{\min}(\nabla^2 f_v(x_\star)) > 0$. Let $v_\star$ attain this maximum, $\rho = \|x - x_\star\|$, and $a_{\mathrm{P}} = M_{\mathrm{f}}\sqrt{\beta}$. Then*

$$M(x) \ge \Delta_{v_\star}(x) \ge \frac{\mu_{\mathrm{L}}\rho^2}{2(1 + a_{\mathrm{P}}\rho)}. \tag{A.14}$$

Proof. Lemma C.1 gives $v_\star$. We derive the self-concordant lower estimate directly from Eq. (1), following the differential calculus in Nesterov (2018, Chapter 5). Put $d = x - x_\star$ and $\omega_i(t) = d^\top\nabla^2 f_i(x_\star + td)d$ for $t \in [0,1]$. Convexity gives $\omega_i(t) \ge 0$, and self-concordance gives $|\omega_i'(t)| \le 2M_{\mathrm{f}}\omega_i(t)^{3/2}$. On every interval where $\omega_i > 0$, differentiation of $\omega_i^{-1/2}$ shows

$$\omega_i(t) \ge \frac{\omega_i(0)}{(1 + M_{\mathrm{f}} t\sqrt{\omega_i(0)})^2}.$$

For $\omega_i(0) > 0$, this lower bound prevents a finite-time first zero, so it holds on the entire segment. If $\omega_i(0) = 0$, the required lower bound is simply zero. Set $b_i = M_{\mathrm{f}}\sqrt{\omega_i(0)}$. For $M_{\mathrm{f}} > 0$, stationarity $\nabla f_i(x_\star) = 0$ and Taylor's integral formula give

$$\Delta_i(x) = \int_0^1 (1-t)\omega_i(t)\,dt \geq \frac{b_i - \log(1+b_i)}{M_{\mathrm{f}}^2} \geq \frac{\omega_i(0)}{2(1+b_i)}.$$

If $M_{\mathrm{f}} = 0$, Eq. (1) makes ω_i constant, and the integral is $\omega_i(0)/2$ directly. If $\omega_i(0) = 0$, use the nonnegative integral and the zero lower bound, without dividing by $\omega_i(0)$ or M_{f}. The elementary inequality $b - \log(1+b) \geq b^2/[2(1+b)]$ follows because the difference vanishes at zero and has derivative $b^2/[2(1+b)^2] \geq 0$. Smoothness gives $b_i \leq a_{\mathrm{P}}\rho$. Multiplying the individual lower bounds by $v_{\star,i}$ and summing gives

$$\Delta_{v_\star}(x) \geq \frac{d^\top \nabla^2 f_{v_\star}(x_\star) d}{2(1+a_{\mathrm{P}}\rho)} \geq \frac{\mu_{\mathrm{L}}\rho^2}{2(1+a_{\mathrm{P}}\rho)}.$$

Finally $M(x) \geq \Delta_{v_\star}(x)$. Integrating componentwise before mixing preserves μ_{L} without a minimum-weight or extra task-count loss in this geometric inequality. The stochastic selection allowance still depends on m. □

Lemma C.1 also proves uniqueness of the common minimizer.

Smoothness and positive curvature do not imply classical self-concordance. The convex function $f(x) = \log(\cosh x)$ is globally 1-smooth and has positive curvature at zero. However, its self-concordance ratio is unbounded, since

$$\frac{|f'''(x)|}{2f''(x)^{3/2}} = |\sinh x|.$$

Thus it fails Assumption 2, which the global convex corollary does not require.

A.3 Two-phase convergence with complete dependence and stopping proofs

For this proof, abbreviate $A_k = \Delta_{w_k}(x_k)$, $d_k = x_k - x_\star$ and $h_k = x_{k+1} - x_k$. Recall that $V_k = \|d_k\|^2$. We first establish the shared one-lag inequality, then prove Theorem A.1 and derive the main minibatch theorem. The exact-feedback limit and its comparison with PAMOO follow the stochastic proof (Appendix A.8).

A.3.1 Predictability, movement energy and the exact one-lag inequality

Lemma A.3 (Predictable descent and movement)**.** *Under alignment, global smooth convexity and* (A.9)*, for* $0 < \eta_k \leq 1/\beta$*,*

$$\mathbb{E}_k V_{k+1} \leq V_k - 2\eta_k A_k + \eta_k^2 \sigma^2, \tag{A.15}$$

$$\mathbb{E}_k \|h_k\|^2 \leq 2\beta\eta_k^2 A_k + \eta_k^2 \sigma^2. \tag{A.16}$$

Proof. Expanding the unprojected squared distance and conditioning on $\mathcal{F}_k$ gives

$$\mathbb{E}_k V_{k+1} = V_k - 2\eta_k \langle \bar{g}_k, d_k \rangle + \eta_k^2 \|\bar{g}_k\|^2 + \eta_k^2 \mathbb{E}_k \|\zeta_k\|^2.$$

Use (A.11) to upper bound this by $V_k - 2\eta_k A_k + (\eta_k^2 - \eta_k/\beta)\|\bar{g}_k\|^2 + \eta_k^2\sigma^2$. The coefficient of $\|\bar{g}_k\|^2$ is nonpositive. For movement, conditional mean zero and (A.10) give $\mathbb{E}_k\|h_k\|^2 = \eta_k^2(\|\bar{g}_k\|^2 + \mathbb{E}_k\|\zeta_k\|^2)$. □

Lemma A.4 (Pathwise transport of a selected gap)**.** *Under alignment and Assumption 1, for every* $x, y \in \mathbb{R}^n$ *and every* $w \in \Delta_m$*,*

$$\Delta_w(y) \geq \tfrac{2}{3}\Delta_w(x) - \beta\|y - x\|^2.$$

Proof. Smoothness applied at y, Cauchy–Schwarz and (A.10) give

$$\begin{aligned}\Delta_w(x) &\leq \Delta_w(y) + \sqrt{2\beta\Delta_w(y)}\|x - y\| + \tfrac{\beta}{2}\|x - y\|^2 \\ &\leq \tfrac{3}{2}\Delta_w(y) + \tfrac{3\beta}{2}\|x - y\|^2.\end{aligned}$$

The last step uses Young's inequality with $\sqrt{2\beta\Delta_w(y)}\|x-y\| \le \Delta_w(y)/2+\beta\|x-y\|^2$. Rearranging proves the claim. The statement is deterministic for every triple (x, y, w) and thus also holds for jointly random triples. □

For the one-dimensional quadratic $f(u) = \beta u^2/2$, $x = 3$ and $y = 2$ give equality, so the factor $2/3$ is sharp with movement coefficient β fixed. In particular $w = w_{k+1}$ and $y = x_{k+1}$ may be functions of the same batch. The estimate accommodates an arbitrary weight change.

Lemma A.5 (One-lag inequality with exact coefficients)**.** *Assume alignment, Assumption 1, and the predictable update-oracle conditions Eq.* (A.9) *at each step of the unprojected normalized-weight recursion. For deterministic steps* $0 < \eta_k, \eta_{k+1} \le 1/\beta$*, put*

$$c_k = 2\beta^2\eta_k\eta_{k+1}, \qquad B_k = \sigma^2\big(\eta_{k+1}^2 + 2\beta\eta_{k+1}\eta_k^2 + c_k\eta_k^2\big).$$

Then

$$\mathbb{E}_k V_{k+2} \le (1-c_k)\mathbb{E}_k V_{k+1} + c_k V_k - \tfrac{4}{3}\eta_{k+1}\mathbb{E}_k\Delta_{w_{k+1}}(x_k) + B_k. \tag{A.17}$$

Proof. Apply (A.15) at $k+1$ and the tower property. Fresh-noise unbiasedness is applied at the pre-batch sigma-field $\mathcal{F}_{k+1}$. Lemma A.4 then gives

$$\begin{aligned}\mathbb{E}_k V_{k+2} \le \mathbb{E}_k V_{k+1} &- \tfrac{4}{3}\eta_{k+1}\mathbb{E}_k\Delta_{w_{k+1}}(x_k)\\ &+ 2\beta\eta_{k+1}\mathbb{E}_k\|h_k\|^2 + \eta_{k+1}^2\sigma^2.\end{aligned}$$

By (A.16), the movement term is at most $4\beta^2\eta_{k+1}\eta_k^2 A_k + 2\beta\eta_{k+1}\eta_k^2\sigma^2$. Rearranging the previous step's descent gives

$$2\eta_k A_k \le V_k - \mathbb{E}_k V_{k+1} + \eta_k^2\sigma^2.$$

Multiplication by the nonnegative coefficient $c_k = 2\beta^2\eta_k\eta_{k+1}$ supplies precisely the required upper bound on $4\beta^2\eta_{k+1}\eta_k^2 A_k$. Collect terms to obtain (A.17). This derivation remains valid when $1 - c_k < 0$ because it multiplies the descent inequality only by $c_k \ge 0$. □

The key accounting is that movement energy is controlled by the preceding step's decrease. Bounding it only by $\beta^2\eta_k^2 V_k$ would unnecessarily trade step size against the structural curvature.

A.3.2 The conditional energy process

Fix deterministic x_0, $w_0 = \mathbf{1}/m$, and $0 < \eta \le 1/\beta$. Independent initialization can instead be conditioned on before the run. A process is adapted here if its value at time k is $\mathcal{F}_k$-measurable. Fresh-batch independence and conditional-expectation rules give versions of the following inequalities that hold almost surely at every integer time (Durrett, 2019, Section 4.1). Nonnegativity and the descent bound imply finite second moments at each finite time.

Lemma A.6 (Adapted comparison and global nonlinear drift)**.** *Under Theorem A.1's hypotheses,* $\mathcal{L}_k$ *is nonnegative, adapted and integrable, satisfies Eq.* (A.5)*, and has* $0 < 2q < 1$*. If* $M_{\mathrm{f}} > 0$*, then*

$$\mathbb{E}_k\mathcal{L}_{k+1} \le \mathcal{L}_k - \frac{2q\mathcal{L}_k}{1+\sqrt{\mathcal{L}_k}/R} + q\mathcal{E}_\infty. \tag{A.18}$$

If $M_{\mathrm{f}} = 0$*, the denominator is one and the same inequality holds globally. At the basic step, this gives Eq.* (14)*.*

Proof. Only within this drift calculation, abbreviate

$$t = \beta\eta, \qquad c = 2t^2, \qquad c_+ = 1 + c, \qquad c_- = (1-t)^2 + c, \qquad v_\eta = \eta^2\sigma^2,$$
$$Z_k = \mathbb{E}_k V_{k+1} + cV_k = c_-\mathcal{L}_k, \qquad a_{\mathrm{P}} = M_{\mathrm{f}}\sqrt{\beta}.$$

Write $\bar{g}_k = \nabla f_{w_k}(x_k)$ and $\zeta_k = g_k - \bar{g}_k$. Since w_k is predictable, $\mathbb{E}_k\zeta_k = 0$ and $\mathbb{E}_k\|\zeta_k\|^2 \le \sigma^2$. The unprojected update gives

$$\mathbb{E}_k V_{k+1} = \|x_k - x_\star - \eta\bar{g}_k\|^2 + \eta^2\mathbb{E}_k\|\zeta_k\|^2. \tag{A.19}$$

Smoothness and alignment imply $\|\bar{g}_k\| \le \beta\|x_k - x_\star\|$. The reverse triangle inequality, with $t \le 1$, gives $\mathbb{E}_k V_{k+1} \ge (1-t)^2 V_k$. Lemma A.3 supplies the upper comparison, so

$$c_- V_k \le Z_k \le c_+ V_k + v_\eta.$$

The identities $c_+ - c_- = 2t - t^2 \geq 0$, $2c_- - c_+ = (1-2t)^2 \geq 0$ and $c_- \geq 2/3$ give $1 \leq c_+/c_- \leq 2$. Dividing the comparison by c_- proves Eq. (A.5). Also $\mu_{\mathrm{L}} \leq \beta$ and $\chi \leq 1$ imply $0 < 2q \leq 2t/[3(1+2t^2)] < 1$.

For constant steps, Lemma A.5, the selection certificate and Lemma A.2 give

$$\begin{aligned}\mathbb{E}_k Z_{k+1} &= \mathbb{E}_k V_{k+2} + c\mathbb{E}_k V_{k+1} \\ &\leq Z_k - \tfrac{4}{3}\eta \mathbb{E}_k \Delta_{w_{k+1}}(x_k) + v_\eta(1+2t+c) \\ &\leq Z_k - \frac{2\eta\chi\mu_{\mathrm{L}}}{3}\frac{V_k}{1+a_{\mathrm{P}}\sqrt{V_k}} + \tfrac{4}{3}\eta\varepsilon + v_\eta(1+2t+c). \end{aligned} \tag{A.20}$$

For $v \geq 0$, set $H(v) = v/(1+a_{\mathrm{P}}\sqrt{v})$. This function is increasing and 1-Lipschitz, because

$$H'(v) = \frac{1+a_{\mathrm{P}}\sqrt{v}/2}{(1+a_{\mathrm{P}}\sqrt{v})^2} \in (0,1] \quad (v>0), \qquad H'(0) = 1.$$

With $(z)_+ = \max\{z, 0\}$, the upper comparison gives $V_k \geq (Z_k - v_\eta)_+/c_+$. Hence

$$\begin{aligned} H(V_k) &\geq H((Z_k - v_\eta)_+/c_+) \\ &\geq H(Z_k/c_+) - v_\eta/c_+ \\ &= \frac{Z_k}{c_+\{1+a_{\mathrm{P}}\sqrt{Z_k/c_+}\}} - \frac{v_\eta}{c_+}. \end{aligned}$$

Since $2\eta\chi\mu_{\mathrm{L}}/(3c_+) = 2q$, substitution into Eq. (A.20) yields

$$\mathbb{E}_k Z_{k+1} \leq Z_k - \frac{2qZ_k}{1+a_{\mathrm{P}}\sqrt{Z_k/c_+}} + v_\eta(1+2t+c+2q) + \tfrac{4}{3}\eta\varepsilon.$$

Divide by c_-. For $M_{\mathrm{f}} > 0$, the denominator becomes $1 + \sqrt{\mathcal{L}_k}/R$, and Eq. (A.4) makes the additive term exactly $q\mathcal{E}_\infty$. For $M_{\mathrm{f}} = 0$, $a_{\mathrm{P}} = 0$ and the denominator is one. This proves Eq. (A.18). By definition, $\mathcal{L}_k$ is $\mathcal{F}_k$-measurable; the upper comparison and finite-time descent ensure integrability. □

The proof closes the drift for the conditional energy itself. For the auxiliary function H in the preceding proof, $\mathbb{E}H(V_k) \leq H(\mathbb{E}V_k)$ by concavity, which is the wrong direction for replacing a negative nonlinear drift by a function of the MSE. The lower comparison in Eq. (A.19) also uses the unprojected update in Algorithm 1.

A.3.3 Far-phase stopping and finite expected entry

Proof of Theorem A.1, far branch and part (i). Assume $M_{\mathrm{f}} > 0$ and write $\gamma = q(R^2 - \mathcal{E}_\infty)/(2R)$, which may have either sign at this point. On $\{\mathcal{L}_k > R^2\}$, conditional Jensen and the tangent bound for the square root give

$$\begin{aligned}\mathbb{E}_k\sqrt{\mathcal{L}_{k+1}} &\leq \sqrt{\mathcal{L}_k - \frac{2q\mathcal{L}_k}{1+\sqrt{\mathcal{L}_k}/R} + q\mathcal{E}_\infty} \\ &\leq \sqrt{\mathcal{L}_k} - \frac{q\sqrt{\mathcal{L}_k}}{1+\sqrt{\mathcal{L}_k}/R} + \frac{q\mathcal{E}_\infty}{2\sqrt{\mathcal{L}_k}} \\ &\leq \sqrt{\mathcal{L}_k} - \gamma. \end{aligned}$$

The square-root argument is nonnegative since $2q < 1$. The last line uses $\sqrt{\mathcal{L}_k}/(1+\sqrt{\mathcal{L}_k}/R) \geq R/2$ and $1/\sqrt{\mathcal{L}_k} \leq 1/R$, and it does not need the small-error condition.

Adaptedness makes $T = \inf\{k : \mathcal{L}_k \leq R^2\}$ a stopping time. The energy and its bounded-time stopped values are integrable by Lemma C.3. On $\{T > k\}$, the displayed drift includes possible entry at the next step. On $\{T \leq k\}$, the stopped value stays fixed. Thus

$$\mathbb{E}_k\sqrt{\mathcal{L}_{(k+1)\wedge T}} \leq \sqrt{\mathcal{L}_{k\wedge T}} - \gamma \mathbf{1}_{\{T>k\}}.$$

For every integer $K \geq 0$, summation gives

$$\mathbb{E}\sqrt{\mathcal{L}_{K\wedge T}} \leq \sqrt{\mathcal{L}_0} - \gamma\,\mathbb{E}(K \wedge T),$$

using $\sum_{k<K} \Pr(T > k) = \mathbb{E}(K \wedge T)$. This is the far branch of Eq. (8) with the general constants, valid for every $K \geq 0$ and for either sign of γ.

Now assume $\mathcal{E}_\infty < R^2$, so that $\gamma > 0$. Nonnegativity of the left side bounds $\mathbb{E}(K \wedge T)$ by $\sqrt{\mathcal{L}_0}/\gamma$. Monotone convergence then yields

$$\mathbb{E}T \leq \frac{2R\sqrt{\mathcal{L}_0}}{q(R^2 - \mathcal{E}_\infty)} < \infty,$$

proving Eq. (A.6) and $T < \infty$ almost surely. If entry holds at initialization, $T = 0$ and the stopped bound is immediate. For $K \geq 1$, Markov's inequality also gives

$$\Pr(T > K) \leq \min\left\{1, \frac{\sqrt{\mathcal{L}_0}}{\gamma K}\right\}.$$

□

The positive-noise numerator is $\sqrt{\mathcal{L}_0}$ because the first-entry energy can lie anywhere below the threshold. The exact-feedback corollary obtains a sharper deterministic entry bound.

A.3.4 Near-phase localization and the first exit

Proof of Theorem A.1, part (ii). We work on $\{T < \infty\}$, without requiring part (i)'s small-error condition. Define $T_{\text{exit}} = \inf\{j \geq T : \mathcal{L}_j > R^2\}$, so that necessarily $T_{\text{exit}} > T$. For integer $n \geq 0$, first restrict to $\{T \leq n\}$. For every $s \geq 0$, the entry and shifted times are then bounded by $n + s + 1$, so Lemma C.3 justifies the conditional inequalities and tower property. Put

$$Y_s^{(n)} = \mathbb{E}\big[\mathcal{L}_{T+s}\mathbf{1}_{\{T\leq n\}}\mathbf{1}_{\{T_{\text{exit}}\geq T+s\}} \mid \mathcal{F}_T\big],$$

where the integrand is defined to be zero on $\{T = \infty\}$. Since $\{T \leq n\} \in \mathcal{F}_T$ and $\mathcal{L}_T \leq R^2$ on entry, $Y_0^{(n)} = \mathcal{L}_T\mathbf{1}_{\{T\leq n\}}$ is bounded.

To control $Y_{s+1}^{(n)}$, use the predictable event

$$\{T_{\text{exit}} \geq T + s + 1\} = \{T_{\text{exit}} > T + s\}.$$

On this event, $\mathcal{L}_{T+s} \leq R^2$, and the drift gives

$$\mathbb{E}[\mathcal{L}_{T+s+1} \mid \mathcal{F}_{T+s}] \leq (1-q)\mathcal{L}_{T+s} + q\mathcal{E}_\infty.$$

Consequently,

$$\begin{aligned} Y_{s+1}^{(n)} &\leq (1-q)\mathbb{E}\big[\mathcal{L}_{T+s}\mathbf{1}_{\{T\leq n\}}\mathbf{1}_{\{T_{\text{exit}}>T+s\}} \mid \mathcal{F}_T\big] \\ &\quad + q\mathcal{E}_\infty \Pr(T \leq n,\ T_{\text{exit}} > T + s \mid \mathcal{F}_T) \\ &\leq (1-q)Y_s^{(n)} + q\mathcal{E}_\infty\mathbf{1}_{\{T\leq n\}}. \end{aligned}$$

Here the strict survival event is a subset of $\{T_{\text{exit}} \geq T + s\}$, and $1 - q > 0$. Induction and geometric summation give

$$Y_s^{(n)} \leq \mathbf{1}_{\{T\leq n\}}\left[(1-q)^s\mathcal{L}_T + \mathcal{E}_\infty\{1 - (1-q)^s\}\right].$$

Conditional Fatou's lemma as $n \to \infty$ yields, on $\{T < \infty\}$,

$$\mathbb{E}\big[\mathcal{L}_{T+s}\mathbf{1}_{\{T_{\text{exit}}\geq T+s\}} \mid \mathcal{F}_T\big] \leq (1-q)^s\mathcal{L}_T + \mathcal{E}_\infty\{1 - (1-q)^s\}.$$

This argument establishes finite conditional bounds without assuming integrability of an arbitrary unbounded stopped value. Since $V_{T+s} \leq \mathcal{L}_{T+s}$, it proves Eq. (A.7).

The first-exit term $T_{\text{exit}} = T + s$ is included. The expression is therefore distinct from MSE conditioned on survival through the current point. Ordinary conditional MSE additionally contains $\mathbb{E}[V_{T+s}\mathbf{1}_{\{T_{\text{exit}}<T+s\}} \mid \mathcal{F}_T]$. □

The argument conditions only on histories and uses predictable survival events, and it never conditions the oracle on future survival. The localized bound also applies to runs that enter when part (i)'s sufficient condition is unavailable.

A.3.5 FINITE-HORIZON EXIT PROBABILITY AND CONDITIONAL ERROR

The localized second moment need not describe the surviving paths by itself, because its size also depends on their probability. The following consequence gives that probability a finite-horizon lower bound.

Corollary A.2 (Finite-horizon exit control)**.** *Under Theorem A.1 with $M_{\mathrm{f}} > 0$, work on $\{T < \infty\}$. For every integer $h \geq 0$, set $\alpha_h = (\mathcal{L}_T + q\mathcal{E}_\infty h)/R^2$. Then*

$$\mathbb{P}(T_{\mathrm{exit}} \leq T + h \mid \mathcal{F}_T) \leq \min\{1, \alpha_h\}. \tag{A.21}$$

On histories with $\alpha_h < 1$, the MSE conditional on strict survival satisfies

$$\mathbb{E}[V_{T+h} \mid \mathcal{F}_T,\ T_{\mathrm{exit}} > T + h] \leq \frac{(1-q)^h \mathcal{L}_T + \mathcal{E}_\infty\{1 - (1-q)^h\}}{1 - \alpha_h}. \tag{A.22}$$

Here conditioning on strict survival means dividing the conditional moment with indicator $\mathbf{1}_{\{T_{\mathrm{exit}} > T+h\}}$ by its conditional probability. The same statements hold with the basic constants of Theorem 1.

Proof. Put $S = T_{\mathrm{exit}} - T$ and $U_j = \mathcal{L}_{T+(j\wedge S)}$, with $j \wedge \infty = j$. Before exit, Lemma A.6 gives $\mathbb{E}[\mathcal{L}_{T+j+1} \mid \mathcal{F}_{T+j}] \leq \mathcal{L}_{T+j} + q\mathcal{E}_\infty$. After exit the stopped value stays fixed, so

$$\mathbb{E}[U_{j+1} \mid \mathcal{F}_{T+j}] \leq U_j + q\mathcal{E}_\infty \mathbf{1}_{\{S>j\}}.$$

For unbounded T, first multiply by $\mathbf{1}_{\{T\leq n\}}$ and use Lemma C.3. Induction and the tower property give

$$\mathbb{E}[U_h \mathbf{1}_{\{T\leq n\}} \mid \mathcal{F}_T] \leq \mathbf{1}_{\{T\leq n\}}(\mathcal{L}_T + q\mathcal{E}_\infty h).$$

Conditional monotone convergence as $n \to \infty$ supplies the same bound on $\{T < \infty\}$, with finite right-hand side because $\mathcal{L}_T \leq R^2$. On $\{S \leq h\}$, $U_h = \mathcal{L}_{T_{\mathrm{exit}}} > R^2$. Conditional Markov's inequality proves Eq. (A.21). If $\alpha_h < 1$, strict survival has conditional probability at least $1 - \alpha_h > 0$. Its moment numerator is at most the through-first-exit moment in Eq. (A.7), and dividing proves Eq. (A.22). □

Entry near the boundary can make the exit bound trivial. The next result first waits for entry into a smaller energy region, providing explicit room below the outer threshold. This is an analysis-only stopping time, not a new algorithmic rule.

Corollary A.3 (Inner entry and finite-horizon survival)**.** *Under Theorem 1, suppose $M_{\mathrm{f}} > 0$ and $\mathcal{E}_\infty < R^2$. Choose $0 < r < R$ such that*

$$\delta_r := q\left(\frac{r}{1 + r/R} - \frac{\mathcal{E}_\infty}{2r}\right) > 0.$$

Define

$$S_r = \inf\{k \geq 0 : \mathcal{L}_k \leq r^2\}, \qquad \tau_r = \inf\{j \geq S_r : \mathcal{L}_j > R^2\}.$$

Then $S_r < \infty$ almost surely and $\mathbb{E}S_r \leq \sqrt{\mathcal{L}_0}/\delta_r$. For every integer $h \geq 0$, almost surely,

$$\begin{aligned} \mathbb{E}[V_{S_r+h}\mathbf{1}_{\{\tau_r \geq S_r+h\}} \mid \mathcal{F}_{S_r}] &\leq (1-q)^h r^2 + \mathcal{E}_\infty\{1 - (1-q)^h\}, \\ \mathbb{P}(\tau_r \leq S_r + h \mid \mathcal{F}_{S_r}) &\leq \min\{1, \alpha_h\}, \qquad \alpha_h := \frac{r^2 + q\mathcal{E}_\infty h}{R^2}. \end{aligned} \tag{A.23}$$

When $\alpha_h < 1$, the MSE conditioned on strict survival satisfies

$$\frac{\mathbb{E}[V_{S_r+h}\mathbf{1}_{\{\tau_r > S_r+h\}} \mid \mathcal{F}_{S_r}]}{\mathbb{P}(\tau_r > S_r + h \mid \mathcal{F}_{S_r})} \leq \frac{(1-q)^h r^2 + \mathcal{E}_\infty\{1 - (1-q)^h\}}{1 - \alpha_h}.$$

The denominator on the left is at least $1 - \alpha_h > 0$.

Proof. **Inner entry.** The drift in Eq. (14) holds globally. For $\mathcal{L}_k > 0$, the tangent inequality for the square root gives

$$\mathbb{E}_k\sqrt{\mathcal{L}_{k+1}} \leq \sqrt{\mathcal{L}_k} + \frac{\mathbb{E}_k\mathcal{L}_{k+1} - \mathcal{L}_k}{2\sqrt{\mathcal{L}_k}} \leq \sqrt{\mathcal{L}_k} - q\left(\frac{\sqrt{\mathcal{L}_k}}{1 + \sqrt{\mathcal{L}_k}/R} - \frac{\mathcal{E}_\infty}{2\sqrt{\mathcal{L}_k}}\right).$$

For $u > 0$, the derivative of $u/(1+u/R) - \mathcal{E}_\infty/(2u)$ is $(1+u/R)^{-2} + \mathcal{E}_\infty/(2u^2) > 0$. Thus the expected decrease before S_r is at least δ_r. Stopping and summing over deterministic times yields

$$\mathbb{E}\sqrt{\mathcal{L}_{k\wedge S_r}} \le \sqrt{\mathcal{L}_0} - \delta_r \mathbb{E}(k \wedge S_r).$$

Finite-time integrability follows from Lemma C.3. Nonnegativity and monotone convergence as $k \to \infty$ give the bound on $\mathbb{E}S_r$ and hence almost-sure finite entry.

Post-entry bounds. Let $\mathcal{G}_j = \mathcal{F}_{S_r+j}$, and define the inclusive localized and stopped energies

$$Z_j = \mathcal{L}_{S_r+j}\mathbf{1}_{\{\tau_r \ge S_r+j\}}, \qquad Y_j = \mathcal{L}_{(S_r+j)\wedge\tau_r}.$$

On $\{\tau_r > S_r + j\}$, the energy is at most R^2, so the global drift implies $\mathbb{E}[\mathcal{L}_{S_r+j+1} \mid \mathcal{G}_j] \le (1-q)\mathcal{L}_{S_r+j} + q\mathcal{E}_\infty$. This random-time inequality follows by decomposing over $\{S_r = t\}$, as in Lemma C.3. Since $\{\tau_r \ge S_r + j + 1\} = \{\tau_r > S_r + j\} \in \mathcal{G}_j$ and $1 - q \ge 0$,

$$\mathbb{E}[Z_{j+1} \mid \mathcal{G}_j] \le (1-q)Z_j + q\mathcal{E}_\infty, \qquad \mathbb{E}[Y_{j+1} \mid \mathcal{G}_j] \le Y_j + q\mathcal{E}_\infty.$$

The second inequality discards the nonpositive descent before exit and uses the constant stopped value afterwards. To justify iteration without assuming integrability at an unbounded stopping time, first multiply both recursions by $\mathbf{1}_{\{S_r \le n\}}$. For fixed n, j, all relevant times are bounded by $n + j + 1$, so the tower property applies by Lemma C.3. Using $Z_0 = Y_0 = \mathcal{L}_{S_r} \le r^2$, geometric summation and ordinary summation give

$$\mathbb{E}[Z_h\mathbf{1}_{\{S_r\le n\}} \mid \mathcal{F}_{S_r}] \le \mathbf{1}_{\{S_r\le n\}}\big[(1-q)^h r^2 + \mathcal{E}_\infty\{1-(1-q)^h\}\big],$$
$$\mathbb{E}[Y_h\mathbf{1}_{\{S_r\le n\}} \mid \mathcal{F}_{S_r}] \le \mathbf{1}_{\{S_r\le n\}}(r^2 + q\mathcal{E}_\infty h).$$

Conditional monotone convergence as $n \to \infty$ removes the truncation, with finite right-hand sides. Now $V_k \le \mathcal{L}_k$ proves the first bound in Eq. (A.23). On $\{\tau_r \le S_r + h\}$, $Y_h = \mathcal{L}_{\tau_r} > R^2$; conditional Markov's inequality proves the second. Finally, strict survival is a subset of the inclusive event. Its moment is therefore no larger than the first bound, and its probability is at least $1 - \alpha_h$. Division proves the conditional-MSE claim. □

Choosing the inner region and horizon. The condition $\delta_r > 0$ is equivalent to $2r^2 > \mathcal{E}_\infty(1 + r/R)$, or

$$r > \frac{\mathcal{E}_\infty/R + \sqrt{(\mathcal{E}_\infty/R)^2 + 8\mathcal{E}_\infty}}{4}.$$

The lower endpoint is below R because $\mathcal{E}_\infty < R^2$, so an inner radius exists. A nontrivial survival-conditioned bound additionally requires $\alpha_h < 1$, equivalently $r^2 + q\mathcal{E}_\infty h < R^2$. This need not hold for any positive integer h when noise is large. Neither corollary establishes permanent containment or global ordinary MSE. Exact feedback instead has the invariance in Corollary A.4.

A.3.6 Quadratic objectives

Proof of Theorem A.1, quadratic clause. When $M_{\mathrm{f}} = 0$, Lemma A.6 gives the global recurrence

$$\mathbb{E}_k\mathcal{L}_{k+1} \le (1-2q)\mathcal{L}_k + q\mathcal{E}_\infty.$$

Taking expectations and summing the geometric recurrence yields

$$\mathbb{E}\mathcal{L}_k \le (1-2q)^k\mathcal{L}_0 + \tfrac{1}{2}\mathcal{E}_\infty\{1-(1-2q)^k\}.$$

Use $V_k \le \mathcal{L}_k$ to obtain Eq. (A.8). The recurrence holds on the full space and therefore controls the error after exits as well. □

A.3.7 The basic fixed-batch theorem

Proof of Theorem 1. For the fixed minibatch size N, Proposition A.1 supplies Assumption A.1 with $\sigma^2 = \sigma_{\mathrm{g}}^2/N$. At $\eta = 1/(2\beta)$, its selection certificate is Eq. (A.1) with

$$\chi = \frac{3}{4}, \qquad \varepsilon = \frac{\nu_\Delta}{\sqrt{N}} + \frac{\sigma_{\mathrm{g}}^2}{8\beta N}.$$

Fresh sampling and the stored weight give Eq. (A.9), so Theorem A.1 applies. Substituting these values into Eq. (A.3)–(A.4) gives Eq. (5)–(6), including the stated residual. Thus Eq. (A.5)–(A.8) give exactly Eq. (7), (9), (8) and Eq. (10), with the same entry and exit events. The stopped square-root bound in the proof of part (i) gives the far branch of Eq. (8) with $\gamma = q(R^2 - \mathcal{E}_\infty)/(2R)$, and the localization argument gives the near branch. Corollary A.1 records the variable-batch version. □

A.4 A NONQUADRATIC MODEL WITH LOCAL MIXTURE CURVATURE

This example has positive local mixture curvature but zero global adaptive curvature, defined here by $\mu_{\mathrm{G}} = \inf_x \max_{w\in\Delta_m} \lambda_{\min}(\nabla^2 f_w(x))$. For $m \geq 2$, $s \geq 0$ and $x \in \mathbb{R}^m$, define

$$\phi(t) = t\arctan t - \tfrac{1}{2}\log(1+t^2), \qquad f_i(x) = \phi(x_i), \qquad \ell_i(x;Z) = f_i(x) + sZ_i \sin x_i,$$

where the coordinates of Z are independent Rademacher random variables and observations are i.i.d. across the batch. This provides a nonquadratic instance with known $x_\star = 0$ and $f_i^\star = 0$, global smoothness, local mixture curvature and uniform noisy-oracle moments. The derivatives are

$$\phi'(t) = \arctan t, \qquad \phi''(t) = \frac{1}{1+t^2}, \qquad \phi'''(t) = -\frac{2t}{(1+t^2)^2}.$$

Thus ϕ is convex, has its unique minimum at zero, and is globally 1-smooth. Moreover,

$$\frac{|\phi'''(t)|}{2\phi''(t)^{3/2}} = \frac{|t|}{\sqrt{1+t^2}} \leq 1.$$

For a direction $h \in \mathbb{R}^m$, multiplication by $|h_i|^3$ proves Assumption 2 with $M_{\mathrm{f}} = 1$ for each f_i. Hence $\beta = M_{\mathrm{f}} = 1$. Each objective has flat directions, while the common population minimizer is unique. At any x,

$$\nabla^2 f_w(x) = \operatorname{diag}\left(\frac{w_i}{1+x_i^2}\right)_{i=1}^m, \qquad \max_{w\in\Delta_m} \lambda_{\min}(\nabla^2 f_w(x)) = \frac{1}{m+\|x\|^2}.$$

To verify the maximum, a lower eigenvalue bound u requires $w_i \geq u(1+x_i^2)$ for every i, so $u \leq 1/(m+\|x\|^2)$. Equality holds with weights proportional to $1+x_i^2$. Consequently $\mu_{\mathrm{L}} = 1/m > 0$ and $\mu_{\mathrm{G}} = 0$.

Centering of Z_i gives unbiased losses and gradients. Uniformly in x and i,

$$\mathbb{E}|\ell_i(x;Z) - f_i(x)|^2 = s^2\sin^2 x_i \leq s^2, \qquad \mathbb{E}\|\nabla\ell_i(x;Z) - \nabla f_i(x)\|^2 = s^2\cos^2 x_i \leq s^2.$$

For a simplex mixture, orthogonality of the coordinate directions bounds the single-observation gradient variance by $s^2\sum_i w_i^2 \leq s^2$. The per-loss sufficient condition holds with $\nu_f = \sigma_{\mathrm{g}} = s$, and Proposition A.1 applies with $\nu_\Delta = \sqrt{m}\,s$. For $s > 0$, $\nabla\ell_i(x_\star;Z) = sZ_ie_i$ is nonzero, demonstrating population alignment with nonzero sample gradients at the population minimizer.

Batch size and the certified residual. For the basic QP under Proposition A.1, fix $0 < \eta \leq 1/\beta$, batch size N and $\chi = 1 - \eta\beta/2$. Equation (A.4) gives

$$\mathcal{E}_\infty = \frac{\frac{4}{3}\eta\nu_\Delta/\sqrt{N} + \eta^2\sigma_{\mathrm{g}}^2\{\frac{4}{3} + 2\beta\eta + 2(\beta\eta)^2 + 2q\}/N}{\{(1-\beta\eta)^2 + 2(\beta\eta)^2\}q}.$$

At fixed N, decreasing η suppresses both gradient-noise contributions, including the raw-Gram term, but the certificate tends to $4\nu_\Delta/(\mu_{\mathrm{L}}\sqrt{N})$. Thus small steps alone do not remove noisy loss selection.

For fixed batch size N and $\eta = 1/2$, take

$$\sigma^2 = \frac{s^2}{N}, \qquad \chi = \frac{3}{4}, \qquad \varepsilon = s\sqrt{\frac{m}{N}} + \frac{s^2}{8N}.$$

Substituting these constants into Eq. (A.4) gives

$$q = \frac{1}{12m}, \qquad R^2 = 2, \qquad \mathcal{E}_\infty = \frac{32}{3}ms\sqrt{\frac{m}{N}} + \frac{34m+2}{3}\frac{s^2}{N}. \tag{A.24}$$

For any target $0 < \epsilon_{\mathrm{cert}} < 2$, a sufficient integer batch size for $\mathcal{E}_\infty \leq \epsilon_{\mathrm{cert}}$ is

$$N \geq \max\left\{1, \left\lceil\frac{4096m^3s^2}{9\epsilon_{\mathrm{cert}}^2}\right\rceil, \left\lceil\frac{2(34m+2)s^2}{3\epsilon_{\mathrm{cert}}}\right\rceil\right\}.$$

Each of the two terms in Eq. (A.24) is then at most $\epsilon_{\mathrm{cert}}/2$. This budget certifies Theorem 1's expected entry and localized bound, not point convergence or the actual limiting error.

An explicit admissible parameter choice. For $m = 2$, $s = 0.01$ and $N = 1$, substitution into Eq. (A.24) gives $\mathcal{E}_\infty < 0.305 < R^2 = 2$, certifying expected entry. The uniform mixture also attains μ_{L} here, and this example establishes applicability when $\mu_{\mathrm{G}} = 0$.

A.5 Oracle and solver qualifications for DunAMOO

This subsection connects data and solver output to the selection certificate Eq. (A.1) for the basic QP Eq. (4). Lemma A.8 in Appendix A.5.4 also quantifies optional ridge regularization. At selection k, the score uses the already-known next step η_{k+1}, while the current parameter update uses η_k and the stored weight w_k.

A.5.1 The Gram of mean gradients

The basic score uses the Gram of the *minibatch-mean gradients*,

$$(\widehat{J}_k^\top \widehat{J}_k)_{ij} = \frac{1}{N_k^2} \sum_{s,t=1}^{N_k} \langle \nabla_x \ell_i(x_k; \mathsf{Z}_{k,s}), \nabla_x \ell_j(x_k; \mathsf{Z}_{k,t}) \rangle.$$

This generally differs from the mean of the per-observation Grams. Writing $\Xi_k = \widehat{J}_k - J(x_k)$, conditional unbiasedness and expansion give

$$\mathbb{E}_k[\widehat{J}_k^\top \widehat{J}_k] = J(x_k)^\top J(x_k) + \mathbb{E}_k[\Xi_k^\top \Xi_k].$$

Delay removes the same-batch covariance but does not make this Gram estimator unbiased. Proposition A.1 and Lemma A.7 control its contribution through a history-fixed comparator gradient.

Computational cost includes individual objective-gradient evaluations, typically m backward passes rather than one for the fixed average. Runtime comparisons must include these evaluations, Gram formation and the QP solve. The QP has m variables once the Gram is formed, and delaying the weight adds storage but no additional gradient evaluation.

Fresh sampling is a separate requirement. The update oracle must satisfy conditional unbiasedness given $\mathcal{F}_k$. Epoch-wise sampling without replacement does not automatically have this property for the full empirical objective, because the remaining sample pool depends on the history. One-iteration delay alone does not repair that dependence. Sampling used to construct a dataset is distinct from sampling minibatches during optimization, and a training implementation must specify the latter before invoking the fresh-i.i.d. guarantees.

A.5.2 Optimal values, common offsets and differential errors

On Δ_m, replacing $\widehat{\Delta}_k$ by $\widehat{\Delta}_k + b\mathbf{1}$ adds the constant $2b$ to the score and changes no feasible-direction derivative. Every maximizer and first-order residual is therefore unchanged, even for a batch-dependent common offset. The optimal values are needed only up to a common additive offset, so if $f_i^\star = \kappa_i + c$ for every i with known κ_i, the constant c need not be known. Alignment alone does not imply equal optimal values.

For supplied estimates $\widehat{f}^\star_{k,i}$, write

$$\widehat{f}_{k,i} = \frac{1}{N_k} \sum_{s=1}^{N_k} \ell_i(x_k; \mathsf{Z}_{k,s}), \qquad \widehat{\Delta}_{k,i} = \widehat{f}_{k,i} - \widehat{f}^\star_{k,i},$$

$$\delta_i = \widehat{f}_{k,i} - f_i(x_k), \qquad a_i = f_i^\star - \widehat{f}^\star_{k,i}, \qquad \widehat{\Delta}_k = \Delta(x_k) + \delta + a.$$

Define the range of the value errors by $\operatorname{osc}(a) = \max_i a_i - \min_i a_i$. For any $W, v \in \Delta_m$,

$$a^\top (W - v) \le \operatorname{osc}(a).$$

This pathwise bound gives the differential-value allowance below, including when the estimates use the current batch. A common value error costs zero. Raw mean losses implement the same selector when all $f_i^\star$ are equal, but nonnegative losses or arbitrary lower bounds do not justify ignoring differential optimal values. The basic gap observations remain signed, and an estimated-target rule requires a bound on its resulting differential error.

A.5.3 AN OBSERVABLE SOLVER CERTIFICATE

Write the full empirical score as

$$\widehat{Q}_k(w) = 2\widehat{\Delta}_k^\top w - \eta_{k+1}\|\widehat{J}_k w\|^2, \qquad w \in \Delta_m.$$

A numerical solver, such as projected gradient ascent on Δ_m, must return a feasible measurable weight. For such a w, compute

$$p = \nabla \widehat{Q}_k(w) = 2\widehat{\Delta}_k - 2\eta_{k+1}\widehat{J}_k^\top \widehat{J}_k w.$$

Its full-score first-order residual has the observable formula

$$r_{\widehat{Q}_k}(w) := \max_{u\in\Delta_m} \langle p, u - w\rangle = \max_i p_i - p^\top w \geq 0. \tag{A.25}$$

Concavity and the first-order optimality condition give (Boyd & Vandenberghe, 2004) $0 \leq \max_{u\in\Delta_m} \widehat{Q}_k(u) - \widehat{Q}_k(w) \leq r_{\widehat{Q}_k}(w)$, and the residual is zero exactly at maximizers. Compactness of the simplex ensures that a finite-valued score attains its maximum, and Lemma C.2 supplies a measurable exact rule even for a singular Gram. Report the realized residual, numerical tolerance and convergence flag at the iteration cap. An objective-value gap alone does not justify the same numerical bound on this first-order residual.

Lemma A.7 (Finite-tolerance certificate for the basic empirical QP)**.** *Assume the population, sampling and moment hypotheses of Proposition A.1. Let η_{k+1} be $\mathcal{F}_k$-measurable with $0 < \eta_{k+1} \leq \eta_{\max} < 2/\beta$ for a deterministic $\eta_{\max}$. Let $W = w_{k+1} \in \Delta_m$ be any feasible measurable output for the score above, using the value estimates defined in Appendix A.5.2. If $r_{\widehat{Q}_k}(W)$ and $\operatorname{osc}(a)$ have finite conditional expectations, then*

$$\mathbb{E}_k \Delta_W(x_k) \geq \left(1 - \frac{\beta\eta_{k+1}}{2}\right) M(x_k) - \varepsilon_k,$$

$$\varepsilon_k = \frac{\nu_\Delta}{\sqrt{N_k}} + \frac{\eta_{k+1}\sigma_{\mathrm{g}}^2}{4N_k} + \frac{1}{2}\mathbb{E}_k r_{\widehat{Q}_k}(W) + \mathbb{E}_k \operatorname{osc}(a).$$

Consequently Eq. (A.1) *holds with the fixed fraction $\chi = 1 - \beta\eta_{\max}/2 > 0$ and this allowance.*

Proof. Condition on $\mathcal{F}_k$ and let j be the smallest population maximum-gap index. The residual inequality with comparator e_j gives the same square completion as in Proposition A.1, with one additional term,

$$\widehat{\Delta}_{k,j} - \widehat{\Delta}_k^\top W \leq \frac{\eta_{k+1}}{4}\|\widehat{g}_{k,j}\|^2 + \tfrac{1}{2} r_{\widehat{Q}_k}(W).$$

In the decomposition $\widehat{\Delta}_k = \Delta(x_k) + \delta + a$, the selected loss-contrast term is bounded by $\nu_\Delta/\sqrt{N_k}$ exactly as in that proposition. The additional value error satisfies $a^\top(W - e_j) \leq \operatorname{osc}(a)$ pathwise. Finally, $\mathbb{E}_k\|\widehat{g}_{k,j}\|^2 \leq 2\beta M(x_k) + \sigma_{\mathrm{g}}^2/N_k$. Taking conditional expectations proves the displayed allowance, including the factor $1/2$ on the full-score residual. Since η_{k+1} is history-measurable and bounded by $\eta_{\max}$, and $M(x_k) \geq 0$, the fixed fraction follows. No independence between W and its current-batch inputs is used. □

The factor $1/2$ multiplying the residual is due to the full-score normalization in Eq. (A.25). A solver stopping rule $r_{\widehat{Q}_k}(W) \leq \epsilon_{\mathrm{sol},k}$ with a past-measurable budget $\epsilon_{\mathrm{sol},k}$ contributes at most $\epsilon_{\mathrm{sol},k}/2$ to ε_k. The convergence theorems require the resulting conditional allowance to satisfy their stated bounds, and logging a realized residual alone does not supply a uniform allowance. With exact optimization and values up to a common offset, both additional terms vanish and the lemma reduces to Proposition A.1. Under the per-objective loss-moment alternative, take $\nu_\Delta = \sqrt{m}\,\nu_f$.

A.5.4 OPTIONAL RIDGE REGULARIZATION AS A CERTIFIED SELECTION ERROR

A ridge term changes the selected weight, not the target population problem. It can be accounted for through the same first-order residual, leaving the unregularized score in Algorithm 1 and the basic theorem unchanged. For a nonnegative, $\mathcal{F}_k$-measurable coefficient λ_k, define

$$\widehat{Q}_{k,\lambda}(w) = \widehat{Q}_k(w) - \eta_{k+1}\lambda_k\|w\|^2, \qquad w \in \Delta_m.$$

Use the residual definition in Eq. (A.25) for either score.

Lemma A.8 (Ridge-to-basic-score certificate)**.** *Under the hypotheses of Lemma A.7, let $W = w_{k+1} \in \Delta_m$ be a feasible measurable output, and assume the ridge-score residual and value-error range have finite conditional expectations. Then*

$$\frac{1}{2} r_{\widehat{Q}_k}(W) \le \frac{1}{2} r_{\widehat{Q}_{k,\lambda}}(W) + \eta_{k+1}\lambda_k \frac{m-1}{4m}. \tag{A.26}$$

Consequently the selection certificate holds with fraction $\chi = 1 - \beta\eta_{\max}/2$ and allowance

$$\varepsilon_k = \frac{\nu_\Delta}{\sqrt{N_k}} + \frac{\eta_{k+1}\sigma_{\mathrm{g}}^2}{4N_k} + \frac{1}{2}\mathbb{E}_k r_{\widehat{Q}_{k,\lambda}}(W) + \eta_{k+1}\lambda_k \frac{m-1}{4m} + \mathbb{E}_k \operatorname{osc}(a). \tag{A.27}$$

Proof. Fix the realized inputs and a feasible w. Since $\nabla \widehat{Q}_k(w) = \nabla \widehat{Q}_{k,\lambda}(w) + 2\eta_{k+1}\lambda_k w$, maximization over $u \in \Delta_m$ gives

$$\begin{aligned} r_{\widehat{Q}_k}(w) &\le r_{\widehat{Q}_{k,\lambda}}(w) + 2\eta_{k+1}\lambda_k \max_{u\in\Delta_m} \langle w, u - w\rangle \\ &= r_{\widehat{Q}_{k,\lambda}}(w) + 2\eta_{k+1}\lambda_k \big(\max_i w_i - \|w\|^2\big). \end{aligned}$$

The last parenthesis is nonnegative because $\sum_i w_i = 1$. For $m > 1$, put $t = \max_i w_i$. Cauchy–Schwarz on the other $m-1$ coordinates gives

$$\begin{aligned} \max_i w_i - \|w\|^2 &\le t - t^2 - \frac{(1-t)^2}{m-1} \\ &= \frac{m-1}{4m} - \frac{m}{m-1}\left(t - \frac{m+1}{2m}\right)^2 \le \frac{m-1}{4m}. \end{aligned}$$

For $m = 1$ both sides are zero. Divide the residual bound by two to obtain Eq. (A.26). Conditional expectation and Lemma A.7 give Eq. (A.27), and predictability permits moving $\eta_{k+1}\lambda_k$ outside the expectation. No independence of W and the current inputs is used. □

Propagation to the convergence bound. At the basic step $\eta = 1/(2\beta)$ and fixed batch size N, suppose $0 \le \lambda_k \le \bar{\lambda}$, $r_{\widehat{Q}_{k,\lambda}}(W) \le \epsilon_{\mathrm{sol}}$ and $\operatorname{osc}(a) \le \epsilon_{\mathrm{val}}$ almost surely, with deterministic nonnegative budgets. Theorem A.1 gives the main theorem's conclusions after replacing $\mathcal{E}_\infty$ by

$$\mathcal{E}_\infty + \frac{4\bar{\lambda}(m-1)}{3\beta\mu_{\mathrm{L}} m} + \frac{16\epsilon_{\mathrm{sol}}}{3\mu_{\mathrm{L}}} + \frac{32\epsilon_{\mathrm{val}}}{3\mu_{\mathrm{L}}}. \tag{A.28}$$

Indeed, the additional allowance is at most $\bar{\lambda}(m-1)/(8\beta m) + \epsilon_{\mathrm{sol}}/2 + \epsilon_{\mathrm{val}}$, and Eq. (A.4) multiplies it by $32/(3\mu_{\mathrm{L}})$ at the basic step. The rate q, threshold R and energy comparison are unchanged, and the entry condition must use the enlarged residual in Eq. (A.28). Exact ridge optimization makes $\epsilon_{\mathrm{sol}} = 0$ but does not remove the ridge allowance. These are sufficient uniform budgets, not inferred tolerances of a particular run. The bridge accounts for numerical selection only, and it does not establish the population geometry or exact optimal values for a neural-network experiment.

A.6 Proof of the global convex maximum-gap guarantee

We first retain the general-selector guarantee, then specialize it to the fixed-batch QP in Corollary 1.

General bound. Assume alignment, Assumptions 1 and A.1, and the selection certificate Eq. (A.1), with fixed $0 < \chi \le 1$. For a fixed initial point x_0 and any constant $0 < \eta \le 1/\beta$, use the unprojected recursion $x_{k+1} = x_k - \eta \widehat{J}_k w_k$, where $w_k \in \Delta_m$ is $\mathcal{F}_k$-measurable and each update uses fresh oracle randomness. For $K \ge 1$, set $\bar{x}_K = K^{-1}\sum_{k<K} x_k$ and $\bar{\varepsilon}_K = K^{-1}\sum_{k<K}\mathbb{E}\varepsilon_k$. For any $V_0 \ge 0$, writing $c = 2(\beta\eta)^2$ gives

$$\mathbb{E}M(\bar{x}_K) \le \frac{3\{(1+c)V_0 + \eta^2\sigma^2\}}{4\eta\chi K} + \frac{\bar{\varepsilon}_K}{\chi} + \frac{3\eta\sigma^2}{4\chi}(1 + 2\beta\eta + c). \tag{A.29}$$

Neither self-concordance nor positive mixture curvature is required.

For $V_0 > 0$, choose $\eta = \min\{1/(2\beta), \sqrt{V_0}/(\sigma\sqrt{K})\}$ when $\sigma > 0$, and $\eta = 1/(2\beta)$ when $\sigma = 0$. The resulting rate for the chosen run is

$$\mathbb{E}M(\bar{x}_K) = O\left(\frac{\beta V_0/K + \sigma\sqrt{V_0/K} + \bar{\varepsilon}_K}{\chi}\right). \tag{A.30}$$

If only a known upper bound $B_0 \geq V_0$ with $B_0 > 0$ is available, replacing V_0 by B_0 in the step formula gives Eq. (A.30) with B_0 in place of V_0. When $V_0 = 0$, the exact bound Eq. (A.29) still applies for every allowed positive step, and the distance-based tuning formula is not used.

Proof of the general bound and Corollary 1. For the constant step η, let $B = \eta^2\sigma^2(1 + 2\beta\eta + c)$ and $U_k = \mathbb{E}V_{k+1} + c\mathbb{E}V_k$. Taking expectations in Eq. (A.17) and adding $c\mathbb{E}V_{k+1}$ gives

$$\begin{aligned} U_{k+1} &\leq U_k - \tfrac{4}{3}\eta\mathbb{E}\Delta_{w_{k+1}}(x_k) + B \\ &\leq U_k - \tfrac{4}{3}\eta\chi\mathbb{E}M(x_k) + \tfrac{4}{3}\eta\mathbb{E}\varepsilon_k + B. \end{aligned}$$

Sum from $k = 0$ to $K - 1$ and use $U_K \geq 0$. The function $M = \max_i \Delta_i$ is convex, so pathwise Jensen and expectation give

$$\mathbb{E}M(\bar{x}_K) \leq \frac{1}{K}\sum_{k<K}\mathbb{E}M(x_k) \leq \frac{3U_0}{4\eta\chi K} + \frac{1}{\chi K}\sum_{k<K}\mathbb{E}\varepsilon_k + \frac{3B}{4\eta\chi}.$$

Lemma A.3 gives $U_0 \leq (1 + c)V_0 + \eta^2\sigma^2$, proving Eq. (A.29). The proof evaluates the one-lag inequality up to index $K - 1$, which involves the iterate x_{K+1} of the infinite recursion, while the reported average uses $x_0, \ldots, x_{K-1}$.

For $V_0 > 0$ and $\sigma > 0$, the horizon-tuned step has $c \leq 1/2$, so

$$\frac{3U_0}{4\eta\chi K} + \frac{3B}{4\eta\chi} \leq \frac{1}{\chi}\left[\frac{9V_0}{8\eta K} + \frac{3\eta\sigma^2}{4K} + \frac{15}{8}\eta\sigma^2\right].$$

Considering which term determines η bounds these terms by $O((\beta V_0/K + \sigma\sqrt{V_0/K})/\chi)$. Adding $\bar{\varepsilon}_K/\chi$ gives Eq. (A.30). For $\sigma = 0$, the choice $\eta = 1/(2\beta)$ instead gives $O(\beta V_0/(\chi K) + \bar{\varepsilon}_K/\chi)$. Replacing V_0 by a known positive upper bound in this argument proves the stated tuning alternative.

For basic DUNAMOO with fixed batch size N and the unbiased primitive moments Eq. (2), take $\sigma = \sigma_{\mathrm{g}}/\sqrt{N}$. With optimal values supplied up to a common offset, Proposition A.1 gives

$$\chi = 1 - \eta\beta/2 \geq 3/4, \qquad \bar{\varepsilon}_K = \frac{\nu_\Delta}{\sqrt{N}} + \frac{\eta\sigma^2}{4}.$$

For $\sigma > 0$, the chosen step satisfies $\eta\sigma^2 \leq \sigma\sqrt{V_0/K}$, and for $\sigma = 0$ the additional gradient term is zero. Substitution into Eq. (A.30) gives exactly the fixed-batch rate Eq. (15). The step becomes $\min\{1/(2\beta), \sqrt{NV_0}/(\sigma_{\mathrm{g}}\sqrt{K})\}$ when $\sigma_{\mathrm{g}} > 0$, as stated in Corollary 1. The gradient contribution to selection shrinks with the run's step, while the fixed-batch loss-contrast term remains. □

The bound holds on the full space. If common minimizers are nonunique, it controls objective gaps relative to any fixed common minimizer.

A.7 MAXIMUM-GAP SELECTION AND GRADIENT-AWARE MIXING

The exact-feedback maximum-gap rule is MG-AMOO's choice (Kretzu et al., 2025, Section 5.2). Its stochastic stored-weight version chooses the smallest $\iota \in \arg\max_i \widehat{\Delta}_{k,i}$ and stores e_ι. With optimal values supplied up to a common offset, let j be the history-fixed population maximum-gap index and $d = P\delta$, where δ is the minibatch-mean centered loss vector. Then $\Delta_\iota(x_k) \geq M(x_k) + d_j - d_\iota$. Since $\mathbb{E}_k d_j = 0$, the contrast moments give

$$\mathbb{E}_k\Delta_\iota(x_k) \geq M(x_k) - \mathbb{E}_k\|d\|_\infty \geq M(x_k) - \frac{\nu_\Delta}{\sqrt{N_k}}.$$

Thus the simpler rule has the better generic fraction $\chi = 1$ in Eq. (A.1), so the selection certificate alone gives no advantage to the QP.

Why retain the QP? The certified fraction χ measures how much loss gap is retained, but discards how the chosen gradients combine. The QP uses this additional directional information to maximize a lower bound on one-step squared-distance decrease. With exact feedback,

$$\|x - \eta J(x)w - x_\star\|^2 \le \|x - x_\star\|^2 - \eta\{2\Delta(x)^\top w - \eta\|J(x)w\|^2\},$$

which follows from convexity. The QP maximizes this computable decrease certificate at the selection point. Delayed application still requires the transport argument.

An explicit range where mixing helps at the same delay. Let $f_1(x) = x_1^2/2$, $f_2(x) = x_2^2/2$, $\eta = 1/2$ and $x_0 = (\sqrt{a}, 1)$ with $a > 1$. Use exact feedback and the same initial stored weight $w_0 = (1/2, 1/2)$ for both rules. The maximum-gap choice at x_0 is uniquely e_1. Writing $w = (t, 1-t)$, the QP score is $1/2 + at - (a+1)t^2/2$, uniquely maximized at $t = a/(a+1)$. Both rules first reach $x_1 = 3x_0/4$, and then

$$x_2^{\mathrm{QP}} = \left(\frac{3\sqrt{a}(a+2)}{8(a+1)}, \frac{3(2a+1)}{8(a+1)}\right), \qquad x_2^{\mathrm{MG}} = \left(\frac{3\sqrt{a}}{8}, \frac{3}{4}\right).$$

Direct subtraction gives

$$\|x_2^{\mathrm{MG}}\|^2 - \|x_2^{\mathrm{QP}}\|^2 = \frac{9(3-2a)}{64(a+1)}.$$

Thus mixing strictly improves the two-step squared distance for $1 < a < 3/2$, ties at $a = 3/2$, and is worse for $a > 3/2$. For $a = 5/4$, it improves both squared distance and maximum gap, since

$$\|x_2^{\mathrm{QP}}\|^2 = \frac{181}{256} < \frac{189}{256} = \|x_2^{\mathrm{MG}}\|^2, \qquad M(x_2^{\mathrm{QP}}) = \frac{845}{4608} < \frac{9}{32} = M(x_2^{\mathrm{MG}}).$$

When the two coordinates have similar gaps, spreading the delayed update can reduce both errors more evenly, whereas when the first gap is sufficiently larger, the concentrated update has the smaller two-step distance. This identifies when mixing is useful in this family, not a general rate or runtime advantage. Maximum-gap selection avoids forming the Gram and solving the QP, so the additional cost is justified only when the gradient information is useful. The general theorem covers either rule through its own selection certificate.

a	V_2 (QP)	V_2 (maximum gap)	$M(x_2)$ (QP)	$M(x_2)$ (maximum gap)
9/8	0.671071	0.720703	0.171067	0.281250
5/4	0.707031	0.738281	0.183377	0.281250
3/2	0.773438	0.773438	0.206719	0.281250
2	0.890625	0.843750	0.250000	0.281250
4	1.265625	1.125000	0.405000	0.281250

Table A.2: **Exact matched-delay diagnostic.** Both rules use $\eta = 1/2$ and $w_0 = (1/2, 1/2)$ on the coordinate quadratics above. Values are computed by two explicit updates in rational arithmetic and rounded only for display. There is no sampling uncertainty. The comparison includes both the range favoring QP mixing and the range favoring maximum-gap selection.

The supplementary diagnostic checks the subtraction identity on the fixed rational grid $a = 101/100, 102/100, \ldots, 400/100$ and reports the representative values in Table A.2. This verifies the calculation, not a claim of empirical acceleration on Multi-MNIST.

A.8 Exact-feedback limit and the radial-scale trade-off

We compare both methods with exact population gradients and objective gaps, exact score optimization, and deterministic tie choices. For DunAMOO, this gives $\sigma = \varepsilon = 0$, whereas setting gradient variance to zero alone does not remove loss-selection, value-input, or solver errors. The comparison uses original PAMOO (Efroni et al., 2025, Algorithm 3 and Section 5).

Corollary A.4 (Exact-feedback two-phase limit)**.** *Under Theorem A.1, use exact gradients, exact population-score optimization and deterministic tie-breaking. For basic* DunAMOO *with constant*

$0 < \eta \le 1/\beta$, *take* $\chi = 1 - \eta\beta/2$. *Then* $\varepsilon = \sigma = \mathcal{E}_\infty = 0$. *Use the energy, rate and threshold in Eq.* (A.3)–(A.4). *If* $M_{\mathrm{f}} > 0$, *the actual entry index* $T_0 = \inf\{k : \mathcal{L}_k \le R^2\}$ *satisfies*

$$T_0 \le \left\lceil \frac{2}{q}\left(\frac{\sqrt{\mathcal{L}_0}}{R} - 1\right)_+ \right\rceil,$$
$$\sqrt{\mathcal{L}_k} \le \begin{cases} \sqrt{\mathcal{L}_0} - kqR/2, & 0 \le k < T_0, \\ \sqrt{\mathcal{L}_{T_0}}(1-q)^{(k-T_0)/2}, & k \ge T_0, \end{cases} \tag{A.31}$$

where $(z)_+ = \max\{z, 0\}$. *Moreover,* $\|x_k - x_\star\| \le \sqrt{\mathcal{L}_k}$, *and if* $M_{\mathrm{f}} = 0$ *then* $\mathcal{L}_k \le (1-2q)^k \mathcal{L}_0$ *for every* $k \ge 0$.

Proof. Exact feedback gives $\varepsilon = \sigma = 0$, so Lemma A.6 reduces to the deterministic recurrence

$$\mathcal{L}_{k+1} \le \mathcal{L}_k - \frac{2q\mathcal{L}_k}{1 + \sqrt{\mathcal{L}_k}/R} \le \mathcal{L}_k \qquad (M_{\mathrm{f}} > 0).$$

Before entry, the far-phase square-root drift is at least $qR/2$. Summing gives the first branch. If entry had not occurred by the displayed ceiling, this bound would already give $\sqrt{\mathcal{L}_k} \le R$ at that index, a contradiction. This proves the entry bound in Eq. (A.31). Monotonicity makes the entered sublevel set invariant, and there $1 + \sqrt{\mathcal{L}_k}/R \le 2$, giving the second branch. The energy comparison gives the distance bound. At $\eta = 1/(2\beta)$, this is the main-text specialization Eq. (16). For $M_{\mathrm{f}} = 0$, the recurrence contracts globally by $1 - 2q$. □

Applying the same geometric estimate to the cone method. Write $Q_x(u) = 2\Delta(x)^\top u - \|J(x)u\|^2$. PAMOO maximizes Q_x over $u \ge 0$ and takes $x^+ = x - J(x)u$. Convexity and alignment give $\|x^+ - x_\star\|^2 \le \|x - x_\star\|^2 - Q_x(u)$. Let $v_\star \in \Delta_m$ maximize the local mixture curvature. For $x \ne x_\star$, set $\rho = \|x - x_\star\|$, $A = \Delta_{v_\star}(x) > 0$, and $B = \|J(x)v_\star\|^2 > 0$. Here positivity follows from the lower bound Eq. (A.14) and convexity, while smoothness gives $B \le 2\beta A$. Optimizing the total weight along $v_\star$ therefore yields

$$\max_{u \ge 0} Q_x(u) \ge Q_x\left(\frac{A}{B} v_\star\right) = \frac{A^2}{B} \ge \frac{A}{2\beta}$$
$$\ge \frac{\mu_{\mathrm{L}} \rho^2}{4\beta(1 + M_{\mathrm{f}}\sqrt{\beta}\,\rho)}. \tag{A.32}$$

Thus the componentwise geometric estimate also improves the cone analysis and is not itself an advantage conferred by normalization or delay.

Matched quadratic comparison. For $M_{\mathrm{f}} = 0$ and exact feedback, Table A.3 compares the available squared-error factors.

Certificate	Geometric factor
General factor stated by Efroni et al. (2025), restricted to quadratics	$1 - 3\mu_{\mathrm{L}}/(32\beta)$
PAMOO, using Eq. (A.32)	$1 - \mu_{\mathrm{L}}/(4\beta)$
Basic DUNAMOO, $\eta = 1/(2\beta)$	$1 - \mu_{\mathrm{L}}/(6\beta)$

Table A.3: Deterministic quadratic certificates with exact feedback and score optimization. The PAMOO rows use its original nonnegative-cone score.

The refined cone certificate is stronger on this subclass, since it starts from V_0, whereas the basic DUNAMOO bound uses $\mathcal{L}_0 \le 2V_0$. Comparing only against the weaker stated cone factor would not establish deterministic acceleration. For nonlinear objectives the exact-feedback DUNAMOO near factor is $1 - \mu_{\mathrm{L}}/(12\beta)$, whereas the stated cone factor is $1 - 3\mu_{\mathrm{L}}/(32\beta)$, with different certified entry regions. These are sufficient upper bounds, and their ordering does not order actual trajectories or training time.

Why fixing the weight scale is a trade-off. Let $0 < \mu < \beta/2$ and

$$f_1(x) = \tfrac{1}{2}(\mu x_1^2 + \beta x_2^2), \qquad f_2(x) = \tfrac{1}{2}(\mu x_1^2 + (\beta/2) x_2^2), \qquad x_0 = (r_0, 0), \quad r_0 > 0.$$

Here $\mu_{\mathrm{L}} = \mu$. Every normalized mixture has gradient μx_1 on this invariant coordinate, so basic DUNAMOO has $\|x_k\| = (1 - \mu/(2\beta))^k r_0$ at its default step, regardless of the weights or delay. In contrast, the cone score for total mass s is $x_1^2(\mu s - \mu^2 s^2)$, maximized at $s = 1/(2\mu)$, giving $\|x_k\| = 2^{-k} r_0$. Universal acceleration is therefore false. Normalization provides a compact feasible set and controlled total mass, while transferring update scaling from the cone optimizer to the external learning rate. This exact-feedback comparison does not establish stochastic performance.

Linked score coefficient and common rescaling. Section 4 derives the score coefficient from the application step. This linkage also yields invariance to common loss rescaling. If every sample loss and supplied optimal value is multiplied by a common $c > 0$, and η_k is replaced by η_k/c, then

$$2(c\widehat{\Delta}_k)^\top w - \frac{\eta_{k+1}}{c}\|c\widehat{J}_k w\|^2 = c\{2\widehat{\Delta}_k^\top w - \eta_{k+1}\|\widehat{J}_k w\|^2\}.$$

With coupled observations, identical initial weights and matching tie choices, the maximizing sets, stored weights and parameter trajectory are unchanged. A common additive offset in the supplied optimal values also cancels, but differential errors do not. This is invariance to common loss rescaling, not arbitrary per-objective rescaling or individual-gradient normalization.

A.9 THE FIXED-AVERAGE EW-SGD BASELINE

Let $\bar{f} = m^{-1}\sum_i f_i$ and $\mu_{\mathrm{U},L} = \lambda_{\min}(\nabla^2 \bar{f}(x_\star))$. Let $\sigma_{\mathrm{U}}^2 \le \sigma^2$ uniformly bound the conditional second moment of the average-gradient error $m^{-1}\sum_i(\widehat{g}_{k,i} - \nabla f_i(x_k))$ for the update batches actually used. The fixed-mixture one-step bound is

$$\mathbb{E}_k V_{k+1} \le V_k - \eta\mu_{\mathrm{U},L}\frac{V_k}{1 + M_{\mathrm{f}}\sqrt{\beta}\sqrt{V_k}} + \eta^2\sigma_{\mathrm{U}}^2. \tag{A.33}$$

Lemma A.9 (EW-SGD without adaptive selection bias)**.** *Let $\bar{f} = m^{-1}\sum_i f_i$ and $\mu_{\mathrm{U},L} = \lambda_{\min}(\nabla^2\bar{f}(x_\star))$. Under the population and gradient-oracle assumptions of Theorem A.1, EW-SGD with weights $1/m$ and $0 < \eta \le 1/\beta$ satisfies Eq.* (A.33)*, with the batch-variance bound σ_{U}^2 above and $\mu_{\mathrm{L}}/m \le \mu_{\mathrm{U},L} \le \mu_{\mathrm{L}}$. For quadratics, set $p_{\mathrm{U}} = (1 - \eta\mu_{\mathrm{U},L})^2$. For every $k \ge 0$, with $p_{\mathrm{U}}^0 = 1$, the global bound is*

$$\mathbb{E}V_k \le p_{\mathrm{U}}^k V_0 + \frac{\eta\sigma_{\mathrm{U}}^2}{\mu_{\mathrm{U},L}(2 - \eta\mu_{\mathrm{U},L})}(1 - p_{\mathrm{U}}^k). \tag{A.34}$$

Proof. Let $\Delta_{\mathrm{unif}}(x) = \bar{f}(x) - \bar{f}(x_\star)$. The fixed weight is predictable, so Lemma A.3 gives $\mathbb{E}_k V_{k+1} \le V_k - 2\eta\Delta_{\mathrm{unif}}(x_k) + \eta^2\sigma_{\mathrm{U}}^2$. In the componentwise proof of Lemma A.2, average with the uniform vector. This gives $\Delta_{\mathrm{unif}}(x) \ge \mu_{\mathrm{U},L}\|x - x_\star\|^2/[2(1 + M_{\mathrm{f}}\sqrt{\beta}\,\|x - x_\star\|)]$. Write $H_i = \nabla^2 f_i(x_\star)$ and use $A \succeq B$ when $A - B$ is positive semidefinite. For a maximizing $w_\star$, $m^{-1}\sum_i H_i \succeq m^{-1}\sum_i w_{\star,i} H_i \succeq (\mu_{\mathrm{L}}/m)I$ at $x_\star$. Uniform feasibility gives the upper inequality. Substitution proves Eq. (A.33). Convexity of squared norm gives the admissible bound $\sigma_{\mathrm{U}}^2 \le \sigma^2$.

For quadratics, the constant Hessian $H_{\mathrm{U}} = \nabla^2\bar{f}$ satisfies $\mu_{\mathrm{U},L} I \preceq H_{\mathrm{U}} \preceq \beta I$. Define the centered average-gradient error $e_k = m^{-1}\sum_i(g_i(x_k, \mathcal{B}_k) - \nabla f_i(x_k))$. Then $\mathbb{E}_k e_k = 0$, $\mathbb{E}_k\|e_k\|^2 \le \sigma_{\mathrm{U}}^2$, and

$$\mathbb{E}_k V_{k+1} = \|(I - \eta H_{\mathrm{U}})(x_k - x_\star)\|^2 + \eta^2\mathbb{E}_k\|e_k\|^2 \le (1 - \eta\mu_{\mathrm{U},L})^2 V_k + \eta^2\sigma_{\mathrm{U}}^2.$$

Let $\|\cdot\|_{\mathrm{op}}$ denote the Euclidean operator norm. Then $0 < \eta \le 1/\beta$ ensures $\|I - \eta H_{\mathrm{U}}\|_{\mathrm{op}} = 1 - \eta\mu_{\mathrm{U},L}$. Geometric summation, with $1 - p_{\mathrm{U}} = \eta\mu_{\mathrm{U},L}(2 - \eta\mu_{\mathrm{U},L})$, proves Eq. (A.34). □

Comparing quadratic contraction certificates. At the common step $\eta = 1/(2\beta)$, the basic-QP certificate in Theorem 1 has factor $p_{\mathrm{D}} = 1 - \mu_{\mathrm{L}}/(6\beta)$, while Lemma A.9 gives $p_{\mathrm{U}} = (1 - \mu_{\mathrm{U},L}/(2\beta))^2$. Direct rearrangement gives

$$p_{\mathrm{D}} < p_{\mathrm{U}} \quad \Longleftrightarrow \quad \frac{\mu_{\mathrm{L}}}{\mu_{\mathrm{U},L}} > 6\left(1 - \frac{\mu_{\mathrm{U},L}}{4\beta}\right).$$

At least six objectives are necessary for this comparison to favor basic DUNAMOO. To see this, set $a = \mu_{\mathrm{L}}/\beta \in (0, 1]$ and $b = \mu_{\mathrm{U},L}/\beta$, so $a/m \le b \le a \le 1$. Since $b - b^2/4$ is increasing on $[0, 1]$, the inequality $p_{\mathrm{D}} < p_{\mathrm{U}}$ requires

$$\frac{a}{6} > b - \frac{b^2}{4} \ge \frac{a}{m} - \frac{a^2}{4m^2}, \qquad \text{hence} \qquad m^2 - 6m + \tfrac{3}{2}a > 0.$$

For integer $1 \le m \le 5$ and $a \le 1$, this is impossible. Thus the certificate establishes no contraction advantage in that range, including both two and four objectives.

In the selection example of Efroni et al. (2025) with $\Delta = 0$, where $m-1$ objectives are x_1^2 and one objective is $\|x\|^2$, one has $\mu_{\mathrm{L}} = 2$, $\mu_{\mathrm{U},L} = 2/m$ and $\beta = 2$, so the condition reads $m > 6 - 3/(2m)$ and holds for $m \ge 6$. Full error targets also depend on initialization and on gradient and selection residuals. This comparison of upper bounds neither proves that DUNAMOO is slower nor orders actual trajectories or training time.

The experiments of Efroni et al. (2025, Appendix A) use all-one weights for EW, which equals the average with an m-fold larger step. State this step rescaling when comparing conventions.

A.10 LIMITS OF FIXED-BATCH GUARANTEES

The following examples explain two distinctions in Theorem 1. Removing current-gradient covariance does not eliminate loss-selection error, and a localized second-moment bound does not control contributions after exit. Both examples satisfy the stated population and minibatch moment conditions.

Proposition A.2 (Loss-selection error can suppress informative objectives)**.** *There are aligned, smooth convex objectives with positive local mixture curvature and nonnegative smooth convex sample losses for which basic* DUNAMOO *permanently stops at a nonoptimal point after its first update. This occurs with exact gradients, exact optimal-value inputs, unbiased bounded loss errors and the default step $\eta = 1/(2\beta)$. It also occurs for nonlinear self-concordant objectives with arbitrarily small loss errors satisfying $\mathcal{E}_\infty < R^2$.*

Proof. Let $x = (u, v) \in \mathbb{R}^2$, $m = 3$, $a > 0$, and take either $\psi(t) = t^2/2$ or $\psi(t) = \phi(t) := t \arctan t - \frac{1}{2}\log(1+t^2)$. Define

$$f_1(x) = f_2(x) = a + \psi(v), \qquad f_3(x) = \psi(u), \qquad \ell_1 = f_1 + aZ, \quad \ell_2 = f_2 - aZ, \quad \ell_3 = f_3,$$

where each batch consists of one fresh Rademacher variable Z. All sample losses are nonnegative smooth convex functions. The unique common population minimizer is $x_\star = (0, 0)$, with known optimal values $(a, a, 0)$. The geometry verified for ϕ in Appendix A.4 gives $\beta = 1$ and $M_{\mathrm{f}} = 1$ in the nonlinear case, while the quadratic case has $\beta = 1$ and $M_{\mathrm{f}} = 0$. In both cases the mixture Hessian at $x_\star$ is $\mathrm{diag}(w_3, w_1 + w_2)$, hence $\mu_{\mathrm{L}} = 1/2$. Every sample gradient is exact, so $\sigma = \sigma_{\mathrm{g}} = 0$. The loss-error vector $(aZ, -aZ, 0)$ is centered, already has zero coordinate sum, and has squared norm $2a^2$. Thus the contrast moment condition holds with $\nu_\Delta = \sqrt{2}a$, and the individual loss variances are also uniformly at most a^2.

Use $\eta = 1/2$, uniform w_0, and $x_0 = (r, 0)$ with $r > 0$ and $\psi(r) < a$. At any selection point $(u, 0)$ with $\psi(u) < a$, the empirical gaps are $(aZ, -aZ, \psi(u))$. For a fixed $w_3 = t \in [0, 1]$, maximizing over the other coordinates puts all remaining mass on the objective whose empirical gap is $+a$. The resulting score is

$$2a + 2\{\psi(u) - a\}t - \eta\psi'(u)^2 t^2.$$

Its derivative is strictly negative on $[0, 1]$, so the unique maximizer has $w_3 = 0$ and selects an objective with zero gradient at $(u, 0)$. This selection is a measurable function of the finite-valued observation. The initial uniform-weight update gives

$$x_1 = (r_1, 0), \qquad r_1 = r - \psi'(r)/6.$$

Since $0 < \psi'(r) \le r$, one has $0 < r_1 < r$ and $\psi(r_1) < \psi(r) < a$. The weight selected at x_0 therefore gives $x_2 = x_1$, and induction gives $x_k = x_1$ for every $k \ge 1$. In particular $M(x_k) = \psi(r_1) > 0$. Changing the subsequent positive step schedule and using its next step in each score still leaves this derivative negative and these updates zero, including under Robbins–Monro steps.

For an explicit quadratic instance take $a = r = 1$. Then $x_k = (5/6, 0)$ and $M(x_k) = 25/72$ for all $k \ge 1$. For a nonlinear instance take any $0 < a < 1/32$ and $r = \sqrt{a}$. Smoothness and $\phi(0) = \phi'(0) = 0$ give $\phi(r) \le r^2/2 = a/2 < a$. Proposition A.1 supplies $\chi = 3/4$ and $\varepsilon = \sqrt{2}a$, so the default constants are

$$R^2 = 2, \qquad \mathcal{E}_\infty = \frac{64\sqrt{2}}{3}a < 2, \qquad \mathcal{L}_0 \le 2V_0 = 2a < 2.$$

Thus entry occurs at $T = 0$, while the iterates remain at a nonoptimal point. Taking $a \downarrow 0$ makes the selection allowance arbitrarily small. □

The selection residual can therefore reflect an actual obstruction to point convergence. Predictability keeps the common minimizer a zero of every population mixture, but another point can be stationary for a selected singular mixture. This example explains why basic normalized weighting and a nonzero selection allowance do not by themselves guarantee point convergence.

Proposition A.3 (Finite variance does not imply global nonlinear MSE control)**.** *For every $p \in (2,3)$ there are nonlinear population objectives satisfying the assumptions of Theorem 1, with uniformly bounded centered loss and gradient second moments and zero loss-contrast error, such that the default fixed-batch algorithm starts inside the near region and satisfies $\mathcal{E}_\infty < R^2$, but*

$$\mathbb{E}\|x_K - x_\star\|^2 \geq cK^{3-p} \longrightarrow \infty$$

for some $c > 0$ and all sufficiently large K. Every finite-time second moment is finite, and the oracle variances and certified residual can be made arbitrarily small.

Proof. Let $n = 1$, $m = 2$, and $f_1 = f_2 = \phi$, with ϕ as above. Then $x_\star = 0$, $f_i^\star = 0$, and $\beta = M_{\mathrm{f}} = \mu_{\mathrm{L}} = 1$. Let $H \geq 1$ have tail $\mathbb{P}(H > h) = h^{-p}$ for $h \geq 1$, and let U be uniform on $[0, 2\pi]$, independently of H. Use independent copies (H_k, U_k) across batches, $N_k = 1$, and sample losses

$$\ell_1(x; H, U) = \ell_2(x; H, U) = \phi(x) + sH\cos(x + U), \qquad s > 0.$$

Here H is finite almost surely, $\mathbb{E}H^2 = p/(p-2) < \infty$, and $\mathbb{E}H < \infty$. The losses and gradients are measurable and integrable at every fixed x, and the bound $|\partial_x \ell_i(x; H, U)| \leq \pi/2 + sH$ also justifies differentiation under expectation. Uniform averaging over U centers both errors, and

$$\mathbb{E}|\ell_i(x; H, U) - f_i(x)|^2 = \mathbb{E}|\partial_x \ell_i(x; H, U) - f_i'(x)|^2 = \frac{s^2 \mathbb{E}H^2}{2} =: v$$

for every x. The population objectives have the required convexity and self-concordance, and no samplewise convexity or interpolation is assumed in the theorem. Both empirical gaps and gradient columns are identical within each batch, so all normalized weights have the same score and update. One may use the measurable uniform tie rule. The loss-contrast error is exactly zero, while $\sigma^2 = \sigma_{\mathrm{g}}^2 = v$. With $\eta = 1/2$, Proposition A.1 gives $\chi = 3/4$ and $\varepsilon = v/8$. Starting at $x_0 = 0$ yields

$$R^2 = 2, \qquad \mathcal{E}_\infty = 12v, \qquad \mathcal{L}_0 = v/3.$$

Choose s so that $0 < v < 1/6$. Then $T = 0$ and the small-error condition holds. For example, $p = 5/2$ and $s = 10^{-2}$ give $v = 1/4000$ and $\mathcal{E}_\infty = 3/1000$.

The scalar recursion, with $\eta = 1/2$, is

$$x_{k+1} = x_k - \eta \arctan x_k + Y_k, \qquad Y_k = \eta s H_k \sin(x_k + U_k).$$

Conditionally on the pre-batch history, x_k is fixed and $(x_k + U_k) \bmod 2\pi$ is fresh uniform, independently of H_k. Consequently the Y_k are i.i.d., symmetric, centered, and have variance $\eta^2 v$ despite their state-dependent representation. For $y \geq \eta s$, the event $\sin U \geq 1/2$, which has probability $1/3$, and the bound $|\sin U| \leq 1$ give

$$\frac{(\eta s)^p}{3 \cdot 2^p} y^{-p} \leq \mathbb{P}(Y_k > y) \leq \mathbb{P}(|Y_k| > y) \leq (\eta s)^p y^{-p}.$$

Put $A = \eta\pi/2 > 0$ and $S_K = \sum_{k<K} Y_k$. Since $|\arctan x| \leq \pi/2$, summing the recursion gives $S_K - AK \leq x_K \leq S_K + AK$. In particular $\mathbb{E}x_K^2 \leq 2K\eta^2 v + 2A^2K^2 < \infty$ for each fixed K.

For $j = 0, \ldots, K-1$ define

$$E_{j,K} = \left\{ Y_j > 3AK, \quad \left| \sum_{i<K,\, i\neq j} Y_i \right| \leq AK, \quad \max_{i<K,\, i\neq j} |Y_i| \leq 3AK \right\}.$$

These events are pairwise disjoint, since each excludes a second innovation larger than $3AK$. Independence, Chebyshev's inequality and a union bound imply

$$\begin{aligned}\mathbb{P}(E_{j,K}) &\geq \mathbb{P}(Y_j > 3AK)\left\{1 - \frac{(K-1)\eta^2 v}{A^2K^2} - (K-1)\mathbb{P}(|Y_0| > 3AK)\right\} \\ &\geq c_0 K^{-p}\end{aligned}$$

for some $c_0 > 0$ and all sufficiently large K, since the bracket tends to one. On $E_{j,K}$, one has $S_K > 2AK$ and hence $x_K > AK$. Disjointness therefore yields

$$\mathbb{E}x_K^2 \geq A^2K^2 \sum_{j<K} \mathbb{P}(E_{j,K}) \geq c_0 A^2 K^{3-p} \longrightarrow \infty.$$

The choice of $s > 0$ only changes the positive constant and the starting horizon of this bound, so it can be arbitrarily small. □

This example does not contradict the localized second-moment bound, which omits the contribution after a previous exit, or the averaged maximum-gap bound, which controls a different quantity. It shows why ordinary nonlinear MSE requires additional tail or invariance conditions, or an explicit contribution from paths that exit the near region.

Why bounded variance is not an invariance certificate. Already for one objective $f(x) = x^2/2$, the sole simplex weight is one. With $0 < \eta < 1$ and i.i.d. standard Gaussian noise ξ_k, $x_{k+1} = (1-\eta)x_k - \eta\xi_k$. For any finite radius R, whenever $|x_k| \leq R$ and $|\xi_k| > (2-\eta)R/\eta$, the next iterate satisfies $|x_{k+1}| > R$. Its fresh-noise probability is the constant $p_R := \mathbb{P}\{|\xi_k| > (2-\eta)R/\eta\} > 0$. The probability of staying inside the ball for all of the next N steps is therefore at most $(1-p_R)^N$, which vanishes as $N \to \infty$. Thus bounded variance permits exits even in a smooth strongly convex single-objective problem. The global quadratic MSE result remains applicable because it controls the error after such exits as well.

B SAME-BATCH NORMALIZED WEIGHTING

This appendix isolates the two stochastic obstructions addressed by DUNAMOO, namely covariance from immediate application of a batch-selected weight and failure of the noisy cone score to attain a finite maximum. The same-batch control below uses the learning-rate score of Algorithm 1, and the two propositions give explicit examples under globally uniform oracle moments.

B.1 MATCHED CONTROL AND CONDITIONAL COVARIANCE

Use the aligned population objectives, normalized set Δ_m, and notation of Section 3. Let $\mathcal{F}_k$ contain the history before the current batch and write $\mathbb{E}_k = \mathbb{E}[\cdot \mid \mathcal{F}_k]$. The iterate x_k and positive steps η_k, η_{k+1} are $\mathcal{F}_k$-measurable. Use deterministic batch sizes $N_k \geq 1$ and, as in Algorithm 1, average fresh i.i.d. joint observations to form

$$\widehat{\Delta}_{k,i} = \frac{1}{N_k} \sum_{s=1}^{N_k} \ell_i(x_k; \mathsf{Z}_{k,s}) - f_i^\star,$$

$$\widehat{g}_{k,i} = \frac{1}{N_k} \sum_{s=1}^{N_k} \nabla_x \ell_i(x_k; \mathsf{Z}_{k,s}), \qquad \widehat{J}_k = [\widehat{g}_{k,1}, \ldots, \widehat{g}_{k,m}].$$

The observations are independent of $\mathcal{F}_k$, while tasks within an observation may be dependent. Exact optimal values, or values correct up to a common additive offset, are supplied. All arrays are finite-valued and jointly measurable, and exact QPs use the measurable tie rule of Lemma C.2. The matched same-batch control is

$$\boxed{\begin{aligned} &W_k \in \operatorname*{arg\,max}_{w \in \Delta_m} \{2\widehat{\Delta}_k^\top w - \eta_{k+1} \|\widehat{J}_k w\|^2\}, \\ &G_k = \widehat{J}_k W_k, \qquad x_{k+1} = x_k - \eta_k G_k. \end{aligned}} \tag{B.1}$$

It uses the same score, normalized feasible set, batch construction and current update step as DUNAMOO, with the selected weight applied immediately. In particular, a deterministic schedule retains coefficient η_{k+1} at selection index k and update step η_k. For constant steps, both equal η.

Assume unbiased sample gradients and the gradient moments in Eq. (2). Write $\widehat{J}_k = J(x_k) + \Xi_k$, with columns $\xi_{k,i} = \widehat{g}_{k,i} - \nabla f_i(x_k)$ and $\mathbb{E}_k \Xi_k = 0$. For a scalar U and vector Y, set $\mathrm{Cov}_k(U, Y) = \mathbb{E}_k[(U - \mathbb{E}_k U)(Y - \mathbb{E}_k Y)]$. A current-batch normalized selector satisfies

$$\mathbb{E}_k[\widehat{J}_k W_k] = J(x_k)\mathbb{E}_k W_k + b_k, \qquad b_k = \mathbb{E}_k[\Xi_k W_k] = \sum_i \mathrm{Cov}_k(W_{k,i}, \widehat{g}_{k,i}). \tag{B.2}$$

The primitive gradient moments in Eq. (2) give $\mathbb{E}_k\|\xi_{k,i}\|^2 \le \sigma_g^2/N_k$ by independent averaging.

Lemma B.1 (Conditional covariance under the main gradient moments)**.** *Assume fresh i.i.d. minibatch sampling, unbiased sample gradients and the gradient moment bound in Eq.* (2)*. For any measurable current-batch normalized selector W_k, with Ξ_k and b_k defined above,*

$$\|b_k\|^2 \le \frac{(m-1)\sigma_g^2}{N_k}, \qquad \mathbb{E}_k\|\Xi_k W_k\|^2 \le \frac{m\sigma_g^2}{N_k}.$$

These bounds allow dependence between the selected weight and every oracle entry.

Proof. Let $\bar{W} = \mathbb{E}_k W_k$ and $P = I - \mathbf{1}\mathbf{1}^\top/m$. Since $\mathbf{1}^\top(W_k - \bar{W}) = 0$ and $\mathbb{E}_k\Xi_k = 0$, $b_k = \mathbb{E}_k[\Xi_k P(W_k - \bar{W})]$. Cauchy–Schwarz and the Frobenius norm $\|A\|_F = (\sum_{i,j} A_{ij}^2)^{1/2}$ give

$$\|b_k\|^2 \le \mathbb{E}_k\|\Xi_k P\|_F^2\, \mathbb{E}_k\|W_k - \bar{W}\|^2 \le \frac{m\sigma_g^2}{N_k}\left(1 - \frac{1}{m}\right).$$

Here $\|\Xi_k P\|_F \le \|\Xi_k\|_F$, while normalized feasibility gives $\mathbb{E}_k\|W_k - \bar{W}\|^2 = \mathbb{E}_k\|W_k\|^2 - \|\bar{W}\|^2 \le 1 - 1/m$. For the second bound, convexity of squared norm yields pathwise $\|\Xi_k W_k\|^2 \le \sum_i W_{k,i}\|\xi_{k,i}\|^2 \le \sum_i \|\xi_{k,i}\|^2$, and then take conditional expectations. □

For the stored, $\mathcal{F}_k$-measurable weight w_k used by DUNAMOO, conditional unbiasedness instead gives $\mathbb{E}_k[\Xi_k w_k] = 0$ and $\mathbb{E}_k\|\Xi_k w_k\|^2 \le \sigma_g^2/N_k$. The identity $b_k = \mathbb{E}_k[\Xi_k P(W_k - \mathbb{E}_k W_k)]$ also shows that common gradient noise in all columns contributes no covariance.

Identical selection accuracy does not make the same-batch update unbiased. Under the loss-contrast and gradient moments of Proposition A.1, the matched selector has the same selection certificate as the DUNAMOO QP, with $\chi = 3/4$ and the allowance in Eq. (A.2) at $\eta = 1/(2\beta)$. This certificate concerns selection at x_k, whereas immediate application still has the covariance in Eq. (B.2). The next proposition realizes a nonzero shifted equilibrium with this matched score.

B.2 PERSISTENT BIAS WITH THE MATCHED LEARNING-RATE SCORE

Proposition B.1 (Same-batch bias with the learning-rate score and uniform oracle moments)**.** *There are nonnegative smooth convex sample losses satisfying the population assumptions of Theorem 1 and the globally uniform loss and gradient moments of Proposition A.1, for which the matched same-batch rule Eq.* (B.1) *has a shifted equilibrium when the score coefficient equals the learning rate. Specifically, let Z be a Rademacher variable and take*

$$\ell_{1,2}(x; Z) = \frac{x^2}{2} \pm Z\left(\frac{1}{2}\sin x + 1\right) + 2.$$

Both population objectives are $f_i(x) = x^2/2 + 2$, with $x_\star = 0$, $f_i^\star = 2$, $\beta = \mu_L = 1$ and $M_f = 0$. Use one fresh observation per update, exact optimal values and exact normalized weight selection, with constant step and score coefficient η, $0 < \eta \le 1/2$ and $x_0 = 0$. The same-batch iterates converge deterministically to the unique $\bar{x} \in [-1/2, 0]$ solving

$$\bar{x} + \tfrac{1}{2}\cos\bar{x} = 0, \qquad \bar{x} \approx -0.450184.$$

This limit is independent of η. For DUNAMOO *on the same oracle with the same parameters and uniform initial weights, instead*

$$\mathbb{E}x_k = 0, \qquad \mathbb{E}x_k^2 \le \frac{\eta}{4(2-\eta)}\left\{1 - (1-\eta)^{2k}\right\} \quad (k \ge 0).$$

Thus its mean-square error envelope vanishes as $\eta \downarrow 0$, whereas the same-batch squared-distance limit is $\bar{x}^2 > 0$.

Proof. Each sample loss is at least $1/2$ and has second derivative $1 \mp \frac{1}{2}Z\sin x \in [1/2, 3/2]$, so it is convex and globally $3/2$-smooth. Taking expectations gives the stated aligned quadratic population objectives. Moreover, for each task and every x,

$$\mathbb{E}|\ell_i(x; Z) - f_i(x)|^2 = \left(\tfrac{1}{2}\sin x + 1\right)^2 \le \frac{9}{4}, \qquad \mathbb{E}|\nabla\ell_i(x; Z) - \nabla f_i(x)|^2 = \frac{1}{4}\cos^2 x \le \frac{1}{4}.$$

The sample gradients are unbiased. The per-loss sufficient condition has $\nu_f = 3/2$ and $\sigma_g = 1/2$, hence Eq. (2) holds with $\nu_\Delta = 3/\sqrt{2}$.

For a realized Z, write $u = Z(w_1 - w_2) \in [-1, 1]$. The empirical gap and gradient of the mixture are

$$\widehat{\Delta}^\top w = \frac{x^2}{2} + u\left(\frac{1}{2}\sin x + 1\right), \qquad \widehat{J}w = x + \frac{u}{2}\cos x.$$

Define $A_\eta(x) = 1 + \frac{1}{2}\sin x - \frac{\eta}{2}x\cos x$. With the score coefficient tied to η, the normalized-weight score becomes

$$(1-\eta)x^2 + 2A_\eta(x)u - \frac{\eta}{4}\cos^2 x\, u^2.$$

On $[-1/2, 0]$, $A_\eta(x) \geq 3/4$, since $\sin x \geq -1/2$ and $x\cos x \leq 0$. Its derivative with respect to u therefore satisfies, for every $0 < \eta \leq 1/2$,

$$2A_\eta(x) - \frac{\eta}{2}\cos^2 x\, u \geq \frac{3}{2} - \frac{\eta}{2} \geq \frac{5}{4} > 0 \qquad (-1 \leq u \leq 1).$$

The unique maximizing aggregate is $u = 1$. Hence the same-batch update is the deterministic map

$$T_\eta(x) = (1-\eta)x - \frac{\eta}{2}\cos x.$$

For $x \in [-1/2, 0]$ and $0 < \eta \leq 1/2$,

$$0 \leq 1 - \frac{3}{2}\eta \leq T'_\eta(x) = 1 - \eta + \frac{\eta}{2}\sin x \leq 1 - \eta < 1.$$

The map is increasing, with

$$T_\eta(-1/2) = -\frac{1}{2} + \frac{\eta}{2}\{1 - \cos(1/2)\} \geq -\frac{1}{2}, \qquad T_\eta(0) = -\frac{\eta}{2} \leq 0.$$

It consequently maps $[-1/2, 0]$ into itself. The function $x + \frac{1}{2}\cos x$ is strictly increasing there and has opposite signs at the two endpoints, so it has the unique zero $\bar{x}$ stated above. The mean-value theorem yields $|T_\eta(x) - \bar{x}| \leq (1-\eta)|x - \bar{x}|$ on this interval. Starting from zero, the iterates therefore converge to $\bar{x}$ for every permitted η, independently of the batch signs.

For the delayed rule, let $a_k = w_{k,1} - w_{k,2} \in [-1, 1]$. Predictability and fresh sampling give

$$G_k = x_k + \frac{a_k Z_k}{2}\cos x_k, \qquad \mathbb{E}_k G_k = x_k, \qquad \mathbb{E}_k|G_k - x_k|^2 \leq \frac{1}{4}.$$

These bounds hold on all of $\mathbb{R}$, so no invariant interval is needed for this part. The squared update satisfies

$$\mathbb{E}_k x_{k+1}^2 \leq (1-\eta)^2 x_k^2 + \frac{\eta^2}{4}.$$

Since $x_0 = 0$, geometric summation gives the stated mean-square bound, and $\mathbb{E}_k x_{k+1} = (1-\eta)x_k$ gives $\mathbb{E}x_k = 0$. Uniform initial weights also make the first update noise-free. The conclusion is a mean-square bound for this example, and it does not assert fixed-batch point convergence of the general normalized-weight algorithm. □

A finite-horizon check with matched learning rates. Table B.1 evaluates the proposition's specified stochastic oracle at $x_0 = 0$ for all $2^{12} = 4096$ equally likely sign histories, using the same signs and learning rate for both rules. The two-task QP is solved in closed form, without ridge or estimated optimal values. Each delayed weight uses only the previous history. This calculation distinguishes the bias mechanism from the independently tuned neural-network comparison, since at $x = 0$ the same-batch combined gradient is $1/2$, whereas the delayed conditional mean gradient is zero. The table is an exhaustive finite-history average in double precision, not a Monte Carlo estimate, and therefore has no sampling confidence interval.

η	Same-batch $\mathbb{E}x_{12}^2$	Delayed $\mathbb{E}x_{12}^2$	Delayed bound
0.5	0.20265876	0.07718248	0.08333333
0.1	0.12073433	0.01175183	0.01210834
0.02	0.01154226	0.00090570	0.00097025

Table B.1: **Matched-step diagnostic on the stated analytic oracle.** The last column is $\eta\{1-(1-\eta)^{24}\}/[4(2-\eta)]$. The delayed mean has magnitude below 2×10^{-17} in all three cases. The closed-form weights satisfy simplex feasibility and the Frank–Wolfe residual check to below 10^{-12}. These checks concern this diagnostic only, not the Multi-MNIST solver.

B.3 Well-posedness of the cone score

Proposition B.2 (Cone maximizers under exact alignment versus noisy gaps)**.** *For exact gradients and exact gaps of differentiable convex aligned objectives, the score $2\Delta(x)^\top u - \|J(x)u\|^2$ of Efroni et al. (2025) attains a finite maximum on $u\in\mathbb{R}^m_+$. In contrast, unbiased gradient and gap estimates with globally uniform centered second moments, even from nonnegative convex smooth sample losses, need not give a finite maximum for its stochastic score. The normalized-weight scores in this paper always attain a finite maximum for finite-valued input arrays.*

Proof. For exact feedback fix $\mathbf{d}=x-x_\star$, $g_i=\nabla f_i(x)$ and $\Delta_i=f_i(x)-f_i(x_\star)$. Convexity and alignment give $0\le\Delta_i\le\langle\mathbf{d},g_i\rangle$. Consider the finitely generated cone

$$\mathcal{C}=\left\{(z,a)=\sum_i u_i(g_i,\Delta_i):u_i\ge0\right\}.$$

The cone is closed. Indeed, if the active generators in a conic representation are linearly dependent, choose a nonzero dependence and reverse its sign if necessary so that some coefficient is positive. Subtracting the largest feasible multiple of that dependence preserves the represented vector, keeps all coefficients nonnegative, and removes an active generator. Repetition leaves a linearly independent active set. For a convergent sequence in $\mathcal{C}$, finitely many possible active sets permit passage to a subsequence with a fixed such set. Its continuous linear inverse yields nonnegative limiting coefficients, so the limit is also in $\mathcal{C}$.

On $\mathcal{C}$, $0\le a\le\langle\mathbf{d},z\rangle$. The nonempty superlevel set $\mathcal{K}=\{(z,a)\in\mathcal{C}:2a-\|z\|^2\ge0\}$ is closed and bounded, because its points satisfy $\|z-\mathbf{d}\|^2\le\|\mathbf{d}\|^2$ and $a\le\|\mathbf{d}\|\|z\|$. It is therefore compact. The continuous score attains a maximum on $\mathcal{K}$. This is a global maximum because all omitted points have negative score whereas the origin has score zero. A finite conic representation supplies an optimal u. At $x=x_\star$ the score is identically zero. The argument permits singular Jacobians. The underlying rule is due to Efroni et al. (2025).

For the stochastic assertion let S,Z be independent Rademacher variables and use

$$\ell_{1,2}(x;S,Z)=\tfrac12x^2\ \pm Z\sin x+S+3.$$

Each sample loss is at least one and has second derivative $1\mp Z\sin x\in[0,2]$, so it is convex and globally 2-smooth. The population objectives remain $f_1=f_2=x^2/2+3$, with $x_\star=0$, $\mu_{\mathrm{L}}=\beta=1$, and $M_{\mathrm{f}}=0$. Their centered loss variance is $1+\sin^2x\le2$, and their centered gradient variance is $\cos^2x\le1$, uniformly in x. Thus the example meets the uniform primitive-moment conditions of Proposition A.1. Use the known population minimum 3 in the empirical gaps. At $x=0$,

$$\widehat{\Delta}=(S,S)^\top,\qquad\widehat{J}=(Z,-Z),\qquad\mathbb{E}\widehat{\Delta}=0,\quad\mathbb{E}\widehat{J}=0.$$

At this point both estimates are unbiased and each gradient column has variance one, and the bounds above hold globally. On $S=1$, along the nonnegative ray $u=(t,t)$,

$$2\widehat{\Delta}^\top u-\|\widehat{J}u\|^2=4t\longrightarrow\infty.$$

Thus with probability $1/2$ the stochastic cone argmax does not exist. The noisy gaps violate the exact-alignment inequality $\widehat{\Delta}_i\le\langle\mathbf{d},\widehat{g}_i\rangle$ used in the deterministic result. Finally, each normalized-weight score is a continuous function on the compact set Δ_m, so it attains a finite maximum. Lemma C.2 supplies a measurable choice. □

This obstruction concerns precisely the noisy cone score displayed above. Simplex feasibility additionally gives a uniform weight-norm bound. Clipping negative empirical gaps does not resolve the example, since both gaps are positive on the diverging event.

C SHARED ANALYTIC AND PROBABILISTIC TOOLS

This appendix records the geometry, measurable selection and random-time conditioning used in the proofs. Standard smooth-convex and conditional-expectation rules follow Boyd & Vandenberghe (2004); Durrett (2019). Initial states and step schedules are deterministic unless conditioned on an independent initialization.

C.1 OBJECTIVE GEOMETRY AND MEASURABLE SELECTION

Lemma C.1 (Elementary objective geometry)**.** *Under alignment and Assumption 1, every fixed simplex mixture is convex and β-smooth, with $0 \preceq \nabla^2 f_w \preceq \beta I$ and $\nabla f_w(x_\star) = 0$. The maximum defining μ_{L} is attained. If $\mu_{\mathrm{L}} > 0$, the common minimizer is unique. Under Assumption 2 with $M_{\mathrm{f}} = 0$, the objectives are quadratic, possibly with singular Hessians.*

Proof. The mixture claims follow by taking convex combinations of the component Hessians. The minimum eigenvalue is continuous, so compactness of Δ_m gives an attaining mixture. Its positive curvature at $x_\star$ precludes a second common minimizer, since convexity would make this mixture constant on the joining segment. When $M_{\mathrm{f}} = 0$, the third directional derivatives vanish, and integration along lines gives a quadratic representation. □

Lemma C.2 (A measurable exact rule on a compact weight set)**.** *Let $C \subset \mathbb{R}^m$ be nonempty, compact and convex, and let $Q(z, w)$ be jointly continuous in finite-dimensional data z and $w \in C$, and concave in w. Its maximizers form a nonempty compact convex set. Selecting their unique minimum-Euclidean-norm element defines a Borel-measurable map of z.*

Proof. For $e > 0$, let $w_e(z)$ maximize $Q(z, w) - e\|w\|^2$ over C. Strict concavity makes this maximizer unique. If $z_j \to z$, compactness and the optimality inequalities imply that every convergent subsequence of $w_e(z_j)$ is optimal at z. Uniqueness therefore makes w_e continuous. Let $v(z)$ be the minimum-norm maximizer of $Q(z, \cdot)$, unique because norm squared is strictly convex on the convex maximizer set. Optimality gives

$$0 \le \max_C Q(z, \cdot) - Q(z, w_e(z)) \le e\{\|v(z)\|^2 - \|w_e(z)\|^2\}.$$

Thus $\|w_e(z)\| \le \|v(z)\|$, and every limit as $e \downarrow 0$ is an unregularized maximizer of minimum norm. Hence $w_{1/j}(z) \to v(z)$, a pointwise limit of continuous maps and therefore Borel measurable. General measurable-maximization results appear in Aliprantis & Border (2006, Chapter 18). □

The empirical simplex score satisfies these hypotheses even when its Gram is singular. Jointly measurable oracle arrays can therefore be composed with this exact tie rule. Inexact solvers instead return feasible measurable weights satisfying the residual condition in Appendix A.5.3.

C.2 FINITE MOMENTS AND CONDITIONING AT RANDOM TIMES

For an integer-valued stopping time S, the stopped history $\mathcal{F}_S$ consists of ambient-measurable events A such that $A \cap \{S = j\} \in \mathcal{F}_j$ for every finite j.

Lemma C.3 (Moment existence and conditioning at integer-valued random times)**.** *Under alignment and Assumption 1, consider the unprojected simplex updates used here. Suppose $\mathbb{E}V_0 < \infty$, the predictable steps are bounded, and the gradient-noise columns have uniformly bounded conditional second moments. Every finite-time squared distance and* DUNAMOO *conditional energy is integrable. For an integer-valued stopping time S and nonnegative adapted Y_j,*

$$\mathbb{E}[Y_{S+1}\mathbf{1}_{\{S<\infty\}} \mid \mathcal{F}_S] = \sum_{j=0}^{\infty} \mathbf{1}_{\{S=j\}}\mathbb{E}[Y_{j+1} \mid \mathcal{F}_j].$$

Random-index variables are defined as zero on $\{S = \infty\}$, and the identity initially permits nonnegative extended conditional expectations.

Proof. Write $\widehat{J}_k = J(x_k) + \Xi_k$, with columns $\xi_{k,i}$ in Ξ_k. For any simplex weight W_k, including a current-batch selector, let $e_k = \Xi_k W_k$. Smoothness and alignment give $\|J(x_k)W_k\| \le \beta\sqrt{V_k}$, while simplex feasibility gives pathwise

$$\|e_k\|^2 \le \sum_i W_{k,i}\|\xi_{k,i}\|^2 \le \sum_i \|\xi_{k,i}\|^2.$$

Consequently

$$\mathbb{E}_k V_{k+1} \le 2(1+\beta\eta_k)^2 V_k + 2\eta_k^2 \mathbb{E}_k\|e_k\|^2.$$

The tower property proves finite-time integrability by induction. For predictable W_k, the standing variance bound gives the sharper $\mathbb{E}_k\|e_k\|^2 \le \sigma^2$. Conditional expectation preserves integrability, so $\mathcal{L}_k$ is integrable. Finite sums of these nonnegative quantities dominate bounded-horizon stopped values.

The random-time identity follows by integration over the stopped history. The displayed sum is $\mathcal{F}_S$-measurable. Integrating it over any $A \in \mathcal{F}_S$, exchanging the nonnegative sum and integral, and applying the defining property of conditional expectation yields $\sum_j \mathbb{E}[Y_{j+1}\mathbf{1}_{A\cap\{S=j\}}]$. This equals the integral of the left-hand random variable on A, proving the identity. Applying this argument at $T+n$ and then using the tower property justifies the stopping-time conditioning in Appendix A.3. No unconditional moment at an unbounded random time is assumed, because nonnegative conditional expectations apply first, and the bounded entry energy and ensuing recurrence supply finite conditional bounds. See Durrett (2019, Sections 4.1 and 4.8). □

D Experimental details

D.1 Multi-MNIST

Let m denote the number of positional digit-classification objectives. The implementation supports $m \in \{2, 4, 8\}$. Each example Z consists of a composite image containing m non-overlapping MNIST digits and a label vector $y_{\mathsf{Z}} = (y_{\mathsf{Z},1}, \ldots, y_{\mathsf{Z},m})$. Objective i predicts the digit at the corresponding position. Figures D.1–D.3 show one example for each variant, using the label-grid order.

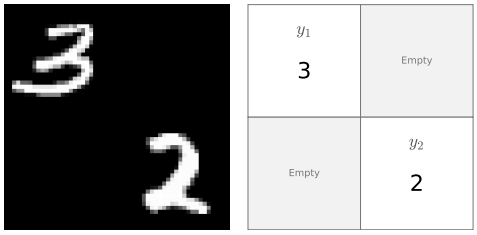


Figure D.1: Example of a two-digit Multi-MNIST image (left) and its label (right).

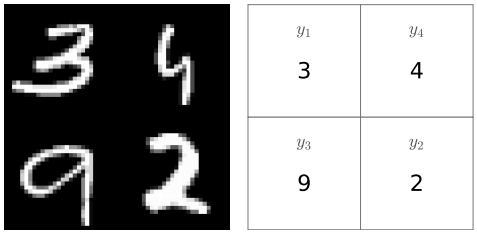


Figure D.2: Example of a four-digit Multi-MNIST image (left) and its label (right).

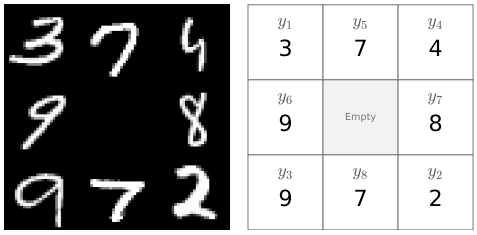


Figure D.3: Example of an eight-digit Multi-MNIST image (left) and its label (right).

Data construction and randomization. Let N_{tr} denote the number of training examples. For each seed, we construct a training dataset $\mathcal{D}_{\mathrm{tr}} = \{(\mathsf{Z}_n, y_n)\}_{n=1}^{N_{\mathrm{tr}}}$ from the official MNIST training partition. Across all variants $m \in \{2, 4, 8\}$, the top-left digit, corresponding to objective 1, is sampled from MNIST images with labels in $\{1, 3, 5, 7, 9\}$, while the bottom-right digit, corresponding to objective 2, is sampled from images with labels in $\{0, 2, 4, 6, 8\}$. For $m \in \{4, 8\}$, the remaining positions are sampled from the full MNIST training pool without parity restrictions. These restrictions apply to data construction, and every objective retains ten output logits and predicts the original digit identity. Source images are sampled without replacement from the corresponding digit pool for each position, while reuse across different positions is permitted, so the requested sample count must not exceed the pool size available for any position.

Pixel intensities are converted to `float32` and divided by 255.

The training dataset is regenerated for every experiment seed. Data construction, minibatch sampling, and model initialization use separate deterministic seed streams. Within each dataset variant and seed, all methods share the same constructed training examples, sample-loss multipliers, initial model parameters, and complete minibatch index sequence.

Model architecture. We use a convolutional neural network (CNN) that processes the entire composite image with a shared feature extractor and a fully connected layer producing ten logits per objective. Let $h_x(\mathsf{Z}) \in \mathbb{R}^{d_m}$ denote the shared feature representation. The final fully connected layer produces

$$z_x(\mathsf{Z}) = W h_x(\mathsf{Z}) + b \in \mathbb{R}^{10m}, \qquad W \in \mathbb{R}^{10m \times d_m}, \qquad b \in \mathbb{R}^{10m}.$$

This output is reshaped into

$$Z_x(\mathsf{Z}) = \mathrm{reshape}_{m \times 10}\big(z_x(\mathsf{Z})\big),$$

where the j-th row contains the ten class logits for objective j. Thus, the final layer produces 20, 40, and 80 logits for $m = 2$, 4, and 8, respectively. For a minibatch of size B, the output tensor has shape

$$B \times m \times 10.$$

For $m = 2$ and $m = 4$, the shared feature extractor consists of three convolutional blocks with 8, 16, and 32 channels. For $m = 8$, an additional block with 64 channels is used. Each block comprises a 3×3 convolution with padding 1, a ReLU activation, and 2×2 max pooling. The resulting feature maps have dimensions $32 \times 7 \times 7$ for the 56×56 inputs ($m = 2, 4$), and $64 \times 5 \times 5$ for the 84×84 inputs ($m = 8$). These features are flattened and directly mapped to $10m$ logits by a single fully connected layer, without an intermediate fully connected hidden layer. The models contain 37,268, 68,648, and 152,464 trainable parameters for $m = 2$, 4, and 8, respectively.

Sample-dependent loss multipliers. For an input Z_n, let $Z_x(\mathsf{Z}_n) \in \mathbb{R}^{m \times 10}$ denote the logit matrix produced by the network. The predictive distribution for objective $i \in \{1, \ldots, m\}$ is

$$p_{x,i}(k \mid \mathsf{Z}_n) = \frac{\exp\big([Z_x(\mathsf{Z}_n)]_{i,k}\big)}{\sum_{r=1}^{10} \exp\big([Z_x(\mathsf{Z}_n)]_{i,r}\big)}, \qquad k \in \{1, \ldots, 10\}.$$

The unweighted per-example cross-entropy is

$$\ell_i(x; \mathsf{Z}_n) = -\log p_{x,i}(y_{\mathsf{Z}_n,i} \mid \mathsf{Z}_n).$$

Let $c \in [0, 1]$ denote the sample-loss contrast and $q^{(m)} \in \mathbb{R}^m$ the fixed contrast-coefficient vector in objective order.

Each training example receives a fixed regime variable $r_n \in \{-1, +1\}$, assigned without using its digit labels. Let ρ denote the reversed-regime fraction, fixed at $\rho = 0.9$ in the dataset constructor. The numbers of reversed and forward examples are

$$N_- = \left\lfloor \rho N_{\mathrm{tr}} + \frac{1}{2} \right\rfloor, \qquad N_+ = N_{\mathrm{tr}} - N_-.$$

These assignments are randomly shuffled and remain fixed throughout training. The sample-loss multiplier for example Z and objective i is

$$a_{\mathsf{Z},i} = 1 + c r_{\mathsf{Z}} q_i^{(m)}, \qquad i = 1, \ldots, m.$$

The forward regime has descending multipliers in objective order, and the reversed regime has the opposite ordering. Since the coefficients sum to zero, every example satisfies

$$\sum_{i=1}^{m} a_{\mathsf{Z},i} = m.$$

These multipliers change the relative training contributions of all objectives without changing the images or labels. They are distinct from the optimizer-selected objective weights and do not prescribe an ordering of the unweighted losses. For $c = 0$, all sample-loss multipliers equal one. The empirical training objectives are

$$\widehat{f}_i(x) = \frac{1}{N_{\mathrm{tr}}} \sum_{n=1}^{N_{\mathrm{tr}}} a_{\mathsf{Z}_n,i} \ell_i(x; \mathsf{Z}_n).$$

Our experiments use the following contrast-coefficient vectors.

$$\begin{aligned} q^{(2)} &= (1, -1), \\ q^{(4)} &= \left(1, \frac{1}{2}, -\frac{1}{2}, -1\right), \\ q^{(8)} &= \left(1, \frac{3}{4}, \frac{1}{2}, \frac{1}{4}, -\frac{1}{4}, -\frac{1}{2}, -\frac{3}{4}, -1\right). \end{aligned}$$

Each vector is ordered by objective index, is symmetric about zero, and satisfies $\sum_{i=1}^{m} q_i^{(m)} = 0$.

We consider two contrast levels, $c \in \{0.5, 0.8\}$. For $c = 0.5$, the multipliers range from 0.5 to 1.5, giving a largest-to-smallest multiplier ratio of 3. For $c = 0.8$, they range from 0.2 to 1.8, increasing this ratio to 9. Thus, larger contrast introduces stronger differences in the relative training contributions of objectives while preserving the total multiplier $\sum_i a_{\mathsf{Z},i} = m$ for every example. The reversed-regime fraction remains fixed at $\rho = 0.9$ across all contrast settings.

Unweighted evaluation and objective-wise learning. The sample-dependent multipliers introduce unequal loss scales and minibatch fluctuations during training, and they are not intended to encode a preference for improving some positional tasks at the expense of others. We therefore distinguish the objectives used for optimization from the quantities used to monitor task-wise learning. All methods train on the multiplier-weighted objectives $\widehat{f}_i$, and the adaptive weight solvers use their corresponding losses and gradients. For Figure 4 and the objective-wise plots in Appendix D.3.2, however, we report the unweighted training cross-entropy

$$\bar{\ell}_i(x) := \frac{1}{N_{\mathrm{tr}}} \sum_{n=1}^{N_{\mathrm{tr}}} \ell_i(x; \mathsf{Z}_n).$$

This evaluates each positional classifier on the original cross-entropy scale, without including the imposed sample-loss multipliers in the reported metric. A small multiplier-weighted loss can reflect downweighting as well as accurate prediction, while an aggregate loss can obscure persistent errors on individual objectives. The unweighted objective-wise trajectories therefore allow us to examine whether learning progresses across all tasks, including those receiving weaker training contributions.

Configuration	Setting
Objectives	Positional digit-classification tasks with $m \in \{2, 4, 8\}$ objectives
Training data	1024 composite images
Sample-loss contrast	$c \in \{0.5, 0.8\}$, approximately 90:10 regime allocation
Shared model	Three ($m = 2, 4$) or four ($m = 8$) convolutional blocks and a shared linear classifier
Training batch size	32 composite images
Training budget	50 epochs
Configured seeds	10 seeds, 123456–123465
Learning-rate	$\{0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1\}$

Table D.1: Configuration of the Multi-MNIST experiments.

The final-worst-objective summary in Figure 4 complements these trajectories by following the task with the largest seed-averaged unweighted loss at the end of training. Removing the multipliers from evaluation does not remove their influence on the optimization trajectory.

Relation to the AMOO setting. This evaluation is motivated by the AMOO goal of minimizing all objectives simultaneously, rather than selecting a preferred trade-off among tasks. The multiplier construction leaves the images and prediction targets unchanged. Moreover, for the tested contrasts $c \in \{0.5, 0.8\}$, every multiplier satisfies $0 < 1-c \le a_{\mathsf{Z},i} \le 1+c$. Since cross-entropy is nonnegative, we have, for every objective and parameter vector,

$$(1-c)\bar{\ell}_i(x) \;\le\; \widehat{f}_i(x) \;\le\; (1+c)\bar{\ell}_i(x).$$

Consequently, along any parameter sequence, all weighted training losses approach zero if and only if all unweighted training losses approach zero. Thus, whenever simultaneous near-zero prediction losses are attainable, the multipliers preserve this common limiting target while changing the relative loss and gradient contributions encountered during training. The unweighted plots assess progress toward this simultaneous-learning goal under deliberately imbalanced optimization conditions. This motivates the experiment as an AMOO-oriented setting, but does not establish an attained common minimizer for the shared CNN or verify the assumptions of the convergence theorem.

D.2 Methods and optimization settings

We compare EW-SGD, a ridge-regularized cone implementation of PAMOO-SGD (Efroni et al., 2025), and our implementation of DunaMOO-SGD. Table D.1 summarizes the configuration. The same optimization protocol is used for the two-digit, four-digit, and eight-digit Multi-MNIST variants, with $m \in \{2, 4, 8\}$ denoting the number of positional digit-classification objectives. At the pre-update parameters x_k and minibatch $\mathcal{B}_k$, the weight solvers use the multiplier-weighted minibatch losses $\widehat{f}_{k,i}(x_k)$ of Section 6 and their gradients,

$$\widehat{g}_{k,i} = \nabla_x \widehat{f}_{k,i}(x_k) = \frac{1}{|\mathcal{B}_k|} \sum_{\mathsf{Z} \in \mathcal{B}_k} a_{\mathsf{Z},i} \nabla_x \ell_i(x_k; \mathsf{Z}), \qquad \widehat{J}_k = [\widehat{g}_{k,1}, \ldots, \widehat{g}_{k,m}].$$

The gradients are taken with respect to all shared model parameters. The weight solvers use zero as the supplied optimal values for all objectives, so that

$$\widehat{\Delta}_k = \big(\widehat{f}_{k,1}(x_k), \ldots, \widehat{f}_{k,m}(x_k)\big)^\top.$$

Zero is a lower-bound reference for cross-entropy, not a verified optimal value for the shared neural-network objectives.

Training protocol. All algorithms use SGD without momentum or weight decay. With 1024 training examples and a minibatch size of 32, each epoch contains 32 updates, and training lasts for 50 epochs. Minibatch indices are sampled independently and uniformly with replacement, so an epoch is a block of updates rather than a without-replacement pass through the training set. The default configuration uses 10 experiment seeds, 123456–123465. Within each seed, all conditions share the constructed training dataset, initial parameter values, and complete minibatch index sequence.

Learning-rate schedule. For each $m \in \{2, 4, 8\}$, the current sweep configuration combines initial learning rates $\eta_0 \in \{0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1\}$ with sample-loss contrasts $c \in \{0.5, 0.8\}$. Each defines a separate comparison in which every method uses the same external SGD learning rate and decay schedule. The learning rate is halved after epochs 20 and 40, with the reduced rate taking effect in the following epoch. DUNAMOO-SGD uses the known next-update rate η_{k+1}, including when the current minibatch is the last one before a scheduled decay.

EW-SGD. Equal weighting uses the fixed vector $\bar{w} = \mathbf{1}/m$ and updates

$$x_{k+1} = x_k - \eta_k \widehat{J}_k \bar{w}.$$

The implementation computes the mean of the weighted training objectives and performs a single backward pass. It does not construct a gradient Gram matrix for this condition.

PAMOO-SGD (implemented cone variant). PAMOO-SGD selects nonnegative weights by solving

$$u_k = \arg\max_{u \in \mathbb{R}^m_+} \left\{ 2\widehat{\Delta}_k^\top u - u^\top (\widehat{J}_k^\top \widehat{J}_k + \lambda I) u \right\}, \qquad \lambda = 10^{-4}.$$

The ridge $\lambda > 0$ makes the cone score strictly concave and bounded above, so a maximizer exists and is unique even when the empirical Gram is singular or the noisy gaps would otherwise leave the unregularized score unbounded (Proposition B.2). It is a numerical safeguard rather than part of the analyzed rule. The selected weights are applied immediately to the current minibatch,

$$x_{k+1} = x_k - \eta_k \widehat{J}_k u_k.$$

The weights are not normalized, so their magnitude enters the update magnitude directly.

Remark D.1. This implementation uses an active-set quadratic-program solver with full parameter gradients. It differs from the projected-gradient solver, positive coordinate floor and separate external learning rate of the published PAMOO-SGD experimental baseline (Efroni et al., 2025, Appendix A). Accordingly, this PAMOO-SGD condition is an implementation variant of that baseline.

DUNAMOO-SGD. DUNAMOO-SGD initializes $w_0 = \bar{w}$ and selects

$$w_{k+1} = \arg\max_{w \in \Delta_m} \left\{ 2\widehat{\Delta}_k^\top w - \eta_{k+1} w^\top (\widehat{J}_k^\top \widehat{J}_k + \lambda I) w \right\},$$

where η_{k+1} is the learning rate of the next update, as in Algorithm 1, and $\lambda = 10^{-4}$. The current update uses the stored weight,

$$x_{k+1} = x_k - \eta_k \widehat{J}_k w_k.$$

Both computations use the same current minibatch and pre-update parameters. The newly selected weight is stored for the next update, including across epoch boundaries. Only weight application is delayed, and gradients are evaluated at the current parameters. The simplex constraint is imposed directly in the optimization problem rather than through post-normalization.

The implemented selector therefore differs from the score of Algorithm 1 only by the ridge, which makes the exact optimizer unique when the empirical Gram is singular. With the halving schedule, η_{k+1} differs from η_k only at the two decay epochs.

Same-batch simplex control. The control selects

$$W_k = \arg\max_{w \in \Delta_m} \left\{ 2\widehat{\Delta}_k^\top w - \eta_{k+1} w^\top (\widehat{J}_k^\top \widehat{J}_k + \lambda I_m) w \right\},$$

and uses it in the current update,

$$x_{k+1} = x_k - \eta_k \widehat{J}_k W_k.$$

This is the ridge-regularized counterpart of the matched rule Eq. (B.1). It retains the same next-step coefficient η_{k+1} as DUNAMOO-SGD, even at a decay boundary where $\eta_{k+1} \neq \eta_k$. The score, constraint, ridge, solver, learning-rate decay schedule, initialization, and minibatch sequence are matched, the initial learning rate is selected separately for each rule as in Appendix D.4.1, and the prescribed difference is whether the selected weight is applied immediately or one iteration later.

Quadratic-program solver. We solve the ridge-regularized QPs in `float64` using an active-set search with support warm starts and enumeration as a fallback. The solver returns the first candidate satisfying its numerical Karush–Kuhn–Tucker conditions.

Differentiation and computational work. Each adaptive step computes the m objective gradients once and reuses them for weight selection and the SGD update, without an additional backward pass. EW-SGD uses a single backward pass. No gradient is propagated through the QP solution.

D.3 Experimental results

D.3.1 Additional details on Figure 4

For clarity, we specify how the curves in Figure 4 are constructed from the objective-wise losses across seeds. For a fixed method and initial learning rate, let $x_t^{(s)}$ denote the model parameters after epoch t for the seed s where $s \in \{1, \dots, 10\}$, and let $\mathcal{D}_{\mathrm{tr}}^{(s)}$ denote its training dataset. Making the seed dependence of the notation in Section 6 explicit, the unweighted training cross-entropy of objective i is

$$\bar{\ell}_i^{(s)}\left(x_t^{(s)}\right) = \frac{1}{N_{\mathrm{tr}}} \sum_{Z \in \mathcal{D}_{\mathrm{tr}}^{(s)}} \ell_i\left(x_t^{(s)}; Z\right).$$

This evaluation excludes the sample-dependent multipliers $a_{Z,i}$, so the reported losses are measured on the unweighted cross-entropy scale.

We first compute the arithmetic mean of each objective's loss across the ten seeds. We then select the objective with the largest seed-averaged loss at the final epoch:

$$i^\star \in \arg\max_{i \in [m]} \frac{1}{10} \sum_{s=1}^{10} \bar{\ell}_i^{(s)}\left(x_{50}^{(s)}\right).$$

Although it may differ between methods and experimental configurations, the selected index $i^\star$ is fixed across epochs. The corresponding solid curve in Figure 4 is therefore

$$t \longmapsto \frac{1}{10} \sum_{s=1}^{10} \bar{\ell}_{i^\star}^{(s)}\left(x_t^{(s)}\right), \qquad t = 0, \dots, 50.$$

Thus, the curve follows the learning trajectory of the objective that remains hardest on average at the end of training.

All panels in Figure 4 use the sample-loss contrast $c = 0.5$. For EW-SGD and DunAMOO-SGD, we set the initial learning rate to $\eta_0 = 0.1$ for all $m \in \{2, 4, 8\}$. For PAMOO-SGD, we select the initial learning rate separately for each m from the grid specified in Appendix D.2, choosing the value that minimizes the largest seed-averaged objective loss at epoch 50, as defined above. All methods follow the learning-rate decay schedule described in Appendix D.2.

D.3.2 Objective-wise Cross-entropy loss

Figures D.4–D.9 show the objective-wise cross-entropy losses at the initial learning rate $\eta_0 = 0.2$ for every method, where a panel labelled f_i shows the unweighted training cross-entropy $\bar{\ell}_i$ of objective i, and Tables D.2–D.7 report the final maximum loss across objectives for every initial learning rate, with the best value per method in bold.

Each entry is the mean over the ten seeds of the final unweighted loss across objectives, with the standard deviation over seeds in parentheses. This differs from Figure 4, which follows the objective with the largest seed-averaged loss.

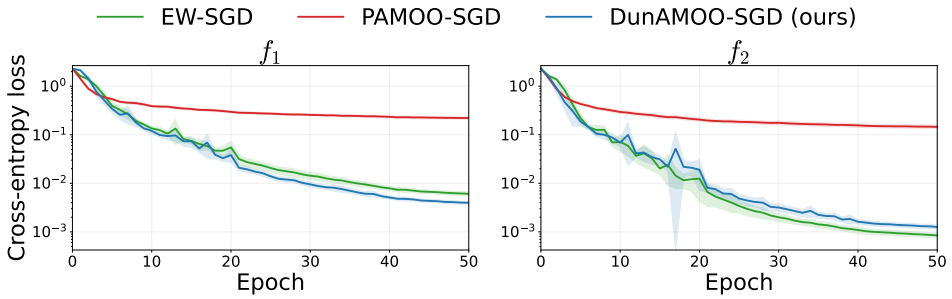


Figure D.4: Objective-wise cross-entropy losses, $m = 2$, $c = 0.5$, $\eta_0 = 0.2$.

Method / Initial LR	0.001	0.005	0.01	0.02	0.05	0.1	0.2	0.5	1
EW-SGD	1.6685 (±0.0250)	0.7889 (±0.1782)	0.3328 (±0.0480)	0.1514 (±0.0107)	0.0306 (±0.0031)	0.0061 (±0.0005)	**0.0057** (±0.0010)	1.4620 (±0.2121)	1.6146 (±0.0007)
PAMOO-SGD	1.5971 (±0.0069)	1.4194 (±0.0529)	1.1956 (±0.0727)	0.8876 (±0.0505)	0.5329 (±0.0270)	0.3634 (±0.0158)	**0.2215** (±0.0120)	0.2375 (±0.2137)	1.5320 (±0.3373)
DunAMOO-SGD (ours)	2.3269 (±0.0299)	0.6345 (±0.1533)	0.2733 (±0.0211)	0.1224 (±0.0079)	0.0210 (±0.0019)	**0.0040** (±0.0003)	0.0040 (±0.0046)	2.2086 (±0.0118)	2.1359 (±0.0173)

Table D.2: Maximum loss across objectives ($m = 2$, $c = 0.5$, average over 10 seeds).

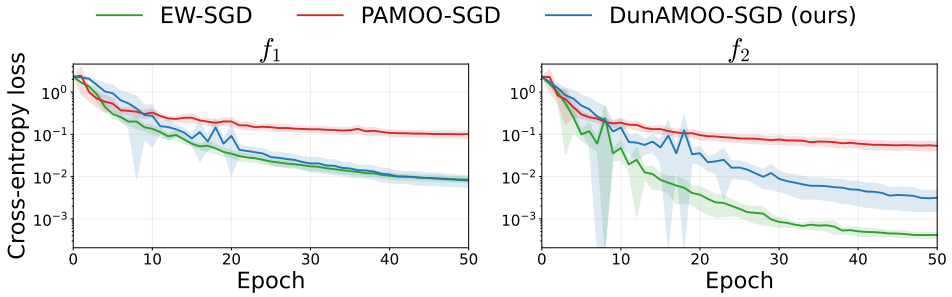


Figure D.5: Objective-wise cross-entropy losses, $m = 2$, $c = 0.8$, $\eta_0 = 0.2$.

Method / Initial LR	0.001	0.005	0.01	0.02	0.05	0.1	0.2	0.5	1
EW-SGD	1.7275 (±0.0494)	0.8599 (±0.2167)	0.4117 (±0.0726)	0.2193 (±0.0217)	0.0786 (±0.0089)	0.0269 (±0.0037)	**0.0085** (±0.0012)	1.5379 (±0.1882)	1.6238 (±0.0015)
PAMOO-SGD	1.6352 (±0.0044)	1.5068 (±0.0403)	1.3662 (±0.0705)	1.1233 (±0.0732)	0.7142 (±0.0490)	0.4393 (±0.0321)	0.2585 (±0.0191)	**0.1011** (±0.0196)	1.6011 (±0.3052)
DunAMOO-SGD (ours)	2.3433 (±0.0447)	0.7451 (±0.1567)	0.3564 (±0.0402)	0.1847 (±0.0157)	0.0571 (±0.0066)	0.0161 (±0.0019)	**0.0081** (±0.0024)	2.2491 (±0.0089)	2.1882 (±0.0195)

Table D.3: Maximum loss across objectives ($m = 2$, $c = 0.8$, average over 10 seeds).

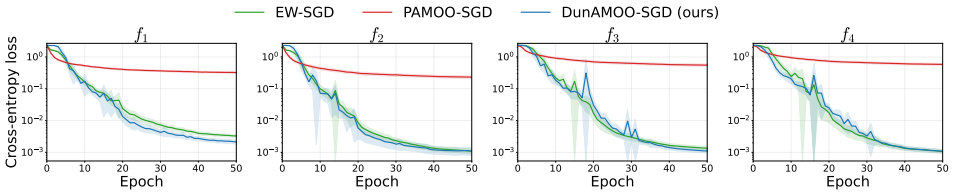


Figure D.6: Objective-wise cross-entropy losses, $m = 4$, $c = 0.5$, $\eta_0 = 0.2$.

Method / Initial LR	0.001	0.005	0.01	0.02	0.05	0.1	0.2	0.5	1
EW-SGD	2.2969 (±0.0021)	1.9440 (±0.2234)	0.8158 (±0.2888)	0.1748 (±0.0422)	0.0166 (±0.0033)	0.0033 (±0.0003)	**0.0010** (±0.0001)	0.4601 (±0.5701)	2.0735 (±0.4204)
PAMOO-SGD	2.2591 (±0.0121)	2.0900 (±0.0517)	1.8589 (±0.0619)	1.5164 (±0.0563)	0.9428 (±0.0278)	**0.5842** (±0.0415)	0.9295 (±0.5713)	2.4107 (±0.0242)	2.5605 (±0.0772)
DUNAMOO-SGD (ours)	2.3079 (±0.0098)	2.0434 (±0.2899)	0.5921 (±0.1878)	0.1360 (±0.0382)	0.0110 (±0.0022)	0.0021 (±0.0003)	**0.0006** (±0.0001)	2.1201 (±0.3386)	2.3013 (±0.0006)

Table D.4: Maximum loss across objectives ($m = 4$, $c = 0.5$, average over 10 seeds).

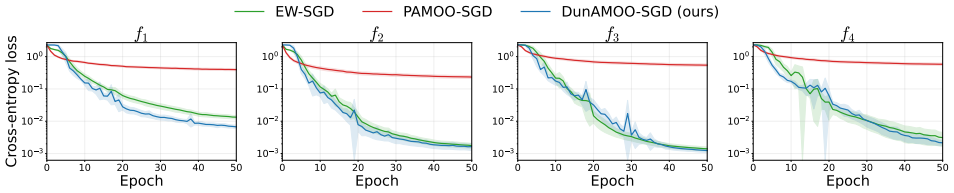


Figure D.7: Objective-wise cross-entropy losses, $m = 4$, $c = 0.8$, $\eta_0 = 0.2$.

Method / Initial LR	0.001	0.005	0.01	0.02	0.05	0.1	0.2	0.5	1
EW-SGD	2.2973 (±0.0019)	1.9938 (±0.2071)	0.8537 (±0.3193)	0.2622 (±0.0438)	0.0552 (±0.0088)	0.0134 (±0.0016)	**0.0047** (±0.0007)	0.0218 (±0.0241)	2.2977 (±0.0011)
PAMOO-SGD	2.2602 (±0.0119)	2.0951 (±0.0521)	1.8601 (±0.0608)	1.5051 (±0.0525)	0.9287 (±0.0318)	**0.5814** (±0.0484)	1.3298 (±0.6202)	2.1858 (±0.4222)	2.5469 (±0.0759)
DUNAMOO-SGD (ours)	2.3083 (±0.0104)	2.0722 (±0.2731)	0.6712 (±0.1572)	0.1853 (±0.0290)	0.0289 (±0.0049)	0.0067 (±0.0008)	**0.0023** (±0.0003)	2.3009 (±0.0012)	2.3018 (±0.0006)

Table D.5: Maximum loss across objectives ($m = 4$, $c = 0.8$, average over 10 seeds).

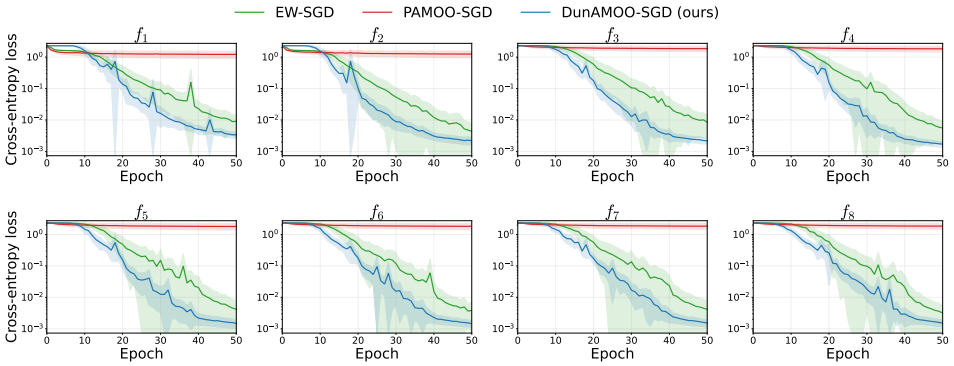


Figure D.8: Objective-wise cross-entropy losses, $m = 8$, $c = 0.5$, $\eta_0 = 0.2$.

Method / Initial LR	0.001	0.005	0.01	0.02	0.05	0.1	0.2	0.5	1
EW-SGD	2.3022 (±0.0004)	2.2969 (±0.0017)	2.2815 (±0.0116)	2.0396 (±0.1671)	0.3663 (±0.2834)	0.0092 (±0.0062)	0.0012 (±0.0002)	**0.0004** (±0.0001)	0.4607 (±0.5659)
PAMOO-SGD	2.2897 (±0.0033)	2.2450 (±0.0165)	2.1193 (±0.0147)	1.8657 (±0.0287)	**1.8546** (±0.3466)	2.3192 (±0.0044)	2.3355 (±0.0086)	2.4160 (±0.0414)	2.5658 (±0.0898)
DUNAMOO-SGD (ours)	2.3044 (±0.0052)	2.3052 (±0.0065)	2.3329 (±0.0308)	1.7185 (±0.4587)	0.0580 (±0.0410)	0.0034 (±0.0008)	**0.0009** (±0.0003)	1.1560 (±0.7093)	2.0955 (±0.3857)

Table D.6: Maximum loss across objectives ($m = 8$, $c = 0.5$, average over 10 seeds).

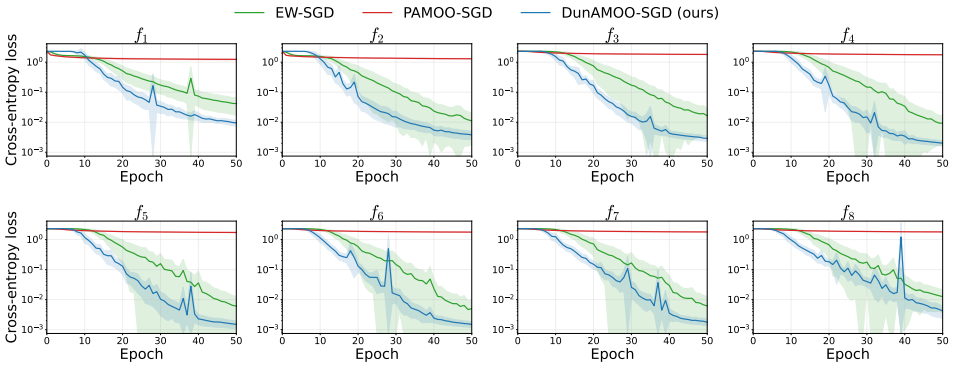


Figure D.9: Objective-wise cross-entropy losses, $m = 8$, $c = 0.8$, $\eta_0 = 0.2$.

Method / Initial LR	0.001	0.005	0.01	0.02	0.05	0.1	0.2	0.5	1
EW-SGD	2.3022 (±0.0004)	2.2991 (±0.0011)	2.2915 (±0.0045)	2.1336 (±0.1184)	0.4955 (±0.3406)	0.0419 (±0.0236)	0.0059 (±0.0013)	**0.0056** (±0.0020)	0.6926 (±0.6480)
PAMOO-SGD	2.2900 (±0.0032)	2.2414 (±0.0171)	2.1005 (±0.0163)	**1.8126** (±0.0357)	2.0883 (±0.2738)	2.3190 (±0.0045)	2.3355 (±0.0103)	2.4209 (±0.0484)	2.5469 (±0.1008)
DUNAMOO-SGD (ours)	2.3045 (±0.0053)	2.3053 (±0.0070)	2.4090 (±0.0988)	1.5227 (±0.3513)	0.0791 (±0.0361)	0.0095 (±0.0020)	**0.0053** (±0.0070)	1.1634 (±0.7027)	2.3034 (±0.0011)

Table D.7: Maximum loss across objectives ($m = 8$, $c = 0.8$, average over 10 seeds).

D.4 Additional Experiments

D.4.1 Same batch vs delayed batch

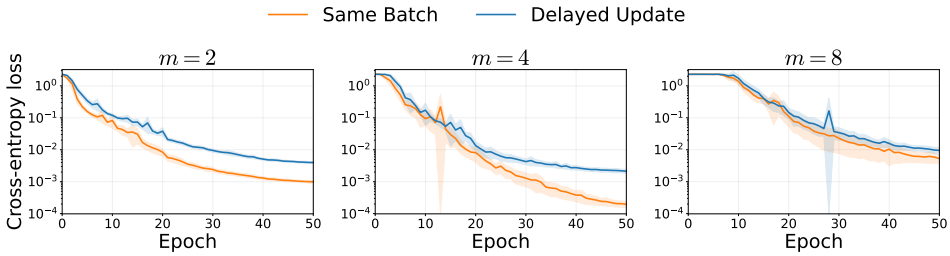


Figure D.10: **Same-batch control versus DunAMOO-SGD.** Training cross-entropy of the objective that is worst at the final epoch, as in Figure 4, averaged over seeds with pointwise approximate 95% confidence intervals, for $m = 2, 4, 8$ at the initial learning rate selected separately for each rule. In the legend, Same Batch is the same-batch control and Delayed Update is DunAMOO-SGD.

We compare the same-batch control of Appendix D.2 with DunAMOO-SGD on Multi-MNIST with $m \in \{2, 4, 8\}$ objectives, selecting the initial learning rate of each rule separately to minimize the seed-averaged final maximum objective loss (Figure D.10). For $m = 2$ and $m = 4$, the same-batch control reaches a lower final loss on every objective. For $m = 8$, the delayed rule reaches a lower final mean loss on objectives 5, 6 and 7, and the control on the remaining objectives. Since learning rates are tuned separately and the canvas layout and network change with m, these results do not isolate the effect of delay or of the objective count alone.

Delayed application is motivated by the convergence guarantee rather than by speed. Fixing the weight before the current minibatch is drawn removes the coupling between weight selection and the current gradient noise, so conditional on the history the applied weight is fixed and the expected update can be analyzed, which is what Theorem 1 uses. The same-batch rule can instead converge to a wrong point with a bias that does not vanish as the step decreases (Proposition B.1). The faster same-batch convergence for $m = 2$ and $m = 4$ therefore does not negate this motivation. It separates finite-budget optimization speed from the conditions under which the guarantee holds.

D.4.2 QUADRATIC PROGRAM VS MAXIMUM-GAP

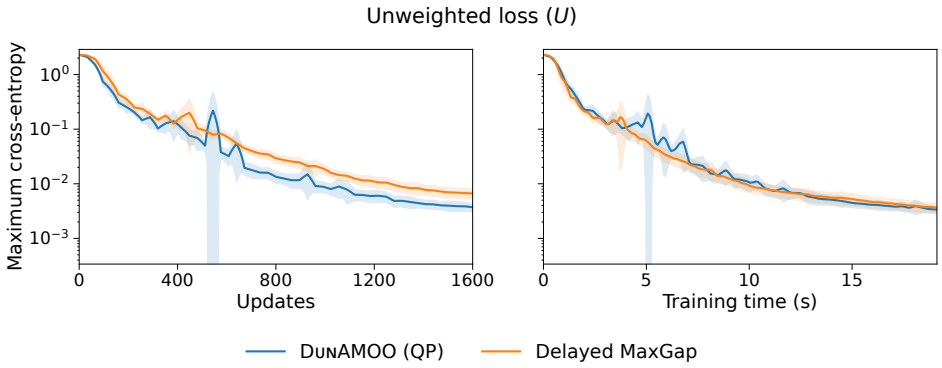


Figure D.11: Comparison of delayed QP-based weighting (DUNAMOO) and the delayed maximum-gap rule, the stored-weight version of the selection of MG-AMOO (Kretzu et al., 2025, Appendix A.7), on Multi-MNIST ($m = 2$, sample-loss contrast $c = 0.5$, initial learning rate $\eta_0 = 0.1$). The left and right panels show convergence against update count and cumulative training time, respectively. Both panels report unweighted training cross-entropy, taking the maximum across objectives. Shaded regions indicate pointwise 95% confidence intervals, and the loss axis uses a logarithmic scale. In the legend, Delayed MaxGap is the delayed maximum-gap rule. QP-based weighting achieves lower final loss under a fixed update budget, while the gap narrows under a matched training-time budget.

Benefit under a fixed update budget. An update denotes one SGD parameter update using a minibatch of 32 examples, rather than an iteration of the weight solver. Both methods apply previously selected weights to fresh minibatch gradients and share the initialization, minibatch sequence, and learning-rate schedule within each seed.

This comparison uses a separate set of 10 paired seeds, so its values differ from the entries of Table D.2. Under an identical budget of 1,600 updates, DUNAMOO reduces the seed-averaged maximum unweighted training loss from 0.006651 for the delayed maximum-gap rule to 0.003721, a reduction of approximately 44.1%. The paired difference between the QP and the maximum-gap rule is -0.002930, with a 95% confidence interval of $[-0.004285, -0.001856]$ from $10{,}000$ bootstrap resamples of paired seed blocks. Thus, the lower final loss is achieved without additional model updates or minibatch draws.

This fixed-update improvement is consistent with the value of incorporating gradient magnitudes and pairwise inner products into weight selection, beyond prioritizing the largest loss gap alone. The benefit concerns optimization quality at a fixed update budget. QP-based selection requires additional gradient and weighting computations, and the observed advantage narrows when the methods are compared under a matched training-time budget.

D.5 Synthetic checks of Theorem 1, Proposition A.2 and Proposition B.1

The Multi-MNIST experiments violate every population assumption of the theory, so they cannot say whether the certified quantities describe the algorithm where the theory applies. This subsection runs DunaMOO on the analytic instances of the paper, where β, M_{f}, μ_{L}, $x_\star$ and the optimal values are known, $\mathcal{L}_k$ can be evaluated along every path from Eq. (5) and the noise structure, and each constant of Theorem 1 is a number. Every weight is the exact maximizer of the simplex score of Algorithm 1 without ridge, obtained in closed form for $m = 2$ and by enumerating supports for $m = 3$, with the minimum-norm tie rule of Lemma C.2. The first-order residual of Eq. (A.25) is below 10^{-13} for every selected weight. All runs use $\eta = 1/2$, $w_0 = \mathbf{1}/m$ and fresh Rademacher batches in float64.

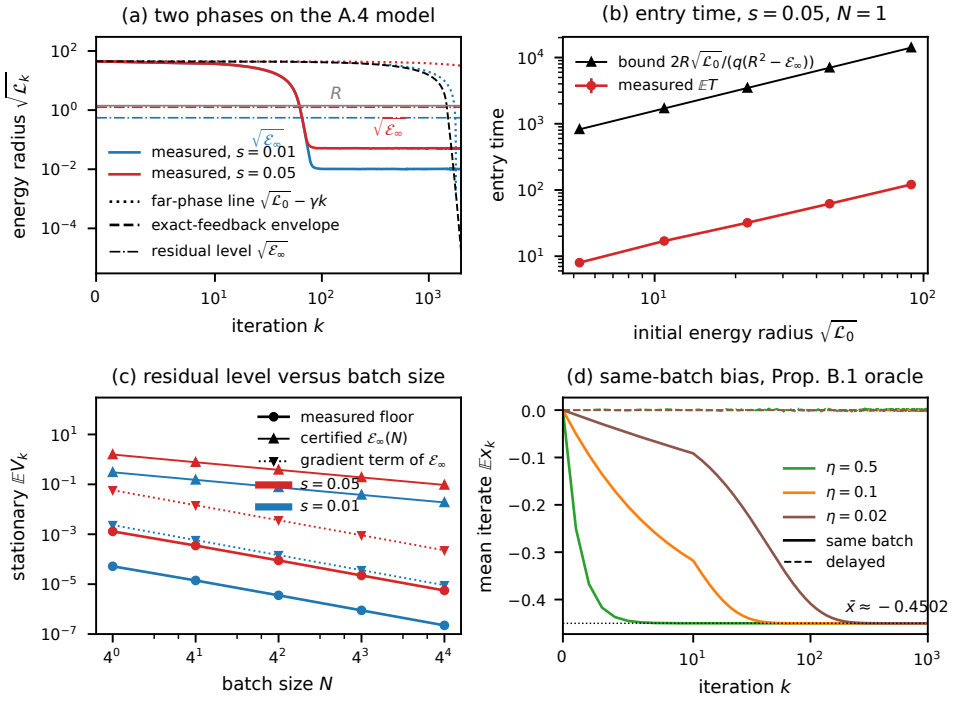


Figure D.12: **Synthetic checks on the analytic instances.** (a) Seed-averaged energy radius $\sqrt{\mathcal{L}_k}$ of DunaMOO on the model of Appendix A.4 with $m = 2$, $N = 1$ and $\sqrt{\mathcal{L}_0} = 44.7$, for two noise levels s, with the far-phase line $\sqrt{\mathcal{L}_0} - \gamma k$ of Eq. (8)(i), the exact-feedback envelope of Eq. (16), the threshold R and the residual levels $\sqrt{\mathcal{E}_\infty}$. (b) Mean entry time T over 2000 seeds against $\sqrt{\mathcal{L}_0}$ for $s = 0.05$, with the bound of Eq. (9). Error bars are two standard errors and are smaller than the markers. (c) Stationary mean squared error, averaged over the second half of 4000 iterations and over 2000 seeds, against the batch size N, with the certified residual $\mathcal{E}_\infty(N)$ of Eq. (A.24) and its gradient-noise term $(34m + 2)s^2/(3N)$. (d) Mean iterate on the oracle of Proposition B.1 over 50,000 sign sequences, for the same-batch rule and for DunaMOO, with the shifted equilibrium $\bar{x}$.

Two phases and entry time. Panel (a) of Figure D.12 starts the model of Appendix A.4 at $x_0 = 32(1, 1)/\sqrt{2}$, where $\sqrt{\mathcal{L}_0} = 44.7$ and $q = 1/24$, $R^2 = 2$, $\mathcal{E}_\infty = 0.304$ for $s = 0.01$ and $\mathcal{E}_\infty = 1.567$ for $s = 0.05$. Both runs enter the near region at $T = 62$ on every seed but a few, then contract by a factor of about 0.6 per iteration in $\mathcal{L}_k$, against the certified factor $1 - q = 0.96$, and reach a floor within about 30 further iterations. Panel (b) varies the initial radius over $\sqrt{\mathcal{L}_0} \in \{5.3, 10.9, 22.1, 44.7, 90.0\}$. The measured mean entry time is 8.0, 17.0, 32.0, 62.0 and 121.0, and the largest entry time over seeds exceeds the mean by at most one. Both the measurement and the bound $2R\sqrt{\mathcal{L}_0}/\{q(R^2 - \mathcal{E}_\infty)\}$ are linear in $\sqrt{\mathcal{L}_0}$, and the bound is larger by a factor between 100 and 117 across the grid. The factor has two sources. The far-phase decrement $\gamma \le qR/2$ uses the worst gap over the self-concordant class, which grows like $\mu_{\mathrm{L}}\rho/(2M_{\mathrm{f}}\sqrt{\beta})$ at distance ρ by Lemma A.2, whereas the gap of this model

grows about four times faster, and the rate $q = \mu_{\mathrm{L}}/(12\beta)$ is the worst case of Lemma A.6. Both are uniform over the class, so the same factor applies at every $\sqrt{\mathcal{L}_0}$.

Residual level. Panel (c) starts at $x_\star$ and reports the stationary mean squared error for $N \in \{1, 4, 16, 64, 256\}$. The measured floor is between $0.52\, s^2/N$ and $0.57\, s^2/N$ for both noise levels, with fitted slope -0.99 in $\log N$. The gradient-noise term of $\mathcal{E}_\infty$ is $(34m+2)s^2/(3N) = 23.3\, s^2/N$, above the floor by a factor of 41 to 45 at every N. The certified residual itself has fitted slope -0.50 and lies above the floor by a factor of 1.2×10^3 at $s = 0.05$ and 5.8×10^3 at $s = 0.01$ for $N = 1$, growing with N, because its loss-contrast term $32\nu_\Delta/(3\mu_{\mathrm{L}}\sqrt{N})$ dominates. That term is inactive on this model, whose loss contrasts $sZ_i \sin x_i$ vanish at $x_\star$, so no selection error survives near the optimum. It cannot be dropped in general. On the instance of Proposition A.2 with $\psi = \phi$, $a = 1/64$ and $r = 1/8$, the run enters at $T = 0$ because $\mathcal{L}_0 \le 2V_0 = 0.031 < R^2$, and every iterate after the first equals $(0.10427, 0)$ with $V_k = 0.01087$ and $M(x_k) = 0.00543$, matching $r_1 = r - \phi'(r)/6$ to machine precision, while $\mathcal{E}_\infty = 0.471$. The quadratic instance with $a = r = 1$ stalls at $(5/6, 0)$ with $M(x_k) = 25/72$. In both cases the selected weight has $w_3 = 0$ at every step, as the proof asserts.

Same-batch bias. Panel (d) and Table D.8 run the oracle of Proposition B.1 for 1000 iterations from $x_0 = 0$ with $\eta \in \{0.5, 0.1, 0.02\}$, solving the same-batch QP of Eq. (B.1) at each step rather than using the closed-form map. The same-batch iterates are identical across all 50,000 sign sequences and converge to $\bar{x} = -0.4501836$ for every η, so the bias is independent of the step, as the proposition states. The mean of the delayed iterate is within two standard errors of zero, and its mean squared error is below the envelope $\eta\{1-(1-\eta)^{2k}\}/\{4(2-\eta)\}$ by 7.6%, 1.0% and 0.1%. The envelope is nearly attained because the selected weights sit at simplex vertices and $\cos x_k \approx 1$.

Check (instance)	Prediction	Measurement
Entry time $\mathbb{E}T$ (App. A.4, $s = 0.05$, $\sqrt{\mathcal{L}_0} = 44.7$)	≤ 7012	62.0 (largest 63)
Entry-time scaling (App. A.4, $\sqrt{\mathcal{L}_0}$ from 5.3 to 90)	linear in $\sqrt{\mathcal{L}_0}$	$1.3\sqrt{\mathcal{L}_0}$ to $1.5\sqrt{\mathcal{L}_0}$
Residual (App. A.4, $s = 0.05$, $N = 1$)	$\le \mathcal{E}_\infty = 1.57$	1.30×10^{-3}
Residual scaling (App. A.4, N from 1 to 256)	$\mathcal{E}_\infty \propto N^{-1/2}$, gradient term $\propto N^{-1}$	floor $\propto N^{-0.99}$
Stall point (Prop. A.2, $\psi = \phi$, $a = 1/64$)	$x_k = (0.10427, 0)$ for $k \ge 1$	exact, $w_3 = 0$
Same-batch limit (Prop. B.1, $\eta = 0.5, 0.1, 0.02$)	$\bar{x} = -0.450184$	-0.450184 for all η
Delayed mean $\mathbb{E}x_{1000}$ (Prop. B.1)	0	-0.0016 ± 0.0025, -0.0002 ± 0.0010, 0.0000 ± 0.0005
Delayed $\mathbb{E}x_{1000}^2$ (Prop. B.1)	$\le 0.0833, 0.0132, 0.00253$	0.0770, 0.0130, 0.00252

Table D.8: **Predicted versus measured quantities on the analytic instances.** Means over 2000 seeds for the model of Appendix A.4 and over 50,000 sign sequences for the oracle of Proposition B.1. The stall and the same-batch limit are deterministic. Uncertainties are two standard errors, rounded.

What the checks establish. Every stated bound holds on every run, the two functional dependences that the theory predicts, entry time linear in $\sqrt{\mathcal{L}_0}$ and a gradient-noise floor of order N^{-1}, are observed, and the two obstructions that motivate delay and the loss-contrast allowance are reproduced exactly. The constants of Theorem 1 are conservative by two orders of magnitude for the entry time and by three or more for the residual on this smooth model, which is the price of a bound that must also cover Proposition A.2. The checks do not test the neural-network setting of Section 6, where the population assumptions fail.