ZoAQ: Adaptive Zeroth-Order Querying via Query-Reuse Coupling

Anonymous authors
Paper under double-blind review

Abstract

Zeroth-order optimization (ZOO) estimates updates from function evaluations, making perturbation queries a primary cost. Fixed budgets spend the same number of queries at every step, while adaptive controllers may offset their savings by using additional oracle calls to test estimator reliability. We introduce ZoAQ[1], an adaptive ZOO method built around query reuse. Rather than discarding past evaluations after each step, ZoAQ makes them useful for both the next update and the decision to query further. This enables adaptive query allocation without extra validation queries. Our analysis characterizes when this agreement identifies an update that supports descent and guides the controller to a sufficient query budget. On synthetic objectives, ZoAQ reduces queries by 43–48% relative to fixed baselines using 1.2M queries. In black-box attacks, it reaches 100% success with 320 and 625 average queries on MNIST and CIFAR-10, respectively. Across four OPT fine-tuning settings, ZoAQ saves 43–46% forward evaluations relative to fixed $K = 4$, with accuracy changes within tasks ranging from -0.018 to $+0.010$.

1 Introduction

Zeroth-order optimization (ZOO) enables learning when gradients are unavailable or too costly to materialize, including black-box adversarial attacks and fine-tuning of large language models (LLMs) using forward passes only (Chen et al., 2017; 2019; Malladi et al., 2023). Each update is estimated from function evaluations along random perturbations, so the local direction count determines both the reliability of the estimate and its cost (Spall, 1992; Nesterov & Spokoiny, 2017). A small budget may produce a noisy direction, whereas a large fixed budget wastes evaluations at iterates where fewer directions suffice. Choosing one budget for the entire trajectory therefore leaves the variation in local requirements unused.

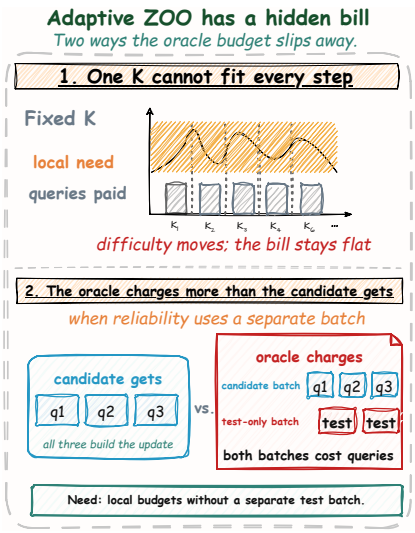


Figure 1: Two hidden query costs in adaptive ZOO.

Figure 1 summarizes the two difficulties in allocating queries that motivate ZoAQ. Adaptive sampling can exploit this variation by adding directions only when the current estimate is unreliable. Existing rules often test estimator variance, norm, or accuracy (Byrd et al., 2012; Bollapragada & Wild, 2023; Bollapragada et al., 2024). To run such a test, a controller may query a separate pilot or validation batch. Those evaluations consume the same oracle budget as the perturbations used to construct an update, but their responses do not improve the candidate estimate if they are used only for validation (Ghadimi & Lan, 2013). When the candidate itself contains only a few directions, even a small validation batch can offset the savings from adaptation (Lin et al., 2025). ZOO estimators based on variance reduction or query reuse can improve estimate quality under a prescribed sampling plan (Liu et al., 2018; Qiu et al.,

[1]An anonymized implementation is available at `https://anonymous.4open.science/r/ZoAQ-312A`.

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.

2025), but do not by themselves determine when the current iterate has received enough directions. A useful adaptive rule should instead make its decision from evidence that also enters the candidate update.

Momentum consistency is a natural candidate because it compares the current estimate with an exponential moving average (EMA) of past estimates. Momentum-based ZOO backbones already maintain this EMA (Chen et al., 2019; Shu et al., 2025), so computing the cosine requires no additional function evaluations. When each step samples perturbation directions independently, however, this cosine is difficult to interpret. A low score may reflect either a change in the underlying gradient or simply a mismatch between the two direction samples.

ZoAQ makes this comparison useful by retaining past query information across consecutive steps. Each record stores its observed scalar response together with a reproducible perturbation seed, so the corresponding direction can be reconstructed without another oracle call. Retained records and newly acquired records form the same candidate estimate, whose agreement with the EMA determines whether ZoAQ accepts the update or adds more queries. After acceptance, the next step returns to the minimum budget. A difficult iterate can therefore spend more queries without making the larger budget the default at the next one. Because retained responses and every new evaluation contribute directly to the candidate update, the controller adapts the local query budget without a separate validation batch.

Our analysis explains when reuse is useful: additional historical records reduce estimation noise, but stale responses also introduce bias. We further show that the EMA error remains controllable even when consecutive estimates share records, so the consistency test can identify updates that support descent and stop expansion by a sufficient local budget.

Our contributions are summarized as follows:

- We propose ZoAQ, which reuses past query information to construct the candidate update and decide whether more queries are needed, enabling adaptive local budgets without a separate validation batch.
- We characterize the trade-off between concentration and staleness induced by stored responses, derive pathwise EMA tracking under overlapping histories, and show that first passage yields a directionally safe update no later than a sufficient local budget.
- On synthetic objectives, ZoAQ reduces queries by 43–48% relative to fixed baselines using 1.2M queries. In black-box attacks, it reaches 100% success with 320 and 625 average queries on MNIST and CIFAR-10, respectively. Across four OPT fine-tuning settings, ZoAQ saves 43–46% forward evaluations relative to fixed $K = 4$, with accuracy changes within tasks ranging from -0.018 to $+0.010$.

2 Related Work and Positioning

Fixed-budget zeroth-order backbones. A first line of work establishes how to make useful updates when gradients are unavailable. Classical zeroth-order methods estimate gradients from randomized function evaluations (Spall, 1992; Nesterov & Spokoiny, 2017), and the broader derivative-free optimization literature develops principled search and trust-region machinery for function-value oracles (Conn et al., 2000; 2009). Modern ZOO methods bring this viewpoint to black-box learning and attacks through coordinate or random-direction estimators and adaptive-momentum updates (Chen et al., 2017; 2019), while R-AdaZO stabilizes the high-dimensional adaptive backbone used in our experiments (Shu et al., 2025). Forward-only LLM fine-tuning gives the same query-accounting issue at model scale, with MeZO showing that language models can be tuned through forward passes and LoRA reducing the number of trainable parameters (Malladi et al., 2023; Hu et al., 2022). These methods solve the backbone problem: they specify how function evaluations become update directions. Their local query budget, however, is usually chosen outside the online accept-or-expand decision. ZoAQ keeps this backbone view and studies the missing allocation layer: when a step has enough query evidence to move. Normalized momentum provides a first-order precedent for separating directional progress from update magnitude (Cutkosky & Mehta, 2020; 2021); our normalized-update protocol follows this analytical convention to isolate directional reliability from update magnitude.

History reuse and variance reduction. A second line of work improves zeroth-order estimates by reducing variance or reusing information. ZO-SVRG adapts variance-reduction ideas to nonconvex ZOO (Liu et al., 2018), and ZoAR shows that query reuse can produce more stable zeroth-order estimators (Qiu et al., 2025). These methods improve the estimator built from a collection of function evaluations, but they do not by themselves determine when sampling should stop at the current step. ZoAQ uses retained acquisition records to enlarge the candidate estimate and compares that candidate with an EMA of previously accepted estimates. Its distinct question is how the concentration gained from retained evidence can support local budget allocation without letting response staleness dominate.

Adaptive sampling. A third line of work chooses sample sizes online rather than fixing one budget for the whole trajectory. This idea is classical in first-order stochastic optimization (Byrd et al., 2012), and derivative-free or zeroth-order variants control stochastic approximation accuracy through norm-, variance-, or accuracy-based criteria (Bollapragada & Wild, 2023; Bollapragada et al., 2024). Such criteria are effective when their reliability statistics can be estimated affordably. In low-query ZOO, additional oracle calls used only to assess reliability may consume a material fraction of the local budget. ZoAQ instead computes its momentum-consistency score from the candidate update and an EMA already maintained by the optimizer. Its novelty lies in combining stored-response estimation with an accept-or-expand allocation rule and characterizing when the resulting score is directionally meaningful.

3 Problem Setup

3.1 Zeroth-Order Estimation and Local Budgets

The zeroth-order problem is to minimize $F(\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \mathbb{R}^d$ using only function-value queries. For a smoothing radius $\mu > 0$, define the Gaussian-smoothed objective

$$F_\mu(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\boldsymbol{u}\sim\mathcal{N}(0,\boldsymbol{I}_d)}[F(\boldsymbol{\theta} + \mu\boldsymbol{u})]. \tag{1}$$

For an exogenous direction $\boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I}_d)$, two function evaluations define the scalar response and its single-direction vector estimate:

$$r_\mu(\boldsymbol{\theta}, \boldsymbol{u}) \triangleq \frac{F(\boldsymbol{\theta} + \mu\boldsymbol{u}) - F(\boldsymbol{\theta} - \mu\boldsymbol{u})}{2\mu}, \qquad \boldsymbol{g}_\mu(\boldsymbol{\theta}, \boldsymbol{u}) \triangleq r_\mu(\boldsymbol{\theta}, \boldsymbol{u})\boldsymbol{u}. \tag{2}$$

For fixed $\boldsymbol{\theta}$, the Gaussian smoothing identity gives $\mathbb{E}_{\boldsymbol{u}}[\boldsymbol{g}_\mu(\boldsymbol{\theta}, \boldsymbol{u})] = \nabla F_\mu(\boldsymbol{\theta})$, so each direction yields an unbiased estimate of the smoothed gradient.

At step t, a fresh-query budget K averages K such estimates:

$$\boldsymbol{g}_t^{\mathrm{fresh}}(K) \triangleq \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{g}_\mu(\boldsymbol{\theta}_t, \boldsymbol{u}_{t,k}), \tag{3}$$

where each $\boldsymbol{u}_{t,k}$ has marginal distribution $\mathcal{N}(0, \boldsymbol{I}_d)$. The controller accepts K_t from $\mathcal{K} = \{K_{\min}, K_{\min} + \Delta K, \ldots, K_{\max}\}$, while a fixed-budget method uses the same K at every step. ZoAQ combines the K newly acquired records with retained records to form $\mathcal{H}_t(K)$ and the candidate estimate $\boldsymbol{g}_t(K)$, with $n_t(K) = |\mathcal{H}_t(K)|$. Thus, K counts fresh directions acquired at step t, while $n_t(K)$ counts the records used by the candidate estimate. Only the K fresh directions incur new oracle calls at step t. Section 4 gives the exact construction.

3.2 Directional Reliability

We assess a local budget by whether its estimated direction supports descent. To isolate this directional requirement, the analysis uses the following normalized-update protocol.

Assumption 3.1 (Normalized-Update Analysis Protocol)**.** For a nonzero candidate estimate, let $\boldsymbol{d}_t(K) = \boldsymbol{g}_t(K)/\|\boldsymbol{g}_t(K)\|$. The analyzed update is $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \alpha_t \boldsymbol{d}_t(K)$.

Under this protocol, a candidate budget satisfies the directional reliability condition when

$$\cos(\boldsymbol{d}_t(K), \nabla F_\mu(\boldsymbol{\theta}_t)) \geq \delta, \qquad \delta > 0, \tag{4}$$

Definition 3.2 (Oracle Alignment Budget)**.** For $\delta > 0$, when the following set is nonempty, define the oracle alignment budget

$$K_t^\circ(\delta) \triangleq \min\{K \in \mathcal{K} : \cos(\boldsymbol{d}_t(K), \nabla F_\mu(\boldsymbol{\theta}_t)) \geq \delta\}. \tag{5}$$

If no grid point satisfies the condition, we set $K_t^\circ(\delta) = K_{\max}$ to keep the budget on the finite grid.

The oracle budget depends on $\nabla F_\mu(\boldsymbol{\theta}_t)$ and cannot be evaluated directly. ZoAQ instead chooses K from a momentum-consistency score computed from the candidate estimate and recent accepted estimates. Every new record acquired during expansion enters the same candidate estimate.

4 ZoAQ: Query-Reuse Adaptive Querying

Figure 2 summarizes how ZoAQ builds one nested candidate, expands it when needed, and commits the accepted update and controller state.

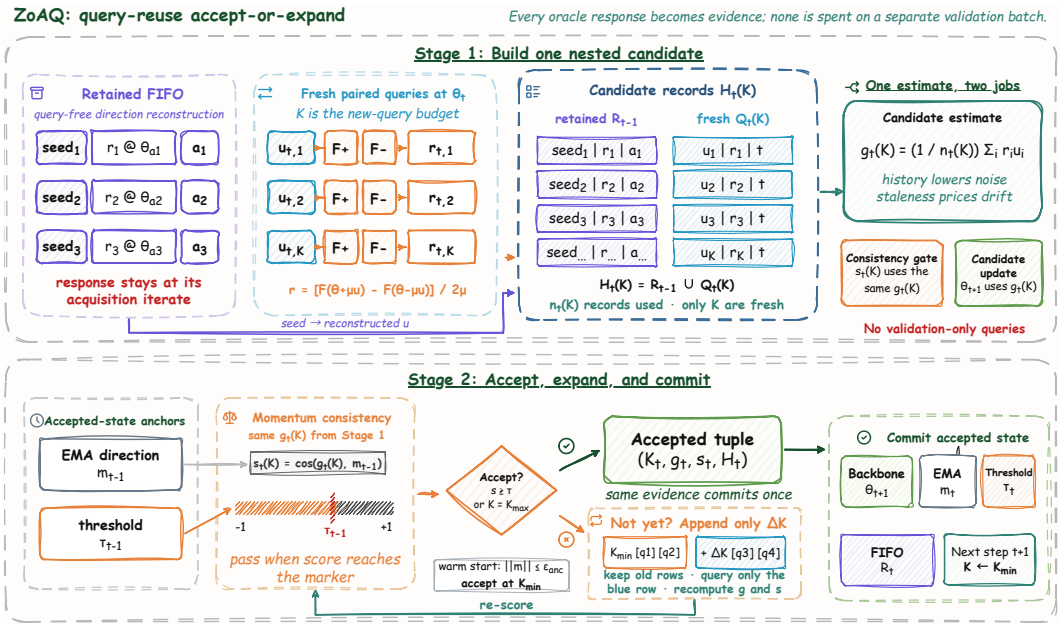


Figure 2: ZoAQ at one optimizer step. Retained records and K fresh directions form a nested candidate estimate. The same estimate drives both the backbone update and momentum-consistency gate; a failed test appends only ΔK fresh records, whereas acceptance commits the update, EMA, threshold, and FIFO history.

4.1 Query-Response Reuse

ZoAQ decides whether to expand the current query budget by comparing the candidate estimate with an exponential moving average (EMA) of previously accepted estimates. With fresh directions alone, a low consistency score can reflect either local gradient change or sampling noise. ZoAQ reduces this ambiguity by allowing retained query-response records to enter the candidate estimate.

Record i stores a reproducible seed for the direction $\boldsymbol{u}_i$, its acquisition step a_i, and the paired scalar response $r_i = r_\mu(\boldsymbol{\theta}_{a_i}, \boldsymbol{u}_i)$ defined in Section 3.1. The seed reconstructs $\boldsymbol{u}_i$ without a new query, while r_i remains the response observed at its acquisition iterate. Let $\mathcal{R}_{t-1}$ be the FIFO history and $\mathcal{Q}_t(K)$ the K fresh records acquired at $\boldsymbol{\theta}_t$. Their union $\mathcal{H}_t(K) = \mathcal{R}_{t-1} \cup \mathcal{Q}_t(K)$ defines the candidate estimate

$$\boldsymbol{g}_t(K) = \frac{1}{n_t(K)} \sum_{i=1}^{n_t(K)} r_i \boldsymbol{u}_i, \qquad n_t(K) = |\mathcal{H}_t(K)|. \tag{6}$$

The sets $\mathcal{Q}_t(K)$ are nested, so expanding from K to $K + \Delta K$ queries only the additional ΔK directions. After acceptance, FIFO retention keeps the most recent records for subsequent steps.

Query reuse and momentum average information at different stages. The candidate in Eq. 6 pools acquisition records before a budget is accepted, changing both the evidence available to the gate and the estimate sent to the backbone. The EMA instead pools previously accepted estimates after those decisions and supplies the reference at the next step. The first operation trades concentration against response staleness, whereas the second trades update noise against trajectory lag. They are complementary rather than duplicate smoothing operations; Appendix I gives the corresponding decomposition.

4.2 Momentum-Consistency Gate

For each candidate budget, ZoAQ computes

$$s_t(K) \triangleq \cos(\boldsymbol{g}_t(K), \boldsymbol{m}_{t-1}), \tag{7}$$

where $\boldsymbol{m}_{t-1}$ is the EMA of previously accepted estimates. Starting from $K_{\min}$, the controller accepts the first budget whose score reaches the current threshold:

$$K_t = \min(\{K \in \mathcal{K} : s_t(K) \geq \tau_{t-1}\} \cup \{K_{\max}\}). \tag{8}$$

If $\|\boldsymbol{m}_{t-1}\| \leq \varepsilon_{\text{anc}}$, ZoAQ takes a warm-start update at $K_{\min}$. If no score passes, it accepts the candidate at $K_{\max}$. Otherwise, every failed test adds fresh records to the same candidate estimate before the score is recomputed. The added queries therefore improve the candidate that can be accepted at the same step. After acceptance, ZoAQ passes $\boldsymbol{g}_t$ to the ZOO backbone, updates the estimate EMA and score threshold, and restarts the next step from $K_{\min}$.

This control flow separates evidence acquisition within a step from state adaptation across steps. Within a step, rejection changes only the fresh budget: it neither advances $\boldsymbol{\theta}_t$ nor updates the EMA, threshold, or FIFO buffer. Because the candidate sets are nested, each retest preserves all records already collected at that step and pays only for the next ΔK fresh directions. Acceptance is the commit point at which the chosen candidate updates the backbone and persistent controller state. Restarting from $K_{\min}$ then lets every new iterate test whether the cheapest budget is adequate, while the calibrated threshold carries information about the consistency of previously accepted estimates. Accordingly, K_t is a local stopping outcome rather than a persistent operating level: expansion at one iterate does not force the next iterate to inherit the larger budget.

In Algorithm 1, SCORE returns τ_{t-1} during warm start, returns $-\infty$ when the candidate estimate is zero and the EMA is nonzero, and otherwise returns the cosine score above. EXTENDHISTORY carries retained records forward and adds only the new records from the current step needed to reach K. For a finite accepted score, CALIBRATE applies $\tau_t = \beta_\tau \tau_{t-1} + (1 - \beta_\tau) s_t$; it leaves the threshold unchanged when $s_t = -\infty$.

The same candidate records determine both the update and the accept-or-expand score, so the controller requires no separate validation queries. Appendix G.1 gives the function evaluation accounting and update backbone for each protocol. Section 5 analyzes the stored-response controller and uses the normalized-update protocol in Assumption 3.1 for its descent result.

5 Theory: Estimation, Consistency, and Local Budgets

ZoAQ reuses scalar responses at their acquisition iterates. This increases the evidence available to a candidate without new function evaluations, but older records target earlier local gradients. We analyze this stored-response estimator and its accept-or-expand controller by quantifying the trade-off between concentration and staleness, propagating the accuracy of accepted candidates through the EMA, and characterizing safe first passage. The resulting estimation, tracking, and gate guarantees apply directly to the controller; the final descent statement uses the normalized-update protocol in Assumption 3.1. Appendices A–E give the full assumptions and proofs.

Algorithm 1 ZoAQ Accept-or-Expand with Query Reuse

Input: oracle F, initial point $\boldsymbol{\theta}_0$, horizon T
Parameters: grid $(K_{\min}, \Delta K, K_{\max})$, N_{hist}, and $(\tau_0, \beta_1, \beta_\tau, \varepsilon_{\text{anc}})$
Initialize: retained buffer $\mathcal{H}_0 \leftarrow \emptyset$ and EMA $\boldsymbol{m}_0 \leftarrow \mathbf{0}$
for $t = 1, \dots, T$ **do**
  $K \leftarrow K_{\min}$
  while true **do**
    $\mathcal{H}_t(K) \leftarrow \textsc{ExtendHistory}(F, \boldsymbol{\theta}_t, \mathcal{H}_{t-1}, K)$
    $\boldsymbol{g}_t(K) \leftarrow \textsc{PairedEstimate}(\mathcal{H}_t(K))$ Eq. 6
    $s_t(K) \leftarrow \textsc{Score}(\boldsymbol{g}_t(K), \boldsymbol{m}_{t-1}, \tau_{t-1}; \varepsilon_{\text{anc}})$
    if $s_t(K) \geq \tau_{t-1}$ or $K = K_{\max}$ **break; else** $K \leftarrow \min\{K + \Delta K, K_{\max}\}$
  end while
  $(K_t, \boldsymbol{g}_t, s_t, \mathcal{H}_t) \leftarrow (K, \boldsymbol{g}_t(K), s_t(K), \mathcal{H}_t(K))$
  $\boldsymbol{\theta}_{t+1} \leftarrow \textsc{BackboneStep}(\boldsymbol{\theta}_t, \boldsymbol{g}_t)$
  $\boldsymbol{m}_t \leftarrow \beta_1 \boldsymbol{m}_{t-1} + (1 - \beta_1)\boldsymbol{g}_t$
  $\tau_t \leftarrow \textsc{Calibrate}(\tau_{t-1}, s_t; \beta_\tau)$
  $\mathcal{H}_t \leftarrow \text{FIFO}(\mathcal{H}_t, N_{\text{hist}})$
end for
return $\boldsymbol{\theta}_T$ and $\{K_t\}_{t=1}^T$

5.1 Stored Responses Trade Concentration for Staleness

For a candidate built from $n_t(K)$ records, define its total acquisition staleness

$$A_t(K) \triangleq \sum_{j \in \mathcal{J}_t(K)} \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{a_j}\|. \tag{9}$$

This quantity measures staleness by optimization movement, so an old record need not be harmful when the iterate has moved little. The candidate differs from the current smoothed gradient $h_t = \nabla F_\mu(\boldsymbol{\theta}_t)$ through acquisition noise and gradient movement.

Lemma 5.1 (Selection-Safe Stored-Response Control)**.** *On the simultaneous candidate event established in Appendix B, every candidate inspected by the controller satisfies*

$$\|\boldsymbol{g}_t(K) - h_t\| \leq e_t(K) \triangleq q(n_t(K), \delta) + \frac{L A_t(K)}{n_t(K)}. \tag{10}$$

The statement remains valid for the random first-passing candidate.

Under a common local noise scale, the radius q decreases as records accumulate, whereas the second term grows with the distance traveled since acquisition. At a common fresh budget K, reuse improves the error certificate whenever

$$\frac{L A_t(K)}{n_t(K)} \leq q(K, \delta) - q(n_t(K), \delta). \tag{11}$$

Thus history is useful when its concentration gain exceeds its staleness cost. Under a full FIFO buffer, the leading envelope has the form $a_t/\sqrt{C} + b_t C$, yielding the interior capacity scale

$$C_{\text{opt}} \asymp \left(\frac{a_t K_{\min}}{L\bar{\ell}}\right)^{2/3}. \tag{12}$$

More history is therefore not uniformly better: noisier problems favor larger buffers, while faster movement favors smaller ones.

5.2 Momentum Consistency as a Directional Certificate

Because cosine is invariant to positive scaling, the implemented score is unchanged if the raw EMA is replaced by its debiased form $\widetilde{\boldsymbol{m}}_{t-1} = \boldsymbol{m}_{t-1}/(1 - \beta_1^{t-1})$. If accepted estimates satisfy $\|\boldsymbol{g}_r - h_r\| \leq \varepsilon_r$, Appendix C.1 proves the pathwise bound

$$\|\widetilde{\boldsymbol{m}}_{t-1} - h_t\| \leq \sum_{r<t} \bar{w}_{r,t}\varepsilon_r + L D_t^{\text{ema}}. \tag{13}$$

This avoids independence assumptions between overlapping accepted buffers. Let κ_t denote the right-hand side divided by $\|h_t\|$.

Theorem 5.2 (Soundness and Completeness of Momentum Consistency). *Suppose $\kappa_t < 1$ and a nonzero candidate passes $\cos(\boldsymbol{g}_t(K), \boldsymbol{m}_{t-1}) \geq \tau_{t-1}$. Then*

$$\cos(\boldsymbol{g}_t(K), h_t) \geq \Delta(\tau_{t-1}, \kappa_t) \triangleq \tau_{t-1}\sqrt{1-\kappa_t^2} - \kappa_t\sqrt{1-\tau_{t-1}^2}. \tag{14}$$

Conversely, a candidate is guaranteed to pass whenever

$$\|\boldsymbol{g}_t(K) - h_t\| \leq \eta_\star(\tau_{t-1}, \kappa_t)\|h_t\|, \tag{15}$$

where

$$\eta_\star(\tau, \kappa) \triangleq \sqrt{1-\tau^2}\sqrt{1-\kappa^2} - \tau\kappa. \tag{16}$$

Both directions are geometrically sharp. They coexist in the nonempty band

$$\kappa_t < \tau_{t-1} < \sqrt{1-\kappa_t^2}, \qquad \kappa_t < \frac{1}{\sqrt{2}}. \tag{17}$$

Within the band, every pass is descent-relevant and sufficiently accurate candidates can pass. Our experiments use Algorithm 1's unprojected EMA, so these guarantees apply when its threshold lies in the band; Appendix C.3 gives the projected calibration.

5.3 Safe First Passage and Query Control

Let $\bar{\kappa} < 1/\sqrt{2}$ uniformly bound κ_t, and choose $\bar{\kappa} < \underline{\tau} \leq \tau_{t-1} \leq \overline{\tau} < \sqrt{1-\bar{\kappa}^2}$. Define

$$\underline{\Delta} \triangleq \Delta(\underline{\tau}, \bar{\kappa}), \qquad \underline{\eta} \triangleq \eta_\star(\overline{\tau}, \bar{\kappa}), \tag{18}$$

and the first sufficient grid point

$$K_t^{\text{suf}} \triangleq \min\left\{K \in \mathcal{K} : e_t(K) \leq \underline{\eta}\|h_t\|\right\}. \tag{19}$$

Theorem 5.3 (Safe First-Passing Local Allocation). *If $K_t^{\text{suf}} \leq K_{\max}$, the increasing-grid search passes no later than K_t^{suf}. Its first passing candidate satisfies*

$$K_t \leq K_t^{\text{suf}}, \qquad \cos(\boldsymbol{g}_t(K_t), h_t) \geq \underline{\Delta} > 0. \tag{20}$$

Hence capped fallback does not occur on the simultaneous event. Under the normalized-update protocol in Assumption 3.1,

$$F_\mu(\boldsymbol{\theta}_{t+1}) - F_\mu(\boldsymbol{\theta}_t) \leq -\alpha_t\underline{\Delta}\|h_t\| + \frac{L\alpha_t^2}{2}. \tag{21}$$

Soundness and completeness play different roles in this stopping argument. Soundness evaluates whichever candidate happens to pass, including one below K_t^{suf}, and guarantees its alignment from the observed score rather than from its budget. Completeness supplies the stopping envelope: once the candidate error reaches the sufficient level in Eq. 19, the score must pass. Their combination therefore does not require the empirical score to increase with K, nor does it require every earlier candidate to fail. An irregular score sequence may pass early, but any such pass remains covered by the same soundness guarantee; otherwise, the search stops by K_t^{suf}.

Reuse enters this conclusion through the candidate error $e_t(K)$. When Eq. 11 holds at the sufficient budget based on fresh directions alone, retained records meet the same accuracy requirement with no larger fresh budget. The comparison therefore holds locally under the stated reuse condition. Over a window before stationarity, the cost of fresh directions is bounded by $\sum_t K_t^{\text{suf}}$ and hence by $K_{\max}$ times the window length. These guarantees cover steps satisfying the movement, tracking, threshold, and cap conditions, while the experiments measure realized allocation with the optimization backbone used in each task.

6 Experiments: Mechanism-to-Scale Evidence

The experiments measure the oracle cost incurred by ZoAQ together with the resulting task outcome. Synthetic objectives expose how its scores and direction budgets evolve, black-box attacks compare queries at a common success rate, and OPT fine-tuning reports forward evaluations alongside task metrics. Mechanism diagnostics help interpret these results but are not used to assign all end-to-end gains to the gate alone. Appendix G and Table 6 provide the full configurations, accounting conventions, baseline standardization, and evidence-to-claim map.

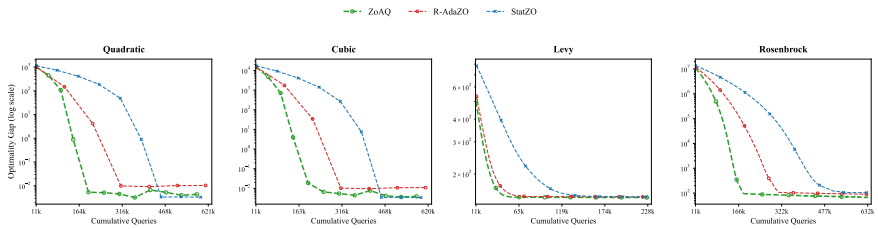


Figure 3: Synthetic convergence under the protocol of 60k steps, averaged over five runs. On the query axis, ZoAQ reaches regions of low error with fewer queries than baselines based on fixed budgets or variance.

Table 1: Black-box attack query efficiency (mean±std over five runs). Full protocol details are deferred to Appendix G.

Dataset	R-AdaZO ($K = 2$)				StatZO ($K_{\max} = 2$)				ZoAQ ($K_t \in [1, 2]$)			
	Success	Avg	Std	Speedup	Success	Avg	Std	Speedup	Success	Avg	Std	Speedup
MNIST	100%	1,527	701	1.00×	100%	1,996	1,231	0.77×	100%	320	45	4.77×
CIFAR-10	100%	1,006	100	1.00×	100%	964	28	1.04×	100%	625	15	1.61×

6.1 Synthetic Mechanism Tests

The synthetic regime makes local allocation directly visible. Figure 3 shows the main effect under the protocol of 60k steps: the fixed baselines spend 1.2M queries, whereas ZoAQ uses about 0.62M–0.68M on the displayed objectives, a 43–48% reduction. The fixed budget sweep in Table 15 controls for budget choice: changing the budget trades cost against final gap on individual tasks, but no single fixed value reproduces the adaptive balance between cost and quality across the suite. ZoAQ instead revisits $K_{\min}$ after every accepted update and acquires more evidence only when the current candidate fails the gate.

The auxiliary controls support the controller design. ZoAQ-Norm needs 28.1% more queries to reach the same Rosenbrock target and shows a weaker descent rate than ZoAQ (1.3 vs. 1.9 loss drop per 1k queries); Appendix H.4 and Figure 12 provide the corresponding diagnostic. Figure 13 checks the gate statistic directly: the displayed baselines stay near zero or decay toward it, whereas ZoAQ remains more positive through most of the trajectory. Appendix H reports the reset diagnostics, and Appendix I compares direction reuse with fresh independent sampling.

6.2 Black-Box Attacks under Query Constraints

Black-box attacks test query allocation in a regime where $d \gg K_t$, local budgets are often one or two directions, and every additional oracle call changes the realized cost. Under a common stabilized R-AdaZO backbone, Table 1 gives 100% success for all methods and separates them by required queries. ZoAQ reduces average queries to 320 on MNIST and 625 on CIFAR-10, corresponding to 4.77× and 1.61× speedups over fixed budget; StatZO does not recover comparable savings in this regime. Figure 14 provides a complementary diagnostic: ZoAQ maintains a more stable momentum-consistency score under the same attack regime. At the same 100% success rate, history reuse alone does not recover the full gain: relative to ZoAR with fixed $K = 2$, ZoAQ uses 56.9% fewer queries on MNIST and 17.3% fewer on CIFAR-10; Table 11 reports the full comparison.

Table 1 keeps the common stabilized R-AdaZO backbone explicit, with fixed $K = 2$ for R-AdaZO, best swept $K_{\max} = 2$ for StatZO, and $K_t \in [1, 2]$ for ZoAQ. Appendix G.10 justifies $N_{\text{hist}} = 8$, and Appendix F reports both the original and stabilized StatZO behavior.

6.3 LLM Fine-Tuning with Forward Passes Only

LLM fine-tuning tests the same accounting when each function evaluation is a forward pass through a large model. We fine-tune OPT-1.3B and OPT-13B on SST-2 and COPA for 5,000 steps with LoRA

Table 2: Main LLM fine-tuning summary over three seeds. FE denotes forward evaluations; full metrics and the $K_t \in [1, 2]$ endpoint are deferred to Table 13.

MODEL	TASK	METHOD	TOTAL FE	FE SAVINGS	BEST EVAL LOSS	ACCURACY
OPT-1.3B	SST-2	R-ADAZO	40000 ± 0	$0.00 \pm 0.00\%$	0.198 ± 0.002	0.920 ± 0.004
		STATZO	30746 ± 412	$23.14 \pm 1.03\%$	0.206 ± 0.005	0.910 ± 0.007
		ZOAQ	21714 ± 318	$\mathbf{45.72 \pm 0.80\%}$	0.195 ± 0.003	0.920 ± 0.005
	COPA	R-ADAZO	40000 ± 0	$0.00 \pm 0.00\%$	0.431 ± 0.004	0.780 ± 0.010
		STATZO	30498 ± 505	$23.76 \pm 1.26\%$	0.461 ± 0.007	0.790 ± 0.014
		ZOAQ	21946 ± 345	$\mathbf{45.14 \pm 0.86\%}$	0.485 ± 0.005	0.790 ± 0.012
OPT-13B	SST-2	R-ADAZO	40000 ± 0	$0.00 \pm 0.00\%$	0.185 ± 0.002	0.938 ± 0.003
		STATZO	31350 ± 390	$21.63 \pm 0.98\%$	0.202 ± 0.004	0.908 ± 0.006
		ZOAQ	22245 ± 290	$\mathbf{44.39 \pm 0.73\%}$	0.192 ± 0.003	0.920 ± 0.005
	COPA	R-ADAZO	40000 ± 0	$0.00 \pm 0.00\%$	0.405 ± 0.005	0.798 ± 0.012
		STATZO	31480 ± 480	$21.30 \pm 1.20\%$	0.452 ± 0.009	0.785 ± 0.016
		ZOAQ	22780 ± 360	$\mathbf{43.05 \pm 0.90\%}$	0.428 ± 0.006	0.792 ± 0.011

under a two-sided ZOO protocol that uses forward passes only, with three seeds. Table 2 summarizes the resulting trade-off between FE and quality.

ZoAQ with $K_t \in [1, 4]$ is the main operating point. Across the four OPT/task pairs in Table 2, it saves 43–46% FE relative to fixed $K = 4$. The allocation is repeated at each of the 5,000 steps rather than chosen once as a smaller fixed direction count. On OPT-1.3B SST-2, it matches R-AdaZO accuracy and slightly improves best evaluation loss; on COPA, it keeps accuracy close while accepting higher loss; on OPT-13B SST-2, it trails R-AdaZO with a fixed budget in both metrics. These results show FE savings relative to fixed $K = 4$, while the quality trade-off varies by task. Reporting both quantities makes the saved computation and resulting task quality directly comparable. Appendix G.6 gives FE accounting, hyperparameters, full metrics, and the aggressive $K_t \in [1, 2]$ endpoint; Appendix F, Sections G.9–G.10, and Appendix H provide the related stability and sensitivity checks.

Reuse and staleness across regimes. Sweeps over the history window expose the balance predicted by the analysis of stored responses. On MNIST attacks, average queries decrease from 374.4 at $N_{\text{hist}} = 4$ to 319.6 at $N_{\text{hist}} = 8$, then increase to 349.2 at $N_{\text{hist}} = 12$. On OPT-1.3B, enlarging the main window from 15 to 24 reduces FE by 3.00% on SST-2 and 3.67% on COPA, with changes in loss and accuracy that vary by task. The best intermediate window is consistent with the capacity trade-off in Eq. 12. Together, these results show that the balance between reuse and staleness depends on the regime; Tables 18 and 14 report the full sweeps.

The controlled objective switch probes the complementary case in which retained records suddenly become stale. When the objective changes from Rosenbrock to Levy, the effective threshold drops with the consistency mismatch, failed tests add fresh evidence, and ZoAQ reaches a lower final loss than R-AdaZO in this test. The controller therefore refreshes its evidence instead of remaining locked to the previous history. Together, the window and switch tests show that useful reuse depends on current movement rather than a fixed history length, consistent with the staleness quantity in Eq. 9. Appendix H.3 gives the protocol and trajectories.

7 CONCLUSION

ZoAQ uses acquired responses as evidence for both updates and allocation. Retained records join fresh directions in each candidate, and failed EMA consistency tests trigger additional queries. Our analysis identifies when the gain in concentration exceeds staleness and when the first passing candidate supports descent under a sufficient local budget. ZoAQ reduces queries by 43–48% on synthetic objectives, delivers 4.77× and 1.61× attack speedups at 100% success, and saves 43–46% FE in OPT fine-tuning with quality changes reported for each task. The history window and controlled objective switch show how the controller exploits useful records and refreshes stale ones, supporting query reuse as a practical basis for adapting local ZOO budgets.

REFERENCES

Raghu Bollapragada and Stefan M. Wild. Adaptive sampling quasi-Newton methods for zeroth-order stochastic optimization. *Mathematical Programming Computation*, 15(2):327–364, 2023. doi: 10.1007/s12532-023-00233-9. URL https://doi.org/10.1007/s12532-023-00233-9.

Raghu Bollapragada, Cem Karamanli, and Stefan M. Wild. Derivative-free optimization via adaptive sampling strategies, 2024. URL https://arxiv.org/abs/2404.11893.

Richard H. Byrd, Gillian M. Chin, Jorge Nocedal, and Yuchen Wu. Sample size selection in optimization methods for machine learning. *Math. Program.*, 134(1):127–155, 2012. doi: 10.1007/S10107-012-0572-5. URL https://doi.org/10.1007/s10107-012-0572-5.

Pin-Yu Chen, Huan Zhang, Yash Sharma, Jinfeng Yi, and Cho-Jui Hsieh. Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. In *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, AISec '17, pp. 15–26, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450352024. doi: 10.1145/3128572.3140448. URL https://doi.org/10.1145/3128572.3140448.

Xiangyi Chen, Sijia Liu, Kaidi Xu, Xingguo Li, Xue Lin, Mingyi Hong, and David Cox. ZO-AdaMM: zeroth-order adaptive momentum method for black-box optimization. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019. Curran Associates Inc.

Andrew R. Conn, Nicholas I. M. Gould, and Philippe L. Toint. *Trust Region Methods*. Society for Industrial and Applied Mathematics, 2000. doi: 10.1137/1.9780898719857. URL https://epubs.siam.org/doi/abs/10.1137/1.9780898719857.

Andrew R. Conn, Katya Scheinberg, and Luis N. Vicente. *Introduction to Derivative-Free Optimization*. Society for Industrial and Applied Mathematics, 2009. doi: 10.1137/1.9780898718768. URL https://epubs.siam.org/doi/abs/10.1137/1.9780898718768.

Ashok Cutkosky and Harsh Mehta. Momentum improves normalized SGD. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 2260–2268. PMLR, 2020.

Ashok Cutkosky and Harsh Mehta. High-probability bounds for non-convex stochastic optimization with heavy tails. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013. doi: 10.1137/120880811. URL https://doi.org/10.1137/120880811.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL https://api.semanticscholar.org/CorpusID:6628106.

Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. URL https://api.semanticscholar.org/CorpusID:18268744.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Wei Lin, Qingyu Song, and Hong Xu. The multi-query paradox in zeroth-order optimization, 2025. URL https://arxiv.org/abs/2509.15552.

Sijia Liu, Bhavya Kailkhura, Pin-Yu Chen, Paishun Ting, Shiyu Chang, and Lisa Amini. Zeroth-order stochastic variance reduction for nonconvex optimization, 2018. URL https://arxiv.org/abs/1805.10367.

Sadhika Malladi, Tianyu Gao, Eshaan Nichani, Alex Damian, Jason D. Lee, Danqi Chen, and Sanjeev Arora. Fine-tuning language models with just forward passes. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=Vota6rFhBQ.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Found. Comput. Math.*, 17(2):527–566, April 2017. ISSN 1615-3375. doi: 10.1007/s10208-015-9296-2. URL https://doi.org/10.1007/s10208-015-9296-2.

Junbin Qiu, Zhengpeng Xie, Xiangda Yan, Yongjie Yang, and Yao Shu. Zeroth-order optimization is secretly single-step policy optimization. In *Tiny Titans: The next wave of On-Device Learning for Foundational Models (TTODLer-FM)*, 2025. URL https://openreview.net/forum?id=HQwuvp7nc0.

Melissa Roemmele, Cosmin Adrian Bejan, and Andrew S. Gordon. Choice of plausible alternatives: An evaluation of commonsense causal reasoning. In *AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*, 2011. URL https://people.ict.usc.edu/~gordon/copa.html.

Yao Shu, Qixin Zhang, Kun He, and Zhongxiang Dai. Refining adaptive zeroth-order optimization at ease. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=NpIIbrg361.

J.C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992. doi: 10.1109/9.119632.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In Tal Linzen, Grzegorz Chrupała, and Afra Alishahi (eds.), *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pp. 353–355, Brussels, Belgium, November 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-5446. URL https://aclanthology.org/W18-5446/.

Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, Todor Mihaylov, Myle Ott, Sam Shleifer, Kurt Shuster, Daniel Simig, Punit Singh Koura, Anjali Sridhar, Tianlu Wang, and Luke Zettlemoyer. OPT: Open pre-trained transformer language models, 2022. URL https://arxiv.org/abs/2205.01068.

A Assumptions and Preliminaries

This appendix fixes the notation and probability interfaces used throughout the proofs. The analysis uses paired Gaussian two-sided responses stored at their acquisition iterates. Evaluation counts for each protocol are reported with the experiments rather than folded into the estimator definition.

A.1 Proof Notation

Objective and gradients. $F : \mathbb{R}^d \to \mathbb{R}$ is the black-box objective, $\boldsymbol{\theta}_t$ is the iterate, and

$$F_\mu(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\boldsymbol{u}\sim\mathcal{N}(0,\boldsymbol{I}_d)}[F(\boldsymbol{\theta}+\mu\boldsymbol{u})] \tag{22}$$

is its Gaussian smoothing at radius $\mu > 0$. We write $h_t \triangleq \nabla F_\mu(\boldsymbol{\theta}_t)$. The constants L_f and L are gradient-Lipschitz constants for F and F_μ, respectively.

Acquisition records. Directions are indexed in acquisition order. The jth record is sampled at step a_j after conditioning on the pre-sampling information $\mathcal{F}_{j-1}$. It contains $\boldsymbol{u}_j \mid \mathcal{F}_{j-1} \sim \mathcal{N}(0, \boldsymbol{I}_d)$ and the paired response $r_j \triangleq r_\mu(\boldsymbol{\theta}_{a_j}, \boldsymbol{u}_j)$. Its vector contribution, conditional mean, and centered acquisition noise are

$$X_j \triangleq r_j\boldsymbol{u}_j, \qquad h_j \triangleq \nabla F_\mu(\boldsymbol{\theta}_{a_j}), \qquad \xi_j \triangleq X_j - h_j. \tag{23}$$

Candidate estimates. At step t, K is the cumulative number of fresh directions acquired so far, $\mathcal{H}_t(K)$ is the candidate record set, and $\mathcal{J}_t(K)$ contains its acquisition indices. With $n_t(K) = |\mathcal{J}_t(K)|$,

$$\boldsymbol{g}_t(K) \triangleq \frac{1}{n_t(K)} \sum_{j\in\mathcal{J}_t(K)} X_j. \tag{24}$$

The controller searches $\mathcal{K} = \{K_{\min}, K_{\min}+\Delta K, \dots, K_{\max}\}$ and accepts K_t, after which $\boldsymbol{g}_t \triangleq \boldsymbol{g}_t(K_t)$.

Momentum and path. Accepted estimates form $\boldsymbol{m}_t = \beta_1\boldsymbol{m}_{t-1} + (1-\beta_1)\boldsymbol{g}_t$. Its debiased version is $\widetilde{\boldsymbol{m}}_t \triangleq \boldsymbol{m}_t/(1-\beta_1^t)$. The score is $s_t(K) \triangleq \cos(\boldsymbol{g}_t(K), \boldsymbol{m}_{t-1})$, and τ_{t-1} is the threshold used at step t. The realized path increment is $\ell_t \triangleq \|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t\|$.

A.2 Regularity and Smoothing

Assumption A.1 (Smooth Objective and Surrogate)**.** The original objective F is differentiable with L_f-Lipschitz gradient, and F_μ has an L-Lipschitz gradient on the region analyzed. Hence

$$F_\mu(\boldsymbol{\theta}') \le F_\mu(\boldsymbol{\theta}) + \langle\nabla F_\mu(\boldsymbol{\theta}), \boldsymbol{\theta}'-\boldsymbol{\theta}\rangle + \frac{L}{2}\|\boldsymbol{\theta}'-\boldsymbol{\theta}\|^2. \tag{25}$$

Lemma A.2 (Smoothing Bias)**.** *Under Assumption A.1,*

$$\|\nabla F_\mu(\boldsymbol{\theta}) - \nabla F(\boldsymbol{\theta})\| \le \mu L_f\sqrt{d}. \tag{26}$$

Thus an ϵ-stationary point of F_μ is an $(\epsilon + \mu L_f\sqrt{d})$-stationary point of F.

Proof. Differentiation under the Gaussian expectation gives $\nabla F_\mu(\boldsymbol{\theta}) = \mathbb{E}[\nabla F(\boldsymbol{\theta}+\mu\boldsymbol{u})]$. Gradient Lipschitzness and $\mathbb{E}\|\boldsymbol{u}\| \le \sqrt{\mathbb{E}\|\boldsymbol{u}\|^2} = \sqrt{d}$ prove Eq. 26. □

A.3 Paired Gaussian Contributions

For $\boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I}_d)$, define

$$r_\mu(\boldsymbol{\theta}, \boldsymbol{u}) \triangleq \frac{F(\boldsymbol{\theta}+\mu\boldsymbol{u}) - F(\boldsymbol{\theta}-\mu\boldsymbol{u})}{2\mu}, \qquad \boldsymbol{g}_\mu(\boldsymbol{\theta}, \boldsymbol{u}) \triangleq r_\mu(\boldsymbol{\theta}, \boldsymbol{u})\boldsymbol{u}. \tag{27}$$

An acquired record stores a reproducible seed for $\boldsymbol{u}$ and its scalar response. Reuse reconstructs the direction and reuses the scalar observed at acquisition without evaluating F again.

Lemma A.3 (Mean Paired Contribution)**.** *Whenever Gaussian differentiation is valid,*

$$\mathbb{E}_{\boldsymbol{u}}[\boldsymbol{g}_\mu(\boldsymbol{\theta}, \boldsymbol{u})] = \nabla F_\mu(\boldsymbol{\theta}). \tag{28}$$

Proof. Gaussian symmetry gives

$$\mathbb{E}_{\boldsymbol{u}}[\boldsymbol{g}_\mu(\boldsymbol{\theta}, \boldsymbol{u})] = \frac{1}{\mu}\mathbb{E}_{\boldsymbol{u}}[F(\boldsymbol{\theta} + \mu\boldsymbol{u})\boldsymbol{u}]. \tag{29}$$

Stein's identity identifies the right-hand side with $\nabla F_\mu(\boldsymbol{\theta})$. □

Lemma A.4 (Local Second-Moment Scale)**.** *Under Assumption A.1,*

$$\mathbb{E}_{\boldsymbol{u}}\|\boldsymbol{g}_\mu(\boldsymbol{\theta}, \boldsymbol{u})\|^2 \le 2(d+2)\|\nabla F(\boldsymbol{\theta})\|^2 + \frac{L_f^2\mu^2}{2}d(d+2)(d+4). \tag{30}$$

Consequently, for a universal constant $C > 0$,

$$\mathbb{E}_{\boldsymbol{u}}\|\boldsymbol{g}_\mu(\boldsymbol{\theta}, \boldsymbol{u}) - \nabla F_\mu(\boldsymbol{\theta})\|^2 \le C\left(d\|\nabla F_\mu(\boldsymbol{\theta})\|^2 + L_f^2\mu^2 d^3\right). \tag{31}$$

Proof. Let $g^\star = \nabla F(\boldsymbol{\theta})$. The fundamental theorem of calculus gives

$$r_\mu(\boldsymbol{\theta}, \boldsymbol{u}) = \langle g^\star, \boldsymbol{u}\rangle + \zeta(\boldsymbol{\theta}, \boldsymbol{u}), \qquad |\zeta(\boldsymbol{\theta}, \boldsymbol{u})| \le \frac{L_f\mu}{2}\|\boldsymbol{u}\|^2. \tag{32}$$

Therefore

$$\|\boldsymbol{g}_\mu(\boldsymbol{\theta}, \boldsymbol{u})\|^2 \le 2\langle g^\star, \boldsymbol{u}\rangle^2\|\boldsymbol{u}\|^2 + \frac{L_f^2\mu^2}{2}\|\boldsymbol{u}\|^6. \tag{33}$$

For a standard Gaussian vector,

$$\mathbb{E}[\langle g^\star, \boldsymbol{u}\rangle^2\|\boldsymbol{u}\|^2] = (d+2)\|g^\star\|^2, \qquad \mathbb{E}\|\boldsymbol{u}\|^6 = d(d+2)(d+4). \tag{34}$$

This proves Eq. 30. Lemma A.3, Eq. 26, and $(a+b)^2 \le 2a^2 + 2b^2$ then give Eq. 31 after collecting constants. □

A.4 Probability Interfaces

Assumption A.5 (Second Moments at Acquisition)**.** For every acquired record,

$$\mathbb{E}[\xi_j \mid \mathcal{F}_{j-1}] = 0, \qquad \mathbb{E}[\|\xi_j\|^2 \mid \mathcal{F}_{j-1}] \le \sigma_j^2, \tag{35}$$

where σ_j is $\mathcal{F}_{j-1}$-measurable. Lemma A.4 permits

$$\sigma_j^2 = C\left(d\|h_j\|^2 + L_f^2\mu^2 d^3\right) \tag{36}$$

for the paired Gaussian model.

Assumption A.6 (Conditional Bernstein Tails)**.** There are $\mathcal{F}_{j-1}$-measurable $\nu_j, b_j > 0$ such that, for every deterministic unit vector $\boldsymbol{v}$ and $|\lambda| < 1/b_j$,

$$\mathbb{E}[\exp(\lambda\langle \boldsymbol{v}, \xi_j\rangle)|\mathcal{F}_{j-1}] \le \exp\left(\frac{\lambda^2\nu_j^2}{2(1 - b_j|\lambda|)}\right). \tag{37}$$

The second-moment route is complete under Assumption A.5. Assumption A.6 is a separate stronger premise used only for concentration with logarithmic dependence on confidence; it is not inferred from a variance bound.

B Stored-Response Estimation

This section establishes the estimation guarantee that supports the rest of the theory. A stored-response candidate differs from the current smoothed gradient through two terms: acquisition noise, whose concentration can improve as records accumulate, and staleness, which increases as retained records move farther from the current iterate.

The analysis uses the objects at acquisition defined in Appendix A and FIFO pruning by pairs after acceptance. The main event is uniform over all candidates that the controller may inspect, so it remains valid after selection of the first passing candidate. We give a complete second-moment bound and a sharper version with logarithmic dependence on confidence under the additional conditional tail assumption.

B.1 A Uniform Error Bound for Stored-Response Candidates

The controller inspects a nested sequence of candidates and accepts the first one that passes. We therefore control every candidate that can be formed, rather than analyze a fixed budget and substitute the selected budget afterward.

Under FIFO pruning by pairs after acceptance, each candidate index set $\mathcal{J}_t(K)$ is an interval in acquisition order. Up to a horizon T, let $\mathfrak{I}_T$ contain every interval of acquisition indices that can occur for any $t \le T$ and $K \in \mathcal{K}$, and write $M_T \triangleq |\mathfrak{I}_T|$. Since at most $TK_{\max}$ directions are acquired, one may use the deterministic bound

$$M_T \le \frac{TK_{\max}(TK_{\max}+1)}{2}. \tag{38}$$

For a realized candidate, define its total staleness

$$A_t(K) \triangleq \sum_{j \in \mathcal{J}_t(K)} \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{a_j}\|. \tag{39}$$

Lemma B.1 (Uniform Candidate Error). *Suppose Assumptions A.1 and A.5 hold. For each admissible candidate, let $\overline{V}_t(K)$ be a deterministic upper bound on $\sum_{j \in \mathcal{J}_t(K)} \sigma_j^2$, and define*

$$q_2(t,K,\delta) \triangleq \frac{1}{n_t(K)} \sqrt{\frac{M_T \overline{V}_t(K)}{\delta}}. \tag{40}$$

Then, with probability at least $1-\delta$, every candidate inspected up to step T satisfies

$$\|\boldsymbol{g}_t(K) - h_t\| \le q_2(t,K,\delta) + \frac{LA_t(K)}{n_t(K)}. \tag{41}$$

If Assumption A.6 also holds, let $\overline{V}_t^{\nu}(K)$ and $\overline{b}_t(K)$ be deterministic bounds on $\sum_{j \in \mathcal{J}_t(K)} \nu_j^2$ and $\max_{j \in \mathcal{J}_t(K)} b_j$. There is a universal constant $c > 0$ such that Eq. 41 also holds with q_2 replaced by

$$q_{\mathrm{B}}(t,K,\delta) \triangleq \frac{c}{n_t(K)} \left[\sqrt{\overline{V}_t^{\nu}(K) \left(d + \log \frac{M_T}{\delta} \right)} + \overline{b}_t(K) \left(d + \log \frac{M_T}{\delta} \right) \right]. \tag{42}$$

Proof. For any candidate, the definitions in Appendix A give the exact decomposition

$$\boldsymbol{g}_t(K) - h_t = \frac{1}{n_t(K)} \sum_{j \in \mathcal{J}_t(K)} \xi_j + \frac{1}{n_t(K)} \sum_{j \in \mathcal{J}_t(K)} (h_j - h_t). \tag{43}$$

The second term is controlled pathwise. The L-smoothness of F_μ yields

$$\left\| \frac{1}{n_t(K)} \sum_{j \in \mathcal{J}_t(K)} (h_j - h_t) \right\| \le \frac{LA_t(K)}{n_t(K)}. \tag{44}$$

Fix an acquisition interval $I \in \mathfrak{I}_T$. The sequence $\{\xi_j\}$ is a martingale-difference sequence with respect to the acquisition filtration, so its cross terms vanish and

$$\mathbb{E} \left\| \sum_{j \in I} \xi_j \right\|^2 \le \mathbb{E} \sum_{j \in I} \sigma_j^2 \le \overline{V}_t(K), \tag{45}$$

where (t,K) denotes the candidate associated with I. Markov's inequality with failure probability δ/M_T, followed by a union bound over $\mathfrak{I}_T$, gives Eq. 40 simultaneously for every admissible interval.

Under Assumption A.6, scalar martingale Bernstein applied to $\langle \boldsymbol{v}, \sum_{j \in I} \xi_j \rangle$ gives the usual variance-plus-range bound for each fixed unit vector $\boldsymbol{v}$. A $1/2$-net of the unit sphere has at most 5^d elements. Taking a union bound over this net and over $\mathfrak{I}_T$ yields Eq. 42. Combining either noise radius with Eq. 44 proves the result. Because the event holds for every admissible interval, it also holds for the adaptively selected first-passing candidate. □

B.2 When Stored Responses Reduce the Required Fresh Budget

History is useful only when the reduction in sampling error exceeds the staleness it introduces. The next result states this comparison at a common fresh budget.

Proposition B.2 (Reuse-Benefit Condition). *Fix a step t and fresh budget K. Suppose the same nonincreasing noise radius $q(n, \delta)$ bounds an average of n admissible records. The candidate using only fresh directions contains K records from the current step and has error certificate $q(K, \delta)$. The candidate with stored responses uses $n_t(K) \geq K$ records and has the certificate from Lemma B.1. If*

$$\frac{LA_t(K)}{n_t(K)} \leq q(K, \delta) - q(n_t(K), \delta), \tag{46}$$

then the candidate with stored responses has an error certificate no larger than the candidate using only fresh directions at the same fresh query cost.

Proof. The candidate using only fresh directions has no acquisition staleness, so its certificate is $q(K, \delta)$. Lemma B.1 bounds the candidate with stored responses by

$$q(n_t(K), \delta) + \frac{LA_t(K)}{n_t(K)}. \tag{47}$$

Applying Eq. 46 proves the claim. □

FIFO pruning turns the staleness term into an explicit capacity constraint.

Lemma B.3 (FIFO Staleness Envelope). *Let the retained buffer contain at most C direction records, with pruning applied after each accepted step. Suppose every accepted step acquires at least $K_{\min}$ fresh directions and $\ell_r \leq \bar{\ell}$ over the steps that can contribute to the current buffer. Once the buffer is full, every retained record is at most*

$$H_C \triangleq \left\lceil \frac{C}{K_{\min}} \right\rceil \tag{48}$$

accepted steps old. Before pruning after acceptance, a candidate with fresh budget K therefore satisfies

$$n_t(K) = C + K, \qquad A_t(K) \leq CH_C\bar{\ell}. \tag{49}$$

Proof. During any H_C consecutive accepted steps, at least $H_C K_{\min} \geq C$ newer records arrive. FIFO pruning by pairs must therefore remove every older record. A retained record acquired r accepted steps earlier is at distance at most $\sum_s \ell_s \leq r\bar{\ell}$ from $\boldsymbol{\theta}_t$. Summing this bound over at most C retained records gives Eq. 49; the K fresh records contribute zero staleness. □

Corollary B.4 (Interior Capacity Scale). *Suppose the leading sampling radius for a full buffer has the homogeneous form $a_t(\delta)/\sqrt{C}$, and the average FIFO age is $c_{\text{age}}C/K_{\min}$ for a constant $c_{\text{age}} > 0$. The leading certificate is then*

$$e_t(C) = \frac{a_t(\delta)}{\sqrt{C}} + \frac{c_{\text{age}}L\bar{\ell}}{K_{\min}}C. \tag{50}$$

Its continuous minimizer is

$$C_t^\star = \left(\frac{a_t(\delta)K_{\min}}{2c_{\text{age}}L\bar{\ell}} \right)^{2/3}. \tag{51}$$

Thus the useful history size grows with local noise and refresh rate, but shrinks with smoothness-weighted movement. More retained records are not uniformly better.

Proof. Differentiate Eq. 50 with respect to C and set the derivative to zero:

$$-\frac{a_t(\delta)}{2C^{3/2}} + \frac{c_{\text{age}}L\bar{\ell}}{K_{\min}} = 0. \tag{52}$$

Solving for C gives Eq. 51. The implementable capacity is the nearest admissible integer. □

B.3 Outputs of the Stored-Response Analysis

The results above reduce the effect of query reuse to one candidate-specific error certificate. For either probability route, write

$$\varepsilon_t(K) \triangleq q(t, K, \delta) + \frac{LA_t(K)}{n_t(K)}, \tag{53}$$

where q is the applicable radius from Lemma B.1. On the simultaneous event, $\|\boldsymbol{g}_t(K) - h_t\| \le \varepsilon_t(K)$ for every inspected candidate, including the first one accepted by the controller.

Two consequences feed the remaining proofs. First, Eq. 46 identifies when retained responses improve the certificate at a fixed fresh budget. Second, Eq. 51 shows how FIFO capacity balances sampling error against staleness induced by movement. Appendix C.1 next propagates the certificates for accepted candidates through the EMA and derives its tracking accuracy.

Movement enters through staleness. For every retained record,

$$\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{a_j}\| \le \sum_{r=a_j}^{t-1} \ell_r. \tag{54}$$

Thus $A_t(K)$ already carries the complete path dependence needed by the stored-response estimator. Once a candidate is accepted, Appendix C.1 uses its realized certificate $\varepsilon_t(K_t)$ together with the EMA-weighted path length; no separate covariance-transport margin is required.

Probability interface. The analytical estimator directly averages the paired contributions $X_j = r_j \boldsymbol{u}_j$ defined in Eq. 23. Its statistical input is therefore contained in Lemma B.1: second moments at acquisition yield the radius q_2 with polynomial dependence on confidence, while conditional Bernstein tails yield the radius q_{B} with logarithmic dependence. All subsequent results condition on the corresponding simultaneous candidate event.

C Momentum Tracking and Directional Control

This section converts the candidate certificates from Appendix B into a certificate for the momentum-consistency gate. The argument is pathwise after the simultaneous candidate event is fixed, so overlapping FIFO buffers and adaptive first-passing selection introduce no additional independence requirement.

C.1 Debiased EMA Tracking

Assume $\boldsymbol{m}_0 = 0$ and $\boldsymbol{m}_r = \beta_1 \boldsymbol{m}_{r-1} + (1-\beta_1)\boldsymbol{g}_r$. For $t \ge 2$, define

$$z_{t-1} \triangleq 1 - \beta_1^{t-1}, \qquad \widetilde{\boldsymbol{m}}_{t-1} \triangleq \frac{\boldsymbol{m}_{t-1}}{z_{t-1}}, \qquad \bar{w}_{r,t} \triangleq \frac{(1-\beta_1)\beta_1^{t-1-r}}{z_{t-1}}. \tag{55}$$

The weights are nonnegative and sum to one. Moreover, $z_{t-1} > 0$, so replacing $\boldsymbol{m}_{t-1}$ by $\widetilde{\boldsymbol{m}}_{t-1}$ leaves every cosine score unchanged.

Proposition C.1 (Pathwise Debiased-EMA Tracking). *Suppose the final estimate accepted at step r satisfies*

$$\|\boldsymbol{g}_r - h_r\| \le \varepsilon_r. \tag{56}$$

Then

$$\|\widetilde{\boldsymbol{m}}_{t-1} - h_t\| \le \sum_{r<t} \bar{w}_{r,t}\varepsilon_r L D_t^{\mathrm{ema}}, \qquad D_t^{\mathrm{ema}} \triangleq \sum_{r<t} \bar{w}_{r,t} \sum_{s=r}^{t-1} \ell_s. \tag{57}$$

In particular, if $\varepsilon_r \le \bar{\varepsilon}$ and $\ell_s \le \bar{\ell}$ over the contributing steps, then

$$\|\widetilde{\boldsymbol{m}}_{t-1} - h_t\| \le \bar{\varepsilon} + L\bar{\ell} A_{t,\beta_1} \le \bar{\varepsilon} + \frac{L\bar{\ell}}{1-\beta_1}, \tag{58}$$

where

$$A_{t,\beta_1} = \frac{1 - t\beta_1^{t-1} + (t-1)\beta_1^t}{(1-\beta_1)(1-\beta_1^{t-1})}. \tag{59}$$

Proof. Equation 55 gives

$$\widetilde{\boldsymbol{m}}_{t-1} - h_t = \sum_{r<t} \bar{w}_{r,t}(\boldsymbol{g}_r - h_r) + \sum_{r<t} \bar{w}_{r,t}(h_r - h_t). \tag{60}$$

The first sum is bounded by $\sum_{r<t} \bar{w}_{r,t}\varepsilon_r$. Smoothness gives $\|h_r - h_t\| \le L\sum_{s=r}^{t-1} \ell_s$, which yields Eq. 57. Under uniform bounds, the first weighted sum is at most $\bar{\varepsilon}$, while the second is $L\bar{\ell}\sum_{r<t} \bar{w}_{r,t}(t-r)$. Evaluating the finite geometric sum gives A_{t,β_1}, and $A_{t,\beta_1} \le (1-\beta_1)^{-1}$. □

The proposition uses errors of the final estimates, not an independence assumption between them. On the simultaneous event of Lemma B.1, one may take

$$\varepsilon_r = q(n_r(K_r), \delta) + \frac{LA_r(K_r)}{n_r(K_r)}. \tag{61}$$

Thus acquisition noise and response staleness enter the EMA with their realized geometric weights.

Corollary C.2 (Relative Tracking Level)**.** *At any pre-stationary step with $\|h_t\| > 0$, define*

$$\kappa_t \triangleq \frac{\sum_{r<t} \bar{w}_{r,t}\varepsilon_r + LD_t^{\text{ema}}}{\|h_t\|}. \tag{62}$$

Then $\|\widetilde{\boldsymbol{m}}_{t-1} - h_t\| \le \kappa_t\|h_t\|$. If $\|h_t\| \ge \epsilon$, $\varepsilon_r \le \bar{\varepsilon}$, and $\ell_s \le \bar{\ell}$, the explicit sufficient bound is

$$\kappa_t \le \frac{\bar{\varepsilon} + L\bar{\ell}/(1-\beta_1)}{\epsilon}. \tag{63}$$

C.2 Sharp Gate Geometry

The next result characterizes both directions of the consistency test. Soundness states what a passing score guarantees; completeness states how accurate a candidate must be in order to pass.

Theorem C.3 (Soundness and Completeness of Momentum Consistency)**.** *Let $h \neq 0$, let an anchor $m \neq 0$ satisfy $\|m - h\| \le \kappa\|h\|$ for $0 \le \kappa < 1$, and let $g \neq 0$.*

If $\cos(g, m) \ge \tau$, then

$$\cos(g, h) \ge \Delta(\tau, \kappa) \triangleq \tau\sqrt{1-\kappa^2} - \kappa\sqrt{1-\tau^2}. \tag{64}$$

In particular, every passing candidate is positively aligned with h whenever $\tau > \kappa$.

Conversely, if $\|g - h\| \le \eta\|h\|$ for $0 \le \eta < 1$, then

$$\cos(g, m) \ge \sqrt{1-\eta^2}\sqrt{1-\kappa^2} - \eta\kappa. \tag{65}$$

Therefore the gate must pass whenever

$$\eta \le \eta_\star(\tau, \kappa) \triangleq \sqrt{1-\tau^2}\sqrt{1-\kappa^2} - \tau\kappa. \tag{66}$$

Proof. The relative anchor error implies $\angle(m, h) \le \arcsin\kappa$: among all m in the radius-$\kappa\|h\|$ ball centered at h, the largest angle is attained by a tangent ray. A passing score gives $\angle(g, m) \le \arccos\tau$. The spherical triangle inequality therefore yields

$$\angle(g, h) \le \arccos\tau + \arcsin\kappa. \tag{67}$$

Taking the cosine proves Eq. 64.

Similarly, $\|g - h\| \le \eta\|h\|$ implies $\angle(g, h) \le \arcsin\eta$. Hence

$$\angle(g, m) \le \arcsin\eta + \arcsin\kappa, \tag{68}$$

whose cosine gives Eq. 65. Solving the resulting inequality for η yields Eq. 66. □

Soundness and nontrivial completeness coexist exactly in the band

$$\kappa < \tau < \sqrt{1-\kappa^2}, \tag{69}$$

which is nonempty if and only if $\kappa < 1/\sqrt{2}$. Both constants are sharp: coplanar vectors placed at the tangent angles used in the proof attain equality.

C.3 Threshold Interface

The geometry requires the threshold to remain inside a feasible band; it does not require a particular calibration statistic. A fixed threshold suffices. An adaptive rule can preserve the same guarantee through projection,

$$\tau_t = \Pi_{[\underline{\tau},\overline{\tau}]}\left(\beta_\tau \tau_{t-1} + (1-\beta_\tau) s_t^{\mathrm{ref}}\right), \tag{70}$$

where, for a uniform tracking level $\bar{\kappa} < 1/\sqrt{2}$,

$$\bar{\kappa} < \underline{\tau} \le \overline{\tau} < \sqrt{1-\bar{\kappa}^2}. \tag{71}$$

Projection separates the guarantee from the selection bias of the score written after first passage. Without a fixed upper bound, updating the threshold only from passing scores makes it nondecreasing and can eventually destroy completeness.

D First-Passing Budget Control

The controller inspects a finite increasing grid $\mathcal{K}$ and stops at the first candidate whose momentum-consistency score reaches the current threshold. The proof below controls this stopping rule directly through the candidate error envelope and the gate geometry. It does not introduce an expected score curve or require earlier candidates to be separated from an oracle threshold.

D.1 Sufficient Local Budget

Fix a step t and work on the simultaneous candidate event from Lemma B.1. Write

$$e_t(K) \triangleq q(n_t(K),\delta) + \frac{LA_t(K)}{n_t(K)}, \qquad K \in \mathcal{K}, \tag{72}$$

so that $\|\boldsymbol{g}_t(K) - h_t\| \le e_t(K)$ for every inspected candidate. The first-passing result below uses the first grid point that meets a sufficient error level; it does not require the empirical score or a candidate-specific concentration radius to be monotone in K.

Suppose the debiased anchor satisfies the uniform relative tracking bound

$$\|\widetilde{\boldsymbol{m}}_{t-1} - h_t\| \le \bar{\kappa}\|h_t\|, \qquad 0 \le \bar{\kappa} < \frac{1}{\sqrt{2}}, \tag{73}$$

and the threshold lies in

$$\bar{\kappa} < \underline{\tau} \le \tau_{t-1} \le \overline{\tau} < \sqrt{1-\bar{\kappa}^2}. \tag{74}$$

Define the uniform geometry margins

$$\underline{\Delta} \triangleq \Delta(\underline{\tau},\bar{\kappa}) > 0, \qquad \underline{\eta} \triangleq \eta_\star(\overline{\tau},\bar{\kappa}) > 0. \tag{75}$$

Definition D.1 (Sufficient Grid Budget)**.** For $\|h_t\| > 0$, let

$$K_t^{\mathrm{suf}} \triangleq \min\left\{K \in \mathcal{K} : e_t(K) \le \underline{\eta}\|h_t\|\right\}, \tag{76}$$

whenever the set is nonempty.

Lemma D.2 (A Sufficient Budget Exists Below the Cap)**.** *If*

$$e_t(K_{\max}) \le \underline{\eta}\|h_t\|, \tag{77}$$

then K_t^{suf} exists and $K_t^{\mathrm{suf}} \le K_{\max}$. On a pre-stationary region $\|h_t\| \ge \epsilon$, it is sufficient to replace the right-hand side of Eq. 77 by $\underline{\eta}\epsilon$.

Proof. Equation 77 places $K_{\max}$ in the defining set of Eq. 76; the minimum therefore exists on the finite grid. The pre-stationary statement follows from $\|h_t\| \ge \epsilon$. □

D.2 Safe First Passage

Proposition D.3 (First-Passing Control). *Assume Eqs. 73–74, and suppose K_t^{suf} exists. Let K_t be the first grid point accepted by the consistency test. Then*

$$K_t \le K_t^{\mathrm{suf}}, \qquad \cos(\boldsymbol{g}_t(K_t), h_t) \ge \underline{\Delta} > 0. \tag{78}$$

Consequently, if Eq. 77 holds, the search passes before capped fallback. If a continuous sufficient level is rounded upward to the next grid point, the corresponding implementation bound incurs at most one increment ΔK.

Proof. At K_t^{suf}, the simultaneous candidate event and Definition D.1 give

$$\|\boldsymbol{g}_t(K_t^{\mathrm{suf}}) - h_t\| \le \underline{\eta}\|h_t\|. \tag{79}$$

Because $\tau_{t-1} \le \overline{\tau}$ and the actual tracking level is no larger than $\bar{\kappa}$, the completeness part of Theorem C.3 forces this candidate to pass. The increasing search therefore stops no later, so $K_t \le K_t^{\mathrm{suf}}$.

For the selected candidate, passage gives $\cos(\boldsymbol{g}_t(K_t), \widetilde{\boldsymbol{m}}_{t-1}) \ge \tau_{t-1} \ge \underline{\tau}$. The soundness part of Theorem C.3 then yields

$$\cos(\boldsymbol{g}_t(K_t), h_t) \ge \Delta(\tau_{t-1}, \kappa_t) \ge \Delta(\underline{\tau}, \bar{\kappa}) = \underline{\Delta}. \tag{80}$$

Thus an earlier accidental pass can reduce the budget but cannot invalidate the directional guarantee. □

The proposition deliberately provides an upper bound rather than exact recovery of an oracle budget. Exact equality would require a lower separation condition for every earlier score, which is unnecessary for either safety or query savings.

D.3 Full Buffer and Query Accounting

Suppose the FIFO buffer after acceptance is full with capacity C, each accepted step adds at least $K_{\min}$ fresh records, and the relevant path increments satisfy $\ell_s \le \bar{\ell}$. When a common nonincreasing radius $q(n, \delta)$ controls averages of n admissible records, let

$$H_C \triangleq \left\lceil \frac{C}{K_{\min}} \right\rceil. \tag{81}$$

Before pruning after acceptance, a candidate with K fresh records has $n_t(K) = C + K$ and $A_t(K) \le C H_C \bar{\ell}$. Hence

$$e_C(K) \triangleq q(C + K, \delta) + \frac{L C H_C \bar{\ell}}{C + K} \tag{82}$$

is a deterministic sufficient envelope. The cap condition $e_C(K_{\max}) \le \underline{\eta}\epsilon$ makes the no-fallback statement operational using buffer, movement, confidence, and cap parameters alone.

Corollary D.4 (Controller Query Cost). *Let $\mathcal{G}$ contain the steps on which the simultaneous candidate event, the tracking bound, the threshold band, and cap sufficiency all hold. Then*

$$\sum_{t=1}^{T} K_t \le \sum_{t \in \mathcal{G}} K_t^{\mathrm{suf}} + K_{\max}(T - |\mathcal{G}|). \tag{83}$$

If the joint conditions hold throughout a window, the second term vanishes. Moreover, whenever the reuse condition in Eq. 46 holds at the sufficient budget for fresh directions alone, the sufficient budget with reuse is no larger.

Proof. Proposition D.3 gives $K_t \le K_t^{\mathrm{suf}}$ on $\mathcal{G}$. Every other step is bounded by the deterministic cap $K_{\max}$. Summing these two cases proves Eq. 83. The final comparison follows by evaluating the reuse envelope at the sufficient budget for fresh directions alone and applying Eq. 46. □

E Descent and Query Accounting

This section closes the theorem chain for the normalized-update model. The result is local to a pre-stationary window on which the estimation, movement, threshold, and cap conditions established above hold. It does not identify coordinatewise adaptive preconditioning with global Euclidean normalization.

E.1 Normalized Descent

Assumption E.1 (Normalized-Update Analysis Protocol)**.** For the descent analysis, the selected nonzero candidate defines

$$\boldsymbol{d}_t \triangleq \frac{\boldsymbol{g}_t}{\|\boldsymbol{g}_t\|}, \qquad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \alpha_t \boldsymbol{d}_t, \qquad \|\boldsymbol{d}_t\| = 1. \tag{84}$$

This is the formal optimizer scope of the descent theorem. The experiments use the stated R-AdaZO backbone; a coordinatewise second-moment preconditioner need not preserve Euclidean angles and is not covered without an additional spectral condition.

Lemma E.2 (Normalized Descent)**.** *Under Assumption A.1, if* $\cos(\boldsymbol{d}_t, h_t) \geq \delta_t > 0$*, then*

$$F_\mu(\boldsymbol{\theta}_{t+1}) - F_\mu(\boldsymbol{\theta}_t) \leq -\alpha_t \delta_t \|h_t\| + \frac{L\alpha_t^2}{2}. \tag{85}$$

Proof. The smoothness inequality and Eq. 84 give

$$F_\mu(\boldsymbol{\theta}_{t+1}) - F_\mu(\boldsymbol{\theta}_t) \leq -\alpha_t \langle h_t, \boldsymbol{d}_t \rangle + \frac{L\alpha_t^2}{2}\|\boldsymbol{d}_t\|^2 \tag{86}$$

$$\leq -\alpha_t \delta_t \|h_t\| + \frac{L\alpha_t^2}{2}. \tag{87}$$

□

E.2 Operational Pre-Stationary Window

The next theorem gathers the preceding interfaces into one statement. Its probability is exactly that of the simultaneous candidate event used to instantiate the error radii; after conditioning on that event, every step of the argument is deterministic.

Theorem E.3 (Certified Descent and Local Query Control)**.** *Consider a finite window beginning after warm-up in which* $\|h_t\| \geq \epsilon$ *and the candidate event of Lemma B.1 holds simultaneously. Suppose:*

1. *the final estimates entering the debiased EMA satisfy Eq. 61;*
2. *the resulting tracking levels obey* $\kappa_t \leq \bar{\kappa} < 1/\sqrt{2}$*;*
3. *the threshold is fixed or projected into the band in Eq. 74;*
4. *the cap satisfies* $e_t(K_{\max}) \leq \underline{\eta}\epsilon$*; and*
5. *the update follows Eq. 84 with* $0 < \alpha_t \leq \underline{\Delta}\epsilon/L$*.*

Then every search in the window passes by $K_t^{\mathrm{suf}} \leq K_{\max}$*, and its selected direction satisfies*

$$\cos(\boldsymbol{d}_t, h_t) \geq \underline{\Delta}. \tag{88}$$

Consequently,

$$F_\mu(\boldsymbol{\theta}_{t+1}) - F_\mu(\boldsymbol{\theta}_t) \leq -\frac{\alpha_t \underline{\Delta}\epsilon}{2}. \tag{89}$$

Proof. Proposition C.1 and the certificates for accepted candidates give the stated relative tracking level. The threshold band and cap condition then satisfy the premises of Proposition D.3, which rules out capped fallback, gives $K_t \leq K_t^{\mathrm{suf}}$, and proves Eq. 88. Lemma E.2 yields

$$F_\mu(\boldsymbol{\theta}_{t+1}) - F_\mu(\boldsymbol{\theta}_t) \leq -\alpha_t \underline{\Delta}\epsilon + \frac{L\alpha_t^2}{2}. \tag{90}$$

The step-size condition bounds the curvature term by $\alpha_t \underline{\Delta}\epsilon/2$, proving Eq. 89. □

E.3 Finite-Horizon Accounting

Corollary E.4 (Pre-Stationary Steps and Fresh Directions). *Suppose the conditions of Theorem E.3 hold until the first step T_ϵ with $\|h_{T_\epsilon}\| < \epsilon$, and suppose F_μ is bounded below by $F_\star$. If $\alpha_t = \alpha = c\epsilon$ with $0 < c \le \underline{\Delta}/L$, then*

$$T_\epsilon - t_1 \le \frac{2\big(F_\mu(\boldsymbol{\theta}_{t_1}) - F_\star\big)}{c\underline{\Delta}\epsilon^2}. \tag{91}$$

The cumulative number of fresh directions satisfies

$$\sum_{t=t_1}^{T_\epsilon-1} K_t \le \sum_{t=t_1}^{T_\epsilon-1} K_t^{\text{suf}} \le (T_\epsilon - t_1)K_{\max}. \tag{92}$$

Proof. Summing Eq. 89 from t_1 to $T_\epsilon - 1$ and using the lower bound $F_\star$ gives Eq. 91. Proposition D.3 gives $K_t \le K_t^{\text{suf}} \le K_{\max}$ at every step in the window, which proves Eq. 92. □

The sharper quantity in Eq. 92 is the sum of local sufficient budgets, not the cap product. For the full-buffer envelope in Eq. 82, each K_t^{suf} is obtained by inverting

$$q(C + K, \delta) + \frac{LCH_C\bar{\ell}}{C + K} \le \underline{\eta}\epsilon. \tag{93}$$

This expression exposes the two ways in which queries help: they add acquisition evidence and dilute the staleness carried by the retained records.

Corollary E.5 (Reuse Versus Fresh-Only Accounting). *Let $K_{t,\text{fresh}}^{\text{suf}}$ be the first grid point satisfying $q(K, \delta) \le \underline{\eta}\|h_t\|$. If the reuse-benefit condition in Eq. 46 holds at this budget, then*

$$K_t^{\text{suf}} \le K_{t,\text{fresh}}^{\text{suf}}. \tag{94}$$

The inequality is strict whenever an earlier reuse candidate already satisfies the sufficient-error condition.

Proof. At $K_{t,\text{fresh}}^{\text{suf}}$, Eq. 46 makes the reuse envelope no larger than the fresh-only envelope, which is at most $\underline{\eta}\|h_t\|$. This grid point is therefore feasible for Definition D.1; taking the minimum proves the result. □

The conclusion is a conditional local saving, not an unconditional dominance statement. When response staleness exceeds the concentration gain, the comparison reverses, which is precisely why Appendix B predicts an interior useful history capacity.

F Diagnostics for Variance-Based Adaptation

The main text compares strategies for query allocation under a common stabilized backbone. This section explains why adaptation based on variance remains costly in the tested regimes after separating instability in the optimization backbone from the sampling rule itself.

F.1 StatZO in Ultra-High-Dimensional Fine-Tuning

The original implementation of StatZO (Bollapragada et al., 2024) relies on standard zeroth-order optimizers (e.g., ZO-SGD or ZO-AdaMM). While theoretically sound for lower-dimensional problems, this configuration exhibits severe instability in ultra-high-dimensional non-convex landscapes ($d \approx 10^5$).

As illustrated in Figure 4 (referencing the ablation in Section 6.3), the original StatZO configuration fails to descend effectively on SST-2, stagnating at high loss values (> 3.5). In this regime, estimating the population variance σ^2 from a mini-batch in such high dimensions produces unstable SNR signals, which in turn destabilize the optimizer's second-moment estimates.

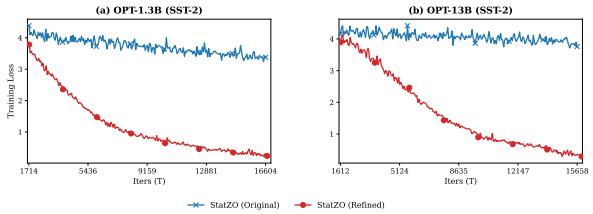


Figure 4: Behavior of the original StatZO configuration on SST-2. Without the stabilized update rule used in our main experiments, the original StatZO configuration stagnates at a high loss, motivating the standardized optimizer backbone used in the main comparison.

F.2 Stabilized StatZO in Black-Box Attacks

To evaluate the adaptive sampling strategy more cleanly, we examine StatZO in two stages: its original formulation and a stabilized version using the same backbone as the main comparisons. This two-step analysis helps separate backbone effects from sampling effects.

First, we evaluate the original StatZO equipped with the standard ZO-AdaMM backbone. As shown in Table 3, the configuration requiring the fewest queries among the tested caps is $K_{\max} = 4$, with 4,382 queries. However, the overall efficiency is hampered by the instability of the standard update rule in high-dimensional settings ($d = 784$).

Table 3: Original StatZO hyperparameter sweep. All configurations use $K_{\min} = 2$ and variance-to-norm threshold 0.75. Among the tested settings for this version, $K_{\max} = 4$ gives the lowest query cost, but the absolute cost remains high in the reported setting.

$K_{\max}$	Success	Avg	Std	Speedup vs ZoAQ
2	100%	6,212	675	0.05×
4	100%	4,382	511	0.07×
6	100%	7,349	1,714	0.04×
8	100%	7,424	588	0.04×
10	100%	10,022	1,316	0.03×

To ensure that the comparison targets the sampling strategy rather than optimizer defects, we also evaluate a stabilized StatZO variant with the R-AdaZO backbone. Table 4 shows the effect of this standardization: the query cost drops from roughly 4,382 to 1,996. The tested cap requiring the fewest queries is $K_{\max} = 2$, while increasing the adaptive budget ($K_{\max} \geq 4$) degrades performance. Even this strongest tested stabilized configuration remains well above the reported ZoAQ cost of 320 queries, suggesting that the query overhead required to validate geometry can outweigh the benefit of variance reduction in this setting.

Cross-dataset pattern: CIFAR-10. We replicate the StatZO sweep on CIFAR-10 ($d = 3072$) to examine whether the same pattern persists across dimensions. As shown in Table 5, the tested configuration requiring the fewest queries remains $K_{\max} = 2$, consistent with MNIST. Even with this tuned setting, StatZO achieves 964 queries, still 1.54× slower than ZoAQ (625 queries with $N_{\text{hist}} = 8$).

This analysis across datasets suggests that the same overhead from validation persists across problem scales: StatZO's variance estimation overhead prevents it from matching ZoAQ's efficiency in the reported settings.

Table 4: Stabilized StatZO hyperparameter sweep. Equipping StatZO with the R-AdaZO backbone improves performance relative to Table 3. Among the tested caps, the lowest query count occurs at $K_{\max} = 2$, and the method remains less query-efficient than ZoAQ in this setting.

$K_{\max}$	SUCCESS	AVG	STD	SPEEDUP VS ZOAQ
2	100%	1,996	1,231	0.16×
4	100%	4,089	953	0.08×
6	100%	5,137	2,004	0.06×
8	100%	3,937	1,802	0.08×
10	100%	4,572	2,189	0.07×

Table 5: StatZO hyperparameter sweep on CIFAR-10 ($d = 3072$). Consistent with MNIST, the tested configuration requiring the fewest queries is $K_{\max} = 2$, yet StatZO still lags behind ZoAQ (625 queries) in the reported setting.

$K_{\max}$	SUCCESS	AVG QUERIES	STD	SPEEDUP VS ZOAQ
2	100%	964	28	0.65×
4	100%	1,134	73	0.55×
6	100%	1,148	37	0.54×
8	100%	1,229	68	0.51×
10	100%	1,361	41	0.46×

F.3 VALIDATION OVERHEAD

Remark: Why history reuse is not folded into StatZO. A natural question is whether history reuse could also be added to StatZO. That modification is not theoretically neutral. As formalized by Bollapragada et al. (2024), the adaptive sampling analysis behind the Norm Condition is built around fresh i.i.d. samples for estimating local accuracy and variance quantities.

Incorporating historical queries $\{(\boldsymbol{u}, y) \in \mathcal{H}_{t-1}\}$ into the current estimate $\boldsymbol{g}_t$ introduces cross-step correlations and transport bias because $\boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t-\tau}$. It would therefore require a different variance-estimation analysis rather than being a drop-in improvement to StatZO's sample-variance estimator. ZoAQ's momentum-consistency condition is instead built around this correlation structure.

The empirical gap is therefore not only an implementation issue; it also reflects a structural bottleneck from validation overhead.

To estimate the variance $\mathbb{V}[\boldsymbol{g}]$, StatZO requires a reliable estimate of the function value $F(\boldsymbol{\theta}_t)$ at the current center point. This incurs a fixed cost of at least one extra query per iteration purely for statistical estimation.

- Overhead ratio: In a low-budget one-sided regime with two perturbed directional calls, an additional center-value query changes the accounting from 2 directional calls to 2 directions+ 1 baseline = 3 oracle calls, a 50% overhead relative to the directional calls alone.
- ZoAQ: The momentum-consistency test uses $\boldsymbol{m}_{t-1}$, a historical variable stored in memory, and requires no additional queries to validate the search direction.

This analysis makes the practical implication explicit: when only a few queries are used, the cost of validating geometry can itself become a meaningful part of the total budget.

F.4 Reference Stability

Beyond query overhead, there is a second statistical reason for StatZO's weaker performance in this regime: the instability of the reference signal itself. Both methods rely on a reference vector to judge the quality of the current estimate $\boldsymbol{g}_t$.

- StatZO uses the sample variance V_k (or effectively, the batch mean $\bar{\boldsymbol{g}}_k$ relative to variance) estimated from a small batch of size k.
- ZoAQ uses the historical momentum $\boldsymbol{m}_{t-1}$ as the reference anchor.

The comparisons above are empirical diagnostics of the tested configurations. The pathwise treatment of the EMA under overlapping accepted histories is given in Appendix C.

G Experimental Protocols

This appendix collects the implementation details, protocol standardization choices, and additional result tables referenced in the main text. Its role is to make the experimental comparison reproducible without expanding the main paper's evidence budget.

Table 6: Experimental evidence map. Each regime tests a distinct part of the mechanism chain, with claims stated at the level supported by the source ledger.

Regime	Evidence used	Claim boundary
Synthetic mechanisms	Convergence traces, sweep over fixed budgets, ablation of the norm trigger, and cosine diagnostics.	Query reductions are measured relative to the fixed protocol using 1.2M queries; cosine diagnostics measure observable consistency with the anchor.
Black-box attacks	MNIST and CIFAR-10 attacks under a common stabilized R-AdaZO backbone.	The result concerns black-box optimization under query constraints, with success and query counts reported under the stated protocol.
LLM fine-tuning	OPT-1.3B and OPT-13B LoRA fine-tuning measured by cumulative forward evaluations and task metrics.	Savings are FE reductions with loss and accuracy trade-offs that vary by task.
Stress checks	Controlled objective switch, history window, reset policy, threshold, and StatZO diagnostics.	These checks show how the mechanism behaves across the tested settings and parameter choices.

G.1 Algorithm and Cost Accounting

Algorithm 1 in the main paper is the authoritative pseudocode. This appendix section records the implementation-level accounting assumptions used by that algorithm in our experiments, together with the protocol choices needed for reproducibility.

Query accounting. The momentum-consistency test itself introduces no additional oracle queries beyond those already used to construct $\boldsymbol{g}_t$. In attacks, a one-sided forward-difference step with K_t sampled directions requires one baseline evaluation plus K_t perturbed evaluations. In synthetic benchmarks and LLM fine-tuning, the two-sided protocol requires two evaluations per sampled direction. The practical saving provided by ZoAQ is therefore a reduction in oracle calls rather than auxiliary computation: the extra cost is primarily memory, since the method retains recent direction-response pairs in order to induce query-reuse coupling. In seed-based implementations, this footprint can be reduced because directions need not be stored explicitly.

These accounting rules vary by protocol. Within each regime, all methods are compared under the same estimator protocol and the same cost definition; across regimes, the reported synthetic query

savings, attack query counts, and LLM savings in forward evaluations measure efficiency for their respective tasks rather than one absolute cost scale.

G.2 Compute, Assets, and Scope

Compute resources. Synthetic and attack experiments were run on standard CPU-accessible hardware or a single GPU. The main LLM fine-tuning runs were executed on a server with four NVIDIA RTX 4090 GPUs (24GB each), which was sufficient for the OPT-1.3B and OPT-13B LoRA runs using forward passes only. We report oracle calls or forward evaluations as the primary cost metric because the paper studies query allocation rather than systems throughput; elapsed time also depends on model loading, the data pipeline, and scheduling across runs.

Existing assets and licenses. The experimental study uses standard public datasets and model checkpoints, including MNIST, CIFAR-10, GLUE/SST-2, COPA, and OPT checkpoints, under their respective public terms of use. We rely on these assets only for benchmarking and do not redistribute modified versions in the paper artifact.

Table 7: Existing assets used in the experiments. When the original source or benchmark metadata does not state a single explicit license, we report that status rather than assigning a license name.

Asset	Version	Source	License	Terms-of-use note
MNIST	Original train/test split	MNIST database (LeCun et al., 1998)	Source terms; no explicit single license	Public benchmark use; no data redistribution
CIFAR-10	CIFAR-10 Python version	CIFAR page and report (Krizhevsky & Hinton, 2009)	Source terms; no explicit single license	Public benchmark use; no data redistribution
GLUE/SST-2	GLUE SST-2 task	GLUE benchmark (Wang et al., 2018)	GLUE/SST-2 source terms	Downstream evaluation only
COPA	COPA benchmark task	COPA release (Roemmele et al., 2011)	COPA source terms	Downstream evaluation only
OPT checkpoints	OPT-1.3B and OPT-13B	Meta OPT release (Zhang et al., 2022); HF cards	OPT model license	LoRA fine-tuning; no checkpoint redistribution

Broader impact. The attack experiments are included to study query efficiency in a controlled black-box setting, not to advocate misuse. The same techniques are also relevant to robustness evaluation and to black-box optimization problems where gradients are unavailable for benign reasons. We therefore view the main broader-impact consideration as responsible disclosure and evaluation practice rather than a new capability claim.

G.3 Baseline Standardization

Unified Optimizer Backbone. As noted in Section 6, standard zeroth-order optimizers often suffer from diverging second-moment estimates in high dimensions. To reduce this confounding factor in the main attack and stabilized adaptive comparisons, we use the R-AdaZO rectified update rule as the common backbone for R-AdaZO, stabilized StatZO, ZoAR-style reuse diagnostics, and ZoAQ. Legacy ZO-AdaMM and MeZO appear only in the protocol roles listed in Table 8. This standardization makes the comparison focus on sampling strategy rather than claiming uniform dominance over every original adaptive ZOO implementation.

Baselines Configuration. We compare ZoAQ against the following representative methods:

- R-AdaZO (Fixed-Budget SOTA): Uses a static query budget K per iteration. We set $K = 10$ for synthetic benchmarks and $K = 2$ or 4 for high-dimensional tasks, following the optimal settings reported in (Shu et al., 2025).

- Refined StatZO (Norm-Based SOTA): A stabilized version of StatZO (Bollapragada et al., 2024). We integrate it with the R-AdaZO backbone to prevent divergence. It adapts $K_t \in [K_{\min}, K_{\max}]$ based on the norm condition $\|\boldsymbol{g}_t - \boldsymbol{m}_{t-1}\|/\|\boldsymbol{m}_{t-1}\| \leq c_{\text{norm}}$, with $c_{\text{norm}} = 0.5$ as the default tolerance.
- ZoAQ (Ours): Utilizes the momentum-consistency condition $\cos(\boldsymbol{g}_t, \boldsymbol{m}_{t-1}) \geq \tau_{t-1}$ to trigger budget expansion. The effective threshold τ_t is updated via EMA.

Table 8 lists every baseline family that appears in the figures or tables and records its protocol role. When a baseline has a tunable cap, we report the sweep or selection rule rather than comparing against an unreported hand-picked setting.

Table 8: Baseline and protocol map. The table separates method role, update backbone, query rule, and selection rule so that the experiments compare strategies for query allocation under the intended protocol rather than mixing incompatible baselines.

Method	Role in experiments	Backbone/update rule	Query rule	Selection/reporting rule
ZO-AdaMM	Legacy synthetic reference	Adam-style ZOO	Fixed K in source run	Reference curve only
R-AdaZO	Fixed-budget baseline	Rectified adaptive ZOO	Fixed K: synthetic 10, attack 2, LLM 4	Protocol-fixed
ZoAR	Reuse diagnostic	History reuse without ZoAQ gate	Source-recorded reuse/fixed budget	Separates reuse from control
StatZO	Statistics-based adaptive baseline	Stabilized R-AdaZO; original diagnostic only	Synthetic $[1, 10]$; attack cap sweep; LLM $[2, 4]$	Best cap in sweep
MeZO	LLM fixed-budget baseline	Forward-only MeZO/LoRA	Fixed $K = 4$	LLM FE reference
ZoAQ	Ours	Stabilized adaptive ZO + query-reuse gate	Synthetic $[1, 10]$; attack $[1, 2]$; LLM $[1, 4]$ or $[1, 2]$	Matched seeds; no extra validation

G.4 Hyperparameter Settings

Synthetic experiments were conducted over $T = 60,000$ iterations in a $d = 100$ dimensional space, with results averaged over 5 independent runs. Random seeds are matched across methods within each synthetic run; figures plot mean trajectories without uncertainty bands for readability, while tables report means and standard deviations where the corresponding columns are shown. Attack and LLM configurations are specified in their dedicated subsections below.

Gradient Estimator Configuration. Consistent with the protocol defined in Sec. 3:

- Synthetic Benchmarks ($d = 100$): We use the two-sided symmetric-difference estimator to minimize gradient approximation error. Thus, R-AdaZO with $K = 10$ directions corresponds to 20 queries/step (Total $\approx$ 1.2M queries over 60k steps).
- High-Dimensional Attacks ($d \geq 784$): Following the R-AdaZO protocol (Shu et al., 2025), we use the one-sided forward-difference estimator. For example, R-AdaZO with $K = 2$ directions corresponds to 3 queries/step (1 baseline + 2 perturbed evaluations).
- LLM Fine-Tuning: We use the two-sided protocol with forward passes only and report forward evaluations (FE). A step with K_t directions costs $2K_t$ FE, as detailed in Appendix G.6.

Common Hyperparameters. A set of common hyperparameters was shared across most algorithms to ensure a fair comparison. These are summarized in Table 9.

Algorithm-Specific Configurations. The specific settings for each algorithm, including both non-adaptive and adaptive methods, are detailed in Table 10. The query budget per iteration is denoted by K_t. For non-adaptive methods, this is a fixed value. For adaptive methods, it varies within a specified range.

Table 9: Common hyperparameters used across all algorithms.

PARAMETER	DESCRIPTION	VALUE
T	TOTAL ITERATIONS	60,000
d	DIMENSION	100
NUM_RUNS	NUMBER OF INDEPENDENT RUNS	5
η (LR)	LEARNING RATE	0.001
(β_1, β_2)	MOMENTUM COEFFICIENTS	(0.9, 0.99)
μ	SMOOTHING PARAMETER	0.005
ϵ (EPS)	TERM FOR NUMERICAL STABILITY	1E-8

Table 10: Algorithm-specific hyperparameter settings.

ALGORITHM	PARAMETER	VALUE
NON-ADAPTIVE BASELINES		
R-ADAZO	FIXED QUERIES PER ITERATION, K_t	10
STATISTICS-BASED ADAPTIVE SOTA		
STATZO	QUERY RANGE, $K_t \in [K_{\min}, K_{\max}]$	[1, 10]
	VARIANCE-TO-NORM THRESHOLD	0.5
OUR PROPOSAL		
ZOAQ	QUERY RANGE, $K_t \in [K_{\min}, K_{\max}]$	[1, 10]
	INITIAL EMA THRESHOLD, τ_0	1.0
	EMA FACTOR, β_τ	0.9
	HISTORY WINDOW SIZE, N_{hist}	8
	BUDGET RESET POLICY	'MIN' (GREEDY RESET)

G.5 ATTACK PROTOCOL

Experimental Protocol. For the black-box adversarial attack experiments (Section 6.2), we follow the experimental protocol established in (Shu et al., 2025). We evaluate on two standard vision benchmarks:

- MNIST ($d = 784$) (LeCun et al., 1998): A CNN with two convolutional layers (32 and 64 filters, kernel size 3×3) followed by max-pooling and two fully connected layers (128 and 10 units), achieving 98.5% test accuracy. Perturbation constraint $\epsilon = 0.2$.
- CIFAR-10 ($d = 3072$) (Krizhevsky & Hinton, 2009): A deeper CNN achieving 85% test accuracy. Perturbation constraint $\epsilon = 0.03$.

The attack objective is to find a minimal perturbation $\boldsymbol{\delta}$ such that the model misclassifies $\boldsymbol{x} + \boldsymbol{\delta}$ while satisfying the ℓ_∞ constraint.

MNIST Hyperparameter Settings. All methods use identical common hyperparameters: learning rate $\eta = 0.01$, momentum coefficients $(\beta_1, \beta_2) = (0.9, 0.99)$, perturbation scale $\mu = 0.5$, and Adam's numerical stability constant $\varepsilon_{\text{Adam}} = 10^{-8}$ (Kingma & Ba, 2014). We evaluate over 5 independent runs with different random seeds (101, 202, 303, 404, 505), with a maximum of 20,000 iterations per attack.

For specific algorithms:

- Non-adaptive methods (R-AdaZO, ZO-AdaMM) use a fixed query budget of $K = 2$ per iteration.
- Methods with history reuse (ZoAQ and ZoAR) use a history buffer size $N_{\text{hist}} = 8$ (the tested value requiring the fewest queries in Table 18).
- ZoAQ is configured with $K_{\min} = 1$, $K_{\max} = 2$, initial threshold $\tau_0 = 1.0$, EMA factor $\beta_\tau = 0.9$, and the 'min' (greedy) reset policy.

CIFAR-10 Hyperparameter Settings. For the CIFAR-10 dataset ($d = 3072$), we use the following configurations:

- Learning rate: $\eta = 0.0001$ (lower than MNIST due to higher sensitivity)
- Perturbation scale: $\mu = 0.005$ (smaller due to higher sensitivity)
- History window: $N_{\text{hist}} = 8$ for ZoAQ (unified with MNIST for simplicity)
- StatZO: tested setting $K_{\max} = 2$ (the cap requiring the fewest queries in Table 5)
- All other parameters remain identical to MNIST

The CIFAR-10 table uses the same five-run reporting protocol as MNIST, with matched random seeds across methods within each run.

Fixed-Budget History-Reuse Control. History reuse improves query efficiency, but it does not recover the full benefit of adaptive allocation. Table 11 compares fixed-budget ZoAR with R-AdaZO and ZoAQ under the common stabilized backbone. All methods achieve 100% success. Relative to ZoAR at $K = 2$, ZoAQ uses 56.9% fewer queries on MNIST and 17.3% fewer on CIFAR-10.

Table 11: Fixed-budget history-reuse control for black-box attacks. Values are average queries $\pm$ standard deviation; all methods achieve 100% success.

DATASET	R-ADAZO ($K = 2$)	ZOAR ($K = 1$)	ZOAR ($K = 2$)	ZOAQ
MNIST	1527 ± 701	1089.2 ± 92.3	742.8 ± 99.9	320 ± 45
CIFAR-10	1006 ± 100	802.0 ± 38.0	755.4 ± 39.1	625 ± 15

G.6 LLM FINE-TUNING PROTOCOL

For the LLM fine-tuning experiments described in Section 6.3, we use the hyperparameter settings listed in Table 12. The base models are OPT checkpoints (Zhang et al., 2022), and the downstream tasks include SST-2 from GLUE (Wang et al., 2018) and COPA.

The main ZoAQ configuration uses the moderate cap $K_t \in [1, 4]$, which shares the same maximum direction budget as the fixed $K = 4$ baselines while allowing adaptive reductions. We also report $K_t \in [1, 2]$ as an aggressive endpoint with a smaller budget to show the trade-off between budget and quality.

All LLM metrics are reported as mean$\pm$std over three seeds. Total FE is the cumulative number of forward evaluations per run of 5,000 steps, not the sum across seeds. Best Eval Loss is the lowest held-out evaluation loss over checkpoints, Last Eval Loss is the final checkpoint loss, Accuracy is the primary task accuracy summary for the selected run/checkpoint protocol, and Dev Acc is the development-set accuracy reported by the training pipeline. We report both loss and accuracy because lower FE can affect them differently, especially on COPA.

Accounting for Forward Evaluations. We detail the exact FE accounting used to derive the total costs reported in the LLM experiments. All LLM fine-tuning tasks run for 5,000 training steps and use the two-sided protocol with forward passes only. A step with local direction budget K_t costs

$$\mathsf{FE}_t = 2K_t, \tag{95}$$

where each direction uses two perturbed forward evaluations. Thus fixed $K = 4$ costs 8 FE per step and 40,000 FE over 5,000 steps.

- Fixed baselines: MeZO and R-AdaZO use $K = 4$, so each complete run uses 40,000 FE.
- StatZO: The direction budget varies in $K_t \in [2, 4]$; total FE is the cumulative FE per 5,000-step run, reported as mean$\pm$std over three seeds.
- ZoAQ: The main setting uses $K_t \in [1, 4]$, while the endpoint with a smaller budget uses $K_t \in [1, 2]$. Total FE is the cumulative FE per run of 5,000 steps, reported as mean$\pm$std over three seeds, and the reported average K is computed as $\mathsf{FE}_{\text{tot}}/(2 \cdot 5000)$.

Table 12: Hyperparameter configurations for LLM fine-tuning tasks. All LLM runs use the two-sided zeroth-order protocol with forward passes only and report forward evaluations (FE) as the cost metric.

CATEGORY	PARAMETER	VALUE
COMMON	MODELS	OPT-1.3B, OPT-13B
	TASKS	SST-2, COPA
	SEEDS	3
	TRAINING STEPS	5,000
	EVALUATION INTERVAL	250 STEPS
	BATCH SIZE	16
	LEARNING RATE η	5×10^{-5}
	PERTURBATION SCALE	10^{-2}
	PRECISION	FP16
	ADAPTATION	LORA
MEZO / R-ADAZO	FIXED DIRECTION BUDGET	$K = 4$
STATZO	DIRECTION-BUDGET RANGE	$K_t \in [2, 4]$
ZOAQ	MAIN DIRECTION-BUDGET RANGE	$K_t \in [1, 4]$
	LOW-BUDGET VARIANT	$K_t \in [1, 2]$
	RESET POLICY	MIN (GREEDY)
	INITIAL EMA THRESHOLD τ_0	1.0
	EMA FACTOR β_τ	0.9
	HISTORY WINDOW SIZE, N_{hist}	15

Table 13: Complete LLM fine-tuning results over three seeds. Total FE is the cumulative number of forward evaluations under the two-sided protocol with forward passes only; FE savings are relative to fixed $K = 4$, which costs 40,000 FE over 5,000 steps.

MODEL	TASK	METHOD	TOTAL FE	AVG K	FE SAVINGS	BEST EVAL LOSS	LAST EVAL LOSS	ACCURACY	DEV ACC	TRAIN LOSS
OPT-1.3B	SST-2	MEZO $K = 4$	40000 ± 0	4.00 ± 0.00	$0.00 \pm 0.00\%$	0.240 ± 0.003	0.240 ± 0.003	0.910 ± 0.005	0.910 ± 0.005	0.476 ± 0.012
		R-ADAZO $K = 4$	40000 ± 0	4.00 ± 0.00	$0.00 \pm 0.00\%$	0.198 ± 0.002	0.199 ± 0.003	0.920 ± 0.004	0.910 ± 0.006	0.288 ± 0.008
		STATZO $K_t \in [2, 4]$	30746 ± 412	3.07 ± 0.04	$23.14 \pm 1.03\%$	0.206 ± 0.005	0.240 ± 0.006	0.910 ± 0.007	0.880 ± 0.010	0.302 ± 0.015
		ZOAQ $K_t \in [1, 2]$	12662 ± 235	1.27 ± 0.02	$68.35 \pm 0.59\%$	0.209 ± 0.004	0.237 ± 0.005	0.870 ± 0.011	0.910 ± 0.008	0.335 ± 0.014
		ZOAQ $K_t \in [1, 4]$	21714 ± 318	2.17 ± 0.03	$45.72 \pm 0.80\%$	0.195 ± 0.003	0.198 ± 0.004	0.920 ± 0.005	0.910 ± 0.005	0.285 ± 0.010
	COPA	MEZO $K = 4$	40000 ± 0	4.00 ± 0.00	$0.00 \pm 0.00\%$	0.532 ± 0.006	0.532 ± 0.006	0.770 ± 0.012	0.750 ± 0.015	0.537 ± 0.018
		R-ADAZO $K = 4$	40000 ± 0	4.00 ± 0.00	$0.00 \pm 0.00\%$	0.431 ± 0.004	0.476 ± 0.005	0.780 ± 0.010	0.780 ± 0.012	0.423 ± 0.010
		STATZO $K_t \in [2, 4]$	30498 ± 505	3.05 ± 0.05	$23.76 \pm 1.26\%$	0.461 ± 0.007	0.461 ± 0.007	0.790 ± 0.014	0.780 ± 0.018	0.459 ± 0.015
		ZOAQ $K_t \in [1, 2]$	12740 ± 288	1.27 ± 0.03	$68.15 \pm 0.72\%$	0.490 ± 0.008	0.490 ± 0.008	0.790 ± 0.016	0.770 ± 0.015	0.499 ± 0.017
		ZOAQ $K_t \in [1, 4]$	21946 ± 345	2.19 ± 0.03	$45.14 \pm 0.86\%$	0.485 ± 0.005	0.492 ± 0.006	0.790 ± 0.012	0.770 ± 0.010	0.460 ± 0.014
OPT-13B	SST-2	MEZO $K = 4$	40000 ± 0	4.00 ± 0.00	$0.00 \pm 0.00\%$	0.230 ± 0.003	0.231 ± 0.003	0.915 ± 0.004	0.912 ± 0.004	0.450 ± 0.010
		R-ADAZO $K = 4$	40000 ± 0	4.00 ± 0.00	$0.00 \pm 0.00\%$	0.185 ± 0.002	0.188 ± 0.002	0.938 ± 0.003	0.925 ± 0.005	0.260 ± 0.007
		STATZO $K_t \in [2, 4]$	31350 ± 390	3.14 ± 0.04	$21.63 \pm 0.98\%$	0.202 ± 0.004	0.238 ± 0.005	0.908 ± 0.006	0.885 ± 0.008	0.295 ± 0.012
		ZOAQ $K_t \in [1, 2]$	12920 ± 260	1.29 ± 0.03	$67.70 \pm 0.65\%$	0.208 ± 0.004	0.230 ± 0.005	0.875 ± 0.009	0.910 ± 0.007	0.325 ± 0.011
		ZOAQ $K_t \in [1, 4]$	22245 ± 290	2.22 ± 0.03	$44.39 \pm 0.73\%$	0.192 ± 0.003	0.198 ± 0.004	0.920 ± 0.005	0.918 ± 0.004	0.278 ± 0.009
	COPA	MEZO $K = 4$	40000 ± 0	4.00 ± 0.00	$0.00 \pm 0.00\%$	0.495 ± 0.008	0.498 ± 0.008	0.775 ± 0.015	0.760 ± 0.018	0.505 ± 0.016
		R-ADAZO $K = 4$	40000 ± 0	4.00 ± 0.00	$0.00 \pm 0.00\%$	0.405 ± 0.005	0.425 ± 0.006	0.798 ± 0.012	0.785 ± 0.014	0.395 ± 0.011
		STATZO $K_t \in [2, 4]$	31480 ± 480	3.15 ± 0.05	$21.30 \pm 1.20\%$	0.452 ± 0.009	0.458 ± 0.009	0.785 ± 0.016	0.790 ± 0.015	0.445 ± 0.014
		ZOAQ $K_t \in [1, 2]$	13250 ± 310	1.33 ± 0.03	$66.88 \pm 0.78\%$	0.482 ± 0.010	0.488 ± 0.010	0.770 ± 0.014	0.765 ± 0.016	0.490 ± 0.018
		ZOAQ $K_t \in [1, 4]$	22780 ± 360	2.28 ± 0.04	$43.05 \pm 0.90\%$	0.428 ± 0.006	0.435 ± 0.007	0.792 ± 0.011	0.788 ± 0.012	0.420 ± 0.012

Sensitivity to the LLM History Window. The main OPT configuration uses $N_{\text{hist}} = 15$. Table 14 reports the changes obtained with windows of 8 and 24 on OPT-1.3B. Increasing the window to 24 further reduces FE on both tasks, with changes in loss and accuracy that vary by task. This comparison measures how the reported outcomes vary with the history window.

G.7 BUDGET SWEEPS

This subsection collects supplementary sweeps that isolate the role of query budget choice and controller design. They complement the evaluation in the main text without changing the main experimental protocol.

G.8 CONTROLLER VARIANTS

This diagnostic checks whether the query savings come from the full accept/expand/reset controller rather than from disabling one side of its decision rule. These are ablation-only controllers, not

Table 14: Sensitivity to the history window on OPT-1.3B relative to the main setting $N_{\text{hist}} = 15$. Negative ΔFE denotes fewer forward evaluations.

TASK	N_{hist}	ΔFE VS. 15	Δ BEST LOSS VS. 15 ↓	Δ ACCURACY VS. 15	IMMEDIATE-ACCEPT RATE
SST-2	8	+2326 (+10.78%)	−0.0056	+0.02	69.8%
SST-2	24	−648 (−3.00%)	−0.0060	+0.04	74.7%
COPA	8	+1366 (+6.14%)	+0.0078	+0.03	69.7%
COPA	24	−818 (−3.67%)	+0.0522	+0.01	73.1%

Table 15: Sweep over fixed budgets on synthetic benchmarks. The comparison shows how ZoAQ combines the efficiency of smaller budgets with the performance of larger budgets without manual tuning of the query count.

FUNCTION	METHOD	FINAL GAP	QUERIES
QUADRATIC	ZOAQ (ADAPTIVE)	6.35E-04	590,024
	R-ADAZO ($K = 2$)	7.52E-02	240,000
	R-ADAZO ($K = 4$)	7.35E-03	480,000
	R-ADAZO ($K = 8$)	6.15E-03	960,000
LEVY	ZOAQ (ADAPTIVE)	1.50E+02	588,591
	R-ADAZO ($K = 2$)	3.45E+02	240,000
	R-ADAZO ($K = 4$)	2.15E+02	480,000
	R-ADAZO ($K = 8$)	1.62E+02	960,000

additional branches used by the final ZoAQ algorithm. All variants share the same history-reuse estimator and update backbone; they differ only in how the momentum-consistency test is acted upon.

- ZoAQ-Adaptive: the reported controller, which accepts the first budget whose history-reuse estimate passes the momentum-consistency gate and otherwise expands the local budget up to the cap.
- ZoAQ-AlwaysMomentum: a diagnostic that bypasses rejection by the gate and therefore keeps the behavior with few queries driven by history reuse even when the consistency score would have triggered expansion.
- ZoAQ-AlwaysGradient: a diagnostic that suppresses the history-reuse acceptance shortcut and relies on the expanded current-budget estimate instead.

Table 16 shows the results under the neutral threshold setting $\tau_0 = 0.0$ used for this diagnostic. The full controller achieves the best balance: the ablation using fewer queries can save queries but reaches worse final gaps, while the ablation using the expanded estimate spends more queries without matching the adaptive controller. Figure 5 shows the corresponding convergence dynamics.

G.9 THRESHOLD SENSITIVITY

To assess the sensitivity of ZoAQ to the initial consistency threshold τ_0, we evaluated the algorithm under different threshold settings. We use a conservative initialization of $\tau_0 = 1.0$ for the main benchmarks (to enforce strict directional consistency at the start), and include that default alongside a neutral-range sweep of $\tau_0 \in \{-0.5, -0.3, -0.1, 0.0, 0.1, 0.3, 0.5\}$.

Table 17 shows representative numerical results, and Figure 6 visualizes the convergence dynamics across all four benchmark functions. The main stable quantity in this sweep is query allocation: total queries vary little across the listed thresholds. Final gaps remain in the same qualitative regime but are not threshold-invariant, especially for stricter positive thresholds on Quadratic. This is the intended conclusion of the diagnostic: the EMA mechanism reduces sensitivity to initialization, but it does not make the threshold irrelevant.

Table 16: Diagnostic of controller actions. The full accept/expand/reset controller balances query efficiency and final accuracy across diverse landscapes.

FUNCTION	VARIANT	FINAL GAP	TOTAL QUERIES
QUADRATIC	ADAPTIVE	3.16E-03	623,893
QUADRATIC	ALWAYSMOMENTUM	6.90E-02	573,652
QUADRATIC	ALWAYSGRADIENT	6.49E-03	605,827
ROSENBROCK	ADAPTIVE	66.89	632,483
ROSENBROCK	ALWAYSMOMENTUM	94.50	568,096
ROSENBROCK	ALWAYSGRADIENT	71.21	615,656

Table 17: Numerical sensitivity analysis of the initial threshold τ_0. Query counts are stable across the tested values, while final gaps can vary; the diagnostic supports reduced sensitivity from EMA calibration rather than threshold invariance.

FUNCTION	τ_0	FINAL GAP	TOTAL QUERIES
QUADRATIC	-0.5	3.16E-03	623,893
QUADRATIC	-0.1	3.16E-03	623,893
QUADRATIC	0.0	3.16E-03	623,893
QUADRATIC	0.3	3.53E-03	623,749
QUADRATIC	0.5	7.33E-03	622,775
QUADRATIC	1.0 (DEFAULT)	6.35E-04	590,024
ROSENBROCK	-0.5	64.48	631,996
ROSENBROCK	-0.1	67.66	632,016
ROSENBROCK	0.0	66.89	632,483
ROSENBROCK	0.3	67.66	632,602
ROSENBROCK	0.5	66.83	631,655
ROSENBROCK	1.0 (DEFAULT)	58.75	569,556

G.10 SENSITIVITY TO THE HISTORY WINDOW

The history window controls the expected trade-off between reuse and staleness. Very small windows are noisy, while overly long windows introduce stale search directions. On MNIST adversarial attacks, the empirical curve is convex and reaches its best point at the default $N_{\text{hist}} = 8$ used in the main attack experiments.

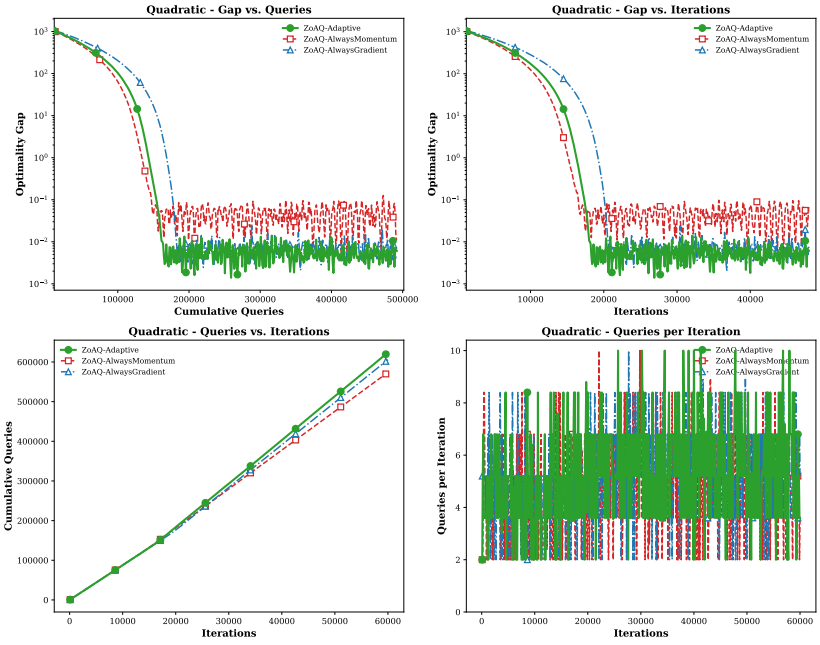
G.11 SYNTHETIC CONVERGENCE CURVES

Figures 7 and 8 present the complete convergence dynamics for the Quadratic, Cubic, and Levy functions, complementing the Rosenbrock results shown in the main paper. The plots support three qualitative patterns in the tested suite:

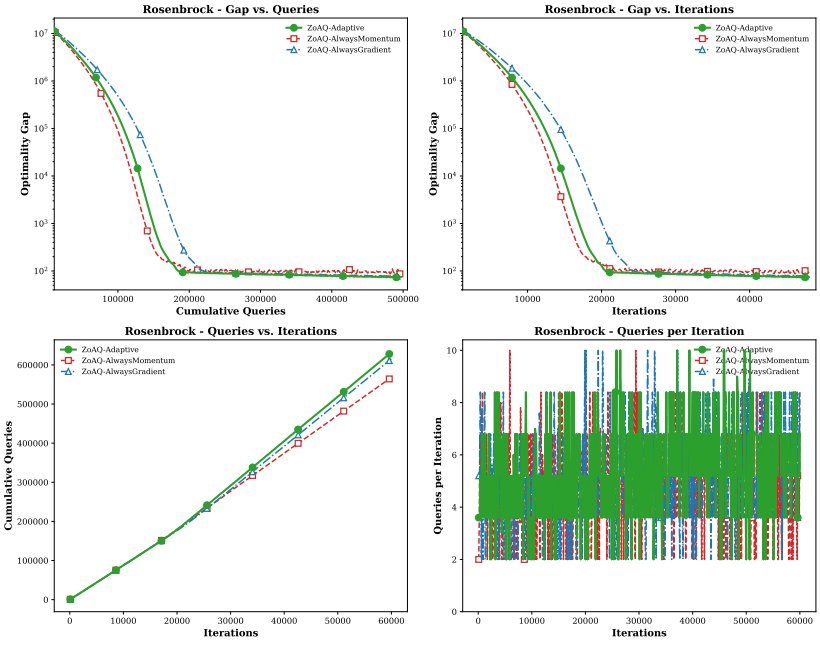
- Efficiency: ZoAQ (green) often reaches a low-error region with fewer queries than the fixed-budget baselines in the ‘Gap vs. Queries‘ plots.
- Overhead: StatZO (blue) shows an initial lag on the Quadratic and Cubic functions, consistent with the diagnosed overhead from validation.
- Cost: Methods with fixed budgets consume the query budget set by the protocol regardless of local difficulty.

H ADDITIONAL MECHANISM DIAGNOSTICS

This section gathers the secondary mechanism diagnostics referenced from the main synthetic discussion. The subsections below separate the effects of reset policy, threshold dynamics, stress tests under changing objectives, and the failure mode of the norm trigger so that each diagnostic supports a distinct question.



(a) Quadratic

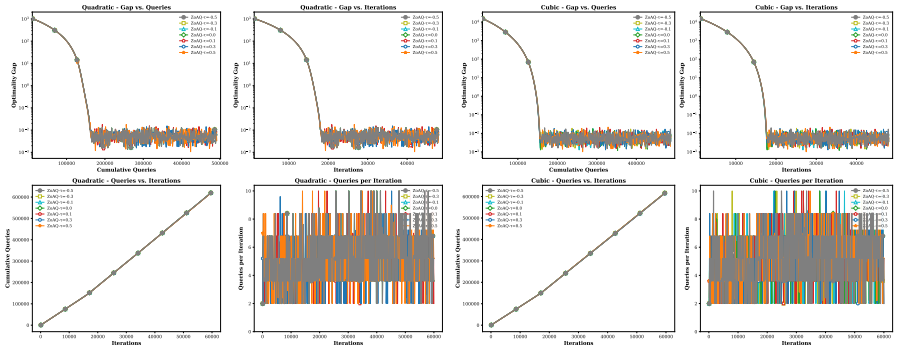


(b) Rosenbrock

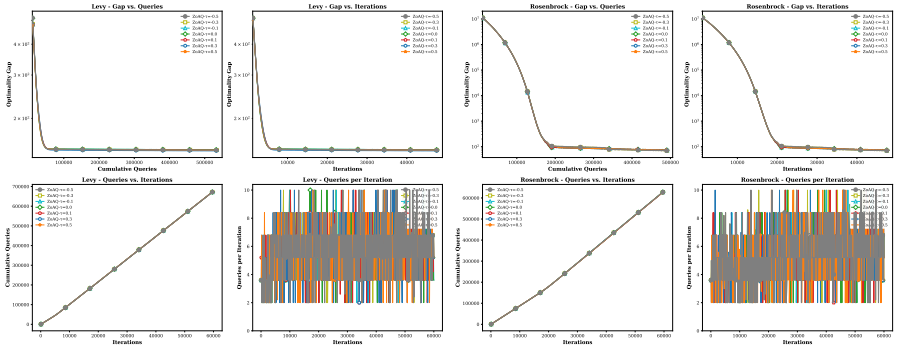
Figure 5: Convergence comparison across variants of the controller actions. The full controller (ZoAQ-Adaptive) provides the best balance between query efficiency and final accuracy.

H.1 Reset Ablation

Table 19 presents the numerical comparison of different budget reset strategies. After an accepted step, 'min' resets the next search to $K_{\min}$, 'half' restarts from the midpoint between $K_{\min}$ and the accepted budget, and 'linear' continues from the previous accepted budget and therefore reduces reset aggressiveness. The 'Min' policy (Greedy Reset) is the most query-efficient among the tested reset strategies across the benchmark functions.



(a) Quadratic (b) Cubic



(c) Levy (d) Rosenbrock

Figure 6: Sensitivity analysis of the initial threshold τ_0 across four benchmarks. The trajectories are broadly stable in the tested range, while the table shows that stricter positive initial thresholds can still affect final gaps.

Table 18: Sensitivity analysis for history window size N_{hist} on MNIST adversarial attacks. The default $N_{\text{hist}} = 8$ gives the lowest average query count among the tested settings.

N_{hist}	AVG QUERIES	STD QUERIES
4	374.4	126.9
6	362.0	57.6
8	319.6	44.8
10	335.6	65.7
12	349.2	88.5

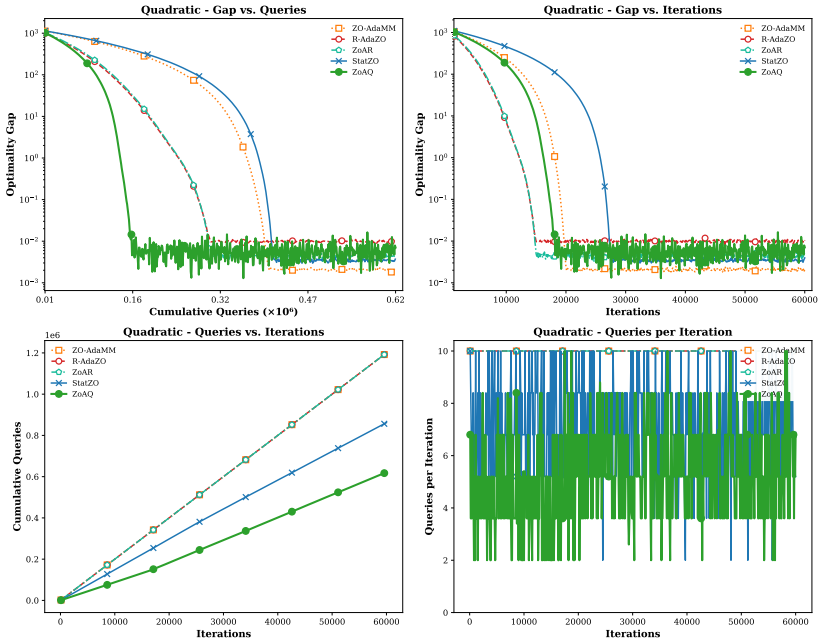
H.2 THRESHOLD DYNAMICS

Figure 9 illustrates the dynamic evolution of τ_t. The EMA threshold records the recent scale of the momentum-consistency score: lower values relax the gate after inconsistent evidence, while higher values require stronger agreement before accepting a low-budget step.

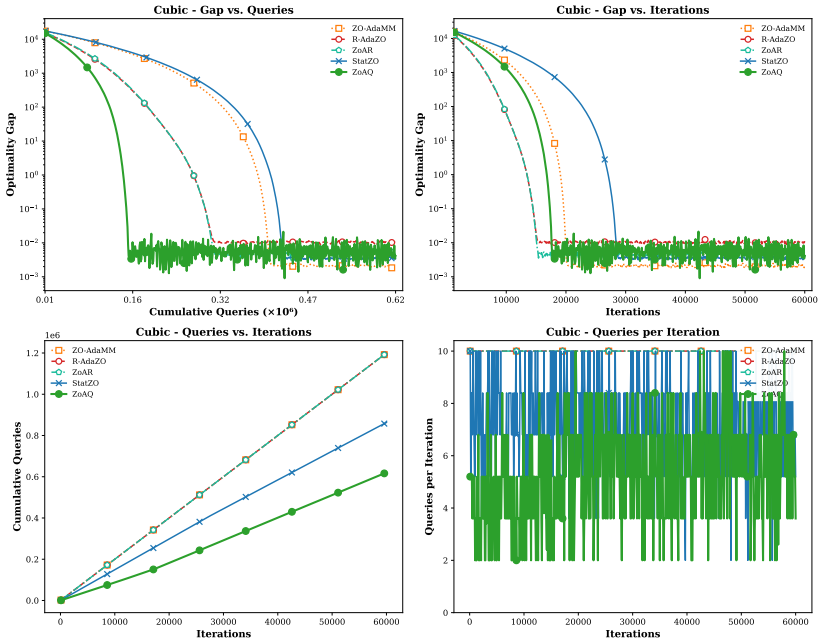
H.3 CONTROLLED OBJECTIVE SWITCH

A key concern with history reuse is whether stale gradients could trap the optimizer when the landscape changes abruptly. We design a Sharp Turn stress test to directly evaluate ZoAQ's drift-detection capability.

Protocol. We optimize the Rosenbrock function ($d = 100$) for 15,000 iterations, then instantly switch the objective to the Levy function (same d) and continue for another 15,000 iterations. This



(a) Quadratic Function

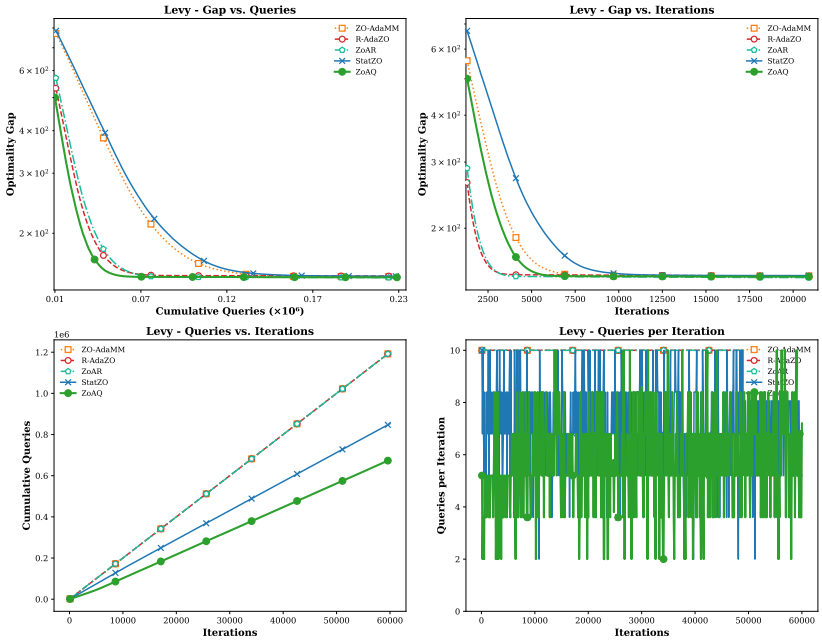


(b) Cubic Function

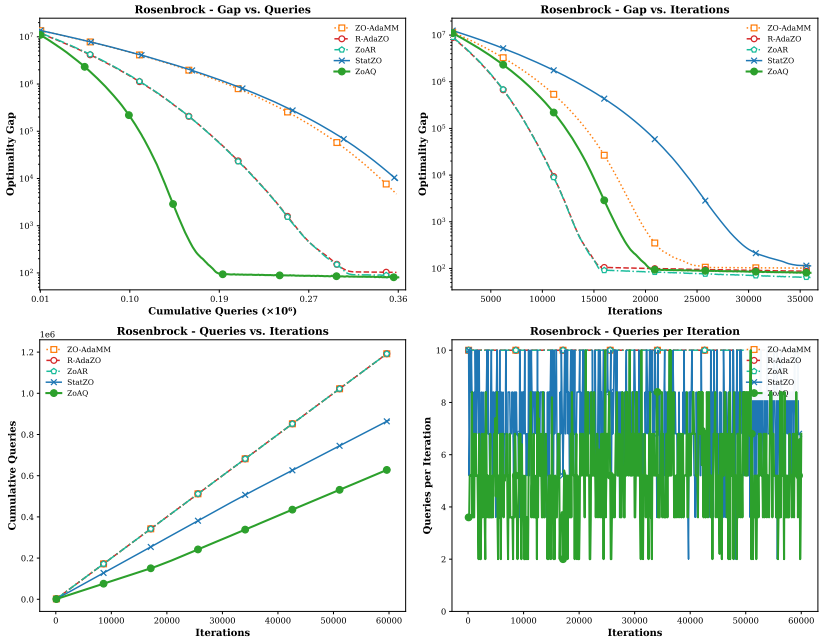
Figure 7: Full convergence dynamics on smooth landscapes (Quadratic and Cubic). The curves show ZoAQ (green) reaching low objective gaps with fewer queries in these tested settings.

simulates an abrupt distribution shift where accumulated momentum points in a fundamentally wrong direction.

Results. Figure 10 shows the optimality gap. The apparent drop at step 15,000 is expected because Rosenbrock has naturally larger objective values than Levy. After the switch, ZoAQ reaches a lower final loss than R-AdaZO in this test, indicating that the retained buffer does not trap the method in the old landscape under this protocol.



(a) Levy Function



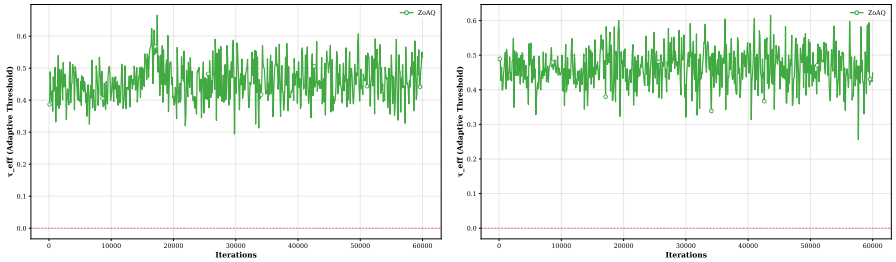
(b) Rosenbrock Function

Figure 8: Full convergence dynamics on difficult landscapes (Levy and Rosenbrock). The plots test whether the same adaptive-query behavior persists beyond smooth objectives.

Mechanism: τ_t as a drift-sensitive trigger. Figure 11 shows that, at the switch point, τ_t drops sharply toward zero, indicating a mismatch between the momentum accumulated on Rosenbrock and the new Levy geometry. The failed-consistency phase refreshes the buffer through added queries, while the lower EMA threshold records the new local score scale rather than treating the old high-consistency regime as still valid.

Table 19: Ablation study: performance across reset strategies (summarized). 'Min' minimizes total queries across all functions; 'Linear' often yields the smallest final gap but at ~1.06M queries.

FUNCTION	STRATEGY	FINAL GAP	TOTAL QUERIES
QUADRATIC	MIN	6.345E-04	590,024
QUADRATIC	HALF	6.194E-04	676,337
QUADRATIC	LINEAR	3.131E-04	1,064,733
CUBIC	MIN	6.460E-04	590,098
CUBIC	HALF	6.100E-04	674,920
CUBIC	LINEAR	3.588E-04	1,063,694
LEVY	MIN	1.505E+02	588,591
LEVY	HALF	1.510E+02	683,240
LEVY	LINEAR	1.506E+02	1,064,713
ROSENBROCK	MIN	5.875E+01	569,556
ROSENBROCK	HALF	5.523E+01	657,151
ROSENBROCK	LINEAR	3.465E+01	1,066,003



(a) Quadratic (b) Cubic

(c) Levy (d) Rosenbrock

Figure 9: Dynamics of the effective threshold τ_t across functions. The EMA adapts the gate to the observed momentum-consistency scale, lowering it after inconsistent evidence and keeping it higher when recent accepted scores remain strong.

Takeaway. This experiment shows that ZoAQ refreshes its history after the objective changes. The momentum-consistency trigger detects the mismatch, adds fresh evidence during failed tests, and recalibrates the acceptance scale.

H.4 TRIGGER METRIC ABLATION

The main text compares the momentum-consistency trigger with a relative norm trigger,

$$\frac{\|\boldsymbol{g}_t - \boldsymbol{m}_{t-1}\|}{\|\boldsymbol{m}_{t-1}\|} \leq c_{\text{norm}}. \tag{96}$$

This ablation asks whether matching vector magnitude is a useful allocation signal in the reported high-dimensional setting. It is not used as a theoretical lower bound on norm-based adaptive sampling.

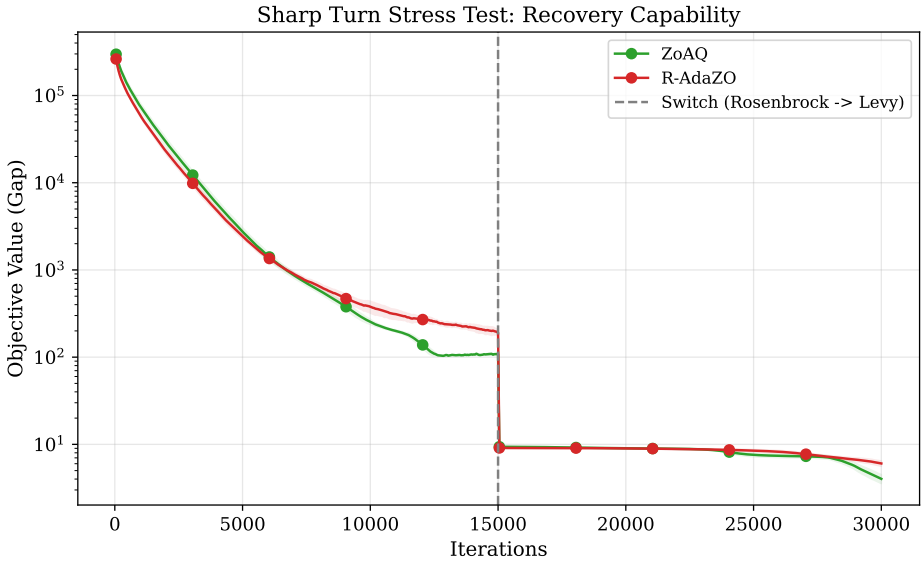


Figure 10: Optimality Gap during the Sharp Turn test. After the switch at $T = 15$k, ZoAQ adapts faster to the new Levy landscape, reaching a lower final objective than R-AdaZO.

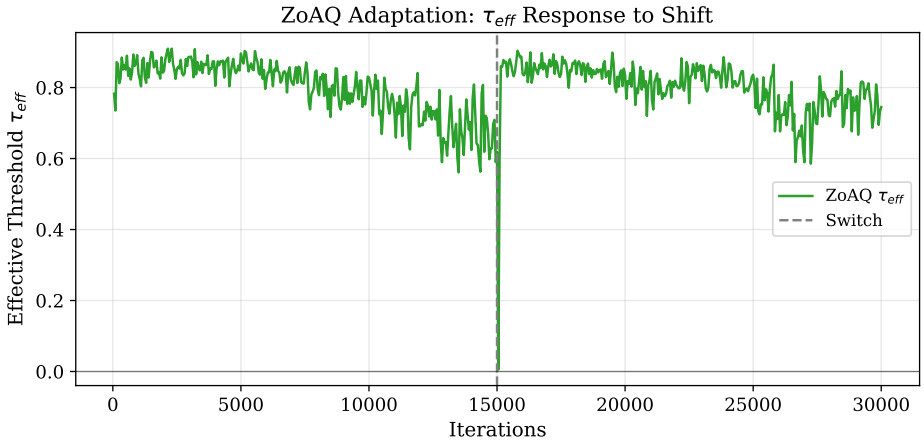


Figure 11: ZoAQ's effective threshold τ_t during the Sharp Turn. The drop at $T = 15$k shows that the momentum-consistency trigger reacts to the abrupt landscape change.

Two effects can make Eq. 96 more demanding than a cosine trigger. First, the paired Gaussian contribution has the local second-moment scale in Eq. 31, so finite-direction estimates can fluctuate substantially in magnitude even when their directions are useful. Second, the denominator becomes sensitive when the momentum norm is small. A relative norm trigger may therefore expand the budget in response to radial fluctuation as well as directional disagreement.

The formal descent result in Appendix E isolates this distinction with a globally normalized proxy update: its one-step bound depends on angular alignment and not on the candidate norm. This statement should not be extended automatically to the experimental R-AdaZO update. Coordinatewise second-moment preconditioning need not preserve Euclidean angles; the ablation below is empirical evidence about the tested controller and backbone, not a proof of scale invariance for Adam-style updates.

Figure 12 reports the comparison. On Rosenbrock with $d = 10{,}000$, the relative-norm trigger uses 28.1% more queries to reach the same target loss, while ZoAQ obtains a larger loss drop per 1k queries (1.9 versus 1.3). The result supports the use of directional consistency in this operating regime without implying that norm-based criteria are universally inferior.

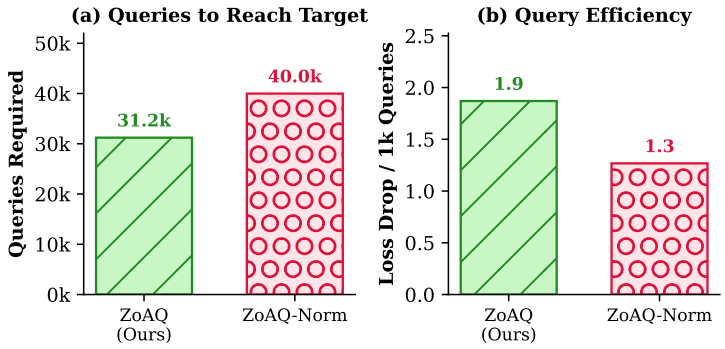


Figure 12: Trigger-metric ablation on Rosenbrock ($d = 10{,}000$). In this run, the relative-norm trigger requires 28.1% more queries to reach the same target, and its loss drop per 1k queries is 1.3 versus 1.9 for momentum consistency.

I Query-Reuse Diagnostics

This section connects the stored-response estimator analyzed in Appendix B to the empirical momentum-consistency traces. Its purpose is diagnostic: the formal guarantees come from candidate accuracy and debiased-EMA tracking, not from treating overlap-induced agreement as a certificate by itself.

I.1 Fresh and Stored-Response Estimation

A fresh spatial estimator uses only directions acquired at the current step,

$$\boldsymbol{g}_t^{\text{fresh}} = \frac{1}{K_t} \sum_{k=1}^{K_t} \boldsymbol{g}_\mu(\boldsymbol{\theta}_t, \boldsymbol{u}_{t,k}). \tag{97}$$

ZoAQ instead averages the contributions observed at acquisition in its candidate record set,

$$\boldsymbol{g}_t^{\text{hist}}(K) = \frac{1}{n_t(K)} \sum_{j \in \mathcal{J}_t(K)} X_j, \qquad X_j = r_\mu(\boldsymbol{\theta}_{a_j}, \boldsymbol{u}_j)\boldsymbol{u}_j. \tag{98}$$

Retained records therefore increase the effective sample count without new function evaluations, but they target gradients at their acquisition iterates. The exact trade-off is the error envelope

$$\|\boldsymbol{g}_t^{\text{hist}}(K) - h_t\| \le q(n_t(K), \delta) + \frac{LA_t(K)}{n_t(K)}. \tag{99}$$

The first term decreases with accumulated evidence; the second prices movement since acquisition. Their balance is the mechanism used by the theory.

Proposition I.1 (When Momentum Consistency Is Interpretable). *Suppose the simultaneous candidate event holds, and let accepted estimates satisfy $\|\boldsymbol{g}_r - h_r\| \le \varepsilon_r$. Then Proposition C.1 gives*

$$\|\widetilde{\boldsymbol{m}}_{t-1} - h_t\| \le \sum_{r<t} \bar{w}_{r,t}\varepsilon_r + LD_t^{\text{ema}}. \tag{100}$$

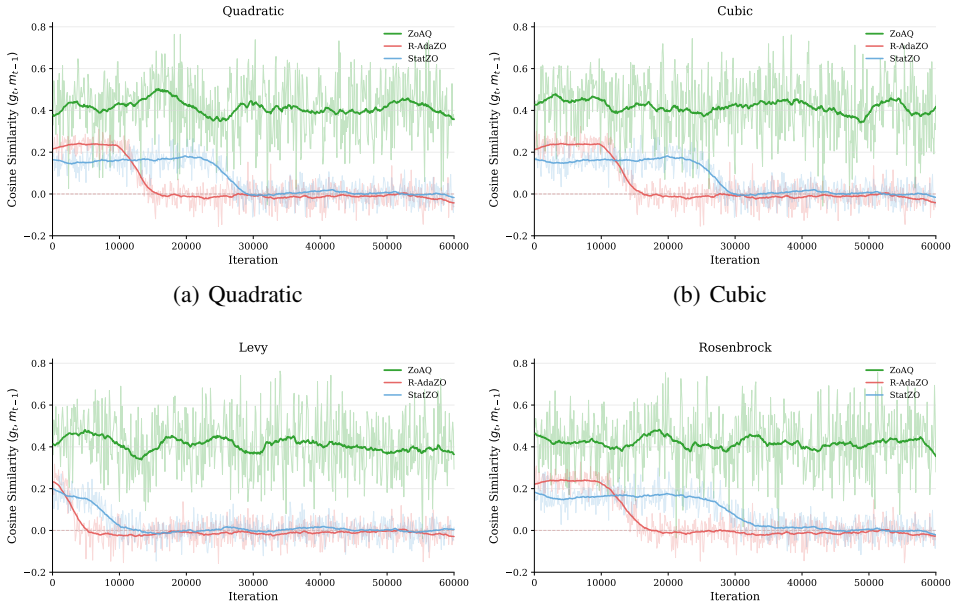
If the right-hand side is at most $\kappa_t\|h_t\|$, a passing score obeys the alignment bound in Eq. 64. Thus history reuse can stabilize the score through improved candidate accuracy, while FIFO staleness and EMA lag determine whether that score still represents current geometry.

Proof. Equation 100 is Eq. 57 with the certificates for accepted candidates substituted. The final statement is the soundness part of Theorem C.3. □

The proposition also explains why raw agreement alone is insufficient. A large stale buffer may make consecutive estimates visually stable while tracking an outdated direction. The path-length terms in Eqs. 99–100 exclude precisely this failure mode.

I.2 Cosine Stability

We record the cosine similarity between the candidate estimate and the previous momentum throughout optimization. Figure 13 reports the synthetic diagnostics.



(a) Quadratic (b) Cubic

(c) Levy (d) Rosenbrock

Figure 13: Synthetic cosine similarity between the candidate estimate and previous momentum under the 60k-step protocol. Curves are recorded every 100 iterations and averaged over five runs; faint lines show the recorded means and dark lines show smoothed trends. ZoAQ remains more persistently positive than the displayed baselines in these runs.

The comparison is descriptive rather than a direct measurement of the theorem's tracking ratio. A positive trace is consistent with a stable reference, but the guarantee additionally requires the candidate and path-length bounds above.

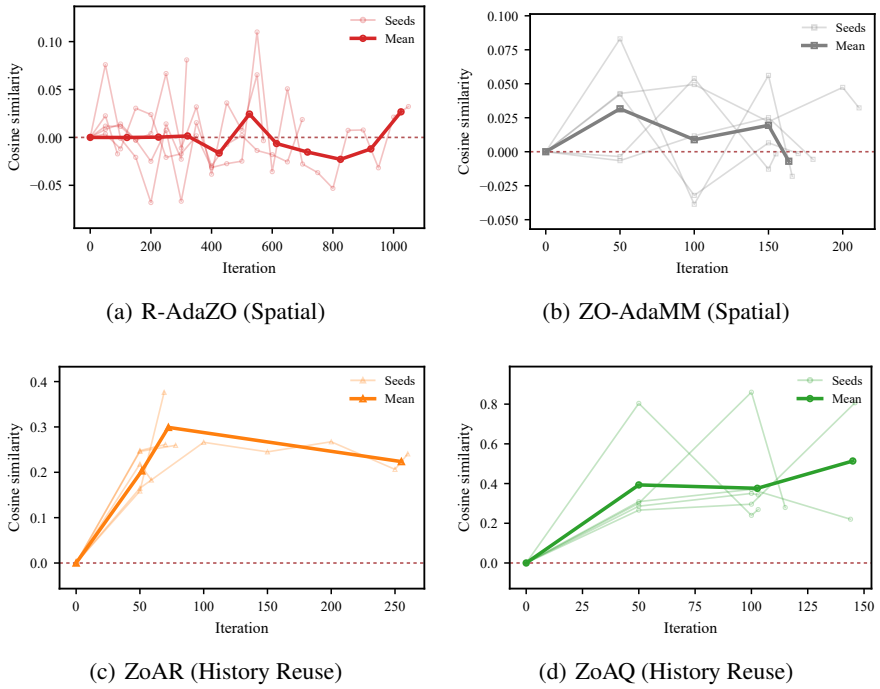
Figure 14 gives the corresponding black-box attack diagnostic. The same qualitative pattern appears in this setting: the history-reuse variants maintain a more persistent positive similarity signal, whereas the spatial variants fluctuate more strongly.

I.3 Query-Space and Update-Space Averaging

History reuse and momentum average at different stages. Equation 98 pools acquisition records before a candidate is accepted. Momentum then averages the sequence of accepted estimates:

$$\boldsymbol{m}_t = (1-\beta_1)\sum_{r=1}^{t}\beta_1^{t-r}\boldsymbol{g}_r. \tag{101}$$

The first operation trades sampling concentration against response staleness; the second trades estimate noise against trajectory lag. Their effects are complementary but not interchangeable, and both costs appear explicitly in the proofs.



(a) R-AdaZO (Spatial)

(b) ZO-AdaMM (Spatial)

(c) ZoAR (History Reuse)

(d) ZoAQ (History Reuse)

Figure 14: Momentum-consistency traces during black-box adversarial attacks on MNIST. History-reuse variants show a more persistently positive signal than the spatial variants in the reported runs.

This view suggests a practical design rule. A longer history is useful in noisy, slowly moving regions, whereas faster movement favors a shorter history. The capacity scale in Eq. 51 makes this dependence explicit. Similarly, the momentum parameter should be interpreted through the lag factor $(1-\beta_1)^{-1}$ in Eq. 58, rather than as cost-free smoothing.

I.4 Diagnostic Scope

The figures support two empirical observations: candidates built from stored responses yield a more stable score in the reported regimes, and failed tests add fresh evidence before the candidate is accepted. They do not show that overlap guarantees correctness. Formal interpretability requires the same ingredients as the main theorem. These are candidate concentration valid under selection, bounded acquisition staleness, debiased EMA tracking, and a feasible threshold band.

J Limitations and Future Directions

We currently retain a fixed number N_{hist} of recent records and initialize the EMA threshold with a warm start. Our window sweep and controlled objective switch show when reused responses remain helpful and when they become stale. Adapting or reweighting the retained history as the trajectory changes is therefore a natural next step. The present analysis assumes explicit bounds on iterate movement, EMA tracking, thresholds, and local query budgets; extending it across initialization, resets, and record pruning would give a fuller account of transient behavior. We evaluate forward-pass fine-tuning on OPT-1.3B and OPT-13B with SST-2 and COPA. Broader model families, tasks, and training horizons remain to be studied.