ACCELERATED METHODS FOR NONCONVEX OPTIMIZATION USING COMPARISONS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study the problem of finding stationary points of non-convex functions when access to the objective is provided only through a comparison oracle that, given two points, outputs which has the larger function value. For a twice differentiable $f\colon \mathbb{R}^n \to \mathbb{R}$ with Lipschitz gradient and Hessian, we develop an algorithm that outputs a list of points containing at least one ϵ-stationary point using $\widetilde{O}(n/\epsilon^{7/4})$ queries, matching the state-of-the-art result of zeroth-order methods with function evaluations. To the best of our knowledge, our algorithm is the first one that achieves Nesterov-type acceleration with comparison queries. Technically, our approach combines restarted accelerated gradient descent with comparison-based estimates of gradient directions. Specifically, we design a geometric grid of guesses for gradient norms at reference points and an anchor-based procedure for estimating gradient norm ratios, enabling accelerated updates under controlled errors without direct access to gradient magnitudes.

1 INTRODUCTION

Finding stationary points of a nonconvex function $f\colon \mathbb{R}^n \to \mathbb{R}$ is a fundamental problem in continuous optimization and a natural starting point for analyzing optimization algorithms in machine learning. Global optimization for general nonconvex functions is NP-hard in the worst case, and efficient global guarantees typically require additional structures (Murty & Kabadi, 1987). A quantifiable alternative is to find an ϵ-stationary point, namely a point x satisfying

$$\|\nabla f(x)\| \le \epsilon. \tag{1}$$

Although a first-order stationary point does not need to be a local minimum, this condition ensures that the first-order variation of the objective function is small in all directions, and it is a standard criterion for studying smooth non-convex optimization (Nocedal & Wright, 2006; Nesterov, 2013; Ghadimi & Lan, 2013; Carmon et al., 2017; 2020).

When gradients are available, the complexity of finding stationary points has a well-established baseline. Suppose that f has an L-Lipschitz gradient, meaning that

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \qquad \forall\, x, y \in \mathbb{R}^n, \tag{2}$$

and let $\Delta_f = f(x_0) - \inf_x f(x)$ denote the finite initial objective gap. Gradient descent finds an ϵ-stationary point using $O(L\Delta_f \epsilon^{-2})$ gradient queries (Nesterov, 2013).

An important additional assumption is Lipschitz continuity of the Hessian:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \le \rho\|x - y\|, \qquad \forall\, x, y \in \mathbb{R}^n. \tag{3}$$

This condition controls how quickly local curvature changes and ensures accurate local quadratic approximations. When second-order derivatives are available, the cubic regularization Newton method of Nesterov & Polyak (2006) reaches an ϵ-stationary point in $O(\Delta_f \sqrt{\rho}\, \epsilon^{-3/2})$ iterations.

The benefits of higher-order smoothness extend beyond explicit second-order methods. Even with only a first-order oracle, an algorithm can still exploit the slowly varying curvature to improve its dependence on ϵ without computing the full Hessian. Nonconvex accelerated methods (Carmon et al., 2017) obtain an $O(\epsilon^{-7/4} \log(1/\epsilon))$ bound for finding approximate first-order stationary points.

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.

Restarted accelerated gradient descent methods (RAGD) (Li & Lin, 2023) remove the additional polylogarithmic factors. Marumo & Takeda (2024) further achieve $O(\epsilon^{-7/4})$ function and gradient evaluations without prior knowledge of the smoothness constants or target accuracy. Jin et al. (2018) show that perturbed accelerated gradient descent attains approximate second-order stationarity with high probability in $\widetilde{O}(\epsilon^{-7/4})$ iterations. More recently, Zhou (2026) proved a matching lower bound $\Omega(\epsilon^{-7/4})$ for functions with Lipschitz gradients and Hessians using deterministic first-order methods.

These developments show that higher-order smoothness can support acceleration even without direct access to the Hessian. However, the methods above still require gradients or other numerical information. More recently, optimization and learning with limited feedback have attracted increasing attention. For example, methods using gradient signs admit convergence guarantees and have been applied to machine learning tasks (Bernstein et al., 2018; Li et al., 2023; Liu et al., 2019). Meanwhile, reinforcement learning from human feedback (RLHF) (Christiano et al., 2017; Fan et al., 2025; Gaur et al., 2025), which only compares and determines which is better using human preferences, has played an important role in training and aligning large language models (LLMs) (Ziegler et al., 2019; Stiennon et al., 2020; Ouyang et al., 2022; Bai et al., 2022). Theoretical work also has established guarantees for learning from preference feedback and related optimization problems (Novoseller et al., 2020; Xu et al., 2020; Chen et al., 2022; Saha et al., 2023; Zhu et al., 2023; Tang et al., 2023). In particular, Wang et al. (2023) show that, under appropriate modeling assumptions, preference-based RL incurs little or no additional cost relative to reward-based RL.

Taken together, the acceleration guarantees and the growing interest in limited-feedback learning raise a natural question: can comparisons retain the benefits of higher-order smoothness for accelerated nonconvex optimization? We consider access to $f\colon \mathbb{R}^n \to \mathbb{R}$ through a comparison oracle:

$$O_f^{\text{comp}}(x,y) = \begin{cases} 1, & f(x) \geq f(y) \\ -1, & f(x) \leq f(y) \end{cases} \tag{4}$$

with either output allowed when $f(x) = f(y)$.

Optimization using the comparisons of queried points has a long history. Classical direct-search methods accept or reject trial steps by such comparisons (Kolda et al., 2003; Audet & Dennis Jr, 2006). The Nelder–Mead method, for example, updates a simplex through comparisons among candidate function values (Nelder & Mead, 1965). However, this method does not guarantee convergence to stationary points even on continuously differentiable convex examples (McKinnon, 1998). Subsequent work has developed complexity guarantees for a range of query models and objective settings. Jamieson et al. (2012) study derivative-free query complexity, using Boolean comparison feedback and comparison-based line search. Gradientless Descent (Golovin et al., 2020) analyzes derivative-free optimization exploiting invariance to monotone transformations. Saha et al. (2024) study batched and multiway preference feedback. More recent work extends dueling optimization to Riemannian manifolds (Ren et al., 2026) or formulates optimization directly through preference relations without assuming an underlying scalar objective (Scheinberg & Xiong, 2026).

For convex objectives, several methods provide convergence guarantees under comparison or preference feedback. Karabag et al. (2021) develop ellipsoid-based methods for smooth convex optimization using comparison or directional-preference information. The stochastic three points method (STP) of Bergou et al. (2020), which also applies beyond convex objectives, samples a direction s_k and selects the point with the smallest objective value among x_k and $x_k \pm \alpha_k s_k$. Gorbunov et al. (2020) incorporate momentum through the stochastic momentum three points method (SMTP). In the order-oracle framework, Lobanov et al. (2024) establish acceleration guarantees under strong convexity. Dueling convex optimization studies convex minimization using pairwise preference feedback (Saha et al., 2021; 2025), while Tang et al. (2023) construct stochastic gradient estimators from rankings.

For nonconvex objectives with Lipschitz gradients, STP yields an $O(n\epsilon^{-2})$ comparison bound for finding an approximate stationary point in expectation (Bergou et al., 2020). Under Lipschitz gradients and Hessians, Wang et al. (2026) show that comparisons can support second-order geometric recovery and give an algorithm that visits an ϵ-first-order stationary point with constant success probability using $\widetilde{O}\big(n^2\Delta_f\sqrt{\rho}\,\epsilon^{-3/2}\big)$ comparison queries. Their result achieves the precision dependence of second-order nonconvex optimization up to logarithmic factors, but recovering the Hessian incurs a quadratic dependence on n. A natural next question is whether comparison feedback can also support the acceleration mechanisms of nonconvex first-order methods:

Can comparison-based methods match the $\epsilon^{-7/4}$ dependence of accelerated first-order methods while achieving nearly linear query complexity in the dimension?

Results. In this paper, we answer this question in the affirmative up to poly-logarithmic factors.

Theorem 1 (Main result, informal)**.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be bounded below by $f^\star$, L-smooth, and ρ-Hessian-Lipschitz. For any $\epsilon \in (0,1)$ and $\delta \in (0,1)$, there exists an algorithm* `Comparison-RAGD` *which, with probability at least $1-\delta$, returns a list of points containing an $\widehat{x}$ satisfying*

$$\|\nabla f(\widehat{x})\| \le \epsilon,$$

using $\widetilde{O}\left(\frac{n\sqrt{L}\rho^{1/4}\left(f(x_0)-f^\star\right)}{\epsilon^{7/4}}\right)$ comparison queries.[1]

To the best of our knowledge, this is the first comparison-based method to attain the accelerated $\epsilon^{-7/4}$ dependence for first-order stationarity in general smooth nonconvex optimization with nearly linear dependence on the dimension. Our result matches the dependence on dimension and accuracy achieved by accelerated zeroth-order methods (Zhang & Gu, 2022), despite using only comparisons and receiving no quantitative information such as function values or derivatives. This establishes, for the first time, that comparison feedback alone suffices to achieve momentum-based acceleration.

Technically, our algorithm is built upon the restarted accelerated gradient descent (RAGD) (Li & Lin, 2023), using the comparison-based gradient direction estimator of Tao et al. (2026) as a primitive. Assume that the gradient norm is bounded below by given constant (this is guaranteed in our setting, otherwise we have found an approximate stationary point), then this primitive estimates the normalized gradient to accuracy ϵ with constant success probability using $O(n\log(1/\epsilon))$ comparison queries. The main difficulty is that accelerated updates require gradient magnitudes, whereas comparisons generally reveal only directions; moreover, estimation errors accumulate through momentum.

Figure 1 illustrates the framework of our algorithm. We allow $O(1+\Delta_f\sqrt{\rho}/\epsilon^{3/2})$ restart epochs, with iterates remaining within distance $O(\sqrt{\epsilon/\rho})$ of the epoch initial point before the restart. For a fixed anchor x_0, we decompose the gradient of one iterate as

$$\nabla f(y_k) = \|\nabla f(x_0)\| \cdot r_k u_k, \quad r_k = \frac{\|\nabla f(y_k)\|}{\|\nabla f(x_0)\|}, \quad u_k = \frac{\nabla f(y_k)}{\|\nabla f(y_k)\|}.$$

We estimate the direction, anchor scale, and norm ratio separately. A constant-factor anchor-scale guess induces a fixed effective step size, while direction and ratio errors yield relative and additive errors. We control accumulated errors through cumulative squared step length: within $O(\sqrt{L}/(\rho\epsilon)^{1/4})$ steps, an epoch either triggers a restart with $\Omega(\epsilon^{3/2}/\sqrt{\rho})$ decrease or yields an ϵ-stationary point.

To recover the magnitude $\|\nabla f(x_0)\|$, we enumerate a geometric grid at the start of each epoch and run a candidate trajectory for each guess, including one constant-factor approximation to the true norm. The gradient-norm ratios are estimated from directions at an anchor, the current point and their midpoint. The construction exploits the midpoint identity for quadratic gradients, with approximation errors controlled by Hessian Lipschitzness. We further supplement the accelerated trajectories with line search for nearly rank-one Hessians and a normalized gradient step in low-curvature regions. These candidates address regimes where the gradient norm ratio estimator is degenerate.

Finally, we retain the required candidates and select a minimum-value candidate to initialize the next epoch. Each epoch has $\widetilde{O}(\sqrt{L}/(\rho\epsilon)^{1/4})$ steps with $\tilde{O}(n)$ comparisons per step, and decreases the objective by $\Omega(\epsilon^{3/2}/\sqrt{\rho})$ (or retains an ϵ-stationary point already). This yields a total comparison complexity of $\widetilde{O}(n\sqrt{L}\rho^{1/4}/\epsilon^{7/4})$ for finding ϵ-stationary points.

Open questions. On the one hand, it is natural to explore whether our comparison-based method can find an approximate second-order stationary point while retaining the accelerated $\widetilde{O}(n/\epsilon^{7/4})$ rate. On the other hand, the tradeoff between dimension and accuracy, such as between our $\widetilde{O}(n/\epsilon^{7/4})$ result and the $\widetilde{O}(n^2/\epsilon^{3/2})$ result by Wang et al. (2026), is worth investigation.

[1] We use $\widetilde{O}(\cdot)$ to hide poly-logarithmic factors in n, ϵ^{-1}, L, ρ, δ^{-1} and Δ.

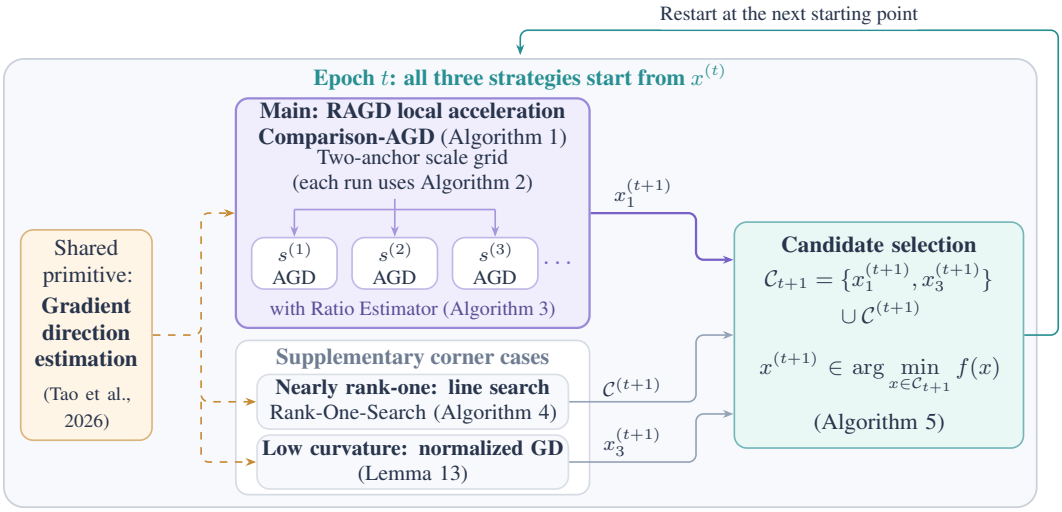

Figure 1: One epoch of our accelerated gradient descent algorithm by comparison.

2 Preliminaries

Basic notation. The quadratic Taylor expansion of f at any point $x_0 \in \mathbb{R}^n$ is

$$q(x) := f(x_0) + \langle g_0,\, x - x_0 \rangle + \frac{1}{2} \langle x - x_0,\, H_0(x - x_0) \rangle .$$

Throughout the paper, for $x \in \mathbb{R}^n$, we denote $g(x) = \nabla f(x)$ and $H(x) = \nabla^2 f(x)$, and omit $(\cdot)$ when x is obvious from the context. For any vector v and matrix A, let $\|v\|$ and $\|A\|$ be their ℓ_2 norms. We denote $\mathbb{B}_R^n(x) := \{y \in \mathbb{R}^n : \|y - x\| \le R\}$ and $\mathbb{S}_R^n(x) := \{y \in \mathbb{R}^n : \|y - x\| = R\}$.

Gradient direction estimation using comparisons. Recently, Tao et al. (2026) developed classical and quantum algorithms for gradient direction estimation using comparisons. These algorithms serve as basic building blocks in our algorithms, stated below:

Theorem 2 (Gradient direction estimation, Theorem 3.3 in Tao et al. 2026). *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ has L_1-Lipschitz gradient. Given query access to a comparison oracle $\mathcal{O}_f^{comp}$ (Equation 4) and $x \in \mathbb{R}^n$ with $\|\nabla f(x)\| \ge \gamma$, for any precision ϵ, there exists an algorithm $\texttt{ComparisonGE}(x, \epsilon, \gamma)$ outputs a unit vector $\widehat{g}$ such that $\left\| \widehat{g} - \frac{\nabla f(x)}{\|\nabla f(x)\|} \right\| \le \epsilon$ with success probability at least $2/3$, using $O(n \log(1/\epsilon))$ queries.*

3 Framework of Local Acceleration of Gradient Descent with Comparison

In this section, we first formally give a structure of one single epoch of Comparison-RAGD algorithm (Algorithm 1), where we use NormEstimator (Definition 1) to roughly guess the norm of gradient norm of initial points. Every guess of norm pairs corresponds to an RAGD epoch, giving an RAGD trajectory, and at last we choose the best one. As a subroutine of Algorithm 1, the structure of RAGD iteration is given by Algorithm 2, where we use RatioEstimator (Definition 2) combining the guess of initial norm pairs to compute a guess of norm of the current gradient. Here we give definitions of NormEstimator and RatioEstimator, and we will give instances of the two definitions in Section 3.3.

Definition 1. *We call a procedure a $(G_{max}, G_{min}, \delta_s)$-NormEstimator, if for any inputs x, it outputs a set $\mathcal{S}$ such that $\exists s \in \mathcal{S}$ satisfying $\left| \frac{s}{\|\nabla f(x)\|} - 1 \right| \le \delta_s$, and all the elements $s \in \mathcal{S}$ satisfy $G_{min} \le s \le G_{max}$.*

Definition 2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic function satisfying $\|H(x)\| \le 1$ and $\sigma_2(H(x)) \ge \delta > 0$ for any $x \in \mathbb{R}^n$, and fix $\epsilon \in (0, 1]$. We call a procedure an ϵ-RatioEstimator if for any inputs $x, y, z \in \mathbb{R}^n$, it outputs two ratios r_1 and r_2 satisfying $\left| r_1 - \frac{\|\nabla f(z)\|}{\|\nabla f(x)\|} \right| < \epsilon$ and $\left| r_2 - \frac{\|\nabla f(z)\|}{\|\nabla f(y)\|} \right| < \epsilon$.*

3.1 ALGORITHM STRUCTURE OF RAGD

In this subsection, we give Algorithm 1 as a general epoch structure to find an ϵ-stationary point or give $\Omega(\epsilon^{3/2})$ decrease to function value. By the standard AGD iterate formula in Line 6 of Algorithm 2, we require information about $\nabla f(y_k)$. Since in the comparison oracle model, the gradient norm $\|\nabla f(y_k)\|$ cannot be directly estimated from comparison queries, even when a valid smoothness upper bound L is known, we can only estimate the direction of the gradient $\nabla f(y_k)/\|\nabla f(y_k)\|$. However, we can create a geometry grid (see Lemma 1) to guess the norm of the gradient, obtaining a multiplicative error between the estimation and the real norm, costing a logarithmic complexity.

Algorithm 1: A Restart Step of Comparison Restarted Accelerated Gradient Descent `Comparison-AGD`(x_0, ϵ)

Input: Starting point x_0, accuracy ϵ, smoothness L, Hessian-Lipschitz constant ρ.
Output: A candidate point set $\mathcal{X}_{\text{out}}$.

1 Set $\eta \leftarrow \frac{1}{4L}, R \leftarrow \frac{1}{16}\sqrt{\frac{\epsilon}{\rho}}, \theta \leftarrow (4\epsilon\rho\eta^2)^{1/4}, \beta \leftarrow 1-\theta, K \leftarrow \left\lceil \frac{1}{\theta} \right\rceil, S_0 \leftarrow 0$, $\delta_g \leftarrow \frac{b\epsilon^2}{200(2+L/\epsilon+\epsilon/L)}$, $\mathcal{S} \leftarrow (2L, \epsilon, 1/2)$-NormEstimator.
2 Sample $v \sim \text{Unif}(\mathbb{S}^{n-1})$ and set $x_1 \leftarrow x_0 + Rv$.
3 Set $\widehat{g}_0 \leftarrow \texttt{ComparisonGE}(x_0, \delta_g, \epsilon), \widehat{g}_1 \leftarrow \texttt{ComparisonGE}(x_1, \delta_g, \epsilon)$.
4 **for each** $s = (s_0, s_1) \in \mathcal{S}^2$ **do**
5 $\quad \{x_0, x_1, \ldots, x_{k+1}\} =$`AGD-Tracking`$(x_0, s, \epsilon)$.
6 $\quad m = \text{argmin}_{\lfloor K/2 \rfloor \le k \le K-1} \left\| x_{k+1}^{(s)} - x_k^{(s)} \right\|$.
7 $\quad \widehat{y}^{(s)} = \frac{1}{m+1}\sum_{k=0}^{m} y_k^{(s)}$.
8 **return** $\mathcal{X}_{\text{out}} = \{\widehat{y}^{(s)}, x_K^{(s)}\}$.

Next, we give Algorithm 2 to track the trajectory of one single epoch in Algorithm 1, given a specific guess of initial gradient norm pair. To circumvent guessing the gradient norm in each loop and generating exponentially many trajectories, we estimate the gradient norm ratio between initial point and current points. Combining this ratio with the initial norm estimation, we estimate the current gradient norm, which is then used to perform the iterative update. At each iteration, we prune the candidate trajectories by retaining the one that achieves the greatest decrease in the objective value, thereby preserving the algorithm's performance guarantees.

Algorithm 2: AGD Iterate Formula `AGD-Tracking`(x_0, s, ϵ)

Input: Starting point x_0, accuracy ϵ, guessed norm pair s.

1 Set $x_{-1}^{(s)} \leftarrow x_0, x_0^{(s)} \leftarrow x_0, k \leftarrow 0, \delta_g \leftarrow \frac{b\epsilon^2}{200(2+L/\epsilon+\epsilon/L)}$.
2 **while** $k < K$ **do**
3 $\quad y_k^{(s)} \leftarrow x_k^{(s)} + \beta\left(x_k^{(s)} - x_{k-1}^{(s)}\right)$.
4 $\quad \widehat{g}_k^{(s)} \leftarrow \texttt{ComparisonGE}\left(y_k^{(s)}, \delta_g, \epsilon\right)$.
5 $\quad (\widehat{r}_{k,0}^{(s)}, \widehat{r}_{k,1}^{(s)}) \leftarrow \texttt{ComparisonRatioEstimator}\left(x_0, y_k^{(s)}, R, \frac{\theta}{400}\right)$.
6 $\quad x_{k+1}^{(s)} \leftarrow \arg\min\{f(y_k^{(s)} - \eta s_0 \widehat{r}_{k,0}^{(s)} \widehat{g}_k^{(s)}), f(y_k^{(s)} - \eta s_1 \widehat{r}_{k,1}^{(s)} \widehat{g}_k^{(s)})\}$.
7 $\quad S_{k+1} = S_k + \|x_{k+1}^{(s)} - x_k^{(s)}\|^2$
8 $\quad$ **if** $(k+1)S_{k+1} > B^2$ **then**
9 $\quad\quad$ Set $x_{-1}^{(s)} \leftarrow x_{k+1}^{(s)}, x_0^{(s)} \leftarrow x_{k+1}^{(s)}, k \leftarrow 0$.
10 $\quad\quad$ **return** $\{x_0^{(s)}, x_1^{(s)}, \ldots, x_{k+1}^{(s)}\}$
11 $\quad k \leftarrow k+1$.

3.2 Stationarity or Decreasing guarantee of Comparison RAGD

In this subsection, we give Theorem 5 as a performance guarantee of our comparison RAGD algorithm, which claims that Algorithm 1 outputs a set of points containing an ϵ-stationary point or a point gives $\Omega(\epsilon^{3/2})$ function value decrease whenever the setting of Definition 1 ($\sigma_2(H) \geq \delta > 0$) is satisfied. To facilitate the error analysis, we first specify the parameter setting and introduce an assumption to bound the error of NormEstimator and RatioEstimator. In Section 3.3, we show how to choose the parameters to satisfy this assumption.

Parameter setting. Fix $\delta_s = 1/2$. Assume $L, \rho > 0$, f has L-Lipschitz gradient and ρ-Lipschitz Hessian, and choose

$$\eta = \frac{1}{4L}, \quad \theta = \sqrt{2}(\eta\sqrt{\rho\epsilon})^{1/2} \leq \frac{1}{2}, \quad \beta = 1-\theta, \quad K = \lceil 1/\theta \rceil, \quad B = b\sqrt{\epsilon/\rho}. \tag{5}$$

Assumption 1. *Suppose $x_{-1} = x_0$, $y_k = x_k + \beta(x_k - x_{k-1})$, $0 \leq \beta \leq 1$, and $x_{k+1} = y_k - h(g_k + w_k)$, where $g_k = \nabla f(y_k)$, the perturbed stepsize $h \in [\eta, 3\eta/2]$ and*

$$\|w_k\| \leq \nu\|g_k\| + a_k, \qquad 0 \leq \nu < \theta/100, \quad 0 \leq a_k \leq b\epsilon/1000.$$

We call Line 8 as a trigger. Note that Algorithm 2 stops under two cases (whether Line 8 triggers or not), we give two conclusions Theorem 3 (proof deferred to Appendix B.2) and Theorem 4 (proof deferred to Appendix B.3) to claim the ability of restart epoch respectively. If the trigger occurs, then we have passed enough cumulative stepsize, thus we can get a decrease guarantee in Theorem 4. Else the trigger does not occur, which means that momentum in AGD steps are relatively small, we have reached in a neighborhood of a stationary point, where the precision is guaranteed by Theorem 3.

Theorem 3 (Complete epoch stationarity)**.** *Under the setting of Equation 5 and Assumption 1, if no trigger occurs through K, select $m = \operatorname{argmin}_{\lfloor K/2 \rfloor + 1 \leq j \leq K} \|x_j - x_{j-1}\|$, $\widehat{y} = \frac{1}{m+1}\sum_{k=0}^{m} y_k$. Then*

$$\|\nabla f(\widehat{y})\| \leq \frac{2\sqrt{2}B}{hK^2} + \frac{2\theta B}{hK} + 8\rho B^2 + \frac{2\sqrt{2}\nu B}{h(1-\nu)K} + \frac{\sum_{k<m} a_k}{m(1-\nu)}.$$

Theorem 4 (Restart decrease)**.** *Under the setting of Equation 5 and Assumption 1, then at the first trigger,*

$$\begin{aligned} f(x_T) - f(x_0) &\leq -\frac{\theta}{h}\left(\frac{1}{8} - \frac{64\nu^2}{\theta^2(1-\nu)^2}\right) S_T + \frac{16h}{\theta(1-\nu)^2}\sum_{k<T} a_k^2 + \frac{32hK}{\theta}\rho^2 B^4 + 24\rho B^3 \\ &\leq -\frac{b^2}{96}\frac{\epsilon^{3/2}}{\sqrt{\rho}}. \end{aligned} \tag{6}$$

Theorem 5 combines two theorems above and gives that whenever Algorithm 1 terminates, it outputs a set of points containing an ϵ-stationary point or a point gives $\Omega(\epsilon^{3/2})$ function value decrease.

Theorem 5. *Under Equation 5, when an epoch stops at the first $TS_T > B^2$, or after K updates, if the first event occurs, the retained endpoint decreases f by at least $b^2\epsilon^{3/2}/(96\sqrt{\rho})$. Otherwise the retained average in Theorem 3 is ϵ-stationary. At most K updates are executed.*

Proof. If the trigger occurs, Theorem 4 guarantees the decrease Equation 6. If the trigger does not occur, then $jS_j \leq B^2$ for all $j \leq K$; Theorem 3 applies and gives the stationarity conclusion. These events exhaust the algorithm's two stopping rules. Finally $K \leq 1/\theta + 1 \leq 2/\theta$, and $1/\theta = \sqrt{6L}/(\rho\epsilon)^{1/4}$, proving the update count. □

3.3 Instances of Techniques Used in RAGD Structure

In this subsection, we introduce Lemma 1 as an instance of Definition 1 and Algorithm 3 as an instance of Definition 2 . We first use a logarithmic geometric grid of possible step lengths, to obtain only a constant multiplicative factor estimation of the gradient norm. We will prove the result later that RAGD is robust to constant-factor step-size perturbations.

Lemma 1 (Geometric grid partition of gradient-scale). *For $0 < G_{\min} \le G_{\max}$ and $0 < \delta_s < 1$, set*

$$m = \left\lceil \frac{\log(G_{\max}/G_{\min})}{\log(1+\delta_s)} \right\rceil, \qquad \mathcal{S} = \{G_{\min}(1+\delta_s)^j : 0 \le j \le m\}.$$

For each $G \in [G_{\min}, G_{\max}]$, some $s \in \mathcal{S}$ satisfies $1 \le s/G \le 1+\delta_s$. Also

$$|\mathcal{S}| = O\left(1 + \delta_s^{-1} \log(G_{\max}/G_{\min})\right).$$

Proof. Choose $j = \lceil \log(G/G_{\min})/\log(1+\delta_s) \rceil$. Then $0 \le j \le m$, and the corresponding grid value has the claimed ratio. The cardinality follows from $\log(1+\delta_s) \ge \delta_s/(1+\delta_s)$. □

Corollary 1. *Take $G_{\min} = \epsilon$, $G_{\max} = 2L$, and $\delta_s = 1/2$. If $G_0 < \epsilon$, then x_0 is already an ϵ-stationary point. Else, the partition in Lemma 1 is a $(2L, \epsilon, 1/2)$-NormEstimator per Definition 1.*

We construct a triangle to compute the gradient norm ratio in Algorithm 3, and give an error analysis in Lemma 11. Note that if the direction of two gradients are relatively close, the denominators in Line 3 would tend to 0, which may bring unbounded error. We deal with this case by choosing two initial points with separated gradient directions, thus we can guarantee that at least one of the initial point gradient is separated from the current point gradient, and choosing the better one gives a better decrease than the exact AGD update. By Lemma 10 (proof given in Appendix C.1), through randomly sampling two points in a local field, we can find two separated directions of gradients.

Algorithm 3: Two-Anchor Ratio Estimator `ComparisonRatioEstimator`(x_0, x_k R, ϵ)

Input: Initial point x_0; Randomly chosen point x_1; Current point x_k; restart radius R; smoothness constant L; momentum parameter β; target accuracy ϵ.

1 Set $\delta_g \leftarrow \frac{b\epsilon^2}{200(2+L/\epsilon+\epsilon/L)}$.

2 Set $\hat{g}_0 \leftarrow$ `ComparisonGE`$(x_0, \delta_g, \epsilon)$,$\hat{g}_1 \leftarrow$ `ComparisonGE`$(x_1, \delta_g, \epsilon)$, $\hat{g}_k \leftarrow$ `ComparisonGE`$(x_k, \delta_g, \epsilon)$, $\hat{g}_{k,0} \leftarrow$ `ComparisonGE`$(\frac{x_0+x_k}{2}, \delta_g, \epsilon)$, $\hat{g}_{k,1} \leftarrow$ `ComparisonGE`$(\frac{x_1+x_k}{2}, \delta_g, \epsilon)$;

3 $\hat{r}_{k,0} \leftarrow \frac{\sin\angle(\hat{g}_0,\hat{g}_{k,0})}{\sin\angle(\hat{g}_k,\hat{g}_{k,0})}$, $\hat{r}_{k,1} \leftarrow \frac{\sin\angle(\hat{g}_1,\hat{g}_{k,1})}{\sin\angle(\hat{g}_k,\hat{g}_{k,1})}$.

4 **return** Two ratios $\hat{r}_{k,0}$ and $\hat{r}_{k,1}$.

Corollary 2 (Corollary of Lemma 11). *Algorithm 3 is a $\frac{2t(1+r)^2}{rs_*}$-RatioEstimator per Definition 2.*

The proof is deferred to Appendix C.2.

4 Restarted Accelerated Gradient Descent by Comparison

In this section, we combine the local Comparison-RAGD epoch developed in Section 3 with a corner case handling algorithm (Algorithm 4) and a normalized-gradient fallback to obtain a global optimization algorithm. Given the local estimation and trajectory arguments introduced in Section 3, we establish the descent mechanism and the overall complexity here.

4.1 Hessian Nearly Rank-One Regime

In this subsection, we handle a corner case that the Hessian at initial epoch point is nearly rank-one (i.e. the setting of Definition 2 is not satisfied). Throughout this subsection, we work inside the local sphere $\mathbb{B}_R(x_0)$, where $R = \sqrt{\epsilon/\rho}/16$. The separation argument in Lemma 10 requires $\sigma_2(H)$ to be bounded away from zero. When $\sigma_2(H)$ is small, the gradient directions generated by different random perturbations may fail to be separated. In this case, rather than estimating gradient-norm ratios, we exploit the nearly rank-one structure of the Hessian directly, as in the following result:

Lemma 2 (Direction recovery for a nearly rank-one Hessian). *Let $H = \lambda_1 uu^\top + E$, where $\|u\| = 1$, $\|E\| \le \kappa$, and $\lambda := |\lambda_1| > 0$. In Line 4, with probability at least $2/3$, one can construct a unit vector $\widehat{u}$ from $\widehat{g}_0, \widehat{g}_+, \widehat{g}_-$ such that*

$$\min\{\|\widehat{u} - u\|, \|\widehat{u} + u\|\} \le C\left(\delta_g + \frac{8\sqrt{n}\kappa}{\lambda} + \frac{8\sqrt{n}\rho r}{\lambda}\right)$$

Algorithm 4: `Rank-One-Search`(x_0, ϵ)

Input: Current point x_0, accuracy ϵ, Hessian Lipschitz constant ρ, constants $c_R, c_\lambda, c_\kappa, \delta_g$.
Output: A set of point $\mathcal{C}$ such that there exist an ϵ-stationary point or a point x_1 satisfying $f(x_1) < f(x_0)$.

1 Set $R \leftarrow \frac{1}{16}\sqrt{\frac{\epsilon}{\rho}}$, $r \leftarrow \frac{1}{8}R$, $\eta_u = \min\left\{\frac{1}{4}, \frac{\epsilon}{32LR}\right\}$, $\delta_g = \frac{\eta_u}{256(1+Lr/\epsilon)}$.
2 Sample $v \sim \mathrm{Unif}(\mathbb{S}^{n-1})$.
3 $\widehat{g}_0 \leftarrow$ `ComparisonGE`$(x_0, \delta_g, \epsilon)$, $\widehat{g}_+ \leftarrow$ `ComparisonGE`$(x_0 + rv, \delta_g, \epsilon)$, $\widehat{g}_- \leftarrow$ `ComparisonGE`$(x_0 - rv, \delta_g, \epsilon)$.
4 $\widehat{u} = \pm\frac{(\widehat{g}_+ - \widehat{g}_-)}{\|\widehat{g}_+ - \widehat{g}_-\|}$.
5 $x_+ \leftarrow x_0 + R\widehat{u}$, $x_- \leftarrow x_0 - R\widehat{u}$, $\widehat{x} = \arg\min\{f(x_+), f(x_-)\}$.
6 Perform comparison-based one-dimensional line search on $\phi(t) := f(x_0 + t\widehat{u})$, $t \in [-R, R]$.
7 Let $\widehat{t}_u$ be the point returned by the line search and set $x_u \leftarrow x_0 + \widehat{t}_u\widehat{u}$.
8 Set $q \leftarrow (I - \widehat{u}\widehat{u}^\top)\widehat{g}_0$.
9 **if** $\|q\| > 0$ **then**
10 Set $x_w \leftarrow x_0 - R\frac{q}{\|q\|}$.
11 **return** $\mathcal{C} = \{x_0, \widehat{x}, x_u, x_w\}$.

for a universal constant C.

Remark 1. *There is a benign degeneracy which should not be interpreted as a failure of the method. If $g_0 \parallel u$, then $g_0 \pm au \parallel u$. Thus the three endpoint gradient directions are collinear and the rank-one reconstruction is ill-conditioned. However, in this case the normalized gradient direction itself is already the dominant Hessian direction, namely $g_0/\|g_0\| = \pm u$. Therefore the algorithm uses $\widehat{g}_0$ directly instead of reconstructing u.*

After recovering the dominate eigenvector direction of the rank-one Hessian in Lemma 2, we can form the candidate $\widehat{x}$ in Line 5 and do line search in Line 6 of Algorithm 4, combining Lemma 12 gives the function value decrease guarantee as follows:

Theorem 6 (Rank-one candidate guarantee)**.** *Suppose*

$$H_0 = \lambda_1 uu^T + F, \quad \|u\| = 1, \quad \|F\| \le \kappa, \quad \sqrt{\rho\epsilon} \le \lambda = |\lambda_1| \le L, \quad \kappa \le \sqrt{\rho\epsilon}/8.$$

Suppose the recovered direction satisfies $\min\{\|\widehat{u} - u\|, \|\widehat{u} + u\|\} \le \eta_u$, where $\eta_u \le \min\{1/4, \epsilon/(32LR)\}$. Then Algorithm 4 retains an ϵ-stationary point or a point decreasing f by at least

$$\frac{1}{1024}\frac{\epsilon^{3/2}}{\sqrt{\rho}}.$$

Its line search uses $O(1 + \log_+(LR/\epsilon))$ comparisons, where $\log_+ t = \max\{0, \log t\}$.

The proof this theorem is deferred to Appendix D.1.

4.2 Complexity of Finding Stationary Point Guaranty

In this subsection we give a global algorithm (Algorithm 5), which outputs a list of candidates containing at least one ϵ-stationary point. We first give a general guaranty in Theorem 7 that in each epoch (`Comparison-AGD`, `Rank-One-Search` or NGD fallback), whichever the case we can visit an ϵ-stationary point or get an $\Omega(\epsilon^{3/2})$ function value decrease with query complexity at most $\widetilde{O}(1/\epsilon^{1/4})$. Using contradiction we can get the general query complexity result in Theorem 8.

Theorem 7 (Per epoch stationary-or-decrease guarantee)**.** *With success probability at least 2/3, at every epoch the candidate set $\mathcal{C}_t$ in Algorithm 5 contains an ϵ-stationary point or satisfies*

$$\min_{x \in \mathcal{C}_t} f(x) \le f(x^{(t)}) - c_{\mathrm{dec}}\frac{\epsilon^{3/2}}{\sqrt{\rho}}.$$

Algorithm 5: Restarted AGD by Comparison `Comparison-RAGD(`x_0`, `ϵ`)`

Input: Randomly initial point $x^{(0)}$; momentum parameter β; target accuracy ϵ.

for $t = 0, 1, \ldots, N = O(1 + \sqrt{\rho}/\epsilon^{3/2})$ **do**

- $\mathcal{X}_{\text{out}}^{(t+1)} = \texttt{Comparison-AGD}(x^{(t)}, \epsilon)$.
- $\mathcal{C}^{(t+1)} = \texttt{Rank-One-Search}(x^{(t)}, \epsilon)$.
- $\widehat{g}_t = \texttt{ComparisonGE}(x^{(t)})$.
- $x_3^{(t+1)} = x^{(t)} - \frac{1}{L}\widehat{g}_t$.
- $x^{(t+1)} = \arg\min_{x_2^{(t+1)} \in \mathcal{C}^{(t+1)}} \{f(x_1^{(t+1)}), f(x_2^{(t+1)}), f(x_3^{(t+1)})\}$.
- $\mathcal{C}_{t+1} \leftarrow \{x_3^{(t+1)}\} \cup \mathcal{C}^{(t+1)} \cup \mathcal{X}_{\text{out}}^{(t+1)}$

Output: A list $\mathcal{C} \leftarrow \cup_{t=1}^{N} \mathcal{C}_t$.

Proof. If any retained candidate is stationary, we have finished. Otherwise, in particular $G_0 > \epsilon$. If $\|H_0\| \le \sqrt{\rho\epsilon}$, Lemma 13 in Appendix D.2 gives a candidate with decrease at least $\epsilon^{3/2}/(32\sqrt{\rho})$, which is at least the stated decrease by the definition of c_{dec}.

Now, suppose $\|H_0\| > \sqrt{\rho\epsilon}$. Since H_0 is symmetric, choose an orthonormal eigenbasis and an eigenvalue λ_1 with maximal absolute value. Let u be its unit eigenvector and $F = H_0 - \lambda_1 uu^T$. In this basis the eigenvalues of F are zero and the remaining eigenvalues of H_0; thus $\|F\| = \sigma_2(H_0)$, and $\sqrt{\rho\epsilon} < |\lambda_1| \le L$. If $\sigma_2(H_0) \le \sqrt{\rho\epsilon}/8$, Theorem 6 guarantees that its decrease is at least $\epsilon^{3/2}/(1024\sqrt{\rho})$, with success probability at least 2/3. This point belongs to $\mathcal{C}_t$, and its decrease is at least the stated one. If $\sigma_2(H_0) > \sqrt{\rho\epsilon}/8$ instead, Theorem 5 guarantees an AGD candidate with decrease at least $b^2\epsilon^{3/2}/(96\sqrt{\rho})$, since the stationary alternative is excluded. The cases are exhaustive, including equality at each threshold. A minimum over a set is at most the value of the successful candidate, proving the inequality. The total success probability is at least 2/3. □

Theorem 7 can directly lead to the query complexity upper bound of Algorithm 5:

Theorem 8 (Global complexity of Comparison RAGD)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be bounded below by $f^\star$, L-smooth, and ρ-Hessian-Lipschitz. For any $\epsilon \in (0, 1)$ and $\delta \in (0, 1)$, the output of Algorithm 5 contains an ϵ-stationary point with probability at least $1 - \delta$. The total query complexity is*

$$\widetilde{O}\left(\frac{n\sqrt{L}\,\rho^{1/4}\Delta}{\epsilon^{7/4}} + \frac{n\sqrt{L}}{(\rho\epsilon)^{1/4}}\right).$$

Proof. Suppose that the output list $\mathcal{C}$ of Algorithm 5 contains no stationary point. It contains every $\mathcal{C}_t$, so at every epoch, Theorem 7 guarantees the decrease of function value, i.e.

$$f(x^{(t+1)}) - f(x^{(t)}) \le -c_{\text{dec}}\epsilon^{3/2}/\sqrt{\rho}.$$

Summing telescopes to

$$f(x^{(N)}) \le f(x^{(0)}) - Nc_{\text{dec}}\epsilon^{3/2}/\sqrt{\rho} < f(x^{(0)}) - \Delta \le f_*.$$

This contradicts $f \ge f_*$, so $\mathcal{C}$ contains at least one ϵ-stationary point. Probability amplification would lift the success probability to $1 - \delta$ using an extra $O(\log 1/\delta)$ overhead. The cost is

$$\widetilde{O}(NK) = \widetilde{O}\left[\left(1 + \frac{\Delta\sqrt{\rho}}{\epsilon^{3/2}}\right)\frac{n\sqrt{L}}{(\rho\epsilon)^{1/4}}\right] = \widetilde{O}\left(\frac{n\sqrt{L}\,\rho^{1/4}\Delta}{\epsilon^{7/4}} + \frac{n\sqrt{L}}{(\rho\epsilon)^{1/4}}\right).$$

□

AI USE STATEMENT

In this work, we used generative AI tools for language editing. Additionally, we have used generative AI tools for some detailed calculations in appendices, and we have checked them to ensure that there is no logical error. We have not used generative AI tools for proposing the main idea and organizing the structure of the paper. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Charles Audet and John E. Dennis Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 17(1):188–217, 2006.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022. arXiv:2204.05862.

El Houcine Bergou, Eduard Gorbunov, and Peter Richtárik. Stochastic three points method for unconstrained smooth minimization. *SIAM Journal on Optimization*, 30(4):2726–2749, 2020. arXiv:1902.03591.

Jeremy Bernstein, Yu-Xiang Wang, Kamyar Azizzadenesheli, and Animashree Anandkumar. signSGD: Compressed optimisation for non-convex problems. In *International Conference on Machine Learning*, pp. 560–569. PMLR, 2018. arXiv:1802.04434.

Yair Carmon, John C. Duchi, Oliver Hinder, and Aaron Sidford. "Convex Until Proven Guilty": Dimension-free acceleration of gradient descent on non-convex functions. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 654–663. PMLR, 2017. URL https://proceedings.mlr.press/v70/carmon17a.html.

Yair Carmon, John C. Duchi, Oliver Hinder, and Aaron Sidford. Lower bounds for finding stationary points I. *Mathematical Programming*, 184(1):71–120, 2020. arXiv:1710.11606.

Xiaoyu Chen, Han Zhong, Zhuoran Yang, Zhaoran Wang, and Liwei Wang. Human-in-the-loop: Provably efficient preference-based reinforcement learning with general function approximation. In *International Conference on Machine Learning*, pp. 3773–3793. PMLR, 2022. arXiv:2205.11140.

Paul F. Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems*, 30, 2017. arXiv:1706.03741.

Flint Xiaofeng Fan, Cheston Tan, Yew-Soon Ong, Roger Wattenhofer, and Wei-Tsang Ooi. FedRLHF: A convergence-guaranteed federated framework for privacy-preserving and personalized RLHF. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, pp. 713–721, 2025. arXiv:2412.15538.

Mudit Gaur, Utsav Singh, Amrit Singh Bedi, Raghu Pasupathy, and Vaneet Aggarwal. On the sample complexity bounds of bilevel reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 38, pp. 67538–67570. Curran Associates, Inc., 2025. doi: 10.52202/085713-2265. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/617ffb01ea5b57769b0d63d5e9fefd3f-Abstract-Conference.html. arXiv:2503.17644.

Saeed Ghadimi and Guanghui Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013. arXiv:1309.5549.

Daniel Golovin, John Karro, Greg Kochanski, Chansoo Lee, Xingyou Song, and Qiuyi Zhang. Gradientless descent: High-dimensional zeroth-order optimization. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=Skep6TVYDB.

Eduard Gorbunov, Adel Bibi, Ozan Sener, El Houcine Bergou, and Peter Richtarik. A stochastic derivative free optimization method with momentum. In *International Conference on Learning Representations*, 2020. arXiv:1905.13278.

Kevin G. Jamieson, Robert Nowak, and Ben Recht. Query complexity of derivative-free optimization. *Advances in Neural Information Processing Systems*, 25, 2012. arXiv:1209.2434.

Chi Jin, Praneeth Netrapalli, and Michael I. Jordan. Accelerated gradient descent escapes saddle points faster than gradient descent. In *Conference on Learning Theory*, pp. 1042–1085, 2018. arXiv:1711.10456.

Mustafa O. Karabag, Cyrus Neary, and Ufuk Topcu. Smooth convex optimization using sub-zeroth-order oracles. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5):3815–3822, 2021. arXiv:2103.00667.

Tamara G. Kolda, Robert Michael Lewis, and Virginia Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003.

Huan Li and Zhouchen Lin. Restarted nonconvex accelerated gradient descent: No more polylogarithmic factor in the $O(\epsilon^{-7/4})$ complexity. *Journal of Machine Learning Research*, 24(157):1–37, 2023. URL `https://jmlr.org/papers/v24/22-0522.html`.

Xiuxian Li, Kuo-Yi Lin, Li Li, Yiguang Hong, and Jie Chen. On faster convergence of scaled sign gradient descent. *IEEE Transactions on Industrial Informatics*, 2023. arXiv:2109.01806.

Sijia Liu, Pin-Yu Chen, Xiangyi Chen, and Mingyi Hong. signSGD via zeroth-order oracle. In *International Conference on Learning Representations*, 2019.

Aleksandr Lobanov, Alexander Gasnikov, and Andrei Krasnov. Acceleration exists! optimization problems when oracle can only compare objective function values. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Naoki Marumo and Akiko Takeda. Parameter-free accelerated gradient descent for nonconvex minimization. *SIAM Journal on Optimization*, 34(2):2093–2120, 2024. doi: 10.1137/22M1540934. arXiv:2212.06410.

Ken I. M. McKinnon. Convergence of the Nelder–Mead simplex method to a nonstationary point. *SIAM Journal on Optimization*, 9(1):148–158, 1998. doi: 10.1137/S1052623496303482.

Katta G. Murty and Santosh N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987. doi: 10.1007/BF02592948.

John A. Nelder and Roger Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

Yurii Nesterov and Boris T. Polyak. Cubic regularization of Newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006.

Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer, 2006.

Ellen Novoseller, Yibing Wei, Yanan Sui, Yisong Yue, and Joel Burdick. Dueling posterior sampling for preference-based reinforcement learning. In *Conference on Uncertainty in Artificial Intelligence*, pp. 1029–1038. PMLR, 2020. arXiv:1908.01289.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35: 27730–27744, 2022. arXiv:2203.02155.

Yuxuan Ren, Abhishek Roy, and Shiqian Ma. Riemannian dueling optimization. *arXiv preprint arXiv:2603.00023*, 2026. arXiv:2603.00023.

Aadirupa Saha, Tomer Koren, and Yishay Mansour. Dueling convex optimization. In *International Conference on Machine Learning*, pp. 9245–9254. PMLR, 2021.

Aadirupa Saha, Aldo Pacchiano, and Jonathan Lee. Dueling RL: Reinforcement learning with trajectory preferences. In *International Conference on Artificial Intelligence and Statistics*, pp. 6263–6289. PMLR, 2023. arXiv:2111.04850.

Aadirupa Saha, Vitaly Feldman, Yishay Mansour, and Tomer Koren. Faster convergence with multiway preferences. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 433–441. PMLR, 2024.

Aadirupa Saha, Tomer Koren, and Yishay Mansour. Dueling convex optimization with general preferences. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 52552–52564. PMLR, 2025. URL https://proceedings.mlr.press/v267/saha25a.html. arXiv:2210.02562.

Katya Scheinberg and Zikai Xiong. Function-free optimization via comparison oracles. *arXiv preprint arXiv:2604.26867*, 2026. arXiv:2604.26867.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021, 2020. arXiv:2009.01325.

Zhiwei Tang, Dmitry Rybin, and Tsung-Hui Chang. Zeroth-order optimization meets human feedback: Provable learning via ranking oracles, 2023. arXiv:2303.03751.

Xiwen Tao, Chenyi Zhang, Helin Wang, Yexin Zhang, and Tongyang Li. Optimal classical and quantum algorithms for gradient testing and estimation by comparisons, 2026. arXiv:2405.11454v3.

Helin Wang, Chenyi Zhang, Xiwen Tao, Yexin Zhang, and Tongyang Li. Finding stationary points by comparisons, 2026. URL https://arxiv.org/abs/2606.27082v1. arXiv:2606.27082.

Yuanhao Wang, Qinghua Liu, and Chi Jin. Is RLHF more difficult than standard RL? A theoretical perspective. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. arXiv:2306.14111.

Yichong Xu, Ruosong Wang, Lin Yang, Aarti Singh, and Artur Dubrawski. Preference-based reinforcement learning with finite-time guarantees. *Advances in Neural Information Processing Systems*, 33:18784–18794, 2020. arXiv:2006.08910.

Hualin Zhang and Bin Gu. Faster gradient-free methods for escaping saddle points. In *The Eleventh International Conference on Learning Representations*, 2022.

Dongruo Zhou. Sharp first-order lower bounds for higher-order smooth nonconvex optimization, 2026. URL https://arxiv.org/abs/2606.05438. arXiv:2606.05438.

Banghua Zhu, Jiantao Jiao, and Michael Jordan. Principled reinforcement learning with human feedback from pairwise or k-wise comparisons. In *ICLR 2023 Workshop on Mathematical and Empirical Understanding of Foundation Models*, 2023. arXiv:2301.11270.

Daniel M. Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B. Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*, 2019. arXiv:1909.08593.

A Auxiliary Lemmas

In this appendix, we collect all auxiliary lemmas needed for our proofs.

Lemma 3 (Normalization and sine perturbations)**.** *For nonzero vectors p, q,*

$$\left| \frac{p}{\|p\|} - \frac{q}{\|q\|} \right| \leq \frac{2\|p-q\|}{\|q\|}.$$

For unit vectors $u, v, \widehat{u}, \widehat{v}$,

$$|\sin\angle(u,v) - \sin\angle(\widehat{u},\widehat{v})| \leq \|u-\widehat{u}\| + \|v-\widehat{v}\|.$$

Proof. Add and subtract $p/\|q\|$. The triangle inequality and $|\|p\| - \|q\|| \leq \|p-q\|$ give

$$\left| \frac{p}{\|p\|} - \frac{q}{\|q\|} \right| \leq \left|1 - \frac{\|p\|}{\|q\|}\right| + \frac{\|p-q\|}{\|q\|} \leq \frac{2\|p-q\|}{\|q\|}.$$

For fixed unit v, $\sin\angle(u,v) = \left\|(I - vv^T)u\right\|$. The reverse triangle inequality and the fact that an orthogonal projection has norm at most one show that changing u changes this sine by at most $\|u-\widehat{u}\|$. By symmetry, changing v changes it by at most $\|v-\widehat{v}\|$. Adding the two bounds proves the claim. □

Lemma 4 (Taylor bounds of function remainder)**.** *Let q be the quadratic Taylor polynomial of f at x_0. For $z = x_0 + d$,*

$$\|\nabla f(z) - \nabla q(z)\| \leq \frac{\rho}{2}\|d^2\|, \qquad |f(z) - q(z)| \leq \frac{\rho}{6}\|d^3\|, \qquad \left\|\nabla^2 f(z) - \nabla^2 f(x_0)\right\| \leq \rho\|d\|.$$

Proof. Write $H_0 = \nabla^2 f(x_0)$. The fundamental theorem of calculus gives

$$\nabla f(z) - \nabla q(z) = \int_0^1 (\nabla^2 f(x_0 + td) - H_0) d \, \mathrm{d}t.$$

Its norm is at most $\rho\|d\|^2 \int_0^1 t \mathrm{d}t$. Applying the fundamental theorem twice to the scalar function $t \mapsto f(x_0 + td)$ gives

$$f(z) - q(z) = \int_0^1 (1-t) d^\top (\nabla^2 f(x_0 + td) - H_0) d \, \mathrm{d}t.$$

The absolute value is at most $\rho\|d\|^3 \int_0^1 t(1-t)\,\mathrm{d}t = \rho\|d\|^3/6$. The Hessian bound is its Lipschitz property. □

Lemma 5 (Localization of the momentum points)**.** *If $jS_j \leq B^2$ for $1 \leq j \leq k$, then*

$$\|x_j - x_0\| \leq B, \qquad \|y_j - x_0\| \leq 2B \qquad (0 \leq j \leq k).$$

At a first trigger T, these bounds hold through $T-1$, not necessarily at x_T.

Proof. Cauchy–Schwarz gives

$$\|x_j - x_0\|^2 \leq jS_j \leq B^2.$$

Also $\|x_j - x_{j-1}\| \leq B$, and $y_j = x_j + (1-\theta)(x_j - x_{j-1})$. □

Lemma 6 (Probability of obtaining a nondegenerate perturbation)**.** *Let $u \in \mathbb{S}^{n-1}$ and let v be uniformly distributed on $\mathbb{S}^{n-1}$. Then for every $0 < \tau \leq 1$, $\Pr[|\langle u, v\rangle| \leq \tau] \leq C\tau\sqrt{n}$. Consequently, with probability at least $1 - C\tau\sqrt{n}$, $|\langle u, v\rangle| \geq \tau$.*

Proof. By rotational invariance, we may assume $u = w_0$. If $v = g/\|g\|$ with $g \sim N(0, I_n)$, then $\langle u, v\rangle = g_1/\|g\|$. Since $\|g\| \leq 2\sqrt{n}$ with constant probability and the density of g_1 is uniformly bounded, the probability that $|g_1|/\|g\| \leq \tau$ is bounded by $C\tau\sqrt{n}$. The stated inequality follows. □

B Proof Details of Local AGD under Inexact Gradients

In this appendix, we analyze local accelerated gradient descent under the error assumption. We bound the accumulated errors and prove the stationarity and function-decrease guatantees for the two possible outcomes of an epoch.

B.1 Accumulated Error Bounds

In this subsection, we bound the accumulated gradient errors. These bounds are used in the following proofs of stationarity (Appendix B.2) and function-decrease (Appendix B.3).

Lemma 7 (Mixed errors controlled by steps)**.** *Under Assumption 1, for every $m \geq 1$,*

$$\sum_{k<m} \|w_k\|^2 \leq \frac{8\nu^2}{h^2(1-\nu)^2} S_m + \frac{2}{(1-\nu)^2} \sum_{k<m} a_k^2,$$

$$\frac{1}{m} \sum_{k<m} \|w_k\| \leq \frac{2\nu}{h(1-\nu)} \sqrt{\frac{S_m}{m}} + \frac{1}{m(1-\nu)} \sum_{k<m} a_k.$$

Proof of Lemma 7. Write $p_k = (x_{k+1} - x_k) - \beta(x_k - x_{k-1}) = -h(g_k + w_k)$. Since $\|g_k\| \leq \|p_k\|/h + \|w_k\|$,

$$\|w_k\| \leq \frac{\nu}{h(1-\nu)} \|p_k\| + \frac{a_k}{1-\nu}.$$

Moreover

$$\sum_{k<m} \|p_k\|^2 \leq 2 \sum_{k<m} \|x_{k+1} - x_k\|^2 + 2\beta^2 \sum_{k<m} \|x_k - x_{k-1}\|^2 \leq 4S_m.$$

Squaring with $(u+v)^2 \leq 2u^2 + 2v^2$ gives

$$\sum_{k<m} \|w_k\|^2 \leq \frac{2\nu^2}{h^2(1-\nu)^2} \sum_{k<m} \|p_k\|^2 + \frac{2}{(1-\nu)^2} \sum_{k<m} a_k^2,$$

and substituting $\sum \|p_k\|^2 \leq 4S_m$ proves the first bound. Summing the unsquared inequality gives

$$\frac{1}{m} \sum_{k<m} \|w_k\| \leq \frac{\nu}{h(1-\nu)m} \sum_{k<m} \|p_k\| + \frac{\sum_{k<m} a_k}{m(1-\nu)}.$$

Cauchy–Schwarz gives $\sum_{k<m} \|p_k\| \leq \sqrt{m \sum_{k<m} \|p_k\|^2} \leq 2\sqrt{mS_m}$. Substitution proves the second bound. □

B.2 Stationarity after an Epoch without Restart

In this subsection, we prove that if an epoch completes all K updates without triggering a restart, the retained average is an ϵ-stationary point under the stated parameter and error conditions.

Proof of Theorem 3. No trigger implies $S_K \leq B^2/K$. There are $K - \lfloor K/2 \rfloor \geq K/2$ eligible indices, so the minimum squared displacement satisfies

$$\|x_m - x_{m-1}\|^2 \leq \frac{\sum_{j=\lfloor K/2 \rfloor + 1}^{K} \|x_j - x_{j-1}\|^2}{K - \lfloor K/2 \rfloor} \leq \frac{2S_K}{K} \leq \frac{2B^2}{K^2}.$$

Also $m \geq K/2$. Since $u_{-1} = u_0$, $\sum_{k<m}(x_k - x_{k-1}) = x_{m-1} - x_0$ and $\sum_{k<m}(x_{k+1} - x_k) = x_m - x_0$. Summing $h\nabla f(y_k) = \beta(x_k - x_{k-1}) - (x_{k+1} - x_k) - hw_k$ therefore gives

$$h \sum_{k<m} \nabla f(y_k) = \theta(x_0 - x_{m-1}) - (x_m - x_{m-1}) - h \sum_{k<m} w_k.$$

Divide by hm, take norms, and use localization to obtain

$$\left\| \frac{1}{m} \sum_{k<m} \nabla f(y_k) \right\| \le \frac{\theta B}{hm} + \frac{\sqrt{2}B}{hKm} + \frac{1}{m} \sum_{k<m} \|w_k\|.$$

Lemma 7, $m \ge K/2$, and $S_m \le S_K \le B^2/K$ bound the last term by

$$\frac{2\nu}{h(1-\nu)} \sqrt{\frac{B^2}{Km}} + \frac{\sum_{k<m} a_k}{m(1-\nu)} \le \frac{2\sqrt{2}\nu B}{h(1-\nu)K} + \frac{\sum_{k<m} a_k}{m(1-\nu)}.$$

The first two terms are at most $2\theta B/(hK)$ and $2\sqrt{2}B/(hK^2)$. Each y_k lies within $2B$ of x_0; their average does as well by convexity of the sphere. Thus $\|y_k - \widehat{y}\| \le 4B$. Lemma 4 applied at $\widehat{y}$ gives

$$\nabla f(y_k) = \nabla f(\widehat{y}) + \nabla^2 f(\widehat{y})(y_k - \widehat{y}) + e'_k, \qquad \|e'_k\| \le 8\rho B^2.$$

Since $\sum_{k<m}(y_k - \widehat{y}) = 0$, averaging and taking norms yield

$$\|\nabla f(\widehat{y})\| \le \left\| \frac{1}{m} \sum_{k<m} \nabla f(y_k) \right\| + 8\rho B^2.$$

Substituting the preceding bounds proves the first conclusion. Finally, $K^{-1} \le \theta$ and $h \ge \eta$ give $B\theta^2/h \le B\sqrt{\rho\epsilon} = b\epsilon$. Thus the first two terms contribute at most $(2\sqrt{2}+2)b\epsilon$, and $8\rho B^2 = 8b^2\epsilon$. Under the concrete accuracy choices, the direction term is at most $4\sqrt{2}b\epsilon/100$, using $\nu \le \theta/100$ and $(1-\nu)^{-1} \le 2$; the additive mean is at most $2b\epsilon/1000$. Adding yields

$$\|\nabla f(\widehat{y})\| \le \left((2\sqrt{2}+2)b + 8b^2 + \frac{4\sqrt{2}b}{100} + \frac{2b}{1000} \right) \epsilon < \epsilon.$$

□

B.3 Function Decrease at a Restart

In this subsection, we prove that the function value decreases sufficiently when the restart condition is met.

Lemma 8 (Quadratic energy inequality with arbitrary errors)**.** *Let $0 < h \le 1/(4L)$, $0 < \theta \le 1$, and $\beta = 1 - \theta$. Suppose $x_{-1} = x_0$ and*

$$y_k = x_k + \beta(x_k - x_{k-1}), \qquad x_{k+1} = y_k - h(\nabla q(y_k) + z_k).$$

Then for every integer $T \ge 1$,

$$q(x_T) - q(x_0) \le -\frac{3\theta}{8h} \sum_{k=0}^{T-1} \|x_{k+1} - x_k\|^2 + \frac{2h}{\theta} \sum_{k=0}^{T-1} \|z_k\|^2. \tag{7}$$

Proof of Lemma 8. Diagonalize H by an orthogonal change of coordinates. It suffices to prove scalar estimates for each coordinate, with $q_j(t) = \lambda t^2/2 + \ell t + c$, and then add them. In the scalar calculation write u_k for the coordinate, and $y_k = u_k + \beta(u_k - u_{k-1})$, and z_k for its scalar perturbation.

First suppose $\lambda \ge -\theta/h$. Exact quadratic expansion, using $\lambda \le L$ at y_k for the new point and $\lambda \ge -\theta/h$ for the old point, gives

$$q_j(u_{k+1}) - q_j(u_k) \le q'_j(y_k)(u_{k+1} - u_k) + \frac{L}{2}((u_{k+1} - u_k) - \beta(u_k - u_{k-1}))^2 + \frac{\theta\beta^2}{2h}(u_k - u_{k-1})^2.$$

Since $q'_j(y_k) = (\beta(u_k - u_{k-1}) - (u_{k+1} - u_k))/h - z_k$, this is

$$-\frac{(u_{k+1} - u_k)^2}{2h} + \frac{(1+\theta)\beta^2}{2h}(u_k - u_{k-1})^2 - \frac{1 - hL}{2h}(u_{k+1} - u_k - \beta(u_k - u_{k-1}))^2 - z_k(u_{k+1} - u_k).$$

Discard the nonpositive square and use

$$|z_k(u_{k+1} - u_k)| \le \frac{\theta}{8h}(u_{k+1} - u_k)^2 + \frac{2h}{\theta} z_k^2.$$

Set $a = (1+\theta)\beta^2$ and $F_k = q_j(u_k) + a(u_k - u_{k-1})^2/(2h)$. Then

$$F_{k+1} - F_k \le -\frac{1}{h}\left(\frac{1}{2} - \frac{\theta}{8} - \frac{a}{2}\right)(u_{k+1} - u_k)^2 + \frac{2h}{\theta} z_k^2.$$

Because

$$\frac{1}{2} - \frac{\theta}{8} - \frac{(1+\theta)(1-\theta)^2}{2} = \frac{3\theta}{8} + \frac{\theta^2(1-\theta)}{2} \ge \frac{3\theta}{8},$$

summing and using $u_{-1} = u_0$ and $a \ge 0$ proves the desired scalar bound.

Now suppose $\lambda < -\theta/h$. Write $q_j(u) = \lambda(u - u_*)^2/2 + c_*$, and let $t_k = u_k - u_*$. The recurrence becomes

$$(u_{k+1} - u_k) = \beta(u_k - u_{k-1}) - h\lambda(t_k + \beta(u_k - u_{k-1})) - hz_k.$$

Define $A_k = \lambda(u_{k+1} - u_k)t_k$. Multiplying the recurrence by λt_k gives

$$A_k = \beta\lambda(u_k - u_{k-1})t_k - h\lambda^2 t_k^2 - h\beta\lambda^2(u_k - u_{k-1})t_k - h\lambda z_k t_k.$$

The two cross terms obey

$$h\beta\lambda^2|(u_k - u_{k-1})t_k| \le \frac{h\lambda^2}{4}t_k^2 + h\beta^2\lambda^2(u_k - u_{k-1})^2, \qquad h|\lambda z_k t_k| \le \frac{h\lambda^2}{4}t_k^2 + hz_k^2.$$

For $k \ge 1$, $t_k = t_{k-1} + (u_k - u_{k-1})$, so

$$A_k \le \beta A_{k-1} + \beta\lambda(1 + h\beta\lambda)(u_k - u_{k-1})^2 - \frac{h\lambda^2}{2}t_k^2 + hz_k^2 \le \beta A_{k-1} + hz_k^2.$$

The last inequality uses $\lambda < 0$ and $1 + h\beta\lambda \ge 1 - hL \ge 3/4$. At $k = 0$, since $v_0 = 0$,

$$A_0 = -h\lambda^2 t_0^2 - h\lambda z_0 t_0 \le \frac{h}{4}z_0^2 \le hz_0^2.$$

Thus $A_k \le h\sum_{i=0}^{k}\beta^{k-i}z_i^2$. Finally,

$$q_j(u_{k+1}) - q_j(u_k) = \frac{\lambda}{2}(u_{k+1} - u_k)^2 + A_k \le -\frac{\theta}{2h}(u_{k+1} - u_k)^2 + A_k.$$

Sum and use $\sum_{j\ge 0}\beta^j = 1/\theta$. This gives

$$q_j(u_T) - q_j(u_0) \le -\frac{\theta}{2h}\sum_{k<T}(u_{k+1} - u_k)^2 + \frac{h}{\theta}\sum_{k<T} z_k^2.$$

Adding the two types of coordinates and weakening to the common constants $3/8$ and 2 proves Equation 7. □

Lemma 9 (A direct smoothness estimate)**.** *Let f be L-smooth, $hL \le 1/4$, $0 \le \beta \le 1$, and $u_{-1} = u_0$. For*

$$y_k = x_k + \beta(x_k - x_{k-1}), \qquad x_{k+1} = y_k - h(\nabla f(y_k) + w_k),$$

one has

$$f(x_T) - f(x_0) \le \frac{1}{4h}\sum_{k=0}^{T-2}\|x_{k+1} - x_k\|^2 - \frac{3}{8h}\|x_T - x_{T-1}\|^2 + 2h\sum_{k<T}\|w_k\|^2. \tag{8}$$

The empty sum is zero when $T = 1$.

Proof of Lemma 9. Using upper smoothness at x_{k+1} and lower smoothness at x_k, both expanded from y_k, imply

$$f(x_{k+1}) - f(x_k) \le \langle\nabla f(y_k), x_{k+1} - x_k\rangle + \frac{L}{2}\|x_{k+1} - y_k\|^2 + \frac{L\beta^2}{2}\|x_k - x_{k-1}\|^2.$$

Substitute $\nabla f(y_k) = -(x_{k+1} - y_k)/h - w_k$, use

$$-\langle x_{k+1} - y_k, x_{k+1} - x_k\rangle = -\tfrac{1}{2}\|x_{k+1} - x_k\|^2 + \tfrac{\beta^2}{2}\|x_k - x_{k-1}\|^2 - \tfrac{1}{2}\|x_{k+1} - y_k\|^2,$$

and bound $|\langle w_k, x_{k+1} - x_k\rangle| \le \|x_{k+1} - x_k\|^2/(8h) + 2h\|w_k\|^2$. Dropping the nonpositive $x_{k+1} - y_k$ term yields

$$f(x_{k+1}) - f(x_k) \le -\frac{3}{8h}\|x_{k+1} - x_k\|^2 + \frac{5}{8h}\|x_k - x_{k-1}\|^2 + 2h\|w_k\|^2.$$

Sum and use $u_{-1} = u_0$. □

The next proposition combines the two preceding lemmas to bound the function decrease over an epoch with a restart.

Proposition 1 (Perturbed restart energy estimate)**.** *Let f have L-Lipschitz gradient and ρ-Lipschitz Hessian, where $L, \rho > 0$. Under the setting of Equation 5, for perturbed stepsize $h \in [\eta, 3\eta/2]$, initialize $x_{-1} = x_0$, and run*

$$y_k = x_k + (1-\theta)(x_k - x_{k-1}), \qquad x_{k+1} = y_k - h(\nabla f(y_k) + w_k).$$

Let $S_j = \sum_{k<j} \|x_{k+1} - x_k\|^2$, and let $T \le K$ be the first index satisfying $TS_T > B^2$. Define the quadratic Taylor model q at x_0 and

$$r_k = \nabla f(y_k) - \nabla q(y_k), \qquad z_k = r_k + w_k.$$

Then

$$f(x_T) - f(x_0) \le -\frac{\theta}{8h} S_T + \frac{4h}{\theta} \sum_{k<T} \|r_k + w_k\|^2 + 24\rho B^3. \tag{9}$$

Proof. For $j < T$, Cauchy–Schwarz and the absence of earlier triggers give $\|x_j - x_0\| \le B$ and $\|y_j - x_0\| \le 2B$. Therefore $\|r_k\| \le 2\rho B^2$ for $k < T$. Also $S_{T-1} \le B^2$, with $S_0 = 0$.

First suppose $\|x_T - x_{T-1}\| \le 2B$. Then $\|x_T - x_0\| \le 3B$, including when $T = 1$. Lemma 8 applies to the exact representation $x_{k+1} = y_k - h(\nabla q(y_k) + z_k)$. Taylor's theorem at x_T gives

$$f(x_T) - f(x_0) \le -\frac{3\theta}{8h} S_T + \frac{2h}{\theta} \sum_{k<T} \|z_k\|^2 + \frac{9}{2}\rho B^3,$$

which implies the asserted weaker estimate.

Now suppose $D := \|x_T - x_{T-1}\| > 2B$. Lemma 9 gives

$$f(x_T) - f(x_0) \le -\frac{5}{16h} D^2 + 2h \sum_{k<T} \|w_k\|^2,$$

since $S_{T-1} \le B^2 < D^2/4$. Furthermore $S_T < 5D^2/4$ and $\theta \le 1/2$, so

$$-\frac{5}{16h} D^2 \le -\frac{\theta}{8h} S_T.$$

Since $w_k = z_k - r_k$,

$$2h \sum_{k<T} \|w_k\|^2 \le 4h \sum_{k<T} \|z_k\|^2 + 4h \sum_{k<T} \|r_k\|^2 \le \frac{4h}{\theta} \sum_{k<T} \|z_k\|^2 + 16hT\rho^2 B^4.$$

Finally $T \le K \le 2/\theta$, $h/\eta \le 3/2$, and $\rho B = b\sqrt{\rho\epsilon} = b\theta^2/\eta$, so

$$16hT\rho^2 B^4 = \rho B^3 (16hT\rho B) \le 48 b\theta\, \rho B^3 \le 24\rho B^3.$$

This proves the estimate without a localization assumption on the triggering endpoint. □

Proof of Theorem 4. Localization and Lemma 4 give $\|\tau_k\| \le 2\rho B^2$ for $k < T$. Therefore

$$\sum_{k<T} \|\tau_k + w_k\|^2 \le 8T\rho^2 B^4 + 2 \sum_{k<T} \|w_k\|^2 \le 8K\rho^2 B^4 + \frac{16\nu^2}{h^2(1-\nu)^2} S_T + \frac{4}{(1-\nu)^2} \sum_{k<T} a_k^2.$$

Multiplying by $4h/\theta$, the error coefficient in Proposition 1, gives respectively

$$\frac{32hK}{\theta}\rho^2 B^4, \qquad \frac{64\nu^2}{h\theta(1-\nu)^2} S_T, \qquad \frac{16h}{\theta(1-\nu)^2} \sum_{k<T} a_k^2.$$

Adding its negative term $-\theta S_T/(8h)$ and remainder $24\rho B^3$, and factoring $-\theta S_T/h$ from the two terms in S_T, proves the first inequality.

For the concrete choices, $\nu \le 1/2$ and

$$\frac{64\nu^2}{\theta^2(1-\nu)^2} \le \frac{256}{10000} < \frac{1}{16}.$$

Put $D = \epsilon^{3/2}/\sqrt{\rho}$. Using $S_T > B^2/K$, $K \le 2/\theta$, and $h \le 3\eta/2$, the negative term is at most

$$-\frac{\theta B^2}{16hK} \le -\frac{\theta^2 B^2}{32h} \le -\frac{b^2}{48}D.$$

The relations $K \le 2/\theta$, $h \le 3\eta/2$, and $\theta^2 = \eta\sqrt{\rho\epsilon}$ imply

$$\frac{hK}{\theta} \le \frac{2h}{\theta^2} \le \frac{3\eta}{\eta\sqrt{\rho\epsilon}} = \frac{3}{\sqrt{\rho\epsilon}}.$$

As $(1-\nu)^{-2} \le 4$, $T \le K$, and $a_k^2 \le b^2\epsilon^2/10^6$,

$$\frac{16h}{\theta(1-\nu)^2}\sum_{k<T} a_k^2 \le \frac{64hK}{\theta}\frac{b^2\epsilon^2}{10^6} \le \frac{192b^2}{10^6}D.$$

Also $\rho^2 B^4 = b^4\epsilon^2$, so $32hK\rho^2B^4/\theta \le 96b^4D$, and $24\rho B^3 = 24b^3D$. These are all the positive terms. Since $192/10^6 + 96b^2 + 24b < 1/96$ at $b = 10^{-4}$, the result follows. □

C Proof Details of Ratio Estimator Techniques

C.1 Sampling Two Separated Gradients

Here, we prove that when the rank of Hessian is larger than one, with high probability we can get two separate gradients:

Lemma 10. *Let $H \in \mathbb{R}^{n\times n}$ be symmetric and satisfy*

$$\|H\| \le 1, \qquad \sigma_2(H) \ge \delta > 0,$$

where $\sigma_2(H)$ denotes the second largest singular value of H. Let x, y be independent uniformly distributed points on $\mathbb{S}^{n-1}$, and let

$$\theta = \angle(Hx, Hy).$$

Then, for every $0 < \eta \le \frac{1}{4}$ we have

$$\Pr\left[\theta \le \eta \text{ or } \theta \ge \pi - \eta\right] \le C\frac{\eta}{\delta},$$

where $C > 0$ is a universal constant.

In particular, for every fixed $a > 1$,

$$\Pr\left[\delta^a \le \theta \le \pi - \delta^a\right] \ge 1 - C\delta^{a-1}.$$

Proof. Let $g, h \sim N(0, I_n)$ be independent. Since $g/\|g\|$ and $h/\|h\|$ are independent and uniform on S^{n-1}, it suffices to consider the angle between Hg and Hh.

Fix $u \in S^{n-1}$. Since the second largest eigenvalue of H^2 is at least δ^2, by the min–max principle there exists a unit vector $q \perp u$ such that $q^\top H^2 q \ge \delta^2$. Let $A = \langle Hg, u\rangle$ and $B = \langle Hg, q\rangle$. Then (A, B) is jointly Gaussian, with $\mathrm{Var}(A) \le 1$ and $\mathrm{Var}(B) \ge \delta^2$.

If $\angle(Hg, u) \le \eta$, then $|B| \le \tan(\eta)|A|$. We claim that for every $t > 0$,

$$\Pr[|B| \le t|A|] \le C\frac{t}{\delta}.$$

To see this, write $B = \sigma X$ and $A = \alpha X + \beta Y$, where $X, Y \sim N(0,1)$ are independent, $\sigma^2 = \mathrm{Var}(B) \ge \delta^2$, and $\alpha^2 + \beta^2 = \mathrm{Var}(A) \le 1$. Let $s = t/\sigma$. If $s \ge 1/2$, the claim is immediate. Otherwise, $|B| \le t|A|$ implies

$$|X| \le s(|X| + |Y|) \le \frac{1}{2}|X| + s|Y|,$$

and hence $|X| \le 2s|Y|$. Since X/Y is standard Cauchy,

$$\Pr[|B| \le t|A|] \le \Pr[|X| \le 2s|Y|] = \frac{2}{\pi}\arctan(2s) \le Cs \le C\frac{t}{\delta}.$$

Taking $t = \tan\eta \le 2\eta$ gives

$$\Pr[\angle(Hg, u) \le \eta] \le C\frac{\eta}{\delta}.$$

Now condition on x. Since $\sigma_2(H) > 0$, $\ker H$ is a proper subspace, so $Hx \neq 0$ almost surely. Applying the above bound with $u = Hx/\|Hx\|$ and with $-u$, respectively, gives

$$\Pr[\theta \le \eta \mid x] \le C\frac{\eta}{\delta}, \qquad \Pr[\theta \ge \pi - \eta \mid x] \le C\frac{\eta}{\delta}.$$

Averaging over x and using the union bound proves

$$\Pr[\theta \le \eta \text{ or } \theta \ge \pi - \eta] \le C\frac{\eta}{\delta}.$$

Finally, if $\delta^a \le 1/4$, taking $\eta = \delta^a$ gives

$$\Pr[\delta^a \le \theta \le \pi - \delta^a] \ge 1 - C\delta^{a-1}.$$

If $\delta^a > 1/4$, then $\delta \le 1$ implies $\delta^{a-1} = \delta^a/\delta > 1/4$. Hence, for a universal constant $C \ge 4$, we have $1 - C\delta^{a-1} \le 0$, and the same bound follows from the nonnegativity of probability. □

C.2 Two-anchor ratio estimator

In this subsection, we analyze gradient-norm ratio estimation using gradient directions at two endpoints and their midpoint. We bound the estimation error under the stated angular nondegeneracy condition.

Lemma 11 (Two-anchor ratio estimator)**.** *Let $a, b \in \mathbb{R}^n$, $m = (a+b)/2$, $g_a = \nabla f(a)$, $g_b = \nabla f(b)$, and $g_m = \nabla f(m)$, all nonzero. Set*

$$r = \frac{\|g_b\|}{\|g_a\|}, \qquad u_a = \frac{g_a}{\|g_a\|}, \qquad u_b = \frac{g_b}{\|g_b\|}.$$

Suppose $\sin\angle(u_a, u_b) \ge s_* > 0$, and unit direction estimates at a, b, m have error at most ω. Define

$$t = 2\omega + \frac{\rho\|a-b\|^2}{\|g_a + g_b\|}, \qquad \widehat{r} = \frac{\sin\angle(\widehat{u}_a, \widehat{u}_m)}{\sin\angle(\widehat{u}_b, \widehat{u}_m)}.$$

If $t \le s_*/(2(1+r))$, the denominator is positive and

$$\left|\frac{\widehat{r}}{r} - 1\right| \le \frac{2t(1+r)^2}{rs_*}. \tag{10}$$

Proof of Lemma 11. Taylor expansion of each endpoint gradient at m cancels the linear terms on averaging, giving $\|g_m - v\| \le E$. By Lemma 3, for nonzero vectors p, q,

$$\left\|\frac{p}{\|p\|} - \frac{q}{\|q\|}\right\| \le \frac{2\|p-q\|}{\|q\|};$$

this follows by adding and subtracting $p/\|q\|$ and using the reverse triangle inequality. Thus the estimated midpoint direction is within $\omega + 2E/M$ of $u_v = v/M$.

For unit vectors, the sine of the angle is Lipschitz in each argument with constant one. Indeed, fixing one argument, it is the norm of the orthogonal projection of the other; symmetry handles the second argument. Consequently, both estimated sines differ from their virtual exact values by at most t.

Write $\phi = \angle(u_a, u_b)$ and $D = \|u_a + ru_b\|$. The exact numerator and denominator are

$$A = \frac{r\sin\phi}{D}, \qquad B_* = \frac{\sin\phi}{D}.$$

These identities follow by orthogonal projection of $u_a + ru_b$. In particular $A = rB_*$ and $B_* \ge s_*/(1+r)$, since $D \le 1 + r$. The assumed bound on t implies $\widehat{B}_* \ge B_*/2$. Hence

$$|\widehat{r} - r| = \frac{|\widehat{A} - r\widehat{B}_*|}{\widehat{B}_*} \le \frac{2t(1+r)}{B_*} \le \frac{2t(1+r)^2}{s_*}.$$

Dividing the inequality above by $r > 0$ gives the original claim. □

D Proof Details of Corner Cases

This appendix collects how we handle the two corner cases. In particular, Appendix D.1 treats a nearly rank-one Hessian component, while Appendix D.2 analyzes normalized-gradient descent in the low-curvature regime.

D.1 Nearly Rank-One Hessians

Proof of Lemma 2. By Lemma 6, with probability at least a constant, $v \in \mathbb{S}^{n-1}$ satisfy $|\langle u, v\rangle| \geq \frac{1}{8\sqrt{n}}$. By Taylor's theorem and the ρ-Lipschitzness of the Hessian,

$$\nabla f(x_0 + h) = g_0 + Hh + R(h),$$

where $\|R(h)\| \leq \rho\|h\|^2/2$. Taking $h = \pm rv$ gives

$$\nabla f(x_\pm) = g_0 \pm rHv + R(\pm rv).$$

Since $H = \lambda_1 uu^\top + E$, we have

$$Hv = \lambda_1\langle u, v\rangle u + Ev.$$

Consequently,

$$\nabla f(x_\pm) = g_0 \pm r\lambda_1\langle u, v\rangle u + e_\pm,$$

where $e_\pm = \pm rEv + R(\pm rv)$. Therefore,

$$\|e_\pm\| \leq r\|E\| + \rho r^2/2 \leq r\kappa + \rho r^2/2.$$

Define $a := r\lambda_1\langle u, v\rangle$. The ideal rank-one model is therefore $\widetilde{g}_0 = g_0$, $\widetilde{g}_+ = g_0 + au$, and $\widetilde{g}_- = g_0 - au$. Let $w_0 := g_0/\|g_0\|$. Write $\widetilde{g}_+ = \alpha_+\widetilde{w}_+$ and $\widetilde{g}_- = \alpha_-\widetilde{w}_-$, where $\widetilde{w}_\pm := \widetilde{g}_\pm/\|\widetilde{g}_\pm\|$ and $\alpha_\pm := \|\widetilde{g}_\pm\|$. Then

$$\alpha_+\widetilde{w}_+ + \alpha_-\widetilde{w}_- = 2\|g_0\|w_0, \qquad \alpha_+\widetilde{w}_+ - \alpha_-\widetilde{w}_- = 2au.$$

Thus, in the exact rank-one case, we have

$$u = \pm\frac{(\alpha_+\widetilde{w}_+ - \alpha_-\widetilde{w}_-)}{\|\alpha_+\widetilde{w}_+ - \alpha_-\widetilde{w}_-\|}.$$

All three vectors lie in the two-dimensional plane $\mathrm{span}\{w_0, u\}$. Let e_2 be a unit vector in the direction of $P_{w_0^\perp}u$. Write $u = \cos\alpha\, w_0 + \sin\alpha\, e_2$. Then

$$\widetilde{g}_\pm = (\|g_0\| \pm a\cos\alpha)w_0 \pm a\sin\alpha\, e_2.$$

Hence, if $t_\pm := \langle\widetilde{w}_\pm, e_2\rangle/\langle\widetilde{w}_\pm, w_0\rangle$, then

$$t_+ = a\sin\alpha/(\|g_0\| + a\cos\alpha), \qquad t_- = -a\sin\alpha/(\|g_0\| - a\cos\alpha).$$

Adding and subtracting these equations gives

$$\|g_0\|/(a\sin\alpha) = \frac{1}{2}(1/t_+ - 1/t_-), \qquad \cos\alpha/\sin\alpha = \frac{1}{2}(1/t_+ + 1/t_-).$$

Thus α, and hence the direction of u, is determined by the three endpoint directions.

It remains to bound the perturbation caused by E. Since $|a| = r\lambda|\langle u, v\rangle| \geq r\lambda/8\sqrt{n}$, the perturbation of the Hessian image satisfies

$$\|rEv\|/|a| \leq 8\sqrt{n}\kappa/(\lambda).$$

Similarly,

$$\|R(\pm rv)\|/|a| \leq 4\sqrt{n}\rho r/(\lambda).$$

Therefore the normalized endpoint directions differ from those of the exact rank-one model by at most $C(8\sqrt{n}\kappa/(\lambda) + 8\sqrt{n}\rho r/(\lambda))$, provided the endpoint gradients remain nondegenerate.

The comparison oracle contributes an additional error δ_g. The reconstruction map from the three directions to the projective direction of u is locally Lipschitz under the stated nondegeneracy assumption, with Lipschitz constant at most C_{geo}. Therefore $\min\{\|\widehat{u} - u\|, \|\widehat{u} + u\|\} \le CC_{\mathrm{geo}}(\delta_g + 8\sqrt{n}\kappa/(\lambda) + 8\sqrt{n}\rho r/(\lambda))$. This proves the lemma. □

Lemma 12 (Interior or boundary consequence of line search)**.** *Suppose $\phi(t) = f(x_0 + tv)$ satisfies $\mu \le \phi''(t) \le L$ on $[-R, R]$, with $\mu > 0$. Let t_* minimize ϕ on that interval. Algorithm 4 retains either a point with $|\phi'(\widehat{t})| \le \zeta$ when $t_* \in (-R, R)$, or an endpoint with decrease at least $\mu R^2/2$ when $t_* \in \{-R, R\}$.*

Proof. Continuity gives a minimizer on the compact interval. Since $\phi'' \ge \mu > 0$, ϕ' is strictly increasing and ϕ is strictly convex, so the minimizer is unique. We prove that each retained interval contains it. If $\phi(p) \le \phi(q)$ but $t_* > q$, the derivative is negative throughout $[p, q]$: for an interior minimizer it is less than $\phi'(t_*) = 0$, and for the right endpoint minimizer it is less than or equal to its endpoint derivative, which is nonpositive. Integration would give $\phi(q) < \phi(p)$, a contradiction. Thus deleting $(q, r]$ is safe. If $\phi(p) > \phi(q)$ but $t_* < p$, the derivative is positive throughout $[p, q]$, which contradicts that inequality. Ties are covered by the first argument. This proves the invariant by induction.

Each update multiplies interval length by $2/3$, so $|\widehat{t} - t_*| \le R(2/3)^{J_{\mathrm{ls}}} \le \zeta/L$. The last inequality follows directly from the ceiling definition when $LR > \zeta$; if $LR \le \zeta$, it follows with zero iterations. For an interior minimizer, Fermat's condition gives $\phi'(t_*) = 0$. Integrating $|\phi''| \le L$ then gives $|\phi'(\widehat{t})| \le L|\widehat{t} - t_*| \le \zeta$.

If $t_* = R$, differentiability and endpoint minimality give $\phi'(R) \le 0$; otherwise sufficiently small leftward motion would decrease ϕ. Integrating the lower second-derivative bound gives

$$\phi(0) \ge \phi(R) - R\phi'(R) + \mu R^2/2 \ge \phi(R) + \mu R^2/2.$$

If $t_* = -R$, endpoint minimality gives $\phi'(-R) \ge 0$, and the same integration gives $\phi(0) \ge \phi(-R) + R\phi'(-R) + \mu R^2/2$. Both endpoints were retained, proving the claim. □

Proof of Theorem 6. If $G_0 \le \epsilon$, the retained x_0 proves the result. Assume $G_0 > \epsilon$. Change the analytical sign of u, if necessary, so $\|\widehat{u} - u\| \le \eta_u$. This does not change uu^T or the algorithm's candidates. For unit vectors, $\langle u, \widehat{u}\rangle = 1 - \|u - \widehat{u}\|^2/2$, hence $\langle u, \widehat{u}\rangle^2 \ge 1 - \eta_u^2 \ge 15/16$. The assumptions give

$$\kappa \le \lambda/8, \qquad \rho R \le \lambda/16, \qquad \lambda\eta_u R \le \epsilon/32, \qquad \kappa R \le \epsilon/128, \qquad \rho R^2 = \epsilon/256.$$

The bound $\lambda\eta_u R \le \epsilon/32$ uses the stated hypothesis $\lambda \le L$. It holds in particular for the spectral decomposition of H_0 used in the regime partition below.

First let $\lambda_1 = -\lambda$. Then

$$\widehat{u}^T H_0 \widehat{u} \le -15\lambda/16 + \kappa \le -13\lambda/16.$$

Choose $s \in \{-1, 1\}$ so $s\langle g_0, \widehat{u}\rangle \le 0$. Taylor's theorem gives

$$f(x_0 + sR\widehat{u}) - f(x_0) \le -13\lambda R^2/32 + \rho R^3/6 \le \left(-\frac{13}{32} + \frac{1}{96}\right)\lambda R^2 \le -\lambda R^2/4.$$

This point is retained. Since $\lambda \ge \sqrt{\rho\epsilon}$ and $R^2 = \epsilon/(256\rho)$, its decrease is at least $\epsilon^{3/2}/(1024\sqrt{\rho})$.

Now let $\lambda_1 = \lambda$. On the line segment, Hessian Lipschitzness gives

$$\phi''(t) = \widehat{u}^T \nabla^2 f(x_0 + t\widehat{u})\widehat{u} \ge 15\lambda/16 - \kappa - \rho R \ge 3\lambda/4 \ge \lambda/2.$$

Global smoothness gives $\phi''(t) \le L$. Thus Lemma 12 applies with $\mu = \lambda/2$. If the constrained minimizer is a boundary point, a retained endpoint decreases f by at least $\lambda R^2/4$, which is the required amount.

It remains to consider an interior minimizer. Write $P = I - \widehat{u}\widehat{u}^T$, $b_\perp = \|Pg_0\|$, and let $x_\ell = x_0 + \widehat{t}\widehat{u}$ be the retained line-search point. The line-search lemma gives $|\langle \nabla f(x_\ell), \widehat{u}\rangle| \le \epsilon/8$. If $b_\perp \le \epsilon/2$, Taylor expansion of the gradient gives

$$\|P\nabla f(x_\ell)\| \le b_\perp + R\|PH_0\widehat{u}\| + \rho R^2/2.$$

Since $P\widehat{u} = 0$, $\|Pu\| \le \|u - \widehat{u}\| \le \eta_u$, and $\|PF\widehat{u}\| \le \kappa$, we have $\|PH_0\widehat{u}\| \le \lambda\eta_u + \kappa$. Consequently

$$\|\nabla f(x_\ell)\| \le \epsilon/8 + \epsilon/2 + \epsilon/32 + \epsilon/128 + \epsilon/512 < \epsilon.$$

The retained line-search point is stationary.

If $b_\perp > \epsilon/2$, let $q = Pg_0$ and $\widehat{q} = G_0 P\widehat{g}_0$. Then $\|\widehat{q} - q\| \le \epsilon/16 < b_\perp$, so $\widehat{q} \ne 0$ and the algorithm generates $w = \widehat{q}/\|\widehat{q}\|$. Lemma 3 gives

$$\|w - q/b_\perp\| \le 2(\epsilon/16)/b_\perp < 1/4.$$

Since $w \perp \widehat{u}$, $\langle g_0, w\rangle = \langle q, w\rangle \ge b_\perp - b_\perp\|w - q/b_\perp\| \ge 3b_\perp/4$. Moreover $|\langle u, w\rangle| = |\langle u - \widehat{u}, w\rangle| \le \eta_u$, so $w^T H_0 w \le \lambda\eta_u^2 + \kappa$. Taylor's theorem now gives

$$\begin{aligned} f(x_0 - Rw) - f(x_0) &\le -3b_\perp R/4 + (\lambda\eta_u^2 + \kappa)R^2/2 + \rho R^3/6 \\ &\le \epsilon R\left(-\frac{3}{8} + \frac{1}{64} + \frac{1}{256} + \frac{1}{1536}\right) \le -\epsilon R/3. \end{aligned}$$

We used $\eta_u^2 \le \eta_u$ and the displayed parameter bounds. Since $\epsilon R/3 = \epsilon^{3/2}/(48\sqrt{\rho})$, this is stronger than the asserted decrease. All cases are covered. The line-search budget is the explicit ceiling in Algorithm 4 with $\zeta = \epsilon/8$, which gives the stated count. □

D.2 Low-Curvature Regime and Normalized-Gradient Descent

The normalized gradient candidate in Algorithm 5 serves as a fallback for the regime in which the Hessian is smaller than the target curvature scale. The accelerated candidate is responsible for the nontrivial regime $\|\nabla^2 f(x_0)\| \ge \sqrt{\rho\epsilon}$. The following lemma formalizes the fallback.

Lemma 13 (Normalized-gradient fallback)**.** *Suppose $\|H_0\| \le \sqrt{\rho\epsilon}$. If $G_0 \le \epsilon$, the retained x_0 is stationary. If $G_0 > \epsilon$ and $\|\widehat{g}_0 - g_0/G_0\| \le 1/4$, then Normalized Gradient Descent retains a point with*

$$f(x_0 - R\widehat{g}_0) - f(x_0) \le -\frac{1}{2}\epsilon R = -\frac{1}{32}\frac{\epsilon^{3/2}}{\sqrt{\rho}}.$$

Without any restriction on H_0, if $G_0 \ge LR$ and the same direction guarantee holds, the decrease is at least $15G_0R/32$.

Proof. For two unit vectors, $\langle u, v\rangle = 1 - \|u - v\|^2/2$; hence $\langle g_0, \widehat{g}_0\rangle \ge 31G_0/32$. Taylor's theorem gives

$$f(x_0 - R\widehat{g}_0) - f(x_0) \le -\frac{31}{32}G_0R + \frac{1}{2}\sqrt{\rho\epsilon}R^2 + \frac{\rho}{6}R^3.$$

Using $G_0 > \epsilon$, $\sqrt{\rho\epsilon}R = \epsilon/16$, and $\rho R^2 = \epsilon/256$, this is at most

$$\epsilon R\left(-\frac{31}{32} + \frac{1}{32} + \frac{1}{1536}\right) \le -\frac{1}{2}\epsilon R.$$

For the last assertion, global smoothness instead gives $-31G_0R/32 + LR^2/2 \le -15G_0R/32$ because $LR \le G_0$. The case $G_0 \le \epsilon$ follows from the definition of stationarity. □