# Stochastic Optimization with Random Search

**El Mahdi Chayti**[1]* **Taha El Bakkali**[2] **Omar Saadi**[3] **Martin Jaggi**[1]
[1]Machine Learning and Optimization Laboratory (MLO), EPFL
[2]Mohammed VI Polytechnic University (UM6P)
[3]College of Computing, Mohammed VI Polytechnic University (UM6P)

## Abstract

We study stochastic random search for nonconvex optimization with only noisy function evaluations. We propose Mi2P, a two-point variant of stochastic three-point methods, and analyze it under generalized $(L_0, L_1)$-smoothness. Two regimes are considered. When the mean objective $f$ is $(L_0, L_1)$-smooth and component variance is bounded, Mi2P achieves $\widetilde{\mathcal{O}}(d^3\sigma_0^2/\varepsilon^6)$, matching the rate of Boucherouite et al. [2024] under a strictly weaker assumption (they require each $f_\xi$ to be $L$-smooth). When an $L^1$-average $(L_0, L_1)$-smoothness holds across $\xi$, Mi2P attains $\widetilde{\mathcal{O}}(d\sigma^2/\varepsilon^4)$, the best known rate, matching gradient-estimation methods such as RSGF [Ghadimi and Lan, 2013]; the per-sample $L$-smoothness these methods assume (each $f_\xi$ is $L$-smooth, $L_1 = 0$ pointwise in $\xi$) is a special case of this condition. For finite sums with $G$-Lipschitz components, a translation-symmetric variance-reduction scheme yields $\widetilde{\mathcal{O}}(\min\{d^{4/3}n^{2/3}/\varepsilon^{8/3},\ dn/\varepsilon^2\})$ without storing snapshots. We also extend the framework to $\delta$-inexact comparison feedback (e.g., RLHF, A/B testing), giving an $\mathcal{O}(\sqrt{d\delta})$ accuracy floor. Our analysis is presented for directional distributions with bounded support; the same rates hold for any rotation-invariant distribution with sub-Gaussian tails on $\|s\|$ (including Gaussian) under a step-size restriction that does not affect the asymptotic complexity.

## 1 Introduction

Zeroth-order, also known as derivative-free, optimization minimizes an objective using only function evaluations, possibly noisy. Such methods are used when gradients are unavailable, unreliable, or expensive to compute, with applications including black-box adversarial attacks [Chen et al., 2017, Ughi et al., 2022], hyperparameter tuning [Koch et al., 2018, Turner et al., 2021], reinforcement learning through rollouts [Mania et al., 2018, Malik et al., 2020], and simulation-based optimization [Conn et al., 2009].

Two paradigms dominate the zeroth-order literature. *Gradient-approximation* methods estimate directional derivatives from finite differences and feed them into first-order updates [Nesterov and Spokoiny, 2017, Ghadimi and Lan, 2013, Liu et al., 2018], achieving strong oracle complexities at the cost of a discretization bias controlled by a smoothing parameter. *Direct-search* methods compare function values at a few trial points and update whenever an improvement is found [Bergou et al., 2020, Golovin et al., 2020]; they avoid discretization but have historically lacked comparable guarantees under stochastic noise.

The Stochastic Three Points (STP) method of Bergou et al. [2020] samples a random direction $s_t \sim \mathcal{D}$, evaluates $f$ at $\{x_t - \eta_t s_t,\ x_t,\ x_t + \eta_t s_t\}$, and selects the best of the three. In the deterministic setting STP achieves $\mathcal{O}(d/\varepsilon^2)$ in the smooth nonconvex case, matching the best rates among gradient-estimation methods. In the stochastic setting, Boucherouite et al. [2024] replaced $f$ with a minibatch

*Corresponding author: el-mahdi.chayti@epfl.ch

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

estimate $f_{\mathcal{C}_t}$ in the same three-point comparison (MiSTP) and proved a complexity bound of $\widetilde{\mathcal{O}}(d^3/\varepsilon^6)$ assuming each $f_\xi$ is $L$-smooth. This rate is significantly worse than the $\widetilde{\mathcal{O}}(d/\varepsilon^4)$ obtained by gradient-estimation methods such as RSGF [Ghadimi and Lan, 2013] under the same assumption. Closing this gap is the starting point of our work.

A second concern is the smoothness assumption itself. Most random-search analyses adopt a global Lipschitz form of $(L_0, L_1)$-smoothness, which excludes basic objectives such as polynomials and exponentials. We adopt instead a globally-valid form of generalized smoothness implied by the local Hessian condition of Zhang et al. [2020], which captures these standard examples; we discuss this point in Section 3.1.

**Contributions.** We introduce Mi2P, analyze it under generalized smoothness, and prove the following rate results, summarized in Table 1.

- **Algorithmic.** Mi2P is a two-point variant of STP that compares only $x_t^+$ and $x_t^-$ via the sign of $M_t^+ - M_t^-$ and updates accordingly. STP and MiSTP additionally evaluate $f_{\mathcal{C}_t}(x_t)$ and pick the best of three; in the stochastic setting this baseline evaluation breaks the monotone descent guarantee that motivated it and adds an irrelevant noisy comparison. Mi2P removes it.
- **Same rate as MiSTP under a weaker assumption.** Under generalized $(L_0, L_1)$-smoothness of the mean $f$ and bounded component variance, Mi2P attains $\widetilde{\mathcal{O}}(d^3\sigma_0^2/\varepsilon^6)$. MiSTP [Boucherouite et al., 2024] attains the same rate under the stronger assumption that each $f_\xi$ is $L$-smooth, whereas our condition requires only generalized smoothness of the mean $f$.
- **Improved rate under a weaker assumption.** Under $L^1$-average $(L_0, L_1)$-smoothness across $\xi$, Mi2P attains $\widetilde{\mathcal{O}}(d\sigma^2/\varepsilon^4)$, the best known rate, matching gradient-estimation methods such as RSGF [Ghadimi and Lan, 2013] but without discretization bias. RSGF assumes per-sample $L$-smoothness (each $f_\xi$ is $L$-smooth, i.e., $L_1 = 0$ enforced pointwise in $\xi$); our assumption is strictly weaker on two counts: it is an average over $\xi$ rather than pointwise, and it allows a gradient-norm-dependent constant.
- **Variance reduction for finite sums.** For $f(x) = \frac{1}{n}\sum_i f_i(x)$ with $G$-Lipschitz components, a translation-symmetric scheme attains $\widetilde{\mathcal{O}}\big(\min\{d^{4/3}n^{2/3}/\varepsilon^{8/3},\, dn/\varepsilon^2\}\big)$, without storing snapshots.
- **Helper or human feedback.** We extend the framework to $\delta$-inexact comparison feedback (RLHF [Christiano et al., 2017, Ouyang et al., 2022], A/B testing [Kohavi et al., 2009]) and prove convergence up to an $\mathcal{O}(\sqrt{d\delta})$ floor.
- **Bounded vs. sub-Gaussian directions.** The analysis is presented for distributions with bounded support (e.g., uniform on the unit sphere). The same rates apply to any rotation-invariant distribution with sub-Gaussian tails on $\|s\|$, including Gaussian $\mathcal{N}(0, I_d/d)$, under a step-size restriction $L_1\eta \lesssim 1$, which does not affect the asymptotic complexity in the natural step-size regime; details are in Appendix H.

Table 1: Complexity comparison for stochastic zeroth-order optimization on nonconvex problems. Rates count function evaluations to reach $\frac{1}{T}\sum_t \mathbb{E}\,\|\nabla f(x_t)\| \le \varepsilon$. The shorthand $\sigma$ stands for the variance scale appropriate to each method (gradient variance $\sigma$ or function variance $\sigma_0$). The constant $\kappa_{\mathcal{D}} \le e$ from generalized smoothness is omitted in the rates as it does not affect $\varepsilon, d, n$ scaling.

| Method | Setting | Smoothness assumption | Complexity |
|---|---|---|---|
| RSGF [Ghadimi and Lan, 2013] | Stochastic | each $f_\xi$ is $L$-smooth | $\widetilde{\mathcal{O}}(d\sigma^2/\varepsilon^4)$ |
| MiSTP [Boucherouite et al., 2024] | Stochastic | each $f_\xi$ is $L$-smooth | $\widetilde{\mathcal{O}}(d^3\sigma_0^2/\varepsilon^6)$ |
| **Mi2P** (Theorem 4.2) | Stochastic | $f$ is $(L_0, L_1)$-smooth | $\widetilde{\mathcal{O}}(d^3\sigma_0^2/\varepsilon^6)$ |
| **Mi2P** (Theorem 4.5) | Stochastic | $\mathbb{E}_\xi$-avg $(L_0, L_1)$-smooth | $\widetilde{\mathcal{O}}(d\sigma^2/\varepsilon^4)$ |
| **Mi2P-VR** (Theorem 5.1) | Finite sum, $G$-Lip $f_i$ | $f$ is $(L_0, L_1)$-smooth | $\widetilde{\mathcal{O}}\Big(\min\Big\{\frac{d^{4/3}n^{2/3}}{\varepsilon^{8/3}}, \frac{dn}{\varepsilon^2}\Big\}\Big)$ |

## 2 Related Work

**Direct search and STP.** Classical direct-search methods [Conn et al., 2009, Vicente, 2013] use deterministic polling sets and scale poorly in dimension, with $\mathcal{O}(d^2/\varepsilon^2)$ in the smooth nonconvex

case [Vicente, 2013, Konečný and Richtárik, 2014]. Random search variants improve this dimension dependence. Bergou et al. [2020] introduced STP and proved $\mathcal{O}(d/\varepsilon^2)$; El Bakkali and Saadi [2025] extended this to $\mathcal{O}(d/\varepsilon)$ in the convex case; importance sampling [Bibi et al., 2020] and momentum [Gorbunov et al., 2020] variants have been studied in the deterministic setting only.

**Stochastic random search.** The only prior analysis of STP for stochastic objectives is that of Bouchererouite et al. [2024], which obtains $\widetilde{\mathcal{O}}(d^3/\varepsilon^6)$ assuming each $f_\xi$ is $L$-smooth. We obtain the same rate under generalized $(L_0, L_1)$-smoothness of $f$, and the improved $\widetilde{\mathcal{O}}(d/\varepsilon^4)$ rate under average $(L_0, L_1)$-smoothness across $\xi$.

**Gradient-estimation methods.** Two-point estimators such as RSGF [Ghadimi and Lan, 2013] achieve $\widetilde{\mathcal{O}}(d/\varepsilon^4)$ under per-sample $L$-smoothness. These methods rely on a smoothing parameter $\mu$ that introduces a bias requiring careful tuning. Mi2P avoids smoothing entirely while attaining the same rate under a weaker assumption. Liu et al. [2018] study zeroth-order variance-reduced methods (ZO-SVRG) in the finite-sum setting; their rates are variant-dependent and not directly comparable to the stochastic-setting bounds above.

**Generalized smoothness.** The $(L_0, L_1)$-smoothness condition was introduced by Zhang et al. [2020] as a local property of the Hessian to justify gradient clipping. Subsequent global formulations vary; the condition we adopt (Assumption 3.1) is implied by their local condition and admits standard non-Lipschitz examples such as polynomials and exponentials.

## 3 Algorithmic Framework

We consider the stochastic minimization problem

$$\min_{x\in\mathbb{R}^d} f(x), \qquad f(x) := \mathbb{E}_{\xi\sim\mathcal{P}}\big[f_\xi(x)\big],$$

where $(\Xi, \mathcal{F}, \mathcal{P})$ is a probability space, $F(x,\xi) = f_\xi(x)$ is differentiable in $x$ for each $\xi$ and measurable in $\xi$ for each $x$, and $f$ is finite-valued and bounded below. The finite-sum setting corresponds to $\Xi = \{1, \dots, n\}$ with uniform $\mathcal{P}$. Throughout, $\|\cdot\|$ denotes the Euclidean norm.

### 3.1 Smoothness

**Assumption 3.1** (Generalized $(L_0, L_1)$-smoothness)**.** There exist $L_0, L_1 \geq 0$ such that for all $x, y \in \mathbb{R}^d$,

$$\|\nabla f(y) - \nabla f(x)\| \leq \big(L_0 + L_1 \|\nabla f(x)\|\big)\, e^{L_1\|y-x\|}\, \|y - x\|\,. \tag{1}$$

This condition is implied by the local Hessian condition $\big\|\nabla^2 f(x)\big\| \leq L_0 + L_1 \|\nabla f(x)\|$ of Zhang et al. [2020] via Grönwall integration; see Lemma B.1 in Appendix B. It admits standard non-Lipschitz objectives such as polynomials $\|x\|^p$ and exponentials $e^{a^\top x}$ (Example B.2). Standard $L$-smoothness corresponds to $L_1 = 0$.

Assumption 3.1 implies the following descent inequality (Lemma B.3 in Appendix B): for all $x, y \in \mathbb{R}^d$,

$$f(y) \leq f(x) + \langle \nabla f(x),\, y - x\rangle + \tfrac{1}{2}\big(L_0 + L_1 \|\nabla f(x)\|\big)\, e^{L_1\|y-x\|}\, \|y - x\|^2\,. \tag{2}$$

### 3.2 Random directions

The exponential factor in (2) is bounded as soon as $\|y - x\|$ is bounded. We therefore impose a bounded-support condition on the directional distribution.

**Assumption 3.2** (Exploration property of $\mathcal{D}$)**.** The distribution $\mathcal{D}$ on $\mathbb{R}^d$ satisfies:

1. $\mathbb{E}_{s\sim\mathcal{D}} \|s\|^2 = 1$;
2. there exists $\mu_{\mathcal{D}} > 0$ such that $\mathbb{E}_{s\sim\mathcal{D}} |\langle g,\, s\rangle| \geq \mu_{\mathcal{D}} \|g\|$ for all $g \in \mathbb{R}^d$;
3. $\|s\| \leq R$ almost surely, for some $R \geq 1$.

The uniform distribution on the unit sphere $\mathbb{S}^{d-1}$ satisfies all three with $\mu_{\mathcal{D}} \asymp 1/\sqrt{d}$ and $R = 1$. The bounded-support condition (item 3) can be relaxed to a sub-Gaussian tail bound on $\|s\|$, which covers Gaussian directions $s \sim \mathcal{N}(0, I_d/d)$ and any rotation-invariant distribution whose radius concentrates around its mean (rescaled chi distributions, truncated Gaussians, etc.); we show in Appendix H that all main results carry over to this broader class under a step-size restriction that does not affect the asymptotic complexity.

### 3.3 Algorithm

Given stochastic estimates $M_t^{\pm}$ of $f(x_t^{\pm})$, Mi2P updates

$$x_{t+1} = x_t - \eta_t \operatorname{sign}\big(M_t^+ - M_t^-\big)\, s_t.$$

**Algorithm 1** Mi2P: stochastic random search

**Require:** initial point $x_0 \in \mathbb{R}^d$, step sizes $\{\eta_t\}_{t\geq 0}$, distribution $\mathcal{D}$

1: **for** $t = 0, 1, 2, \ldots$ **do**
2: sample $s_t \sim \mathcal{D}$, set $x_t^{\pm} = x_t \pm \eta_t s_t$
3: obtain stochastic estimates $M_t^{\pm}$ of $f(x_t^{\pm})$
4: $x_{t+1} \leftarrow x_t - \eta_t \operatorname{sign}(M_t^+ - M_t^-)\, s_t$
5: **end for**

**Comparison with STP and MiSTP.** Deterministic STP [Bergou et al., 2020] updates to the best of three: $x_{t+1} \in \arg\min_{x\in\{x_t^-, x_t, x_t^+\}} f(x)$. Including $x_t$ guarantees monotone descent, $f(x_{t+1}) \leq f(x_t)$. MiSTP [Boucherouite et al., 2024] applies the same three-point rule to a minibatch surrogate $f_{\mathcal{C}_t}$. The minibatch comparison no longer guarantees descent on $f$, and the evaluation at $x_t$ contributes a noisy comparison without the descent benefit it was meant to provide. Mi2P drops it: only the two trial points $x_t^{\pm}$ are evaluated, and the update is determined by the sign of their estimated difference. Beyond the conceptual simplification, this also reduces the per-iteration query cost from three evaluations to two, a $33\%$ reduction at no theoretical loss. The estimators $M_t^{\pm}$ can be constructed in three ways: by subsampling (Section 4), variance reduction (Section 5), or helper feedback (Section 6).

### 3.4 Descent bound

The standard route to a descent bound for stochastic random search, taken by Boucherouite et al. [2024], treats $M_t^+$ and $M_t^-$ as independent noisy estimates of $f(x_t^+)$ and $f(x_t^-)$ and bounds the per-iteration error by the sum of the two approximation errors $\mathbb{E}|M_t^+ - f(x_t^+)| + \mathbb{E}|M_t^- - f(x_t^-)|$. Each term is at best $\mathcal{O}(\sigma_0/\sqrt{b})$ under bounded variance, and this is what limits the analysis to the $\widetilde{\mathcal{O}}(d^3/\varepsilon^6)$ rate.

A simple but consequential observation is that the algorithm is invariant under shifts of the estimators: replacing $(M_t^+, M_t^-)$ by $(M_t^+ + c, M_t^- + c)$ leaves the update unchanged. We may therefore optimize the per-iteration error over $c \in \mathbb{R}$. The minimizer collapses the sum of two errors into a single error on the difference:

$$\inf_{c\in\mathbb{R}} \big(|M_t^+ - c - f(x_t^+)| + |M_t^- - c - f(x_t^-)|\big) = \big|\,(M_t^+ - M_t^-) - (f(x_t^+) - f(x_t^-))\,\big|. \tag{3}$$

The right-hand side can be much smaller than either of the two original errors. When $M_t^+$ and $M_t^-$ are formed from the same minibatch, the noise in the two evaluations is highly correlated: most of the per-point fluctuation $f_\xi(x_t^{\pm}) - f(x_t^{\pm})$ reflects the value of $f_\xi$ at a single iterate (shared between the two evaluations) rather than the local variation of $f_\xi$ across the perturbation $\eta_t s_t$. The shared component cancels in the difference, and only the residual variation matters. Building this cancellation into the analysis from the outset yields the following descent bound, which is the workhorse of all subsequent results. Define

$$\kappa_{\mathcal{D}} := e^{L_1 \eta_t R}.$$

Under Assumption 3.2, $\|s_t\| \leq R$ a.s., so the exponential factor in (2) along an iteration is bounded by $\kappa_{\mathcal{D}}$.

**Lemma 3.3** (Descent bound)**.** *Suppose $f$ satisfies Assumption 3.1 and $s_t \sim \mathcal{D}$ satisfies Assumption 3.2. If $\eta_t \le 1/(L_1 R)$ (so $\kappa_{\mathcal{D}} \le e$) and $\eta_t \le \mu_{\mathcal{D}}/(\kappa_{\mathcal{D}} L_1)$, then*

$$\mathbb{E}[f(x_{t+1})] \le \mathbb{E}[f(x_t)] - \tfrac{\mu_{\mathcal{D}}}{2}\eta_t \,\mathbb{E}\,\|\nabla f(x_t)\| + \tfrac{\kappa_{\mathcal{D}} L_0}{2}\eta_t^2 + 2\,\mathbb{E}\big|\, M_t^+ - M_t^- - (f(x_t^+) - f(x_t^-))\,\big|. \quad (4)$$

The proof, in Appendix C, follows the structure of [Bergou et al., 2020, Boucherouite et al., 2024] with (2) replacing the global Lipschitz quadratic bound, $\|s_t\| \le R$ a.s. controlling the exponential factor, and the translation invariance (3) reducing the noise term. The remaining task in each subsequent section is to bound the difference error $\mathbb{E}|(M_t^+ - M_t^-) - (f(x_t^+) - f(x_t^-))|$ for the relevant choice of estimators.

## 4 Stochastic Subsampling

We consider the general stochastic problem with minibatch estimators

$$M_t^{\pm} := \tfrac{1}{b}\sum_{j=1}^{b} f_{\xi_j}(x_t^{\pm}), \qquad \{\xi_j\}_{j=1}^{b} \text{ i.i.d.}$$

We analyze two regimes, distinguished by where the smoothness assumption is imposed: on the mean $f$ alone, or as an average of per-sample smoothness across the components $f_\xi$. The two regimes exploit different amounts of structure in the components and therefore obtain different rates: when only the mean is smooth, the noise in the difference $M_t^+ - M_t^-$ is controlled by the variance of the function values; when smoothness holds across $\xi$, that noise can be further reduced through directional averaging on the sphere, improving $\widetilde{\mathcal{O}}(d^3/\varepsilon^6)$ to $\widetilde{\mathcal{O}}(d/\varepsilon^4)$.

### 4.1 Smoothness on the mean

In this regime we assume only that the mean $f$ is $(L_0, L_1)$-smooth, plus a bounded variance condition on the components.

**Assumption 4.1** (Bounded component variance)**.** There exists $\sigma_0^2 < \infty$ such that $\mathbb{E}[(f_\xi(x) - f(x))^2] \le \sigma_0^2$ for all $x \in \mathbb{R}^d$.

**Theorem 4.2.** *Let $F_0 := f(x_0) - f^\star < \infty$. Under Assumptions 3.1, 3.2 and 4.1, with constant step size $\eta \le \mu_{\mathcal{D}}/(\kappa_{\mathcal{D}} L_1)$,*

$$\tfrac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\,\|\nabla f(x_t)\| = \tfrac{1}{\mu_{\mathcal{D}}}\,\mathcal{O}\left(\tfrac{\kappa_{\mathcal{D}} L_1 F_0}{\mu_{\mathcal{D}} T} + \sqrt{\tfrac{\kappa_{\mathcal{D}} L_0 F_0}{T}} + \tfrac{\sqrt{\kappa_{\mathcal{D}} L_0 \sigma_0}}{b^{1/4}}\right).$$

*With $\mu_{\mathcal{D}} \asymp 1/\sqrt{d}$, choosing $T = \mathcal{O}(dL_1/\varepsilon + dL_0F_0/\varepsilon^2)$ and $b = \mathcal{O}((dL_0\sigma_0)^2/\varepsilon^4)$ yields a total complexity of $\widetilde{\mathcal{O}}(d^3\sigma_0^2/\varepsilon^6)$ function evaluations.*

The proof is in Appendix D.1. The rate matches that of MiSTP [Boucherouite et al., 2024], but our assumption is strictly weaker: Boucherouite et al. require each $f_\xi$ to be $L$-smooth, whereas Assumption 3.1 imposes only generalized smoothness of the mean $f$.

**Where the $d^3/\varepsilon^6$ rate comes from.** Three sources of dimension-dependence stack up. The exploration constant $\mu_{\mathcal{D}} \asymp 1/\sqrt{d}$ contributes a factor of $\sqrt{d}$ to the iteration count $T$. The minibatch size $b$ scales as $(dL_0\sigma_0)^2/\varepsilon^4$, since the noise in $M_t^+ - M_t^-$ is $\mathcal{O}(\sigma_0/\sqrt{b})$ and must be small compared to the $\eta\,\|\nabla f\|$ signal, with $\eta \lesssim \mu_{\mathcal{D}} \asymp 1/\sqrt{d}$. Multiplying $T$ and $b$ gives the $d^3$ factor. The $\varepsilon^{-6}$ exponent decomposes as $\varepsilon^{-2}$ from $T$ and $\varepsilon^{-4}$ from $b$, in line with the general principle that controlling a comparison-based oracle to accuracy $\eta\varepsilon$ on the gradient requires noise $\mathcal{O}(\eta\varepsilon)$, hence $b \gtrsim \sigma_0^2/(\eta\varepsilon)^2$.

### 4.2 Average smoothness across $\xi$

To obtain the $\widetilde{\mathcal{O}}(d/\varepsilon^4)$ rate, the best known rate, matching gradient-estimation methods such as RSGF, we strengthen the smoothness condition slightly, but in a way that remains weaker than per-sample $L$-smoothness.

**Assumption 4.3** (Average $(L_0, L_1)$-smoothness across $\xi$)**.** There exist $L_0, L_1 \geq 0$ such that for all $x, y \in \mathbb{R}^d$,

$$\mathbb{E}_\xi \left\|\nabla f_\xi(y) - \nabla f_\xi(x)\right\| \leq \left(L_0 + L_1 \left\|\nabla f(x)\right\|\right) e^{L_1\|y-x\|} \left\|y - x\right\|. \tag{5}$$

**Assumption 4.4** (Bounded gradient variance)**.** There exists $\sigma \geq 0$ such that $\mathbb{E}[\nabla f_\xi(x)] = \nabla f(x)$ and $\mathbb{E}\left\|\nabla f_\xi(x) - \nabla f(x)\right\|^2 \leq \sigma^2$.

Per-sample $L$-smoothness (each $f_\xi$ is $L$-smooth) implies Assumption 4.3 via Jensen's inequality and Assumption 4.4; see Lemma D.1. The converse is false: Assumption 4.3 accommodates components $f_\xi$ that are potentially individually non-smooth, provided the averaged modulus $\mathbb{E}_\xi \left\|\nabla f_\xi(y) - \nabla f_\xi(x)\right\|$ remains controlled as in (5). Per-sample $L$-smoothness is therefore strictly stronger than Assumption 4.3, which in turn is stronger than Assumption 3.1 alone (Corollary D.2: Assumption 4.3 implies $f$ is itself $(L_0, L_1)$-smooth).

**Theorem 4.5.** *Let $F_0 := f(x_0) - f^\star < \infty$. Under Assumption 3.2 (with $\mathcal{D}$ uniform on the unit sphere), Assumption 4.3, and Assumption 4.4, with step size $\eta \leq \mu_{\mathcal{D}}/(16\,\kappa_{\mathcal{D}} L_1)$,*

$$\tfrac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\left\|\nabla f(x_t)\right\| = \tfrac{1}{\mu_{\mathcal{D}}}\,\mathcal{O}\left(\tfrac{\kappa_{\mathcal{D}} L_1 F_0}{\mu_{\mathcal{D}} T} + \sqrt{\tfrac{\kappa_{\mathcal{D}} L_0 F_0}{T}} + \tfrac{\sigma}{\sqrt{db}}\right).$$

*With $\mu_{\mathcal{D}} \asymp 1/\sqrt{d}$, choosing $T = \mathcal{O}(dL_1/\varepsilon + dL_0F_0/\varepsilon^2)$ and $b = \mathcal{O}(\sigma^2/\varepsilon^2)$ yields a total complexity of $\widetilde{\mathcal{O}}(d\sigma^2/\varepsilon^4)$ function evaluations.*

The proof is in Appendix D.2. The rate matches RSGF [Ghadimi and Lan, 2013], but that analysis assumes per-sample $L$-smoothness, which is strictly stronger than Assumption 4.3. Mi2P also avoids the discretization bias of finite-difference gradient estimators.

**Where the improvement over $\varepsilon^{-6}$ comes from.** Compared to Theorem 4.2, the noise term in the rate decreases from $\mathcal{O}(\sigma_0/(\eta\sqrt{b}))$ to $\mathcal{O}(\sigma/\sqrt{db})$. Two facts drive this. First, Assumption 4.3 pins the per-sample function difference $f_\xi(x_t^+) - f_\xi(x_t^-)$ to within $\mathcal{O}(\eta^2)$ of the population difference, so the residual noise is no longer a function-value variance but a gradient projection $\langle \nabla f_\xi(x_t) - \nabla f(x_t),\, s_t\rangle$. Second, since $s_t$ is uniform on the sphere, $\mathbb{E}\left\langle g,\, s_t\right\rangle^2 = \left\|g\right\|^2/d$ (Lemma A.3), giving a $1/\sqrt{d}$ factor not present in the function-value variance bound. Together, these turn $\sigma_0/\sqrt{b}$ into $\sigma/\sqrt{db}$ with $\sigma$ now a gradient variance, so the required $b$ drops to $\mathcal{O}(\sigma^2/\varepsilon^2)$, an $\varepsilon^{-2}$ improvement over Theorem 4.2 and a $d$ improvement, exactly accounting for the gap from $d^3/\varepsilon^6$ to $d/\varepsilon^4$.

## 5 Variance Reduction in the Finite-Sum Setting

We now consider $f(x) = \frac{1}{n}\sum_{i=1}^n f_i(x)$, where each $f_i$ is $G$-Lipschitz: $\|\nabla f_i\| \leq G$. We apply variance reduction by periodically evaluating the full dataset. For a fixed epoch length $m \geq 1$,

$$M_t^\pm = \begin{cases} f(x_t^\pm), & t \equiv 0 \pmod m, \\ \frac{1}{b}\sum_{j=1}^b f_{\xi_j}(x_t^\pm), & \text{otherwise.} \end{cases}$$

**Snapshots: standard scheme vs. Mi2P.** Standard variance-reduction schemes such as SVRG [Johnson and Zhang, 2013] and ZO-SVRG [Liu et al., 2018] maintain a snapshot iterate $\tilde{x}$, refreshed every $m$ iterations together with the corresponding full-batch quantity. Between refreshes, each minibatch estimate is corrected by the snapshot: in the zeroth-order analogue, one would use

$$M_t^\pm = f(\tilde{x}^\pm) + \tfrac{1}{b}\sum_{j=1}^b \big(f_{\xi_j}(x_t^\pm) - f_{\xi_j}(\tilde{x}^\pm)\big),$$

which forms a control variate that reduces the variance of $M_t^\pm$ as an estimator of $f(x_t^\pm)$. Implementing this requires storing both the snapshot iterate and its full-batch values, a non-trivial memory cost in large-scale settings.

For Mi2P this storage is unnecessary. By translation invariance (3), the algorithm depends only on the difference $M_t^+ - M_t^-$. Choosing the same snapshot $\tilde{x}^+ = \tilde{x}^-$, the terms $f(\tilde{x}^\pm)$ are constants independent of the trial points $x_t^\pm$, so they enter $M_t^+$ and $M_t^-$ as a common shift and cancel in the difference. The plain minibatch estimator $M_t^\pm = \frac{1}{b}\sum_j f_{\xi_j}(x_t^\pm)$ already produces the same variance-reduced bound on $M_t^+ - M_t^-$ that the snapshot-corrected version produces, without implementation overhead.

**Theorem 5.1.** *Under Assumptions 3.1 and 3.2 and the G-Lipschitz finite-sum setting, with step size $\eta \le \mu_{\mathcal{D}}/(\kappa_{\mathcal{D}} L_1)$,*

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\,\|\nabla f(x_t)\| = \frac{1}{\mu_{\mathcal{D}}}\,\mathcal{O}\left(\frac{\kappa_{\mathcal{D}} L_1 F_0}{\mu_{\mathcal{D}} T} + \sqrt{\frac{\kappa_{\mathcal{D}} L_0 F_0}{T}} + \frac{Gm}{\sqrt{b}}\right).$$

*Choosing $T = \mathcal{O}(dL_1/\varepsilon + dL_0F_0/\varepsilon^2)$, $b(m) = \mathcal{O}(dG^2m^2/\varepsilon^2)$, and optimizing over $m$ subject to $b(m) \le n$,*

$$\textsc{Calls}^\star = \mathcal{O}\left(\min\left\{\frac{d^{4/3}n^{2/3}L_0^{1/3}G^{2/3}F_0}{\varepsilon^{8/3}},\ \frac{dnL_0F_0}{\varepsilon^2}\right\}\right).$$

The proof is in Appendix E. The complexity improves on the deterministic $\mathcal{O}(dnL_0F_0/\varepsilon^2)$ whenever $n \gtrsim dG^2/\varepsilon^2$.

**Role of the epoch length $m$.** The parameter $m$ controls a trade-off between two competing costs. With $m = 1$, every iteration uses an exact full-batch evaluation: the difference estimator is exact, but each iteration costs $n$ queries, recovering the deterministic complexity $\mathcal{O}(dnL_0F_0/\varepsilon^2)$. With $m \to \infty$, the algorithm reduces to pure subsampling: cheap iterations but the iterates drift $\Theta(m\eta)$ away from the snapshot, inflating the variance of the difference estimator. Specifically, the bound on $\mathbb{E}|\Delta_t|$ scales with $m\eta/\sqrt{b}$, which forces $b = \Theta(dG^2m^2/\varepsilon^2)$, while the amortized full-pass cost is $n/m$ per iteration. Optimizing balances these two contributions and produces $m^\star = \Theta((n\varepsilon^2/dG^2)^{1/3})$ for $n$ in the relevant range, giving the $d^{4/3}n^{2/3}/\varepsilon^{8/3}$ scaling. The minimum with $dn/\varepsilon^2$ in Theorem 5.1 captures the regime where $b(m^\star) \ge n$ and full-batch evaluations become cheaper than the variance-reduced minibatch.

## 6 Inexact Comparison Feedback

In settings such as RLHF [Christiano et al., 2017, Ouyang et al., 2022] and A/B testing [Kohavi et al., 2009], exact function evaluations are unavailable; instead one observes a scalar signal $h$ whose differences approximate those of $f$.

**Assumption 6.1** ($\delta$-inexact comparison feedback)**.** There exists $\delta \ge 0$ such that for all $x, y \in \mathbb{R}^d$, $\mathbb{E}\,|\,h(x) - h(y) - (f(x) - f(y))\,| \le \delta$.

This is in the spirit of the helper frameworks of Chayti and Karimireddy [2024] for first-order and Chayti et al. [2024] for second-order methods. Setting $M_t^\pm = h(x_t^\pm)$ in Algorithm 1:

**Theorem 6.2.** *Under Assumptions 3.1, 3.2 and 6.1, with $\eta \le \mu_{\mathcal{D}}/(\kappa_{\mathcal{D}} L_1)$,*

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\,\|\nabla f(x_t)\| = \frac{1}{\mu_{\mathcal{D}}}\,\mathcal{O}\left(\frac{\kappa_{\mathcal{D}} L_1 F_0}{\mu_{\mathcal{D}} T} + \sqrt{\frac{\kappa_{\mathcal{D}} L_0 F_0}{T}} + \sqrt{\kappa_{\mathcal{D}} L_0\,\delta}\right).$$

*Choosing $T = \mathcal{O}(dL_1/\varepsilon + dL_0F_0/\varepsilon^2)$ ensures $\frac{1}{T}\sum_t \mathbb{E}\,\|\nabla f(x_t)\| \le \varepsilon + \mathcal{O}(\sqrt{d\delta})$.*

The proof is in Appendix F. Convergence is therefore guaranteed up to an $\mathcal{O}(\sqrt{d\delta})$ accuracy floor reflecting the helper's inexactness.

**Why $\sqrt{d\delta}$ and not $\delta$.** The helper noise contributes a constant $\delta$ to the per-iteration descent budget regardless of $\eta$, while the signal $\mathbb{E}|\,\langle \nabla f(x_t),\, \eta s_t\rangle\,|$ is $\mu_{\mathcal{D}}\eta\,\|\nabla f(x_t)\| \asymp \eta\,\|\nabla f(x_t)\|\,/\sqrt{d}$. Setting noise equal to signal yields $\|\nabla f(x_t)\| \asymp \sqrt{d}\,\delta/\eta$, and balancing with the smoothness term $L_0\eta$ gives $\eta \asymp \sqrt{\delta/L_0}$ and $\|\nabla f(x_t)\| \asymp \sqrt{dL_0\delta}$. The $\sqrt{d}$ factor is therefore the unavoidable cost of inferring a $d$-dimensional gradient from one-dimensional comparisons in a random direction.

## 7 Limitations of Momentum

A natural question is whether momentum-based variance reduction techniques developed for first-order methods (Heavy-Ball [Polyak, 1964], STORM [Cutkosky and Orabona, 2019], implicit gradient transport [Arnold et al., 2019]) can be adapted to Mi2P by replacing stochastic gradients with function differences. We show in Appendix G that none of these adaptations improves the rate. The structural obstruction is that, unlike a stochastic gradient, a finite difference $f_\xi(x_t^+) - f_\xi(x_t^-)$ has signal and noise that both scale linearly in the step size $\eta$. Exponential averaging, therefore, cannot trade variance for bias the way it does for gradient oracles, and the optimum of the resulting bias-variance trade-off is $\beta = 1$, i.e., no momentum.

## 8 Experiments

Two further experiments are reported in the appendix: the dependence of the query complexity on the dimension (Appendix I) and the finite-sum setting of Section 5 (Appendix J).

### 8.1 Logistic regression

We compare Mi2P, RSGF, and a coordinate-wise zeroth-order method (ZO-CD) on $\ell_2$-regularized logistic regression on the Breast Cancer Wisconsin dataset ($n = 455$, $d = 30$) [Wolberg et al., 1995]. The objective is

$$f(x) = \tfrac{1}{n} \sum_{i=1}^{n} \log\bigl(1 + e^{-y_i \langle a_i,\, x\rangle}\bigr) + \tfrac{\lambda}{2n} \|x\|^2 .$$

For each batch size $b \in \{1, 5, 10, 25, 50, 100\}$, we run 20 trials. Step sizes are tuned once per $b$ in a pilot phase, then fixed. We compare under matched function-query budgets and report $f(x)$ versus queries (mean $\pm$ one s.d.).

Each iteration of RSGF samples $u$ on the unit sphere and forms $g = \mu^{-1}(F_B(x + \mu u) - F_B(x))\, u$; ZO-CD uses $g = \sum_{i=1}^{d} (2\mu)^{-1}(F_B(x + \mu e_i) - F_B(x - \mu e_i))\, e_i$. Mi2P uses only the sign of $F_B(x + \alpha s) - F_B(x - \alpha s)$. Per iteration, Mi2P and RSGF cost $2b$ queries; ZO-CD costs $2bd$.

Results are shown in Figure 1. For $b \in \{1, 5\}$, Mi2P underperforms RSGF: noise dominates the $\Theta(\eta)$ signal in $f_B(x_t^+) - f_B(x_t^-)$, making the sign comparison unreliable. For $b \geq 10$, Mi2P matches or surpasses RSGF and consistently outperforms ZO-CD, in line with the $\widetilde{\mathcal{O}}(d\sigma^2/\varepsilon^4)$ rate of Theorem 4.5.

### 8.2 Policy search with preference feedback

We evaluate Mi2P, MiSTP, and Zeroth-Order Policy Gradient (ZPG) [Zhang and Ying, 2025] in a policy-search setting with pairwise preference feedback. ZPG applies an RSGF-style gradient estimator to a preference-based objective. For two policies $\pi_\theta, \pi_{\theta'}$ with returns $R(\pi_\theta), R(\pi_{\theta'})$, we generate Bradley–Terry preferences via

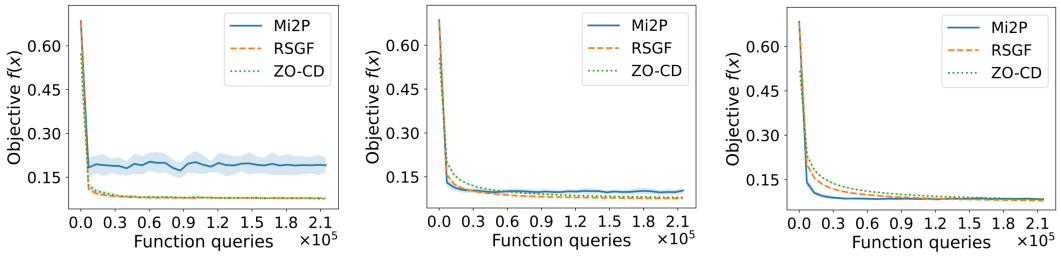
$$\pi_{\theta'} \succ \pi_\theta \iff \mathrm{Bernoulli}\bigl(\sigma(R(\pi_{\theta'}) - R(\pi_\theta))\bigr) = 1,$$

where $\sigma(t) = 1/(1 + e^{-t})$. Each comparison aggregates $N = 64$ trajectories.

We use three Gymnasium environments [Towers et al., 2024]: `CartPole-v1`, `InvertedPendulum-v5`, `Swimmer-v5`. All methods use identical two-layer MLPs with 64 hidden units and tanh activations. Hyperparameters: ZPG smoothing $\mu = 10^{-1}$ and step size $\alpha = 10^{-3}$; Mi2P step size $10^{-1}$; MiSTP step size $2.5 \cdot 10^{-2}$. Step sizes were tuned on a small grid. We run each method for a fixed evaluation budget; Table 2 reports mean $\pm$ standard deviation of final returns over 10 seeds.

Table 2: Final returns (mean $\pm$ s.d.) on three Gymnasium environments.

| Env | Mi2P | MiSTP | ZPG |
|---|---|---|---|
| `CartPole-v1` | $\mathbf{473.9 \pm 19.7}$ | $464.1 \pm 63.4$ | $223.0 \pm 166.7$ |
| `InvertedPendulum-v5` | $\mathbf{353.0 \pm 190.2}$ | $330.2 \pm 232.9$ | $257.9 \pm 214.6$ |
| `Swimmer-v5` | $\mathbf{258.8 \pm 62.0}$ | $208.8 \pm 39.4$ | $71.4 \pm 78.2$ |




(a) $b = 1$ (b) $b = 5$ (c) $b = 10$

(d) $b = 25$ (e) $b = 50$ (f) $b = 100$

Figure 1: Logistic regression on Breast Cancer Wisconsin: $f(x)$ vs. function queries for different batch sizes. Mean $\pm$ one s.d. over 20 runs.

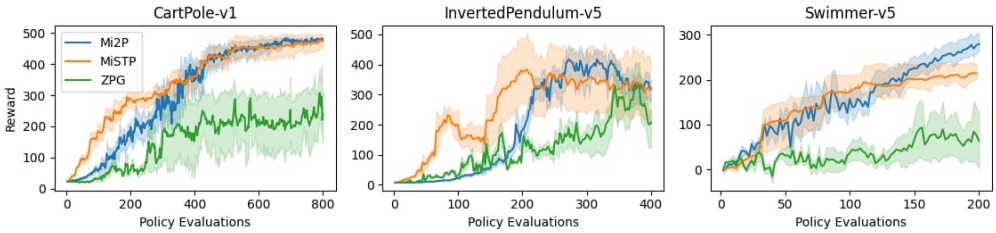



Figure 2: Learning curves on `CartPole-v1` (left), `InvertedPendulum-v5` (middle), `Swimmer-v5` (right). Shaded regions show $\pm 1$ s.d. over 10 seeds.

Mi2P consistently obtains the best mean return across the three environments. MiSTP performs comparably on `CartPole-v1`; the gap widens on `InvertedPendulum-v5` and `Swimmer-v5`. ZPG exhibits high variance and lower mean returns, particularly on `Swimmer-v5`.

### 8.3 Hyperparameter tuning with a cheap helper

We illustrate the helper setting of Section 6, where the objective of interest is expensive to evaluate and a cheap proxy provides the signal $h$. The task is to tune group-wise $\ell_2$ regularization for logistic regression on Breast Cancer Wisconsin. The $d = 30$ features are partitioned into $G = 6$ groups; group $g$ is penalized by $\lambda_g = e^{\theta_g}$, and the tuned variable is $\theta \in \mathbb{R}^G$, which plays the role of the iterate in Algorithm 1. For a given $\theta$, a model is trained by gradient descent (learning rate $0.5$) on a $60\%$ training split, and its logistic loss on the remaining $40\%$ defines the objective $f(\theta)$. The true objective uses full training (300 steps on the full training set); the helper $h(\theta)$ uses short training on a random subsample (50 steps, 200 examples) and is about ten times cheaper per evaluation. The helper is biased by early stopping and stochastic through the subsample, yet its differences track those of $f$ up to a bounded expected deviation, which is the regime of Assumption 6.1. Writing $\theta^{\pm} = \theta \pm \eta s$ for a base point $\theta$ and a direction $s$, we estimate $\delta \approx 0.013$ for this helper by averaging $|(h(\theta^+) - h(\theta^-)) - (f(\theta^+) - f(\theta^-))|$ over 2000 such pairs, with $\theta$ drawn uniformly in $[-2, 2.5]^G$, the region visited by the optimizer.

We run Algorithm 1 with $M_t^{\pm} = h(\theta_t^{\pm})$, where $\theta_t^{\pm} = \theta_t \pm \eta s_t$, $\eta = 0.3$ and $s_t$ is uniform on the unit sphere of $\mathbb{R}^G$, starting from the over-regularized point $\theta_0 = 2.5 \cdot \mathbf{1}$, and track the true loss $f(\theta_t)$; Figure 3 reports the median over 6 seeds. Using only the cheap helper, Mi2P reduces the true loss

from $0.121$ to about $0.095$, closing more than half of the gap to the best reachable value $0.073$ at roughly ten times lower per-evaluation cost, and then settles at a floor set by the helper's inexactness. Tuning directly on $f$ reaches the optimum at full cost. The residual gap above the optimum is the accuracy floor of Theorem 6.2. A controlled synthetic experiment (Appendix K) confirms that this floor scales as $\sqrt{d\delta}$ in both the dimension $d$ and the helper deviation $\delta$.

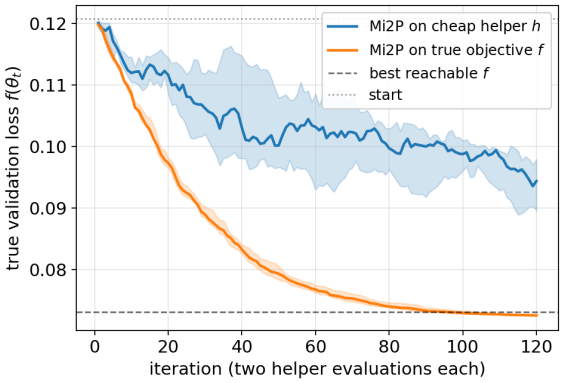



Figure 3: Hyperparameter tuning with a cheap helper. Mi2P tunes group-wise $\ell_2$ regularization using only the cheap proxy $h$, and settles at a floor above the optimum set by the helper deviation $\delta$ (Theorem 6.2). Median over 6 seeds, interquartile band.

## 9 Limitations

The step sizes prescribed by our theorems depend on the smoothness parameters $L_0$ and $L_1$, which are typically unknown in black-box settings. The same holds for RSGF and MiSTP, whose step sizes depend on $L$. In our experiments, step sizes are selected by a small grid search rather than set from theory. Designing adaptive variants of Mi2P that do not require knowledge of $L_0$ and $L_1$ while retaining the rates of Theorems 4.2, 4.5, 5.1 and 6.2 is an interesting open question.

## 10 Conclusion

We analyzed Mi2P, a two-point variant of stochastic random search, under generalized $(L_0, L_1)$-smoothness. The analysis matches the $\widetilde{\mathcal{O}}(d^3/\varepsilon^6)$ complexity of MiSTP under a strictly weaker assumption, and obtains $\widetilde{\mathcal{O}}(d/\varepsilon^4)$, the best known rate, matching RSGF, under an average smoothness condition that is itself implied by, but strictly weaker than, the per-sample smoothness assumed by gradient-estimation methods. We also gave a snapshot-free variance-reduction scheme for finite sums and an extension to inexact comparison feedback. The same rates hold for sub-Gaussian directional distributions (including Gaussian) under a step-size restriction that does not affect asymptotic complexity. Open directions include momentum-like schemes that account for the $\Theta(\eta)$ scaling of both signal and noise in comparison-based oracles, and large-scale evaluation in human-in-the-loop settings.

## Acknowledgements and Disclosures

We thank Rayane Bouftini for help with the reinforcement learning experiments. The authors gratefully acknowledge access to the computing resources of the Toubkal Supercomputer at UM6P (Morocco). The authors declare no competing interests. No specific funding supported this work.

## References


S. M. R. Arnold, P.-A. Manzagol, R. Babanezhad, I. Mitliagkas, and N. Le Roux. Reducing the variance in online optimization by transporting past gradients. *NeurIPS*, 2019.

E. H. Bergou, E. Gorbunov, and P. Richtárik. Stochastic three points method for unconstrained smooth minimization. *SIAM J. Optim.*, 30(4):2726–2749, 2020.

A. Bibi, E. H. Bergou, O. Sener, B. Ghanem, and P. Richtárik. A stochastic derivative-free optimization method with importance sampling. *AAAI*, 2020.

S. Boucherouite, G. Malinovsky, P. Richtárik, and E. H. Bergou. Minibatch stochastic three points method for unconstrained smooth minimization. *AAAI*, 2024.

E. M. Chayti and S. P. Karimireddy. Optimization with access to auxiliary information. *TMLR*, 2024.

E. M. Chayti, M. Jaggi, and N. Doikov. Unified convergence theory of stochastic and variance-reduced cubic Newton methods. *TMLR*, 2024.

P.-Y. Chen, H. Zhang, Y. Sharma, J. Yi, and C.-J. Hsieh. ZOO: Zeroth-order optimization based black-box attacks. *Workshop on AI and Security*, 2017.

P. F. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei. Deep reinforcement learning from human preferences. *NeurIPS*, 2017.

A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. SIAM, 2009.

A. Cutkosky and F. Orabona. Momentum-based variance reduction in non-convex SGD. *NeurIPS*, 2019.

T. El Bakkali and O. Saadi. On the almost sure convergence of the stochastic three points algorithm. *ICLR*, 2025.

S. Ghadimi and G. Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM J. Optim.*, 23(4):2341–2368, 2013.

D. Golovin, J. Karro, G. Kochanski, C. Lee, X. Song, and Q. Zhang. Gradientless descent: high-dimensional zeroth-order optimization. *ICLR*, 2020.

E. Gorbunov, A. Bibi, O. Sener, E. H. Bergou, and P. Richtárik. A stochastic derivative-free optimization method with momentum. *ICLR*, 2020.

R. Johnson and T. Zhang. Accelerating stochastic gradient descent using predictive variance reduction. *NeurIPS*, 2013.

P. Koch, O. Golovidov, S. Gardner, B. Wujek, J. Griffin, and Y. Xu. Autotune: a derivative-free optimization framework. *KDD*, 2018.

R. Kohavi, R. Longbotham, D. Sommerfield, and R. M. Henne. Controlled experiments on the web: survey and practical guide. *KDD*, 2009.

J. Konečný and P. Richtárik. Simple complexity analysis of simplified direct search. arXiv:1410.0390, 2014.

S. Liu, B. Kailkhura, P.-Y. Chen, P. Ting, S. Chang, and L. Amini. Zeroth-order stochastic variance reduction for nonconvex optimization. *NeurIPS*, 2018.

D. Malik, A. Pananjady, K. Bhatia, K. Khamaru, P. L. Bartlett, and M. J. Wainwright. Derivative-free methods for policy optimization. *JMLR*, 21:1–51, 2020.

H. Mania, A. Guy, and B. Recht. Simple random search of static linear policies is competitive for reinforcement learning. *NeurIPS*, 2018.

Y. Nesterov and V. Spokoiny. Random gradient-free minimization of convex functions. *Found. Comput. Math.*, 17(2):527–566, 2017.

L. Ouyang, J. Wu, X. Jiang, D. Almeida, et al. Training language models to follow instructions with human feedback. *NeurIPS*, 2022.

B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.*, 4(5):1–17, 1964.

M. Towers, J. K. Terry, A. Kwiatkowski, J. U. Balis, et al. Gymnasium: A standard interface for reinforcement learning environments. arXiv:2407.17032, 2024.

R. Turner, D. Eriksson, M. McCourt, J. Kiili, E. Laaksonen, Z. Xu, and I. Guyon. Bayesian optimization is superior to random search for ML hyperparameter tuning: Analysis of the Black-Box Optimization Challenge 2020. *NeurIPS Competition Track*, 2021.

G. Ughi, V. Abrol, and J. Tanner. An empirical study of derivative-free-optimization algorithms for targeted black-box attacks. *Optim. Eng.*, 23(3):1319–1346, 2022.

R. Vershynin. *High-Dimensional Probability*. Cambridge University Press, 2018.

L. N. Vicente. Worst case complexity of direct search. *EURO J. Comput. Optim.*, 1(1):143–153, 2013.

W. H. Wolberg, W. N. Street, and O. L. Mangasarian. Breast Cancer Wisconsin (Diagnostic) Data Set. *UCI Machine Learning Repository*, 1995. DOI: `https://doi.org/10.24432/C5DW2B`.

J. Zhang, T. He, S. Sra, and A. Jadbabaie. Why gradient clipping accelerates training: a theoretical justification for adaptivity. *ICLR*, 2020.

Q. Zhang and L. Ying. Zeroth-order policy gradient for reinforcement learning from human feedback without reward inference. *ICLR*, 2025.

# A Preliminaries

**Lemma A.1** (Variance of a minibatch average)**.** *Let $(Z_i)_{i=1}^b$ be i.i.d. with $\mathbb{E}[Z_1] = 0$ and $\mathrm{Var}(Z_1) \leq \sigma^2$. Then $\bar{Z} := \frac{1}{b}\sum_i Z_i$ satisfies $\mathbb{E}\bar{Z} = 0$ and $\mathrm{Var}(\bar{Z}) = \mathrm{Var}(Z_1)/b \leq \sigma^2/b$. The same bound holds in $\mathbb{R}^d$ with $\mathbb{E}\left\|\bar{z}\right\|^2 \leq \sigma^2/b$ when $\mathbb{E}\left\|z_1\right\|^2 \leq \sigma^2$.*

*Proof.* By the linearity of the expectation operator, we have:

$$\mathbb{E}[\bar{Z}] = \mathbb{E}\left[\frac{1}{b}\sum_{i=1}^{b} Z_i\right] = \frac{1}{b}\sum_{i=1}^{b}\mathbb{E}[Z_i] = 0.$$

To compute the variance, we expand the square of the sum. Since the mean is zero, $\mathrm{Var}(\bar{Z}) = \mathbb{E}[(\bar{Z})^2]$:

$$\begin{aligned}\mathrm{Var}(\bar{Z}) &= \mathbb{E}\left[\left(\frac{1}{b}\sum_{i=1}^{b} Z_i\right)^2\right] = \frac{1}{b^2}\mathbb{E}\left[\left(\sum_{i=1}^{b} Z_i\right)\left(\sum_{j=1}^{b} Z_j\right)\right] \\ &= \frac{1}{b^2}\mathbb{E}\left[\sum_{i=1}^{b} Z_i^2 + \sum_{i\neq j} Z_i Z_j\right].\end{aligned}$$

By the linearity of expectation and the fact that the variables $Z_i$ and $Z_j$ are independent for $i \neq j$, the cross-terms factorize and vanish: $\mathbb{E}[Z_i Z_j] = \mathbb{E}[Z_i]\mathbb{E}[Z_j] = 0 \cdot 0 = 0$. Thus, the expression collapses to:

$$\mathrm{Var}(\bar{Z}) = \frac{1}{b^2}\sum_{i=1}^{b}\mathbb{E}[Z_i^2] = \frac{1}{b^2}\sum_{i=1}^{b}\mathrm{Var}(Z_i) = \frac{b\,\mathrm{Var}(Z_1)}{b^2} = \frac{\mathrm{Var}(Z_1)}{b} \leq \frac{\sigma^2}{b}.$$

For the vector case in $\mathbb{R}^d$, we apply this argument coordinate-wise. Let $z_i^{(k)}$ be the $k$-th coordinate of the vector $z_i$. The expected squared norm is the sum of the variances of its coordinates:

$$\mathbb{E}\left\|\bar{z}\right\|^2 = \sum_{k=1}^{d}\mathrm{Var}(\bar{z}^{(k)}) = \sum_{k=1}^{d}\frac{\mathrm{Var}(z_1^{(k)})}{b} = \frac{1}{b}\mathbb{E}\left\|z_1\right\|^2 \leq \frac{\sigma^2}{b}.$$

□

**Lemma A.2** (Jensen)**.** *For $\phi : \mathbb{R} \to \mathbb{R}$ convex and $X$ integrable, $\phi(\mathbb{E}X) \leq \mathbb{E}\phi(X)$. In particular $|\mathbb{E}X| \leq \mathbb{E}|X|$ and $\mathbb{E}|X| \leq \sqrt{\mathbb{E}X^2}$.*

*Proof.* By the definition of convexity, for the point $x_0 = \mathbb{E}[X]$, there exists a subgradient $m \in \mathbb{R}$ such that the tangent line globally underestimates the function: $\phi(x) \geq \phi(\mathbb{E}[X]) + m(x - \mathbb{E}[X])$ for all $x$. Substituting the random variable $X$ for $x$ and taking expectations:

$$\mathbb{E}[\phi(X)] \geq \mathbb{E}[\phi(\mathbb{E}[X]) + m(X - \mathbb{E}[X])] = \phi(\mathbb{E}[X]) + m(\mathbb{E}[X] - \mathbb{E}[X]) = \phi(\mathbb{E}[X]).$$

For the first particular case, choose $\phi(x) = |x|$, which is convex. Jensen's inequality directly yields $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$. For the second case, choose $\phi(x) = x^2$, which is strictly convex. Jensen's gives $(\mathbb{E}|X|)^2 \leq \mathbb{E}(|X|^2) = \mathbb{E}(X^2)$. Taking the square root of both sides (since $y \mapsto \sqrt{y}$ is monotonically increasing for $y \geq 0$) yields $\mathbb{E}|X| \leq \sqrt{\mathbb{E}X^2}$. □

**Lemma A.3** (Sphere identity)**.** *Let $u \in \mathbb{S}^{d-1}$ uniform and $v \in \mathbb{R}^d$. Then $\mathbb{E}\left\langle v,\, u\right\rangle^2 = \left\|v\right\|^2/d$.*

*Proof.* We can rewrite the squared inner product as a quadratic form:

$$\mathbb{E}\left\langle v,\, u\right\rangle^2 = \mathbb{E}[(v^\top u)(u^\top v)] = \mathbb{E}[v^\top(uu^\top)v].$$

Since expectation is a linear operator and $v$ is deterministic, we move the expectation to the random vector $u$:

$$\mathbb{E}\left\langle v,\, u\right\rangle^2 = v^\top\mathbb{E}[uu^\top]v.$$

Let $M = \mathbb{E}[uu^\top]$ be the covariance matrix of the uniform distribution on the unit sphere. By rotational symmetry, $M$ must be invariant under any orthogonal transformation, meaning $M = cI_d$ for some constant $c$. To find $c$, we take the trace of both sides:

$$\text{Tr}(M) = \mathbb{E}[\text{Tr}(uu^\top)] = \mathbb{E}[\text{Tr}(u^\top u)] = \mathbb{E}[\|u\|^2].$$

Since $u \in \mathbb{S}^{d-1}$, $\|u\|^2 = 1$ almost surely, so $\text{Tr}(M) = 1$. Concurrently, $\text{Tr}(cI_d) = c \cdot d$. Equating these gives $c = 1/d$. Thus, $M = \frac{1}{d} I_d$. Substituting this back into our quadratic form:

$$\mathbb{E}\,\langle v,\, u\rangle^2 = v^\top \left(\frac{1}{d} I_d\right) v = \frac{\|v\|^2}{d}.$$

□

## B Generalized Smoothness: Implication, Examples, Descent Inequality

**Lemma B.1** (Local Hessian condition implies Assumption 3.1)**.** *If $f$ is twice differentiable and $\left\|\nabla^2 f(x)\right\| \le L_0 + L_1 \|\nabla f(x)\|$ for all $x \in \mathbb{R}^d$, then $f$ satisfies Assumption 3.1 with the same constants.*

*Proof.* Fix $x, y \in \mathbb{R}^d$, set $C = \|y - x\|$, and define the straight-line path between them: $z(t) = x + t(y - x)$ for $t \in [0, 1]$. Let the gradient along this path be $g(t) = \nabla f(z(t))$, and define its magnitude as $r(t) = \|g(t)\|$.

By the multivariate chain rule, the derivative of $g(t)$ with respect to $t$ is:

$$g'(t) = \nabla^2 f(z(t))(y - x).$$

Taking the norm of both sides and applying the local Hessian assumption:

$$\|g'(t)\| \le \left\|\nabla^2 f(z(t))\right\| \|y - x\| \le (L_0 + L_1 \|g(t)\|)C = (L_0 + L_1 r(t))C.$$

Now consider the derivative of the norm $r(t) = (g(t)^\top g(t))^{1/2}$. By the chain rule and Cauchy–Schwarz:

$$r'(t) = \frac{g(t)^\top g'(t)}{\|g(t)\|} \le \frac{\|g(t)\| \, \|g'(t)\|}{\|g(t)\|} = \|g'(t)\|.$$

Combining these inequalities gives a linear differential inequality: $r'(t) \le (L_0 + L_1 r(t))C$. To solve this, we rearrange to isolate $t$:

$$\frac{r'(t)}{L_0/L_1 + r(t)} \le L_1 C.$$

Integrating from 0 to $t$ yields:

$$\ln\left(\frac{L_0/L_1 + r(t)}{L_0/L_1 + r(0)}\right) \le L_1 C t \implies L_0 + L_1 r(t) \le \big(L_0 + L_1 r(0)\big) e^{L_1 C t}.$$

To bound the total change in the gradient, we use the fundamental theorem of calculus:

$$\|\nabla f(y) - \nabla f(x)\| = \left\| \int_0^1 g'(t)\, dt \right\| \le \int_0^1 \|g'(t)\|\, dt.$$

Substitute $\|g'(t)\| \le (L_0 + L_1 r(t))C$ and use our exponential bound:

$$\begin{aligned}
\|\nabla f(y) - \nabla f(x)\| &\le \int_0^1 \big(L_0 + L_1 r(0)\big) e^{L_1 C t} C\, dt \\
&= \big(L_0 + L_1 \|\nabla f(x)\|\big) C \left[\frac{e^{L_1 C t}}{L_1 C}\right]_{t=0}^{t=1} \\
&= \frac{L_0 + L_1 \|\nabla f(x)\|}{L_1} \big(e^{L_1 C} - 1\big).
\end{aligned}$$

Finally, using the basic calculus inequality $e^z - 1 \le z e^z$ for $z \ge 0$ (with $z = L_1 C$):

$$\|\nabla f(y) - \nabla f(x)\| \le (L_0 + L_1 \|\nabla f(x)\|)\, C\, e^{L_1 C}.$$

□

**Example B.2.**
*Polynomials.* For $f(x) = \|x\|^p$ with $p \geq 2$, calculating the derivatives yields $\left\|\nabla^2 f(x)\right\| \leq p(p-1)\|x\|^{p-2}$ and $\|\nabla f(x)\| = p\|x\|^{p-1}$. The ratio $\left\|\nabla^2 f(x)\right\| / \|\nabla f(x)\|$ scales as $1/\|x\|$, meaning for large $x$ it is bounded by a constant, satisfying the local Hessian condition.

*Exponentials.* For $f(x) = e^{a^\top x}$, the Hessian is $\nabla^2 f(x) = aa^\top e^{a^\top x}$, so its spectral norm is $\|a\|^2 e^{a^\top x}$. The gradient is $\nabla f(x) = ae^{a^\top x}$, so its norm is $\|a\| e^{a^\top x}$. Thus, $\left\|\nabla^2 f(x)\right\| = \|a\| \cdot \|\nabla f(x)\|$, exactly matching the local condition with $L_0 = 0, L_1 = \|a\|$.

Both functions are excluded by the standard global Lipschitz form of $(L_0, L_1)$-smoothness but are perfectly admitted by Assumption 3.1.

**Lemma B.3** (Generalized descent inequality)**.** *Under Assumption 3.1, the inequality* (2) *holds for all* $x, y \in \mathbb{R}^d$.

*Proof.* By the fundamental theorem of calculus applied to the scalar function $h(t) = f(x+t(y-x))$, the exact function difference is:

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + t(y-x)),\, y - x\rangle \, dt.$$

Subtract the linear approximation $\langle \nabla f(x),\, y - x\rangle$ (which integrates over $[0,1]$ as a constant):

$$f(y) - f(x) - \langle \nabla f(x),\, y - x\rangle = \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x),\, y - x\rangle \, dt.$$

Apply the Cauchy–Schwarz inequality to bound the inner product:

$$\leq \int_0^1 \|\nabla f(x + t(y-x)) - \nabla f(x)\| \, \|y - x\| \, dt.$$

Next, we invoke the generalized smoothness condition (Assumption 3.1) on the points $x + t(y-x)$ and $x$. The distance between them is $\|t(y-x)\| = t\|y-x\|$. This gives:

$$\leq \int_0^1 \big(L_0 + L_1 \|\nabla f(x)\|\big) \cdot \big(t\|y-x\|\big) \cdot e^{L_1 t \|y-x\|} \cdot \|y-x\| \, dt.$$

We factor out everything independent of the integration variable $t$:

$$= \big(L_0 + L_1 \|\nabla f(x)\|\big) \|y-x\|^2 \int_0^1 t e^{L_1 \|y-x\| t} \, dt.$$

We bound the integral using integration by parts, giving $\int_0^1 te^{at}\,dt = \frac{e^a(a-1)+1}{a^2} \leq \frac{1}{2}e^a$ for $a \geq 0$. Setting $a = L_1\|y-x\|$:

$$\leq \frac{1}{2}\big(L_0 + L_1\|\nabla f(x)\|\big) e^{L_1\|y-x\|} \|y-x\|^2.$$

□

## C Proof of Lemma 3.3: General Descent Bound

*Proof.* We evaluate the generalized descent inequality (Lemma B.3) at the perturbed points $x_t^\pm = x_t \pm \eta_t s_t$. The distance is $\|y - x\| = \eta_t \|s_t\|$. Substituting this in directly gives:

$$f(x_t^\pm) \leq f(x_t) \pm \eta_t \langle \nabla f(x_t),\, s_t\rangle + \tfrac{1}{2}(L_0 + L_1\|\nabla f(x_t)\|)\, e^{L_1 \eta_t \|s_t\|} \eta_t^2 \|s_t\|^2.$$

By Assumption 3.2, the search direction has strictly bounded support, $\|s_t\| \leq R$ almost surely. Therefore, we can uniformly bound the exponential factor: $e^{L_1\eta_t\|s_t\|} \leq e^{L_1\eta_t R} =: \kappa_{\mathcal{D}}$.

To account for the fact that the algorithm uses noisy function evaluations $M_t^\pm$ instead of $f(x_t^\pm)$, we add the absolute error $|M_t^\pm - f(x_t^\pm)|$ to both sides of the upper bound for $M_t^\pm$:

$$M_t^\pm \leq f(x_t) \pm \eta_t \langle \nabla f(x_t),\, s_t\rangle + \tfrac{\kappa_{\mathcal{D}}}{2}(L_0 + L_1\|\nabla f(x_t)\|)\, \eta_t^2 \|s_t\|^2 + |M_t^\pm - f(x_t^\pm)|.$$

The logic of the random search update dictates that we step to $x_t^+$ if $M_t^+ \le M_t^-$, and to $x_t^-$ otherwise. Case 1: Suppose $M_t^+ \le M_t^-$, meaning the algorithm chooses $x_{t+1} = x_t^+$. Because $M_t^+$ is the smaller of the two, $M_t^+ \le \frac{1}{2}(M_t^+ + M_t^-)$. Using our analytic bounds for $M_t^+$ and $M_t^-$ cancels the linear inner product term, establishing a bound that isolates the negative absolute inner product:

$$M_t^+ \le f(x_t) - \eta_t |\langle \nabla f(x_t),\, s_t \rangle| + \tfrac{\kappa_{\mathcal{D}}}{2}(L_0 + L_1 \|\nabla f(x_t)\|)\, \eta_t^2 \|s_t\|^2 + \sum_{i \in \{\pm\}} |M_t^i - f(x_t^i)|.$$

The true function value at the chosen point is bounded by the noisy estimate plus its noise margin: $f(x_{t+1}) = f(x_t^+) \le M_t^+ + |M_t^+ - f(x_t^+)|$. Substituting the bound for $M_t^+$:

$$f(x_{t+1}) \le f(x_t) - \eta_t |\langle \nabla f(x_t),\, s_t \rangle| + \tfrac{\kappa_{\mathcal{D}}}{2}(L_0 + L_1 \|\nabla f(x_t)\|)\, \eta_t^2 \|s_t\|^2 + 2 \sum_{i \in \{\pm\}} |M_t^i - f(x_t^i)|.$$

Case 2: If $M_t^- < M_t^+$, the chosen point is $x_{t+1} = x_t^-$. The exact same symmetrical logic applies, leading to the identical inequality.

Taking expectations on both sides over the random direction $s_t$ and the noise, we utilize the directional moments: $\mathbb{E}\|s_t\|^2 = 1$ and $\mathbb{E}|\langle \nabla f(x_t),\, s_t \rangle| \ge \mu_{\mathcal{D}} \mathbb{E}\|\nabla f(x_t)\|$. Furthermore, the algorithm's decision rule is invariant to uniform translations of the function estimates ((3)). Thus, we can replace the sum of absolute errors with the pure difference error $\Delta_t = (M_t^+ - M_t^-) - (f(x_t^+) - f(x_t^-))$:

$$\mathbb{E} f(x_{t+1}) \le \mathbb{E} f(x_t) - \eta_t \mu_{\mathcal{D}} \mathbb{E}\|\nabla f(x_t)\| + \tfrac{\kappa_{\mathcal{D}}}{2}\eta_t^2 \big(L_0 + L_1 \mathbb{E}\|\nabla f(x_t)\|\big) + 2\mathbb{E}|\Delta_t|.$$

Regrouping the terms proportional to the gradient norm:

$$\mathbb{E} f(x_{t+1}) \le \mathbb{E} f(x_t) - \eta_t \big(\mu_{\mathcal{D}} - \tfrac{\kappa_{\mathcal{D}} L_1}{2}\eta_t\big) \mathbb{E}\|\nabla f(x_t)\| + \tfrac{\kappa_{\mathcal{D}} L_0}{2}\eta_t^2 + 2\mathbb{E}|\Delta_t|.$$

If we constrain the step size such that $\eta_t \le \mu_{\mathcal{D}}/(\kappa_{\mathcal{D}} L_1)$, then $\frac{\kappa_{\mathcal{D}} L_1}{2}\eta_t \le \mu_{\mathcal{D}}/2$. Thus, the descent coefficient $\mu_{\mathcal{D}} - \frac{\kappa_{\mathcal{D}} L_1}{2}\eta_t$ is strictly bounded from below by $\mu_{\mathcal{D}}/2$, yielding (4). □

# D Stochastic Subsampling: Proofs

## D.1 Proof of Theorem 4.2

Let $X_j = f_{\xi_j}(x_t^+) - f_{\xi_j}(x_t^-) - (f(x_t^+) - f(x_t^-))$. Then $\mathbb{E}X_j = 0$. By the standard inequality $(a-b)^2 \le 2a^2 + 2b^2$ and Assumption 4.1, the variance is bounded by $\mathbb{E}X_j^2 \le 4\sigma_0^2$. By Lemma A.1 and Lemma A.2, the mean absolute error is: $\mathbb{E}|\Delta_t| \le \sqrt{\mathbb{E}\big(\frac{1}{b}\sum_j X_j\big)^2} \le 2\sigma_0/\sqrt{b}$.

Plugging this into the general descent bound (4) and telescoping the sum over iterations $t = 1, \dots, T$:

$$\tfrac{\mu_{\mathcal{D}}}{2T} \sum_{t=1}^{T} \mathbb{E}\|\nabla f(x_t)\| \le \tfrac{F_0}{\eta T} + \tfrac{\kappa_{\mathcal{D}} L_0}{2}\eta + \tfrac{4\sigma_0}{\eta\sqrt{b}}.$$

Optimizing the step size $\eta$ subject to $\eta \le \mu_{\mathcal{D}}/(\kappa_{\mathcal{D}} L_1)$ and using $\mu_{\mathcal{D}} \asymp 1/\sqrt{d}$, $\kappa_{\mathcal{D}} \le e$, we choose $T = \mathcal{O}(dL_1/\varepsilon + dL_0F_0/\varepsilon^2)$ and $b = \mathcal{O}((dL_0\sigma_0)^2/\varepsilon^4)$. The total oracle complexity is $\textsc{Calls} = 2bT = \widetilde{\mathcal{O}}(d^3\sigma_0^2/\varepsilon^6)$. □

## D.2 Proof of Theorem 4.5

**Lemma D.1** (Pointwise smoothness implies Assumption 4.3)**.** *If each $f_\xi$ satisfies Assumption 3.1 a.s. with constants $(L_0, L_1)$ and Assumption 4.4 holds, then Assumption 4.3 holds with constants $(L_0 + L_1\sigma, L_1)$.*

*Proof.* By assumption, the generalized smoothness inequality holds per-sample:

$$\|\nabla f_\xi(y) - \nabla f_\xi(x)\| \le (L_0 + L_1 \|\nabla f_\xi(x)\|)\, e^{L_1\|y-x\|}\, \|y - x\|.$$

Take the expectation $\mathbb{E}_\xi$ on both sides. To evaluate the right side, we use Jensen's inequality (Lemma A.2) on the norm, combined with the variance assumption $\mathbb{E}\left\|\nabla f_\xi(x) - \nabla f(x)\right\|^2 \le \sigma^2$:

$$\begin{aligned}\mathbb{E}\left\|\nabla f_\xi(x)\right\| &\le \sqrt{\mathbb{E}\left\|\nabla f_\xi(x)\right\|^2} = \sqrt{\mathbb{E}\left\|\nabla f_\xi(x) - \nabla f(x) + \nabla f(x)\right\|^2} \\ &= \sqrt{\mathbb{E}\left\|\nabla f_\xi(x) - \nabla f(x)\right\|^2 + \left\|\nabla f(x)\right\|^2} \le \sqrt{\sigma^2 + \left\|\nabla f(x)\right\|^2}.\end{aligned}$$

Using the algebraic inequality $\sqrt{a^2+b^2} \le a + b$ for non-negative $a, b$:

$$\mathbb{E}\left\|\nabla f_\xi(x)\right\| \le \sigma + \left\|\nabla f(x)\right\|.$$

Substituting this back into our expected smoothness bound yields the new constants $(L_0 + L_1\sigma, L_1)$. □

**Corollary D.2.** *Under Assumption 4.3, $f$ itself satisfies Assumption 3.1 with the same constants.*

*Proof.* Since $f(x) = \mathbb{E}_\xi[f_\xi(x)]$, moving the norm inside the expectation via Jensen's inequality yields:

$$\left\|\nabla f(y) - \nabla f(x)\right\| = \left\|\mathbb{E}_\xi[\nabla f_\xi(y) - \nabla f_\xi(x)]\right\| \le \mathbb{E}_\xi\left\|\nabla f_\xi(y) - \nabla f_\xi(x)\right\|.$$

Applying the bound from Assumption 4.3 immediately gives $\left(L_0 + L_1\left\|\nabla f(x)\right\|\right) e^{L_1\|y-x\|}\ \|y - x\|$. □

**Lemma D.3** (Difference approximation error)**.** *Under Assumption 3.2 (with $\mathcal{D}$ uniform on the unit sphere), Assumption 4.3, and Assumption 4.4,*

$$\mathbb{E}\big[|\Delta_t| \,\big|\, x_t\big] \le \tfrac{2\eta\sigma}{\sqrt{db}} + 2\kappa_{\mathcal{D}}\eta^2\big(L_0 + L_1\,\mathbb{E}[\|\nabla f(x_t)\| \mid x_t]\big).$$

*Proof.* Let the empirical gradient from the minibatch be $\bar{g}(x) = \frac{1}{b}\sum_{j=1}^{b}\nabla f_{\xi_j}(x)$. By the fundamental theorem of calculus applied to the path $x_t + u\eta s_t$ for $u \in [-1, 1]$:

$$f_\xi(x_t^+) - f_\xi(x_t^-) = \int_{-1}^{1}\langle\nabla f_\xi(x_t + u\eta s_t),\, \eta s_t\rangle\ du.$$

Averaging over the minibatch $\xi_j$ and subtracting the exact population difference gives the error $\Delta_t$:

$$\Delta_t = \eta\int_{-1}^{1}\langle\bar{g}(y_u) - \nabla f(y_u),\, s_t\rangle\ du, \qquad \text{where } y_u := x_t + u\eta s_t. \tag{6}$$

We strategically decompose the integrand $\bar{g}(y_u) - \nabla f(y_u)$ into three terms:

$$\bar{g}(y_u) - \nabla f(y_u) = (\mathrm{I}) + (\mathrm{II}) - (\mathrm{III}),$$

where:

- $(\mathrm{I}) = \bar{g}(x_t) - \nabla f(x_t)$ is the base noise at the current iterate.
- $(\mathrm{II}) = \bar{g}(y_u) - \bar{g}(x_t)$ is the stochastic variation along the search vector.
- $(\mathrm{III}) = \nabla f(y_u) - \nabla f(x_t)$ is the population variation.

*Term (I).* Using Cauchy–Schwarz and the Sphere identity (Lemma A.3):

$$\mathbb{E}\big[|\,\langle(\mathrm{I}),\, s_t\rangle\,| \,\big|\, x_t, \xi\big] \le \frac{\|(\mathrm{I})\|}{\sqrt{d}} = \frac{\|\bar{g}(x_t) - \nabla f(x_t)\|}{\sqrt{d}}.$$

Taking the expectation over the minibatch $\xi$, we apply Lemma A.2 and Lemma A.1:

$$\mathbb{E}[\|\bar{g}(x_t) - \nabla f(x_t)\| \mid x_t] \le \sqrt{\mathbb{E}[\|\bar{g}(x_t) - \nabla f(x_t)\|^2]} \le \frac{\sigma}{\sqrt{b}}.$$

Thus, $\mathbb{E}[|\,\langle(\mathrm{I}),\, s_t\rangle\,| \mid x_t] \le \frac{\sigma}{\sqrt{db}}$.

*Term (II).* Conditional on $x_t$ and $s_t$, the samples are i.i.d. By the triangle inequality on the sum, we bound the expected norm using the average generalized smoothness (Assumption 4.3):

$$\begin{aligned}\mathbb{E}\big[\|\bar{g}(y_u) - \bar{g}(x_t)\| \mid x_t, s_t\big] &\le \mathbb{E}_\xi \|\nabla f_\xi(y_u) - \nabla f_\xi(x_t)\| \\ &\le (L_0 + L_1 \|\nabla f(x_t)\|)\, e^{L_1\|y_u - x_t\|} \, \|y_u - x_t\|.\end{aligned}$$

Substituting $\|y_u - x_t\| = |u|\eta\,\|s_t\|$ and applying Cauchy–Schwarz:

$$\mathbb{E}[|\langle (\mathrm{II}),\, s_t\rangle| \mid x_t, s_t] \le (L_0 + L_1 \|\nabla f(x_t)\|)\, e^{L_1|u|\eta\|s_t\|}\, |u|\eta\,\|s_t\|^2.$$

*Term (III).* By Corollary D.2, the population gradient satisfies the exact same smoothness bound:

$$\|(\mathrm{III})\| \le (L_0 + L_1 \|\nabla f(x_t)\|)\, e^{L_1|u|\eta\|s_t\|}\, |u|\eta\,\|s_t\|,$$

which gives the same bound for $|\langle (\mathrm{III}),\, s_t\rangle|$ as for (II).

*Combining and Integrating.* Since $s_t$ is drawn uniformly from the unit sphere, $\|s_t\| = 1$. The exponential term simplifies: $e^{L_1|u|\eta\|s_t\|} \le e^{L_1\eta} = \kappa_{\mathcal{D}}$. Now we integrate over $u \in [-1, 1]$. For Term (I), $\int_{-1}^{1} \frac{\sigma}{\sqrt{db}}\, du = 2\frac{\sigma}{\sqrt{db}}$. For Terms (II) and (III), we integrate the absolute value: $\int_{-1}^{1} |u|\, du = [\frac{u^2}{2}]_0^1 - [\frac{u^2}{2}]_{-1}^0 = 1$. Summing the contributions yields the final bound:

$$\mathbb{E}\big[|\Delta_t| \,\big|\, x_t\big] \le 2\eta\frac{\sigma}{\sqrt{db}} + 2\eta\left(\kappa_{\mathcal{D}}\eta\big(L_0 + L_1\,\mathbb{E}[\|\nabla f(x_t)\| \mid x_t]\big)\right)\cdot 1.$$

□

*Proof of Theorem 4.5.* Substitute Lemma D.3 into (4): for $\eta \le \mu_{\mathcal{D}}/(\kappa_{\mathcal{D}} L_1)$,

$$\begin{aligned}\mathbb{E}f(x_{t+1}) &\le \mathbb{E}f(x_t) - \tfrac{\mu_{\mathcal{D}}}{2}\eta\mathbb{E}\,\|\nabla f(x_t)\| + \tfrac{\kappa_{\mathcal{D}} L_0}{2}\eta^2 + \tfrac{4\eta\sigma}{\sqrt{db}} + 4\kappa_{\mathcal{D}}\eta^2(L_0 + L_1\mathbb{E}\,\|\nabla f(x_t)\|) \\ &= \mathbb{E}f(x_t) - \eta\big(\tfrac{\mu_{\mathcal{D}}}{2} - 4\kappa_{\mathcal{D}} L_1\eta\big)\mathbb{E}\,\|\nabla f(x_t)\| + \tfrac{9}{2}\kappa_{\mathcal{D}} L_0\eta^2 + \tfrac{4\eta\sigma}{\sqrt{db}}.\end{aligned}$$

With $\eta \le \mu_{\mathcal{D}}/(16\kappa_{\mathcal{D}} L_1)$, the gradient coefficient is $\ge \mu_{\mathcal{D}}/4$. Telescoping and dividing by $T$ and $\mu_{\mathcal{D}}\eta/4$:

$$\tfrac{1}{T}\sum_{t=1}^{T}\mathbb{E}\,\|\nabla f(x_t)\| \le \tfrac{4F_0}{\mu_{\mathcal{D}}\eta T} + \tfrac{18\kappa_{\mathcal{D}} L_0}{\mu_{\mathcal{D}}}\eta + \tfrac{16\sigma}{\mu_{\mathcal{D}}\sqrt{db}}.$$

Optimizing $\eta = \min\{\mu_{\mathcal{D}}/(16\kappa_{\mathcal{D}} L_1), \sqrt{F_0/(\kappa_{\mathcal{D}} L_0 T)}\}$ yields the rate of Theorem 4.5. With $\mu_{\mathcal{D}} \asymp 1/\sqrt{d}$ and $\kappa_{\mathcal{D}} \le e$: $T = \Theta(dL_1/\varepsilon + dL_0F_0/\varepsilon^2)$ controls the first two terms; $b = \Theta(\sigma^2/\varepsilon^2)$ controls the third. Total $\textsc{Calls} = 2bT = \widetilde{\mathcal{O}}(dL_0F_0\sigma^2/\varepsilon^4)$. □

# E Proof of Theorem 5.1: Variance Reduction

*Proof.* Let $\tilde{x}_t$ denote the exact snapshot evaluated at the beginning of the current epoch (the iterate from the most recent full pass, which is at most $m-1$ iterations earlier). We define the per-sample variance reduction error as $X_j = f_{\xi_j}(x_t^+) - f_{\xi_j}(x_t^-) - (f(x_t^+) - f(x_t^-))$. Translation symmetry implies that shifting each $f_{\xi_j}$ by the constant $f_{\xi_j}(\tilde{x}_t)$ does not change the behavior of the algorithm.

Using the $G$-Lipschitz property of each component $f_i$, we bound the variance:

$$\mathbb{E}X_j^2 \le 8G^2\big(\mathbb{E}\,\big\|x_t^+ - \tilde{x}_t\big\|^2 + \mathbb{E}\,\big\|x_t^- - \tilde{x}_t\big\|^2\big).$$

Because the step size is fixed to $\eta$, iterates drift by at most $m\eta$ from the snapshot (meaning $\|x_t - \tilde{x}_t\| \le m\eta$). Thus, $\mathbb{E}X_j^2 \le 8G^2 \cdot 2(m\eta)^2 \le (4Gm\eta)^2$. By Lemma A.1 and Lemma A.2, the mean absolute error is bounded by $\mathbb{E}|\Delta_t| \le 4\eta Gm/\sqrt{b}$.

Plugging this modified noise term into (4) and proceeding exactly as in Appendix D.1 (replacing $4\sigma_0/\sqrt{b}$ with $8\eta Gm/\sqrt{b}$):

$$\tfrac{1}{T}\sum_t \mathbb{E}\,\|\nabla f(x_t)\| = \tfrac{1}{\mu_{\mathcal{D}}}\,\mathcal{O}\left(\tfrac{\kappa_{\mathcal{D}} L_1 F_0}{\mu_{\mathcal{D}} T} + \sqrt{\tfrac{\kappa_{\mathcal{D}} L_0 F_0}{T}} + \tfrac{Gm}{\sqrt{b}}\right).$$

Assuming standard dimension scaling $\mu_{\mathcal{D}} \asymp 1/\sqrt{d}$, hitting $\varepsilon$-accuracy requires $T = \Theta(dL_0F_0/\varepsilon^2 + dL_1F_0/\varepsilon)$ iterations and minibatch size $b(m) = \Theta(dG^2m^2/\varepsilon^2)$.

The total evaluation cost per block of $m$ iterations involves one full dataset pass ($n$ calls) and $m-1$ minibatch steps of size $b(m)$. The amortized cost function is: $\textsc{Calls}(m) = \frac{T}{m}(n + (m-1)b(m))$. Optimizing this over the epoch length $m$ subject to $b(m) \le n$ yields the optimal block size $m^\star = \min\{(n\varepsilon^2/dG^2)^{1/3}, (n\varepsilon^2/dG^2)^{1/2}\}$. Substitution of $m^\star$ yields the stated $\widetilde{\mathcal{O}}(d^{4/3}n^{2/3}/\varepsilon^{8/3})$ complexity. □

## F Proof of Theorem 6.2: Inexact Comparison Feedback

*Proof.* Setting the estimates $M_t^{\pm} = h(x_t^{\pm})$ in our central inequality (4), Assumption 6.1 configured with points $(x, y) = (x_t^+, x_t^-)$ directly ensures bounded difference error: $\mathbb{E}|\Delta_t| \le \delta$. Substituting this into the descent lemma:

$$\mathbb{E}f(x_{t+1}) \le \mathbb{E}f(x_t) - \tfrac{\mu_{\mathcal{D}}}{2}\eta\mathbb{E}\left\|\nabla f(x_t)\right\| + \tfrac{\kappa_{\mathcal{D}}L_0}{2}\eta^2 + 2\delta.$$

Telescoping the sum over $T$ steps and dividing by the learning rate factor $\eta T$:

$$\tfrac{1}{T}\sum_t \mathbb{E}\left\|\nabla f(x_t)\right\| \le \tfrac{1}{\mu_{\mathcal{D}}}\big(2F_0/(\eta T) + \kappa_{\mathcal{D}}L_0\eta + 4\delta/\eta\big).$$

Optimizing $\eta$ subject to $\eta \le \mu_{\mathcal{D}}/(\kappa_{\mathcal{D}}L_1)$ and adopting $\mu_{\mathcal{D}} \asymp 1/\sqrt{d}$ with $T = \mathcal{O}(dL_1/\varepsilon + dL_0F_0/\varepsilon^2)$ yields the rate $\varepsilon + \mathcal{O}(\sqrt{d\delta})$. □

## G Failure of Classical Momentum

We consider three momentum variants adapted to function differences. Let $s_t \sim \mathcal{D}$, with the stochastic step difference $\Delta f_\xi^t = f_\xi(x_t^+) - f_\xi(x_t^-)$, and the population difference $\Delta f^t = f(x_t^+) - f(x_t^-)$:

$$\begin{aligned}
\text{(Heavy-Ball)} \quad & M_t = (1-\beta)M_{t-1} + \beta\,\Delta f_\xi^t, \\
\text{(MVR)} \quad & M_t = (1-\beta)(M_{t-1} + \Delta f_\xi^t - \Delta f_\xi^{t-1}) + \beta\,\Delta f_\xi^t, \\
\text{(Transport)} \quad & M_t = (1-\beta)M_{t-1} + \beta\big(f_\xi(\tilde{x}_t^+) - f_\xi(\tilde{x}_t^-)\big), \quad \tilde{x}_t^{\pm} = x_t^{\pm} + \tfrac{1-\beta}{\beta}(x_t^{\pm} - x_{t-1}^{\pm}).
\end{aligned}$$

**Heavy-Ball: bias-variance breakdown.** We define the tracking error of the momentum buffer relative to the true population difference as $e_t = M_t - \Delta f^t$. We decompose this recursively:

$$\begin{aligned}
e_t &= (1-\beta)M_{t-1} + \beta\Delta f_\xi^t - \Delta f^t \\
&= (1-\beta)(e_{t-1} + \Delta f^{t-1}) + \beta(\Delta f^t + v_t) - \Delta f^t \\
&= (1-\beta)e_{t-1} + (1-\beta)(\Delta f^{t-1} - \Delta f^t) + \beta v_t.
\end{aligned}$$

Defining the population drift $b_t = \Delta f^t - \Delta f^{t-1}$ and the noise $v_t = \Delta f_\xi^t - \Delta f^t$, we write:

$$e_t = (1-\beta)(e_{t-1} + b_t) + \beta v_t.$$

Assuming $G$-Lipschitz components, the function difference between consecutive steps (separated by distance bounded by $2\eta$) is strictly bounded, giving $|b_t| \le 2G\eta$. The variance of the noise is bounded by $\mathbb{E}v_t^2 \le 4G^2\eta^2$.

Unrolling the recurrence relation for $e_t$ gives:

$$e_t = (1-\beta)^t e_0 + \sum_{k=1}^{t}(1-\beta)^{t-k+1}b_k + \beta\sum_{k=1}^{t}(1-\beta)^{t-k}v_k.$$

Taking the absolute value, passing to expectations, and applying the triangle inequality:

$$\mathbb{E}|e_t| \le (1-\beta)^t\mathbb{E}|e_0| + \sum_{k=1}^{t}(1-\beta)^{t-k+1}\mathbb{E}|b_k| + \beta\mathbb{E}\left|\sum_{k=1}^{t}(1-\beta)^{t-k}v_k\right|.$$

Evaluating the geometric sum for the bias term yields an upper bound of $\frac{1-\beta}{\beta}2G\eta$. Since the noise components $v_k$ are zero-mean and mutually independent, the standard deviation of the final sum scales as $\sqrt{\sum(1-\beta)^{2(t-k)}\mathbb{E}v_k^2} \leq \frac{1}{\sqrt{\beta}}2G\eta$. Multiplied by the front factor $\beta$, the aggregated noise term evaluates to $\sqrt{\beta}\cdot 2G\eta$. Thus:

$$\mathbb{E}|e_t| \leq (1-\beta)^t\mathbb{E}|e_0| + \tfrac{(1-\beta)\cdot 2G\eta}{\beta} + 2G\eta\sqrt{\beta}.$$

Crucially, both the bias and variance terms scale linearly as $\Theta(\eta)$. In traditional gradient momentum, the noise scales as $\Theta(1)$, allowing $\beta < 1$ to trade bias for reduced variance. Here, optimizing $\beta$ cannot decouple the $\eta$-dependence. The same conclusion holds for MVR, where the tracking correction limits the bias, but both correction and variance terms remain structurally tied to $\mathcal{O}(G\eta)$, leaving no improvement over the non-momentum baseline.

The implicit transport momentum fails for a different reason. It requires extrapolated states $\tilde{x}_t^{\pm} = x_t^{\pm} + \frac{1-\beta}{\beta}(x_t^{\pm} - x_{t-1}^{\pm})$ which introduces an additional displacement of order $\frac{1-\beta}{\beta}\eta$. Since the variance of the difference $f_\xi(\tilde{x}_t^+) - f_\xi(\tilde{x}_t^-)$ depends intimately on the proximity of these states (the distance between points being evaluated), this spatial extrapolation directly increases the spatial distance by a factor of $\frac{1}{\beta}$. Consequently, the overall noise term scales as $\mathcal{O}(\frac{\sigma}{\sqrt{\beta}})$, perfectly offsetting the potential $\sqrt{\beta}$ gain from temporal averaging. Once again, the best choice is $\beta = 1$, implying that not using momentum is preferable.

# H Sub-Gaussian Directions

The body of the paper assumes the directional distribution has strictly bounded support (Assumption 3.2, item 3). We rigorously demonstrate here that all main results carry over to a strictly larger class of distributions whose radial part has sub-Gaussian tails, with zero change to asymptotic complexity.

## H.1 The sub-Gaussian condition

We replace item 3 of Assumption 3.2 by the following weighted-moment condition.

3′. (*Sub-Gaussian radius.*) There exist constants $C_0, C_1, C_2 > 0$ such that for all $\alpha \in [0,1]$ and $k \in \{0,1,2\}$,

$$\mathbb{E}_{s\sim\mathcal{D}}\big[\|s\|^k\ e^{\alpha\|s\|}\big] \leq C_k.$$

A sufficient condition for item 3′ is that $\|s\|$ is sub-Gaussian around $\mathbb{E}\,\|s\|$, meaning $\Pr(|\,\|s\| - \mathbb{E}\,\|s\|\,| \geq u) \leq 2e^{-cdu^2}$ for some constant $c > 0$. The bounded-support condition is merely the degenerate special case where the sub-Gaussian constant is zero.

**Lemma H.1** (Gaussian directions satisfy item 3′)**.** *For $s \sim \mathcal{N}(0, I_d/d)$ and any $\alpha \geq 0$, $\mathbb{E}[e^{\alpha\|s\|}] \leq 2e^{\alpha+\alpha^2/(2d)}$. For $\alpha \in [0,1]$ and $k \in \{0,1,2\}$, $\mathbb{E}[\|s\|^k e^{\alpha\|s\|}] \leq C_k$ for an absolute constant $C_k$.*

*Proof.* For a scaled Gaussian vector $s \sim \mathcal{N}(0, I_d/d)$, the modified variable $\sqrt{d}\ \|s\|$ represents the Euclidean norm of a standard normal random vector in $\mathbb{R}^d$. This follows a $\chi$-distribution. By standard concentration of measure [Vershynin, 2018, Theorem 3.1.1], the tail decays sub-Gaussianly:

$$\Pr(\sqrt{d}\ \|s\| \geq \sqrt{d} + t) \leq e^{-t^2/2}.$$

Dividing the inequality inside the probability by $\sqrt{d}$ gives $\Pr(\|s\| \geq 1 + t/\sqrt{d}) \leq e^{-t^2/2}$. Letting $Z = \|s\| - 1$, the bound becomes $\Pr(Z \geq u) \leq e^{-du^2/2}$ for $u \geq 0$. This implies that the positive part $Z_+ := \max(Z, 0)$ is sub-Gaussian with variance proxy $1/d$. The Moment Generating Function (MGF) is therefore bounded: $\mathbb{E}[e^{\alpha Z_+}] \leq e^{\alpha^2/(2d)}$ for $\alpha \geq 0$.

For the full exponential expectation, since $Z \geq -1$ almost surely:

$$\mathbb{E}[e^{\alpha\|s\|}] = e^{\alpha}\,\mathbb{E}[e^{\alpha Z}] \leq e^{\alpha}(1 + \mathbb{E}[e^{\alpha Z_+}]) \leq 2e^{\alpha+\alpha^2/(2d)}.$$

The polynomial-weighted moments follow by an application of the Cauchy–Schwarz inequality:

$$\mathbb{E}[\|s\|^k e^{\alpha\|s\|}] \le \sqrt{\mathbb{E}\|s\|^{2k}}\sqrt{\mathbb{E}[e^{2\alpha\|s\|}]}.$$

The first factor is a standard $\chi$-distribution moment (bounded by an absolute constant for $k \le 2$), and the second is bounded by our MGF estimate $e^{2\alpha+2\alpha^2/d}$. For $\alpha \in [0,1]$, this product is globally bounded by a uniform constant $C_k$. □

### H.2 Descent bound and main theorems

**Lemma H.2** (Descent under sub-Gaussian directions)**.** *Suppose $f$ satisfies Assumption 3.1 and $s_t \sim \mathcal{D}$ satisfies items 1 and 2 of Assumption 3.2 together with item 3′, and $L_1\eta \le 1$. Then the descent bound* (4) *holds with $\kappa_{\mathcal{D}}$ replaced by $\widetilde{\kappa_{\mathcal{D}}} := C_2$ (an absolute constant), provided $\eta \le \mu_{\mathcal{D}}/(\widetilde{\kappa_{\mathcal{D}}} L_1)$.*

*Proof.* Reviewing the proof of Lemma 3.3, the strict bounded support condition $\|s_t\| \le R$ was invoked purely to limit the growth of two specific expectations that emerged from the generalized descent inequality: $\mathbb{E}[e^{L_1\eta\|s_t\|}\|s_t\|^2]$ and $\mathbb{E}[e^{L_1\eta\|s_t\|}\|s_t\|^2\|\nabla f(x_t)\|]$.

Under the sub-Gaussian regime, provided the learning rate is chosen such that $L_1\eta \le 1$, we can map the exponent scaling to $\alpha = L_1\eta \in [0,1]$. Using item 3′ evaluated at $k = 2$, the first expectation integrates directly to a finite upper bound $\widetilde{\kappa_{\mathcal{D}}} = C_2$. The second expectation factorizes into $\|\nabla f(x_t)\|\,\mathbb{E}[e^{L_1\eta\|s_t\|}\|s_t\|^2]$ due to the explicit independence of the search direction $s_t$ and the current iterate $x_t$. This yields exactly the same bound format as before, just substituting $\widetilde{\kappa_{\mathcal{D}}}$ for $\kappa_{\mathcal{D}}$. □

**Corollary H.3** (Sub-Gaussian rates)**.** *Under items 1, 2, 3′ of Assumption 3.2 and the additional restriction $L_1\eta \le 1$, all results of Theorem 4.2, Theorem 4.5, Theorem 5.1, and Theorem 6.2 hold with $\kappa_{\mathcal{D}}$ replaced by $\widetilde{\kappa_{\mathcal{D}}}$, an absolute constant. The asymptotic complexities are unchanged. In particular, this covers Gaussian directions $s \sim \mathcal{N}(0, I_d/d)$ via Lemma H.1.*

*Proof.* The fundamental descent bound is seamlessly provided by Lemma H.2. When applying this structure to the difference approximation error in Lemma D.3, the strictly bounded support was only used to integrate the expected local variation over the path segment $u \in [-1,1]$, specifically the term $\mathbb{E}[\|s_t\|^2 e^{L_1|u|\eta\|s_t\|}]$. Setting our scaling factor to $\alpha = L_1|u|\eta \le 1$, item 3′ uniformly bounds this specific integral by the same absolute constant $C_2$.

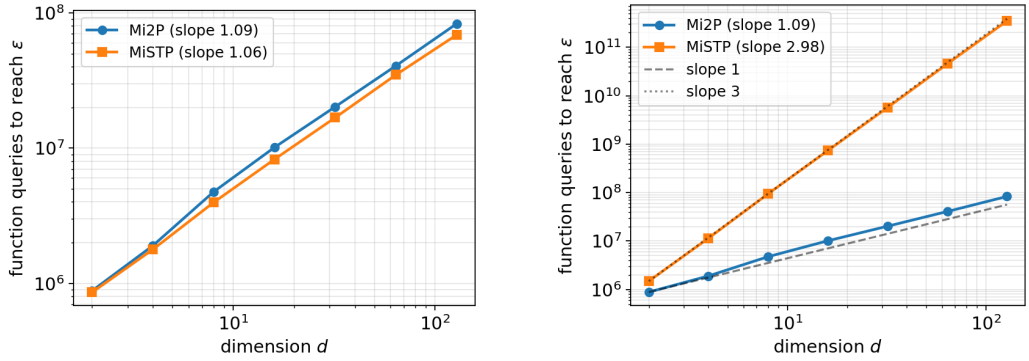
All downstream theorems rely exclusively on the primary moment conditions ($\mathbb{E}\|s_t\|^2 = 1$ and $\mathbb{E}|\langle g, s_t\rangle| \ge \mu_{\mathcal{D}}\|g\|$), combined with the established descent bound. In the natural step-size regime employed by these convergence theorems ($\eta \asymp \mu_{\mathcal{D}}/L_1 \asymp 1/(L_1\sqrt{d})$ for standard isotropic distributions $\mathcal{D}$), the supplementary condition $L_1\eta \le 1$ natively holds, meaning the sub-Gaussian substitution seamlessly preserves all asymptotic $\tilde{\mathcal{O}}$ complexities. □

## I Dimension Dependence: Mi2P versus MiSTP

**Problem.** We use the stochastic quadratics $f_\xi(x) = \frac{1}{2}\|x-\xi\|^2$ with $\xi \sim \mathcal{N}(0, \kappa I_d)$, so that $f(x) = \frac{1}{2}\|x\|^2 + d\kappa/2$ and $\nabla f(x) = x$. Each $f_\xi$ is 1-smooth, so Assumptions 3.1 and 4.3 both hold. We consider two noise regimes. In regime (i), $\kappa = \sigma_c^2/d$, so the gradient variance $\mathbb{E}\|\nabla f_\xi(x) - \nabla f(x)\|^2 = \sigma_c^2$ is independent of $d$ (Assumption 4.4). In regime (ii), $\kappa = \sigma_c^2$, so the part of the function-value noise that depends on $x$, namely $-\langle x, \xi\rangle$, has variance $\sigma_c^2\|x\|^2$, independent of $d$; this is the regime of Theorem 4.2. We take $\sigma_c = 1$. Since $M_t^{\pm}$ and the MiSTP center value are averages over the same minibatch, the term $\frac{1}{2}\|\xi\|^2$ is common to all evaluated points and does not affect either update; the remaining part is linear in $\xi$, so we sample the minibatch averages from their exact joint Gaussian law.

**Setup.** Both methods use directions uniform on the unit sphere and step size $\eta = \varepsilon/\sqrt{d}$, the scaling $\eta \asymp \mu_{\mathcal{D}}$. Mi2P uses the constant batch size $b = 400$, whose independence from $d$ is the prescription of Theorem 4.5. In regime (ii), MiSTP uses $b \approx 178\, d^2$, the $d^2$ growth prescribed by Theorem 4.2, with the constant calibrated at the smallest dimension. We start from $\|x_0\| = 3$, stop once the running average $\frac{1}{T}\sum_t \|\nabla f(x_t)\|$ falls below $\varepsilon = 0.15$, and count $2b$ function evaluations per iteration of Mi2P and $3b$ per iteration of MiSTP. We sweep $d \in \{2, 4, \ldots, 128\}$ and report the median over 6 seeds.

**Results.** Figure 4 reports the total number of function evaluations. In panel (a), both methods run in regime (i) with the same batch size; the fitted slopes are 1.09 for Mi2P and 1.06 for MiSTP, so the two updates behave similarly on the same problem. In panel (b), each method runs in the regime and with the batch size of its theorem; the fitted slopes are 1.09 for Mi2P and 2.98 for MiSTP, in agreement with the $d$ and $d^3$ dependence of Theorems 4.2 and 4.5. In both panels the number of iterations grows linearly in $d$; the gap in panel (b) comes from the batch size each analysis requires.




(a) Regime (i), same batch size

(b) Each method in the regime of its theorem

Figure 4: Function evaluations to reach $\frac{1}{T}\sum_t \|\nabla f(x_t)\| \le \varepsilon$ versus dimension on stochastic quadratics. Median over 6 seeds; slopes from a least-squares fit in log-log scale.

## J Variance Reduction: Mi2P-VR versus ZO-SVRG and ZO-GD

**Problem.** We minimize $f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x)$ with $f_i(x) = \log\big(1 + \exp(-y_i \langle a_i, x\rangle)\big)$, whose components are $G$-Lipschitz since $\|\nabla f_i(x)\| \le \|a_i\|$. We take $d = 20$, features $a_i \sim \mathcal{N}(0, I_d)$, a planted parameter $x^\star$ with $\|x^\star\| = 3$, and labels $y_i \in \{-1, 1\}$ with $\Pr(y_i = 1) = \sigma(\langle a_i, x^\star\rangle)$, where $\sigma(t) = 1/(1 + e^{-t})$. We sweep $n \in \{64, 128, \ldots, 4096\}$.

**Methods and setup.** Mi2P-VR is the scheme of Theorem 5.1: an exact full-batch comparison every $m$ iterations and minibatch comparisons of size $b$ otherwise, with no stored snapshot. ZO-SVRG uses the two-point random-direction estimator with smoothing radius $10^{-3}$; at each snapshot it averages $Q \in \{2, 5\}$ directions over the full batch, and each inner step uses the snapshot control variate with one direction. ZO-GD applies the same two-point estimator to the full batch at every iteration. For both variance-reduced methods, $m = \mathrm{round}(\sqrt{n})$ and $b = \mathrm{round}(n/m)$. Step sizes (and $Q$ for ZO-SVRG) are tuned per method and per $n$ on a separate pilot instance. We start from $x_0 = 0$ and stop once $\frac{1}{T}\sum_t \|\nabla f(x_t)\| \le \varepsilon = 0.08$. Each evaluation of a component counts as one query, so a full-batch pair costs $2n$ queries, a minibatch pair $2b$, and an inner ZO-SVRG step $4b$. We report the median and interquartile range over 15 seeds; the range is conditional on the pilot-tuned step sizes.

**Results.** Table 3 and Figure 5 report the queries needed to reach $\varepsilon$. Full-batch ZO-GD is cheapest for small $n$, and the three methods are comparable around $n = 512$. From $n = 1024$ on, Mi2P-VR needs the fewest queries, with interquartile ranges separated from both baselines: its query count grows slowly in $n$, while that of ZO-GD and ZO-SVRG grows roughly linearly. At $n = 4096$, Mi2P-VR uses about $4\times$ fewer queries than ZO-SVRG and $5\times$ fewer than ZO-GD, without storing snapshots.

Table 3: Function queries to reach $\varepsilon$, in units of $10^5$: median [first quartile, third quartile] over 15 seeds. Bold marks the best method in rows where its interquartile range does not overlap those of the other methods.

| $n$ | Mi2P-VR | ZO-SVRG | ZO-GD |
|---|---|---|---|
| 64 | 1.40 [1.26, 2.54] | 0.79 [0.71, 0.92] | 0.71 [0.56, 0.73] |
| 128 | 2.00 [1.62, 2.33] | 1.11 [0.97, 1.22] | 1.10 [0.96, 1.15] |
| 256 | 3.15 [3.04, 3.25] | 1.84 [1.72, 2.12] | 1.85 [1.78, 2.08] |
| 512 | 4.09 [3.96, 4.24] | 3.95 [3.43, 4.23] | 3.72 [3.55, 4.06] |
| 1024 | **4.32 [3.93, 5.07]** | 4.85 [4.49, 5.78] | 7.13 [6.79, 7.76] |
| 2048 | **4.66 [4.40, 4.99]** | 12.0 [9.90, 12.3] | 14.1 [13.2, 14.5] |
| 4096 | **5.25 [4.73, 5.47]** | 20.0 [18.3, 22.6] | 28.5 [25.9, 31.3] |

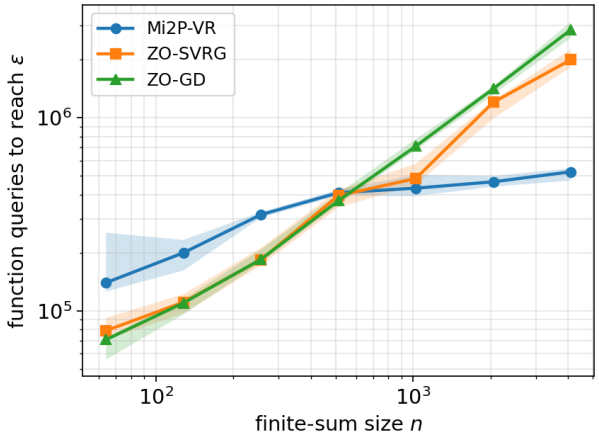



Figure 5: Function queries to reach $\varepsilon$ versus the finite-sum size $n$ on logistic regression with $d = 20$. Median over 15 seeds with interquartile band.

## K Helper Feedback: Synthetic Verification of the Floor

We verify the $\sqrt{d\delta}$ accuracy floor of Theorem 6.2 on a controlled problem. We take $f(x) = \frac{1}{2}\|x\|^2$, so that $L_0 = 1$, $L_1 = 0$ and $\|\nabla f(x)\| = \|x\|$, together with a stochastic scalar helper $h(x) = f(x) + \zeta$, where $\zeta \sim \mathcal{N}(0, \sigma_h^2)$ is drawn independently at each evaluation. For this helper, $\mathbb{E}|h(x) - h(y) - (f(x) - f(y))| = \mathbb{E}|\zeta_x - \zeta_y| = 2\sigma_h/\sqrt{\pi} =: \delta$, so Assumption 6.1 holds with $\delta \propto \sigma_h$. We run Algorithm 1 with $M_t^{\pm} = h(x_t^{\pm})$, measure the floor as the tail average of $\|x_t\|$ once the iterate has plateaued, and tune the step size to minimize the floor at each setting.

Figure 6 reports two sweeps. Fixing $\delta$ and varying $d$, the floor grows as $d^{0.57}$, consistent with the $\sqrt{d}$ prediction. Fixing $d = 16$ and varying $\delta$, the floor grows as $\delta^{0.50}$, matching the $\sqrt{\delta}$ dependence. Together these confirm the $\sqrt{d\delta}$ floor.

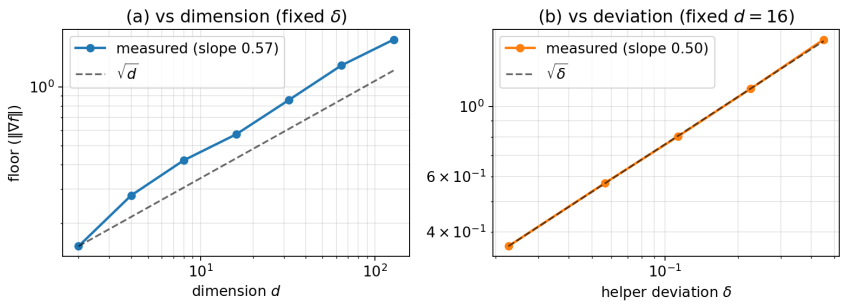



Figure 6: Synthetic verification of the Theorem 6.2 floor on $f(x) = \frac{1}{2}\|x\|^2$ with a noisy scalar helper. (a) floor versus dimension at fixed $\delta$; (b) floor versus helper deviation at fixed $d$. Dashed lines show the $\sqrt{d}$ and $\sqrt{\delta}$ references.

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our theoretical claims regarding iteration complexities under generalized smoothness and our empirical claims regarding convergence speedups are directly supported by the derivations and experiments in subsequent sections.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 9 discusses the dependence of the step sizes on the smoothness parameters. Section 7 shows that standard momentum techniques do not improve the rate, and Theorem 6.2 establishes an accuracy floor due to helper inexactness.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.

- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All assumptions are stated in Sections 3 to 6, and complete proofs of all results are provided in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Hyperparameters, tuning procedures, network architectures, and dataset and environment properties are given in Section 8 and Appendices I, J and K.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case

of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The datasets and environments we use are publicly available and cited. We do not release code: the method is a few lines (Algorithm 1), its implementation is straightforward, and all experimental details needed to reproduce the results are given in Section 8 and the appendix.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Batch sizes, step sizes, tuning grids, and evaluation budgets are given in Section 8 and Appendices I, J and K.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Each experiment states its number of independent seeds and its error measure: mean $\pm 1$ standard deviation for logistic regression (20 runs) and policy search (10 seeds), median and interquartile range for hyperparameter tuning (6 seeds) and the finite-sum comparison (15 seeds), and median over 6 seeds for the dimension study.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: The paper evaluates algorithmic efficiency purely in terms of exact function queries (oracle complexity) rather than wall-clock compute time, so hardware specifications do not affect the claims.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: This research introduces mathematical optimization algorithms and conforms fully to the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This paper is strictly focused on foundational mathematical optimization algorithms and has no direct path to negative societal impacts requiring mitigation discussion.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper proposes a numerical optimization algorithm and poses no high risk for misuse requiring controlled release or safety filters.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We properly credit the creators of the datasets and environments (e.g., Gymnasium, Breast Cancer Wisconsin) by citing the standard references.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: We do not introduce new datasets or assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing or human subjects research.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This paper does not involve research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were not used as a methodological component in the algorithmic development of this work. Only for writing and editing.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.