PARTIALLY INFORMED STOCHASTIC OPTIMIZATION VIA ZEROTH-ORDER RESIDUAL TRACKING

Anonymous authors
Paper under double-blind review

ABSTRACT

Many optimization problems provide direct access to only part of the gradient, while the remaining component must be inferred from function evaluations. This structure arises in performative prediction, performative reinforcement learning, and bilevel optimization with a black-box lower level. We introduce Partially Informed Stochastic Optimization (PISO), which combines the available gradient component with a running estimate of the residual. By centring each zeroth-order observation on this estimate, PISO makes the variance due to random probing depend on the tracking error rather than the residual's magnitude. This yields a randomized sketch iteration with an explicit contraction mechanism and a principled interpretation of momentum. For smooth nonconvex objectives, we establish convergence guarantees that depend on the residual's variation and initial magnitude, without requiring a uniform bound on the unavailable gradient component. With independently sampled function evaluations and Gaussian or spherical probes, PISO attains leading function-evaluation complexities of $O(d^2\varepsilon^{-6})$ under Lipschitz gradients and $O(d^2\varepsilon^{-5})$ under an additional Lipschitz-Hessian assumption. These rates require only one probe direction, two function samples, and one known-component sample per update. The analysis also quantifies how batching trades observations for fewer deployments. Across five settings, PISO variants improve on the tested baselines with comparable oracle access in three, reach a comparable solution faster in bilevel routing, and perform best among methods without response-model derivatives in performative reinforcement learning.

1 INTRODUCTION

Let $F : \mathbb{R}^d \to \mathbb{R}$ be a function whose gradient decomposes as

$$\nabla F(x) \;=\; g_{\mathrm{K}}(x) \;+\; g_{\mathrm{U}}(x), \tag{1}$$

where g_{K} can be estimated by direct sampling and g_{U} cannot: the only access to g_{U} is through noisy evaluations of F. Problems of this form arise in several areas.

Performative prediction. The decision changes the data distribution (Perdomo et al., 2020): one minimizes $F(x) = \mathbb{E}_{\xi\sim\mathcal{D}(x)}[\ell(x,\xi)]$, where $\mathcal{D}(x)$ is the distribution the deployed model x induces. Differentiating under the two dependencies makes $g_{\mathrm{K}}(x) = \mathbb{E}_{\xi\sim\mathcal{D}(x)}[\nabla_x \ell(x,\xi)]$ the ordinary loss gradient on the population currently observed, available by sampling, and g_{U} the term involving the derivative of the map $x \mapsto \mathcal{D}(x)$, which the practitioner does not possess. A lender adjusting a credit-scoring rule can compute the first, but learns how applicants adapt their reported features only by deploying a rule and observing outcomes (Levanon & Rosenfeld, 2021).

Performative reinforcement learning. The environment reacts to the deployed policy (Bell et al., 2021; Mandal et al., 2023). The performative policy gradient theorem of Basu et al. (2025) splits $\nabla_\theta V_{\pi_\theta}^{\pi_\theta}$ into a score term in $\nabla_\theta \log \pi_\theta$, the familiar policy gradient and our g_{K}, plus terms in $\nabla_\theta \log P_{\pi_\theta}$ and $\nabla_\theta r_{\pi_\theta}$, our g_{U}, which describe how the kernel and reward respond to the policy and require a model of that response.

Bilevel problems with an opaque lower level. Two parties act in sequence. The *leader* moves first, chooses x and pays the objective; the *followers* take x as given and settle on a response $y^\star(x)$, for instance the equilibrium of a game among themselves (Cui et al., 2023; Fischer & Paccagnan,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2026). With $F(x) = f(x, y^\star(x))$ and f known to the leader,

$$\nabla F(x) = \underbrace{\nabla_x f\big(x, y^\star(x)\big)}_{g_{\mathrm{K}}(x)} + \underbrace{\big[\partial_x y^\star(x)\big]^\top \nabla_y f\big(x, y^\star(x)\big)}_{g_{\mathrm{U}}(x)} :$$

the leader can evaluate both partial derivatives once $y^\star(x)$ is observed, so what is missing is the response Jacobian. In toll setting, the leader is the road authority and the followers are the drivers who reroute; here g_{K} is exact, whereas in the two performative settings it must be sampled. The same structure appears whenever a differentiable model is composed with a black-box simulator, or with a legacy or proprietary subsystem.

Prior approaches and this work. Treating F as a black box and applying zeroth-order optimization to all of ∇F (Nesterov & Spokoiny, 2017; Ghadimi & Lan, 2013) discards g_{K}. Recent work in each area exploits it: Hikima et al. (2025) use the computable component as a guide in performative prediction and Hikima & Takeda (2025a) attain $\mathcal{O}(d^2\varepsilon^{-6})$ samples by averaging many probe directions per step; Fischer & Paccagnan (2026) keep the leader's partial gradients exact and estimate only the followers' response Jacobian. These methods share a template, estimating a gradient or a Jacobian at the current iterate and then reducing its variance, and a scope, each being developed for one area. We keep the premise that g_{K} should be used and change the object being estimated: we maintain a running estimate of the opaque part and centre every zeroth-order observation on it, so that each observation reports how wrong the current estimate is rather than a fresh guess. The update is then a randomized fixed-point iteration whose error contracts, with stochastic error proportional to the current estimation error rather than to the size of g_{U}, and because it needs only (1), one algorithm and one analysis cover all three areas.

Contributions.

- We formulate optimization under (1) as *residual identification* and derive PISO and PISO^2 (§2), showing that centring the zeroth-order observation on the running estimate turns gradient estimation into an error-feedback contraction.
- We give a single convergence analysis (§3) whose complexity is governed by the smoothness L_v and initial magnitude $\|v(x_0)\|$ of the residual, with no dependence on a uniform bound on $\|g_{\mathrm{U}}\|$ and no bounded-gradient, bounded-domain, convexity or PL assumption. It covers the two stochastic oracle regimes and any probe distribution satisfying three moment conditions, matching the $d^2\varepsilon^{-6}$ and $d^2\varepsilon^{-5}$ rates of Hikima & Takeda (2025a) with L replaced by L_v, one probe per update instead of $\Omega(d)$, and one sample of each oracle per iteration.
- The analysis assigns each parameter a role and a prescription: the momentum ρ is the step size of the sketch and has a stability window of width $2/c_d$; a second momentum ρ_K on the known component is what makes one sample per iteration admissible; and the weight γ is a control-variate coefficient interpolating between black-box and white-box treatment. We also separate the cost of *deploying* a decision from that of observing outcomes under it, and show the deployment count can be cut from $\Theta(d^2\varepsilon^{-6})$ to $\Theta(d^2\varepsilon^{-2})$ at no cost in total observations, and to $\Theta(d\,\varepsilon^{-2})$ at the cost of a factor d in observations.
- We evaluate on five settings spanning the three areas (§5). PISO is best on pricing, strategic classification and a security game, matches the strongest bilevel baseline at convergence on a routing game while reaching it earlier, and on performative RL is best among methods that query the environment rather than differentiating a model of it. A momentum sweep locates the optimum at the predicted $1 - 1/c_d$ and exhibits the predicted divergence below $1 - 2/c_d$.

2 FROM GRADIENT ESTIMATION TO RESIDUAL TRACKING

Setting and notation. We minimize a function $F : \mathbb{R}^d \to \mathbb{R}$ satisfying (1), given two oracles. The first returns $\widehat{g}_{\mathrm{K}}(x)$, unbiased for $g_{\mathrm{K}}(x)$ with variance σ_K^2/n_K, where n_K is a batch size; we write $\delta_k := \gamma\big(\widehat{g}_{\mathrm{K}}(x_k) - g_{\mathrm{K}}(x_k)\big)$ for its error scaled by a weight $\gamma > 0$ introduced below. The second returns noisy values $\widehat{F}(x)$, each an average of n_F samples. At iteration k we draw a *probe* $u_k \in \mathbb{R}^d$ with a *score* $s_k \in \mathbb{R}^d$ satisfying

$$\mathbb{E}\left[s_k u_k^\top\right] = I_d, \qquad \text{so that} \qquad \mathbb{E}\left[\langle z, u_k\rangle\, s_k\right] = z \quad \text{for every fixed } z \in \mathbb{R}^d, \tag{2}$$

for instance $u_k = s_k \sim \mathcal{N}(0, I_d)$ or $u_k = s_k = \sqrt{d}\, w_k$ with w_k uniform on the unit sphere. Let c_d be the smallest constant with $\mathbb{E}\,\|\langle z, u_k\rangle s_k\|^2 \le c_d\|z\|^2$; it equals $d+2$ and d for those two choices. From the two function queries we form the symmetric two-point directional derivative, which decomposes as

$$D_k \;:=\; \frac{\widehat{F}(x_k + \mu_k u_k) - \widehat{F}(x_k - \mu_k u_k)}{2\mu_k} \;=\; \langle \nabla F(x_k), u_k\rangle + b_k + \xi_k, \tag{3}$$

where b_k is the finite-difference bias and ξ_k has conditional mean zero. An iteration costs $2n_F$ function evaluations and n_K known-component samples while touching only the three points x_k and $x_k \pm \mu_k u_k$; in a performative problem each of those is a *deployed decision*, and Appendix K takes up the distinction. Throughout, η is the step size, μ_k the probe radius, $\rho, \rho_K \in [0,1)$ two momentum parameters, $\beta := 1-\rho$, $\beta_K := 1-\rho_K$, and m_k, h_k the states maintained by the method.

The tracked object. For the weight $\gamma > 0$ define the *residual*

$$v(x) \;:=\; \nabla F(x) - \gamma\, g_{\mathrm{K}}(x) \;=\; g_{\mathrm{U}}(x) + (1-\gamma)\, g_{\mathrm{K}}(x). \tag{4}$$

At $\gamma = 1$ it is exactly g_{U}, and at $\gamma = 0$ the full gradient. We keep γ general because the two endpoints are two known designs, a fully black-box method and the principle of applying zeroth-order estimation only where it is needed (Fischer & Paccagnan, 2026), and because the analysis identifies a better choice. Our goal is a state m_k tracking $v(x_k)$, since then $\gamma\widehat{g}_{\mathrm{K}}(x_k) + m_k$ estimates $\nabla F(x_k)$.

A first estimator, and its cost. Rearranging (3) with $\nabla F = \gamma g_{\mathrm{K}} + v$ shows what the two queries reveal about v:

$$D_k - \langle \gamma\widehat{g}_{\mathrm{K}}(x_k), u_k\rangle \;=\; \langle v(x_k), u_k\rangle \;-\; \langle \delta_k, u_k\rangle \;+\; b_k + \xi_k. \tag{5}$$

Each iteration supplies one noisy *scalar* linear measurement of the unknown vector $v(x_k)$ along u_k. Multiplying by the score turns it into a vector estimate, and by (2) the estimate $R_k^{\mathrm{fresh}} := (D_k - \langle \gamma\widehat{g}_{\mathrm{K}}(x_k), u_k\rangle)s_k$ is unbiased for $v(x_k)$ up to b_k. Its second moment, however, is of order $c_d\|v(x_k)\|^2$: a single probe inflates the estimate by the dimension, and the error persists at that level however long the method runs. Averaging N probes per iteration divides the inflation by N, but the guarantee of Hikima & Takeda (2025a) requires $N = \Omega(d)$, so $\Omega(d)$ queries precede every update; in performative settings each query is a deployment.

Measuring the discrepancy instead. The estimator R_k^{fresh} discards the state: it reconstructs $v(x_k)$ from scratch even when m_k is already close. Suppose instead we treat m_k as known and estimate only the discrepancy $v(x_k) - m_k$. Subtracting $\langle m_k, u_k\rangle$ from (5) gives a scalar measurement of that discrepancy along u_k; sketching it with s_k estimates the discrepancy as a vector; and adding m_k back recovers an estimate of $v(x_k)$:

$$R_k \;:=\; m_k + \big(D_k - \langle \gamma\widehat{g}_{\mathrm{K}}(x_k) + m_k, u_k\rangle\big)\, s_k. \tag{6}$$

By (2) this has the same target as R_k^{fresh}, since the two occurrences of m_k cancel in expectation. What changes is the scale of the error, now set by $\|v(x_k) - m_k\|$ rather than by $\|v(x_k)\|$, so it shrinks as the state improves. This is the device behind control variates and error feedback: estimate the correction to a known quantity, not the quantity itself.

We call the method *Partially Informed Stochastic Optimization* (PISO), and PISO^2 for the variant with $\lambda > 0$, which reads out the direction from a mixture of the fresh observation R_k and the state. Line 6 tracks the known component as well, because $\widehat{g}_{\mathrm{K}}(x_k)$ enters the step directly and using it fresh injects its full variance $\gamma^2\sigma_K^2$ into G_k; averaging reduces this to $\mathcal{O}(\beta_K\gamma^2\sigma_K^2)$ at the cost of a lag the step size must respect, and $\rho_K = 0$ recovers the fresh estimate. We deliberately do *not* substitute h_{k+1} for $\widehat{g}_{\mathrm{K}}(x_k)$ inside R_k: doing so would change the tracked object to $\nabla F(x_k) - \gamma h_{k+1}$, whose drift constant is $L_v + \gamma L_K$ rather than L_v (Appendix F). With $\tilde{\beta} := 1 - (1-\lambda)\rho$ and $\zeta_k := \gamma(h_k - g_{\mathrm{K}}(x_k))$, both $\mathcal{F}_k$-measurable, the error recursions are as follows.

Lemma 1 (PISO is sketched error feedback)**.** *Let $E_k := m_k - v(x_k)$, $\Delta_k := G_k - \nabla F(x_k)$ and $P_k := s_k u_k^\top$. With $m_0 = 0$: (i) $\lambda R_k + (1-\lambda)m_{k+1} = (1-\tilde{\beta})m_k + \tilde{\beta}R_k$; (ii) $E_{k+1} = (I_d - \beta P_k)E_k + \beta\big(-\langle\delta_k, u_k\rangle + b_k + \xi_k\big)s_k + (v_k - v_{k+1})$; (iii) $\Delta_k = (1-\beta_K)\zeta_k + (I_d - \tilde{\beta}P_k)E_k + \big(\beta_K I_d - \tilde{\beta}P_k\big)\delta_k + \tilde{\beta}(b_k + \xi_k)s_k$; (iv) $\zeta_{k+1} = (1-\beta_K)\zeta_k + \beta_K\delta_k + \gamma\big(g_{\mathrm{K}}(x_k) - g_{\mathrm{K}}(x_{k+1})\big)$.*

Algorithm 1 PISO ($\lambda = 0$) and PISO^2 ($\lambda \in (0, 1]$)

Require: x_0, $m_0 = 0$, $h_0 = \widehat{g}_{\mathrm{K}}(x_0)$, weight $\gamma > 0$, momenta ρ, ρ_K, mixing λ, schedules $\{\eta_k\}, \{\mu_k\}$

1: **for** $k = 0, \dots, T-1$ **do**
2: Draw $\widehat{g}_{\mathrm{K}}(x_k)$; draw a probe direction u_k with score s_k
3: $D_k \leftarrow \big[\widehat{F}(x_k + \mu_k u_k) - \widehat{F}(x_k - \mu_k u_k)\big]/(2\mu_k)$
4: $R_k \leftarrow m_k + \big(D_k - \langle \gamma \widehat{g}_{\mathrm{K}}(x_k) + m_k, u_k \rangle\big) s_k$ ▷ *residual observation, centred on* m_k
5: $m_{k+1} \leftarrow \rho\, m_k + (1-\rho) R_k$
6: $h_{k+1} \leftarrow \rho_K\, h_k + (1-\rho_K)\, \widehat{g}_{\mathrm{K}}(x_k)$ ▷ *known-component average*
7: $G_k \leftarrow \gamma\, h_{k+1} + \lambda R_k + (1-\lambda) m_{k+1}$
8: $x_{k+1} \leftarrow x_k - \eta_k G_k$ ▷ sign reversed for maximization
9: **end for**

Why this contracts. In the idealized case of exact oracles and a stationary iterate, (ii) reduces to the exact identity

$$E_{k+1} = (I_d - \beta P_k) E_k. \tag{7}$$

This is a step of stochastic gradient descent on the quadratic $\frac{1}{2}\|m - v\|^2$ with step size β, the gradient E_k replaced by its sketch $P_k E_k$, which is unbiased by (2). Taking second moments gives the textbook step-size trade-off, a linear descent term against a quadratic variance term: $\mathbb{E}_k \|(I_d - bP_k)z\|^2 \le \theta(b)\|z\|^2$ with

$$\theta(b) \;:=\; 1 - 2b + c_d b^2. \tag{8}$$

So β is the step size of a randomized fixed-point iteration for v: under (D1)–(D2), $\theta(\beta)$ bounds the contraction factor, is minimized at $\beta = 1/c_d$, and guarantees contraction for $0 < \beta < 2/c_d$; for the Gaussian, sphere and coordinate families (D2) is an equality and the statements are exact:

$$\rho^\star = 1 - 1/c_d, \qquad \text{and the residual iteration is neutral at } \rho = 1 - 2/c_d \text{ and } \textit{expands} \text{ for } \rho < 1 - 2/c_d. \tag{9}$$

In particular $\rho = 0$ is not "no momentum" but $\beta = 1$, an expanding map with $\theta(1) = c_d - 1$. The update sits in the family of error-feedback and sketch-and-project methods (Richtárik et al., 2021; Gower & Richtárik, 2015). In (iii) the oracle error δ_k appears only through $\beta_K I_d - \tilde{\beta} P_k$, whose second moment against a mean-zero δ_k is $(\beta_K^2 - 2\beta_K \tilde{\beta} + c_d \tilde{\beta}^2)\, \mathbb{E}\,\|\delta_k\|^2$, and through ζ_k, whose steady state is $\mathcal{O}(\beta_K \gamma^2 \sigma_K^2)$ by (iv). Both are tunable, which is what allows one known-component sample per iteration.

3 CONVERGENCE

Assumption 1 (Smoothness)**.** F is L-smooth and bounded below by $F^\star$, with $\Delta_F := F(x_0) - F^\star$; and g_{K} is L_K-Lipschitz. We write L_v for the smallest constant with $\mathbb{E}_k \|v_{k+1} - v_k\|^2 \le L_v^2 \eta^2\, \mathbb{E}_k \|G_k\|^2$.

Since v is $(L + \gamma L_K)$-Lipschitz, $L_v \le L + \gamma L_K$ always; the rates below improve on black-box ones when L_v is small, which depends on the problem. One sets η with an a-priori upper bound $\bar{L}_v$ (e.g. $L + \gamma L_K$); the trajectory-specific L_v sharpens the bound only retrospectively.

Assumption 2 (Oracles)**.** $\mathbb{E}_k[\widehat{g}_{\mathrm{K}}(x_k)] = g_{\mathrm{K}}(x_k)$ with variance $\sigma_{K,k}^2$, drawn independently of (u_k, s_k); and $\mathbb{E}[\xi_k \mid \mathcal{F}_k, u_k, s_k, \delta_k] = 0$ with $\mathbb{E}[\xi_k^2 \mid \mathcal{F}_k, u_k, s_k] \le \sigma_{D,k}^2$. Write $\Sigma_{b,k}^2 := \mathbb{E}_k[b_k^2 \|s_k\|^2]$.

Assumption 3 (Probe distribution)**.** (u_k, s_k) is drawn i.i.d. given $\mathcal{F}_k$, independently of δ_k, and satisfies (D1) $\mathbb{E}_k[s_k u_k^\top] = I_d$; (D2) $\mathbb{E}_k \|\langle z, u_k \rangle s_k\|^2 \le c_d \|z\|^2$; (D3) moment bounds $\varsigma^2, \varsigma_4^2, \varsigma_6^2, \varsigma_8^2$ on $\|s_k\|^2$ against powers of $\|u_k\|$.

Only (D1)–(D2) enter the contraction; (D3) enters the noise constants. Theorem 2 is stated in terms of the *mean* bias $\bar{b}_k := \mathbb{E}_k[b_k s_k]$ and a second bias vector $\bar{B}_k := \mathbb{E}_k[b_k \|s_k\|^2 u_k]$; the corollaries bound them in one of two ways (Appendix D). If the family satisfies a *smoothing identity* $\mathbb{E}[\frac{F(x+\mu u) - F(x-\mu u)}{2\mu} s] = \nabla F_{\nu, \kappa\mu}(x)$ for a symmetric measure ν and scale κ, then $\|\bar{b}_k\| \le L \mu_k \chi_1$

with $\chi_1 := \kappa\,\mathbb{E}_\nu\,\|w\|$, and $\frac{1}{2}Q\mu_k^2\chi_2$ with $\chi_2 := \kappa^2\,\mathbb{E}_\nu\,\|w\|^2$ under a Q-Lipschitz Hessian (Berahas et al., 2022; Maggiar et al., 2018); Gaussian and sphere probes satisfy it with $\chi_1 = \sqrt{d}$, $\chi_2 = d$. Coordinate probes do not, and a direct bound gives $\chi_1 = d/2$, a factor $\sqrt{d}$ worse, which is the structural cause of the known gap between coordinate and random directions. The distinction between $\bar{b}_k$ and $\Sigma_{b,k}$ controls the dimension dependence: the bias accumulates coherently over the $\sim 1/\beta$ steps of the momentum window, so its mean is undamped while its fluctuation is damped by β like any other noise.

Theorem 2 (Main). *Under Assumptions 1–3, with $\theta := \theta(\tilde{\beta})$, $\beta \le 1/c_d$ and*

$$\eta \;\le\; \min\Big\{\tfrac{1}{2L},\; \tfrac{\beta}{11\sqrt{\theta}\,L_v},\; \tfrac{\beta_K}{9\,\gamma L_K}\Big\}, \tag{10}$$

the iterates of Algorithm 1 run with a constant step size $\eta_k \equiv \eta$ (the radius μ_k may vary) satisfy

$$\begin{aligned}\frac{1}{T}\sum_{k<T}\mathbb{E}\,\|\nabla F(x_k)\|^2 \le \frac{2\Delta_F}{\eta T} + \frac{6\theta\|v(x_0)\|^2}{\beta T} + \frac{6\gamma^2\sigma_K^2}{n_K\beta_K T}\\ + \frac{2}{T}\sum_{k<T}\Big[\big(\tfrac{3}{2}\beta_K^2 + \tfrac{3}{2}c_d\tilde{\beta}^2 + \tfrac{15}{4}\theta\beta c_d + \tfrac{9}{2}\beta_K\big)\gamma^2\sigma_{K,k}^2\\ + \big(\tfrac{3}{2}\tilde{\beta}^2 + \tfrac{15}{4}\theta\beta\big)\big(\varsigma^2\sigma_{D,k}^2 + \Sigma_{b,k}^2\big)\\ + 30\theta\|\bar{b}_k\|^2 + 30\theta\beta^2\|\bar{B}_k\|^2\Big].\end{aligned} \tag{11}$$

No bounded gradients, bounded domain, convexity, PL condition, or uniform bound on $\|g_{\mathrm{U}}\|$ is used: the opaque component enters only through its value at x_0, via $\|v(x_0)\|$, and through its drift L_v. Every occurrence of the known-oracle variance carries a factor βc_d, $\tilde{\beta}^2 c_d$ or β_K, so a single known-component sample per iteration suffices; at $\rho_K = 0$ the last coefficient is $\Theta(1)$ and σ_K^2 becomes a floor removable only by drawing $\Theta(\gamma^2\sigma_K^2/\varepsilon^2)$ samples per iteration (Appendix G). The three bias terms sit at different orders in β, and only the mean $\|\bar{b}_k\|^2$ is undamped.

3.1 Oracle regimes and complexity

Whether the two function queries in (3) can share randomness determines the noise model. If they evaluate the same realization, the noise is dominated by a μ-free gradient term, a finite probe radius of order $\min\{\sigma_g/\ell, \varepsilon/L\}$ suffices, and the method reaches ε-stationarity in $\mathcal{O}\big(\Delta_v d(\gamma^2\sigma_K^2+\sigma_g^2)\varepsilon^{-4}\big)$ iterations with $\Delta_v := L_v\Delta_F + \|v(x_0)\|^2$ (Appendix G, Regime A). In a performative problem the two queries *deploy* $x_k \pm \mu_k u_k$ and sample from two different distributions, so they cannot share randomness and

$$\sigma_{D,k}^2 \;\le\; \frac{\sigma_F^2}{2\mu_k^2\,n_F}, \tag{12}$$

with σ_F^2 a bound on the conditional variance of the objective under the deployed distribution. The probe radius must then balance (12) against the finite-difference bias, and the balance is asymmetric: variance is damped by β in (11), the mean bias is not, which fixes the exponent of ε. Taking $\lambda = 0$, so $\tilde{\beta} = \beta$, and the sharp bound $\|\bar{B}_k\| \le 3dL\mu_k\chi_1$ of the random families under $\beta \le 1/(3d)$, the noise part of (11) is at most

$$\big(11\beta c_d + 12\beta_K\big)\frac{\gamma^2\sigma_K^2}{n_K} + 11\beta\,\varsigma^2\sigma_D^2 + 11\beta\,\Sigma_b^2 + 60\|\bar{b}\|^2 + 60L^2\mu^2\chi_1^2. \tag{13}$$

Corollary 3 (Independent randomization). *Write $C_1 := 102000\,L^2\sigma_F^2\varsigma^2\chi_1^2$ and $C_2 := 1220\,\varsigma^2\sigma_F^2 Q\chi_2$, and let $n_K \ge 1$ and $\beta_K = \min\{1, n_K\varepsilon^2/(72\gamma^2\sigma_K^2)\}$, with the warm start $h_0 = \widehat{g}_{\mathrm{K}}(x_0)$, $\mu_\star$ as in Appendix G, and η* equal to *the right-hand side of* (10)*. Under L-smoothness alone, take $\beta = n_F\varepsilon^4/C_1$ with $1 \le n_F \le n_F^{\max} := C_1 n_K/(66c_d\gamma^2\sigma_K^2\varepsilon^2)$ and $\varepsilon^2 \le 7\gamma^2\sigma_K^2/(dn_K)$; if $\nabla^2 F$ is Q-Lipschitz, take instead $\beta = n_F\varepsilon^3/C_2$ with $n_F \le C_2 n_K/(66c_d\gamma^2\sigma_K^2\varepsilon)$ and $\varepsilon^2 \le \gamma^2\sigma_K^2/(dn_K)$. The total number of function evaluations to reach $\frac{1}{T}\sum_{k<T}\mathbb{E}\,\|\nabla F_k\|^2 \le \varepsilon^2$ then satisfies*

$$\textit{under L-smoothness alone:}\qquad \mathsf{N} = \mathcal{O}\Big(\frac{\Delta_v\,L^2\sigma_F^2\,\varsigma^2\chi_1^2}{\varepsilon^6} + \frac{n_F}{n_K}\cdot\frac{\gamma^3 L_K\sigma_K^2\Delta_F}{\varepsilon^4} + R_{\mathrm{lo}}\Big), \tag{14}$$

$$\textit{if } \nabla^2 F \textit{ is Q-Lipschitz:}\qquad \mathsf{N} = \mathcal{O}\Big(\frac{\Delta_v\,\sigma_F^2\,\varsigma^2 Q\chi_2}{\varepsilon^5} + \frac{n_F}{n_K}\cdot\frac{\gamma^3 L_K\sigma_K^2\Delta_F}{\varepsilon^4} + R_{\mathrm{lo}}\Big), \tag{15}$$

Method	Setting	Complexity	Probes	Additional structure
Two-point ZO (Ghadimi & Lan, 2013)	black box	$\mathcal{O}(d\varepsilon^{-4})$	1	shared randomness
Guided ZO (Hikima et al., 2025)	perf. pred.	$\mathcal{O}(d^4\varepsilon^{-6})$	1	known component
Multi-direction ZO (Hikima & Takeda, 2025a)	perf. pred.	$\mathcal{O}(d^2\varepsilon^{-6})$, $\mathcal{O}(d^2\varepsilon^{-5})$	$\Omega(d)$	Q-Lipschitz (ε^{-5})
PZOS (Fischer & Paccagnan, 2026)	MPEC	$\mathcal{O}(d^{3/2}\varepsilon^{-5})$, Goldstein	1	bounded ∇f, exact $\nabla_x f$
PePG (Basu et al., 2025)	perf. RL	$\Omega(\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert^2\varepsilon^{-2}(1-\gamma)^{-3})$	n/a	response model
PISO (this work)	all three	$\mathcal{O}(d^2\varepsilon^{-6})$, $\mathcal{O}(d^2\varepsilon^{-5})$	1	split (1), g_{K} oracle

Table 1: Complexities as reported by each source; criteria and oracle models differ, so only the last two columns compare directly. Against the guided line we improve $d^4 \to d^2$; against the frontier we match d^2 with $L \to L_v$ and one probe per update; PZOS has a Goldstein guarantee we lack but assumes bounded gradients and a deterministic follower; PePG requires a differentiable response model.

where $R_{\mathrm{lo}} := n_F\gamma^4\sigma_K^4/(n_K^2\varepsilon^4) + n_F L\Delta_F/\varepsilon^2$ collects the initialization and L-branch terms, lower order at $n_F = 1$ and made explicit in Appendix G, where a batched warm start is also shown to remove the former. The leading terms are independent of n_F and n_K and carry no σ_K.

Reading $\varsigma^2\chi_1^2$ and $\varsigma^2\chi_2$ off the probe constants gives d^2 for Gaussian and sphere probes in both cases; coordinate probes, which lack the smoothing identity, are treated by a separate corollary with the direct bias bounds and give d^3 and $d^{5/2}$ (Appendix G). These are the six complexities reported by Hikima & Takeda (2025a) for plain multi-direction methods, obtained here from one analysis and a table of constants. Three things change. *The factor multiplying Δ_F is L_v, not L*: the step-size condition (10) is a minimum of three branches, the L_v/β branch dominates once $\beta = \mathcal{O}(\varepsilon^2)$, and L survives only at order ε^{-2}, while the L^2 in (14) is the finite-difference bias common to any two-point method (Appendix G.4). *One probe per update suffices*, where the guarantee of Hikima & Takeda (2025a) needs $N \geq 81d$, which at $d = 30$ exceeds their own 5000-sample budget (Appendix J). *One sample of each oracle per iteration suffices*: $n_F = n_K = 1$ is admissible, though not the best operating point in a performative problem, since N is flat in n_F while the number of deployed decisions $3T$ falls as $1/n_F$ (§3.2).

The weight γ and the smoothness gain. The analysis assigns each parameter a role and a prescription (Appendix I); we single out γ because it governs the constant that separates this work from prior methods. The drift term in Lemma 1(ii) is, by (4), exactly

$$v_{k+1} - v_k = \mathsf{d}_k^U + (1-\gamma)\,\mathsf{d}_k^K, \qquad \mathsf{d}_k^U := g_{\mathrm{U}}(x_{k+1}) - g_{\mathrm{U}}(x_k), \quad \mathsf{d}_k^K := g_{\mathrm{K}}(x_{k+1}) - g_{\mathrm{K}}(x_k). \tag{16}$$

Proposition 4 (Per-step drift-minimizing weight)**.** *Fix an iteration k and the displacement $x_{k+1} - x_k$, and write $\mathsf{A}_k := \mathbb{E}_k\,\|\mathsf{d}_k^U\|^2$, $\mathsf{B}_k := \mathbb{E}_k\langle \mathsf{d}_k^U, \mathsf{d}_k^K\rangle$, $\mathsf{C}_k := \mathbb{E}_k\,\|\mathsf{d}_k^K\|^2$, with $\varphi_k := \mathsf{B}_k/\sqrt{\mathsf{A}_k\mathsf{C}_k}$. If $\mathsf{C}_k > 0$, the per-step drift $\mathbb{E}_k\,\|v_{k+1} - v_k\|^2 = \mathsf{A}_k + 2(1-\gamma)\mathsf{B}_k + (1-\gamma)^2\mathsf{C}_k$ is a strictly convex quadratic in γ, minimized at $\gamma_k^\star = 1 + \mathsf{B}_k/\mathsf{C}_k$ with minimum value $\mathsf{A}_k(1-\varphi_k^2)$. At $\gamma = 1$ the drift is A_k, the increment of g_{U} alone; at $\gamma = 0$ it is $\mathbb{E}_k\,\|\nabla F(x_{k+1}) - \nabla F(x_k)\|^2$, which the black-box analysis bounds by $L^2\eta^2\,\mathbb{E}_k\,\|G_k\|^2$; equality need not hold, so $L_v(0) \leq L$ rather than $L_v(0) = L$.*

Two qualifications apply. L_v is a *uniform* bound on the drift ratio along a trajectory that itself depends on γ, so the proposition identifies the drift-minimizing weight at each step rather than the minimizer of L_v; and the method needs $\gamma > 0$, which $\gamma_k^\star$ can fail when $\mathsf{B}_k < -\mathsf{C}_k$. With those caveats the gain has two parts. Moving from $\gamma = 0$ to $\gamma = 1$ replaces the increment of the full gradient, at most $L\eta\|G_k\|$, by that of the opaque component alone, which in performative prediction scales with the performative sensitivity and can be far smaller; this is the substitution $L \to L_v$ of Corollary 3, which Hikima et al. (2025) and Fischer & Paccagnan (2026) do not obtain because their estimators are redrawn each iteration and their rates carry L. Moving from $\gamma = 1$ to $\gamma_k^\star$ reduces the per-step drift by a further factor $1-\varphi_k^2$ when the two components' increments are correlated: γ is the least-squares regression coefficient of $-\mathsf{d}_k^U$ on d_k^K, a control-variate coefficient built from uncentred second moments rather than a covariance in the strict sense, and neither endpoint is drift-minimizing unless $\varphi_k = 0$. The remaining parameters have prescriptions of the same character, (9) for ρ and Appendix I for λ and ρ_K, each settable once γ is fixed.

3.2 DEPLOYMENTS, AND HOW THE BATCH SIZES CONTROL THEM

In a performative problem a query is an intervention: evaluating $\widehat{F}$ at a point requires *deploying* that decision and observing outcomes under it. Each iteration deploys three decisions, x_k for the known-component oracle and $x_k \pm \mu_k u_k$ for the two function values, so the total is $3T$ regardless of the batch sizes. Deployments and samples are therefore distinct resources, and where a decision is a price schedule or a lending rule, the number of distinct decisions released is the cost that matters.

The two batch sizes act on this cost in the same direction. Averaging n_F samples per query point divides $\sigma_{D,k}^2$ by n_F and permits a proportionally larger β; averaging n_K samples of $\widehat{g}_{\mathrm{K}}$ permits a larger β_K. Since $T = \Theta(\Delta_v/(\beta\varepsilon^2))$, both enlarge the step and reduce T. How far this goes is governed by two caps on β that sit a factor d apart. Below $\beta_1 := \chi_1^2/\varsigma_6^2 = \Theta(1/d^2)$ the damped second-moment bias $11\beta\Sigma_b^2$ stays dominated by the undamped mean bias, the probe radius balances as in Corollary 3, and N is flat in β; between β_1 and the bias-analysis cap $\beta_2 := 1/(3d)$ the second-moment bias takes over the radius balance and N grows linearly in β.

	β	deployments $3T$	function evals. $2n_FT$	known samples n_KT
$n_F = n_K = 1$	ε^4/C_1	$\Theta(d^2\varepsilon^{-6})$	$\Theta(d^2\varepsilon^{-6})$	$\Theta(d^2\varepsilon^{-6})$
batches raised until $\beta = \beta_1$	$\Theta(1/d^2)$	$\Theta(d^2\varepsilon^{-2})$	$\Theta(d^2\varepsilon^{-6})$	$\Theta(d\,\varepsilon^{-4})$
batches raised until $\beta = \beta_2$	$\Theta(1/d)$	$\Theta(d\,\varepsilon^{-2})$	$\Theta(d^3\varepsilon^{-6})$	$\Theta(d\,\varepsilon^{-4})$

So enlarging the batches up to β_1 is free in function evaluations while cutting the deployment count by ε^{-4}, from $\Theta(d^2\varepsilon^{-6})$ to $\Theta(d^2\varepsilon^{-2})$, plus the additive $\gamma L_K \Delta_F \varepsilon^{-2}$ of the known-component branch (Appendix K). Going further, to β_2, removes one more factor d from the deployment count at the price of a factor d in observations, a trade a practitioner may well accept when a deployment is an intervention on a population. Three points are worth separating. The analysis permits $n_F = n_K = 1$, so no minibatching is required, but $n_F = 1$ is the wrong operating point in a performative problem, since it maximizes the number of distinct decisions released for the same observations. The two caps have different origins, and neither is the stability boundary: $\beta_2 = 1/(3d)$ is the limit of the sharp bias-vector domination argument (Appendix D), while the sketch contraction itself holds up to $2/c_d$ and Theorem 2 uses $\beta \le 1/c_d$; β_1 reflects the generic bound $\Sigma_b^2 \le \frac{1}{4}L^2\mu^2\varsigma_6^2$, which we show is loose by a factor d on examples; a general bound $\Sigma_b^2 = \mathcal{O}(L^2\mu^2 d^2)$ would move β_1 to $\Theta(1/d)$ and make the last row free as well, and we leave it open. And a multi-direction method cannot trade this way at all: its N probes are $2N$ *distinct* decisions per update, which repeated sampling cannot replace, so with $N = \Theta(d^2\varepsilon^{-4})$ its deployment count equals its sample count, $\Theta(d^2\varepsilon^{-6})$.

4 RELATED WORK

Methods that exploit a known gradient component. Using a cheap or partial signal alongside an expensive one is the subject of *optimization with access to auxiliary information* (Chayti & Karimireddy, 2024) and the helper framework extending it to second order (Chayti et al., 2024); our setting is an instance in which the complement of the auxiliary signal is accessible only through function values. Within performative prediction, Ray et al. (2022) and Liu et al. (2024) use one-point estimators with constants in $\sup|f|$, Hikima & Takeda (2025b) replace that by a variance bound, and Hikima et al. (2025); Hikima & Takeda (2025a) are discussed above. Within MPECs, Cui et al. (2023); Cui & Shanbhag (2023) apply zeroth-order tools to the whole composite objective, Fischer & Paccagnan (2026) estimate only the follower Jacobian, and Sow et al. (2022) smooth the inner solution map. All of these estimate a gradient or a Jacobian at the current iterate, leaving a second moment of order c_d redrawn each iteration: in the bound of Fischer & Paccagnan (2026) the leading $\mathcal{O}(d)$ term is common to their partial estimator and to the black-box baseline. Tracking a residual makes the direction error contract instead, and needs no composite structure, which is why it extends to performative RL.

Performative prediction and RL. Perdomo et al. (2020) introduced performative prediction and the stability/optimality distinction; Mendler-Dünner et al. (2020), Mofakhami et al. (2023), Li & Wai (2024) and Jagadeesan et al. (2022) study variants. Methods targeting performative *optima* assume a parametric distribution map (Izzo et al., 2021; 2022; Miller et al., 2021) or plug in an

estimate of it (Lin & Zrnic, 2024); Cyffers et al. (2026) study the choice of regularizer under shift. Strategic classification (Levanon & Rosenfeld, 2021; Cyffers et al., 2024) and Stackelberg games with algorithmic agents (Frankel et al., 2025) supply two of our settings. In performative RL most work targets *stability* (Mandal et al., 2023; Rank et al., 2024; Mandal & Radanovic, 2025); Basu et al. (2025) give the first policy-gradient method for optimality, whose guarantee needs a softmax PeMDP with a known feature map so that the response derivatives are in closed form. Where such a model exists its use is preferable; the settings that motivate our work are those where it does not, and PISO assumes only that the response can be queried.

Zeroth-order estimation, nonsmoothness, and sketching. The estimator machinery is standard (Nesterov & Spokoiny, 2017; Ghadimi & Lan, 2013; Shamir, 2017; Flaxman et al., 2005); recent work studies comparison-based random search (Chayti et al., 2025; El Bakkali et al., 2026a) and heavy-tailed value noise (El Bakkali et al., 2026b), all treating the objective as a single black box. The mean-versus- second-moment distinction for the bias is due to Berahas et al. (2022). For nonsmooth objectives Goldstein stationarity is the appropriate target (Clarke, 1990; Lin et al., 2022; Kornowski & Shamir, 2024), and Fischer & Paccagnan (2026) give a notion tailored to a smooth leader; equilibrium response maps are generally only Lipschitz (Lindberg & Engelson, 2015), so our smooth analysis is complementary there. Recursion (7) is structurally the EF21 mechanism (Richtárik et al., 2021) and a sketch-and-project iteration (Gower & Richtárik, 2015); Huang et al. (2022) study zeroth-order momentum without identifying it as a sketch step size.

5 EXPERIMENTS

We evaluate on five settings spanning the three areas: multiproduct pricing (six weekly instances, $d = 10, 40$ buyers, 100 simulations, 5000 samples) and strategic classification (credit data of Yeh & Lien (2009), $d = 12$, five response strengths) for performative prediction; the 8×8 gridworld of Mandal et al. (2023) for performative RL; and the 225 routing and 427 security instances of Fischer & Paccagnan (2026) for bilevel games. Baselines are the strongest published method for each area operating under (1): GZO-NS and GZO-HS (Hikima et al., 2025), ZO-OGVR (Hikima & Takeda, 2025b), a Gaussian two-point method ZO-TG, ZOS and PZOS (Fischer & Paccagnan, 2026), and a vanilla policy gradient. On the RL setting we additionally report PePG (Basu et al., 2025) as a reference with privileged information: it is given exact reward and transition response Jacobians, so it differentiates the environment's response instead of querying it, which is the assumption our setting is defined by not making. Its curve is not a bound on what other methods can attain, but it indicates what a differentiable response model is worth. We run PISO and PISO^2 with Gaussian probes and with a variant cycling through the columns of a random orthonormal frame. The latter has the sphere family as its one-step marginal but is dependent across consecutive steps, so Assumption 3 does not cover it and the theory applies to it only heuristically. Full protocols, the remaining figures, a metrics table and all parameter ablations are in Appendix L.

On pricing (Figure 1) the ordering is stable across all six instances: the cycling variants of PISO and PISO^2 are best, the Gaussian variants follow, then the guided baselines, then the two methods that ignore g_{K}. Strategic classification is more clear-cut: every PISO variant attains a test loss between 0.618 and 0.641 at every response strength against 0.821 for the best baseline, with accuracy higher by four to ten points, and the methods that ignore g_{K} degrade as the response strengthens while the PISO variants are flat (Table 3). On the bilevel games (Figure 2) two outcomes are worth separating: on routing PISO and PZOS are indistinguishable at convergence, confirmed by per-instance scatter plots lying on the diagonal, and what PISO gains is speed; on security PISO is below PZOS throughout and across the dimension range $d \in [100, 300]$. On performative RL, PISO is best among methods that query the environment, with the skyline faster still (Figure 4).

The ablations (Appendix L.1) test the analysis directly. The ρ sweep locates the optimum at $\rho = 0.92$ in all four variants, the nearest grid point to the predicted $\rho^\star \in \{0.900, 0.917\}$, and every value at or below the predicted neutrality threshold $1 - 2/c_d$ descends and then *rises*, which is the signature of a residual iteration whose sketch step size has left its stability window. The ρ_K sweep is informative in the other direction: with the growing batches used, $\rho_K = 0$ is already among the best choices, which is what the analysis predicts when n_K is large. The components ablation confirms that neither the known component alone nor an uncentred zeroth-order correction is stable on its own.

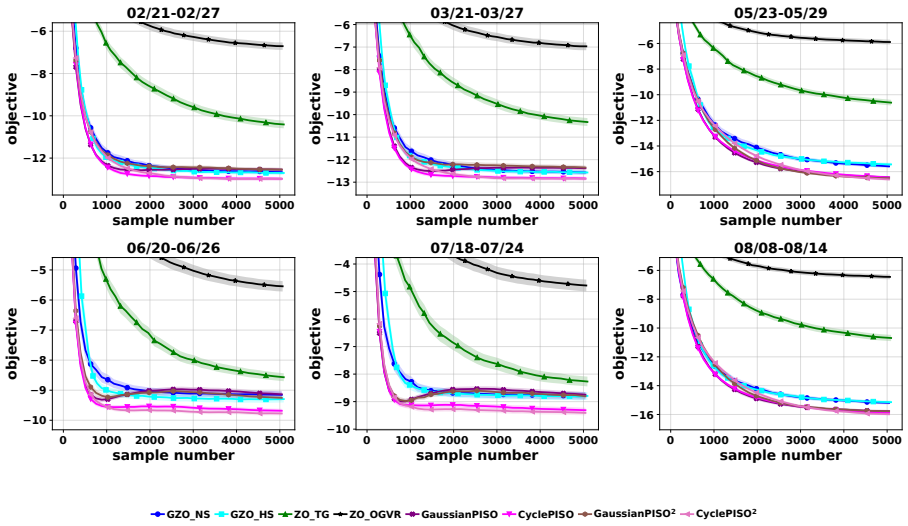
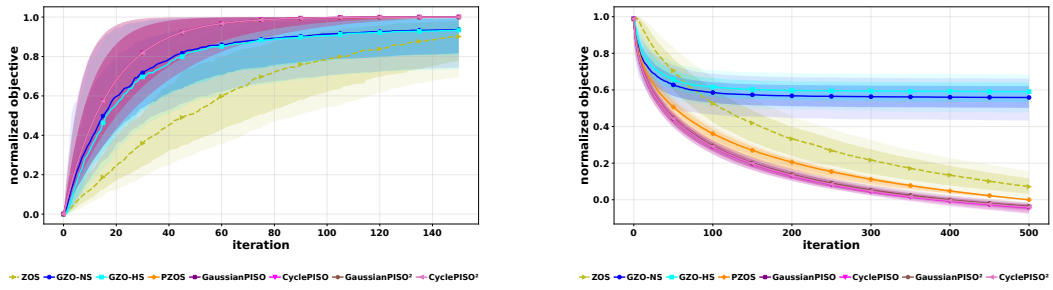


Figure 1: Multiproduct pricing, six weekly instances. Objective (lower is better) against samples consumed; means over 100 simulations with one-standard-error bands. All four PISO variants separate from the guided baselines early in the run and the cycling variants attain the lowest objective on every instance. ZO-TG and ZO-OGVR, which do not use the computable component, remain far behind.



(a) routing, 225 instances

(b) security, 427 instances

Figure 2: Bilevel games with an opaque lower level; per-instance normalized objective against iteration, median with interquartile and 10–90 percentile bands. (a) Routing is a maximization problem: PISO attains the same endpoint as PZOS but reaches it well before the guided baselines. (b) Security is a minimization problem: PISO separates from PZOS from the first iterations and ends below it.

6 Conclusion

We introduced PISO for stochastic optimization with a partially accessible gradient, a structure shared by performative prediction, performative reinforcement learning, and bilevel optimization with a black-box lower level. PISO tracks the residual and centres each zeroth-order observation on its current estimate, making the probing variance depend on the tracking error rather than the residual's magnitude. Our analysis establishes nonconvex convergence guarantees with one probe direction per update and no uniform bound on the unavailable gradient component. These guarantees capture gains from slowly varying residuals, while full-objective smoothness still controls finite-difference bias. Experiments across five settings demonstrate improvements in solution quality or convergence speed over baselines with comparable oracle access.

Our analysis assumes a smooth objective, which equilibrium response maps need not produce. Extending residual tracking to Goldstein stationarity would broaden its applicability to nonsmooth bilevel problems. Analysing the empirically strong orthogonal cycling variants, whose dependent probes fall outside our assumptions, remains open.

AI USE STATEMENT

We used a large language model as an assistant for restructuring and editing the manuscript, checking and tightening proofs, drafting parts of the experiment code, and suggesting related work. The authors verified all such output. The research ideas, the method and the main results are the authors' own, and the authors take full responsibility for the content of this work.

REPRODUCIBILITY STATEMENT

All proofs are in Appendices B–J, with every constant derived explicitly and the assumptions restated in Appendix A. Experimental protocols, problem definitions, baseline configurations, and per-setting hyperparameters are in Appendix L. The classification data is the public credit dataset of Yeh & Lien (2009). Anonymized code for all methods, baselines and experiments is available at `https://anonymous.4open.science/r/PISO-EA02/`.

REFERENCES

Debabrota Basu, Udvas Das, Brahim Driss, and Uddalak Mukherjee. Performative policy gradient: Optimality in performative reinforcement learning. *arXiv preprint arXiv:2512.20576*, 2025.

James Bell, Linda Linsefors, Caspar Oesterheld, and Joar Skalse. Reinforcement learning in newcomblike environments. *Advances in Neural Information Processing Systems*, 34:22146–22157, 2021.

Albert S. Berahas, Liyuan Cao, Krzysztof Choromanski, and Katya Scheinberg. A theoretical and empirical comparison of gradient approximations in derivative-free optimization. *Foundations of Computational Mathematics*, 22(2):507–560, 2022.

El Mahdi Chayti and Sai Praneeth Karimireddy. Optimization with access to auxiliary information. *Transactions on Machine Learning Research*, 2024. arXiv:2206.00395.

El Mahdi Chayti, Martin Jaggi, and Nikita Doikov. Unified convergence theory of stochastic and variance-reduced cubic newton methods. *Transactions on Machine Learning Research*, 2024. arXiv:2302.11962.

El Mahdi Chayti, Taha El Bakkali El Kadi, Omar Saadi, and Martin Jaggi. Stochastic optimization with random search. *arXiv preprint arXiv:2510.15610*, 2025.

Frank H. Clarke. *Optimization and nonsmooth analysis*. SIAM, 1990.

Shisheng Cui and Uday V. Shanbhag. On the computation of equilibria in monotone and potential stochastic hierarchical games. *Mathematical Programming*, 198(2):1227–1285, 2023.

Shisheng Cui, Uday V. Shanbhag, and Farzad Yousefian. Complexity guarantees for an implicit smoothing-enabled method for stochastic MPECs. *Mathematical Programming*, 198(2):1153–1225, 2023.

Edwige Cyffers, Muni Sreenivas Pydi, Jamal Atif, and Olivier Cappé. Optimal classification under performative distribution shift. In *Advances in Neural Information Processing Systems*, volume 37, pp. 68144–68160, 2024.

Edwige Cyffers, Alireza Mirrokni, and Marco Mondelli. Optimal regularization for performative learning. In *Proceedings of the International Conference on Machine Learning*, 2026. arXiv:2510.12249.

Taha El Bakkali, El Mahdi Chayti, and Omar Saadi. Nonsmooth optimization with zeroth order comparison feedback. *arXiv preprint arXiv:2602.05622*, 2026a.

Taha El Bakkali, El Mahdi Chayti, Qiuyi Zhang, Imane Rahali, and Omar Saadi. Stochastic zeroth-order optimization under heavy-tailed noise. *arXiv preprint arXiv:2605.17394*, 2026b.

Miriam Fischer and Dario Paccagnan. Black-box followers, white-box leaders: Partial zeroth-order methods for mpecs. *arXiv preprint arXiv:2605.16685*, 2026.

Abraham D. Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pp. 385–394, 2005.

Ethan Frankel, Kshitij Kulkarni, Dmitriy Drusvyatskiy, Sewoong Oh, and Lillian J. Ratliff. Finite-time convergence rates in stochastic stackelberg games with smooth algorithmic agents. In *Proceedings of the International Conference on Machine Learning*, 2025.

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

Robert M. Gower and Peter Richtárik. Randomized iterative methods for linear systems. *SIAM Journal on Matrix Analysis and Applications*, 36(4):1660–1690, 2015.

Yuya Hikima and Akiko Takeda. Zeroth-order gradient estimators for stochastic problems with decision-dependent distributions. *arXiv preprint arXiv:2510.24929*, 2025a.

Yuya Hikima and Akiko Takeda. Zeroth-order methods for nonconvex stochastic problems with decision-dependent distributions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17195–17203, 2025b.

Yuya Hikima, Hiroshi Sawada, and Akinori Fujino. Guided zeroth-order methods for stochastic non-convex problems with decision-dependent distributions. In *Proceedings of the International Conference on Machine Learning*, 2025.

Feihu Huang, Shangqian Gao, Jian Pei, and Heng Huang. Accelerated zeroth-order and first-order momentum methods from mini to minimax optimization. *Journal of Machine Learning Research*, 23(36):1–70, 2022.

Zachary Izzo, Lexing Ying, and James Zou. How to learn when data reacts to your model: performative gradient descent. In *Proceedings of the International Conference on Machine Learning*, pp. 4641–4650, 2021.

Zachary Izzo, James Zou, and Lexing Ying. How to learn when data gradually reacts to your model. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 3998–4035, 2022.

Meena Jagadeesan, Tijana Zrnic, and Celestine Mendler-Dünner. Regret minimization with performative feedback. In *Proceedings of the International Conference on Machine Learning*, pp. 9760–9785, 2022.

Guy Kornowski and Ohad Shamir. An algorithm with optimal dimension-dependence for zero-order nonsmooth nonconvex stochastic optimization. *Journal of Machine Learning Research*, 2024.

Sagi Levanon and Nir Rosenfeld. Strategic classification made practical. In *Proceedings of the International Conference on Machine Learning*, pp. 6243–6253, 2021.

Qiang Li and Hoi-To Wai. Stochastic optimization schemes for performative prediction with non-convex loss. *Advances in Neural Information Processing Systems*, 37:8673–8697, 2024.

Licong Lin and Tijana Zrnic. Plug-in performative optimization. In *Proceedings of the International Conference on Machine Learning*, pp. 30546–30565, 2024.

Tianyi Lin, Zeyu Zheng, and Michael Jordan. Gradient-free methods for deterministic and stochastic nonsmooth nonconvex optimization. *Advances in Neural Information Processing Systems*, 2022.

P. O. Lindberg and Leonid Engelson. Tolled multi-class traffic equilibria and toll sensitivities. *EURO Journal on Transportation and Logistics*, 4(2), 2015.

Haitong Liu, Qiang Li, and Hoi-To Wai. Two-timescale derivative free optimization for performative prediction with markovian data. In *Proceedings of the International Conference on Machine Learning*, pp. 31425–31450, 2024.

Alvaro Maggiar, Andreas Wächter, Irina S. Dolinskaya, and Jeremy Staum. A derivative-free trust-region algorithm for the optimization of functions smoothed via gaussian convolution using adaptive multiple importance sampling. *SIAM Journal on Optimization*, 28(2):1478–1507, 2018.

Debmalya Mandal and Goran Radanovic. Performative reinforcement learning with linear markov decision process. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 3232–3240. PMLR, 2025.

Debmalya Mandal, Stelios Triantafyllou, and Goran Radanovic. Performative reinforcement learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23642–23680. PMLR, 2023.

Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. In *Advances in Neural Information Processing Systems*, pp. 4929–4939, 2020.

John P. Miller, Juan C. Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *Proceedings of the International Conference on Machine Learning*, pp. 7710–7720, 2021.

Mehrnaz Mofakhami, Ioannis Mitliagkas, and Gauthier Gidel. Performative prediction with neural networks. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 11079–11093, 2023.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the International Conference on Machine Learning*, pp. 7599–7609, 2020.

Ben Rank, Stelios Triantafyllou, Debmalya Mandal, and Goran Radanovic. Performative reinforcement learning in gradually shifting environments. *arXiv preprint arXiv:2402.09838*, 2024.

Mitas Ray, Lillian J. Ratliff, Dmitriy Drusvyatskiy, and Maryam Fazel. Decision-dependent risk minimization in geometrically decaying dynamic environments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 8081–8088, 2022.

Peter Richtárik, Igor Sokolov, and Ilyas Fatkhullin. EF21: A new, simpler, theoretically better, and practically faster error feedback. In *Advances in Neural Information Processing Systems*, volume 34, pp. 4384–4396, 2021.

Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52), 2017.

Daouda Sow, Kaiyi Ji, and Yingbin Liang. On the convergence theory for Hessian-free bilevel algorithms. *Advances in Neural Information Processing Systems*, 35, 2022.

I-Cheng Yeh and Che-hui Lien. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2): 2473–2480, 2009.

APPENDIX CONTENTS

A Overview of the appendix and summary of notation **14**

B The structural identity **15**

B.1 What centring buys 16

C The probe distribution **16**

D The finite-difference bias **17**

E The tracking lemma **19**

F Proof of Theorem 2 **21**

G Complexity **22**

G.1 Regime A: shared randomness 22

G.2 Regime B: independent randomization 24

G.3 Regime C: deterministic oracle with tolerance 26

G.4 Where L, L_v and L_K enter 26

H Initializing the known-component state **27**

I Setting the parameters **28**

I.1 Discussion of the prescriptions 28

I.2 The momentum ρ 29

I.3 The mixing weight λ 29

I.4 The weight γ 30

J Probes per update **31**

K Deployment accounting **31**

K.1 Deployments, the binding cost in performative prediction 31

L Additional experimental results **32**

L.1 Ablations 34

L.2 Experimental hyperparameters 34

A Overview of the appendix and summary of notation

Appendix B proves the structural identity (Lemma 1) and explains what centring buys. Appendix C collects the properties of the probe distribution and proves the constants of Table 2. Appendix D treats the finite-difference bias, separating its mean from its second moment; this separation is what controls the dimension dependence of the rates. Appendix E proves the tracking lemma, the technical core of the analysis. Appendix F proves Theorem 2, and Appendix G the complexity corollaries. Appendix I proves the statements of Appendix I, and Appendix J the probe-count claim of Remark 32.

Every symbol below has one meaning throughout. Pairs that look similar are listed together so they can be checked at a glance.

x_k	iterate	u_k, s_k	probe and score
m_k	residual state	$P_k = s_k u_k^\top$	sketch matrix
R_k	residual observation, (6)	c_d	sketch constant, (19)
G_k	search direction	$\varsigma^2, \varsigma_4^2, \varsigma_6^2, \varsigma_8^2$	probe moments, (20)
$v(x)$	residual, (4)	ν, κ	smoothing measure and scale, (21)
$E_k = m_k - v(x_k)$	tracking error	χ_1, χ_2	smoothing moments, (22)
$\Delta_k = G_k - \nabla F(x_k)$	direction error	b_k	finite-difference bias (scalar)
Φ_k	Lyapunov function	$\bar{b}_k, \bar{B}_k$	bias vectors, (29)
ω	weight on $\|E_k\|^2$ in Φ_k	$\Sigma_{b,k}^2$	$\mathbb{E}_k[b_k^2 \|s_k\|^2]$
$\rho,\ \beta = 1-\rho$	momentum, sketch step	δ_k	known-oracle error, scaled by γ
$\lambda,\ \tilde{\beta} = 1-(1-\lambda)\rho$	mixing, readout gain	$\sigma_{K,k}^2, \sigma_{D,k}^2$	oracle variances
$\theta(b) = 1 - 2b + c_d b^2$	contraction factor	ξ_k	directional-oracle noise
γ	known-component weight	δ_{tol}	solver tolerance (Regime C)
η, μ_k	step size, probe radius	n_F, n_K	per-iteration batch sizes
L, L_K, L_v, Q	smoothness constants	N	total function evaluations

We abbreviate $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot \mid \mathcal{F}_k]$, where $\mathcal{F}_k$ is the σ-field generated by all randomness drawn before iteration k, so that $x_0, \dots, x_k$, $m_0, \dots, m_k$ and $h_0, \dots, h_k$ are all $\mathcal{F}_k$-measurable while the iteration-k draws (u_k, s_k), $\widehat{g}_{\mathrm{K}}(x_k)$ and the function-value noise are not. We write $v_k := v(x_k)$ and $\nabla F_k := \nabla F(x_k)$, and set $\theta := \theta(\tilde{\beta})$.

Standing assumptions. We restate them for reference.

Assumption 4 (Smoothness; restatement of Assumption 1)**.** F is L-smooth and bounded below by $F^\star$, with $\Delta_F := F(x_0) - F^\star$; and g_{K} is L_K-Lipschitz. We write L_v for the smallest constant satisfying

$$\mathbb{E}_k \|v_{k+1} - v_k\|^2 \ \le\ L_v^2\, \eta^2\ \mathbb{E}_k \|G_k\|^2. \tag{17}$$

Finiteness of L_v is implied rather than assumed: $v = \nabla F - \gamma g_{\mathrm{K}}$ is $(L + \gamma L_K)$-Lipschitz, so (17) holds with $L_v \le L + \gamma L_K$ because $\|x_{k+1} - x_k\| = \eta \|G_k\|$. Using the smallest valid constant along the trajectory rather than the global Lipschitz modulus only strengthens the results.

Assumption 5 (Oracles; restatement of Assumption 2)**.** $\mathbb{E}_k[\widehat{g}_{\mathrm{K}}(x_k)] = g_{\mathrm{K}}(x_k)$ and $\mathbb{E}_k \|\widehat{g}_{\mathrm{K}}(x_k) - g_{\mathrm{K}}(x_k)\|^2 \le \sigma_{K,k}^2$, and $\widehat{g}_{\mathrm{K}}(x_k)$ is drawn independently of (u_k, s_k); consequently $\mathbb{E}_k[\delta_k] = 0$ and $\mathbb{E}_k \|\delta_k\|^2 \le \gamma^2 \sigma_{K,k}^2$. In (3), b_k is a deterministic function of (x_k, u_k, μ_k), and $\mathbb{E}[\xi_k \mid \mathcal{F}_k, u_k, s_k, \delta_k] = 0$ with $\mathbb{E}[\xi_k^2 \mid \mathcal{F}_k, u_k, s_k] \le \sigma_{D,k}^2$.

Note that ξ_k is required to be conditionally mean-zero, not independent of u_k. This matters: in a performative problem the noise of $\widehat{F}(x_k \pm \mu_k u_k)$ depends on u_k, because a different decision is deployed at each of the two query points.

Assumption 6 (Probe distribution; restatement of Assumption 3)**.** The pair (u_k, s_k) is drawn i.i.d. given $\mathcal{F}_k$ and independently of δ_k; no independence from ξ_k is assumed, only the conditional mean

and variance conditions of Assumption 5. It satisfies, for every $\mathcal{F}_k$-measurable z,

$$\text{(D1)} \quad \mathbb{E}_k\left[s_k u_k^\top\right] = I_d, \tag{18}$$

$$\text{(D2)} \quad \mathbb{E}_k\left\|\langle z, u_k\rangle s_k\right\|^2 \le c_d\|z\|^2, \tag{19}$$

$$\text{(D3)} \quad \mathbb{E}_k\|s_k\|^2 \le \varsigma^2, \quad \mathbb{E}_k\left[\|s_k\|^2\|u_k\|^2\right] \le \varsigma_4^2, \quad \mathbb{E}_k\left[\|u_k\|^4\|s_k\|^2\right] \le \varsigma_6^2, \quad \mathbb{E}_k\left[\|u_k\|^6\|s_k\|^2\right] \le \varsigma_8^2, \tag{20}$$

$$\text{(D4)} \quad \mathbb{E}\left[\tfrac{F(x+\mu u)-F(x-\mu u)}{2\mu}\, s\right] = \nabla F_{\nu,\kappa\mu}(x), \qquad F_{\nu,\tau}(x) := \mathbb{E}_{w\sim\nu}\left[F(x+\tau w)\right], \tag{21}$$

for some symmetric probability measure ν on $\mathbb{R}^d$ and scale $\kappa > 0$. Given (D4) we set

$$\chi_1 \;:=\; \kappa\, \mathbb{E}_\nu\|w\|, \qquad \chi_2 \;:=\; \kappa^2\, \mathbb{E}_\nu\|w\|^2. \tag{22}$$

Only (D1) and (D2) enter the contraction; (D3) and (D4) enter only through noise and bias constants.

B THE STRUCTURAL IDENTITY

Proof of Lemma 1. (i) Two-timescale readout. Substituting $m_{k+1} = \rho m_k + (1-\rho)R_k$ into the readout,

$$\lambda R_k + (1-\lambda)m_{k+1} = \lambda R_k + (1-\lambda)\big[\rho m_k + (1-\rho)R_k\big] = (1-\lambda)\rho\, m_k + \big[\lambda + (1-\lambda)(1-\rho)\big]R_k.$$

Expanding the bracket, $\lambda + (1-\lambda)(1-\rho) = \lambda + 1 - \rho - \lambda + \lambda\rho = 1 - \rho(1-\lambda) = \tilde\beta$, while the coefficient of m_k is $(1-\lambda)\rho = 1 - \tilde\beta$. This proves (i).

(ii) State recursion. We first identify the innovation. Substituting $\gamma\widehat{g}_{\mathrm{K}}(x_k) = \gamma g_{\mathrm{K}}(x_k) + \delta_k$ and (3) into line 4 of Algorithm 1, and using $\nabla F_k - \gamma g_{\mathrm{K}}(x_k) = v_k$,

$$\begin{aligned} D_k - \langle\gamma\widehat{g}_{\mathrm{K}}(x_k) + m_k, u_k\rangle &= \langle\nabla F_k, u_k\rangle + b_k + \xi_k - \langle\gamma g_{\mathrm{K}}(x_k) + \delta_k + m_k, u_k\rangle \\ &= \langle\nabla F_k - \gamma g_{\mathrm{K}}(x_k) - m_k - \delta_k, u_k\rangle + b_k + \xi_k \\ &= \langle -E_k - \delta_k, u_k\rangle + b_k + \xi_k, \end{aligned} \tag{23}$$

where the last line uses $E_k = m_k - v_k$. Therefore

$$R_k - m_k = \big(\langle -E_k - \delta_k, u_k\rangle + b_k + \xi_k\big)s_k. \tag{24}$$

Since $m_{k+1} - m_k = \beta(R_k - m_k)$ and $\langle E_k, u_k\rangle s_k = s_k u_k^\top E_k = P_k E_k$,

$$m_{k+1} - v_k = (I_d - \beta P_k)E_k + \beta\big(-\langle\delta_k, u_k\rangle + b_k + \xi_k\big)s_k. \tag{25}$$

Adding and subtracting v_{k+1} and using $E_{k+1} = m_{k+1} - v_{k+1}$ gives (ii).

(iii) Direction error. Identity (23) does not involve β, so replacing β by $\tilde\beta$ in the derivation of (25) and using (i) gives

$$\tilde m_{k+1} - v_k = (I_d - \tilde\beta P_k)E_k + \tilde\beta\big(-\langle\delta_k, u_k\rangle + b_k + \xi_k\big)s_k, \tag{26}$$

where $\tilde m_{k+1} := \lambda R_k + (1-\lambda)m_{k+1}$. Since $G_k = \gamma h_{k+1} + \tilde m_{k+1}$ and $\nabla F_k = \gamma g_{\mathrm{K}}(x_k) + v_k$,

$$\Delta_k = \gamma\big(h_{k+1} - g_{\mathrm{K}}(x_k)\big) + \big(\tilde m_{k+1} - v_k\big) = \big[(1-\beta_K)\zeta_k + \beta_K\delta_k\big] + \big(\tilde m_{k+1} - v_k\big),$$

using (27) for the first bracket, and grouping the two δ_k terms as $\beta_K\delta_k - \tilde\beta\langle\delta_k, u_k\rangle s_k = (\beta_K I_d - \tilde\beta P_k)\delta_k$ with (26) gives (iii). Note that δ_k no longer appears standing alone: with the fresh estimate in the step one would obtain the extra term δ_k, and hence $\theta\gamma^2\sigma_{K,k}^2$ in $\mathbb{E}_k\|\Delta_k\|^2$.

(iv) Known-component recursion. From $h_{k+1} = (1-\beta_K)h_k + \beta_K\widehat{g}_{\mathrm{K}}(x_k)$,

$$\gamma\big(h_{k+1} - g_{\mathrm{K}}(x_k)\big) = (1-\beta_K)\gamma\big(h_k - g_{\mathrm{K}}(x_k)\big) + \beta_K\gamma\big(\widehat{g}_{\mathrm{K}}(x_k) - g_{\mathrm{K}}(x_k)\big) = (1-\beta_K)\zeta_k + \beta_K\delta_k, \tag{27}$$

which is the substitution used in (iii). Adding $\gamma\big(g_{\mathrm{K}}(x_k) - g_{\mathrm{K}}(x_{k+1})\big)$ to both sides of (27) gives $\zeta_{k+1} = \gamma(h_{k+1} - g_{\mathrm{K}}(x_{k+1})) = (1-\beta_K)\zeta_k + \beta_K\delta_k + \gamma\big(g_{\mathrm{K}}(x_k) - g_{\mathrm{K}}(x_{k+1})\big)$, which is (iv).

We stress the indexing: ζ_k is defined from h_k, not h_{k+1}, so that it is $\mathcal{F}_k$-measurable. The state h_{k+1} used in G_k involves the step-k sample $\widehat{g}_{\mathrm{K}}(x_k)$, and (27) is what isolates that sample as the mean-zero term $\beta_K\delta_k$. □

family	c_d	ς^2	χ_1	χ_2
Gaussian, $u = s \sim \mathcal{N}(0, I_d)$	$d+2$	d	$\sqrt{d}$	d
Sphere, $u = s = \sqrt{d}\, w$, $\|w\| = 1$	d	d	$\sqrt{d}$	d
Coordinate, $u = s = \sqrt{d}\, e_i$	d	d	$d/2$	$d^{3/2}/3$

Table 2: Constants for three standard families. Gaussian and sphere satisfy the smoothing identity (D4), with ν Gaussian and $\kappa = 1$ for the first and ν uniform on the unit ball and $\kappa = \sqrt{d}$ for the second, and therefore share $\chi_1 = \sqrt{d}$. The coordinate family *fails* (D4): its mean estimator is the coordinate-wise finite-difference gradient, which is not the gradient of any smoothing, and it loses a factor $\sqrt{d}$ in χ_1. This is the structural cause of the known gap between coordinate and random directions.

Lemma 5 (Contraction factor)**.** *For every $\mathcal{F}_k$-measurable z and every $b \in \mathbb{R}$, $\mathbb{E}_k \|(I_d - bP_k)z\|^2 \le \theta(b)\|z\|^2$ with θ as in* (8)*, with equality whenever* (19) *holds with equality. Consequently $\theta(b) < 1$ if and only if $0 < b < 2/c_d$, with $\theta(2/c_d) = 1$ (mean-square neutral) and $\theta(b) > 1$ for $b > 2/c_d$ or $b < 0$; θ is minimized at $b = 1/c_d$; and $\theta(1/c_d) = 1 - 1/c_d$.*

Proof. Expanding and using (18) for the cross term and (19) for the last,

$$\mathbb{E}_k \|(I_d - bP_k)z\|^2 = \|z\|^2 - 2b\, z^\top \mathbb{E}_k[P_k]z + b^2\, \mathbb{E}_k \|\langle z, u_k\rangle s_k\|^2 \le \|z\|^2 - 2b\|z\|^2 + b^2 c_d \|z\|^2.$$

The quadratic $\theta(b) - 1 = b(c_d b - 2)$ is negative exactly for $0 < b < 2/c_d$, giving the stated window; its minimizer is $b = 1/c_d$ with value $1 - 1/c_d$. □

B.1 What centring buys

We justify §B.1. Consider the uncentred observation $R_k^{\mathrm{fresh}} = \big(D_k - \langle \gamma \widehat{g}_{\mathrm{K}}(x_k), u_k\rangle\big) s_k$, in which m_k does not appear inside the inner product. Repeating the computation (23) without the m_k term gives $D_k - \langle \gamma \widehat{g}_{\mathrm{K}}(x_k), u_k\rangle = \langle v_k - \delta_k, u_k\rangle + b_k + \xi_k$, hence $R_k^{\mathrm{fresh}} = P_k v_k + (-\langle \delta_k, u_k\rangle + b_k + \xi_k) s_k$ and, with $m_{k+1} = \rho m_k + (1-\rho) R_k^{\mathrm{fresh}}$,

$$E_{k+1} = \rho E_k + \beta\big(P_k - I_d\big)v_k + \beta\big(-\langle \delta_k, u_k\rangle + b_k + \xi_k\big)s_k + (v_k - v_{k+1}). \tag{28}$$

The sketch now multiplies v_k rather than E_k. Taking second moments, the term $\beta(P_k - I_d)v_k$ contributes $\beta^2(c_d - 1)\|v_k\|^2$ at each step, and since the remaining contraction is only $\rho = 1 - \beta$, its stationary contribution to $\mathbb{E}\|E_k\|^2$ is of order $\beta c_d \|v_k\|^2$. With $G_U := \sup_x \|v(x)\|$, driving this below ε^2 requires $\beta \lesssim \varepsilon^2/(c_d G_U^2)$ in addition to the restrictions of Theorem 2, so the complexity acquires a factor $c_d G_U^2/\varepsilon^2$. The centred recursion of Lemma 1(ii) has E_k in place of v_k and therefore carries no such term. In particular, the uncentred variant is not the $\rho \to 0$ limit of PISO; it is a different algorithm whose complexity depends on the magnitude of the opaque component.

C The probe distribution

Lemma 6 (Constants of Table 2)**.** *Let $d \ge 2$.*

(a) Gaussian*, $u = s \sim \mathcal{N}(0, I_d)$. Then* (18) *holds;* (19) *holds with equality and $c_d = d + 2$; $\varsigma^2 = d$, $\varsigma_4^2 = d(d+2) \le 2d^2$, $\varsigma_6^2 = d(d+2)(d+4) \le 9d^3$ and $\varsigma_8^2 \le 54d^4$; and* (21) *holds with $\nu = \mathcal{N}(0, I_d)$ and $\kappa = 1$, so $\chi_1 \le \sqrt{d}$ and $\chi_2 = d$.*

(b) Sphere*, $u = s = \sqrt{d}\, w$ with w uniform on the unit sphere. Then* (18) *holds;* (19) *holds with equality and $c_d = d$; $\varsigma^2 = d$, $\varsigma_4^2 = d^2$, $\varsigma_6^2 = d^3$, $\varsigma_8^2 = d^4$, all exactly; and* (21) *holds with ν uniform on the unit ball and $\kappa = \sqrt{d}$, so $\chi_1 \le \sqrt{d}$ and $\chi_2 \le d$.*

(c) Coordinate*, $u = s = \sqrt{d}\, e_i$ with i uniform on $[d]$. Then* (18) *holds,* (19) *holds with equality and $c_d = d$, and the ς moments coincide with (b). Property* (21) ***fails****, and the direct bounds are $\chi_1 = d/2$ under L-smoothness and $\chi_2 = d^{3/2}/3$ under a Q-Lipschitz Hessian.*

Proof. (a) $\mathbb{E}[uu^\top] = I_d$ gives (18). For (19), Isserlis' theorem gives $\mathbb{E}[\|u\|^2 uu^\top] = \sum_j \mathbb{E}[u_j^2 uu^\top] = (d+2)I_d$, so $\mathbb{E}\,\|\langle z,u\rangle u\|^2 = \mathbb{E}[\langle z,u\rangle^2\|u\|^2] = z^\top(d+2)I_d\,z = (d+2)\|z\|^2$ exactly. The chi-square moments $\mathbb{E}\,\|u\|^{2k} = d(d+2)\cdots(d+2k-2)$ give the ς values, with $d(d+2) \le 2d^2$, $d(d+2)(d+4) \le 9d^3$ and $d(d+2)(d+4)(d+6) \le 54d^4$ for $d \ge 2$. For (21), Gaussian integration by parts ($\mathbb{E}[g(u)u_i] = \mathbb{E}[\partial_i g(u)]$ for differentiable g of moderate growth) gives $\mathbb{E}[F(x+\mu u)u] = \mu\,\mathbb{E}[\nabla F(x+\mu u)]$; and since the law of u is symmetric, substituting $u \mapsto -u$ gives $\mathbb{E}[F(x-\mu u)u] = -\,\mathbb{E}[F(x+\mu u)u] = -\mu\,\mathbb{E}[\nabla F(x+\mu u)]$. Hence

$$\mathbb{E}\left[\tfrac{F(x+\mu u)-F(x-\mu u)}{2\mu}\,u\right] = \tfrac{1}{2\mu}\big(\mu\,\mathbb{E}[\nabla F(x+\mu u)]+\mu\,\mathbb{E}[\nabla F(x+\mu u)]\big) = \mathbb{E}[\nabla F(x+\mu u)] = \nabla F_{\mathcal{N},\mu}(x),$$

the last equality by differentiating under the expectation. So $\kappa = 1$, and $\chi_1 = \mathbb{E}\,\|u\| \le \sqrt{\mathbb{E}\,\|u\|^2} = \sqrt{d}$ by Jensen, $\chi_2 = \mathbb{E}\,\|u\|^2 = d$.

(b) Rotational invariance gives $\mathbb{E}[ww^\top] = I_d/d$, so $\mathbb{E}[su^\top] = d\,\mathbb{E}[ww^\top] = I_d$. Since $\|u\|^2 = \|s\|^2 = d$ deterministically, $\mathbb{E}\,\|\langle z,u\rangle s\|^2 = d^2\,\mathbb{E}\langle z,w\rangle^2 = d^2\cdot\|z\|^2/d = d\|z\|^2$, exactly, and the ς moments follow by substituting $\|u\|^2 = \|s\|^2 = d$. For (21) write $\tau := \sqrt{d}\,\mu$, so that $\mu u = \tau w$ and

$$\tfrac{F(x+\mu u)-F(x-\mu u)}{2\mu}\,s = \tfrac{\sqrt{d}\,\big(F(x+\tau w)-F(x-\tau w)\big)}{2\mu}\,w = \tfrac{d\,\big(F(x+\tau w)-F(x-\tau w)\big)}{2\tau}\,w.$$

By the identity of Flaxman et al. (2005), $\mathbb{E}\left[\frac{d}{\tau}F(x+\tau w)w\right] = \nabla F_{\mathcal{B},\tau}(x)$ with $\mathcal{B}$ uniform on the unit ball; combining the two query points as above gives (21) with $\nu = \mathcal{B}$ and $\kappa = \sqrt{d}$. Finally $\mathbb{E}_{\mathcal{B}}\,\|w\| = d/(d+1) \le 1$ and $\mathbb{E}_{\mathcal{B}}\,\|w\|^2 = d/(d+2) \le 1$, so $\chi_1 \le \sqrt{d}$ and $\chi_2 \le d$.

(c) $\mathbb{E}[su^\top] = d\,\mathbb{E}[e_i e_i^\top] = d\cdot I_d/d = I_d$. Also $\|\langle z,u\rangle s\|^2 = \langle z,\sqrt{d}e_i\rangle^2\cdot d = d^2 z_i^2$, whose mean is $d^2\|z\|^2/d = d\|z\|^2$. Since $\|u\|^2 = \|s\|^2 = d$, the ς moments match (b). For (D4), writing $\tau = \sqrt{d}\mu$ and repeating the computation in (b),

$$\mathbb{E}_k\left[\tfrac{F(x+\mu u)-F(x-\mu u)}{2\mu}s\right] = \sum_{i=1}^{d}\frac{F(x+\tau e_i)-F(x-\tau e_i)}{2\tau}\,e_i,$$

the coordinate-wise central-difference gradient. This is not the gradient of any smoothing $F_{\nu,\tau}$ in general, so (D4) fails. The direct bounds follow from the scalar estimates $\left|\frac{F(x+\tau e_i)-F(x-\tau e_i)}{2\tau} - \partial_i F(x)\right| \le \frac{L\tau}{2}$ under L-smoothness and $\le \frac{Q\tau^2}{6}$ under a Q-Lipschitz Hessian: summing d coordinates in ℓ_2 gives $\|\bar{b}\| \le \sqrt{d}\cdot\frac{L\tau}{2} = \frac{L\mu d}{2}$ and $\|\bar{b}\| \le \sqrt{d}\cdot\frac{Q\tau^2}{6} = \frac{Q\mu^2 d^{3/2}}{6}$, which are $L\mu\chi_1$ and $\frac{1}{2}Q\mu^2\chi_2$ with $\chi_1 = d/2$ and $\chi_2 = d^{3/2}/3$. □

Remark 7 (Sphere and Gaussian coincide). After the rescaling implicit in $\kappa = \sqrt{d}$, the two random families have the same c_d up to an additive constant, the same ς moments, and the same χ_1, χ_2. Every bound below therefore agrees in order between them. The coordinate family differs only through χ_1, χ_2, and loses a factor $\sqrt{d}$ in χ_1 because (D4) fails.

D THE FINITE-DIFFERENCE BIAS

Two vector-valued quantities are needed:

$$\bar{b}_k \;:=\; \mathbb{E}_k\left[b_k s_k\right], \qquad \bar{B}_k \;:=\; \mathbb{E}_k\left[b_k\|s_k\|^2 u_k\right], \tag{29}$$

alongside the second moment $\Sigma_{b,k}^2 = \mathbb{E}_k[b_k^2\|s_k\|^2]$. The distinction is the crux of the dimension dependence: the bias accumulates coherently across the $\sim 1/\beta$ steps of the momentum window, so the undamped contribution is $\|\bar{b}_k\|^2$, whereas the fluctuation of $b_k s_k$ about its mean is mean-zero and is damped by β like any other noise.

Lemma 8 (Second moment of the bias)**.** *Under* (20)*: (a) if ∇F is L-Lipschitz then $|b_k| \le \frac{L\mu_k}{2}\|u_k\|^2$ and $\Sigma_{b,k}^2 \le \frac{1}{4}L^2\mu_k^2\varsigma_6^2$; (b) if in addition $\nabla^2 F$ is Q-Lipschitz then $|b_k| \le \frac{Q\mu_k^2}{6}\|u_k\|^3$ and $\Sigma_{b,k}^2 \le \frac{1}{36}Q^2\mu_k^4\varsigma_8^2$.*

Proof. (a) By Taylor's theorem with integral remainder, $\left|F(x \pm \mu u) - F(x) \mp \mu\langle \nabla F(x), u\rangle\right| \le \frac{L\mu^2}{2}\|u\|^2$. Subtracting the two inequalities and dividing by 2μ gives $|b_k| \le \frac{L\mu}{2}\|u\|^2$, whence $\Sigma_{b,k}^2 = \mathbb{E}_k[b_k^2\|s_k\|^2] \le \frac{L^2\mu^2}{4}\,\mathbb{E}_k[\|u_k\|^4\|s_k\|^2]$ and (20) applies. (b) A Q-Lipschitz Hessian gives $\left|F(x \pm \mu u) - F(x) \mp \mu\langle \nabla F, u\rangle - \frac{\mu^2}{2}u^\top \nabla^2 F(x) u\right| \le \frac{Q\mu^3}{6}\|u\|^3$; the quadratic terms are equal at the two query points and cancel on subtraction, leaving $|b_k| \le \frac{Q\mu^2}{6}\|u\|^3$, and (20) applies again. □

Lemma 9 (Mean bias)**.** *Suppose* (21) *holds, with* χ_1, χ_2 *as in* (22)*. Then*

$$\bar{b}_k = \mathbb{E}_\nu\left[\nabla F(x_k + \kappa\mu_k w)\right] - \nabla F(x_k), \tag{30}$$

and consequently

$$\|\bar{b}_k\| \le L\,\mu_k\,\chi_1, \qquad \textit{and if } \nabla^2 F \textit{ is } Q\textit{-Lipschitz,} \qquad \|\bar{b}_k\| \le \tfrac{1}{2}\,Q\,\mu_k^2\,\chi_2. \tag{31}$$

Proof. By Assumption 5, $\mathbb{E}_k[\xi_k s_k] = 0$, since conditionally on (u_k, s_k) the factor s_k is determined and ξ_k has mean zero. By (18), $\mathbb{E}_k[\langle \nabla F_k, u_k\rangle s_k] = \nabla F_k$. Subtracting these from $\mathbb{E}_k[D_k s_k]$ and using (3) gives $\bar{b}_k = \mathbb{E}_k[D_k s_k] - \nabla F_k$, which by (21) equals $\nabla F_{\nu,\kappa\mu_k}(x_k) - \nabla F_k$. Differentiating under the expectation, $\nabla F_{\nu,\tau}(x) = \mathbb{E}_\nu[\nabla F(x + \tau w)]$, which is (30).

For the first bound in (31), Jensen and L-Lipschitzness of ∇F give

$$\|\bar{b}_k\| \le \mathbb{E}_\nu\left\|\nabla F(x_k + \kappa\mu_k w) - \nabla F(x_k)\right\| \le L\kappa\mu_k\;\mathbb{E}_\nu\,\|w\| = L\mu_k\chi_1.$$

For the second, symmetry of ν gives $\bar{b}_k = \frac{1}{2}\,\mathbb{E}_\nu\left[\nabla F(x_k + \tau w) + \nabla F(x_k - \tau w) - 2\nabla F(x_k)\right]$ with $\tau = \kappa\mu_k$. Writing $\nabla F(x \pm h) = \nabla F(x) \pm \nabla^2 F(x)h + r_\pm$ with $\|r_\pm\| \le \frac{Q}{2}\|h\|^2$, the linear terms cancel in the sum and $\|\nabla F(x + h) + \nabla F(x - h) - 2\nabla F(x)\| \le Q\|h\|^2$. Hence $\|\bar{b}_k\| \le \frac{1}{2}Q\kappa^2\mu_k^2\,\mathbb{E}_\nu\,\|w\|^2 = \frac{1}{2}Q\mu_k^2\chi_2$. □

Lemma 10 (Weighted mean bias)**.** $\|\bar{B}_k\| \le \Sigma_{b,k}\,\varsigma_4$.

Proof. By Jensen and then Cauchy–Schwarz, grouping $\|s_k\|$ with b_k,

$$\|\bar{B}_k\| \le \mathbb{E}_k\left[|b_k|\,\|s_k\|^2\|u_k\|\right] = \mathbb{E}_k\left[\big(|b_k|\|s_k\|\big)\big(\|s_k\|\|u_k\|\big)\right] \le \sqrt{\mathbb{E}_k[b_k^2\|s_k\|^2]}\;\sqrt{\mathbb{E}_k[\|s_k\|^2\|u_k\|^2]},$$

and (20) bounds the second factor by ς_4. □

Lemma 11 (Sharp bound for the random families)**.** *For the Gaussian and sphere families of Lemma 6,*

$$\|\bar{B}_k\| \;\le\; 3\,d\,L\,\mu_k\,\chi_1 \qquad \textit{and, if } \nabla^2 F \textit{ is } Q\textit{-Lipschitz,} \qquad \|\bar{B}_k\| \;\le\; \tfrac{3}{2}\,d\,Q\,\mu_k^2\,\chi_2. \tag{32}$$

Proof. For the sphere family $\|s_k\|^2 = d$ deterministically, so $\bar{B}_k = \mathbb{E}_k[b_k\|s_k\|^2 u_k] = d\;\mathbb{E}_k[b_k s_k] = d\,\bar{b}_k$ exactly, and (31) gives the claim with room to spare. For the Gaussian family, applying Gaussian integration by parts to $h(u) = F(x + \mu u)\|u\|^2$, whose partial derivative is $\partial_i h = \mu\,\partial_i F(x + \mu u)\|u\|^2 + 2F(x+\mu u)u_i$, gives $\mathbb{E}[F(x+\mu u)\|u\|^2 u] = \mu\,\mathbb{E}[\nabla F(x+\mu u)\|u\|^2] + 2\mu\,\mathbb{E}[\nabla F(x+\mu u)]$; combining the two query points as in Lemma 6(a) and subtracting $\mathbb{E}[\langle \nabla F(x), u\rangle\|u\|^2 u] = (d + 2)\nabla F(x)$ yields

$$\bar{B}_k = \mathbb{E}\left[\big(\nabla F(x_k + \mu_k u) - \nabla F(x_k)\big)\|u\|^2\right] + 2\bar{b}_k.$$

Hence $\|\bar{B}_k\| \le L\mu_k\,\mathbb{E}\,\|u\|^3 + 2L\mu_k\sqrt{d} \le L\mu_k(1.7d^{3/2} + 2\sqrt{d}) \le 3L\mu_k d^{3/2}$ for $d \ge 2$, using $\mathbb{E}\,\|u\|^3 \le (\mathbb{E}\,\|u\|^4)^{3/4} \le 2^{3/4}d^{3/2}$, and $\chi_1 = \sqrt{d}$ gives the stated form. The Q-Lipschitz version follows by symmetrizing as in Lemma 9. □

The generic bound of Lemma 10 needs no structure but is loose by a factor $\Theta(d)$: it gives $\|\bar{B}_k\| \le \frac{1}{2}L\mu_k\varsigma_6\varsigma_4 = \Theta(L\mu_k d^{5/2})$ against the $\Theta(L\mu_k d^{3/2})$ of (32). The difference matters, because it is exactly the gap between admitting $\beta = \Theta(1/d)$ and only $\beta = \Theta(1/d^2)$ in §G, and hence between a minimum deployment count of $\Theta(d\varepsilon^{-2})$ and $\Theta(d^2\varepsilon^{-2})$. We therefore use (32) for the random families and Lemma 10 only when no smoothing identity is available. The following proposition quantifies how far apart the mean and the root-mean-square of the bias can be; it is used only to interpret the rates, not to prove them.

Proposition 12 (The two bias scales differ by at least $\sqrt{d}$)**.** *Suppose $s_k = u_k$ and*

$$\mathbb{E}_k\left[\frac{\langle u_k, z\rangle^2}{\|u_k\|^2}\right] = \frac{\|z\|^2}{d} \qquad \text{for all } z \in \mathbb{R}^d, \tag{33}$$

which holds with equality for all three families of Lemma 6. Then for every bias function b_k,

$$\|\bar{b}_k\| \;\le\; \frac{1}{\sqrt{d}}\,\Sigma_{b,k}. \tag{34}$$

Proof. Fix a unit vector z and group $\|u_k\|$ with b_k:

$$\langle \bar{b}_k, z\rangle = \mathbb{E}_k\left[b_k\langle u_k, z\rangle\right] = \mathbb{E}_k\left[\big(b_k\|u_k\|\big)\cdot\frac{\langle u_k, z\rangle}{\|u_k\|}\right] \le \sqrt{\mathbb{E}_k\left[b_k^2\|u_k\|^2\right]}\,\sqrt{\mathbb{E}_k\left[\frac{\langle u_k, z\rangle^2}{\|u_k\|^2}\right]} = \frac{\Sigma_{b,k}}{\sqrt{d}},$$

by Cauchy–Schwarz and (33). Taking the supremum over unit z gives (34). For (33): for Gaussian u the ratio $u/\|u\|$ is uniform on the unit sphere and a uniform unit vector w satisfies $\mathbb{E}[ww^\top] = I_d/d$; for the sphere family $u = \sqrt{d}\,w$ so that $\langle u, z\rangle^2/\|u\|^2 = \langle w, z\rangle^2$; for the coordinate family $\langle u, z\rangle^2/\|u\|^2 = z_i^2$ with i uniform, whose mean is $\|z\|^2/d$. □

Remark 13 (Sharpness of (31) and (34))*.* Take Gaussian probes, $L = 1$ and $x = 0$. For the separable half-quadratic $F(x) = \frac{1}{2}\sum_i \max(0, x_i)^2$, which is 1-smooth with a Hessian discontinuous on the coordinate hyperplanes, one computes $b_k = \frac{\mu_k}{4}\sum_i \operatorname{sign}(u_i)u_i^2$, so $(\bar{b}_k)_j = \frac{\mu_k}{4}\,\mathbb{E}\,|u_j|^3 = \frac{\mu_k}{4}\cdot 2\sqrt{2/\pi}$ for every j and

$$\|\bar{b}_k\| = \mu_k\sqrt{d/(2\pi)} = \tfrac{1}{\sqrt{2\pi}}\,L\mu_k\sqrt{d}.$$

Hence the first bound in (31) is attained up to the constant $1/\sqrt{2\pi} \approx 0.399$, and the dimension dependence of Corollary 24 cannot be improved by sharpening $\bar{b}_k$. For the same F all cross terms vanish and $\Sigma_{b,k}^2 = \frac{\mu_k^2}{16}d(3d+12)$, so the ratio $\Sigma_{b,k}/\|\bar{b}_k\|$ equals $\frac{\sqrt{2\pi}}{4}\sqrt{3d+12} \to \frac{\sqrt{6\pi}}{4}\sqrt{d} \approx 1.085\sqrt{d}$, within 8.5% of (34). The inequality (34) admits no converse: for $F(x) = c\,x_1x_2x_3$ one has $b_k = c\mu_k^2 u_1u_2u_3$ and $\bar{b}_k = 0$ exactly, since every $\mathbb{E}[u_1u_2u_3u_j]$ vanishes, while $\Sigma_{b,k}^2 = c^2\mu_k^4(d+6) > 0$.

E THE TRACKING LEMMA

Set $\Sigma_{\mathrm{mz},k}^2 := c_d\gamma^2\sigma_{K,k}^2 + \varsigma^2\sigma_{D,k}^2$.

Lemma 14 (Known-component tracking)**.** *Under Assumptions 4–6 and $\beta_K \le 1$,*

$$\mathbb{E}_k\,\|\zeta_{k+1}\|^2 \;\le\; \Big(1-\frac{\beta_K}{2}\Big)\|\zeta_k\|^2 + \frac{3}{2}\beta_K^2\gamma^2\sigma_{K,k}^2 + \frac{3\gamma^2L_K^2\eta^2}{\beta_K}\;\mathbb{E}_k\,\|G_k\|^2. \tag{35}$$

Proof. Write Lemma 1(iv) as $\zeta_{k+1} = a_k + b_k' + c_k$ with $a_k := (1-\beta_K)\zeta_k$, $b_k' := \beta_K\delta_k$ and $c_k := \gamma(g_K(x_k) - g_K(x_{k+1}))$. Here a_k is $\mathcal{F}_k$-measurable and $\mathbb{E}_k[b_k'] = 0$, so $\mathbb{E}_k\,\|a_k + b_k'\|^2 = \|a_k\|^2 + \beta_K^2\,\mathbb{E}_k\,\|\delta_k\|^2$. The term c_k is *not* independent of b_k', since x_{k+1} depends on G_k which contains δ_k; we therefore separate it by Young's inequality rather than by an orthogonality argument. With parameter $\beta_K/2$,

$$\mathbb{E}_k\,\|\zeta_{k+1}\|^2 \le \Big(1+\frac{\beta_K}{2}\Big)\Big[(1-\beta_K)^2\|\zeta_k\|^2 + \beta_K^2\gamma^2\sigma_{K,k}^2\Big] + \Big(1+\frac{2}{\beta_K}\Big)\,\mathbb{E}_k\,\|c_k\|^2.$$

Since g_K is L_K-Lipschitz and $\|x_{k+1} - x_k\| = \eta\|G_k\|$, we have $\mathbb{E}_k\,\|c_k\|^2 \le \gamma^2L_K^2\eta^2\,\mathbb{E}_k\,\|G_k\|^2$, and $1 + 2/\beta_K \le 3/\beta_K$ for $\beta_K \le 1$. Finally $(1-\beta_K)^2 \le 1-\beta_K$ and $(1+\beta_K/2)(1-\beta_K) = 1 - \beta_K/2 - \beta_K^2/2 \le 1 - \beta_K/2$, while $(1+\beta_K/2) \le \frac{3}{2}$. □

Lemma 15 (Residual tracking)**.** *Under Assumptions 4–6 and $\beta \le 1/c_d$,*

$$\mathbb{E}_k\,\|E_{k+1}\|^2 \le \Big(1-\frac{\beta}{2}\Big)\|E_k\|^2 + \frac{5}{4}\beta^2\big(\Sigma_{\mathrm{mz},k}^2 + \Sigma_{b,k}^2\big) + 10\beta\|\bar{b}_k\|^2 + 10\beta^3\|\bar{B}_k\|^2 + \frac{5L_v^2\eta^2}{\beta}\;\mathbb{E}_k\,\|G_k\|^2. \tag{36}$$

Proof. By Lemma 1(ii), decompose $E_{k+1} = a_k + \beta b_k s_k + \mathsf{z}_k + w_k$ with

$$a_k := (I_d - \beta P_k)E_k, \qquad \mathsf{z}_k := \beta\big(-\langle \delta_k, u_k\rangle + \xi_k\big)s_k, \qquad w_k := v_k - v_{k+1}.$$

The bias term $\beta b_k s_k$ is kept together with a_k and its cross term evaluated exactly; this is what makes the mean bias, rather than the second moment, the undamped quantity.

Step 1: z_k decouples. Conditionally on $(\mathcal{F}_k, u_k, s_k)$ the vector $a_k + \beta b_k s_k$ is determined, while z_k is linear in (δ_k, ξ_k), both of which are conditionally mean-zero by Assumption 5. Hence $\mathbb{E}_k\langle a_k + \beta b_k s_k, \mathsf{z}_k\rangle = 0$ and

$$\mathbb{E}_k \left\|a_k + \beta b_k s_k + \mathsf{z}_k\right\|^2 = \mathbb{E}_k \left\|a_k + \beta b_k s_k\right\|^2 + \mathbb{E}_k \left\|\mathsf{z}_k\right\|^2.$$

Step 2: $\mathbb{E}_k \|\mathsf{z}_k\|^2 \le \beta^2 \Sigma^2_{\mathrm{mz},k}$. Expanding the square,

$$\mathbb{E}_k \left\|\mathsf{z}_k\right\|^2 = \beta^2 \Big\{ \mathbb{E}_k \left[\langle \delta_k, u_k\rangle^2 \|s_k\|^2\right] - 2\,\mathbb{E}_k \left[\langle \delta_k, u_k\rangle \xi_k \|s_k\|^2\right] + \mathbb{E}_k \left[\xi_k^2 \|s_k\|^2\right] \Big\}.$$

The middle term vanishes because $\mathbb{E}[\xi_k \mid \mathcal{F}_k, u_k, s_k, \delta_k] = 0$ while the other factors are determined by that conditioning. For the first term, condition on δ_k, which is independent of (u_k, s_k), and apply (19) with $z = \delta_k$: $\mathbb{E}_k \left\|\langle \delta_k, u_k\rangle s_k\right\|^2 \le c_d\, \mathbb{E}_k \left\|\delta_k\right\|^2 \le c_d \gamma^2 \sigma^2_{K,k}$. For the third, $\mathbb{E}_k[\xi_k^2 \|s_k\|^2] \le \sigma^2_{D,k}\, \mathbb{E}_k \left\|s_k\right\|^2 \le \varsigma^2 \sigma^2_{D,k}$.

Step 3: the bias cross term, exactly. Since $\langle (I_d - \beta P_k)E_k, b_k s_k\rangle = b_k\big(\langle E_k, s_k\rangle - \beta \|s_k\|^2 \langle E_k, u_k\rangle\big)$, taking conditional expectations and using the definitions (29),

$$\mathbb{E}_k \langle a_k, b_k s_k\rangle = \langle E_k, \bar{b}_k\rangle - \beta \langle E_k, \bar{B}_k\rangle.$$

Therefore, with Lemma 5 and $\mathbb{E}_k \left\|b_k s_k\right\|^2 = \Sigma^2_{b,k}$,

$$\mathbb{E}_k \left\|a_k + \beta b_k s_k\right\|^2 \le \theta(\beta) \|E_k\|^2 + 2\beta \langle E_k, \bar{b}_k\rangle - 2\beta^2 \langle E_k, \bar{B}_k\rangle + \beta^2 \Sigma^2_{b,k}.$$

Applying $2\langle p, q\rangle \le \frac{\beta}{8}\|p\|^2 + \frac{8}{\beta}\|q\|^2$ with $p = E_k$ and $q = \beta \bar{b}_k$, and then with $q = \beta^2 \bar{B}_k$,

$$2\beta \langle E_k, \bar{b}_k\rangle \le \tfrac{\beta}{8}\|E_k\|^2 + 8\beta \|\bar{b}_k\|^2, \qquad 2\beta^2 \big|\langle E_k, \bar{B}_k\rangle\big| \le \tfrac{\beta}{8}\|E_k\|^2 + 8\beta^3 \|\bar{B}_k\|^2.$$

The asymmetry that drives the whole analysis is visible here: $\bar{b}_k$ enters with a factor β, which after division by the contraction rate $\beta/2$ leaves it at order 1, whereas $\Sigma^2_{b,k}$ enters with β^2 and is damped.

Step 4: drift and assembly. Since $\|x_{k+1} - x_k\| = \eta \|G_k\|$, (17) gives $\mathbb{E}_k \left\|w_k\right\|^2 \le L_v^2 \eta^2\, \mathbb{E}_k \left\|G_k\right\|^2$. Young's inequality $\|X + w\|^2 \le (1 + \alpha)\|X\|^2 + (1 + \alpha^{-1})\|w\|^2$ with $\alpha = \beta/4$, together with $1 + 4/\beta \le 5/\beta$ for $\beta \le 1$, gives

$$\mathbb{E}_k \left\|E_{k+1}\right\|^2 \le \Big(1 + \tfrac{\beta}{4}\Big)\Big[\big(\theta(\beta) + \tfrac{\beta}{4}\big)\|E_k\|^2 + \beta^2 \Sigma^2_{\mathrm{mz},k} + \beta^2 \Sigma^2_{b,k} + 8\beta \|\bar{b}_k\|^2 + 8\beta^3 \|\bar{B}_k\|^2\Big] + \frac{5 L_v^2 \eta^2}{\beta}\, \mathbb{E}_k \left\|G_k\right\|^2.$$

Because $\beta \le 1/c_d$ implies $c_d \beta^2 \le \beta$ and hence $\theta(\beta) \le 1 - \beta$, we have $\theta(\beta) + \frac{\beta}{4} \le 1 - \frac{3\beta}{4}$ and

$$\Big(1 + \frac{\beta}{4}\Big)\Big(1 - \frac{3\beta}{4}\Big) = 1 - \frac{\beta}{2} - \frac{3\beta^2}{16} \le 1 - \frac{\beta}{2}.$$

Bounding the remaining bracketed terms by $(1 + \beta/4) \le \frac{5}{4}$ gives (36). □

Lemma 16 (Direction error)**.** *Under the same hypotheses, with $\theta = \theta(\tilde{\beta})$,*

$$\mathbb{E}_k \left\|\Delta_k\right\|^2 \le 3(1 - \beta_K)^2 \|\zeta_k\|^2 + 3\theta \|E_k\|^2 + 3\big(\beta_K^2 + c_d \tilde{\beta}^2\big)\gamma^2 \sigma^2_{K,k} + 3\tilde{\beta}^2 \varsigma^2 \sigma^2_{D,k} + 3\tilde{\beta}^2 \Sigma^2_{b,k}. \quad (37)$$

Proof. By Lemma 1(iii), $\Delta_k = (1 - \beta_K)\zeta_k + X_k + Y_k$ with

$$X_k := (I_d - \tilde{\beta} P_k)E_k + \big(\beta_K I_d - \tilde{\beta} P_k\big)\delta_k + \tilde{\beta}\xi_k s_k, \qquad Y_k := \tilde{\beta} b_k s_k,$$

so $\|\Delta_k\|^2 \le 3\big((1 - \beta_K)^2 \|\zeta_k\|^2 + \|X_k\|^2 + \|Y_k\|^2\big)$, and ζ_k is $\mathcal{F}_k$-measurable by construction.

For X_k, all three cross terms vanish. The first and second are orthogonal because δ_k is mean-zero and independent of (u_k, s_k), so with $M_k := (I_d - \tilde{\beta}P_k)^\top(\beta_K I_d - \tilde{\beta}P_k)$ we get $\mathbb{E}_k[E_k^\top M_k \delta_k] = E_k^\top \mathbb{E}_k[M_k]\,\mathbb{E}_k[\delta_k] = 0$; the terms involving ξ_k vanish because $\mathbb{E}[\xi_k \mid \mathcal{F}_k, u_k, s_k, \delta_k] = 0$. Hence

$$\mathbb{E}_k \|X_k\|^2 \le \theta\|E_k\|^2 + \mathbb{E}_k \left\|(\beta_K I_d - \tilde{\beta}P_k)\delta_k\right\|^2 + \tilde{\beta}^2\varsigma^2\sigma_{D,k}^2.$$

For the middle term, condition on δ_k and expand using (18)–(19):

$$\mathbb{E}_k \left\|(\beta_K I_d - \tilde{\beta}P_k)\delta_k\right\|^2 = \left(\beta_K^2 - 2\beta_K\tilde{\beta} + c_d\tilde{\beta}^2\right)\mathbb{E}_k\|\delta_k\|^2 \;\le\; \left(\beta_K^2 + c_d\tilde{\beta}^2\right)\gamma^2\sigma_{K,k}^2, \tag{38}$$

the cross term being $-2\beta_K\tilde{\beta}\,\delta_k^\top\mathbb{E}_k[P_k]\delta_k = -2\beta_K\tilde{\beta}\|\delta_k\|^2 \le 0$. Finally $\mathbb{E}_k\|Y_k\|^2 = \tilde{\beta}^2\Sigma_{b,k}^2$. □

Remark 17 (The two damping mechanisms). The quantity $\psi := \beta_K^2 - 2\beta_K\tilde{\beta} + c_d\tilde{\beta}^2$ of (38) is positive for $c_d > 1$, since its discriminant $4\tilde{\beta}^2(1 - c_d)$ is negative, and is at most $\beta_K^2 + c_d\tilde{\beta}^2$. Both terms are under the user's control, which is the point: with a fresh known-component estimate one has $\beta_K = 1$ and $\psi \ge 1 - 2\tilde{\beta} + c_d\tilde{\beta}^2 = \theta$, an $\Theta(1)$ multiple of σ_K^2 that no choice of the remaining parameters can reduce.

F Proof of Theorem 2

Lemma 18 (Descent). *If $\eta \le \frac{1}{2L}$ then $\mathbb{E}_k[F(x_{k+1})] \le F(x_k) - \frac{\eta}{2}\|\nabla F_k\|^2 - \frac{\eta}{4}\mathbb{E}_k\|G_k\|^2 + \frac{\eta}{2}\mathbb{E}_k\|\Delta_k\|^2$.*

Proof. L-smoothness and $x_{k+1} = x_k - \eta G_k$ give $F(x_{k+1}) \le F(x_k) - \eta\langle\nabla F_k, G_k\rangle + \frac{L\eta^2}{2}\|G_k\|^2$. The polarization identity $-\langle p, q\rangle = -\frac{1}{2}\|p\|^2 - \frac{1}{2}\|q\|^2 + \frac{1}{2}\|q - p\|^2$ with $p = \nabla F_k$, $q = G_k$ gives $-\eta\langle\nabla F_k, G_k\rangle = -\frac{\eta}{2}\|\nabla F_k\|^2 - \frac{\eta}{2}\|G_k\|^2 + \frac{\eta}{2}\|\Delta_k\|^2$. Finally $-\frac{\eta}{2} + \frac{L\eta^2}{2} \le -\frac{\eta}{4}$ whenever $L\eta \le \frac{1}{2}$. □

Proof of Theorem 2. Define

$$\Phi_k := F(x_k) - F^\star + \omega\|E_k\|^2 + \omega_K\|\zeta_k\|^2, \qquad \omega := \frac{3\eta\theta}{\beta}, \qquad \omega_K := \frac{3\eta}{\beta_K}.$$

Combining Lemma 18 with (37),

$$\begin{aligned}\mathbb{E}_k[F(x_{k+1})] \le F(x_k) &- \tfrac{\eta}{2}\|\nabla F_k\|^2 - \tfrac{\eta}{4}\mathbb{E}_k\|G_k\|^2 + \tfrac{3\eta}{2}(1-\beta_K)^2\|\zeta_k\|^2 + \tfrac{3\eta\theta}{2}\|E_k\|^2 \\ &+ \tfrac{3\eta}{2}\left(\beta_K^2 + c_d\tilde{\beta}^2\right)\gamma^2\sigma_{K,k}^2 + \tfrac{3\eta\tilde{\beta}^2}{2}\left(\varsigma^2\sigma_{D,k}^2 + \Sigma_{b,k}^2\right).\end{aligned}$$

Add ω times (36) and ω_K times (35), then collect.

Coefficient of $\|E_k\|^2$. It is $\frac{3\eta\theta}{2} + \omega(1 - \frac{\beta}{2}) - \omega = \frac{3\eta\theta}{2} - \frac{\omega\beta}{2} = 0$.

Coefficient of $\|\zeta_k\|^2$. It is $\frac{3\eta}{2}(1-\beta_K)^2 + \omega_K(1 - \frac{\beta_K}{2}) - \omega_K = \frac{3\eta}{2}(1-\beta_K)^2 - \frac{\omega_K\beta_K}{2} \le 0$ for $\omega_K = 3\eta/\beta_K$, since $(1-\beta_K)^2 \le 1$.

Coefficient of $\mathbb{E}_k\|G_k\|^2$. It is

$$-\frac{\eta}{4} + \omega\cdot\frac{5L_v^2\eta^2}{\beta} + \omega_K\cdot\frac{3\gamma^2L_K^2\eta^2}{\beta_K} = -\frac{\eta}{4} + \frac{15\theta L_v^2\eta^3}{\beta^2} + \frac{9\gamma^2L_K^2\eta^3}{\beta_K^2},$$

which is nonpositive as soon as each of the two positive terms is at most $\eta/8$, that is $\eta \le \beta/(\sqrt{120\theta}L_v)$ and $\eta \le \beta_K/(\sqrt{72}\,\gamma L_K)$. Since $\sqrt{120} < 11$ and $\sqrt{72} < 9$, (10) suffices.

Constant terms. We have $\omega\cdot\frac{5}{4}\beta^2 = \frac{15}{4}\eta\theta\beta$, $\omega\cdot 10\beta = 30\eta\theta$, $\omega\cdot 10\beta^3 = 30\eta\theta\beta^2$ and $\omega_K\cdot\frac{3}{2}\beta_K^2 = \frac{9}{2}\eta\beta_K$. Using $\Sigma_{\mathrm{mz},k}^2 = c_d\gamma^2\sigma_{K,k}^2 + \varsigma^2\sigma_{D,k}^2$, the coefficient of $\gamma^2\sigma_{K,k}^2$ is $\frac{3\eta}{2}(\beta_K^2 + c_d\tilde{\beta}^2) + \frac{15}{4}\eta\theta\beta c_d + \frac{9}{2}\eta\beta_K$, that of $\varsigma^2\sigma_{D,k}^2$ and of $\Sigma_{b,k}^2$ is $\frac{3\eta\tilde{\beta}^2}{2} + \frac{15}{4}\eta\theta\beta$, and those of $\|\bar{b}_k\|^2$ and $\|\bar{B}_k\|^2$ are $30\eta\theta$ and $30\eta\theta\beta^2$.

Taking total expectations, telescoping and using $\Phi_T \geq 0$, $\frac{\eta}{2}\sum_{k<T}\mathbb{E}\,\|\nabla F_k\|^2 \leq \Phi_0 + \eta\sum_{k<T}\mathcal{Q}_k$ with $\mathcal{Q}_k$ the bracket above divided by η. Since $m_0 = 0$ we have $E_0 = -v(x_0)$, and the warm start $h_0 = \widehat{g}_{\mathrm{K}}(x_0)$ gives $\zeta_0 = \gamma(\widehat{g}_{\mathrm{K}}(x_0) - g_{\mathrm{K}}(x_0)) = \delta_0$, hence $\mathbb{E}\,\|\zeta_0\|^2 \leq \gamma^2\sigma_K^2$. Therefore $\Phi_0 \leq \Delta_F + \frac{3\eta\theta}{\beta}\|v(x_0)\|^2 + \frac{3\eta}{\beta_K}\gamma^2\sigma_K^2$, and dividing by $\eta T/2$ gives (11). □

Remark 19 (Why the known component is not used inside R_k). Replacing $\widehat{g}_{\mathrm{K}}(x_k)$ by h_{k+1} in line 5 of Algorithm 1 would remove δ_k from the innovation entirely. It would also change the tracked object from $v(x)$ to $\nabla F(x_k) - \gamma h_{k+1}$, whose increment is $(v_{k+1} - v_k) + \gamma\{[g_{\mathrm{K}}(x_{k+1}) - g_{\mathrm{K}}(x_k)] - [h_{k+2} - h_{k+1}]\}$. The two bracketed increments do not cancel, so the drift constant in (17) becomes $L_v + \gamma L_K$ instead of L_v. When g_{U} varies slowly and $\gamma \approx 1$ this is close to L, which would forfeit the substitution described in §G.4. Keeping the fresh estimate inside R_k costs instead a term $c_d\gamma^2\sigma_{K,k}^2$ in $\Sigma_{\mathrm{mz},k}^2$, damped by β; we prefer to pay in noise rather than in smoothness.

Remark 20 (Reading the bias terms). The three bias contributions to (11) sit at different orders in β: the second moment $\Sigma_{b,k}^2$ carries a factor β, the weighted mean $\|\bar{B}_k\|^2$ a factor β^2, and only the mean $\|\bar{b}_k\|^2$ appears at order 1. Since $\|\bar{b}_k\| \leq L\mu_k\chi_1$ while $\Sigma_{b,k} \leq \frac{1}{2}L\mu_k\varsigma_6$, and these differ by a factor of order d for the random families, the undamped part of the bias is much smaller than a second-moment treatment would suggest. Carrying $\Sigma_{b,k}^2$ undamped, as a cruder version of Lemma 15 does, degrades the dimension dependence of Corollary 24 from d^2 to d^4. The bound $\|\bar{b}_k\| \leq L\mu_k\chi_1$ is due to Berahas et al. (2022), following Maggiar et al. (2018), and is the ingredient with which Hikima & Takeda (2025a) reach d^2 for multi-direction methods; what is specific to the present setting is its use inside the recursion (36), which requires the exact cross-term computation of Step 3 and introduces $\bar{B}_k$.

G Complexity

Throughout this section we take $\lambda = 0$, so $\tilde{\beta} = \beta$ and $\theta = \theta(\beta) \leq 1$; see Appendix I for why a fixed $\lambda > 0$ is inadmissible in the limit $\varepsilon \to 0$. Two routes are available for the term $60\theta\beta^2\|\bar{B}_k\|^2$, and we state both because they lead to different admissible ranges for β.

Sharp route (Gaussian and sphere probes). Impose

$$\beta \;\leq\; \frac{1}{3d}. \tag{39}$$

Then Lemma 11 gives $60\beta^2\|\bar{B}_k\|^2 \leq 60\beta^2 \cdot 9d^2L^2\mu_k^2\chi_1^2 \leq 60L^2\mu_k^2\chi_1^2$, a quantity of the same size as the undamped bound $60\|\bar{b}_k\|^2 \leq 60L^2\mu_k^2\chi_1^2$ on the mean bias. The bias contribution is therefore at most $60\|\bar{b}_k\|^2 + 60L^2\mu_k^2\chi_1^2 \leq 120L^2\mu_k^2\chi_1^2$, and under a Q-Lipschitz Hessian at most $30Q^2\mu_k^4\chi_2^2$ by the same computation. This is the route used in the corollaries below, and it admits β up to $\Theta(1/d)$.

Generic route (any family). Impose $\beta\varsigma_4^2 \leq 1$ instead. Then Lemma 10 gives $60\beta^2\|\bar{B}_k\|^2 \leq 60\beta^2\varsigma_4^2\Sigma_{b,k}^2 \leq 60\beta\Sigma_{b,k}^2$, so $\bar{B}_k$ is absorbed into the $\Sigma_{b,k}^2$ term, whose coefficient becomes 71β, while the mean-bias term stays $60\|\bar{b}_k\|^2$. Every rate below holds unchanged along this route, but since $\varsigma_4^2 = \Theta(d^2)$ it only admits β up to $\Theta(1/d^2)$, which is what limits the deployment count of Appendix K to $\Theta(d^2\varepsilon^{-2})$ instead of $\Theta(d\,\varepsilon^{-2})$.

Along the sharp route, with $\tilde{\beta} = \beta$ and $\theta \leq 1$, the coefficients of Theorem 2 obey $3\beta_K^2 + 3c_d\beta^2 + \frac{15}{2}\beta c_d + 9\beta_K \leq 11\beta c_d + 12\beta_K$ and $3\beta^2 + \frac{15}{2}\beta \leq 11\beta$, so the noise part of (11) is at most

$$\big(11\beta c_d + 12\beta_K\big)\frac{\gamma^2\sigma_K^2}{n_K} + 11\beta\,\varsigma^2\sigma_D^2 + 11\beta\,\Sigma_b^2 + 60\|\bar{b}\|^2 + 60L^2\mu^2\chi_1^2, \tag{40}$$

the last two terms being replaced by $30Q^2\mu^4\chi_2^2$ under a Q-Lipschitz Hessian. Only the last term is undamped: every appearance of $\gamma^2\sigma_K^2$ carries a factor βc_d or β_K, which is what permits one known-component sample per iteration. Each regime below optimizes μ in (40).

G.1 Regime A: shared randomness

Suppose both function queries use the same sample ζ and each realization $f(\cdot;\zeta)$ is ℓ-smooth. Writing $D_k = \langle\nabla F_k, u_k\rangle + \langle\nabla f(x_k;\zeta) - \nabla F_k, u_k\rangle + \varpi_k(\zeta)$ with $|\varpi_k(\zeta)| \leq \frac{\ell\mu_k}{2}\|u_k\|^2$, the decomposition of (3) holds with $b_k = \mathbb{E}_\zeta[\varpi_k]$ and $\xi_k = \langle\nabla f - \nabla F_k, u_k\rangle + \varpi_k - \mathbb{E}_\zeta[\varpi_k]$.

Remark 21 (Bookkeeping in this regime). Here ξ_k is linear in u_k and therefore not independent of it, so the bound $\sigma_{D,k}^2 \le \sigma_g^2$ is unavailable and the product $\varsigma^2\sigma_{D,k}^2$ must be replaced by a direct computation. With $\Sigma := \mathrm{Cov}(\nabla f(x_k;\zeta))$ and $\sigma_g^2 := \operatorname{tr}\Sigma$, Gaussian probes give

$$\mathbb{E}\left[\langle \nabla f - \nabla F_k, u\rangle^2 \|u\|^2\right] = \operatorname{tr}\big(\Sigma\, \mathbb{E}[uu^\top \|u\|^2]\big) = (d+2)\operatorname{tr}\Sigma = c_d\sigma_g^2,$$

so every occurrence of $\varsigma^2\sigma_{D,k}^2$ in (11) becomes $c_d\sigma_g^2$. Since $c_d = d+2$ and $\varsigma^2 = d$, this leaves all orders unchanged.

The noise ξ_k does depend on μ_k through the centred remainder $\varpi_k - \mathbb{E}_\zeta\,\varpi_k$, so a finite probe radius must be chosen. Since $|\varpi_k| \le \frac{\ell\mu_k}{2}\|u_k\|^2$, the remainder contributes to $\mathbb{E}_k[\xi_k^2\|s_k\|^2]$ at most its own second moment $\frac{\ell^2\mu_k^2}{4}\,\mathbb{E}_k[\|u_k\|^4\|s_k\|^2] \le \frac{1}{4}\ell^2\mu_k^2\varsigma_6^2$ plus a cross term with the gradient noise bounded by Cauchy–Schwarz by $2\sqrt{c_d\sigma_g^2\cdot\frac{1}{4}\ell^2\mu_k^2\varsigma_6^2} = \ell\mu_k\varsigma_6\sqrt{c_d}\,\sigma_g$. Choosing

$$\mu_k \;\le\; \mu_A := \min\Big\{\frac{\sqrt{c_d}\,\sigma_g}{\ell\,\varsigma_6},\; \frac{\varepsilon}{30\,L\chi_1}\Big\} \tag{41}$$

makes the remainder terms at most $2c_d\sigma_g^2$, so that $\mathbb{E}_k[\xi_k^2\|s_k\|^2] \le 3c_d\sigma_g^2$ replaces $c_d\sigma_g^2$ in Remark 21, and makes the bias terms $60\|\bar b_k\|^2 + 60L^2\mu_k^2\chi_1^2 \le 120L^2\mu_A^2\chi_1^2 \le \frac{120}{900}\varepsilon^2$ and $11\beta\Sigma_{b,k}^2 \le \frac{11}{4}\beta L^2\mu_A^2\varsigma_6^2 \le \frac{11}{3600}\varepsilon^2$ under $\beta\varsigma_6^2 \le \chi_1^2$, together below $\varepsilon^2/7$. The rate below is stated with these choices; the only effect is a constant factor of 3 on σ_g^2 and a seven-way rather than six-way accuracy split.

Corollary 22 (Regime A)**.** *In Regime A, with one known-component sample per iteration, warm start $h_0 = \widehat{g}_{\mathrm{K}}(x_0)$,*

$$\beta = \min\Big\{\frac{1}{c_d},\; \frac{\chi_1^2}{\varsigma_6^2},\; \frac{n_K\varepsilon^2}{77\,c_d(\gamma^2\sigma_K^2 + 3n_K\sigma_g^2)}\Big\},\quad \mu_k = \mu_A,\quad \beta_K = \min\Big\{1,\; \frac{n_K\varepsilon^2}{84\,\gamma^2\sigma_K^2}\Big\},$$

$$\eta = \min\Big\{\tfrac{1}{2L}, \tfrac{\beta}{11\sqrt{\theta}L_v}, \tfrac{\beta_K}{9\gamma L_K}\Big\},$$

one has $\frac{1}{T}\sum_{k<T}\mathbb{E}\,\|\nabla F_k\|^2 \le \varepsilon^2$ after

$$T = \mathcal{O}\Big(\frac{\Delta_v\, d\,(\gamma^2\sigma_K^2 + \sigma_g^2)}{\varepsilon^4} + \frac{\gamma^3 L_K\sigma_K^2\Delta_F}{\varepsilon^4} + \frac{L\Delta_F + \gamma^2\sigma_K^2}{\varepsilon^2}\Big)$$

iterations, where $\Delta_v := L_v\Delta_F + \|v(x_0)\|^2$.

Proof. With $\mu \to 0$ the floor (40) reduces to $(11\beta c_d + 12\beta_K)\gamma^2\sigma_K^2/n_K + 33\beta c_d\sigma_g^2$ plus the bias terms controlled by (41), using Remark 21 and the remainder bound above to replace $\varsigma^2\sigma_D^2$ by $3c_d\sigma_g^2$. The stated caps make each of $11\beta c_d\gamma^2\sigma_K^2/n_K$, $12\beta_K\gamma^2\sigma_K^2/n_K$ and $33\beta c_d\sigma_g^2$ at most $\varepsilon^2/7$, and the choice (41) makes the bias terms at most $\varepsilon^2/7$ together; these four pieces account for $4\varepsilon^2/7$. The remaining three terms of (11) are at most $\varepsilon^2/7$ each once $T \ge 14\Delta_F/(\eta\varepsilon^2)$, $T \ge 42\theta\|v_0\|^2/(\beta\varepsilon^2)$ and $T \ge 42\gamma^2\sigma_K^2/(n_K\beta_K\varepsilon^2)$, so the seven pieces sum to ε^2. Substituting $\eta^{-1} \le 2L + 11L_v/\beta + 9\gamma L_K/\beta_K$ from (10), together with $\beta^{-1} = \mathcal{O}\big(c_d(\gamma^2\sigma_K^2 + \sigma_g^2)/\varepsilon^2\big)$ and $\beta_K^{-1} = \mathcal{O}(\gamma^2\sigma_K^2/\varepsilon^2)$, gives the three displayed terms; the lower bound on T arising from the initialization of h contributes $\mathcal{O}(\gamma^4\sigma_K^4/(n_K^2\varepsilon^4))$ under the single-sample warm start, and $\mathcal{O}(\gamma^2\sigma_K^2/(n_K\varepsilon^2))$ under the batched warm start of Appendix H. □

Regime A: shared randomness. If both queries evaluate the same realization $f(\cdot;\zeta)$ and each realization is smooth, the difference quotient concentrates around a directional derivative of $f(\cdot;\zeta)$, so $\sigma_{D,k}^2 \le \sigma_g^2$ independently of μ_k. One may then send $\mu_k \to 0$, and $\Sigma_{b,k}$, $\bar b_k$ and $\bar B_k$ all vanish. This applies whenever the randomness in F is exogenous to the decision.

Corollary 23 (Regime A, restated)**.** *For any $n_K \ge 1$, with warm start $h_0 = \widehat{g}_{\mathrm{K}}(x_0)$,*

$$\beta = \min\Big\{\frac{1}{c_d},\; \frac{\chi_1^2}{\varsigma_6^2},\; \frac{n_K\varepsilon^2}{66\,c_d(\gamma^2\sigma_K^2 + 3n_K\sigma_g^2)}\Big\},\qquad \mu_k = \mu_A,$$

$$\beta_K = \min\Big\{1,\; \frac{n_K\varepsilon^2}{72\,\gamma^2\sigma_K^2}\Big\},\qquad \eta = \min\Big\{\frac{1}{2L},\; \frac{\beta}{11\sqrt{\theta}L_v},\; \frac{\beta_K}{9\gamma L_K}\Big\}.$$

one has $\frac{1}{T}\sum_{k<T}\mathbb{E}\,\|\nabla F_k\|^2 \le \varepsilon^2$ *after*

$$T = \mathcal{O}\Big(\underbrace{\frac{\Delta_v\, d\,(\gamma^2\sigma_K^2+\sigma_g^2)}{\varepsilon^4}}_{\text{residual tracking}} + \underbrace{\frac{\gamma^3 L_K \sigma_K^2 \Delta_F}{n_K\,\varepsilon^4}}_{\text{known-component tracking}} + \underbrace{\frac{\gamma^4\sigma_K^4}{n_K^2\,\varepsilon^4}}_{\text{initialization of } h} + \underbrace{\frac{L\Delta_F}{\varepsilon^2}}_{\text{lower order}} \Big) \tag{42}$$

iterations, using $2n_F$ *function evaluations and* n_K *known-component samples per iteration, plus the one-time cost of forming* h_0*. The last term comes from that initialization and can be reduced to* $\mathcal{O}(\gamma^2\sigma_K^2\varepsilon^{-2})$ *by a batched warm start (Appendix H).*

The first term matches the $\mathcal{O}(d\sigma^2\varepsilon^{-4})$ benchmark for stochastic zeroth-order optimization (Ghadimi & Lan, 2013), with L replaced by L_v. The middle term is the complexity of momentum SGD with a gradient oracle of variance $\gamma^2\sigma_K^2$ and smoothness L_K. It is the price of a noisy known component, is incurred by any method that uses $\widehat{g}_{\mathrm{K}}$, and vanishes when the known component is exact.

G.2 Regime B: independent randomization

When the two queries deploy different decisions they draw from different distributions and cannot share randomness. With n_F samples per side and μ_k the probe radius,

$$\sigma_{D,k}^2 \le \frac{\sigma_F^2}{2\mu_k^2\, n_F}, \tag{43}$$

where σ_F^2 bounds the conditional variance of the objective under the deployed distribution. Indeed, the two queries are independent, so $\mathrm{Var}(\widehat{F}(x_k+\mu_k u_k)-\widehat{F}(x_k-\mu_k u_k)) \le 2\sigma_F^2/n_F$, and dividing by $(2\mu_k)^2$ gives (43).

Corollary 24 (Regime B, L-smoothness only)**.** *Let* $\lambda = 0$*, let the probe family satisfy the smoothing identity with constants* χ_1, ς *(Gaussian or sphere), and write* $C_1 := 102000\, L^2\sigma_F^2\varsigma^2\chi_1^2$*. Assume* $\varepsilon^2 \le 7\gamma^2\sigma_K^2/(dn_K)$*, let* $n_K \ge 1$ *and* $1 \le n_F \le n_F^{\max} := C_1 n_K/(66 c_d\gamma^2\sigma_K^2\varepsilon^2)$*, initialize* $h_0 = \widehat{g}_{\mathrm{K}}(x_0)$ *and* $m_0 = 0$*, and choose*

$$\beta = \frac{n_F\varepsilon^4}{C_1}, \qquad \beta_K = \min\Big\{1, \frac{n_K\varepsilon^2}{72\gamma^2\sigma_K^2}\Big\}, \qquad \mu_\star^2 = \frac{\sigma_F\varsigma}{L\chi_1}\sqrt{\frac{11\beta}{246\, n_F}}, \qquad \eta = \min\Big\{\frac{1}{2L}, \frac{\beta}{11\sqrt{\theta}L_v}, \frac{\beta_K}{9\gamma L_K}\Big\}.$$

Then $\beta\varsigma_6^2 \le \chi_1^2$ *and* $\beta \le 1/(3d)$ *hold, and the total number of function evaluations to reach* $\frac{1}{T}\sum_{k<T}\mathbb{E}\,\|\nabla F_k\|^2 \le \varepsilon^2$ *is*

$$\mathsf{N} = 2n_F T = \mathcal{O}\left(\frac{\big(L_v\Delta_F + \|v_0\|^2\big)\, L^2\sigma_F^2\,\varsigma^2\chi_1^2}{\varepsilon^6} + \frac{n_F}{n_K}\cdot\frac{\gamma^3 L_K\sigma_K^2\Delta_F}{\varepsilon^4} + \frac{n_F}{n_K^2}\cdot\frac{\gamma^4\sigma_K^4}{\varepsilon^4} + \frac{n_F L\Delta_F}{\varepsilon^2}\right), \tag{44}$$

with n_K *known-component samples per iteration. The first term is independent of* n_F *and* n_K *and carries no* σ_K*. At* $n_F = 1$ *the other three are lower order in* ε*. At general* $n_F \le n_F^{\max}$ *their ratios to the first are at most* $\gamma L_K/(c_d L_v)$*,* $\gamma^2\sigma_K^2/(c_d n_K\Delta_v)$ *and* $Ln_K\varepsilon^2/(c_d\gamma^2\sigma_K^2 L_v)$ *respectively, so the first two are lower order in* d *under* $L_v \ge \gamma L_K/d$ *and* $\Delta_v \ge \gamma^2\sigma_K^2/(dn_K)$*, and the third is lower order in* ε*.*

Proof. Substituting (43), Lemma 8(a) and (31) into (40), the μ-dependent part is

$$\Psi(\mu) = \underbrace{\frac{11\beta\varsigma^2\sigma_F^2}{2\mu^2 n_F}}_{11\beta\varsigma^2\sigma_D^2} + L^2\mu^2\Big[\underbrace{120\chi_1^2}_{60\|\bar{b}\|^2+60L^2\mu^2\chi_1^2} + \underbrace{\tfrac{11}{4}\beta\varsigma_6^2}_{11\beta\Sigma_b^2}\Big] \;\le\; \frac{A}{\mu^2} + B\mu^2,$$

with $A = \frac{11}{2}\beta\varsigma^2\sigma_F^2/n_F$ and $B = 123L^2\chi_1^2$, where the last inequality uses $\beta\varsigma_6^2 \le \chi_1^2$ and $120 + \frac{11}{4} \le 123$. Minimizing $A/\mu^2 + B\mu^2$ gives $\mu_\star^2 = \sqrt{A/B}$, which is the stated value, and $\Psi(\mu_\star) = 2\sqrt{AB} = 2\sqrt{\frac{11}{2}\cdot 123}\; L\sigma_F\varsigma\chi_1\sqrt{\beta/n_F} \le 53\, L\sigma_F\varsigma\chi_1\sqrt{\beta/n_F}$, since $2\sqrt{676.5} < 53$. Requiring $\Psi(\mu_\star) \le \varepsilon^2/6$ gives $\beta/n_F \le \varepsilon^4/(102000\, L^2\sigma_F^2\varsigma^2\chi_1^2)$, which the stated choice satisfies.

It remains to check that the known-oracle constraint does not degrade the leading term. Besides the probe-radius constraint just derived, β must satisfy $11\beta c_d\gamma^2\sigma_K^2 \le \varepsilon^2/6$ from (40), that is $\beta \le$

$n_K\varepsilon^2/(66c_d\gamma^2\sigma_K^2)$. The probe-radius constraint is the binding one precisely when $n_F\varepsilon^4/C_1 \le n_K\varepsilon^2/(66c_d\gamma^2\sigma_K^2)$, that is when $n_F \le n_F^{\max}$ as in the range $n_F \le n_F^{\max}$ of Corollary 3, with $n_F^{\max}$ scaled by n_K; the known-oracle variance then does not enter the leading term. Substituting $\beta = n_F\varepsilon^4/C_1$ into $T = \Theta(\Delta_v/(\beta\varepsilon^2))$ gives $T = \Theta(C_1\Delta_v/(n_F\varepsilon^6))$ and hence $\mathsf{N} = 2n_FT = \Theta(C_1\Delta_v/\varepsilon^6)$, independent of n_F, which is (44). For $n_F > n_F^{\max}$ the second constraint binds, β is capped at $n_K\varepsilon^2/(66c_d\gamma^2\sigma_K^2)$ and $\mathsf{N} = 2n_FT$ grows linearly in n_F.

Two further requirements must be checked. First, $\beta = n_F\varepsilon^4/C_1$ must also respect the moment condition $\beta\varsigma_6^2 \le \chi_1^2$ used in the radius balance, which for Gaussian probes reads $\beta \le 1/(9d^2)$ and implies the sharp-route cap (39). Since $n_F \le n_F^{\max}$ this holds as soon as $n_F^{\max}\varepsilon^4/C_1 \le 1/(9d^2)$, that is $\varepsilon^2 \le 66c_d\gamma^2\sigma_K^2/(9d^2n_K)$, which the hypothesis $\varepsilon^2 \le 7\gamma^2\sigma_K^2/(dn_K)$ of Corollary 3 guarantees because $c_d \ge d$. Second, the remaining lower bounds on T contribute to $\mathsf{N} = 2n_FT$ terms proportional to n_F: the L_K branch of (10) gives $\mathcal{O}(n_F\gamma^3L_K\sigma_K^2\Delta_F/(n_K\varepsilon^4))$, the L branch $\mathcal{O}(n_FL\Delta_F\varepsilon^{-2})$, and the ζ_0 initialization $\mathcal{O}(n_F\gamma^4\sigma_K^4/(n_K^2\varepsilon^4))$. At $n_F = 1$ these are the lower-order terms stated in (44); at $n_F = n_F^{\max} = \Theta(n_K\varepsilon^{-2})$ the first becomes $\mathcal{O}(C_1\gamma L_K\Delta_F/(c_d\varepsilon^6))$, a factor $\gamma L_K/(c_dL_v)$ times the leading term and hence lower order in d whenever $L_v \ge \gamma L_K/c_d$; the other two are smaller still. □

Corollary 25 (Regime B, Lipschitz Hessian)**.** *Under the hypotheses of Corollary 24 with ∇^2F Q-Lipschitz and the smoothing identity constant χ_2, write $C_2 := 1220\,\varsigma^2\sigma_F^2Q\chi_2$ and choose instead*

$$\beta = \frac{n_F\varepsilon^3}{C_2}, \qquad 1 \le n_F \le \frac{C_2n_K}{66c_d\gamma^2\sigma_K^2\varepsilon}, \qquad \varepsilon^2 \le \frac{\gamma^2\sigma_K^2}{d\,n_K}, \qquad \mu_\star^6 = \frac{11\beta\varsigma^2\sigma_F^2}{124\,n_FQ^2\chi_2^2},$$

with β_K and η as in Corollary 24. Then $\beta\varsigma_8^2 \le \chi_2^2$ holds and

$$\mathsf{N} = \mathcal{O}\big((L_v\Delta_F + \|v_0\|^2)\,\sigma_F^2\varsigma^2Q\chi_2\,\varepsilon^{-5}\big)$$

plus the same lower-order terms as in (44)*.*

Proof. Now Lemma 8(b) and the second bound of (31) give

$$\Psi(\mu) = \frac{A}{\mu^2} + Q^2\mu^4\Big[30\chi_2^2 + \tfrac{11}{36}\beta\varsigma_8^2\Big] \le \frac{A}{\mu^2} + B'\mu^4, \qquad B' = 31Q^2\chi_2^2,$$

using $\beta\varsigma_8^2 \le \chi_2^2$ and $30 + \frac{11}{36} \le 31$, and where the bias contribution is $30Q^2\mu^4\chi_2^2$ as computed after (39). Setting $\Psi' = -2A/\mu^3 + 4B'\mu^3 = 0$ gives $\mu_\star^6 = A/(2B')$, the stated value, and $\Psi(\mu_\star) = A^{2/3}B'^{1/3}\big(2^{1/3}+2^{-2/3}\big) \le 19\big(\beta\varsigma^2\sigma_F^2/n_F\big)^{2/3}(Q\chi_2)^{2/3}$, since $(2^{1/3}+2^{-2/3})\cdot(11/2)^{2/3}\cdot 31^{1/3} < 19$. Requiring this to be at most $\varepsilon^2/6$ gives $\beta\varsigma^2\sigma_F^2Q\chi_2/n_F \le \varepsilon^3/114^{3/2}$ and $114^{3/2} < 1220$, which is the stated $\beta = n_F\varepsilon^3/C_2$. The constraint $\beta \le n_K\varepsilon^2/(66c_d\gamma^2\sigma_K^2)$ gives the range of n_F, and $\beta\varsigma_8^2 \le \chi_2^2$ at n_F maximal reduces to the stated condition on ε since $\varsigma_8^2/\chi_2^2 = 54d^2$ for Gaussian probes and $c_d \ge d$. With $T = \Theta(\Delta_v/(\beta\varepsilon^2))$ and $\mathsf{N} = 2n_FT$ the rate follows. □

Remark 26 (Instantiation for the random families). Reading $\varsigma^2\chi_1^2$ and $\varsigma^2\chi_2$ off Table 2 gives d^2 and d^2 for the Gaussian and sphere families in Corollaries 24 and 25. These coincide with the corresponding rates of Hikima & Takeda (2025a) for plain multi-direction methods.

Corollary 27 (Regime B, coordinate probes)**.** *Let $u_k = s_k = \sqrt{d}\,e_{i_k}$ with i_k uniform, so $c_d = \varsigma^2 = d$, $\varsigma_4^2 = d^2$, $\varsigma_6^2 = d^3$ and $\varsigma_8^2 = d^4$, and use the direct mean-bias bounds $\|\bar b_k\| \le L\mu_k\chi_1$ with $\chi_1 = d/2$, respectively $\frac12Q\mu_k^2\chi_2$ with $\chi_2 = d^{3/2}/3$, of Lemma 6(c), together with the generic route $\beta\varsigma_4^2 \le 1$ for $\bar B_k$, so that the bias part of the floor is $60\|\bar b_k\|^2 + 71\beta\Sigma_{b,k}^2$. Under the remaining hypotheses of Corollary 24, with $\beta \le \min\{\chi_1^2/\varsigma_6^2,\, 1/\varsigma_4^2\} = 1/d^2$ (the second condition being the generic-route requirement for $\bar B_k$ and the binding one for $d \ge 4$), and with the constants $60 \to 60$, 71β in place of 11β in the radius balance,*

$$\mathsf{N} = \mathcal{O}\Big(\frac{\Delta_v\,L^2\sigma_F^2\,\varsigma^2\chi_1^2}{\varepsilon^6}\Big) = \mathcal{O}\Big(\frac{\Delta_v\,L^2\sigma_F^2\,d^3}{\varepsilon^6}\Big), \qquad \textit{and} \qquad \mathcal{O}\Big(\frac{\Delta_v\,\sigma_F^2\,d^{5/2}Q}{\varepsilon^5}\Big)$$

under a Q-Lipschitz Hessian, with the same lower-order terms as Corollary 24.

Proof. The proof of Corollary 24 used the smoothing identity only through the bound $\|\bar b_k\| \le L\mu_k\chi_1$ and the sharp route for $\bar B_k$. Replacing the former by the direct bound of Lemma 6(c) and

the latter by the generic route changes only the bracket in the radius balance, from $120\chi_1^2 + \frac{11}{4}\beta\varsigma_6^2$ to $60\chi_1^2 + \frac{71}{4}\beta\varsigma_6^2 \le 78\chi_1^2$ under $\beta\varsigma_6^2 \le \chi_1^2$, and hence only the numerical constant in C_1. The exponents follow by substituting $\varsigma^2\chi_1^2 = d \cdot d^2/4$ and $\varsigma^2\chi_2 = d \cdot d^{3/2}/3$. These are the coordinate-direction rates of Hikima & Takeda (2025a), and the extra factor $\sqrt{d}$ relative to the random families enters solely through χ_1, that is through the failure of the smoothing identity. □

Remark 28 (Probe radius schedule). Corollaries 24 and 25 give $\mu_\star \propto n_F^{-1/4}$ and $\mu_\star \propto n_F^{-1/6}$ respectively. With a batch growing linearly in k this is a power-law decay in k, whereas a geometric schedule shrinks the radius exponentially and inflates the variance term $\beta\varsigma^2\sigma_F^2/\mu_k^2$ accordingly.

G.3 Regime C: deterministic oracle with tolerance

Suppose F carries no stochastic noise but $\widehat{F}$ satisfies $|\widehat{F}(x) - F(x)| \le \delta_{\mathrm{tol}}$ deterministically, as when the objective is evaluated by a solver run to a fixed tolerance. The error is not mean-zero and does not shrink with resampling, so it must be carried in b_k: this contributes at most $\delta_{\mathrm{tol}}/\mu_k$ to $|b_k|$, hence at most $\delta_{\mathrm{tol}}\varsigma/\mu_k$ to $\|\bar{b}_k\|$ by Jensen, with a sign that cannot be assumed to cancel. Adding the finite-difference bias,

$$\|\bar{b}_k\| \le \frac{\delta_{\mathrm{tol}}\varsigma}{\mu_k} + \begin{cases} L\mu_k\chi_1, \\ \frac{1}{2}Q\mu_k^2\chi_2. \end{cases} \tag{45}$$

Letting $\mu_k \to 0$ is now inadmissible.

Corollary 29 (Regime C)**.** *With $\sigma_D = 0$, $\theta \le 1$, and the generic route $\beta\varsigma_4^2 \le 1$ (so that the bias part of the floor is $60\|\bar{b}\|^2 + 71\beta\Sigma_b^2$): (a) under L-smoothness, $\mu_\star = \sqrt{\delta_{\mathrm{tol}}\varsigma/(L\chi_1)}$ gives an attainable accuracy floor $\varepsilon^2 \gtrsim 240\,\delta_{\mathrm{tol}}L\varsigma\chi_1$; (b) with a Q-Lipschitz Hessian, $\mu_\star = \big(\delta_{\mathrm{tol}}\varsigma/(Q\chi_2)\big)^{1/3}$ gives $\varepsilon^2 \gtrsim 135\,(\delta_{\mathrm{tol}}\varsigma)^{4/3}(Q\chi_2)^{2/3}$. Part (b) is the sharper floor whenever $\delta_{\mathrm{tol}} \le 5.6\,L^3\chi_1^3/(\varsigma Q^2\chi_2^2)$.*

Proof. (a) Minimizing $h(\mu) = \delta_{\mathrm{tol}}\varsigma/\mu + L\mu\chi_1$ gives $\mu_\star = \sqrt{\delta_{\mathrm{tol}}\varsigma/(L\chi_1)}$ and $h(\mu_\star) = 2\sqrt{\delta_{\mathrm{tol}}L\varsigma\chi_1}$, so $60\|\bar{b}\|^2 \le 240\,\delta_{\mathrm{tol}}L\varsigma\chi_1$. (b) Minimizing $h(\mu) = \delta_{\mathrm{tol}}\varsigma/\mu + \frac{1}{2}Q\mu^2\chi_2$: setting $h' = -\delta_{\mathrm{tol}}\varsigma/\mu^2 + Q\mu\chi_2 = 0$ gives $\mu_\star^3 = \delta_{\mathrm{tol}}\varsigma/(Q\chi_2)$, at which the two summands equal $(\delta_{\mathrm{tol}}\varsigma)^{2/3}(Q\chi_2)^{1/3}$ and half that, so $h(\mu_\star) = \frac{3}{2}(\delta_{\mathrm{tol}}\varsigma)^{2/3}(Q\chi_2)^{1/3}$ and $60\|\bar{b}\|^2 \le 60 \cdot \frac{9}{4}(\delta_{\mathrm{tol}}\varsigma)^{4/3}(Q\chi_2)^{2/3} = 135(\delta_{\mathrm{tol}}\varsigma)^{4/3}(Q\chi_2)^{2/3}$. Comparing the two floors and cubing gives the threshold. In both cases the term $71\beta\Sigma_b^2$ is of lower order at $\mu_\star$: since $\Sigma_b^2 \le 2(\delta_{\mathrm{tol}}\varsigma/\mu)^2 + 2(\text{finite-difference bias})^2$ with $\mathbb{E}\,\|s\|^2 \le \varsigma^2$, its leading part at $\mu_\star$ is at most 142β times the undamped term $60(\delta_{\mathrm{tol}}\varsigma/\mu_\star)^2$, and is absorbed into the stated floors up to a constant. □

With a deterministic oracle there is nothing to average, so the method converges at the optimization rate $T = \mathcal{O}\big((L_v\Delta_F + \|v_0\|^2)/(\beta\varepsilon^2)\big)$ until ε reaches the floor of Corollary 29, and then stalls.

Remark 30 (Cost of omitting the known-component average). Setting $\rho_K = 0$, hence $\beta_K = 1$, makes the coefficient of $\gamma^2\sigma_K^2/n_K$ in (40) equal to $11\beta c_d + 12$, which is $\Theta(1)$. Driving the floor below ε^2 then requires $n_K = \Omega(\gamma^2\sigma_K^2/\varepsilon^2)$ known-component samples per iteration. In Regime B with $n_F = 1$ the iteration count is $T = \Theta(\varepsilon^{-6})$ in both cases, so the total number of known-component samples is $n_K T = \Theta(\varepsilon^{-8})$ without the average and $T = \Theta(\varepsilon^{-6})$ with it, at $n_K = 1$. The known-component momentum therefore buys a factor ε^{-2} in that resource, by moving the averaging from within an iteration to across iterations.

G.4 Where L, L_v and L_K enter

The step-size condition (10) is a minimum of three branches, so $\eta^{-1} \le 2L + 11\sqrt{\theta}L_v/\beta + 9\gamma L_K/\beta_K$ and the optimization term of (11) splits accordingly into a term in L, one in L_v/β and one in L_K/β_K. Since β and β_K are both driven to $\mathcal{O}(\varepsilon^2)$ or smaller by the noise terms, the last two dominate and L survives only in a lower-order $\mathcal{O}(L\Delta_F\varepsilon^{-2})$ term. The two smoothness constants that do appear at leading order play different roles.

L_v multiplies Δ_F in the zeroth-order part of the complexity. This is the substitution the decomposition buys, and it is what a black-box method pays L for. Two further appearances of the smoothness constants are not affected by it: the factor L^2 in (14) comes from the finite-difference bias of a

two-point estimate and is common to any method using one, and Q in (15) plays the same role one derivative up. Alongside $L \to L_v$, the initialization term is $\|v(x_0)\|^2$ rather than $\|\nabla F(x_0)\|^2$.

L_K multiplies Δ_F in the first-order part. The term $\gamma^3 L_K \sigma_K^2 \Delta_F \varepsilon^{-4}$ is the complexity of momentum SGD on a gradient oracle of variance $\gamma^2\sigma_K^2$ and smoothness L_K. It is lower order in Regime B, is incurred by any method that consumes $\widehat{g}_{\mathrm{K}}$, and disappears when the known component is exact, as it is in Stackelberg problems where the leader's partial gradient is available in closed form.

Remark 31 (When the substitution is a gain). In Regime B the leading term is $\mathcal{O}(\Delta_v L^2 \sigma_F^2 \varsigma^2 \chi_1^2 \varepsilon^{-6})$ and carries no σ_K, so the gain over a black-box method is the substitution $L \to L_v$ outright. In Regime A the comparison is sharper. Setting $\gamma = 0$ makes $v = \nabla F$, $\delta_k \equiv 0$ and $L_v = L$, recovering a black-box method with leading term $L\Delta_F\, d\, \sigma_g^2 \varepsilon^{-4}$; the corresponding term for PISO is $L_v \Delta_F\, d\, (\gamma^2\sigma_K^2 + \sigma_g^2)\varepsilon^{-4}$, because the residual observation subtracts $\langle \gamma \widehat{g}_{\mathrm{K}}, u_k \rangle$ and inherits its noise. The method therefore improves on the black-box baseline exactly when

$$\frac{L_v}{L} \;<\; \frac{\sigma_g^2}{\sigma_g^2 + \gamma^2 \sigma_K^2}. \tag{46}$$

The residual must be smoother than F by at least the factor by which the known oracle inflates the observation noise. Both sides are estimable, and (46) holds with room to spare whenever the known component is accurate relative to the function oracle, which is the case in all of our settings; it degenerates to an equality at $\gamma = 0$, as it should.

Since $v = \nabla F - \gamma g_{\mathrm{K}}$ is $(L + \gamma L_K)$-Lipschitz, L_v is always finite and the rates have a residual-tracking leading term that never exceeds the corresponding black-box term by more than the factor $(L + \gamma L_K)/L$, a constant when L_K/L and γ are bounded; the known-oracle, known-component tracking and initialization terms are additional and are not covered by this comparison. The gain is realized when the opaque component varies slowly along the trajectory: in performative prediction L_v scales with the performative sensitivity, and for a fixed displacement the per-step squared drift at $\gamma_k^\star$ is reduced by the factor $1 - \varphi_k^2$ of Proposition 4; no corresponding ratio for the uniform constant L_v is claimed.

Remark 32 (Probes per update). The requirement $N = \Omega(d)$ in Hikima & Takeda (2025a) is not an artefact of tuning: their bound contains a term proportional to $\frac{d}{N}\|\nabla F(x)\|^2$ that must be absorbed into the descent step, which with their constants forces $N \geq 81d$. At $d = 30$ the resulting query count exceeds their own 5000-sample budget, so their guarantee is vacuous at the budget their experiments run. PISO uses one probe per iteration at any operating point.

H INITIALIZING THE KNOWN-COMPONENT STATE

The state h must be initialized, and the choice enters Theorem 2 through the term $6\gamma^2\sigma_K^2/(n_K\beta_K T)$, which requires

$$T \;\geq\; \frac{36\;\mathbb{E}\,\|\zeta_0\|^2}{\beta_K\,\varepsilon^2} \qquad \text{where} \qquad \zeta_0 = \gamma\big(h_0 - g_{\mathrm{K}}(x_0)\big). \tag{47}$$

Three strategies are available; recall $\beta_K = n_K\varepsilon^2/(72\gamma^2\sigma_K^2)$ in Corollaries 22 and 24.

Zero initialization, $h_0 = 0$. Then $\zeta_0 = -\gamma g_{\mathrm{K}}(x_0)$ and $\mathbb{E}\,\|\zeta_0\|^2 = \gamma^2\|g_{\mathrm{K}}(x_0)\|^2$, with no variance reduction at all. Substituting into (47) gives $T \geq \Theta\big(\gamma^4\sigma_K^2\|g_{\mathrm{K}}(x_0)\|^2/(n_K\varepsilon^4)\big)$. This introduces a dependence on $\|g_{\mathrm{K}}(x_0)\|$, the magnitude of the known gradient at the starting point, which no other term in the analysis carries; we therefore do not use it.

Single-sample warm start, $h_0 = \widehat{g}_{\mathrm{K}}(x_0)$. Then $\zeta_0 = \delta_0$ and $\mathbb{E}\,\|\zeta_0\|^2 \leq \gamma^2\sigma_K^2/n_K$, so (47) becomes

$$T \;\geq\; \frac{36\,\gamma^2\sigma_K^2}{n_K\beta_K\varepsilon^2} = \frac{2592\,\gamma^4\sigma_K^4}{n_K^2\,\varepsilon^4},$$

which is the last term of (42). Its cost is n_K samples, drawn at x_0, which is deployed at iteration 0 in any case for the loop itself; the warm start therefore adds no deployment. Note the n_K^2: this term benefits twice from the known-component batch, once through $\mathbb{E}\,\|\zeta_0\|^2$ and once through β_K.

Batched warm start. Forming h_0 from an average of M independent estimates gives $\mathbb{E}\,\|\zeta_0\|^2 \le \gamma^2\sigma_K^2/(Mn_K)$, and choosing $M = \lceil 1/\beta_K \rceil$ reduces (47) to

$$T \;\ge\; \frac{36\,\gamma^2\sigma_K^2}{n_K\,\varepsilon^2},$$

an ε^{-2} requirement rather than ε^{-4}. The one-time cost is $Mn_K = n_K/\beta_K = 72\gamma^2\sigma_K^2/\varepsilon^2$ samples, independent of n_K, and again a single deployment of x_0. Since the total is $n_K T = \Omega(\varepsilon^{-4})$ in both regimes, this cost is negligible, and it removes the initialization term from the leading order of Corollary 22.

Remark 33 (One-time versus per-iteration batching). The batched warm start is the one place in the analysis where averaging many samples is genuinely useful, and it is a *one-time* cost. It should not be confused with the per-iteration minibatch that the known-component momentum removes (Remark 30): there, n_K samples were required at *every* iteration, contributing a factor ε^{-2} to the total; here a single batch of the same size is drawn once. The same distinction applies to h_0 and to the residual state, for which $m_0 = 0$ is already optimal since $\mathbb{E}\,\|E_0\|^2 = \|v(x_0)\|^2$ involves no oracle noise and no better initialization is available without extra function queries.

I SETTING THE PARAMETERS

I.1 DISCUSSION OF THE PRESCRIPTIONS

The analysis assigns each parameter a role, and in each case a prescription. Proofs are in Appendix I.

The momentum ρ behaves as a sketch step size. With θ as in (8), (D1)–(D2) give $\mathbb{E}_k\,\|(I_d - bP_k)z\|^2 \le \theta(b)\|z\|^2$, with equality for all three probe families, so $\theta(b) < 1$ if and only if $b < 2/c_d$ and θ is minimized at $b = 1/c_d$. Hence

$$\rho^\star = 1 - 1/c_d, \qquad \text{while (7) is neutral at } \rho = 1 - 2/c_d \text{ and } \textit{expands} \text{ for } \rho < 1 - 2/c_d. \tag{48}$$

So ρ must scale with d, and $\rho = 0$ is not a momentum-free special case but the choice $\theta(1) = c_d - 1 \approx d$, an expanding map; the momentum-free comparison is the uncentred estimator of Appendix B.1.

The mixing weight λ plays the role of a Kalman gain. Minimizing the $\tilde\beta$-dependent part of Lemma 16 over the readout gain, keeping the exact second moment ψ of (38), with $N_k := \varsigma^2\sigma_{D,k}^2 + \Sigma_{b,k}^2$ the noise of a single fresh observation,

$$\tilde\beta_k^{\mathrm{u}} = \frac{\|E_k\|^2 + \beta_K\gamma^2\sigma_{K,k}^2}{c_d\big(\|E_k\|^2 + \gamma^2\sigma_{K,k}^2\big) + N_k}, \qquad \lambda_k^\star = \max\{0, (\tilde\beta_k^{\mathrm{u}} - \beta)/\rho\}. \tag{49}$$

As $N_k \to 0$ and $\beta_K \to 1$ this tends to $1/c_d$; as N_k grows it switches the readout off and PISO^2 degenerates to PISO, which is what convergence requires. Up to a correction we bound explicitly, the denominator is the mean squared innovation $\mathbb{E}_k\,\|R_k - m_k\|^2$, so $\tilde\beta_k^\star$ can be approximated from quantities the method already computes; since S_k involves the variance bound $\sigma_{K,k}^2$ while the identity involves the realized $\mathbb{E}_k\,\|\delta_k\|^2$, this is an adaptive heuristic justified by the identity rather than an exactly observable quantity.

The known-component momentum ρ_K is an ordinary momentum. Unlike ρ, it carries no sketch and therefore no dimension-dependent stability window: any $\rho_K \in [0, 1)$ contracts. Its value trades the variance $\mathcal{O}(\beta_K\gamma^2\sigma_K^2)$ of Lemma 1(iv) against the lag $\mathcal{O}(\gamma^2 L_K^2\eta^2/\beta_K^2)$, and through (10) a smaller β_K also forces a smaller step. Balancing the two gives $\beta_K = \Theta(\varepsilon^2/(\gamma^2\sigma_K^2))$, the choice of Corollary 23. The contrast with ρ is worth stating: ρ is the step size of a randomized sketch and diverges outside a window of width $2/c_d$, whereas ρ_K is a plain exponential average.

The weight γ acts as a control variate on the drift. With increments $\mathsf{d}_k^U := g_{\mathrm{U}}(x_{k+1}) - g_{\mathrm{U}}(x_k)$ and $\mathsf{d}_k^K := g_{\mathrm{K}}(x_{k+1}) - g_{\mathrm{K}}(x_k)$, set $\mathsf{B} := \mathbb{E}\langle \mathsf{d}^U, \mathsf{d}^K\rangle$, $\mathsf{C} := \mathbb{E}\,\|\mathsf{d}^K\|^2$ and let φ be their correlation.

Proposition 34 (Per-step drift-minimizing weight, restated). *Fix k and the displacement $x_{k+1}-x_k$. With $\mathsf{A}_k, \mathsf{B}_k, \mathsf{C}_k, \varphi_k$ as in Proposition 4 and $\mathsf{C}_k > 0$, the per-step drift $\mathbb{E}_k \|v_{k+1}-v_k\|^2 = \mathsf{A}_k + 2(1-\gamma)\mathsf{B}_k + (1-\gamma)^2\mathsf{C}_k$ is minimized at $\gamma_k^\star = 1 + \mathsf{B}_k/\mathsf{C}_k$ with minimum $\mathsf{A}_k(1-\varphi_k^2)$.*

So γ is a control-variate coefficient built from uncentred second moments, applied one derivative up: the residual observation is a control variate on gradients, and γ is one on gradient *increments*. The $\gamma = 1$ endpoint, which applies zeroth-order estimation only where it is needed, is optimal only when $\varphi = 0$. The four parameters can be set in sequence once γ is fixed: ρ depends on the probe family and dimension, λ on the signal-to-noise ratio and ρ_K on the known-oracle variance. They are not fully decoupled, since γ changes the tracked object and hence the trajectory, L_v and the noise scales that the other three depend on. We report tuned values alongside the prescriptions in §5.

I.2 The momentum ρ

Lemma 5 gives the whole story: $\beta = 1-\rho$ is the step size of a randomized fixed-point iteration for v, with stability window $\beta \in (0, 2/c_d)$ and optimum $\beta = 1/c_d$. Hence

$$\rho^\star = 1 - \frac{1}{c_d}, \qquad \text{and the iteration (7) expands whenever } \rho \le 1 - \frac{2}{c_d}. \tag{50}$$

Two consequences deserve emphasis. First, ρ must be tied to the dimension, so a constant value across a range of d is necessarily mis-set at one end. Second, $\rho = 0$ is not a momentum-free special case: it corresponds to $\beta = 1$ and $\theta(1) = c_d - 1$, an expanding map. The correct momentum-free comparison is the uncentred estimator of Appendix B.1. With q independent probes averaged per step, (19) holds with c_d replaced by $c_d^{(q)} = 1 + (c_d - 1)/q$, since for $R = \frac{1}{q}\sum_j \langle z, u_j\rangle s_j$ we have $\mathbb{E}_k\|R - z\|^2 = \frac{1}{q}\mathbb{E}_k\|\langle z,u\rangle s - z\|^2 = \frac{c_d-1}{q}\|z\|^2$; so $\rho^\star$ moves toward $1 - q/d$ and the admissible step size grows with q.

I.3 The mixing weight λ

A fixed λ is inadmissible. In (11) the coefficient of $\varsigma^2\sigma_{D,k}^2$ is $\tilde\beta^2 + \frac{5}{2}\theta\beta$. For PISO, $\tilde\beta = \beta$ and this is $\mathcal{O}(\beta)$, so it vanishes as $\beta \to 0$. For PISO^2 with λ held fixed, $\tilde\beta \to \lambda$ as $\beta \to 0$ and the floor stalls at $2\lambda^2\varsigma^2\sigma_D^2 > 0$. The reason is structural: the readout injects a fraction $\tilde\beta$ of a single fresh observation directly into the search direction, and one observation carries irreducible variance. Corollaries 23–29 are therefore stated for $\lambda = 0$, and PISO^2 requires $\lambda \to 0$ as $\varepsilon \to 0$.

Proposition 35 (Optimal readout gain). *Let $V_k := \gamma^2\sigma_{K,k}^2$, $S_k := \|E_k\|^2 + V_k$, $S_k' := \|E_k\|^2 + \beta_K V_k$ and $N_k := \varsigma^2\sigma_{D,k}^2 + \Sigma_{b,k}^2$, and recall that $\lambda \in [0,1]$ constrains $\tilde\beta = \beta + \rho\lambda$ to $[\beta, 1]$. Keeping the exact second moment $\psi(\beta_K, \tilde\beta) = \beta_K^2 - 2\beta_K\tilde\beta + c_d\tilde\beta^2$ of (38) rather than its upper bound, the $\tilde\beta$-dependent part of the direction-error bound is minimized without constraint at*

$$\tilde\beta_k^{\mathrm{u}} = \frac{S_k'}{c_d S_k + N_k} = \frac{\|E_k\|^2 + \beta_K\gamma^2\sigma_{K,k}^2}{c_d\big(\|E_k\|^2 + \gamma^2\sigma_{K,k}^2\big) + N_k}, \qquad \textit{and the feasible optimum is} \qquad \tilde\beta_k^\star = \min\big\{1, \max\{\beta, \tilde\beta_k^{\mathrm{u}}\}\big\}, \tag{51}$$

with $\lambda_k^\star = (\tilde\beta_k^\star - \beta)/\rho$. The numerator collapses to S_k only when $\beta_K = 1$, that is without the known-component average. As $N_k \to 0$ and $\beta_K \to 1$, $\tilde\beta_k^{\mathrm{u}} \to 1/c_d$; as $N_k/S_k \to \infty$, $\tilde\beta_k^{\mathrm{u}} \to 0$ and the lower clip is active, giving $\lambda_k^\star = 0$, that is PISO.

Proof. From Lemma 16, retaining ψ exactly in place of $\beta_K^2 + c_d\tilde\beta^2$, the $\tilde\beta$-dependent part is

$$g(\tilde\beta) = \theta(\tilde\beta)\|E_k\|^2 + \psi(\beta_K, \tilde\beta)V_k + \tilde\beta^2 N_k = \big(\|E_k\|^2 + \beta_K^2 V_k\big) - 2\tilde\beta\big(\|E_k\|^2 + \beta_K V_k\big) + \tilde\beta^2\big(c_d\|E_k\|^2 + c_d V_k + N_k\big),$$

a strictly convex quadratic in $\tilde\beta$ with unconstrained minimizer $S_k'/(c_d S_k + N_k)$. On the interval $[\beta, 1]$ the minimizer of a convex quadratic is its unconstrained minimizer clipped to the interval. The limits are immediate. □

The form $S'/(c_d S + N)$ is a scalar Kalman gain: correctable signal over total signal plus noise, with c_d the variance inflation of the sketch and the known-component variance counted at weight β_K in

the numerator because only that fraction of it reaches the step. Its denominator is approximately observable from the innovation, because it is essentially the mean squared innovation of the method.

Proposition 36 (Innovation identity)**.** *Let* $\widetilde{N}_k := \Sigma_{b,k}^2 + \mathbb{E}_k[\xi_k^2 \|s_k\|^2]$. *If* (19) *holds with equality then*

$$\mathbb{E}_k \left\| R_k - m_k \right\|^2 = c_d \Big(\|E_k\|^2 + \mathbb{E}_k \|\delta_k\|^2 \Big) + \widetilde{N}_k - 2\langle E_k, \bar{B}_k \rangle, \tag{52}$$

and the correction term obeys $2|\langle E_k, \bar{B}_k \rangle| \le 2\varsigma_4 \Sigma_{b,k} \|E_k\|$*, so its size relative to the leading term is at most* $2\varsigma_4 \Sigma_{b,k} / (c_d \|E_k\|)$.

Proof. By (24), $\|R_k - m_k\|^2 = \big(\langle E_k + \delta_k, u_k \rangle - b_k - \xi_k\big)^2 \|s_k\|^2$. Expanding the square gives six terms, which we evaluate individually.

(i) $\mathbb{E}_k[\langle E_k, u_k \rangle^2 \|s_k\|^2] = c_d \|E_k\|^2$, by (19) with equality.

(ii) $\mathbb{E}_k[\langle \delta_k, u_k \rangle^2 \|s_k\|^2] = c_d \, \mathbb{E}_k \|\delta_k\|^2$, by conditioning on δ_k, which is independent of (u_k, s_k), and applying (19) with equality.

(iii) $2\,\mathbb{E}_k[\langle E_k, u_k \rangle \langle \delta_k, u_k \rangle \|s_k\|^2] = 0$, since the expression is linear in the mean-zero δ_k, which is independent of (u_k, s_k).

(iv) $\mathbb{E}_k[b_k^2 \|s_k\|^2] + \mathbb{E}_k[\xi_k^2 \|s_k\|^2] = \widetilde{N}_k$, by definition.

(v) $-2\,\mathbb{E}_k[\langle E_k + \delta_k, u_k \rangle b_k \|s_k\|^2] = -2\langle E_k, \mathbb{E}_k[b_k \|s_k\|^2 u_k] \rangle = -2\langle E_k, \bar{B}_k \rangle$, the δ_k part vanishing as in (iii).

(vi) $-2\,\mathbb{E}_k[\langle E_k + \delta_k, u_k \rangle \xi_k \|s_k\|^2] = 0$ and $2\,\mathbb{E}_k[b_k \xi_k \|s_k\|^2] = 0$, because $\mathbb{E}[\xi_k \mid \mathcal{F}_k, u_k, s_k, \delta_k] = 0$ and the remaining factors are measurable with respect to that conditioning.

Summing gives (52); the final bound is Cauchy–Schwarz together with Lemma 10. □

Consequently $c_d S_k + N_k$ can be approximated by the running mean of $\|R_k - m_k\|^2$, which the method already computes, up to the correction of Proposition 36, which is small whenever the tracking error exceeds the finite-difference bias scale. This gives the denominator of (51) directly. The numerator $S_k' = \|E_k\|^2 + \beta_K V_k$ differs from $S_k = \|E_k\|^2 + V_k$ by $(1 - \beta_K) V_k$, so an estimate or bound $\widehat{V}_k$ of the known-component variance is also needed when $\beta_K < 1$: given $\widehat{N}_k$ and $\widehat{V}_k$, set $\widehat{S}_k = (\widehat{\|R - m\|^2} - \widehat{N}_k)/c_d$, $\widehat{S}_k' = \widehat{S}_k - (1 - \beta_K)\widehat{V}_k$, and read off $\tilde{\beta}_k^\star$ from (51). The gain is thus approximately estimable rather than directly observable. Since $\|E_k\|$ decreases along a successful run, $\tilde{\beta}_k^\star$ decays, so a constant λ tuned on a fixed budget approximates a time-average of a decaying schedule.

I.4 The weight γ

Proof of Proposition 4. By (4), $v_{k+1} - v_k = \mathsf{d}_k^U + (1 - \gamma)\mathsf{d}_k^K$ exactly. With k and the displacement fixed, the increments do not depend on γ and expanding gives

$$\mathbb{E} \left\| \mathsf{d}^U + (1 - \gamma)\mathsf{d}^K \right\|^2 = \mathsf{A} + 2(1 - \gamma)\mathsf{B} + (1 - \gamma)^2 \mathsf{C},$$

with $\mathsf{A} := \mathbb{E} \|\mathsf{d}^U\|^2$, $\mathsf{B} := \mathbb{E}\langle \mathsf{d}^U, \mathsf{d}^K \rangle$ and $\mathsf{C} := \mathbb{E} \|\mathsf{d}^K\|^2$. This is a strictly convex quadratic in γ when $\mathsf{C} > 0$; setting the derivative with respect to $(1 - \gamma)$ to zero gives $(1 - \gamma)^\star = -\mathsf{B}/\mathsf{C}$, that is $\gamma^\star = 1 + \mathsf{B}/\mathsf{C}$, with minimum value $\mathsf{A} - \mathsf{B}^2/\mathsf{C} = \mathsf{A}(1 - \varphi^2)$ where $\varphi = \mathsf{B}/\sqrt{\mathsf{A}\mathsf{C}}$. At $\gamma = 1$ the drift is A, and at $\gamma = 0$ the tracked object is ∇F and the drift is $\mathbb{E} \|\nabla F(x_{k+1}) - \nabla F(x_k)\|^2$, which is at most $L^2 \eta^2 \, \mathbb{E} \|G_k\|^2$ by L-smoothness.

We stress what the proposition does not say. The constant L_v of (17) is a uniform bound over the trajectory, and the trajectory depends on γ, so the per-step minimizer $\gamma_k^\star$ is not the minimizer of $L_v(\gamma)$ and no statement of the form $L_v(\gamma^\star)/L_v(1) = \sqrt{1 - \varphi^2}$ is claimed for the uniform constant. Moreover $\gamma_k^\star > 0$ requires $\mathsf{B}_k > -\mathsf{C}_k$. □

Thus γ is a control-variate coefficient built from uncentred second moments, applied to gradient increments rather than to gradients. It is estimable: running with $\gamma = 1$ so that m_k tracks g_{U}, recording d_k^K and $m_{k+1} - m_k$ and regressing the latter on the former estimates B/C, while the coefficient of determination estimates φ^2.

Remark 37 (Where γ enters, and where it does not). Inspecting (11), γ appears in exactly three places: the noise term $\frac{7}{2}\theta\gamma^2\sigma_{K,k}^2$, which increases in γ; the initialization $\|v(x_0)\|^2 = \|g_{\mathrm{U}}(x_0) + (1-\gamma)g_{\mathrm{K}}(x_0)\|^2$; and L_v, through the step-size restriction. It does not affect c_d, and therefore neither $\rho^\star$ nor $\tilde{\beta}^\star$. Including the noise term, minimizing the per-step drift proxy $\mathsf{A}_k+2(1-\gamma)\mathsf{B}_k+(1-\gamma)^2\mathsf{C}_k$ plus $\nu\gamma^2$ with $\nu \propto \sigma_K^2/(n_F\varepsilon^2)$ shrinks the regression estimate to $\gamma_\nu^\star = \frac{\mathsf{C}}{\mathsf{C}+\nu}\gamma^\star$, recovering $\gamma^\star$ when the known-component oracle is exact and $\gamma \to 0$ when it is very noisy.

Remark 38 (A time-varying γ_k). Allowing γ to vary changes the tracked object, adding $-(\gamma_{k+1} - \gamma_k)g_{\mathrm{K}}(x_{k+1})$ to $v_{k+1} - v_k$, so that $\mathbb{E}_k\|v_{k+1} - v_k\|^2 \le 2L_v^2\eta^2\,\mathbb{E}_k\|G_k\|^2 + 2|\gamma_{k+1} - \gamma_k|^2\,\mathbb{E}_k\|g_{\mathrm{K}}(x_{k+1})\|^2$. The first term is absorbed as before with η reduced by $\sqrt{2}$; the second contributes an additional floor $\frac{20\theta}{\beta^2}\cdot\frac{1}{T}\sum_k|\gamma_{k+1} - \gamma_k|^2\,\mathbb{E}\|g_{\mathrm{K}}\|^2$, so any schedule with $\sum_k|\gamma_{k+1} - \gamma_k|^2 = \mathcal{O}(\beta^2\varepsilon^2 T/\theta)$ leaves the rate unchanged.

J Probes per update

We justify Remark 32. In the analysis of Hikima & Takeda (2025a) for the Gaussian multi-direction estimator with N probes per iteration, the bound on the estimator error contains a term proportional to $\frac{d}{N}\|\nabla F(x)\|^2$, with the stated constant giving $\frac{54d}{N}\|\nabla F(x)\|^2$ after their Lemma 1. For a descent argument to close, this term must be absorbed into the $-\|\nabla F(x)\|^2$ produced by the descent step, which requires $54d/N < 1$, that is $N = \Omega(d)$; obtaining their stated constant requires $N \ge 81d$. Each probe costs two function queries, so at least $162d$ queries precede every parameter update. At $d = 30$ this is 4860 queries, against a total budget of 5000 in their own experiments.

By contrast, PISO attains the rates of Corollaries 24 and 25 with a single probe per iteration, because the averaging that reduces the sketch variance is performed across iterations by the contraction (7) rather than within one iteration. The two designs reach the same ε and d dependence with different granularity: PISO updates after $\mathcal{O}(1)$ deployments, whereas a multi-direction method must release $\Omega(d)$ decisions between updates.

K Deployment accounting

K.1 Deployments, the binding cost in performative prediction

In Regime B a query is not an observation but an intervention: evaluating $\widehat{F}$ at a point requires *deploying* that decision and collecting outcomes from the induced distribution. Each iteration of Algorithm 1 deploys three decisions, namely x_k for the known-component oracle and $x_k \pm \mu_k u_k$ for the two function values, so the total number of deployments is $3T$ regardless of the batch sizes. Deployments and samples are therefore distinct resources, and in an application where a decision is a price schedule, a lending rule or a policy, the number of distinct decisions released is the cost that matters.

The two batch sizes act on this cost in the same direction, through β and β_K. Averaging n_F samples at each query point divides $\sigma_{D,k}^2$ by n_F and permits a proportionally larger β; averaging n_K samples of $\widehat{g}_{\mathrm{K}}$ divides $\sigma_{K,k}^2$ by n_K and permits a proportionally larger β_K. Since $T = \Theta(\Delta_v/(\beta\varepsilon^2))$, both enlarge the step per iteration and so reduce T. Two structural caps limit how far this can go: the sketch stability window $\beta \le 1/c_d$ of Appendix I, and the condition $\beta \le 1/(3d)$ under which the weighted bias vector $\bar{B}_k$ of Lemma 1 stays dominated by the mean bias (Appendix D). Both are $\Theta(1/d)$, so with $C_1 := 102000L^2\sigma_F^2\varsigma^2\chi_1^2 = \Theta(d^2)$ the two extreme operating points are as follows.

the two extreme operating points are compared in the table below.

Two caps on β govern how far this goes. The first is $\beta_1 := \chi_1^2/\varsigma_6^2$, the hypothesis of Corollary 24 under which the damped second-moment bias $11\beta\Sigma_b^2 \le \frac{11}{4}\beta L^2\mu^2\varsigma_6^2$ stays dominated by the undamped mean bias $120L^2\mu^2\chi_1^2$; for Gaussian probes $\beta_1 = 1/(9d^2)$. The second is the

cap $\beta_2 := 1/(3d)$ of (39), which is the limit of the sharp bias-vector domination argument of Lemma 11 and not a stability boundary: the sketch contraction itself holds up to $2/c_d$, and Theorem 2 uses the conservative range $\beta \le 1/c_d$. For $\beta \le \beta_1$ the radius balance is as in Corollary 24 and $\mathsf{N} = 2n_F T$ is independent of β; deployments are $3T = \Theta(\Delta_v/(\beta\varepsilon^2))$ and reach $\Theta(d^2\Delta_v\varepsilon^{-2})$ at β_1, plus the additive term $\gamma L_K \Delta_F \varepsilon^{-2}$ from the known-component branch at $\beta_K = \Theta(1)$, which is absorbed whenever $L_v \ge \gamma L_K/d^2$. For $\beta_1 < \beta \le \beta_2$ the bracket in the radius balance becomes $120\chi_1^2 + \frac{11}{4}\beta\varsigma_6^2 = \Theta(\beta d^3)$, so the required n_F grows as $\beta^2 d^4 \varepsilon^{-4}$ and $\mathsf{N} = 2n_F T = \Theta(\beta d^4 \Delta_v \varepsilon^{-6})$ grows linearly in β, reaching $\Theta(d^3\Delta_v\varepsilon^{-6})$ at β_2, where deployments are $\Theta(d\Delta_v\varepsilon^{-2})$. The minimum deployment count attainable without loss in N is therefore $\Theta(d^2\Delta_v\varepsilon^{-2})$, and $\Theta(d\Delta_v\varepsilon^{-2})$ costs a factor d in observations. The gap between the two caps is exactly the looseness of the generic bound $\Sigma_b^2 \le \frac{1}{4}L^2\mu^2\varsigma_6^2$: on the half-quadratic example of Remark 13 one computes $\Sigma_b^2 = \frac{\mu^2}{16}d(3d+12) = \Theta(L^2\mu^2 d^2)$, a factor d below the bound, and a general bound of that order would place β_1 at $\Theta(1/d)$ and make the last operating point free. We do not have such a bound and leave it open. Third, at that point $\beta_K = \Theta(1)$: the known-component average is close to inactive at the deployment-optimal end, and its role is to make the opposite end, $n_K = 1$, admissible at all (Appendix I). The choice of n_K therefore interpolates between minimizing known-component samples and minimizing deployments.

The operating points compare as follows.

	β	deployments $3T$	function evals. $2n_F T$	known samples $n_K T$
$n_F = n_K = 1$	ε^4/C_1	$\Theta(d^2\varepsilon^{-6})$	$\Theta(d^2\varepsilon^{-6})$	$\Theta(d^2\varepsilon^{-6})$
batches raised until $\beta = \beta_1$	$\Theta(1/d^2)$	$\Theta(d^2\varepsilon^{-2})$	$\Theta(d^2\varepsilon^{-6})$	$\Theta(d\,\varepsilon^{-4})$
batches raised until $\beta = \beta_2$	$\Theta(1/d)$	$\Theta(d\,\varepsilon^{-2})$	$\Theta(d^3\varepsilon^{-6})$	$\Theta(d\,\varepsilon^{-4})$

Comparison. A multi-direction method draws N probes per update, which is $2N$ *distinct* deployed decisions, and this cannot be replaced by repeated sampling because distinct directions are the object. With $N = \Theta(d^2\varepsilon^{-4})$ and $T = \Theta(\varepsilon^{-2})$ (Hikima & Takeda, 2025a), its deployment count coincides with its sample count, $\Theta(d^2\varepsilon^{-6})$. Against this, PISO at β_1 deploys a factor ε^{-4} fewer decisions for the same total number of observations, and at β_2 a factor $d\,\varepsilon^{-4}$ fewer at the cost of a factor d more observations.

L ADDITIONAL EXPERIMENTAL RESULTS

Table 3: Strategic classification: test loss and test accuracy (%) at the end of the sample budget, averaged over 100 simulations, for each strategic-response strength τ. Best per column in bold. All four PISO variants dominate every baseline on both metrics at every τ, by roughly 0.2 in test loss and 4–10 accuracy points. The two methods that ignore the computable component, ZO-TG and ZO-OGVR, degrade as τ grows, that is as the opaque component becomes more influential, whereas the PISO variants are essentially flat in τ; the guided baselines improve with τ but never close the gap. Training metrics are in the accompanying archive.

	$\tau = 0.5$		$\tau = 1$		$\tau = 2$		$\tau = 4$		$\tau = 8$	
Method	loss	acc.	loss	acc.	loss	acc.	loss	acc.	loss	acc.
GZO-NS	0.891	59.1	0.912	60.8	0.899	63.2	0.879	65.0	0.878	65.7
GZO-HS	0.846	57.2	0.862	59.8	0.865	62.0	0.835	63.7	0.821	65.0
ZO-TG	1.428	45.4	1.487	46.0	1.505	45.8	1.558	44.6	1.580	44.1
ZO-OGVR	1.181	51.4	1.243	52.8	1.383	54.3	1.539	54.9	1.534	56.4
GaussianPISO	0.624	**66.6**	**0.620**	67.6	0.632	67.7	0.622	**69.3**	0.631	69.0
CyclePISO	**0.622**	66.6	0.627	**67.6**	**0.620**	**68.3**	**0.619**	69.0	**0.618**	69.5
GaussianPISO^2	0.632	66.2	0.629	66.8	0.641	67.6	0.628	68.2	0.636	69.0
CyclePISO^2	0.623	66.5	0.632	67.0	0.632	68.0	0.627	68.9	0.621	**69.7**

Performative RL. Among methods that only query the environment (Figure 4) the ordering is PISO $\approx$ PISO^2, with GZO-NS closing part of the gap late in the run, followed by the remaining zeroth-order baselines and the vanilla policy gradient. We regard PePG as the informative reference

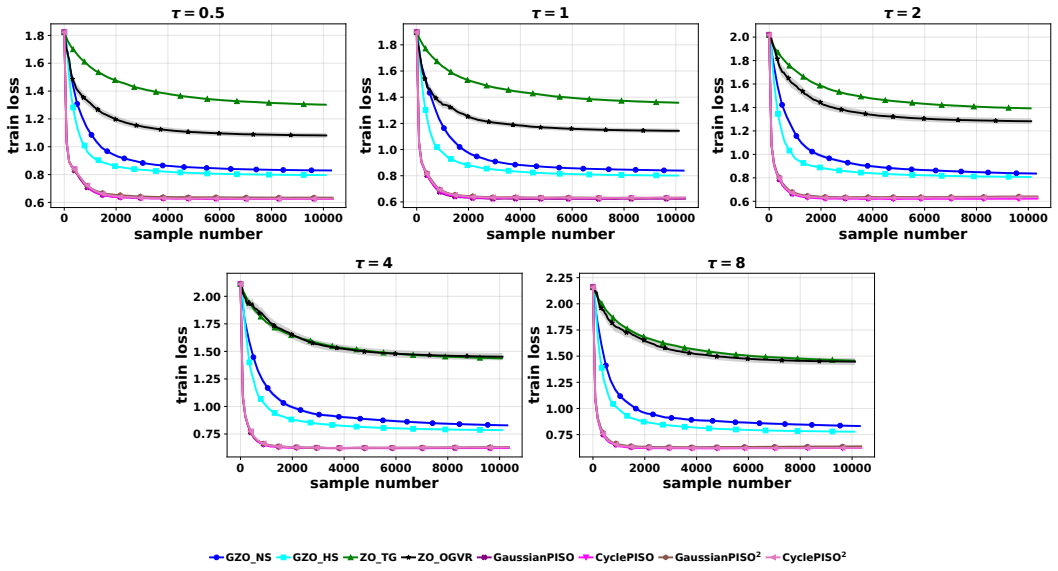


Figure 3: Strategic classification on the credit dataset, $d = 12$, five strategic-response strengths τ. Training loss against samples consumed; means over 100 simulations with one-standard-error bands. All four PISO variants reach a loss of about 0.63 at every τ, while the strongest baseline plateaus near 0.80; the two methods that ignore the computable component degrade sharply as τ increases.

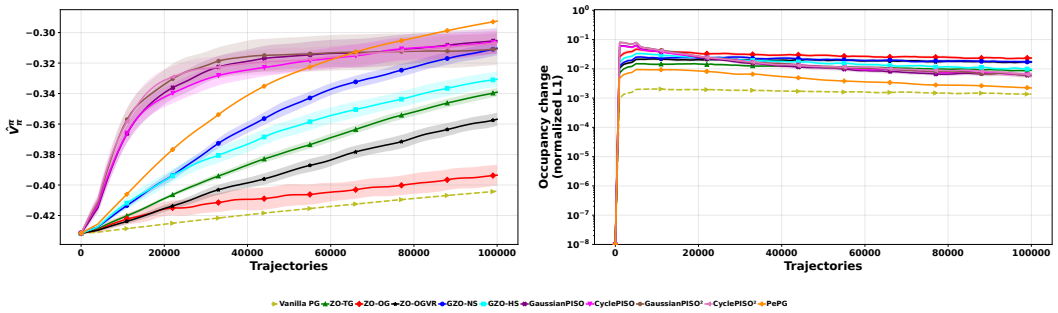


Figure 4: Performative reinforcement learning on an 8×8 gridworld, 20 seeds. Left: performative return $\widehat{V}_{\pi}^{\pi}$ against the number of trajectories consumed, with ± 1.96 standard-error bands. Right: normalized L_1 change of the state occupancy between consecutive updates, on a log scale. PISO and PISO^2 make the fastest early progress among the methods that operate under (1), while GZO-NS closes part of the gap late in the run. PePG is a reference with privileged information: it is supplied with exact reward and transition response Jacobians and exact tabular advantages, so it differentiates the environment's response rather than querying it. Its curve is not a bound on what other methods can attain, but it indicates what a differentiable response model is worth.

rather than a competitor: it is handed the derivative that a query-based method must estimate, and what the figure shows is how much of the remaining range is recovered without any model of the response. The right panel is a diagnostic of how far each method moves the policy per update. PISO's occupancy change is of order 10^{-1} early in the run and decays to about 10^{-2} at the end, while PePG and the vanilla policy gradient make smaller late-stage moves. So PISO takes larger policy steps early in the run, and part of its advantage over the vanilla policy gradient may reflect that; every baseline's step schedule was tuned independently (Appendix L.2), so this is not an artefact of a shared setting, but we do not claim the gap is attributable to the estimator alone. The decay of PISO's occupancy change over the run indicates that the larger steps do not destabilize the policy.

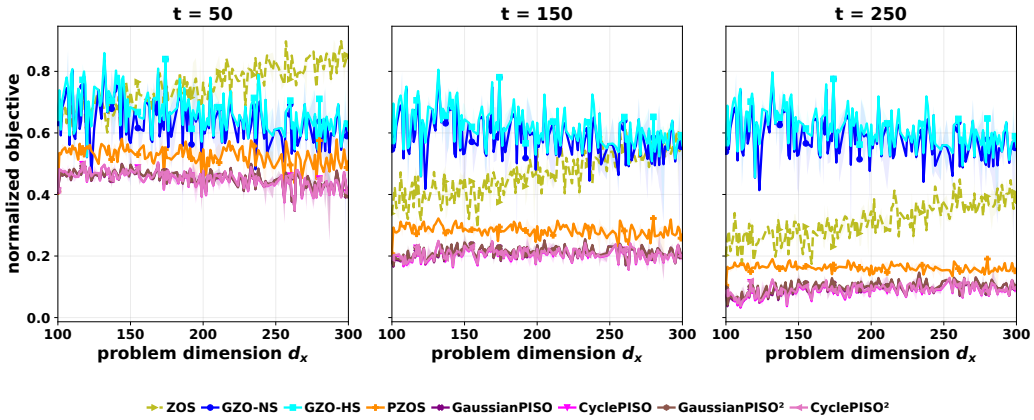


Figure 5: Security game, normalized objective against problem dimension $d_x \in [100, 300]$ at three iteration counts. The separation between PISO and PZOS is stable across the range, and both separate from the guided baselines, whose objective does not decrease appreciably after $t = 50$.

L.1 Ablations

All ablations use the pricing problem of §5, so $d = 10$ and hence $c_d = 12$ for Gaussian probes and $c_d = 10$ for the cycling variant. The batch size starts at 30 and increases by 2 per iteration, so the known-component batch n_K is large throughout; this matters for Figure 9.

L.2 Experimental hyperparameters

For reproducibility, we report the complete hyperparameter configurations used to produce the experimental results. Unless stated otherwise, a decay factor is applied multiplicatively at every update. For the baselines, we use the hyperparameters reported as optimal in their respective references, except for performative RL, where we tune the zeroth-order baselines independently.

Radius and known-part conventions. We denote the perturbation radius by r_k. For ZO-TG, ZO-OG, ZO-OGVR, and the Gaussian and Cycle variants of PISO and PISO^2, we use $\mu_k = r_k/\sqrt{d}$. For GZO-NS, GZO-HS, ZOS, and PZOS, we use $\mu_k = r_k$. Thus, across methods, r_k represents the scale of the actual perturbation after accounting for the normalization of the probe direction. For exponential schedules we write $r_k = r_0 q_r^k$ and $\eta_k = \eta_0 q_\eta^k$, where q_r and q_η are the radius and step decay factors. When a radius floor is used, we denote it by $r_{\min}$. We denote the known-sample fraction by f_K. Unless stated otherwise, the reported configurations use a fresh known part, equivalently $\rho_K = 0$.

L.2.1 Multiproduct pricing

Let $m_k := 30 + 2k$, where k is the current iteration number.

GZO-NS. We set $r_0 := 0.19$, $r_{\min} := 10^{-4}$, $q_r := 0.95$, $\beta_0 := 0.01$, the step decay to 0.95, $\alpha_0 := 0$, and the damping factor to 0.98. The batch size is m_k.

GZO-HS. The r_k and step schedules are the same as for GZO-NS. We use a historical-guide window of 1, $\alpha_0 := 0$, damping 0.98, and batch size m_k.

ZO-TG. We set $r_0 := 0.19$, $r_{\min} := 10^{-4}$, and $q_r := 0.95$; the step starts from $\beta_0 := 0.0031623$ and decays by 0.95. We use the Gaussian two-point estimator with batch size m_k.

ZO-OGVR. We use the same r_k schedule as above, with $\beta_0 := 0.001$ and step decay 0.95. The initial baseline uses 20 samples, the history length is $s_{\max} := 10$, the distance coefficient is $M := 0.1$, and the batch size is m_k.

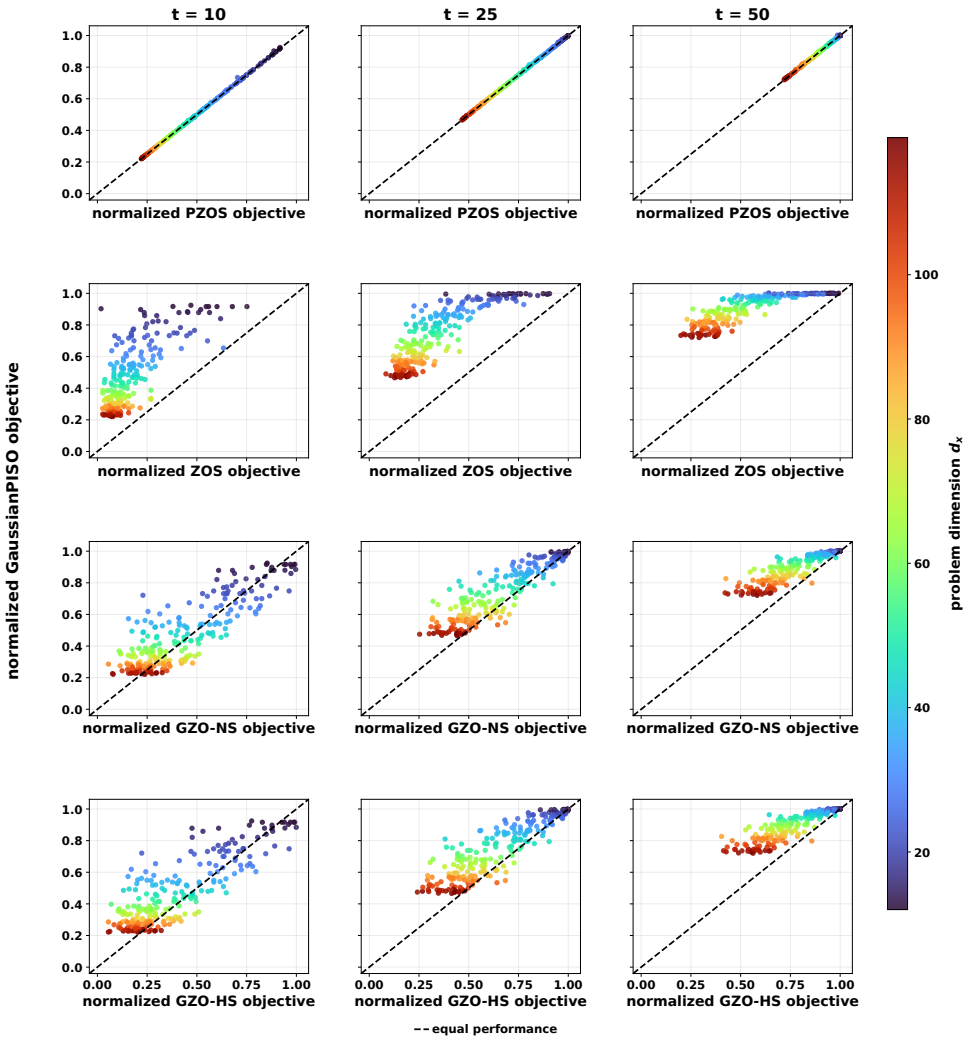


Figure 6: Routing, per-instance comparison of GaussianPISO against each baseline at $t = 10, 25, 50$, coloured by problem dimension. Points above the diagonal favour PISO, which is a maximization problem here. The PZOS row lies on the diagonal at every horizon, so the two methods return the same solution per instance; the ZOS, GZO-NS and GZO-HS rows lie above it, with the margin largest at small t and at small dimension.

Gaussian PISO. We set $r_0 := 0.40$, $q_r := 0.96$, $\eta_0 := 0.005$, and $q_\eta := 0.98$. We use $\gamma := 2.00$, $\rho := 0.92$, and $f_K := 0.50$. The base batch is m_k, giving total PISO cost $3m_k$ per update.

Cycle PISO. The parameters are the same as for Gaussian PISO, with cycle length $L := 10$.

Gaussian PISO^2. We set $r_0 := 0.40$, $q_r := 0.96$, $\eta_0 := 0.005$, and $q_\eta := 1$. We use $\gamma := 2.00$, $\rho := 0.92$, $\lambda := 0.25$, and $f_K := 0.50$. the total cost is $3m_k$ per update.

Cycle PISO^2. The parameters are the same as for Gaussian PISO^2, with cycle length $L := 10$.

We select the PISO hyperparameters using a two-stage grid search. In the first stage, we search over $r_0 \in \{0.20, 0.40, 0.80\}$, $\eta_0 \in \{0.0025, 0.005, 0.010\}$, and $\rho \in \{0.90, 0.92, 0.97\}$ for all four variants. In the second stage, we search over $f_K \in \{0.34, 0.50, 0.67\}$ for all four variants, the cycle length $L \in \{2, 5, 10\}$ for Cycle PISO and Cycle PISO^2, and the mixing weight $\lambda \in \{0.10, 0.25, 0.50\}$ for Gaussian PISO^2 and Cycle PISO^2.

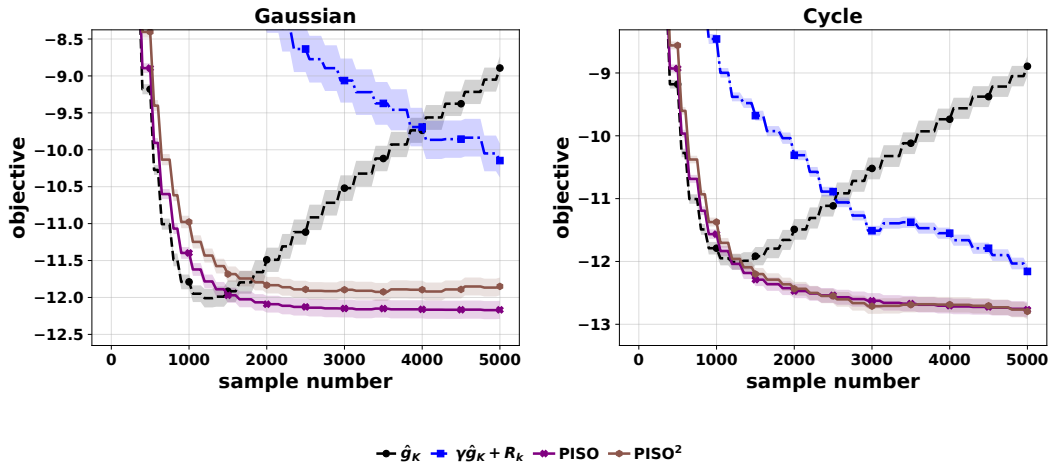


Figure 7: Components. $\widehat{g}_{\mathrm{K}}$ alone (black) uses only the computable part and diverges once the opaque part matters; $\gamma\widehat{g}_{\mathrm{K}} + R_k$ (blue) adds the *uncentred* residual observation of Appendix B.1 with $\rho = 0$, and is unstable; PISO and PISO^2 add the centring and the momentum. The two unstable arms are the two ways of omitting a piece of (6), and confirm that neither the known component nor an uncentred zeroth-order correction suffices on its own.

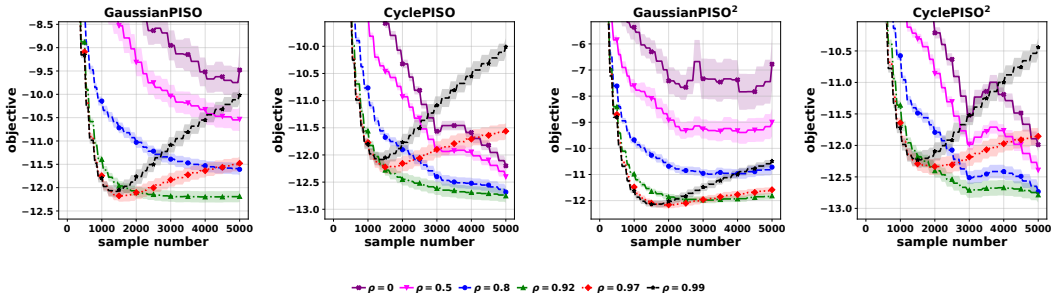


Figure 8: Momentum ρ, the sharpest test of the analysis. Theory predicts an optimum at $\rho^\star = 1 - 1/c_d$, that is 0.917 for the Gaussian panels and 0.900 for the cycling panels, and predicts that the residual iteration is neutral at $\rho = 1 - 2/c_d$ and *expands* below it, that is at 0.833 and 0.800 respectively (Appendix I). Both are visible. The best curve is $\rho = 0.92$ in all four panels, the nearest grid point to the prediction; $\rho \in \{0, 0.5, 0.8\}$ all descend initially and then *rise*, the signature of a tracking error that grows once the sketch step size leaves its stability window; and $\rho = 0.99$ is over-smoothed, descending furthest before turning up as the state lags the moving residual.

L.2.2 Strategic classification

The optimization batch is $30 + 2k$ throughout unless stated otherwise.

GZO-NS. We set $r_0 := 10$, $r_{\min} := 0.10$, $q_r := 0.95$, $\beta_0 := 1.0$, step decay 0.95, $\alpha_0 := 0$, and damping 0.98. The optimization batch is $30 + 2k$ and the guide batch is $100 + 2k$.

GZO-HS. The r_k and step schedules are the same as for GZO-NS. We set $\alpha_0 := 0$, damping to 0.98, and the historical window to 1, with batch size $30 + 2k$.

ZO-TG. We set $r_0 := 10$, $r_{\min} := 0.10$, and $q_r := 0.95$, and use $\beta_0 := 1.0$ with step decay 0.95. The estimator is Gaussian two-point and the batch size is $30 + 2k$.

ZO-OGVR. We use the same r_k schedule, with $\beta_0 := d^{-1/2}$ and step decay 0.95. The initial baseline uses 20 samples, $s_{\max} := 10$, $M := 0.10$, and batch size $30 + 2k$.

Gaussian PISO. We set $r_0 := 3.00$, $q_r := 0.95$, $\eta_0 := 0.70$, and $q_\eta := 0.98$. We use $\gamma := 3.00$, $\rho := 0.98$, and $f_K := 0.67$, with base batch $30 + 2k$.

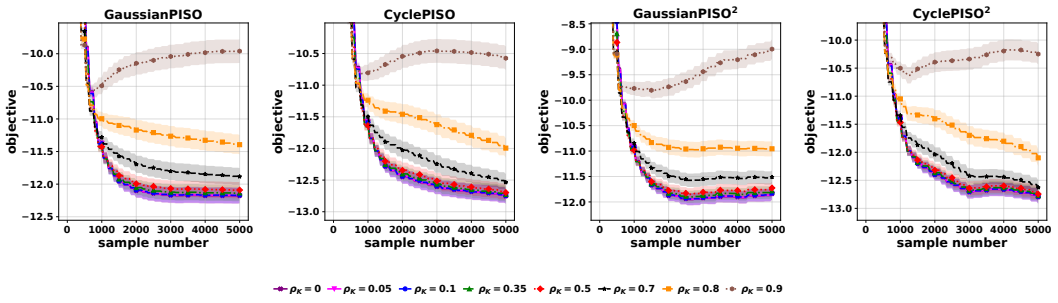


Figure 9: Known-component momentum ρ_K. Here $\rho_K = 0$, that is $\beta_K = 1$ and a fresh estimate, is among the best choices, and performance degrades monotonically for $\rho_K \geq 0.7$. This is what the analysis predicts in this regime: with the growing batch the known-component variance σ_K^2/n_K is already small, so $\beta_K = 1$ satisfies the floor condition of (40) and a smaller β_K only tightens the step-size cap $\eta \leq \beta_K/(9\gamma L_K)$ of (10) and adds lag. The known-component average earns its place when n_K is small, where it is what makes $n_K = 1$ admissible (Remark 30); it is not needed when one can afford a large batch.

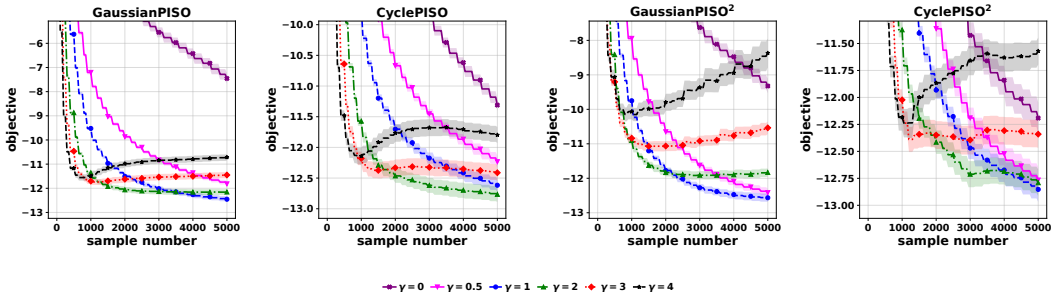


Figure 10: Known-component weight γ. The endpoints behave as §G.4 describes: $\gamma = 0$ tracks ∇F and is worst, and large γ degrades as the noise term $\gamma^2\sigma_K^2$ and the drift constant $L_v(\gamma)$ both grow. The interior optimum is $\gamma \in \{1, 2\}$ depending on the variant, consistent with Proposition 4 placing $\gamma^\star = 1 + \mathsf{B}/\mathsf{C}$ away from both endpoints, though the grid is too coarse to resolve $\gamma^\star$ precisely and we do not attempt to estimate it directly.

Cycle PISO. The parameters are the same as for Gaussian PISO, with cycle length $L := 6$.

Gaussian PISO^2. The r_k and η_k schedules are the same as for Gaussian PISO. We use $\gamma := 3.00$, $\rho := 0.98$, $\lambda := 0.05$, and $f_K := 0.75$, with base batch $30 + 2k$.

Cycle PISO^2. The parameters are the same as for Gaussian PISO^2, with cycle length $L := 6$.

We select the PISO hyperparameters using a two-stage grid search. In the first stage, we search over $r_0 \in \{1.50, 3.00, 6.00\}$, $\eta_0 \in \{0.35, 0.70, 1.40\}$, and $\rho \in \{0.95, 0.98, 0.99\}$ for all four variants. In the second stage, we search over $f_K \in \{0.50, 0.67, 0.75\}$ for all four variants, the cycle length $L \in \{3, 6, 12\}$ for Cycle PISO and Cycle PISO^2, and the mixing weight $\lambda \in \{0.01, 0.05, 0.10\}$ for Gaussian PISO^2 and Cycle PISO^2.

L.2.3 PERFORMATIVE REINFORCEMENT LEARNING

For this experiment, an inverse-square-root schedule with initial value q_0, scale S_q, and minimum $q_{\min}$ means

$$q(N) := \max\left\{q_{\min}, \frac{q_0}{\sqrt{1 + N/S_q}}\right\},$$

where N is the cumulative trajectory count.

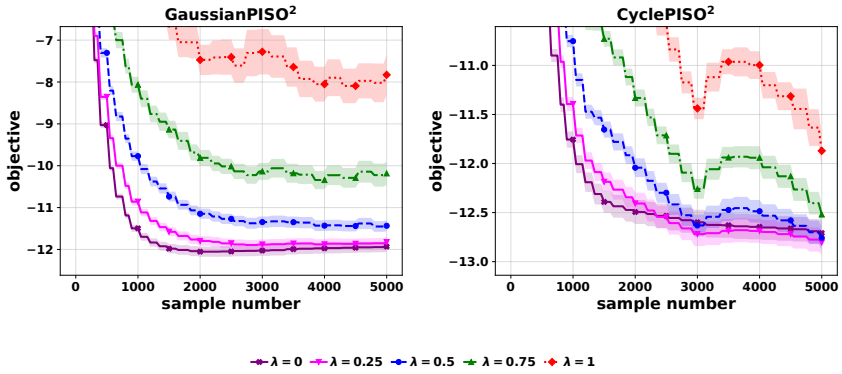


Figure 11: Mixing weight λ. Performance is monotonically decreasing in λ over the grid $\{0, 0.25, 0.5, 0.75, 1\}$, with $\lambda = 1$, which discards the state and uses the fresh observation alone, far worse than $\lambda = 0$. This is consistent with (51): at $\rho = 0.92$ and $c_d \in \{10, 12\}$ the high-signal-to-noise optimum $\lambda^\star = (1/c_d - \beta)/\rho$ is 0.004 and 0.022, both below the smallest nonzero grid point, so the grid cannot separate $\lambda^\star$ from 0 and only the degradation for large λ is resolved. A finer sweep near zero would be needed to test the prescription itself.

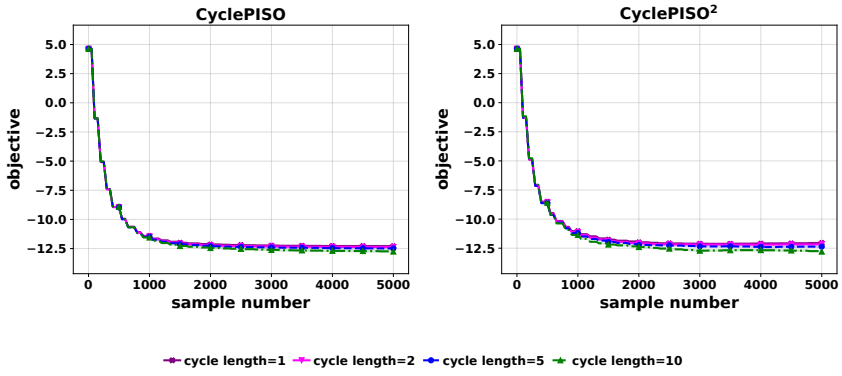


Figure 12: Cycle length L for the orthogonalized probe, at $d = 10$. Longer cycles are mildly better and $L = d = 10$ is best, but the differences are small and within a fraction of the spread of the ρ ablation. Since the per-step marginal is the same for every L, and only the dependence between consecutive probes changes, the analysis of §3 does not distinguish these curves; the ordering is consistent with orthogonality reducing the effective sketch variance, but we do not claim it as a prediction.

Vanilla PG. The step schedule has initial value 0.05, scale 10^5, and minimum 0.02. We use batches of 100 trajectories and clip the gradient norm at 0.25.

ZO-TG. We set $r_0 := 4.0$, $S_r := 2 \cdot 10^5$, and $r_{\min} := 3.0$; the step schedule has initial value 0.40, scale 10^5, and minimum 0.20. We use 50 trajectories per side and gradient clip 0.25.

ZO-OG. We set $r_k := 0.50$ and use a step schedule with initial value 0.08, scale 10^5, and minimum 0.04. We use batch size 100, baseline decay 0.98, and no gradient clipping.

ZO-OGVR. We set $r_0 := 2.0$, $S_r := 8 \cdot 10^5$, and $r_{\min} := 0.5$, and use constant step 0.16. We use batch size 100, initial baseline batch 100, history window 10, distance weight 0, and no gradient clipping.

GZO-NS. We set $r_0 := 4.0$, $S_r := 5 \cdot 10^4$, and $r_{\min} := 1.0$, and use constant step 0.40. We set $\alpha := 0.9$ and use total batch 100, split into 20 known-gradient trajectories and 40 trajectories for each zeroth-order side. The gradient clip is 1.0.

GZO-HS. We set $r_k := 2.0$. The step schedule has initial value 0.40, scale $2.5 \cdot 10^4$, and minimum 0.04. We set $\alpha := 0.9$, history window 1, use 50 trajectories per zeroth-order side, and disable gradient clipping.

Gaussian PISO. We set $r_k := 1.50$, with $\mu_k := 1.50/\sqrt{d}$. For the step schedule, $\eta_0 := 1.80$, $S_\eta := 10^5$, and $\eta_{\min} := 1.00$. We set $\gamma := 0.75$, $\rho := 0.997$, $\rho_K := 0.90$, and $f_K := 0.50$, with total batch 100 and one direction per update.

Cycle PISO. The parameters are the same as for Gaussian PISO, with cycle length $L := 24$.

Gaussian PISO^2. The r_k and η_k schedules are the same as for Gaussian PISO. We use $\gamma := 1.00$, $\rho := 0.997$, $\rho_K := 0.90$, $\lambda := 0.01$, and $f_K := 0.50$, with total batch 100 and one direction per update.

Cycle PISO^2. The parameters are the same as for Gaussian PISO^2, with cycle length $L := 16$.

PePG. The step schedule has initial value 0.5, scale 10^5, and minimum 0.1. We use batch size 100, gradient clip 0.25, and transition-probability floor 10^{-12}. PePG uses exact reward- and transition-response derivatives and exact tabular advantages, while estimating the required expectations from trajectories.

We select the zeroth-order baseline hyperparameters using a two-stage grid search. In the first stage, for ZO-TG we search over $r_0 \in \{1, 2, 4\}$, $\eta_0 \in \{0.10, 0.20, 0.40\}$, and gradient clip in $\{0.25, 0.50, 1.00\}$; for ZO-OG over $r_0 \in \{0.25, 0.50, 1.00\}$, $\eta_0 \in \{0.04, 0.08, 0.16\}$, baseline decay in $\{0.90, 0.98, 0.995\}$, and gradient clip in $\{\text{none}, 0.25, 0.50\}$; for ZO-OGVR over $r_0 \in \{1, 2, 4\}$, $\eta_0 \in \{0.08, 0.16, 0.32\}$, $M \in \{0, 0.1, 1\}$, and gradient clip in $\{\text{none}, 0.25, 0.50\}$; for GZO-NS over $r_0 \in \{2, 4, 8\}$, $\eta_0 \in \{0.20, 0.40, 0.80\}$, known batch in $\{20, 34, 50\}$, and gradient clip in $\{0.25, 0.50, 1.00\}$; and for GZO-HS over $r_0 \in \{1, 2, 4\}$, $\eta_0 \in \{0.20, 0.40, 0.80\}$, historical window in $\{1, 5, 10\}$, and gradient clip in $\{\text{none}, 0.25, 0.50\}$. In the second stage, we search over radius schedule in $\{\text{constant}, \text{inverse-sqrt}\}$, $S_r \in \{5 \cdot 10^4, 2 \cdot 10^5, 8 \cdot 10^5\}$, $r_{\min} \in \{0.5, 1, 3\}$, step schedule in $\{\text{constant}, \text{inverse-sqrt}\}$, $S_\eta \in \{2.5 \cdot 10^4, 10^5, 4 \cdot 10^5\}$, and $\eta_{\min} \in \{0.04, 0.10, 0.20\}$; for GZO-NS and GZO-HS we additionally search over $\alpha \in \{0.75, 0.90, 1.00\}$.

We select the PISO hyperparameters using a two-stage grid search. In the first stage, we search over $r_0 \in \{1.00, 1.50, 2.00\}$, $\eta_0 \in \{1.20, 1.80, 2.40\}$, $\gamma \in \{0.75, 1.00, 1.25\}$, $\rho \in \{0.99, 0.995, 0.997\}$, $\rho_K \in \{0.00, 0.50, 0.90\}$, and $f_K \in \{0.34, 0.50, 0.67\}$. Gaussian PISO and Cycle PISO share these parameters, as do Gaussian PISO^2 and Cycle PISO^2. In the second stage, we search over $S_\eta \in \{5 \cdot 10^4, 10^5, 2 \cdot 10^5\}$, $\eta_{\min} \in \{0.50, 1.00, 1.50\}$, the cycle length $L \in \{16, 24, 32\}$ for Cycle PISO and Cycle PISO^2, and the mixing weight $\lambda \in \{0.01, 0.03, 0.10\}$ for Gaussian PISO^2 and Cycle PISO^2.

L.2.4 BILEVEL ROUTING GAME

For every method with dimension scaling, the base step is divided by the problem dimension d.

ZOS. We set $r_0 := 0.50$ and $q_r := 1$ and use constant step $0.70/d$ with a unit-sphere two-point estimator.

PZOS. We use $r_0 := 0.50$ and $q_r := 1$ and constant step $0.70/d$. The update combines the exact leader partial gradient with a two-point estimate of the follower-response contribution.

GZO-NS. We set $r_0 := 0.50$ and $q_r := 1$ and use constant step $0.70/d$, with $\alpha_0 := 0$, damping 0.98, and history window 5. Update clipping is disabled.

GZO-HS. The parameters are the same as for GZO-NS, using the historical-guide variant.

Gaussian PISO. We set $r_0 := 0.50$, $q_r := 1$, $\eta_0 := 0.70/d$, and $q_\eta := 1$, with $\gamma := 1.00$ and $\rho := 0.98$.

Cycle PISO. The parameters are the same as for Gaussian PISO, with cycle length $L := 10$.

Gaussian PISO^2. The r_k and η_k schedules are the same as for Gaussian PISO. We use $\gamma := 1.00$, $\rho := 0.98$, and $\lambda := 0.10$.

Cycle PISO^2. The parameters are the same as for Gaussian PISO^2, with cycle length $L := 10$.

We select the PISO hyperparameters using a two-stage grid search. In the first stage, we search over $r_0 \in \{0.25, 0.50, 1.00\}$, $a \in \{0.35, 0.70, 1.40\}$ with $\eta_0 = a/d$, and the residual momentum $\rho \in \{0.97, 0.98, 0.99\}$ for all four variants. In the second stage, we search over the cycle length $L \in \{5, 10, 20\}$, with effective cycle length $\min\{L, d\}$, for Cycle PISO and Cycle PISO^2, and the mixing weight $\lambda \in \{0.03, 0.10, 0.30\}$ for Gaussian PISO^2 and Cycle PISO^2.

L.2.5 BILEVEL SECURITY GAME

All step sizes use the inverse-square-root iteration schedule $\eta_k = \eta_0/\sqrt{k+1}$.

Residual norm cap. For security, we apply a norm cap to the residual observation and its momentum state as a numerical safeguard against occasional value explosions induced by unusually large observations. Let g_k^{K} denote the unweighted known gradient, and define

$$B_k^{\mathrm{cap}} := c_{\mathrm{cap}} \max\{\|g_k^{\mathrm{K}}\|_2, a_K\}, \qquad c_{\mathrm{cap}} > 0, \quad a_K > 0.$$

The norm-cap ratio c_{cap} is the multiplier of this reference norm. The known-norm floor a_K is a lower bound on the reference scale, so the cap remains $c_{\mathrm{cap}} a_K$ when the known-gradient norm is small. It does not impose a lower bound on either gradient or residual norm. For $B > 0$, let

$$\mathcal{C}_B(z) := \begin{cases} z, & \|z\|_2 \le B, \\ \dfrac{B}{\|z\|_2} z, & \|z\|_2 > B. \end{cases}$$

Writing the raw residual observation and raw momentum update as

$$R_k^{\mathrm{raw}} := m_k + \big(D_k - \langle \gamma g_k^{\mathrm{K}} + m_k, u_k \rangle\big) s_k, \qquad \widetilde{m}_{k+1} := \rho m_k + (1-\rho) R_k^{\mathrm{raw}},$$

the implementation sets

$$m_{k+1} := \mathcal{C}_{B_k^{\mathrm{cap}}}(\widetilde{m}_{k+1}), \qquad \overline{R}_k := \mathcal{C}_{B_k^{\mathrm{cap}}}(R_k^{\mathrm{raw}}), \qquad G_k := \gamma g_k^{\mathrm{K}} + (1-\lambda) m_{k+1} + \lambda \overline{R}_k.$$

Here $\lambda = 0$ for PISO and $\lambda > 0$ for PISO^2. Both vectors are clipped separately, after forming the raw momentum update and before forming the readout. Consequently, $\|(1-\lambda) m_{k+1} + \lambda \overline{R}_k\|_2 \le B_k^{\mathrm{cap}}$ for $\lambda \in [0, 1]$. The cap uses $\|g_k^{\mathrm{K}}\|_2$, without the factor γ. The security experiment uses $c_{\mathrm{cap}} = a_K = 1$, giving $B_k^{\mathrm{cap}} = \max\{\|g_k^{\mathrm{K}}\|_2, 1\}$.

ZOS. We set $r_0 := 0.10$ and $q_r := 1$. The step base is 0.05 for $d \le 100$ and 0.03 for $d > 100$. The estimator is unit-sphere two-point.

PZOS. We use the same r_k and step schedule as ZOS, combining the exact leader partial gradient with a finite-difference estimate of the follower-response contribution.

GZO-NS. We set $r_0 := 0.10$ and $q_r := 1$ and use the same $0.05/0.03$ step bases as ZOS, with $\alpha_0 := 0$, damping 0.98, and history window 5. Update clipping is disabled.

GZO-HS. The parameters are the same as for GZO-NS, using the historical-guide variant.

Gaussian PISO. We set $r_0 := 0.10$ and $q_r := 1$. We set $\eta_0 := 0.06$ for $d \le 100$ and $\eta_0 := 0.04$ for $d > 100$. We use $\gamma := 1.00$ and $\rho := 0.98$.

Cycle PISO. The parameters are the same as for Gaussian PISO, with cycle length $L := 10$.

Gaussian PISO^2. The r_k and η_k schedules are the same as for Gaussian PISO. We use $\gamma := 1.00$, $\rho := 0.98$, and $\lambda := 0.10$.

Cycle PISO^2. The parameters are the same as for Gaussian PISO^2, with cycle length $L := 10$.

We select the PISO hyperparameters using a two-stage grid search. In the first stage, we search over $r_0 \in \{0.05, 0.10, 0.20\}$, the paired $(a_{d\le 100}, a_{d>100}) \in \{(0.03, 0.02), (0.06, 0.04), (0.12, 0.08)\}$, and the residual momentum $\rho \in \{0.97, 0.98, 0.99\}$ for all four variants. In the second stage, we search over the cycle length $L \in \{5, 10, 20\}$, with effective cycle length $\min\{L, d\}$, for Cycle PISO and Cycle PISO^2, and the mixing weight $\lambda \in \{0.03, 0.10, 0.30\}$ for Gaussian PISO^2 and Cycle PISO^2.

L.2.6 PRICING ABLATIONS COMMON SETTINGS

The paper profile uses 100 simulations, 5000 training samples, 1000 metric samples, and 101 reporting-grid points. The default direction is Gaussian; Cycle panels replace it by the orthogonal cycle. Both PISO and PISO^2 use central differences with independent coupling, $r_0 := 0.40$, $q_r := 0.96$, $\gamma := 2.00$, $\rho := 0.92$, $f_K := 0.50$, an initial batch of 30 increasing by 2 per update, one direction per step, and cycle length 10. For PISO, $\eta_0 := 0.005$ and $q_\eta := 0.98$; for PISO^2, $\eta_0 := 0.005$ and $q_\eta := 1$, and PISO^2 additionally uses $\lambda := 0.25$.