Lower Bounds for Randomized Linear-Oracle Optimization over Strongly Convex Sets

Anonymous authors
Paper under double-blind review

Abstract

Recent results by Halbey et al. (2026a); Grimmer & Liu (2026) have established worst-case objective-error lower bounds of order $1/T^2$ after T linear minimization oracle (LMO) calls, matching the upper bound of Garber & Hazan (2015) in terms of dependence on T and by doing so, resolving a long-standing open question. However, upon close inspection both lower bounds are somewhat brittle: Halbey et al. (2026a) only applies to the standard Frank–Wolfe algorithm, and its prescribed slow exact-line-search trajectories can amplify sufficiently small initialization errors exponentially, corresponding to roughly one additional bit of initialization accuracy per iteration, as we show; Grimmer & Liu (2026) on the other hand obtain lower bounds for any deterministic LMO-based first-order method via a resisting oracle which adversarially adapts to the queries of the algorithm. In both cases it is conceivable that randomization may circumvent these lower bounds and in particular neither establishes a lower bound for unrestricted randomized first-order LMO methods. We close this gap by proving an $\Omega(L\,\mathrm{diam}(S)^2/T^2)$ distributional lower bound for randomized first-order linear minimization oracle methods over strongly convex sets. While our construction reuses the family of Grimmer & Liu (2026), our proof characterizes the conditional distribution of the hidden permutation after adaptive queries and bounds the expected decrease of a combinatorial statistic controlling optimizer variance. We also show that the same error threshold holds with probability at least $3/4$ and yields a stopping-time tail bound.

1 Introduction

We are interested in solving the optimization problem

$$\min_{x\in S} f(x),$$

where f is an L-smooth convex function and S is a compact convex set. The Frank–Wolfe algorithm (Frank & Wolfe, 1956; Levitin & Polyak, 1966) replaces potentially expensive projections by a linear minimization oracle (LMO), $\texttt{LMO}_S(p) \in \arg\min_{z\in S}\langle p, z\rangle$ and its general $O(1/T)$ rate is optimal in the linear-oracle model Jaggi (2013); Lan (2013). However, once we add more structure to both the function and the domain, convergence rates can often be improved; see (Braun et al., 2025) for an overview. In particular, in the case of strongly convex functions and strongly convex sets, a rate of $O(1/T^2)$ was established in Garber & Hazan (2015) and similarly improved convergence rates can be obtained for sharp functions (Kerdreux et al., 2022), uniformly convex sets (Kerdreux et al., 2021; 2026), and even if the objective is just convex but the set satisfies a mild curvature condition, $o(1/T)$ rates can be obtained (Pokutta, 2026). We will be concerned with *distributional* lower bounds for *randomized* LMO-based first-order methods for the case of strongly convex functions over strongly convex sets.

Related work

Most closely related to our work are Halbey et al. (2026a) (and their extension to uniformly convex sets and sharp functions (Halbey et al., 2026b)) as well as Grimmer & Liu (2026); we summarize the differences in Table 1.

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.

Table 1: Comparison of the lower bounds. The rate is objective error after T oracle calls.

Result	Method class	Randomization covered?	Data chosen	Rate
(Halbey et al., 2026a)	Exact-line-search or short-step Frank–Wolfe	Not an unrestricted-method result	Horizon-dependent initialization	T^{-2}
(Grimmer & Liu, 2026), span	LMO-span methods	Yes, for every fixed seed	One fixed ordered instance	T^{-2}
(Grimmer & Liu, 2026), resisting	Unrestricted first-order LMO	Deterministic only	Permutation completed from the transcript	T^{-2}
This paper	Unrestricted first-order LMO with feasible final output on every instance and seed	Yes	Uniform prior fixed before the seed	T^{-2}

Halbey et al. (2026a) give slowly converging trajectories at a rate of $\Omega(1/t^2)$ for exact-line-search or short-step Frank–Wolfe, with a horizon-dependent prescribed initialization. Their result concerns those specific updates, not arbitrary LMO methods. Moreover, as we show in Appendix B.4, tiny changes in the starting point can be strongly amplified along the slow exact-line-search trajectories. Over a fixed horizon of T iterations, sufficiently small initialization errors can be magnified by roughly 2^T, up to polynomial factors. Offsetting this local amplification corresponds to roughly one additional binary digit of initialization accuracy per iteration. This describes sensitivity near a prescribed trajectory; it does not mean that every perturbation or rounding error makes Frank–Wolfe converge faster. We also establish the existence of *infinite* slow-envelope trajectories on their unit-disk instance and show that the initializations with this eventual behavior form an uncountable set of planar Lebesgue measure zero (Theorem B.4); the stated guarantee in Halbey et al. (2026a) is finite-horizon, with an initialization that may depend on the horizon.

Complementing the above, for any *deterministic* LMO-based method returning feasible outputs, Grimmer & Liu (2026) establish an $\Omega(L\,\mathrm{diam}(S)^2/(T+1)^2)$ objective-error lower bound after T oracle calls. Their resisting oracle completes a hard instance *after* observing the algorithm's queries. Applying this argument separately to each random seed may produce a different hard instance for each seed. A randomized lower bound instead requires an instance fixed independently of the realized seed, on which the algorithm has large expected error. On the other hand Grimmer & Liu (2026) also give one fixed instance hard for every LMO-span method, which is a subclass of all LMO-based methods. This result immediately gives rise to a lower bound for randomized methods but *only* those from the restricted class of LMO-span methods, which is a significant restriction as we will discuss in Appendix B.3.

Contribution

We show that randomization cannot improve the corresponding order of convergence, even without restrictions on query directions or intermediate evaluation points. We do this by providing a hard distribution of instances and establishing a corresponding distributional lower bound that then by Yao's principle (Yao, 1977) provides a randomized lower bound and we obtain:

$$\mathbb{E}[f(\hat{x}_T) - f^\star] = \Omega\left(\frac{L\,\mathrm{diam}(S)^2}{(T+1)^2}\right). \tag{1}$$

We follow the general information-theoretic approach from Braun & Pokutta (2015); Braun et al. (2015), where we bound the (expected) information gain in each iteration albeit we do not work with information-theoretic quantities such as mutual information and entropy but rather use a combinatorial statistic Δ through which we lower-bound the conditional optimizer variance.

We retain the feasible sets, common quadratic objective, optimizer and thresholding formula of Grimmer & Liu (2026), and their conversion of optimizer variance into loss. The departure is the *fixed uniform prior* over their feasible sets, which turns out to be our hard distribution. In particular, in contrast to the resisting oracle setup, we can no longer respond adversarially to the queries of the algorithm. As a consequence an answer can now potentially reveal significantly more information than in the resisting oracle setup.

The hidden instance is determined by a permutation of a known list of d coordinate weights. Let Tr_t record the algorithm's seed and its first t queries and answers, and let $m_t \le d$ count the coordinate–weight assignments not yet explicitly disclosed by a nonzero LMO component. The permutations compatible with this transcript form a graph under exchanges of two coordinate weights. We show that this graph has a common outdegree Δ_t, the number of compatible exchanges from any one of its vertices. Our key estimate bounds its expected decrease in one query:

$$\mathbb{E}[\Delta_t - \Delta_{t+1} \mid \mathrm{Tr}_t] \le 6m_t. \tag{2}$$

Initially $m_0 = d$ and $\Delta_0 = \binom{d}{2}$. Since $m_t \le d$, summing the expected decreases gives

$$\mathbb{E}\Delta_T \ge \binom{d}{2} - 6dT.$$

Thus the initial potential is quadratic in d, whereas the expected decrease per query is at most linear in d. Choosing d to be a sufficiently large constant multiple of $T + 1$ leaves substantial conditional optimizer variance, which we convert into optimization error. A single query may disclose several coordinate weights; our argument bounds expected progress rather than imposing a one-coordinate-per-query rule. We also do not require a span restriction. Section 2 collects the results and their short derivation; Sections 3 and 4 identify the conditional distribution and its variance; Section 5 gives the full $6m$ argument.

2 Model, family, and lower bounds

2.1 Oracle model

For a compact convex set $S \subset \mathbb{R}^d$ and differentiable $f : \mathbb{R}^d \to \mathbb{R}$, we consider

$$f^\star \doteq \min_{x \in S} f(x). \tag{3}$$

The function is L-smooth when its gradient is L-Lipschitz, and μ-strongly convex when $f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle + \mu\|y - x\|_2^2/2$ for all x, y.

Definition 2.1 (Strongly convex set)**.** A convex set S is α-strongly convex if, for $x, y \in S$ and $\theta \in [0, 1]$,

$$\theta x + (1 - \theta)y + B\Big(0, \frac{\alpha}{2}\theta(1 - \theta)\|x - y\|_2^2\Big) \subseteq S,$$

where $B(0, r)$ is the closed Euclidean ball.

A *query* consists of any $(x_t, p_t) \in \mathbb{R}^d \times \mathbb{R}^d$ and returns $y_t \doteq (f(x_t), \nabla f(x_t), z_{t+1})$, where $z_{t+1} \doteq \mathtt{LMO}_S(p_t)$. Let U be any *random seed* independent of the hidden instance, and put

$$\mathrm{Tr}_t \doteq (U, (x_s, p_s, y_s)_{s<t}), \qquad \mathscr{H}_t \doteq \sigma(\mathrm{Tr}_t). \tag{4}$$

Here Tr_t is the *transcript*, including the seed, and $\mathscr{H}_t$ is its generated σ-algebra: all information available before query t; conditioning on Tr_t means conditioning on $\mathscr{H}_t$. Fixing a realization $U = u$ makes the algorithm's decision rules deterministic; the transcript may still vary with the hidden instance. Queries, stopping decisions and outputs are arbitrary measurable functions of this history. Our argument is purely information-theoretic with minimal assumptions on the randomized algorithm in question beyond first-order and LMO access. In particular, we do not assume any span, atom-hull, computational, or finite-seed restriction. Evaluation points x_t need not belong to S: the function and gradient are defined on all of $\mathbb{R}^d$. Only the returned output must be feasible, on *every instance and every seed*, not merely almost surely. Ordinary Frank–Wolfe and its feasible variants meet this requirement. The model provides only the oracle responses specified above; an exact-line-search oracle for an unknown objective would be additional access. In our hard family, however, the objective is the same known quadratic for every permutation, with all its parameters independent of the hidden instance. Thus, evaluating this formula, or computing an exact line search from it, requires no additional oracle access. Only the feasible set is hidden in this family. We set $\mathtt{LMO}_{S_\pi}(0) \doteq 0$; unused calls can be padded with zero queries.

2.2 The permuted hard family

The following construction, including its normalization, is due to Grimmer & Liu (2026, Sections 2.1 and 3.1). Fix $d \geq 3$ and define

$$C \doteq \frac{1}{\sqrt{d^2 + d + 2}}, \qquad w_k \doteq C\sqrt{2k}, \quad k = 1, \dots, d. \tag{5}$$

For a permutation π of $\{1, \dots, d\}$, let

$$S_\pi \doteq \left\{ x \in \mathbb{R}^d : \frac{1}{2}\|x\|_2^2 + \sum_{i=1}^{d} w_{\pi(i)}|x_i| \leq C^2 \right\}. \tag{6}$$

Coordinate i carries weight $w_{\pi(i)}$; the integer $\pi(i)$ is its *weight rank*. We have $2C^2 + \sum_k w_k^2 = 1$.

Lemma 2.2 (Geometry of S_π)**.** *Every S_π is 1-strongly convex and* $\operatorname{diam}(S_\pi) = 2C(2 - \sqrt{2})$.

See the strong-convexity and diameter lemmas of Grimmer & Liu (2026, Section 2.1). Next we pick $\rho \in (0, 1)$ to be the unique solution of the boundary equation

$$\sum_{k=1}^{d} \left[\frac{1}{2}\rho^2 (w_d - w_k)^2 + \rho w_k (w_d - w_k) \right] = C^2, \tag{7}$$

whose existence and uniqueness are proved in Grimmer & Liu (2026, Section 3.1). Define

$$M \doteq \frac{\rho}{1 - \rho} w_d, \qquad f(x) \doteq \frac{1}{2}\|x - M\mathbf{1}\|_2^2. \tag{8}$$

This objective is 1-smooth, 1-strongly convex, and used for all instances. Moreover, for every instance π, the choice of ρ in (7) places the constrained minimizer on the boundary of S_π.

Lemma 2.3 (Optimal points)**.** *The minimizer of f over S_π is unique, with*

$$(x_\star^{(\pi)})_i = \rho(w_d - w_{\pi(i)}) \qquad \text{and} \qquad \rho \geq \frac{2}{d^2}. \tag{9}$$

See the permuted-minimizer lemma of Grimmer & Liu (2026, Section 3.1).

Remark 2.4 (Common optimal value). Coordinate permutations preserve the common quadratic objective. Although the optimal points vary with π, all instances have the same optimal value

$$f^\star = \frac{1}{2} \sum_{k=1}^{d} \left(\rho(w_d - w_k) - M \right)^2.$$

For a fixed terminal transcript, including the seed, the output x is fixed: the output rule is a deterministic function of that transcript. Its loss $f(x) - f^\star$ is consequently constant over all compatible instances, so its conditional expected loss equals its actual loss on each completion.

For admissible deterministic methods A with budget T, measurable mixtures ν, and objective-gap loss $\mathscr{L}$, put

$$V_{\mathrm{dist}}(\mu) \doteq \inf_A \mathbb{E}_{I \sim \mu} \mathscr{L}(A, I) \qquad \text{and} \qquad V_{\mathrm{rand}} \doteq \inf_\nu \sup_I \mathbb{E}_{A \sim \nu} \mathscr{L}(A, I).$$

The method A may know μ. Yao's inequality $V_{\mathrm{dist}}(\mu) \leq V_{\mathrm{rand}}$ makes a fixed-prior bound sufficient. Appendix B.1 gives the full minimax formulation and proof.

Proposition 2.5 (Uniform prior)**.** *For the permuted family,*

$$V_{\mathrm{dist}}(\mathrm{Unif}) = V_{\mathrm{rand}}. \tag{10}$$

Proof of Proposition 2.5. See Appendix A.1.

Lemma 2.6 (Rescaling)**.** *For $L, \alpha > 0$, the rescaling $\tilde{S} \doteq \alpha^{-1} S$ and $\tilde{f}(x) \doteq \frac{L}{\alpha^2} f(\alpha x)$ takes a 1-smooth, 1-strongly convex f and 1-strongly convex S to an L-smooth, L-strongly convex objective and an α-strongly convex set. Diameter D and gap g become D/α and Lg/α^2; LMO transcripts correspond under $x \mapsto \alpha x$.*

See the rescaling lemma of (Grimmer & Liu, 2026, Section 2). For completeness we include a proof due to the additional measurable maps that need to be preserved.

Proof of Lemma 2.6. See Appendix A.2.

2.3 Derivation from the two main estimates

In the later sections we will construct a combinatorial statistic $\Delta_t \geq 0$ from a given transcript, with $\Delta_0 = d(d-1)/2$ and $\Delta_{t+1} \leq \Delta_t$. For this quantity we will establish two key estimates

$$\mathbb{E}[\Delta_t - \Delta_{t+1} \mid \mathscr{H}_t] \leq 6d \qquad \text{and} \qquad f(\hat{x}_T) - f^\star \geq \frac{\rho^2 C^2}{48 d^2} \Delta_T^2. \tag{11}$$

The first is Corollary 5.4, using $m_t \leq d$. It bounds the expected decrease of our combinatorial uncertainty statistic, averaging over hidden permutations compatible with the current transcript; the seed and past observations are already fixed. The second follows from Corollary 4.2 and the common optimal value in Remark 2.4. It converts the lower bound on conditional optimizer variance measured by Δ_T into a lower bound on the primal gap; it is trivial when no coordinates remain unresolved. Our main work will be to bound the expected decrease.

2.4 Main results

In the considered model here the method knows the common objective, the scalar parameters and the uniform prior. Only the permutation, sampled independently of the seed, is hidden.

Theorem 2.7 (Distributional lower bound)**.** *For every integer $T \geq 1$, let $d \doteq 64(T+1)$ and let Π be uniform on the $d!$ sets (6). Every randomized first-order LMO method using at most T calls and returning a feasible point on every instance and every seed satisfies*

$$\mathbb{E}_{\Pi,U}[f(\hat{x}_T) - f(x_\star^{(\Pi)})] \geq c \, \frac{\operatorname{diam}(S_\Pi)^2}{(T+1)^2}, \qquad c \doteq (3 + 2\sqrt{2})/(6144 \cdot 64^2). \tag{12}$$

For every $L, \alpha > 0$, positive rescaling gives the right side $cL \operatorname{diam}(S)^2/(T+1)^2$, an L-smooth, L-strongly convex objective and an α-strongly convex feasible set. Some fixed permutation has at least this expected loss over the seed alone.

Proof of Theorem 2.7. See Appendix A.3.

Corollary 2.8 (Query complexity)**.** *There are instances requiring $T = \Omega((L \operatorname{diam}(S)^2/\varepsilon)^{1/2})$ queries to guarantee expected gap at most ε, under the same universal final-feasibility requirement.*

Proof. Rearrange (12): $(T+1)^2 \geq cL \operatorname{diam}(S)^2/\varepsilon$ is necessary. □

Theorem 2.9 (Constant-probability lower bound)**.** *Put $c_0 \doteq \frac{3+2\sqrt{2}}{6144 \cdot 64^2}$. Under the dimension, prior, rescaling and method assumptions of Theorem 2.7,*

$$\mathbb{P}_{\Pi,U}\left[f(\hat{x}_T) - f(x_\star^{(\Pi)}) \geq c_0 \frac{L \operatorname{diam}(S)^2}{(T+1)^2} \right] \geq \frac{3}{4}. \tag{13}$$

Some fixed permutation has probability at least $\frac{3}{4}$ over the seed alone.

Proof of Theorem 2.9. See Appendix A.4.

Corollary 2.10 (Stopping-time tail)**.** *Let N count calls up to the first return, with $N \doteq \infty$ on nontermination. Every returned output must be feasible on every instance and every seed. If, on each instance, the probability of returning an output whose gap is at least the threshold in Theorem 2.9 is at most $\eta < \frac{3}{4}$, then some fixed instance satisfies $\mathbb{P}[N > T] \geq \frac{3}{4} - \eta$.*

Proof. Truncate at T, returning the feasible point 0 if the method has not stopped. The truncated failure event is contained in the union of the original returned-failure event and $\{N > T\}$. Apply Theorem 2.9; nontermination already lies in the tail. □

The fixed permuation witnesses for the expectation in Theorem 2.7 and the probability in Theorem 2.9 may differ; neither depends on the realized seed though. The diameter in our bound is that of the constructed instance. In this family, diameter and set curvature satisfy $\alpha \operatorname{diam}(S) = \Theta(1/d)$, so rescaling cannot adjust them independently. We therefore do not claim optimal dependence on all geometric parameters simultaneously.

3 Exact LMO transcripts under a uniform permutation

For the considered family of instances, the exact thresholding formula captures all information revealed in an LMO answer, including the inequalities implied by its zero coordinates. With this we can describe the entire conditional distribution directly.

3.1 The exact oracle answer

Recall the convention $\texttt{LMO}_{S_\pi}(0) = 0$. A zero query reveals no instance-dependent information.

We recall the exact LMO formula of (Grimmer & Liu, 2026, Sections 2.2 and 3.1) and derive an explicit characterization of the information contained in its answer.

Lemma 3.1 (Exact LMO and the information in its answer)**.** *Fix $p \neq 0$ and let $z \doteq \texttt{LMO}_{S_\pi}(p)$. There is a unique $\lambda > 0$ such that*

$$z_i = -\operatorname{sign}(p_i) \max\left\{0, \frac{|p_i|}{\lambda} - w_{\pi(i)}\right\}. \tag{14}$$

Define the threshold parameter τ *and* active set A *by*

$$\tau \doteq \frac{1}{2C^2\lambda^2}, \qquad A \doteq \{i : \tau p_i^2 > \pi(i)\}. \tag{15}$$

Coordinates in A are active*; the others are* inactive*. Then $A = \{i : z_i \neq 0\}$ and*

$$\sum_{i=1}^{d} (\tau p_i^2 - \pi(i))_+ = 1. \tag{16}$$

The pair (p, z) determines λ, τ, every active rank $\pi(i)$, and the inactive lower bounds

$$\pi(i) \geq \lceil \tau p_i^2 \rceil \qquad (i \notin A). \tag{17}$$

Conversely, the active-rank equalities and inactive-rank lower bounds determine the LMO answer.

We can recover λ and π explicitly from the observed vectors:

$$\lambda = \frac{\sum_i |p_i z_i|}{C^2 + \|z\|_2^2/2}, \qquad \pi(i) = \frac{(|p_i|/\lambda - |z_i|)^2}{2C^2} \quad (z_i \neq 0). \tag{18}$$

The coordinate label i is already known; its assigned weight rank is learned. Every nonzero component supplies an equality, not just one per query; some equalities may repeat previously revealed ranks.

Proof of Lemma 3.1. See Appendix A.5.

3.2 Conditional distributions, Remaining ranks, and compatible assignments

Let Π be uniform over all $d!$ permutations, selected before the method starts. Queries may depend on the complete previous transcript.

Lemma 3.2 (Form of every conditional posterior)**.** *Fix a seed value $U = u$, thereby fixing a deterministic method, and a transcript Tr_t of positive probability under the uniform instance prior for that fixed method. Conditional on this transcript, Π is uniform over all permutations satisfying:*

(i) exact assignments $\Pi(i) = k$ for coordinates active in an earlier LMO answer;

(ii) lower bounds $\Pi(i) \geq b_i$ from coordinates inactive in earlier answers.

There are no additional restrictions in the numerical LMO answers.

Proof of Lemma 3.2. See Appendix A.6.

Fix the seed and a realized transcript. Let K_t be the coordinates whose ranks appeared in nonzero LMO components. From the query and its answer, every nonzero component z_i reveals $w_{\Pi(i)}$, and hence $\Pi(i)$. Let $I_t \doteq [d]\backslash K_t$. Write $m \doteq |I_t|$ and list the unused *original weight ranks* as $r_1 < \cdots < r_m$.

Here j denotes a position in the remaining list, while r_j denotes the corresponding original rank. The remaining weights are therefore $w_{r_1}, \dots, w_{r_m}$. This lets us work with a shorter index set without changing the original ranks or weights. For $v \in I_t$, let b_v be its largest observed lower bound (initially 1). The first remaining rank at least b_v has index $L_v \doteq \min\{j \in [m] : r_j \geq b_v\}$. Realizability ensures the minimum exists. Let $\imath : [m] \to I_t$ list coordinates in increasing order of (L_v, v), so ties retain the original coordinate order, and put $\ell_i \doteq L_{\imath(i)}$. A *reduced permutation* σ means precisely $\Pi(\imath(i)) = r_{\sigma(i)}$. Thus representing the permutations as permutation matrices, $v \doteq \imath(i)$ is a *coordinate* (not be be confused with a list index i here), i its row number, $j \doteq \sigma(i)$ a column number, and r_j the original integer weight rank. Query components below are relabeled by the same map: p_i in a reduced formula means the full query component $p_{\imath(i)}$. See Appendix C for an example.

Conditional on the full transcript, including the seed, the induced reduced permutation Σ is uniformly distributed over the compatible assignments

$$\mathscr{F}(\ell) \doteq \{\sigma \in \mathfrak{S}_m : \sigma(i) \geq \ell_i \text{ for all } i\}. \tag{19}$$

A permutation is represented by one occupied cell in each row and column. Row i permits columns $\ell_i, \dots, m$; these allowed sets are nested as the ℓ_i are increasing and form a Ferrers board. With $\ell_1 \leq \cdots \leq \ell_m$, nonemptiness is equivalent to $\ell_i \leq i$: if $\ell_i > i$, rows $i, \dots, m$ have too few columns; otherwise assign from row m down to row 1. Set $c_i \doteq i - \ell_i + 1$ then we have

Lemma 3.3 (Counting compatible assignments)**.** *For $\mathscr{F}(\ell) \neq \emptyset$ we have*

$$|\mathscr{F}(\ell)| = \prod_{i=1}^{m} c_i. \tag{20}$$

Proof of Lemma 3.3. See Appendix A.7.

4 Conditional variance and optimization error

The deterministic proof in (Grimmer & Liu, 2026, Section 3.3) averages over an unconstrained suffix of unassigned ranks and obtains a lower bound on loss via their optimizer variance. We use the same variance-to-loss argument, however our averaging has to be done more carefully due to the row-dependent restrictions from (19). We define the following estimate to account for them

$$\Delta(\ell) \doteq \sum_{i=1}^{m} (c_i - 1) = \frac{m(m+1)}{2} - \sum_{i=1}^{m} \ell_i. \tag{21}$$

This is $m(m-1)/2$ for the unrestricted permutation and 0 for a unique assignment, including the empty assignment.

Lemma 4.1 (Rank variance under nested lower bounds)**.** *Let $m \geq 1$ and Σ be uniform on $\mathscr{F}(\ell)$. Then*

$$\sum_i \mathrm{Var}[\Sigma(i)] \geq \frac{\Delta^2}{12m} \qquad \textit{and} \qquad \sum_i \mathrm{Var}[\sqrt{r_{\Sigma(i)}}] \geq \frac{\Delta^2}{48md}. \tag{22}$$

Proof of Lemma 4.1. See Appendix A.8.

Corollary 4.2 (Conditional optimization error)**.** *After fixing the seed, at a final transcript h with $m > 0$, any fixed vector x feasible on every compatible instance satisfies*

$$\mathbb{E}[f(x) - f(x_\star^{(\Pi)}) \mid h] \geq \frac{\rho^2 C^2 \Delta^2}{48md}. \tag{23}$$

Proof. Fix the final transcript h and a point x feasible for every instance consistent with h. Although x is fixed, the optimizer $X_\star \doteq x_\star^{(\Pi)}$ may differ between those instances. On each compatible instance, feasibility and constrained first-order optimality give $f(x) - f(X_\star) \geq \|x - X_\star\|_2^2/2$. Moreover,

$$\mathbb{E}[\|x - X_\star\|_2^2 \mid h] = \|x - \mathbb{E}[X_\star \mid h]\|_2^2 + \sum_v \mathrm{Var}[X_{\star,v} \mid h].$$

The first term is nonnegative. Resolved coordinates have zero variance; by (9) the remaining sum is $2\rho^2 C^2 \sum_i \mathrm{Var}[\sqrt{r_{\Sigma(i)}}]$. Apply (22) and the factor $1/2$ above. □

5 Expected decrease in one query

We bound the expected decrease of Δ in one LMO query by counting pairs of compatible assignments that differ by a single swap and produce different answers.

Fix the transcript and the next LMO query direction $p \in \mathbb{R}^d$; the oracle minimizes $\langle p, z\rangle$ over S_Π. If $m = 0$ or $p = 0$, Δ is unchanged; assume $m \geq 1$ and $p \neq 0$. The unresolved entries of this full vector may still all be zero. We build a graph as follows. Each vertex is a compatible assignment $\sigma \in \mathscr{F}(\ell)$, where ℓ is the bound vector *before* the new query's information is incorporated; in the following we use a superscript + to denote quantities after the query, in particular, $\Delta = \Delta(\ell)$ and $\Delta^+ \doteq \Delta(\ell^+)$. For a row i with column $s \doteq \sigma(i)$, choose $a \in \{\ell_i, \dots, s-1\}$ and put $k \doteq \sigma^{-1}(a)$. Swap these columns: $\sigma'(i) \doteq a$, $\sigma'(k) \doteq s$, leaving other rows unchanged. The swap is allowed: row i accepts a, and row k receives a larger column index. It exchanges original ranks r_a, r_s between the fixed physical coordinates $\imath(i), \imath(k)$. Each swap gives a directed edge $\sigma \to \sigma'$; both orientations are counted. Every vertex has the same number of outgoing edges:

$$\sum_i (\sigma(i) - \ell_i) = \frac{m(m+1)}{2} - \sum_i \ell_i = \Delta.$$

An *answer class* consists of the assignments producing the same numerical LMO answer. Applying the earlier indexing construction to the transcript extended by this answer represents the class as $\mathscr{F}(\ell^+)$.

A swap stays within an answer class exactly when it preserves its fixed assignments and new bounds, hence when it remains an allowed swap of the updated family. Each vertex has $\Delta^+ \leq \Delta$ outgoing edges within its class and $\Delta - \Delta^+$ leaving it. Since Σ is uniform on $\mathscr{F}(\ell)$,

$$\mathbb{E}[\Delta - \Delta^+] = \frac{\#\{\text{directed swap edges whose endpoints give different answers}\}}{|\mathscr{F}(\ell)|}. \tag{24}$$

We provide an example in Appendix C.1.

To bound this count, we use the threshold equation. The disclosed coordinates contribute

$$\psi(\tau) \doteq \sum_{v \in K_t} (\tau p_v^2 - \Pi(v))_+.$$

In the formulas for the unresolved row (which gives the bounds) we use p_i as an abbreviation for the full component $p_{\imath(i)}$; alternatively think of it as relabeling so that we can apply the same argument on a smaller set. For $\sigma \in \mathscr{F}(\ell)$, the threshold τ_σ is the unique solution to

$$\psi(\tau) + \sum_{i=1}^{m} (\tau p_i^2 - r_{\sigma(i)})_+ = 1. \tag{25}$$

The function ψ is known, nonnegative, nondecreasing and independent of σ. Below τ_σ, the left side and every summand are below 1. A swap exchanges two ranks and we first count swaps whose larger rank has a nonzero LMO component.

Lemma 5.1 (Active larger ranks)**.** *For the fixed query p and a compatible hidden assignment $\sigma \in \mathscr{F}(\ell)$, define*

$$A_\sigma \doteq \{i \in [m] : \tau_\sigma p_i^2 > r_{\sigma(i)}\}.$$

These are the pre-query unresolved rows whose ranks are disclosed by nonzero components of the new LMO answer. With Σ uniform on $\mathscr{F}(\ell)$, A_Σ denotes the corresponding random set, and

$$\mathbb{E} \sum_{i \in A_\Sigma} (\Sigma(i) - \ell_i) \leq m. \tag{26}$$

Proof. Since Σ is uniform, multiplying the expectation in (26) by $|\mathscr{F}(\ell)|$ counts triples (σ, i, a) with $i \in A_\sigma$ and $\ell_i \leq a < s \doteq \sigma(i)$. Map each triple to $(\sigma', s) \in \mathscr{F}(\ell) \times \{1, \dots, m\}$, where σ' is the compatible assignment obtained by swapping a, s. We show injectivity.

Suppose for contradiction distinct rows i, k give the same target (σ', s). Their columns $a_i \doteq \sigma'(i)$ and $a_k \doteq \sigma'(k)$ are distinct and below s. In the preimage where row i carries r_s, activity gives

$\tau_\sigma p_i^2 > r_s > 0$, hence $p_i \neq 0$; the other preimage similarly gives $p_k \neq 0$. At $\tau = r_s/p_i^2 < \tau_\sigma$, the left side of (25) is below 1, hence $\frac{p_k^2}{p_i^2} < \frac{r_{a_k}+1}{r_s}$. The other preimage gives $p_i^2/p_k^2 < (r_{a_i}+1)/r_s$. Multiplying these two ratio bounds gives

$$1 = \frac{p_k^2}{p_i^2}\frac{p_i^2}{p_k^2} < \frac{(r_{a_k}+1)(r_{a_i}+1)}{r_s^2} < 1,$$

where the last inequality uses that r_{a_i}, r_{a_k} are distinct integers below r_s. This contradiction proves the row is unique; $a = \sigma'(i)$ and the inverse swap recover σ. There are at most $m|\mathscr{F}(\ell)|$ targets. □

An inactive row i has $\tau_\sigma p_i^2 \leq r_{\sigma(i)}$. Its zero LMO component excludes a previously allowed smaller rank r_a exactly when

$$\ell_i \leq a < \sigma(i), \qquad r_a < \tau_\sigma p_i^2 \leq r_{\sigma(i)}. \tag{27}$$

Let $N(\sigma)$ count these pairs (i, a). The condition $\ell_i \leq a$ ensures that the rank was allowed before the answer.

Lemma 5.2 (Reduction to newly excluded pairs)**.** *For the count $N(\sigma)$ defined by (27),*

$$\mathbb{E}[\Delta - \Delta^+] \leq 2m + 2\mathbb{E}N(\Sigma). \tag{28}$$

Proof of Lemma 5.2. See Appendix A.9.

It remains to establish a linear bound on $\mathbb{E}N(\Sigma)$, which we will do via the threshold equation. The right-hand side of (25) is 1, and all terms are nonnegative. Thus for every unresolved row j with $p_j^2 > 0$,

$$\tau_\sigma \leq \frac{r_{\sigma(j)}+1}{p_j^2}. \tag{29}$$

Rows with $p_j = 0$ are omitted when minimizing the quantities on the right. The nonnegative known-coordinate contribution ψ does not change this inequality.

Lemma 5.3 (Linear bound on newly excluded pairs)**.** *For every conditional distribution of the form (19) and every query,*

$$\mathbb{E}N(\Sigma) \leq 2m.$$

Proof. Separate the pairs in (27) according to whether $r_a + 1 < \tau_\sigma p_i^2$. The pairs not satisfying this strict inequality have $\tau_\sigma p_i^2 - 1 \leq r_a < \tau_\sigma p_i^2$. This half-open interval contains at most one integer, so there is at most one such pair per row. This includes equality at the left endpoint and requires no generic-position hypothesis.

For a pair satisfying $r_a + 1 < \tau_\sigma p_i^2$, set $s \doteq \sigma(i) > a$ and transpose indices a, s to obtain σ'. The target is compatible because row i accepts a and the other row moves upward. Map the triple (σ, i, a) to (σ', s).

Here $p_i^2 > 0$ and $u \doteq \frac{r_a+1}{p_i^2} < \tau_\sigma$. By (29), every quantity $(r_{\sigma(j)}+1)/p_j^2$ with $p_j \neq 0$ is strictly greater than u. After the transposition, row i has value u. The value of the other swapped row increases if its query component is nonzero, and all remaining values are unchanged. Thus i is the unique minimizer of $(r_{\sigma'(j)}+1)/p_j^2$ over nonzero unresolved query components. The target σ' determines i and $a = \sigma'(i)$; with s, it recovers σ by transposing back. This proves injectivity without using any property of the new threshold $\tau_{\sigma'}$.

There are at most $m|\mathscr{F}(\ell)|$ targets, so the expected number of pairs satisfying the strict inequality is at most m. Adding the at most one remaining pair per row gives $2m$. If all unresolved query components are zero, no pairs satisfy (27); if $m = 0$, the count is also zero. The zero full query has already been treated by the oracle convention. □

Corollary 5.4 (One-query decrease)**.** *For every adaptive transcript,*

$$\mathbb{E}[\Delta - \Delta^+ \mid \mathscr{H}_t] \leq 6m. \tag{30}$$

AI use statement

Generative AI tools assisted throughout the research process with formulation and revision of mathematical arguments, implementation of verification and plotting scripts, formalization, and manuscript editing. The associated checks include symbolic calculations, numerical and adversarial tests, and machine-checked proofs for the results covered by the formalization. All claims have been independently verified by the authors.

References

Alekh Agarwal, Peter L. Bartlett, Pradeep Ravikumar, and Martin J. Wainwright. Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization. *IEEE Transactions on Information Theory*, 58(5):3235–3249, 2012. doi: 10.1109/TIT.2011.2182178.

Gábor Braun and Sebastian Pokutta. An information diffusion Fano inequality. *arXiv preprint arXiv:1504.05492*, 2015. doi: 10.48550/arXiv.1504.05492.

Gábor Braun, Sebastian Pokutta, and Yao Xie. Info-Greedy Sequential Adaptive Compressed Sensing. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):601–611, 2015.

Gábor Braun, Cristóbal Guzmán, and Sebastian Pokutta. Lower bounds on the oracle complexity of nonsmooth convex optimization via information theory. *IEEE Transactions on Information Theory*, 63(7):4709–4724, 2017. doi: 10.1109/TIT.2017.2701343.

Gábor Braun, Alejandro Carderera, Cyrille W. Combettes, Hamed Hassani, Amin Karbasi, Aryan Mokhtari, and Sebastian Pokutta. *Conditional Gradient Methods: From Core Principles to AI Applications*. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2025. ISBN 978-1-61197-855-1. doi: 10.1137/1.9781611978568.

Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1–2):95–110, 1956.

Dan Garber and Elad Hazan. Faster rates for the Frank–Wolfe method over strongly-convex sets. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pp. 541–549, 2015.

Benjamin Grimmer and Ning Liu. Lower bounds for linear minimization oracle methods optimizing over strongly convex sets. *arXiv preprint arXiv:2602.22608*, 2026. doi: 10.48550/arXiv.2602.22608.

Jannis Halbey, Daniel Deza, Max Zimmer, Christophe Roux, Bartolomeo Stellato, and Sebastian Pokutta. Lower bounds for Frank–Wolfe on strongly convex sets. *arXiv preprint arXiv:2602.04378*, 2026a. doi: 10.48550/arXiv.2602.04378.

Jannis Halbey, Christophe Roux, and Sebastian Pokutta. Curvature-dependent lower bounds for Frank–Wolfe. *arXiv preprint arXiv:2605.10595*, 2026b. doi: 10.48550/arXiv.2605.10595.

Elad Hazan. A linearly convergent projection-free algorithm for smooth convex sets. *arXiv preprint arXiv:2608.04321*, 2026. doi: 10.48550/arXiv.2608.04321.

Martin Jaggi. Revisiting Frank–Wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, pp. 427–435, 2013.

Thomas Kerdreux, Alexandre d'Aspremont, and Sebastian Pokutta. Projection-Free Optimization on Uniformly Convex Sets. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130 of *Proceedings of Machine Learning Research*, pp. 19–27. PMLR, 2021. URL https://proceedings.mlr.press/v130/kerdreux21a.html.

Thomas Kerdreux, Alexandre d'Aspremont, and Sebastian Pokutta. Restarting Frank-Wolfe: Faster Rates under Hölderian Error Bounds. *Journal of Optimization Theory and Applications*, 192: 799–829, 2022.

Thomas Kerdreux, Alexandre d'Aspremont, and Sebastian Pokutta. Local and Global Uniform Convexity Conditions. In *Data Science and Optimization*, volume 91 of *Fields Institute Communications*, pp. 85–120. Springer, 2026. doi: 10.1007/978-3-032-03844-9_4.

Guanghui Lan. The complexity of large-scale convex programming under a linear optimization oracle. *arXiv preprint arXiv:1309.5550*, 2013. doi: 10.48550/arXiv.1309.5550.

Evgeny S. Levitin and Boris T. Polyak. Constrained minimization methods. *U.S.S.R. Computational Mathematics and Mathematical Physics*, 6(5):1–50, 1966. doi: 10.1016/0041-5553(66)90114-5.

Anders Martinsson and Pascal Su. Mastermind with a linear number of queries. *Combinatorics, Probability and Computing*, 33(2):143–156, 2024. doi: 10.1017/S0963548323000366.

Arkadi S. Nemirovski and David B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, New York, 1983.

Sebastian Pokutta. Frank–Wolfe beyond $1/t$ convergence. *arXiv preprint arXiv:2604.28006*, 2026. doi: 10.48550/arXiv.2604.28006.

Maxim Raginsky and Alexander Rakhlin. Information complexity of black-box convex optimization: A new look via feedback information theory. In *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing*, pp. 803–810, 2009. doi: 10.1109/ALLERTON.2009.5394945.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958. doi: 10.2140/pjm.1958.8.171.

Andrew Chi-Chih Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science*, pp. 222–227, 1977. doi: 10.1109/SFCS.1977.24.

Appendix

A Proofs of the main results

A.1 Proof of Proposition 2.5

Proof. For a coordinate permutation g, let P_g be its permutation matrix. Relabeling sends S_π to $P_g S_\pi = S_{\pi \circ g^{-1}}$, and $P_g \mathbf{1} = \mathbf{1}$ makes the objective invariant. To run the conjugated method gA on S, simulate A on $P_g^{-1} S$: transform internal queries (x, p) to $(P_g x, P_g p)$ and transform observed gradients and LMO answers back by P_g^{-1}. Function values stay unchanged, and the final internal output is mapped by P_g. This preserves feasibility, objective gap, and query count. Thus the method class and loss are invariant under a transitive action on the $d!$ instances. The inequality $V_{\mathrm{dist}}(\mathrm{Unif}) \le V_{\mathrm{rand}}$ follows from (36). Conversely, fix a deterministic method A, choose g uniformly, and run gA. On any fixed instance, invariance and transitivity make its expected loss equal to the uniform-prior average loss of A. This randomized method therefore has the same expected loss on every instance. Taking the infimum over A gives $V_{\mathrm{rand}} \le V_{\mathrm{dist}}(\mathrm{Unif})$. □

A.2 Proof of Lemma 2.6

Proof. Fix a seed of a method for the scaled problem and simulate it using the base oracle. Send its query (x, p) to $(\alpha x, p)$; feed back the base value and gradient multiplied by L/α^2 and L/α, and the base LMO answer divided by α. If the scaled method returns $\hat{x}$, the base simulator returns $\alpha\hat{x}$. Its base gap multiplied by L/α^2 is the scaled gap. These measurable maps preserve the first stopping decision and universal final feasibility, without restricting the method's directions. □

A.3 Proof of Theorem 2.7

Proof. Fix the seed. Summing the first estimate gives

$$\mathbb{E}\Delta_T \ge \Delta_0 - 6Td \ge d^2/8 \qquad (d = 64(T+1)), \tag{31}$$

since $\Delta_0 \ge d^2/4$ and $6Td \le d^2/8$. The second estimate and Jensen's inequality then yield

$$\mathbb{E}[f(\hat{x}_T) - f^\star] \ge \frac{\rho^2 C^2 d^2}{3072} \ge \frac{C^2}{768 d^2} = \frac{3 + 2\sqrt{2}}{6144} \frac{\operatorname{diam}(S_\pi)^2}{d^2}.$$

Using $\rho \ge 2/d^2$ and $C^2 = (3 + 2\sqrt{2}) \operatorname{diam}(S_\pi)^2/8$. Substitute $d = 64(T+1)$, average over the independent seed, and then select a fixed instance from the finite average. Lemma 2.6 gives every $L, \alpha > 0$. □

A.4 Proof of Theorem 2.9

Proof. Let $D_T \doteq \Delta_0 - \Delta_T \ge 0$. Equation (11) gives $\mathbb{E}D_T \le 6Td$, also after seed averaging. For $a \doteq \Delta_0 - d^2/8 = d(3d-4)/8 > 0$, Markov's inequality gives

$$\mathbb{P}[\Delta_T < d^2/8] = \mathbb{P}[D_T > a] \le \frac{6Td}{a} = \frac{48T}{3d-4} \le \frac{1}{4}, \tag{32}$$

because $3d - 4 = 192T + 188$. On the complementary event, (11) gives gap at least $\rho^2 C^2 d^2/3072 \ge c_0 \operatorname{diam}(S_\pi)^2/(T+1)^2$. The loss is constant on each seed/transcript class, so this is an actual loss bound, not only a conditional expectation. Positive rescaling and finite averaging give the stated probability and a fixed-instance witness. □

A.5 Proof of Lemma 3.1

Proof. The thresholding formula and uniqueness of its multiplier are the exact-LMO lemma from Grimmer & Liu (2026, Section 2.2), applied after a coordinate permutation. We prove the information statements. For $z_i \ne 0$, put $u_i \doteq |p_i|/\lambda$. The coordinate's constraint contribution is

$$\frac{1}{2}(u_i - w_{\pi(i)})^2 + w_{\pi(i)}(u_i - w_{\pi(i)}) = \frac{1}{2}(u_i^2 - w_{\pi(i)}^2).$$

Using $w_k^2 = 2C^2k$ and dividing the boundary equation by C^2 gives (16). Activity is equivalent to $\tau p_i^2 > \pi(i)$. The left side of (16) is continuous, strictly increasing once positive, and ranges from zero to infinity, so it also determines τ uniquely.

On active coordinates, (14) gives $w_{\pi(i)} = |p_i|/\lambda - |z_i|$. Substitution into the constraint yields

$$C^2 = \frac{1}{\lambda}\sum_i |p_i z_i| - \frac{1}{2}\|z\|^2, \qquad \lambda = \frac{\sum_i |p_i z_i|}{C^2 + \frac{1}{2}\|z\|^2}.$$

In particular, the observed vectors alone give

$$\pi(i) = \frac{(|p_i|/\lambda - |z_i|)^2}{2C^2} \quad (z_i \neq 0).$$

Here i is an already known coordinate label; the learned quantity is the weight rank assigned to it which induces one equality condition for each nonzero component. Some may repeat earlier revealed ranks. For an inactive coordinate, stationarity gives $|p_i| \leq \lambda w_{\pi(i)}$, equivalently $\tau p_i^2 \leq \pi(i)$ and (17).

Conversely, with the active set specified, (16) gives

$$\tau = \frac{1 + \sum_{i\in A} \pi(i)}{\sum_{i\in A} p_i^2}. \tag{33}$$

The active ranks and inactive inequalities reproduce this τ, the same activity pattern, and the vector in (14). □

A.6 Proof of Lemma 3.2

Proof. At time zero there are no constraints. Suppose the statement holds before a query. For fixed U and Tr_t, the pair (x_t, p_t) is fixed and adds no condition on Π. The value and gradient of the common objective (8) are independent of Π. By Lemma 3.1, the LMO answer is equivalent exactly to the active rank equalities and inactive lower bounds. Intersecting them with the previous constraints proves the description by induction. For the fixed seed, the transcript is a deterministic function of Π. Conditioning on it therefore restricts the uniform prior to the compatible permutations, preserving their equal weights. □

A.7 Proof of Lemma 3.3

Proof. Fill rows $m, m-1, \dots, 1$. The $m-i$ already assigned rows have lower bounds at least ℓ_i, so they remove exactly $m-i$ columns from $\ell_i, \dots, m$. Row i therefore has c_i available columns, whatever the preceding choices. Multiplication proves (20). Choosing uniformly among the currently available allowed columns gives every compatible assignment probability $\prod_i c_i^{-1} = |\mathscr{F}(\ell)|^{-1}$. □

A.8 Proof of Lemma 4.1

Proof. Sample Σ by filling rows $m, m-1, \dots, 1$, choosing uniformly among the available allowed columns, as in the proof of Lemma 3.3. Given the assignments to rows $i+1, \dots, m$, $\Sigma(i)$ is uniform on c_i distinct column *indices*, hence on c_i distinct integers. If these are $a_1 < \cdots < a_{c_i}$, then $a_v - a_u \geq v - u$; the pairwise variance formula shows their variance is at least that of $1, \dots, c_i$, namely $(c_i^2 - 1)/12$. Total variance and Cauchy–Schwarz give

$$\sum_i \mathrm{Var}[\Sigma(i)] \geq \frac{1}{12}\sum_i (c_i^2 - 1) \geq \frac{1}{12}\sum_i (c_i - 1)^2 \geq \frac{\Delta^2}{12m}.$$

Original ranks satisfy $|r_a - r_b| \geq |a - b|$. The square-root comparison in (Grimmer & Liu, 2026, Section 3.3, proof of Theorem 1.1) therefore gives $|\sqrt{r_a} - \sqrt{r_b}| \geq |a - b|/(2\sqrt{d})$. Apply this to independent copies and use $\mathrm{Var}(X) = \frac{1}{2}\mathbb{E}(X - X')^2$. The conditional row argument above handles the nested lower bounds in $\mathscr{F}(\ell)$. □

A.9 Proof of Lemma 5.2

Proof. By Lemma 5.1, edges whose larger rank is active at either endpoint contribute at most $2m$ in expectation: apply the bound at each endpoint, reversing orientations for the second.

Now consider an edge swapping $r_a < r_s$, with the larger-rank row inactive at both endpoints. Let Φ, Φ' be their threshold left sides in (25), with roots τ, τ'. Suppose neither endpoint's larger-rank row excludes r_a. At τ, replacing r_s by r_a creates no positive term, and replacing r_a by r_s in the other row cannot increase its term. Hence $\Phi'(\tau) \le \Phi(\tau) = 1$ and $\tau' \ge \tau$. Reversing the swap gives $\Phi(\tau') \le \Phi'(\tau') = 1$ and $\tau \ge \tau'$.

The thresholds agree; both swapped rows are inactive, and all other assignments are unchanged. The LMO answers therefore coincide by Lemma 3.1. Thus every remaining edge between different answers has a pair (i, a) counted by $N(\sigma)$ at some endpoint σ. The endpoint and pair determine the swap, so at most two orientations are counted per pair. This gives the remainder $2\mathbb{E}N(\Sigma)$. □

B Detailed comparison with existing lower bounds

A lower bound's implications for randomization depend on its algorithm class, its horizon-dependent data, and the time at which the instance is fixed relative to the seed. We examine these choices in the two recent $1/T^2$ lower bounds.

B.1 Distributional and randomized values

For a finite instance family $\mathcal{I}$, let $\mathcal{A}_T$ be the admissible deterministic methods using at most T calls. For bounded measurable loss $\mathcal{L} \in [0, M_{\text{loss}}]$, define

$$V_{\text{dist}}(\mu) \doteq \inf_{A \in \mathcal{A}_T} \mathbb{E}_{I\sim\mu} \mathcal{L}(A, I), \tag{34}$$

$$V_{\text{rand}} \doteq \inf_{\nu} \sup_{I \in \mathcal{I}} \mathbb{E}_{A\sim\nu} \mathcal{L}(A, I), \qquad V_{\text{det}} \doteq \inf_{A \in \mathcal{A}_T} \sup_{I} \mathcal{L}(A, I). \tag{35}$$

The deterministic method may know μ; ν ranges over arbitrary measurable mixtures of methods. In our application, compactness, feasibility and finiteness give $M_{\text{loss}} \doteq \max_\pi \max_{x \in S_\pi} (f(x) - f(x_\star^{(\pi)})) < \infty$.

Proposition B.1 (Yao's principle)**.** *For every prior μ, and under closure under arbitrary mixtures,*

$$V_{\text{dist}}(\mu) \le V_{\text{rand}} \le V_{\text{det}}, \tag{36}$$

$$\sup_{\mu} V_{\text{dist}}(\mu) = V_{\text{rand}}. \tag{37}$$

Proof of Proposition B.1. For any μ, ν, boundedness and measurability give $\sup_I \mathbb{E}_\nu \mathcal{L} \ge \mathbb{E}_\nu \mathbb{E}_\mu \mathcal{L} \ge V_{\text{dist}}(\mu)$. Take the infimum over ν; deterministic methods are degenerate mixtures. For equality, take the closure of the deterministic loss vectors in the finite-dimensional cube $[0, M_{\text{loss}}]^{|\mathcal{I}|}$ and its compact convex hull. Closure does not change the infimum of a linear functional, and finite mixtures approximate its points. The finite-dimensional minimax theorem gives the equality (Yao, 1977; Sion, 1958). □

A lower bound for one fixed prior therefore suffices.

B.2 Further related work

Classical and information-theoretic oracle bounds limit what a query reveals (Nemirovski & Yudin, 1983; Raginsky & Rakhlin, 2009; Agarwal et al., 2012; Braun & Pokutta, 2015; Braun et al., 2017). Our approach is somewhat similar but it determines the exact conditional distribution (and provides a combinatorial statistic Δ) rather than converting real-valued answers into a finite alphabet whose information can be measured with information-theoretic notions such as entropy. Adaptive permutation problems such as permutation Mastermind have related conditional-assignment questions (Martinsson & Su, 2024), but the answers of an LMO require rather different threshold constraints. There is also a recent result by Hazan (2026) which establishes linear-convergence for smooth convex sets,

however it uses membership-based supporting-tangent computations and a uniform interior rolling-ball condition; both the oracle and the geometry differ from the nonsmooth-boundary sets considered here.

B.3 (Grimmer & Liu, 2026): deterministic and randomized quantifiers

Grimmer & Liu (2026) prove a sharp $\Omega(L\,\mathrm{diam}(S)^2/T^2)$ rate for so-called *LMO-span methods*, which is a subclass of general LMO-based methods. In the case of *deterministic methods*, this bound can then be converted into one for *general* LMO-based methods using a resisting oracle; see further below.

More precisely, Grimmer & Liu (2026) show that for *LMO-span methods*, one fixed ordered instance is hard for every admissible trajectory. Thus, fixing a randomized method's seed $U = u$ gives another admissible trajectory on that same instance. The lower bound therefore holds for every u and remains valid after averaging over U. Put differently, for randomized LMO-span methods, Grimmer & Liu (2026) already implicitly provide a lower bound on the rate of $\Omega(L\,\mathrm{diam}(S)^2/T^2)$ however not for general randomized LMO-based methods. We will now show that the LMO-span restriction is indeed significant; see also the discussion in Grimmer & Liu (2026).

Remark B.2 (Methods included in and excluded from the span class). Let $z_{k+1} \doteq \mathtt{LMO}_S(p_k)$. The *LMO-span* model of Grimmer & Liu (2026) requires

$$\begin{aligned} p_k &\in \mathrm{span}\big(\{x_\ell - x_0\}_{\ell=1}^k \cup \{z_\ell - x_0\}_{\ell=1}^k \cup \{\nabla f(x_\ell)\}_{\ell=0}^k\big) \setminus \{0\}, \\ x_{k+1} &\in \mathrm{conv}\{x_0, z_1, \dots, z_{k+1}\}. \end{aligned}$$

There is no computational restriction within these sets. The class contains ordinary Frank–Wolfe with open-loop steps, short steps, or exact line search, and the usual away-step, pairwise, fully corrective, and blended pairwise variants: their LMO direction is the gradient or a linear combination of observed vectors, and their feasible iterates remain in the hull of previously obtained atoms. Randomizing their step choices preserves these conditions for every fixed seed.

The broader class of general first-order LMO-based methods also permits directions such as

$$p_k \doteq \nabla f(x_k) + \delta_k e_{J_k} \qquad \text{or} \qquad p_k \doteq H_k \nabla f(x_k),$$

which the span condition excludes when the added coordinate or transformed gradient lies outside the observed span. Such directions can still be used with the feasible update $x_{k+1} \doteq (1-\gamma_k)x_k + \gamma_k z_{k+1}$.

For the following examples, C, w_k, S_π, M, and f are the family parameters in (5)–(8), and $x_\star^{(\pi)}$ is the optimizer in (9). At $x_0 = 0$, $\nabla f(0) = -M\mathbf{1}$, so a random coordinate perturbation already lies outside the permitted span at the first query. The span class also excludes methods that probe the hidden set and then construct a feasible output outside the atom hull. Indeed,

$$\mathtt{LMO}_{S_\pi}(e_i) = -\left(\sqrt{w_{\pi(i)}^2 + 2C^2} - w_{\pi(i)}\right) e_i,$$

and its magnitude is strictly decreasing in $w_{\pi(i)}$. An exact coordinate query thus identifies $\pi(i)$. After d coordinate queries, a method can recover π and output its explicit optimizer. The query atoms are componentwise nonpositive, whereas the optimizer has $d-1$ positive coordinates, so that output is outside their convex hull.

There is also a one-call separation on a specified public instance. Fix S_{π_0} and its optimizer $x_\star^{(\pi_0)}$. A method can store $p^\star \doteq \nabla f(x_\star^{(\pi_0)})$, query $z_1 \doteq \mathtt{LMO}_S(p^\star)$, and output z_1. This output is feasible on every set presented to the method. On S_{π_0}, first-order optimality gives

$$\langle p^\star, x - x_\star^{(\pi_0)} \rangle \geq 0 \quad \text{for every } x \in S_{\pi_0}.$$

Uniqueness of linear minimization over the strongly convex set gives $z_1 = x_\star^{(\pi_0)}$, so the method has zero error on that instance. It is not a LMO-span method though: $p^\star$ has unequal coordinates, whereas the initial permitted span is $\mathrm{span}\{\nabla f(0)\} = \mathrm{span}\{\mathbf{1}\}$.

These are separations between algorithm classes on specified instances, not $o(T^{-2})$ worst-case algorithms. Grimmer & Liu (2026)'s resisting oracle theorem rules out such an improvement for deterministic methods, and our distributional theorem rules it out for randomized methods.

To go from LMO-span methods to unrestricted deterministic LMO-based methods, Grimmer & Liu (2026) instead assign the hidden coordinate ranks gradually in response to the queries. After producing a deterministic transcript, they complete those assignments to a fixed permutation that would have generated the same transcript. This is a resisting-oracle argument.

To make this distinction clear, write $\mathscr{A}_{\text{det}}$ for the deterministic class, I for an instance, $\mathscr{L}$ for the loss, and c_T for the asserted lower threshold. These symbols are schematic here; see also Appendix B.1. The resisting argument gives

$$\forall A \in \mathscr{A}_{\text{det}} \quad \exists I_A \quad \mathscr{L}(A, I_A) \geq c_T. \tag{38}$$

Applying it separately to fixed seeds yields only $\forall u\, \exists I_u$. A randomized lower bound requires

$$\forall \nu \in \Delta(\mathscr{A}_{\text{det}}) \quad \exists I \quad \mathbb{E}_{A\sim\nu}\mathscr{L}(A, I) \geq c_T, \tag{39}$$

where $\Delta(\mathscr{A}_{\text{det}})$ denotes probability distributions over the deterministic methods. The same instance must work after averaging over seeds and cannot depend on the realized seed; this is precisely what makes things much harder.

A sufficient distributional statement is

$$\exists \mu \quad \forall A \in \mathscr{A}_{\text{det}} : \qquad \mathbb{E}_{I\sim\mu}\mathscr{L}(A, I) \geq c_T, \tag{40}$$

with μ fixed before the method runs, i.e., the algorithm may know the distribution but not the realization from that distribution. Yao's inequality then gives (39). We take μ uniform over the coordinate permutations in Grimmer & Liu (2026)'s family and we retain their sets S_π, square-root-spaced weights, common quadratic objective, optimizer and exact-LMO formulas, and conversion of optimizer variation into objective error. However, their resisting argument averages over completions of an unconstrained suffix of ranks, while under our fixed prior, an unrestricted query can instead (i) reveal several ranks simultaneously and (ii) impose row-dependent lower bounds that need to be taken into consideration for the conditional distributions. In Sections 3 to 5 we characterize these conditional distributions, prove their variance bound, and bound the expected decrease of Δ. We emphasize, that our argument is a distributional complexity argument with a fixed prior for unrestricted randomized methods and *not* a randomization of the resisting oracle, which would most likely reduce to their randomized span-class methods lower bound.

B.4 (Halbey et al., 2026a): prescribed trajectories and initialization sensitivity

Halbey et al. (2026a) study

$$\min_{\|x\|_2\leq 1} \|x - p\|_2^2, \qquad \|p\|_2 = 1, \tag{41}$$

for Frank–Wolfe with exact line search or short steps. For every requested horizon, they construct an initialization whose objective gap is $\Theta(1/t^2)$ over a long segment. This proves sharpness of the $O(1/T^2)$ rate for those updates. It does not determine the measure of such initializations or lower-bound unrestricted LMO methods.

We analyze the sensitivity of their exact-line-search backward construction (and hence also short-steps). For the exact-line-search segments analyzed below, the restriction is the slow-envelope condition (43) on the contraction factors at every step; we make no sensitivity claim for portions of a trajectory outside that condition. For a nonoptimal iterate with noncollinear residual, write

$$e \doteq x - p, \qquad r \doteq \|e\|_2, \qquad u \doteq e/r, \qquad \theta \doteq \langle u, p\rangle.$$

Here r is the distance to the optimizer and θ specifies the residual's orientation. Rotational symmetry keeps the trajectory in the plane spanned by x_0 and p. The unit-ball LMO returns $v = -u$. Exact line search along $d \doteq v - x = -(1 + r)u - p$ gives

$$\gamma(r, \theta) = \frac{r(1 + r + \theta)}{(1 + r)^2 + 1 + 2(1 + r)\theta}.$$

Substitution in $x^+ \doteq x + \gamma d$ yields the exact two-dimensional recurrence

$$F(r, \theta) \doteq (r^+, \theta^+) = \big(rs(r, \theta), -(1 + r)s(r, \theta)\big), \qquad s(r, \theta)^2 = \frac{1 - \theta^2}{(1 + r)^2 + 1 + 2(1 + r)\theta}. \tag{42}$$

Its Jacobian describes the first-order propagation of perturbations in these two state coordinates.

On the slow segments constructed by Halbey et al. (2026a), the contraction $s_t \doteq r_{t+1}/r_t$ satisfies

$$s_t = 1 - \frac{4}{3} r_t + c_t r_t^2, \qquad 1 \le c_t \le \frac{5}{2}. \tag{43}$$

The radial recurrence gives $r_t = \Theta(1/t)$. After one forward step, the angular recurrence gives

$$\theta_t = -1 + \frac{1}{3} r_t + O(r_t^2). \tag{44}$$

Along this curve, differentiation of (42) gives

$$DF(r_t, \theta_t) = \begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix} + O(r_t). \tag{45}$$

The limiting matrix arising from $r \to 0^+$ with $\theta_t = -1 + \frac{1}{3} r_t + O(r_t^2) \to -1$ along the trajectory has tangent eigenvector $(3, 1)^\top$ with eigenvalue 1 and transversal eigenvector $(0, 1)^\top$ with eigenvalue -2, i.e., i.e., it leaves $(3, 1)$ invariant, tangent to (44), and multiplies the transverse direction $(0, 1)$ by -2. In particular, a transverse perturbation thus changes sign and approximately doubles in one slow step. The following statement accounts for the variation of the Jacobians along the segment.

Proposition B.3 (Initialization sensitivity along a slow segment)**.** *Reindex a slow segment satisfying* (43) *after entry into a sufficiently small fixed neighborhood of* $(0, -1)$ *and after one forward step, so that* (44) *holds from time zero. Fix entry scales* $0 < r_- \le r_+$ *and assume* $r_- \le r_0 \le r_+$*, with* r_+ *sufficiently small. Put*

$$\mathcal{J}_T \doteq DF(r_{T-1}, \theta_{T-1}) \cdots DF(r_0, \theta_0), \qquad \sigma_T \doteq \|\mathcal{J}_T\|_2.$$

There are constants $a, A, C > 0$*, depending only on the fixed entry scales and slow-envelope bounds, not on the segment length, such that*

$$a\, 2^T (T+1)^{-C} \le \sigma_T \le A\, 2^T (T+1)^C. \tag{46}$$

Consequently $\log_2 \sigma_T = T + O(\log T)$.

Proof Sketch. The full proof can be found in Appendix B.5. In the eigenbasis of the limiting matrix, the Jacobian is $\mathrm{diag}(1, -2) + E_t$, with $\|E_t\| = O(1/t)$. A cone about the transverse direction is invariant. Multiplying its expansion factors $2(1 - O(1/t))$ gives the lower bound; harmonic summation contributes only a polynomial factor. An operator-norm estimate gives the upper bound. □

To state the associated linearized precision requirement, define

$$\Lambda_T \doteq \max_{0 \le t \le T} \|\mathcal{J}_t\|_2. \qquad \text{and} \qquad \mathcal{J}_0 = I.$$

A radius-u uncertainty ball about the initialization has worst-case first-order trajectory error $u\Lambda_T$. For tolerance η, the number of binary digits required in this linearized uncertainty model is

$$B_{\mathrm{lin}}(T, \eta) \doteq \left\lceil \log_2 \frac{\Lambda_T}{\eta} \right\rceil = T + \log_2(1/\eta) + O(\log T). \tag{47}$$

For fixed or polynomially shrinking tolerance, the leading term is one binary digit per slow iteration. The approximately 10^{-300} numerical accuracy reported in Halbey et al. (2026a) for a 1000-step construction is consistent with this scale: $300 \log_2 10 \approx 996.6$. Note, that Proposition B.3 is a worst-case *linearized* statement about small transverse errors along a prescribed trajectory. It does not imply that every lower-precision run fails and in paticular, a specially representable initialization, a tangential perturbation, or favorable rounding cancellation may behave better. Appendix B.5 proves a separate nonlinear bound for additive errors in the polar recurrence.

The exact map also gives a measure-zero result in the plane for the set $\mathcal{E}$ of initializations whose trajectories never reach p in finite time and eventually satisfy the slow-envelope condition. These trajectories approach p asymptotically. The result concerns infinite trajectories, not hardness at a single prescribed horizon.

Theorem B.4 (Zero area for the eventual slow envelope)**.** *Consider exact-line-search Frank–Wolfe on the unit disk in $\mathbb{R}^2$ for $f(x) \doteq \|x - p\|_2^2$, with $\|p\|_2 = 1$, as in (41). The optimizer p is kept fixed once reached. Let $\mathscr{E}$ be the set of $x_0 \in B_1(0)$ whose iterates never reach p and satisfy (43) for every sufficiently large t. Then $\mathscr{E}$ has two-dimensional Lebesgue measure zero, although it is uncountable. For every $x_0 \in \mathscr{E}$,*

$$tr_t \to \frac{3}{4}, \qquad t^2(f(x_t) - f(p)) \to \frac{9}{16}. \tag{48}$$

In particular, an initialization distribution absolutely continuous with respect to area assigns probability zero to $\mathscr{E}$.

Proof Sketch. The proof can be found in Appendix B.5.5. With $q \doteq 1 + \theta$, the map is injective on

$$W \doteq \{(r, q) : 0 < r \le 1/100,\ r/4 \le q \le 2r/5\}$$

and expands (r, q) area by a factor greater than $6/5$. The area of starts remaining there for n updates is therefore at most $\text{area}(W)(5/6)^n$. The envelope forces eventual membership in W, and finite preimages and Cartesian coordinate changes preserve nullity. A nested-interval argument supplies such a trajectory at every initial radius in $(0, 1/100]$, so the statement is not vacuous. □

For exact line search, or equivalently short steps on this instance, an initial point sampled from a distribution with a density on the disk almost surely lies outside $\mathscr{E}$ and therefore does not generate an infinite trajectory satisfying the eventual slow-envelope condition. We establish the existence of such trajectories, together with the measure-zero statement, in Theorem B.4. The stated guarantee in Halbey et al. (2026a) is finite-horizon, with an initialization that may depend on the horizon. This does not rule out long finite slow phases, and Halbey et al. (2026a)'s finite-horizon lower bound is independent of this. Nor does the result establish that every initialization whose trajectory has $\Omega(t^{-2})$ error belongs to $\mathscr{E}$.

Neither the local sensitivity estimate nor Theorem B.4 determines the measure of all hard initializations. Let $T(x_0, \varepsilon)$ be the first time the objective gap is at most ε. For fixed $\kappa > 0$, define

$$H_\varepsilon(\kappa) \doteq \left\{x_0 \in B_1(0) : T(x_0, \varepsilon) \ge \frac{\kappa}{\sqrt{\varepsilon}}\right\}, \qquad q_\varepsilon(\kappa) \doteq \frac{\text{vol}(H_\varepsilon(\kappa))}{\text{vol}(B_1(0))}. \tag{49}$$

The finite-horizon strict inequalities define a relatively open set by continuity. Thus a strict hard point gives a positive-measure subset of $H_\varepsilon(\kappa)$ and $q_\varepsilon(\kappa) > 0$ at the relevant fixed accuracies. This supplies no lower bound uniform as $\varepsilon \downarrow 0$. Independent uniform starts sampled until one lies in $H_\varepsilon(\kappa)$ have a geometric waiting time with mean $1/q_\varepsilon(\kappa)$. A limit $q_\varepsilon(\kappa) \to 0$ would therefore imply a diverging expected number of trials. That limit is unknown: in particular, a positive limit superior or nonconvergence has not been ruled out.

The inverse recurrence explains why the local sensitivity and nonlinear perturbation bounds do not settle this measure question. A valid nonterminating next state (R, Θ), with $R > 0$, determines $s = -\Theta - R$ and $r = R/s$, but the preceding angle may have two values:

$$\theta_\pm = -s^2(1 + r) \pm \sqrt{(1 - s^2)(1 - (1 + r)^2 s^2)}. \tag{50}$$

There can be as many as 2^T histories after T inverse steps, whereas the width of one slow inverse branch is $2^{-T}T^{O(1)}$. Schematically,

$$\underbrace{2^T}_{\text{possible branches}} \times \underbrace{2^{-T}}_{\text{width per branch}} \asymp 1.$$

The elementary doubling map exhibits this cancellation. For the Frank–Wolfe map, a proof that $q_\varepsilon(\kappa) \to 0$ would additionally need a global estimate on escape or on the measures of the inverse histories that remain hard.

B.5 Proofs for the (Halbey et al., 2026a) calculation

In this appendix we derive the exact recurrence and prove the variable-Jacobian estimate in Proposition B.3. We then consider additive nonlinear perturbations and prove Theorem B.4 for the eventual slow envelope. Some of the computations are already contained in (Halbey et al., 2026a) while others are relatively straightforward but tedious. We include both in detail for the sake of completeness.

B.5.1 Exact polar dynamics

Fix $\|p\|_2 = 1$ and $f(x) \doteq \|x - p\|_2^2$ on the unit ball. For $x \neq p$, write

$$e \doteq x - p = ru, \qquad r \doteq \|e\|_2, \qquad \theta \doteq \langle u, p \rangle.$$

The LMO minimizes $\langle e, v \rangle$ over $\|v\|_2 \leq 1$ and returns $v = -u$. Its displacement from the current iterate is

$$d \doteq v - x = -u - (p + ru) = -(1 + r)u - p.$$

Along $x + \gamma d$, exact line search minimizes $\|e + \gamma d\|_2^2$. The unconstrained minimizer is $-\langle e, d \rangle / \|d\|_2^2$, and

$$\langle e, d \rangle = -r(1 + r + \theta), \qquad D(r, \theta) \doteq \|d\|_2^2 = (1 + r)^2 + 1 + 2(1 + r)\theta.$$

Thus

$$\gamma(r, \theta) = \frac{r(1 + r + \theta)}{D(r, \theta)}. \tag{51}$$

For the noncollinear feasible states in the slow construction this step-size belongs to $(0, 1)$, so clipping to the Frank–Wolfe segment has no effect and we can ignore it.

Set $e^+ \doteq e + \gamma d$, $r^+ \doteq \|e^+\|_2$, and $s \doteq r^+ / r$. Substitution gives

$$s^2 = 1 - \frac{\langle e, d \rangle^2}{r^2 \|d\|_2^2} = 1 - \frac{(1 + r + \theta)^2}{D(r, \theta)} = \frac{1 - \theta^2}{D(r, \theta)}.$$

Exact line search also makes e^+ orthogonal to d, so

$$0 = \langle e^+, d \rangle = -(1 + r)\langle e^+, u \rangle - \langle e^+, p \rangle.$$

Direct substitution of (51) gives $\langle e^+, p \rangle = -r(1 + r)s^2$. Dividing by $r^+ = rs$ gives $\theta^+ = -(1 + r)s$. Hence the exact recurrence is

$$F(r, \theta) = (rs, -(1 + r)s), \qquad s^2 = \frac{1 - \theta^2}{D(r, \theta)},$$

as used in Appendix B.4.

B.5.2 The slow curve and its Jacobian

Assume the slow-envelope hypothesis from Halbey et al. (2026a)

$$s_t = 1 - \frac{4}{3} r_t + c_t r_t^2, \qquad 1 \leq c_t \leq \frac{5}{2}.$$

Since $r_{t+1} = r_t s_t$,

$$\frac{1}{r_{t+1}} - \frac{1}{r_t} = \frac{1 - s_t}{r_t s_t} = \frac{4/3 - c_t r_t}{s_t}. \tag{52}$$

For sufficiently small fixed radii, the last expression is bounded above and below by positive constants uniformly over $c_t \in [1, 5/2]$. After reindexing at entry into this neighborhood, summation gives $r_t = \Theta((t + 1)^{-1})$, with the fixed entry-scale bounds from the proposition.

The angular recurrence gives

$$\begin{aligned} \theta_{t+1} &= -(1 + r_t)s_t \\ &= -1 + \frac{1}{3} r_t + \left(\frac{4}{3} - c_t \right) r_t^2 - c_t r_t^3. \end{aligned}$$

Since $r_{t+1} = r_t + O(r_t^2)$, after one forward step this is

$$\theta_{t+1} = -1 + \frac{1}{3} r_{t+1} + O(r_{t+1}^2).$$

The forward recurrence therefore determines the coefficient $1/3$.

We now differentiate at a generic state (r, θ), where $r = \|x - p\|_2 > 0$ and $\theta = \langle (x - p)/r, p \rangle \in (-1, 1)$. Recall the scalar denominator and write the contraction factor as a function of these coordinates:

$$D(r, \theta) \doteq (1 + r)^2 + 1 + 2(1 + r)\theta = \|d\|_2^2,$$

$$s(r, \theta) \doteq \sqrt{\frac{1 - \theta^2}{D(r, \theta)}}.$$

Along the trajectory, $s_t = s(r_t, \theta_t)$. Write $s_r \doteq \partial s/\partial r$ and $s_\theta \doteq \partial s/\partial \theta$, holding the other coordinate fixed. Taking logarithms gives

$$2 \log s = \log(1 - \theta^2) - \log D.$$

Differentiating this identity with respect to r and θ separately, using $\partial_r \log s = s_r/s$ and $\partial_\theta \log s = s_\theta/s$, gives

$$s_r = -s\frac{1 + r + \theta}{D}, \tag{53}$$

$$s_\theta = -s\left(\frac{\theta}{1 - \theta^2} + \frac{1 + r}{D}\right). \tag{54}$$

Thus

$$DF(r, \theta) = \begin{pmatrix} s + rs_r & rs_\theta \\ -s - (1 + r)s_r & -(1 + r)s_\theta \end{pmatrix}. \tag{55}$$

For the slow-curve substitution, set $\beta \doteq (\theta + 1 - r/3)/r^2$; the preceding forward-step expansion gives a fixed bound $|\beta| \le B$. Substituting $\theta = -1 + r/3 + \beta r^2$ in (53)–(54) yields

$$s = 1 + O(r), \qquad s_r = -2 + O(r), \qquad s_\theta = 2 + O(r).$$

Here all three remainders are bounded by Cr in absolute value for $0 < r \le \delta$ and $|\beta| \le B$, with $C, \delta > 0$ depending only on B. Consequently,

$$DF(r_t, \theta_t) = J_\infty + O(r_t), \qquad J_\infty \doteq \begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix}. \tag{56}$$

Its eigenvectors satisfy

$$J_\infty(3, 1)^\top = (3, 1)^\top, \qquad J_\infty(0, 1)^\top = -2(0, 1)^\top.$$

The first is tangent to $\theta = -1 + r/3$; the second changes the angle while fixing the residual radius r to first order.

B.5.3 Products of the varying Jacobians

Proof of Proposition B.3. Equation (52) gives $r_t = \Theta((t + 1)^{-1})$. By (56),

$$J_t \doteq DF(r_t, \theta_t) = J_\infty + O((t + 1)^{-1}).$$

Let

$$P \doteq \begin{pmatrix} 3 & 0 \\ 1 & 1 \end{pmatrix}$$

be the fixed eigenvector matrix of the limiting matrix J_∞ and observe that $P^{-1}J_\infty P = \mathrm{diag}(1, -2)$. There are fixed $K, t_0 > 0$ such that

$$P^{-1}J_t P = \mathrm{diag}(1, -2) + E_t, \qquad |(E_t)_{ij}| \le \epsilon_t \doteq \frac{K}{t + t_0}.$$

Taking a smaller fixed entry neighborhood if needed increases t_0 and ensures $\epsilon_t \le 1/4$ at each reindexed step.

Let us now write a linearized perturbation in the fixed eigenbasis as $h_t = P(a_t, b_t)^\top$. Its update $h_{t+1} = J_t h_t$ has coefficient form

$$\begin{pmatrix} a_{t+1} \\ b_{t+1} \end{pmatrix} = P^{-1}J_t P \begin{pmatrix} a_t \\ b_t \end{pmatrix}.$$

Consider the cone $\mathscr{C} \doteq \{(a, b) \in \mathbb{R}^2 : |a| \le |b|\}$ in coefficient space. For $(a_t, b_t) \in \mathscr{C}$, the entrywise bounds give

$$|a_{t+1}| \le (1 + 2\epsilon_t)|b_t|, \qquad |b_{t+1}| \ge (2 - 2\epsilon_t)|b_t|.$$

Since $\epsilon_t \le 1/4$, the first bound is no larger than the second, so $\mathscr{C}$ is invariant. Starting from $(a_0, b_0) = (0, 1)$ gives

$$|b_T| \ge 2^T \prod_{t<T}(1 - \epsilon_t). \tag{57}$$

For $0 \le u \le 1/2$, $\log(1-u) \ge -2u$. Since $\sum_{t<T} \epsilon_t = O(\log(T+1))$, the product is bounded below by a negative power of $T+1$. Changing coordinates back through P yields

$$\|J_{T-1} \cdots J_0\|_2 \ge a 2^T (T+1)^{-C}.$$

For the upper bound, the induced maximum norm satisfies

$$\|P^{-1} J_t P\|_\infty \le 2(1+\epsilon_t).$$

Multiplication and $\log(1+u) \le u$ yield

$$\|J_{T-1} \cdots J_0\|_2 \le A 2^T (T+1)^C.$$

Equivalence of norms changes only the constants a, A, C, proving (46). □

For each fixed T, differentiability gives the uncertainty interpretation used in (47):

$$\lim_{u \downarrow 0} \frac{1}{u} \sup_{\|h\|_2 \le u} \max_{0 \le t \le T} \|F^t((r_0,\theta_0)+h) - F^t(r_0,\theta_0)\|_2 = \max_{0 \le t \le T} \|D(F^t)(r_0,\theta_0)\|_2.$$

The right side is $2^T T^{O(1)}$ by the proposition. Its binary logarithm gives the linearized precision requirement.

The same leading exponent holds in Cartesian coordinates. Rotate to $p = (1,0)$ and fix one side of the invariant plane. The coordinate map is

$$(r,\theta) \mapsto x \doteq (1 + r\theta,\ r\sqrt{1-\theta^2}).$$

Its Jacobian columns are orthogonal, with singular values 1 and $r/\sqrt{1-\theta^2}$. Along the slow curve the latter is $\Theta(\sqrt{r}) = \Theta(t^{-1/2})$. Thus coordinate conversion contributes only polynomial factors, changing the required number of binary digits by $O(\log T)$.

B.5.4 Nonlinear bounds for additive state perturbations

We next consider a nonlinear additive perturbation error model of the two polar state coordinates.

Theorem B.5 (Finite-horizon additive perturbation bound)**.** *Let $z_t \doteq (r_t, \theta_t)$ be a slow segment from Proposition B.3, reindexed after entering a sufficiently small fixed neighborhood of $(0,-1)$. There are constants $c, \eta_0, \Gamma > 0$, independent of T, with the following property. For every integer $T \ge 1$ and $0 < \eta \le \eta_0$, let*

$$\tilde{z}_0 \doteq z_0 + e_0, \qquad \tilde{z}_{t+1} \doteq F(\tilde{z}_t) + \xi_t, \quad 0 \le t < T,$$

where $e_0, \xi_t \in \mathbb{R}^2$ may be chosen adversarially. If

$$\|e_0\|_\infty + \sum_{t=0}^{T-1} \|\xi_t\|_\infty \le c\eta\, 2^{-T} (T+1)^{-\Gamma}, \tag{58}$$

then every perturbed state is in the domain of the polar map and

$$\|\tilde{z}_t - z_t\|_\infty \le \eta r_t^3 = O\Big(\frac{\eta}{(t+1)^3}\Big), \qquad 0 \le t \le T. \tag{59}$$

In particular, (58) follows from

$$\|e_0\|_\infty \le \frac{c\eta}{2} 2^{-T} (T+1)^{-\Gamma}, \qquad \max_{t<T} \|\xi_t\|_\infty \le \frac{c\eta}{2} 2^{-T} (T+1)^{-(\Gamma+1)}.$$

Proof. Put $q \doteq 1 + \theta$. This translates the second state coordinate, leaving differences and derivative matrices unchanged. Along the slow segment, $q_t = r_t/3 + O(r_t^2)$ and $r_t = \Theta((t+1)^{-1})$. After shrinking the fixed entry neighborhood, choose constants $0 < a < b$ such that

$$a r_t \le q_t \le b r_t$$

at every reindexed state. The numerator and denominator of the contraction factor are

$$1 - \theta^2 = 2q - q^2, \qquad D(r,\theta) = r^2 + 2(1+r)q. \tag{60}$$

Use a larger fixed radial cutoff $r_{\max} \doteq 2r_+$, with r_+ sufficiently small, and consider $0 < r \le r_{\max}$, $ar/2 \le q \le 2br$. This cutoff is distinct from the actual initial radius r_0. Both quantities in (60) are bounded above and below by positive constant multiples of r throughout this larger neighborhood.

In the (r, q) coordinates, with the output's second coordinate translated as well, the same map is represented by

$$F(r,q) = \left(r\sqrt{\frac{2q-q^2}{r^2+2(1+r)q}}, 1-(1+r)\sqrt{\frac{2q-q^2}{r^2+2(1+r)q}} \right).$$

Differentiating twice gives

$$\|D^2F(r,q)\|_\infty \le Kr^{-2} \tag{61}$$

for a fixed K. The first two derivatives of the polynomial numerator and denominator are bounded; differentiating their square-root quotient twice introduces at worst two reciprocal powers of quantities comparable to r.

Write $h_t \doteq \tilde{z}_t - z_t$ and provisionally assume $\|h_t\|_\infty \le \eta r_t^3$. For sufficiently small η_0, the segment from z_t to $z_t + h_t$ stays in the larger neighborhood, with radius comparable to r_t. To see the margins, $r_t \le r_+$ leaves room up to $r_{\max} = 2r_+$ even for outward perturbations at time zero, while $\|h_t\|_\infty / r_t \le \eta_0 r_+^2$ is small relative to the fixed differences between $ar_t \le q_t \le br_t$ and the enlarged angular bounds. The integral mean-value formula therefore gives

$$F(z_t + h_t) - F(z_t) = A_t h_t, \qquad A_t = DF(z_t) + R_t, \qquad \|R_t\|_\infty \le K' r_t^{-2} \|h_t\|_\infty \le K' \eta r_t.$$

In the fixed basis P used above,

$$P^{-1} A_t P = \operatorname{diag}(1,-2) + \widetilde{E}_t, \qquad \|\widetilde{E}_t\|_\infty \le \frac{K''}{t+1}.$$

The induced maximum norm and $\prod_{s=j}^{t-1}(1 + K''/(s+1)) \le K_0(t+1)^{K''}$ yield fixed $K_1, K_2 > 0$ such that every product formed while the provisional bound holds satisfies

$$\|A_{t-1} \cdots A_j\|_\infty \le K_1 2^{t-j} (t+1)^{K_2}. \tag{62}$$

Variation of constants in $h_{t+1} = A_t h_t + \xi_t$ gives

$$\|h_t\|_\infty \le K_3 2^t (t+1)^{K_2} \left(\|e_0\|_\infty + \sum_{j=0}^{t-1} \|\xi_j\|_\infty \right).$$

Take $\Gamma \doteq K_2 + 3$. The bound $r_t \ge c_r/(t+1)$ for a fixed $c_r > 0$ and (58) imply

$$\frac{\|h_t\|_\infty}{\eta r_t^3} \le \frac{K_3 c}{c_r^3} 2^{t-T} \left(\frac{t+1}{T+1} \right)^{K_2+3} \le \frac{K_3 c}{c_r^3}.$$

Choose $c < c_r^3/K_3$. A first-exit argument proves the provisional bound for all $0 \le t \le T$ and keeps every perturbed state in the larger neighborhood. This gives (59). Finally, the separate bounds imply (58) because $T/(T+1) \le 1$. □

The fixed lower bound on the starting distance to the optimizer r_- from Proposition B.3 remains part of the theorem. Its constants need not be uniform when the starting radius tends to zero independently of T; already at $t = 0$ the allowed error is ηr_0^3. The theorem only provides a sufficient bound for the displayed additive-error model.

B.5.5 Zero area for the eventual slow envelope

We prove Theorem B.4. Rotate so $p = (1, 0)$ and put $q \doteq 1 + \theta$. The coordinate maps $\chi_\pm$ recover the original Cartesian point x from (r, q), with + and − selecting the upper and lower half-disks, respectively:

$$\chi_\pm(r,q) \doteq (1 - r + rq,\ \pm r\sqrt{2q-q^2}). \tag{63}$$

The interior of the disk, excluding the horizontal axis, corresponds to

$$\mathcal{D} \doteq \{(r,q) : 0 < r < 2,\ 0 < q < 1 - r/2\}.$$

Translating the second coordinate of (42) gives

$$\Phi(r,q) \doteq (rs, 1-(1+r)s), \qquad s^2 = \frac{N}{D}, \quad N \doteq 2q - q^2, \quad D \doteq r^2 + 2(1+r)q. \tag{64}$$

The exact line-search parameter lies in $(0,1)$ throughout $\mathscr{D}$: its numerator is $r(r+q) > 0$, and its denominator exceeds that numerator by $(r+2)q > 0$. Thus clipping does not change this map. Moreover,

$$D - (1+r)^2 N = ((1+r)q - r)^2 \geq 0. \tag{65}$$

Hence $s \leq 1/(1+r)$ and $r_t \leq (t + 1/r_0)^{-1}$ until termination; in particular, $r_t \to 0$.

Put

$$\delta \doteq \frac{1}{100}, \qquad a \doteq \frac{1}{4}, \qquad b \doteq \frac{2}{5}, \qquad W \doteq \{(r,q) : 0 < r \leq \delta,\ ar \leq q \leq br\} \subset \mathscr{D}.$$

Let $W_0 \doteq W$ and recursively define

$$W_{n+1} \doteq \{z \in W : \Phi(z) \in W_n\}, \qquad W_\infty \doteq \bigcap_{n \geq 0} W_n.$$

Thus W_n requires the first $n+1$ states, including the initialization, to lie in W. Write $\mathscr{L}^2_{r,q}$ for ordinary area in the (r,q) parameter plane, and $\mathscr{L}^2$ for ordinary area in the original (x_1, x_2) plane. The maps $\chi_\pm$ do not preserve area; their absolute Jacobian determinants account for the change in area.

Lemma B.6 (Area of initializations remaining in W)**.** *The map Φ is injective on W, and*

$$\mathscr{L}^2_{r,q}(W_n) \leq \frac{3\delta^2}{40}\left(\frac{5}{6}\right)^n.$$

In particular $\mathscr{L}^2_{r,q}(W_\infty) = 0$.

Proof. Direct differentiation gives

$$\det D\Phi(r,q) = -s\,\partial_q s = \frac{((1+r)q - r)(r+q)}{(r^2 + 2(1+r)q)^2}. \tag{66}$$

With $k \doteq q/r \in [a,b]$, its magnitude on W is

$$J(r,k) \doteq \frac{(1-(1+r)k)(1+k)}{(r+2(1+r)k)^2}.$$

The numerator is positive. For fixed k, it decreases with r, while the denominator increases. Also,

$$\partial_k J(r,k) = -\frac{(r+2)^2}{(r+2(1+r)k)^3} < 0.$$

Therefore

$$|\det D\Phi| \geq J(\delta, b) = \frac{208600}{167281} > \frac{6}{5}. \tag{67}$$

If $(R,Q) = \Phi(r,q)$, then $s = 1 - Q - R > 0$ and $r = R/s$. Once r is recovered, s^2 is strictly increasing in q on $0 < q < r/(1+r)$, since $\partial_q(s^2) = -2\det D\Phi > 0$ there. The whole of W lies in that branch, so q is unique. The same argument gives a C^1 diffeomorphism on the open branch; the change-of-variables formula therefore applies to its measurable subsets. Since $\Phi(W_{n+1}) \subseteq W_n$, (67) implies

$$\frac{6}{5}\mathscr{L}^2_{r,q}(W_{n+1}) \leq \int_{W_{n+1}} |\det D\Phi| = \mathscr{L}^2_{r,q}(\Phi(W_{n+1})) \leq \mathscr{L}^2_{r,q}(W_n).$$

Finally $\mathscr{L}^2_{r,q}(W) = (b-a)\delta^2/2 = 3\delta^2/40$. Iteration proves the bound, whose limit proves nullity. □

The corresponding Cartesian area satisfies

$$\mathscr{L}^2(\chi_+(W_n) \cup \chi_-(W_n)) \leq \frac{3\delta^{5/2}}{10}\left(\frac{5}{6}\right)^n. \tag{68}$$

Indeed $\left|\det D\chi_\pm\right| = r/\sqrt{2q - q^2} \leq 2\sqrt{\delta}$ on W; apply Lemma B.6 to both maps. This bounds the area of initial points x_0 whose reduced states satisfy $(r_t, q_t) \in W$ for every $t = 0, \ldots, n$, including the initial state.

Lemma B.7 (The eventual envelope and its asymptotic error)**.** *For a nonterminating trajectory,* (43) *holds eventually if and only if its* (r_t, q_t) *coordinates eventually remain in* W*. On any such trajectory,*

$$\frac{q_t}{r_t} = \frac{1}{3} - \frac{2}{9}r_t + o(r_t), \qquad s_t = 1 - \frac{4}{3}r_t + (2 + o(1))r_t^2, \qquad tr_t \longrightarrow \frac{3}{4}.$$

Proof. Suppose first that (43) holds. Once $r_t \le \delta$, (64) gives

$$\frac{q_{t+1}}{r_{t+1}} = \frac{1/3 + (4/3 - c_t)r_t - c_t r_t^2}{s_t}.$$

Here $74/75 \le s_t \le 1$, and interval bounds using $1 \le c_t \le 5/2$ give

$$\frac{3857}{12000} \le \frac{q_{t+1}}{r_{t+1}} \le \frac{101}{296}.$$

Both endpoints lie strictly between a and b defined further above. Together with (65), this proves eventual membership in W.

Conversely, suppose the trajectory eventually remains in W. Discard the finite initial segment before this permanent stay and number the remaining states starting at $t = 0$. Set $k_t \doteq q_t/r_t$. The exact recurrence implies

$$\frac{1}{r_{t+1}} - \frac{1}{r_t} = 1 + k_{t+1}, \qquad k_{t+1} = G(r_t, k_t), \tag{69}$$

where the following expression extends smoothly to $r = 0$ for $k \in [a, b]$:

$$G(r,k) \doteq \frac{(1+k)^2}{\sqrt{A}(\sqrt{A} + \sqrt{B})} - 1, \qquad A \doteq k(2 - kr), \quad B \doteq 2k + (1 + 2k)r.$$

Write $\hat{s}(r,k) \doteq \sqrt{A/B}$ for the contraction in these coordinates, so $s_t = \hat{s}(r_t, k_t)$. This distinguishes it from the original function $s(r,\theta)$ in (42). At $r = 0$ the map G is $g(k) \doteq (k-1)^2/(4k)$, so uniformly on this interval $G(r,k) = g(k) + O(r)$. The fixed point $k_* \doteq 1/3$ satisfies

$$g(k_*) = k_*, \qquad g'(k) \le -\frac{21}{16} \quad (a \le k \le b).$$

Set $\lambda \doteq 21/16$ and apply the mean-value theorem to g between k_t and $k_* = 1/3$. Since $g(k_*) = k_*$ and $|g'(k)| \ge \lambda$ on $[a, b]$,

$$\lambda |k_t - k_*| \le |g(k_t) - g(k_*)| \le |k_{t+1} - k_*| + Cr_t.$$

The second inequality uses $k_{t+1} = G(r_t, k_t)$ and the uniform estimate $|G(r,k) - g(k)| \le Cr$. Iterate this inequality forward. Since k_t is bounded and $r_{t+j} \le r_t$, letting the number of iterations tend to infinity yields $|k_t - k_*| \le Cr_t/(\lambda - 1)$. Thus $\beta_t \doteq (k_t - 1/3)/r_t$ is bounded.

Uniform Taylor expansion for bounded β gives

$$G(r, 1/3 + \beta r) = \frac{1}{3} + \left(-2\beta - \frac{2}{3}\right) r + O(r^2),$$

$$\hat{s}(r, 1/3 + \beta r) = 1 - \frac{4}{3}r + \left(2\beta + \frac{22}{9}\right) r^2 + O(r^3).$$

As $r_{t+1}/r_t = 1 + O(r_t)$, the first expansion implies

$$\beta_{t+1} = -2\beta_t - \frac{2}{3} + O(r_t).$$

Let $L \doteq \limsup_t |\beta_t + 2/9| < \infty$. Since $r_t \to 0$, this identity gives $2L \le L$, whence $L = 0$. Thus $\beta_t \to -2/9$, and the second expansion gives $c_t \to 2$, proving eventual (43). Finally averaging (69) gives $(1/r_t)/t \to 1 + 1/3 = 4/3$. □

Lemma B.8 (Existence at every small positive radius)**.** *For every* $r_0 \in (0, \delta]$ *there is a* $q_0 \in [ar_0, br_0]$ *whose trajectory remains in* W *for all iterations.*

Proof. The angular map crosses both ends of the interval:

$$G(r,a) > b, \qquad G(r,b) < a \qquad (0 < r \le \delta). \tag{70}$$

For example $G(r,a) > b$ is equivalent to $B - A(1+(1+b)r)^2 > 0$ at $k = a$. After dividing by r, that difference is $(65 - 322r + 49r^2)/400 > 0$ on the specified interval. Similarly $G(r,b) < a$ follows from

$$\frac{A(1+(1+a)r)^2 - B}{r}\Big|_{k=b} = \frac{4 + 85r - 25r^2}{100} > 0.$$

All factors squared in these comparisons are positive.

Start with the closed interval $J_0 \doteq [a,b]$ of possible k_0 at the fixed r_0. Inductively choose a nonempty closed subinterval J_n such that all states through time n lie in W, and k_n takes the values a, b at its two endpoints, in either order. The update is continuous on J_n. By (70), k_{n+1} takes values on opposite sides of $[a,b]$ at those endpoints. The intermediate-value theorem gives a closed subinterval $J_{n+1} \subseteq J_n$ on which $k_{n+1} \in [a,b]$, with endpoint values a, b. One may take a first crossing of one endpoint value and the last crossing of the other value before it. All previous constraints remain satisfied, and (65) preserves the radial cutoff. Compactness gives a point in $\bigcap_n J_n$. □

Proof of Theorem B.4. On $\mathscr{D}$, the Jacobian (66) vanishes only on the curve $q = r/(1+r)$, a set of area zero. Consequently the preimage of every area-zero set is area zero: away from this curve the inverse-function theorem provides local C^1 inverse maps. Countably many such neighborhoods suffice, and their locally Lipschitz inverse maps preserve null sets. Add the null critical curve to cover the remaining points.

By Lemma B.6, every finite preimage of W_∞ in $\mathscr{D}$, and their countable union, has zero (r,q) area. Lemma B.7 identifies this union with the eventually slow trajectories starting off the horizontal axis and the disk boundary. A trajectory hitting the critical curve reaches the horizontal axis on the next step, and then the optimizer, so it is not an omitted eventually slow trajectory. The Cartesian charts (63) are locally C^1 on $\mathscr{D}$ and preserve null sets. The excluded initial axis and disk boundary are already Cartesian null sets. Thus $\mathscr{L}^2(\mathscr{E}) = 0$.

Lemma B.8 supplies a point of $\mathscr{E}$ at every initial radius $r_0 \in (0, \delta]$, by Lemma B.7. Distinct radii give distinct initializations, proving uncountability. The rate lemma also gives $t^2 f(x_t) = (t r_t)^2 \to 9/16$. Absolute continuity proves the probability claim. □

C Example of compatible assignments and swaps

Example C.1. *Take $d \doteq 5$ and the generating permutation $(2,4,1,3,5)$. Use nonnegative query directions with squared components $(5/2, 9/2, 0, 0, 0)$ and then $(3,0,0,2,0)$. Both observed thresholds are 1. The first answer reveals $\Pi(1) = 2, \Pi(2) = 4$ simultaneously; the second adds $\Pi(4) \ge 2$ while its only active coordinate is already known. The remaining ranks are $(r_1, r_2, r_3) = (1,3,5)$, and sorting the compressed bounds gives $\iota = (3,5,4)$ and $\ell = (1,1,2)$. There are four compatible reduced permutations:* $(1,2,3), (1,3,2), (2,1,3), (3,1,2)$.

C.1 Swap graph and one-query update

We continue Example C.1, with remaining original ranks $(1,3,5)$, physical row order $\iota = (3,5,4)$, and bounds $\ell = (1,1,2)$. The four reduced permutations are the vertices of Figure 2. From $(1,2,3)$, exchanging the assignments in rows $1,2$ swaps original ranks $1,3$ between physical coordinates $3,5$, giving $(2,1,3)$. The swap in rows $2,3$ gives $(1,3,2)$. The swap in rows $1,3$ would give $(3,2,1)$, violating $\ell_3 = 2$. Every vertex has $\Delta = (1+2+3) - (1+1+2) = 2$ outgoing edges, hence eight directed edges in total.

Now query $p \doteq e_3$, the unit vector at physical coordinate 3, which is reduced row 1. Write $k \doteq \Pi(3) = r_{\sigma(1)} \in \{1,3,5\}$. Only this coordinate can be active, so the threshold equation is $(\tau - k)_+ = 1$. Since $C = 1/\sqrt{32}$, the threshold and LMO answer are

$$\tau_\sigma = k + 1, \qquad z(k) \doteq -\frac{\sqrt{k+1} - \sqrt{k}}{4}\, e_3.$$

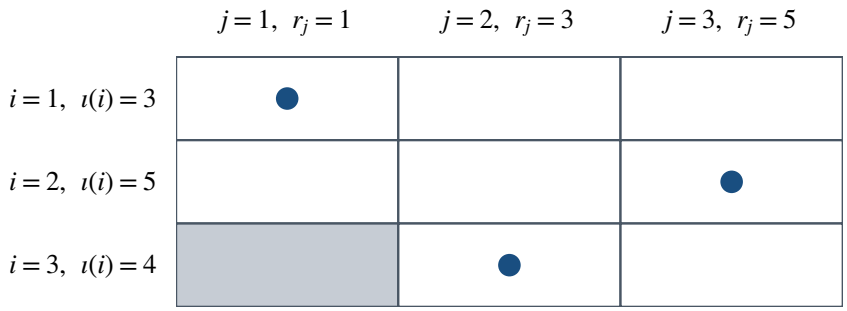

Figure 1: The example's allowed assignments. Columns show both reduced indices j and original ranks r_j; rows show both i and $\iota(i)$. The shaded cell is excluded by the second answer. Dots mark the generating assignment $\sigma = (1, 3, 2)$ (in the reduced coordinates), not the only compatible assignment.

These three numerical answers are distinct. Hence two vertices give the same answer exactly when they have the same first reduced column $\sigma(1)$.

After the answer, remove coordinate 3 and its disclosed rank k. The other query components are zero, so the old bounds $b_4 = 2$ and $b_5 = 1$ are unchanged. Recompute the lower bounds on list indices and sort the rows as before. In the table, the answer classes use the old reduced indices; the last four columns describe the updated state.

k	answer class	remaining original ranks	physical row order	ℓ^+	Δ^+
1	$\{(1,2,3),(1,3,2)\}$	$(3,5)$	$(4,5)$	$(1,1)$	1
3	$\{(2,1,3)\}$	$(1,5)$	$(5,4)$	$(1,2)$	0
5	$\{(3,1,2)\}$	$(1,3)$	$(5,4)$	$(1,2)$	0

For $k = 1$, both bounds select the first remaining rank 3, so both new index bounds are 1. The physical-coordinate tie rule puts coordinate 4 before 5: the new row order is not obtained merely by deleting the first entry of $(3, 5, 4)$. The updated family is $\mathscr{F}((1,1)) = \{(1,2),(2,1)\}$. For $k = 3$ or 5, coordinate 4 must receive the larger remaining rank. Thus $\mathscr{F}((1,2)) = \{(1,2)\}$: two coordinates remain inactive, but their ranks are determined by the constraints.

Under the conditional uniform law on the four vertices, the outcomes $k = 1, 3, 5$ have probabilities $1/2, 1/4, 1/4$. Consequently

$$\mathbb{E}\Delta^+ = \tfrac{1}{2}\cdot 1 + \tfrac{1}{4}\cdot 0 + \tfrac{1}{4}\cdot 0 = \tfrac{1}{2}, \qquad \mathbb{E}[\Delta - \Delta^+] = 2 - \tfrac{1}{2} = \tfrac{3}{2}.$$

Only the two directed edges between $(1, 2, 3)$ and $(1, 3, 2)$ stay within an answer class. The other six cross classes; averaging over the four vertices gives the same decrease $6/4 = 3/2$. The generating permutation has $k = 1$, so its realized decrease is 1, rather than the conditional average $3/2$.

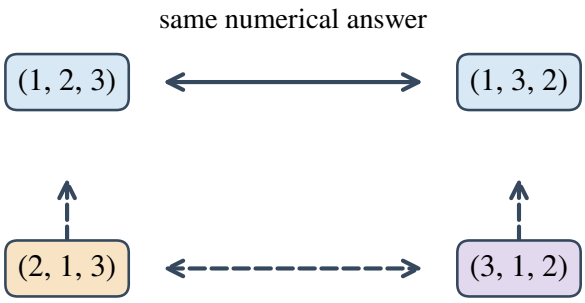

Figure 2: Vertices are the four compatible reduced permutations. Each double-headed connection represents two directed swap edges. Equal colors indicate equal numerical answers to $p = e_3$. Dashed connections cross answer classes; only the top pair remains together.