WHAT DOES ROBUST SELECTION REVEAL ABOUT THE ALTERNATIVES?

Anonymous authors
Paper under double-blind review

ABSTRACT

Choosing a good candidate need not require estimating every candidate well. We ask how much uncertainty can remain after a choice is already justified. We study unit directions whose quality is the squared positive part of their cosine similarity with an unknown vector. We observe that vector through exact linear measurements. A choice is robustly ε-optimal if it stays within ε of the best candidate for every nonzero vector consistent with the measured history. Once the selected direction has been measured, we derive the exact worst-case uncertainty in an alternative's quality, conditional on the selected quality θ. For $0 < \varepsilon \leq 1/2$, this diameter is $\theta + \varepsilon$ below $\theta = \varepsilon/(1+2\varepsilon)$ and $2\sqrt{\varepsilon(1+\varepsilon)\theta(1-\theta)}$ above it. Matching histories attain both branches. For an already valid selected confidence interval, we derive a location-sensitive bound on simultaneous intervals obtained from the same observations and coverage event. A complementary construction shows that, for $0 < \varepsilon \leq 1/10$, randomized adaptive continuation can require $\Omega(\min\{m, \varepsilon/\eta^2\})$ additional queries in the worst case, even with the exact norm and selected quality supplied. Here m is the unmeasured dimension, $\eta > 0$ is the target interval width, the failure probability is fixed below $1/2$, and $\eta/\sqrt{\varepsilon}$ is sufficiently small. These results quantify what robust selection reveals about unselected alternatives and what it can leave unmeasured.

1 INTRODUCTION

An auditor can have enough information to choose an update before knowing the quality of every proposed update. Once the choice is justified, a second question arises: how much can the same measurements already say about the alternatives? This question is relevant when a candidate must later be compared, screened, or reconsidered. A choice is one output; a quantitative description of all candidates is another. This complements the decision-focused question studied in linear best-arm identification (Soare et al., 2014; Fiez et al., 2019).

We study this distinction through retained-descent quality. For a unit direction u and a nonzero gradient g, let

$$\kappa(u; g) = \frac{[\langle g, u\rangle]_+^2}{\|g\|^2}. \tag{1}$$

Here $[a]_+ = \max\{a, 0\}$. The positive part permits a nonnegative step along the update $-u$. The normalization compares its best smoothness-based descent guarantee with that of the gradient direction. The auditor observes g through exact linear measurements. We count these scalar measurements, not execution time.

We require a selected candidate to remain within ε of the best one for *every* nonzero gradient consistent with all observed measurements. Thus no unmeasured possibility can overturn the requested decision accuracy. This robust stopping condition follows the regret-based elicitation principle (Boutilier et al., 2006; Bennouna et al., 2026) and links the decision to the uncertainty left in the measured history.

Our main finding is that the same selection error permits very different certification precision, depending on the selected quality. Suppose the selected direction has been measured and its quality is subsequently known to be θ. Over all compatible gradients, how far can an alternative's quality vary? We give an exact, attained answer. The diameter rises from ε at $\theta = 0$, reaches its maximum

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.

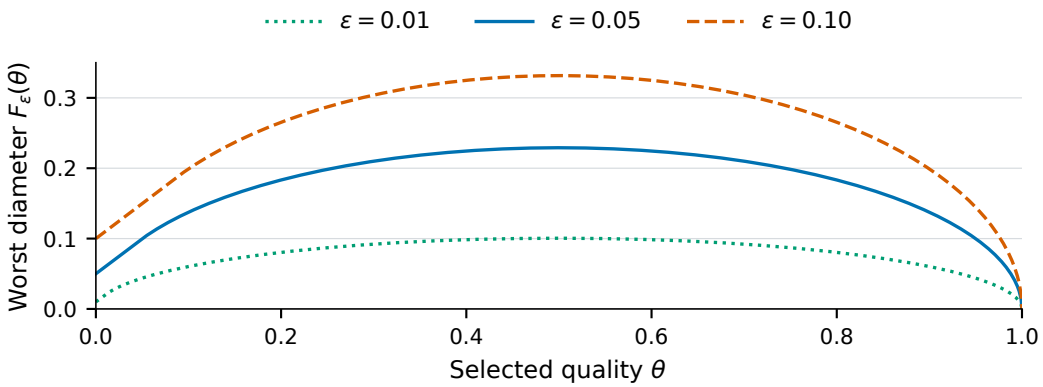

Figure 1: **The same selection accuracy leaves different uncertainty.** The curves show the exact worst compatible candidate-quality diameter $F_\varepsilon(\theta)$ from Theorem 3, conditional on the selected quality θ. They are analytic evaluations, not empirical confidence-interval widths. The selected direction is measured, and selection is valid on the full compatible fiber before conditioning on θ. The uncertainty is largest at intermediate selected quality.

at $\theta = 1/2$, and falls to zero at $\theta = 1$. Low selected quality forces every candidate to be poor; near-perfect alignment leaves little gradient energy unmeasured. Between these extremes, uncertainty can remain of order $\sqrt{\varepsilon}$.

Real audits usually supply an interval rather than an exact quality. As that quality varies, the alternatives' possible intervals move as well as widen. We derive explicit candidate hulls and a finite-interval width bound that accounts for both effects. A standard confidence-set mapping argument (Dufour & Taamouti, 2005) preserves the selected interval's original coverage event. The new quantitative question is how wide the images can be; robust selection provides the geometric constraint that controls this width.

Finally, we ask whether reuse can give arbitrarily fine intervals. We embed a vector-recovery problem in a valid stopping history of a globally valid selector. Even after granting the exact norm and selected quality, fine simultaneous output can require additional measurements. The construction shows how a justified selection can leave a difficult estimation problem.

The paper develops three linked results: an attained conditional precision law, its finite-interval certification consequence, and a continuation obstruction. The proofs are constructive and use the same exact scalar interface. The first two concern compatible ranges; the third permits randomized intervals to fail with a specified probability and therefore requires a separate argument.

2 Model: A Justified Choice and Its Measured History

Fix unit candidates $u_1, \dots, u_K \in \mathbb{R}^d$ before auditing and a nonzero unknown $g \in \mathbb{R}^d$. Write $\kappa_j(g) = \kappa(u_j; g)$. A query in direction v returns the exact scalar $\langle g, v \rangle$. Directions may depend measurably on earlier responses and an independent random seed. Each query direction is represented by a row of the measurement matrix. Every row is retained and every query is counted, even if redundant. The terminal history H includes rows used for selection and any subsequent selected-quality audit. Let W be their span and $p = P_W g$, which is determined by their responses. The nonzero gradients consistent with all responses form the compatible set

$$\mathcal{F}_H = (p + W^\perp) \setminus \{0\}. \tag{2}$$

We refer to this set as the *compatible fiber*.

Definition 1 (Robust selection)**.** The history H certifies candidate k to tolerance ε if

$$\max_j \kappa_j(g') - \kappa_k(g') \le \varepsilon \qquad \text{for every } g' \in \mathcal{F}_H. \tag{3}$$

The selected candidate is *anchored* if $u_k \in W$.

We use $0 \le \varepsilon \le 1/2$ unless a result specifies a smaller range. The fiber in (3) has no norm restriction. If the selected direction is not anchored, one additional query in that direction suffices. Adding this or any audit row shrinks the fiber and preserves (3). This anchors the currently selected candidate.

Why this score? If a differentiable loss has a β-Lipschitz gradient, then for $t \ge 0$ its standard quadratic upper bound along $-u$ decreases by $t\langle g, u\rangle - \beta t^2/2$. Optimizing this bound gives $[\langle g, u\rangle]_+^2/(2\beta)$. Its ratio to the corresponding gradient-direction value $\|g\|^2/(2\beta)$ is (1). The score therefore compares directions under their best nonnegative step sizes in this bound.

A checkable condition. Anchoring makes robust selection testable from the history. Put

$$E = \|p\|^2, \quad A = [\langle p, u_k\rangle]_+^2, \quad a_j = \langle p, u_j\rangle, \quad d_j = \|P_{W^\perp} u_j\|^2. \tag{4}$$

Proposition 2 (Robust stopping predicate)**.** *For $E > 0$ and $\varepsilon > 0$, an anchored history satisfies (3) if and only if, for every j,*

$$d_j \le \varepsilon, \qquad (A + \varepsilon E)(\varepsilon - d_j) \ge \varepsilon a_j^2 \quad \textit{whenever } a_j > 0. \tag{5}$$

For $\varepsilon = 0$, the equivalent conditions are $d_j = 0$ and $[a_j]_+^2 \le A$. If $E = 0$ and the fiber is nonempty, the condition is simply $d_j \le \varepsilon$ for all j.

The proof reduces the worst residual direction to a scalar quadratic. Equation (5) couples the measured and unmeasured candidate components; using only $d_j \le \varepsilon$ loses the precision dependence below.

When can selection stop early? The same condition explains which candidate geometries permit early selection. Let $M = \sum_{j=1}^K u_j u_j^\top$ have decreasing eigenvalues $\lambda_1, \lambda_2, \ldots$, and let $q = \dim W$. The anchored candidate has zero residual. The projection-trace bound and Proposition 2 therefore give

$$\sum_{\ell > q} \lambda_\ell \le \sum_{j=1}^K \|P_{W^\perp} u_j\|^2 \le (K-1)\varepsilon. \tag{6}$$

A slowly decaying candidate spectrum forces a large measured subspace, regardless of query order or responses. For orthogonal candidates, $q \ge K - (K-1)\varepsilon$: with $K = 32$ and $\varepsilon = 0.05$, an anchored history needs at least 31 independent measurement directions, or at least 30 before an optional anchor query. These are necessary dimension bounds; the response-dependent condition in Proposition 2 still determines whether a choice is certified.

By contrast, the candidates in Section 5 share two large components and have small individual residuals. Two queries certify a choice on an open set of responses while many residual directions remain unmeasured. Shared structure can thus permit early selection without determining every alternative precisely. The precision law below describes what can remain uncertain after such a choice. Appendix C.1 gives the short trace argument.

3 An Exact Law for the Uncertainty After Selection

First separate the geometry from the cost of statistical calibration. After robust selection and anchoring, supply the exact selected quality θ. Define the nonempty slice

$$\mathcal{F}_{H,\theta} = \{g' \in \mathcal{F}_H : \kappa_k(g') = \theta\}. \tag{7}$$

Its candidate-j diameter is $\sup_{g' \in \mathcal{F}_{H,\theta}} \kappa_j(g') - \inf_{g' \in \mathcal{F}_{H,\theta}} \kappa_j(g')$. Let $D_*(\varepsilon, \theta)$ be the supremum of these diameters over all finite unit-candidate sets, anchored histories satisfying (3), nonselected candidates, and ambient dimensions. A deterministic interval covering every compatible input must contain this candidate's range. This is a diameter-of-information question (Traub et al., 1988), restricted to histories that already certify a robust choice.

Theorem 3 (Sharp conditional precision)**.** *For $0 < \varepsilon \le 1/2$ and $0 \le \theta \le 1$, let $\tau_\varepsilon = \varepsilon/(1+2\varepsilon)$. Then*

$$D_*(\varepsilon, \theta) = F_\varepsilon(\theta) := \begin{cases} \theta + \varepsilon, & 0 \le \theta \le \tau_\varepsilon, \\ 2\sqrt{\varepsilon(1+\varepsilon)\theta(1-\theta)}, & \tau_\varepsilon \le \theta \le 1. \end{cases} \tag{8}$$

Two unit candidates in three dimensions and a valid two-query stopping history attain the bound. For $\varepsilon = 0$, $D_*(0, \theta) = 0$.

Proof idea. For $\theta > 0$, the slice fixes $T = \|g'\|^2 = A/\theta$. Let $e = E/T \in [\theta, 1]$, $x = a_j/\sqrt{T}$, and $t = \sqrt{d_j(1-e)}$. The candidate's endpoint qualities are $[x-t]_+^2$ and $[x+t]_+^2$. Writing $z = d_j/\varepsilon$, (5) bounds $x \le \sqrt{(\theta+\varepsilon e)(1-z)}$. The hinge-square width is nondecreasing in x. For nonnegative P, Q, the resulting scalar maximization is

$$\max_{0\le z\le 1}\left\{\left(\sqrt{P(1-z)}+\sqrt{Qz}\right)^2-\left[\sqrt{P(1-z)}-\sqrt{Qz}\right]_+^2\right\}=\begin{cases}P+Q, & P\le Q,\\ 2\sqrt{PQ}, & P\ge Q.\end{cases} \tag{9}$$

Set $P = \theta + \varepsilon e$ and $Q = \varepsilon(1-e)$, then optimize over e. This gives (8). The matching construction allocates candidate mass to two measured axes and one residual axis while satisfying (5) at equality. Outside the displayed history, its selector measures the remaining axis and selects exactly. Full details, including $\theta = 0, 1$ and $\varepsilon = 0$, appear in Appendix A.

What the curve says. At fixed ε, the largest diameter occurs at $\theta = 1/2$ and equals $\sqrt{\varepsilon(1+\varepsilon)}$. For example, $\varepsilon = 0.05$ gives diameters approximately 0.060, 0.229, and 0.046 at $\theta = 0.01, 0.50$, and 0.99. The low-quality and near-alignment regimes can be much narrower than a quality-independent square-root bound.

The theorem supplies θ *after* selection to isolate the geometry of the remaining uncertainty. If an audit obtains this value through measurements, those rows remain in H and the upper bound still applies. The next section instead starts from an acquired confidence interval and retains all measurements used to obtain it.

4 From one actual interval to simultaneous precision

Suppose an audit supplies $I_k = [L, U] \subseteq [0, 1]$ with $w = U - L$. For every fixed nonzero g, under its actual selection, stopping, and randomization protocol, assume

$$\Pr_g(E_I) \ge 1-\alpha, \qquad E_I = \{\kappa_k(g) \in I_k\}. \tag{10}$$

All rows used to acquire I_k are part of H. Add a missing anchor if necessary, keep I_k unchanged, and do not condition its coverage anew on the resulting history.

4.1 Compatible ranges and explicit endpoints

The interval and history define

$$C(H, I_k) = \{g' \in \mathcal{F}_H : \kappa_k(g') \in I_k\}. \tag{11}$$

For each candidate, take the smallest closed interval containing its quality values on this set, called its *interval hull*. This construction preserves simultaneous coverage on E_I: the true g belongs to the set on that event. This implication is standard confidence-set mapping and does not itself require robust selection. The role of (3) is to bound the resulting widths.

For $A > 0$ and $W^\perp \ne \{0\}$, put $h = A/E$, $V = \min\{U, h\}$, $\ell = L/h$, $v = V/h$, and $b_j = a_j/\sqrt{E}$. Compatibility is equivalent to $U > 0$ and $L \le h$. The selected quality ranges over $(0, h]$; zero is only a limit in this branch.

Proposition 4 (Candidate-specific interval hulls)**.** *For a compatible input in this branch, each nonselected candidate's exact closed interval hull $J_j = [l_j, r_j]$ is given by*

$$b_j > 0: \quad l_j = [b_j\sqrt{\ell} - \sqrt{d_j(1-\ell)}]_+^2, \quad t_j = \operatorname{clamp}\left(\frac{b_j^2}{b_j^2+d_j}, \ell, v\right),$$
$$r_j = \left(b_j\sqrt{t_j} + \sqrt{d_j(1-t_j)}\right)^2; \tag{12}$$

$$b_j \le 0: \quad l_j = 0, \qquad r_j = [b_j\sqrt{\ell} + \sqrt{d_j(1-\ell)}]_+^2. \tag{13}$$

Here $\operatorname{clamp}(x, a, b) = \min\{b, \max\{a, x\}\}$.

At a fixed selected quality $\theta = ht$, the endpoint expressions are $[b_j\sqrt{t} \pm \sqrt{d_j(1-t)}]_+^2$. The proposition optimizes these expressions over $t \in [\ell, v]$. With one unmeasured dimension, the attainable qualities need not fill the interval hull.

4.2 A LOCATION-SENSITIVE WIDTH BOUND

Define $G_0(L, V) = V - L$. For $0 < \varepsilon \le 1/2$, if $L < \tau_\varepsilon$ set

$$G_\varepsilon(L, V) = \min\{1, V + \varepsilon\}. \tag{14}$$

For $L \ge \tau_\varepsilon$, set

$$\begin{aligned} s_0 &= \frac{(1+\varepsilon)^2(1-L)}{(1+\varepsilon)^2(1-L) + \varepsilon^2 L}, & \widehat{s} &= \mathrm{clamp}(s_0, L, V), \\ S(L, s) &= \sqrt{s(1-L)} + \sqrt{L(1-s)}, \\ \Lambda_\varepsilon(L, s) &= \frac{s - L + \sqrt{(s-L)^2 + 4\varepsilon(1+\varepsilon)S(L,s)^2}}{2}, \\ G_\varepsilon(L, V) &= \min\{1, V + \varepsilon, \Lambda_\varepsilon(L, \widehat{s})\}. \end{aligned} \tag{15}$$

The formula accounts for both interval width and interval location.

Theorem 5 (Finite-interval reuse)**.** *Assume (3) holds at every output history of the protocol in (10). After anchoring, a deterministic postprocessor returns $J_1, \ldots, J_K$, with $J_k = I_k$, using no further query, norm estimate, or error-probability allocation. In the compatible positive-A branch, use Proposition 4 and define $B(H, I_k) = \max\{w, G_\varepsilon(L, V)\}$. For the remaining branches use the rules below. Then*

$$\Pr_g\left\{\kappa_j(g) \in J_j \ \forall j, \quad \max_j \mathrm{width}(J_j) \le B(H, I_k)\right\} \ge 1 - \alpha. \tag{16}$$

The remaining cases are as follows. At full *ambient* measured rank, the gradient is $p \ne 0$; if its selected quality lies in I_k, return exact points for the other candidates and set $B = w$. With a nontrivial unmeasured space and $A = 0$, compatibility is $L = 0$; return $[0, \varepsilon]$ for the other candidates and use $B = \max\{w, \varepsilon\}$. For an incompatible interval, flag incompatibility, return $[0, 1]$ for the others, and set $B = 1$. In every case retain I_k. The model excludes a history compatible only with $g = 0$. Incompatible intervals lie outside E_I.

Why this bound? For $\varepsilon > 0$ and $b_j > 0$, Proposition 2 gives

$$d_j \le \varepsilon, \qquad b_j^2 \le (h + \varepsilon)(1 - d_j/\varepsilon). \tag{17}$$

For a pair of selected-quality slices, the difference between a candidate's upper and lower values becomes a two-dimensional quadratic optimization over this ellipse. Its largest eigenvalue yields Λ_ε after maximizing over $h \le 1$. The remaining scalar maximum occurs at $\widehat{s}$. The elementary bounds 1 and $V + \varepsilon$ give the two other terms. Negative intercepts and clipped lower endpoints are treated explicitly in Appendix A.

The width bound $B(H, I_k)$ is computed from the observed history and input interval, so it can vary across runs. Equation (16) inherits the original fixed-g coverage guarantee; it does not condition that guarantee on the observed history.

Corollary 6 (Exact regimes)**.** *For $0 < \varepsilon \le 1/2$, the worst compatible nonselected diameter given $I_k = [0, U]$ is $\min\{1, U + \varepsilon\}$. If instead $L \ge \gamma_\varepsilon = (1+\varepsilon)/(1+2\varepsilon)$, it is*

$$2\sqrt{\varepsilon(1+\varepsilon)L(1-L)}, \tag{18}$$

independently of $U \ge L$. In the latter regime the bound is also at least $U - L$, so it covers the unchanged selected interval.

Here the worst case is over compatible histories as in Theorem 3, with the indicated interval supplied after selection. The general G_ε is an upper bound, not an exact diameter for every finite interval. The ellipse argument relaxes a unit-candidate constraint; the equality constructions restore it in the stated regimes. At $\varepsilon = 0.05$, the high-L regime begins at $21/22$, so it describes near-perfect alignment.

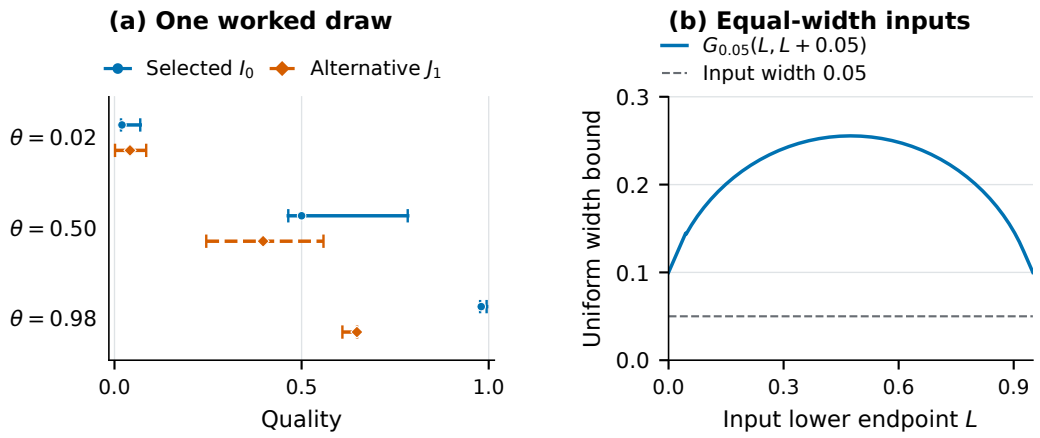

Figure 2: **From one calibrated interval to simultaneous precision.** (a) The 95% selected interval I_0 and the interval hull for the predesignated alternative u_1, after two selection and 16 calibration queries. Marks show the true qualities for reference. This is one reproducible worked draw, not an empirical coverage estimate. (b) Separate analytic bound $G_{0.05}(L, L+0.05)$ for equal-width inputs; this is a uniform bound, not the realized widths in (a).

Why a singleton substitution fails. Consider $W = \mathrm{span}(e_0, e_1)$, $p = e_0$, $u_k = e_0$, and the other candidate $(20e_0 + 15e_1 + e_2)/\sqrt{626}$. For $\varepsilon = 6/625$, the history strictly satisfies (5). With $I_k = [9/20, 11/20]$, the other hull has width

$$\frac{399 + 120\sqrt{11}}{6260} > 0.1, \qquad \sup_{\theta \in I_k} F_\varepsilon(\theta) = \frac{\sqrt{3786}}{625} < 0.1. \tag{19}$$

Even the maximum of the input width and all singleton diameters is too small. The locations of the slices move; their individual diameters cannot bound the diameter of their union.

4.3 A COMPLETE WORKED AUDIT

For a concrete use of Theorem 5, take $\varepsilon = 0.05$ and 65 unit candidates on orthonormal axes $e_0, e_*, e_1, \ldots, e_{64}$ in $\mathbb{R}^{66}$:

$$u_0 = e_0, \qquad u_j = \frac{4}{5}e_0 + \sqrt{\frac{211}{600}}e_* + \frac{1}{\sqrt{120}}e_j, \qquad 1 \le j \le 64. \tag{20}$$

The responses $(1, 0)$ to queries e_0, e_* certify u_0 by Proposition 2. This is the early-stop branch of the selector in Section 5, with 64 residual directions left.

Next make 16 fresh Gaussian queries in that residual space. If S is the sum of their squared responses and $q_{.025}, q_{.975}$ are the corresponding χ^2_{16} quantiles, the selected interval is

$$I_0 = \left[\frac{1}{1 + S/q_{.025}}, \frac{1}{1 + S/q_{.975}}\right]. \tag{21}$$

For every fixed input producing these selection responses, this interval has at least 95% coverage under the Gaussian protocol. All 18 query rows are then retained. Recompute W, p, E, a_j, d_j from all query rows and responses and apply Proposition 4, keeping I_0 unchanged. The resulting family inherits the same coverage event, without another query. The measured rank is 18 almost surely, leaving 48 unmeasured directions.

Figure 2 illustrates this calculation for three fixed inputs using one fixed probe draw. Inference uses only the query rows and responses; true qualities are used afterwards to annotate and check the outputs. Its raw selected intervals need not have equal widths. The separate analytic panel holds that width fixed to isolate the dependence of the uniform bound on interval location. Appendix C.2 specifies the inputs, calibration, and complete accounting.

5 FINER SIMULTANEOUS PRECISION CAN REQUIRE NEW INFORMATION

A compatible-range diameter constrains an interval that must contain *every* compatible input. It is not by itself a lower bound for a randomized procedure that may fail on different inputs. We next impose a fixed-input probability guarantee and bound the number of additional queries.

Theorem 7 (A statistical continuation obstruction)**.** *For every $m \geq 1$ and $0 < \varepsilon \leq 1/10$, there is a fixed candidate family and a globally valid robust selector with a two-query stopping history. At that history, even when the exact norm and selected quality are supplied, the following bound holds. Let a randomized adaptive continuation, with an input-independent seed and worst-case additional-query cap Q, return intervals satisfying, for every fixed compatible input,*

$$\Pr\{\textit{all intervals cover and all have width at most } \eta\} \geq 1 - \alpha, \tag{22}$$

where $\eta > 0$ and $0 < \alpha < 1/2$. There is a constant D_ε with $D_\varepsilon^2 \geq 117\varepsilon/2048$ such that, for each integer $1 \leq n \leq m$ with $n\eta^2 < 4D_\varepsilon^2$,

$$Q \geq (1 - 2\alpha)n\left(1 - \frac{n\eta^2}{4D_\varepsilon^2}\right). \tag{23}$$

In particular, if $n = \min\{m, \lfloor 117\varepsilon/(1024\eta^2)\rfloor\} \geq 1$, then $Q \geq (1 - 2\alpha)n/2$.

A valid stopping history. Use orthonormal axes $e_0, e_*, e_1, \ldots, e_m$ and unit candidates

$$u_0 = e_0, \qquad u_j = ce_0 + be_* + \delta e_j, \quad c = 4/5, \quad \delta^2 = \varepsilon/6, \quad b^2 = 9/25 - \delta^2. \tag{24}$$

Query e_0, e_*, obtaining x, z. If $x > 0$ and $|bz| < x/10$, select u_0; otherwise measure the remaining axes and select exactly. This entire open branch satisfies robust selection. Fix its two-query history $(x, z) = (1, 0)$. Even after supplying $\|g\|^2 = T = 64/25$ and $\kappa_0 = 25/64$, all

$$g = e_0 + vh, \qquad v = \sqrt{39}/5, \quad \|h\|_2 = 1, \qquad \kappa_j(g) = \frac{(c + \delta v h_j)^2}{T} \tag{25}$$

remain possible, with $h \in \operatorname{span}\{e_1, \ldots, e_m\}$. Each extra query is equivalent to one linear query of h.

Why randomization does not remove the obstruction. The coordinate-to-quality map has derivative at least $D_\varepsilon = 2\delta v(c - \delta v)/T > 0$. A simultaneous quality box of width η therefore corresponds to a coordinate set of diameter at most $\sqrt{n}\eta/D_\varepsilon$ on an n-dimensional subproblem. A seed-independent uniform sphere prior remains a residual sphere after adaptive exact queries. If its radius exceeds half that diameter, the box cannot contain both members of any antipodal pair and has conditional success probability at most $1/2$. After padding to $Q < n$ orthonormal measurement directions, the expected measured energy is Q/n; Markov's inequality then yields (23). The argument averages over inputs independently of the seed, so a violation identifies one fixed hard input for the whole randomized procedure. Appendix B gives the full proof and boundary cases.

For fixed α and sufficiently small $\eta/\sqrt{\varepsilon}$, the bound is $\Omega(\min\{m, \varepsilon/\eta^2\})$. The scale follows from coordinate recovery through a smooth transformation of size $\sqrt{\varepsilon}$. This connects the construction to uniform vector recovery (Kunsch & Wnuk, 2026), with the recovery problem remaining after a choice is justified on its full compatible fiber. The lower bound applies to additional queries at this history, with all earlier information retained.

6 RELATED WORK

Pure exploration from linear observations. Linear best-arm identification allocates noisy observations to identify the candidate with the largest linear reward (Soare et al., 2014). Fiez et al. (2019) allow the measurement vectors and candidate items to come from different sets. These works study information acquisition for selection. We study the quality ranges left after robust selection, using exact observations and the normalized positive-square score.

Decision sufficiency and regret-based elicitation. Bennouna et al. (2026) characterize sufficient linear observations for exact linear optimization and formulate generalized approximate sufficiency. Regret-based utility elicitation maintains a compatible parameter set and queries until a decision meets a regret tolerance (Boutilier et al., 2006). Our robust stopping condition uses this established principle. We analyze the quantitative quality ranges left for alternatives after the condition is met. Approximate regret for the normalized positive-square score is not preserved by simply replacing it with a linear ordering objective.

Projection and simultaneous inference. Confidence-set projection gives valid inference for functions of an underlying parameter; explicit quadratic projection geometry is studied by Dufour & Taamouti (2005). Simultaneous inference protects subsequent selection (Berk et al., 2013). In our construction, a selected interval and exact history define a compatible parameter set, and its images share the original coverage event. Robust selection is unnecessary for that event inclusion. It is needed for the precision bounds on those images, including the dependence on the selected interval's location.

Uncertainty and recovery from linear information. Information-based complexity describes uncertainty after partial observations through information radii and diameters (Traub et al., 1988). Here the robust selection condition constrains the compatible sets used in the precision law. Kunsch & Wnuk (2026) give adaptive randomized uniform vector approximation on the Euclidean unit ball, with fixed-input high-probability guarantees and double-logarithmic dimension dependence. Their preprint's Theorem 3.1 is the precise comparison for this information model. In the constructed family (24), quality is a monotone coordinate transform, so classical recovery difficulty supplies the statistical background. Adaptive linear-information lower bounds are also developed by Kunsch (2017). Our proof treats the fixed-confidence interval requirement directly. Its contribution is to place this recovery difficulty within a valid selection history, with all prior observations retained and exact normalization information supplied.

7 Discussion and Conclusion

The precision curve gives a concrete answer to a question that a selection guarantee alone leaves open. At low selected quality, robust selection forces every alternative to be weak. Near perfect alignment, little unmeasured energy remains. Intermediate qualities can leave substantially larger alternative ranges. An actual interval adds a second effect: compatible ranges shift as the selected quality changes. The finite-interval bound captures this movement without spending a new probability budget.

These statements separate three information tasks: justifying a choice, acquiring a selected-quality interval, and refining all candidate intervals. Our zero-additional-query result starts after the first two tasks and any missing anchor measurement. The continuation construction explains why a demand for finer simultaneous precision can remain a separate task.

The analysis assumes exact real-valued measurements and selection valid for every compatible gradient. Extending the precision laws to noisy measurements or selection guarantees that hold only with high probability is an open problem. The worked example provides one calibration protocol; a general analysis of calibration cost and the frequency of early stopping in natural candidate families remains for future work.

A justified choice reveals information about its alternatives. The selected quality and the location of its uncertainty interval determine how precisely those alternatives can already be described.

Reproducibility Statement

The appendices provide the full geometric and probabilistic arguments. The accompanying source includes the analytic figure scripts, the fixed protocol and recorded probe draw for Section 4.3, and all resulting intervals. The grids evaluate the proved formulas, and the recorded draw can be replayed without further random sampling.

AI USAGE STATEMENT

AI tools assisted with the theoretical formulation, derivation and checking of mathematical claims, proof writing, literature search, synthetic worked-example design and implementation, figures, and manuscript preparation. The authors are responsible for the final content, including the claims, proofs, citations, and artifacts.

REFERENCES

Omar Bennouna, Amine Bennouna, Saurabh Amin, and Asuman Ozdaglar. Data informativeness in linear optimization under uncertainty, 2026. URL https://arxiv.org/abs/2602.15365v1.

Richard Berk, Lawrence Brown, Andreas Buja, Kai Zhang, and Linda Zhao. Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837, 2013. doi: 10.1214/12-AOS1077.

Craig Boutilier, Relu Patrascu, Pascal Poupart, and Dale Schuurmans. Constraint-based optimization and utility elicitation using the minimax decision criterion. *Artificial Intelligence*, 170:686–713, 2006. doi: 10.1016/j.artint.2006.02.003.

Jean-Marie Dufour and Mohamed Taamouti. Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica*, 73(4):1351–1365, 2005. doi: 10.1111/j.1468-0262.2005.00618.x.

Tanner Fiez, Lalit Jain, Kevin Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/8ba6c657b03fc7c8dd4dff8e45defcd2-Abstract.html.

Robert J. Kunsch. High-dimensional function approximation: Breaking the curse with Monte Carlo methods, 2017. URL https://arxiv.org/abs/1704.08213v1.

Robert J. Kunsch and Marcin Wnuk. Uniform approximation of vectors using adaptive randomized information. *Journal of Approximation Theory*, 313:106216, 2026. doi: 10.1016/j.jat.2025.106216. URL https://arxiv.org/abs/2408.01098v1.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/hash/f8d84caae48546d0934d637ab54f7086-Abstract.html.

J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-Based Complexity*. Academic Press, 1988.

A PROOFS OF THE PRECISION RESULTS

This appendix proves the anchored stopping criterion, the conditional precision frontier, and the finite-interval bounds. All candidate directions are unit vectors. We retain every measured row, including rows used to obtain the selected confidence interval. Let W be their span, let $p = P_W g$, and write

$$\mathcal{F} = (p + W^\perp) \setminus \{0\}, \qquad \kappa_j(g') = \frac{[\langle g', u_j\rangle]_+^2}{\|g'\|^2}.$$

Throughout the proofs below, the selected direction u_k belongs to W. If it does not initially, measuring it once establishes this condition and shrinks the compatible set. In particular, a previously valid full-fiber selection guarantee remains valid after that measurement. We use

$$E = \|p\|^2, \qquad A = [\langle p, u_k\rangle]_+^2, \qquad a_j = \langle p, u_j\rangle, \qquad d_j = \|P_{W^\perp} u_j\|^2. \tag{26}$$

The residual quantity d_j is a squared norm.

The affine fiber is valid also for adaptive exact queries. To see this, fix the realized internal random seed and the transcript. Every g' satisfying all the recorded linear equations produces the first recorded response; inductively, it then produces the same next query and response. It therefore follows the same history. Conversely, following that history imposes all its linear equations, whose solution set is $p + W^\perp$. No norm constraint is introduced by this argument.

A.1 THE ANCHORED STOPPING CRITERION

We prove Proposition 2, including the sign and zero-energy cases. The worst compatible regret is

$$R_k = \sup_{g' \in \mathcal{F}} \left(\max_j \kappa_j(g') - \kappa_k(g') \right).$$

Since there are finitely many candidates and the selected candidate is among them, we can write $R_k = \max_j L_j$, where

$$L_j = \max\left\{0, \sup_{g' \in \mathcal{F}} (\kappa_j(g') - \kappa_k(g'))\right\}.$$

The nonnegative envelope avoids irrelevant negative pairwise suprema.

Suppose first that $E > 0$, and fix a candidate. Abbreviate $a = a_j$, $d = d_j$, and $v = P_{W^\perp} u_j$. If $d > 0$, decompose an arbitrary residual as $z = tv/\sqrt{d} + z_\perp$, where $z_\perp \perp v$. The pairwise regret is

$$\frac{[a + \sqrt{d}\, t]_+^2 - A}{E + t^2 + \|z_\perp\|^2}.$$

For a positive numerator, setting $z_\perp = 0$ can only increase this ratio. A nonpositive numerator cannot increase its nonnegative envelope. Thus it suffices to consider

$$\frac{[a + \sqrt{d}\, t]_+^2 - A}{E + t^2}, \qquad t \in \mathbb{R}, \tag{27}$$

together with zero. If $d = 0$, the numerator is constant and its positive part is maximized at $z = 0$, so $L_j = \max\{0, ([a]_+^2 - A)/E\}$.

If $a \le 0$ and $d > 0$, an active positive part requires $t \ge -a/\sqrt{d} \ge 0$. On that branch, subtracting $d(E + t^2)$ from the numerator of (27) gives

$$a^2 - A - dE + 2a\sqrt{d}\, t \le -a^2 - A - dE \le 0.$$

The inactive branch is nonpositive, and the ratio tends to d as $t \to +\infty$. Therefore $L_j = d$ when $a \le 0$, including $d = 0$. The supremum need not be attained at a finite gradient.

If $a > 0$ and $d > 0$, the ratio without the positive-part truncation is the Rayleigh quotient, on vectors $(1, t/\sqrt{E})$, of

$$M_j = \begin{pmatrix} (a^2 - A)/E & a\sqrt{d/E} \\ a\sqrt{d/E} & d \end{pmatrix}. \tag{28}$$

Its off-diagonal entries are strictly positive. A top eigenvector has both coordinates strictly positive, so its corresponding t is positive and the positive part is active. The untruncated maximum is therefore attainable in (27). Including the already established $d = 0$ case,

$$L_j = \begin{cases} d_j, & a_j \le 0, \\ \dfrac{B_j + d_j + \sqrt{(B_j - d_j)^2 + 4a_j^2 d_j/E}}{2}, & a_j > 0, \end{cases} \qquad B_j = \frac{a_j^2 - A}{E}. \tag{29}$$

In particular, squaring a negative a_j in the second branch would not be valid: its maximizing eigenvector could deactivate the positive part.

For $a_j > 0$ and $\varepsilon > 0$, the condition $L_j \le \varepsilon$ is equivalent to $\varepsilon I - M_j$ being positive semidefinite. Its determinant, multiplied by E, is

$$(\varepsilon E - a_j^2 + A)(\varepsilon - d_j) - a_j^2 d_j = (A + \varepsilon E)(\varepsilon - d_j) - \varepsilon a_j^2.$$

Consequently the equivalent scalar conditions are

$$d_j \le \varepsilon, \qquad (A + \varepsilon E)(\varepsilon - d_j) \ge \varepsilon a_j^2 \quad \text{if } a_j > 0. \tag{30}$$

For completeness, if $d_j < \varepsilon$, the determinant condition also forces the other diagonal entry to be nonnegative. If $d_j = \varepsilon > 0$, that condition cannot hold for $a_j > 0$. Thus no positive-semidefiniteness condition was dropped. When $a_j \le 0$, only $d_j \le \varepsilon$ is needed.

At $\varepsilon = 0$, (29) instead yields

$$d_j = 0, \qquad [a_j]_+^2 \le A \quad \text{for every } j. \tag{31}$$

This case must be stated separately: substituting zero into (30) would lose the measured-score comparison.

If $E = 0$, then $p = 0$, $A = 0$, and every $a_j = 0$. Provided $W^\perp$ is nontrivial, the selected quality is identically zero and the largest possible candidate quality is d_j, attained along its positive residual direction when $d_j > 0$. Thus $R_k = \max_j d_j$. If W is full dimensional and $E = 0$, there is no nonzero compatible gradient; this is an invalid input. At full measured rank with $E > 0$, all d_j vanish and the formulas reduce to the directly known regret at $g' = p$. This completes the proof of the stopping criterion.

A.2 Exact quality slices and finite-interval hulls

We next prove Proposition 4. Assume $A > 0$ and that $W^\perp$ is nontrivial, and put

$$h = A/E \in (0, 1], \qquad b_j = a_j/\sqrt{E}.$$

A selected quality $\theta > 0$ fixes the total squared norm to A/θ. It is compatible with this history precisely when $0 < \theta \le h$. Writing $t = \theta/h$, the residual norm is $\sqrt{E(1/t - 1)}$. Cauchy–Schwarz, with equality in the two opposite candidate-residual directions, gives the extreme qualities

$$\left[b_j\sqrt{t} - \sqrt{d_j(1-t)}\right]_+^2, \qquad \left[b_j\sqrt{t} + \sqrt{d_j(1-t)}\right]_+^2. \tag{32}$$

If $d_j = 0$, any residual direction of the required norm gives the same quality. If $t = 1$, the residual is zero. Thus the formula includes these degeneracies. A one-dimensional residual sphere can contain only two points; we assert the interval hull of its image, not that the image fills the interval.

For $I_k = [L, U]$, compatibility is $U > 0$ and $L \le h$. Let $V = \min\{U, h\}$, $\ell = L/h$, and $v = V/h$. The value $t = 0$ is used only as a limit in taking a closed hull; $A > 0$ never permits a finite compatible gradient with selected quality zero. In particular, $I_k = \{0\}$ is incompatible in this branch.

When $b_j > 0$, the signed lower expression in (32) is nondecreasing in t. The lower hull endpoint is therefore its positive-part square at ℓ. The signed upper expression is nonnegative and has its maximum at $b_j^2/(b_j^2 + d_j)$ before restriction to $[\ell, v]$. This follows by differentiation, or by maximizing

the inner product of $(b_j, \sqrt{d_j})$ with $(\sqrt{t}, \sqrt{1-t})$. Consequently its constrained maximum occurs at

$$\widehat{t}_j = \min\left\{v, \max\left\{\ell, \frac{b_j^2}{b_j^2 + d_j}\right\}\right\}.$$

When $b_j \le 0$, the lower endpoint is zero and the signed upper expression is nonincreasing in t, so its positive-part square is largest at ℓ. This proves all the stated candidate-specific hull formulas.

The remaining branches do not require division by A. If $A = 0$, the selected quality is identically zero; compatibility requires $L = 0$. Robust selection then puts every other quality in $[0, \varepsilon]$. This is a safe interval, not necessarily the smallest transcript-specific hull. At full measured rank and $E > 0$, check the single selected value A/E against I_k and return the exact other values. An incompatible supplied interval cannot cover the true selected quality. Returning $[0, 1]$ for the other candidates defines an output on such inputs, but does not turn them into successful joint certificates.

A.3 The sharp singleton frontier

We prove Theorem 3. The diameter here is the difference between the supremum and infimum over all compatible gradients with a specified selected quality. Its worst case additionally ranges over admissible candidate sets and histories. It is a deterministic compatibility quantity; no statement about the cost of acquiring a singleton interval is needed in its definition.

Upper bound. If $\theta = 0$, robust selection immediately bounds every candidate quality by ε, giving diameter at most ε. Suppose $\theta > 0$, and write

$$T = A/\theta, \quad e = E/T \in [\theta, 1], \quad x = a_j/\sqrt{T}, \quad t = \sqrt{d_j(1-e)}.$$

By (32), the candidate diameter is $[x+t]_+^2 - [x-t]_+^2$. This expression is nondecreasing in x for fixed $t \ge 0$. For $\varepsilon > 0$, set $z = d_j/\varepsilon \in [0, 1]$ and define

$$P = \theta + \varepsilon e, \qquad Q = \varepsilon(1-e).$$

The stopping criterion implies $x \le \sqrt{P(1-z)}$: this follows from (30) when $a_j > 0$, and is automatic when $a_j \le 0$. Also $t = \sqrt{Qz}$. Hence the diameter is at most

$$H(z) = \left(\sqrt{P(1-z)} + \sqrt{Qz}\right)^2 - \left[\sqrt{P(1-z)} - \sqrt{Qz}\right]_+^2. \tag{33}$$

For $P, Q \ge 0$ the exact scalar maximum is

$$\max_{0 \le z \le 1} H(z) = \begin{cases} P + Q, & P \le Q, \\ 2\sqrt{PQ}, & P \ge Q. \end{cases} \tag{34}$$

Indeed, if $P \le Q$, Cauchy–Schwarz bounds the first square by $P+Q$. Taking $z = Q/(P+Q)$ attains that square and makes the subtracted positive part zero; the zero-sum case is immediate. If $P \ge Q$, put $X = \sqrt{P(1-z)}$ and $Y = \sqrt{Qz}$. In the sector $X \ge Y$, the expression is $4XY \le 2\sqrt{PQ}$, with equality at $z = 1/2$. In the other sector, $z > P/(P+Q) \ge 1/2$. The first square is decreasing beyond its maximizer $Q/(P+Q)$, so its supremum in this sector is its value at $X = Y$, namely $4PQ/(P+Q) \le 2\sqrt{PQ}$. Cases with $PQ = 0$ follow directly.

Let $\tau = \varepsilon/(1+2\varepsilon)$. If $\theta \le \tau$, the robust selection inequality itself gives diameter at most $\theta + \varepsilon$. If $\theta \ge \tau$, then for every $e \ge \theta$,

$$P - Q = \theta - \varepsilon + 2\varepsilon e \ge 0.$$

Equation (34) now bounds the diameter by $2\sqrt{\varepsilon(1-e)(\theta+\varepsilon e)}$. The derivative of $(1-e)(\theta+\varepsilon e)$ is $\varepsilon - \theta - 2\varepsilon e$, which is nonpositive on $e \ge \theta$ in this regime. Its maximum is at $e = \theta$, giving

$$F_\varepsilon(\theta) = \begin{cases} \theta + \varepsilon, & 0 \le \theta \le \tau, \\ 2\sqrt{\varepsilon(1+\varepsilon)\theta(1-\theta)}, & \tau \le \theta \le 1. \end{cases} \tag{35}$$

The branches agree at τ. If $\varepsilon = 0$, the anchored exact stopping condition makes all candidate residuals zero. A positive selected quality then fixes the common normalization, and a zero selected quality forces all qualities to be zero. Every conditional diameter is therefore zero.

Unit candidates attaining the bound. Fix $0 < \varepsilon \le 1/2$ and $0 \le \theta \le 1$. In $\mathbb{R}^3$, use orthonormal axes e_0, e_1, e_2, and define the scalar parameters

$$e = \max\left\{\theta, \frac{\varepsilon - \theta}{2\varepsilon}\right\}, \quad d = \frac{\varepsilon}{2}, \quad x = \sqrt{\frac{\theta + \varepsilon e}{2}}, \quad B = 1 - \varepsilon - \frac{\theta}{2e}.$$

Here $0 < e \le 1$, $\theta/e \le 1$, and $B \ge 1/2 - \varepsilon \ge 0$. Commit

$$u_k = \sqrt{\theta/e}\, e_0 + \sqrt{1 - \theta/e}\, e_1, \qquad u_j = \frac{x}{\sqrt{e}}\, e_0 + \sqrt{B}\, e_1 + \sqrt{d}\, e_2. \tag{36}$$

Both are unit vectors, since $x^2/e + B + d = 1$. Query the two fixed axes e_0, e_1. At responses $(\sqrt{e}, 0)$, one has $E = e$, $A = \theta$, $a_j = x > 0$, and residual squared norm d. In particular,

$$(A + \varepsilon E)(\varepsilon - d) = \varepsilon x^2,$$

so the full-fiber stopping predicate holds at equality.

This history is produced by a globally correct selector. On every input, first query e_0, e_1. Stop with k if its exact predicate passes; otherwise query e_2 and choose an exact winner. The fallback uses all three coordinates and thus guarantees correctness outside the displayed history as well. The candidate set and query rule are fixed before the unknown input.

After selection, supplying θ leaves the two gradients

$$g_+ = \sqrt{e}\, e_0 + \sqrt{1 - e}\, e_2, \qquad g_- = \sqrt{e}\, e_0 - \sqrt{1 - e}\, e_2$$

compatible. Both have norm one and selected quality θ. If $\theta \le \tau$, then $e = (\varepsilon - \theta)/(2\varepsilon)$ and $x^2 = d(1 - e) = (\varepsilon + \theta)/4$. Their competitor qualities are zero and $\theta + \varepsilon$. If $\theta \ge \tau$, then $e = \theta$ and $x^2 = \theta(1 + \varepsilon)/2 \ge d(1 - \theta)$, so the difference between the competitor qualities is $4\sqrt{x^2 d(1 - \theta)} = F_\varepsilon(\theta)$.

At $\theta = 0$, this construction has $e = 1/2$ and $u_k = e_1$; the selected quality is zero on the entire fiber, and the two competitor values are 0 and ε. Thus no division by θ is used in that case. At $\theta = 1$, the two gradients coincide and the frontier is zero. For $\varepsilon = 0$, zero diameter is already both an upper bound and an attained value at every nonempty conditional fiber. These arguments prove sharpness on the entire stated domain. The exact selected quality was supplied only after selection; the construction does not assert that acquiring it is free or that the equality history has positive probability.

A.4 The location-sensitive finite-interval upper bound

We now prove the precision part of Theorem 5. The candidate hulls and the probabilistic event will be combined at the end. It suffices first to consider $\varepsilon > 0$ and the compatible $A > 0$ branch with a nontrivial residual space. The case $\varepsilon = 0$ is treated separately below. For a candidate with $b_j > 0$, write $b = b_j$ and $d = d_j$. The stopping predicate gives

$$0 \le d \le \varepsilon, \qquad b^2 \le (h + \varepsilon)(1 - d/\varepsilon). \tag{37}$$

For this upper bound only, we relax the additional unit-vector constraint $b^2 + d \le 1$. Parameterize the relaxed ellipse by $b = \sqrt{h + \varepsilon}\, c$ and $\sqrt{d} = \sqrt{\varepsilon}\, q$, where $c, q \ge 0$ and $c^2 + q^2 \le 1$. If $b_j \le 0$, its lower endpoint is zero, and replacing b_j by zero enlarges its entire hull. That replacement is included in the same ellipse because $d_j \le \varepsilon$.

The lower endpoint of a nonnegative-intercept hull occurs at L. For an upper slice $s \in [L, V]$, put

$$v_s = \begin{pmatrix} \sqrt{(1 + \varepsilon/h)s} \\ \sqrt{\varepsilon(1 - s/h)} \end{pmatrix}, \qquad w_L = \begin{pmatrix} \sqrt{(1 + \varepsilon/h)L} \\ -\sqrt{\varepsilon(1 - L/h)} \end{pmatrix}.$$

With $z = (c, q)^\top$, the difference of endpoints is $(v_s^\top z)^2 - [w_L^\top z]_+^2$. On the sector $w_L^\top z \ge 0$, this is the quadratic form of $v_s v_s^\top - w_L w_L^\top$. Its trace is $s - L$ and its determinant is $-\varepsilon C_h$, where

$$C_h = (1 + \varepsilon/h)\left(\sqrt{s(1 - L/h)} + \sqrt{L(1 - s/h)}\right)^2. \tag{38}$$

Its largest eigenvalue is therefore

$$\lambda_h(L, s) = \frac{s - L + \sqrt{(s - L)^2 + 4\varepsilon C_h}}{2}. \tag{39}$$

The clipped lower sector. The same eigenvalue bounds the sector $w_L^\top z < 0$ when $L \geq \tau$. Indeed, the unconstrained maximizer of the upper square on the first-quadrant unit circle is parallel to v_s. It lies in the nonnegative-lower sector if and only if

$$(h+\varepsilon)\sqrt{sL} \geq \varepsilon\sqrt{(h-s)(h-L)}. \tag{40}$$

The left side increases with s and the right side decreases, so it suffices to check $s = L$. That condition is $L \geq \varepsilon h/(h+2\varepsilon)$, which follows from $h \leq 1$ and $L \geq \varepsilon/(1+2\varepsilon) = \tau$. Along the quadrant arc beyond the upper-square maximizer, the upper square decreases. Its maximum in the negative-lower sector is consequently reached at the sector boundary. There $w_L^\top z = 0$, so the hinge expression equals the untruncated quadratic form and is bounded by (39). Homogeneity handles vectors inside the unit disk, and continuity includes sector boundaries and degenerate endpoints.

Removing dependence on the measured fraction. For $a \leq h \leq 1$, define $f_a(h) = (1+\varepsilon/h)(1-a/h)$. Equation (38) can be written $C_h = (\sqrt{sf_L(h)} + \sqrt{Lf_s(h)})^2$. If $a \geq \tau$, then

$$f_a'(h) = \frac{(a-\varepsilon)h + 2\varepsilon a}{h^3} \geq 0.$$

For $a \geq \varepsilon$ this is immediate. Otherwise the numerator is smallest at $h = 1$ and equals $a(1+2\varepsilon)-\varepsilon \geq 0$. Both terms in C_h are therefore nondecreasing for $h \in [s, 1]$, giving $C_h \leq C_1$. Define

$$S(L,s) = \sqrt{s(1-L)} + \sqrt{L(1-s)}, \qquad \lambda(L,s) = \frac{s - L + \sqrt{(s-L)^2 + 4\varepsilon(1+\varepsilon)S(L,s)^2}}{2}.$$

Every candidate width is bounded by $\max_{L \leq s \leq V} \lambda(L,s)$ when $L \geq \tau$. Independently, unit quality bounds the width by one, and robust selection bounds every upper endpoint by $V + \varepsilon$. These last two bounds suffice also when $L < \tau$.

Closed scalar maximization. We give the stationary-point argument to justify the clipped maximizer. Assume $\varepsilon > 0$ and $0 < L < 1$, and set

$$a_s = \begin{pmatrix} \sqrt{s} \\ \sqrt{1-s} \end{pmatrix}, \qquad l = \begin{pmatrix} \sqrt{L} \\ -\sqrt{1-L} \end{pmatrix}, \qquad D = \begin{pmatrix} \sqrt{1+\varepsilon} & 0 \\ 0 & \sqrt{\varepsilon} \end{pmatrix}.$$

The matrix $M(s) = D(a_s a_s^\top - ll^\top)D$ has top eigenvalue $\lambda(L,s)$. For $L < s < 1$, its off-diagonal entry is positive and its determinant is negative. The top eigenvalue is simple and positive, with a positive unit eigenvector z. Differentiating that eigenvalue gives

$$\frac{\partial \lambda}{\partial s} = 2(a_s^\top Dz)(a_s'^\top Dz).$$

The first factor is positive. Since a_s' is perpendicular to a_s, an interior stationary point requires $Dz = ta_s$ for some $t > 0$. Define $N = D(I - ll^\top)D$. The alignment identity then implies

$$Nz = M(s)z + D(I - a_s a_s^\top)Dz = M(s)z = \lambda z.$$

The matrix N is rank-one positive semidefinite. Its only positive eigenvalue is $1 + \varepsilon - L$, with positive eigenvector proportional to $(\sqrt{(1+\varepsilon)(1-L)}, \sqrt{\varepsilon L})^\top$. Alignment therefore forces

$$s = s_0 = \frac{(1+\varepsilon)^2(1-L)}{(1+\varepsilon)^2(1-L) + \varepsilon^2 L}. \tag{41}$$

Conversely, if $s_0 \in (L, 1)$, this vector satisfies the alignment identity and is the positive eigenvector of $M(s_0)$, so it gives a stationary point. There is no other interior stationary point.

The derivative at $s = L$ has the sign of

$$\sqrt{L(1-L)} + \sqrt{\varepsilon(1+\varepsilon)}(1-2L).$$

It is positive below $\gamma = (1+\varepsilon)/(1+2\varepsilon)$, zero at γ, and negative above it. The derivative near $s = 1$ is negative. Also $s_0 \geq L$ is equivalent to $L \leq \gamma$. It follows that the maximum over $[L, V]$ occurs at $\widehat{s} = \min\{V, \max\{L, s_0\}\}$. If $L = 1$, necessarily $V = 1$, and the width is zero directly. Combining these facts gives the stated bound

$$G_\varepsilon(L,V) = \begin{cases} \min\{1, V+\varepsilon\}, & L < \tau, \\ \min\{1, V+\varepsilon, \lambda(L,\widehat{s})\}, & L \geq \tau. \end{cases} \tag{42}$$

No unit-feasibility claim follows from the relaxed ellipse optimization. This is a general upper bound, not an assertion of equality for every finite interval.

For $\varepsilon = 0$, (31) gives $d_j = 0$ and $[a_j]_+^2 \le A$. A candidate quality is then $([a_j]_+^2/A)\theta$, with multiplier in $[0, 1]$. Its compatible width is at most $V - L$, proving the separate definition $G_0(L, V) = V - L$.

A.5 The finite-interval regimes with equality

We prove Corollary 6. Equality here is worst-case conditional compatibility diameter, not a claim about the probability or cost of acquiring the supplied interval.

Intervals starting at zero. The upper bound is $\min\{1, U + \varepsilon\}$, including the $A = 0$ branch. For $U > 0$ and $0 \le \varepsilon \le 1/2$, put $h = \min\{U, 1 - \varepsilon\} > 0$ and use in $\mathbb{R}^3$

$$W = \operatorname{span}\{e_0, e_1\}, \quad p = e_0, \quad u_k = \sqrt{h}\, e_0 + \sqrt{1 - h}\, e_1, \quad u_j = \sqrt{h + \varepsilon}\, e_0 + \sqrt{1 - h - \varepsilon}\, e_1.$$

These unit candidates have no residual component. At the two paid responses $(1, 0)$, a compatible gradient is $g' = e_0 + te_2$ and the regret is $\varepsilon/(1 + t^2) \le \varepsilon$. The selected quality ranges over $(0, h]$, so $[0, U]$ is valid throughout this fiber. The competitor quality ranges over $(0, h + \varepsilon]$, whose closed hull has width $h + \varepsilon = \min\{1, U + \varepsilon\}$. The zero endpoint is a limit, which suffices for an exact closed interval hull. A globally correct selector uses the same two fixed queries, checks the stopping predicate, and queries the remaining axis for exact selection if the predicate fails. When $U = 0$ and $\varepsilon > 0$, the $\theta = 0$ construction in (36) gives diameter ε and identically zero selected quality. When also $\varepsilon = 0$, all candidate qualities must be zero and the claimed bound is attained trivially.

High lower endpoints. Let $0 < \varepsilon \le 1/2$ and $L \ge \gamma$. Equation (41) gives $s_0 \le L$, so the bound simplifies to

$$G_\varepsilon(L, V) = 2\sqrt{\varepsilon(1 + \varepsilon)L(1 - L)}. \tag{43}$$

To show equality, take the singleton construction at $\theta = L$. It has $h = 1$, $b_j^2 = (1 + \varepsilon)/2$, $d_j = \varepsilon/2$, and a remaining measured coefficient with square $1/2 - \varepsilon \ge 0$. The upper slice of this candidate peaks at $b_j^2/(b_j^2 + d_j) = \gamma \le L$, while the lower slice increases. Therefore enlarging the supplied set of selected qualities from $\{L\}$ to $[L, U]$ does not change either endpoint of its hull. It attains (43), and is a feasible unit candidate, so the independent caps 1 and $V + \varepsilon$ cannot reduce that expression. For $L = 1$, both sides vanish.

The unchanged selected interval is also covered by this family-width bound. Indeed, $U - L \le 1 - L$ and, for $L \ge \gamma$, $1 - L \le 2\sqrt{\varepsilon(1 + \varepsilon)L(1 - L)}$. For $L < 1$, squaring the latter inequality reduces to $1 - L \le 4\varepsilon(1 + \varepsilon)L$, which follows from $(1 - L)/L \le \varepsilon/(1 + \varepsilon)$; the case $L = 1$ is immediate. This proves the named finite equality regime without claiming sharpness elsewhere.

A.6 Completing the simultaneous reuse statement

Let $\mathcal{E}_I = \{\kappa_k(g) \in I_k\}$ be the original valid selected-interval event. On this event the actual gradient is compatible with the complete retained history and the supplied interval. It therefore belongs to every set whose candidate image defines a hull. All nonselected hulls cover on this same event, and the selected interval remains literally $J_k = I_k$. The deterministic width calculations above give

$$\max_j \operatorname{width}(J_j) \le \max\{U - L, G_\varepsilon(L, V)\}$$

in the compatible positive-A branch. The zero-A branch instead uses $\max\{U - L, \varepsilon\}$, and the compatible full-rank branch uses $U - L$. An incompatible interval is outside $\mathcal{E}_I$; assigning the other outputs $[0, 1]$ there gives the trivial family-width bound one.

Thus the joint event of coverage and the corresponding observable width bound contains $\mathcal{E}_I$. For each fixed nonzero gradient its probability is at least the original $1 - \alpha$. No independence between the interval and the history is used, and no new probability allocation or norm event is introduced. The conclusion is unconditional under the original selection and acquisition protocol, not a fresh conditional-coverage claim at each realized history. The set-image coverage argument itself is general; robust selection is what supplies the quantitative width bounds proved above.

B PROOF OF THE CONTINUATION BOUND

We prove Theorem 7. Throughout this section, $m \geq 1$, $0 < \varepsilon \leq 1/10$, $0 < \alpha < 1/2$, and $\eta > 0$. The continuation algorithm may use arbitrary measurable adaptive linear queries and internal randomization. Its query budget Q is a worst-case cap on the number of additional queries.

B.1 A VALID SELECTION HISTORY

Let $e_0, e_*, e_1, \ldots, e_m$ be orthonormal axes in $\mathbb{R}^{m+2}$ and set

$$\begin{aligned} c = \frac{4}{5}, \qquad \delta = \sqrt{\frac{\varepsilon}{6}}, \qquad b = \sqrt{\frac{9}{25} - \delta^2}, \\ u_0 = e_0, \qquad u_j = ce_0 + be_* + \delta e_j \quad (1 \leq j \leq m). \end{aligned} \tag{44}$$

All candidates have unit norm. Write an arbitrary nonzero gradient as $g = xe_0 + ze_* + \sum_{j=1}^{m} y_j e_j$. The selector first measures x and z. It returns u_0 when

$$x > 0, \qquad |bz| < \frac{x}{10}. \tag{45}$$

Otherwise, it measures all remaining coordinates and returns an exact maximizer of $\kappa_j(g)$, with a fixed rule for ties.

To verify the early-stop branch, fix any responses satisfying (45). For every unmeasured vector y,

$$\langle g, u_j \rangle = cx + bz + \delta y_j \leq \frac{9}{10}x + \delta y_j. \tag{46}$$

If the left side is nonpositive, then $\kappa_j(g) - \kappa_0(g) \leq 0$. Otherwise, both sides of (46) are positive. Weighted Cauchy–Schwarz gives

$$\left(\frac{9}{10}x + \delta y_j\right)^2 \leq x^2 + \frac{\delta^2}{1 - (9/10)^2} y_j^2. \tag{47}$$

Since the selected numerator is x^2, it follows that

$$\kappa_j(g) - \kappa_0(g) \leq \frac{\delta^2}{19/100} \frac{y_j^2}{\|g\|^2} \leq \frac{50}{57}\varepsilon < \varepsilon. \tag{48}$$

Thus selection is ε-accurate on the entire affine fiber of every early-stop history, without a norm restriction. The fallback is exact, so the selector is globally valid. The branch is open and contains the history $(x, z) = (1, 0)$.

Fix this history and retain both measured rows and responses. After selection, give the continuation algorithm the exact squared norm $T = 64/25$ and the exact selected quality $\theta_0 = 25/64$. The compatible gradients are

$$\begin{aligned} g = e_0 + vh, \qquad v = \frac{\sqrt{39}}{5}, \\ h \in \operatorname{span}\{e_1, \ldots, e_m\}, \qquad \|h\| = 1. \end{aligned} \tag{49}$$

The grants are made only after the globally valid selection decision. Every gradient in (49) gives the same selection history, squared norm, and selected quality.

For $t \in [-1, 1]$, define

$$f_\varepsilon(t) = \frac{(c + \delta v t)^2}{T}, \qquad D_\varepsilon = \frac{2\delta v(c - \delta v)}{T}. \tag{50}$$

Here $\delta v < c$, so no positive-part truncation is needed and $\kappa_j(g) = f_\varepsilon(h_j)$ for $j \geq 1$. Moreover, $f'_\varepsilon(t) \geq D_\varepsilon > 0$ throughout $[-1, 1]$. Consequently,

$$|f_\varepsilon(s) - f_\varepsilon(t)| \geq D_\varepsilon |s - t| \quad \text{for all } s, t \in [-1, 1]. \tag{51}$$

B.2 Adaptive queries under a seed-independent prior

Let Ξ contain all randomization used by the continuation algorithm, independently of its input. Let $I_0, \dots, I_m$ be its output intervals and write

$$\mathcal{E}(h, \Xi) = \big\{\kappa_j(e_0 + vh) \in I_j, \quad \text{width}(I_j) \le \eta \\ \text{for all } 0 \le j \le m\big\}. \tag{52}$$

The assumed guarantee is pointwise in the input:

$$\mathbb{P}_\Xi\{\mathcal{E}(h, \Xi)\} \ge 1 - \alpha \qquad \text{for every fixed } h \text{ in (49)}. \tag{53}$$

Coverage and width are part of the same event; the intervals may be wider on a failed run.

Choose any integer $1 \le n \le m$. For the proof, let H be uniform on the unit sphere in $\text{span}\{e_1, \dots, e_n\}$, independently of Ξ. We may also reveal this support subspace to the algorithm. This prior is supported entirely on the fixed history (49); it is specified directly, rather than obtained by conditioning a different input distribution on that history. For $j > n$, the candidate quality is the known constant $c^2/T = 1/4$. Averaging (53) over this prior yields

$$\mathbb{P}_{H,\Xi}\{\mathcal{E}(H, \Xi)\} \ge 1 - \alpha. \tag{54}$$

Every subsequent query in the original space reduces to one linear measurement of H: subtract the known e_0 contribution and divide by v. Components outside its support contain no additional information. Condition on Ξ. Each query direction is then a measurable function of the previous responses. Orthogonalizing its component in the support against the previously measured directions either gives one new unit direction or shows that its response is already determined.

First suppose $Q < n$. We can simulate redundant responses from the known ones and, after the original algorithm stops, add orthogonal queries until exactly Q independent directions have been measured. Padding can use a fixed measurable rule for choosing an unmeasured direction. This produces a more informative experiment that reproduces the original intervals; the original algorithm simply ignores the added answers.

Let W_t be the span of the first t orthogonalized directions, and define

$$p_t = P_{W_t} H, \qquad R_t = 1 - \|p_t\|^2, \qquad N_t = W_t^\perp \cap \text{span}\{e_1, \dots, e_n\}. \tag{55}$$

Conditionally on the seed and the first t responses, the law of H is uniform on

$$p_t + \sqrt{R_t}\,\mathbb{S}(N_t), \qquad \dim N_t = n - t, \tag{56}$$

where $\mathbb{S}(N_t)$ denotes the unit sphere in N_t. This follows by induction using regular conditional distributions. Initially, the claim is the definition of the prior. Given the past, the next direction is fixed, and conditioning a uniform sphere on its coordinate in that direction leaves a uniform spherical slice. The choice of direction adds no information beyond the responses from which it was computed. The conditional statements hold almost surely; when $\dim N_t = 1$, the posterior consists of two equiprobable endpoints.

By conditional isotropy, the squared response along the next unit direction has conditional mean $R_t/(n - t)$. Thus

$$\mathbb{E}[R_{t+1} \mid \Xi, \text{ first } t \text{ responses}] = \frac{n - t - 1}{n - t} R_t. \tag{57}$$

Starting from $R_0 = 1$ and iterating gives

$$\mathbb{E}[R_Q \mid \Xi] = 1 - \frac{Q}{n}, \qquad \mathbb{E}\|p_Q\|^2 = \frac{Q}{n}. \tag{58}$$

These identities therefore also hold after averaging over the seed.

B.3 A narrow quality box cannot cover both antipodes

Condition on the seed and the full padded transcript. The reported intervals are now fixed. If any interval has width greater than η, the joint event (52) is impossible. Otherwise, for $1 \le j \le n$ let

$$B_j = \{t \in [-1, 1] : f_\varepsilon(t) \in I_j\}, \qquad B = B_1 \times \cdots \times B_n. \tag{59}$$

An empty B_j also makes success impossible. In the remaining case, (51) implies

$$\operatorname{diam}(B) \le \frac{\sqrt{n}\,\eta}{D_\varepsilon} = 2a, \qquad a^2 = \frac{n\eta^2}{4D_\varepsilon^2}. \tag{60}$$

If $R_Q > a^2$, each posterior antipodal pair $p_Q + z, p_Q - z$ has distance $2\sqrt{R_Q} > 2a$. At most one member of the pair can lie in B. Since the conditional law in (56) is invariant under this pairing, its probability of covering all first n candidate qualities is at most $1/2$. This includes the two-point posterior when $n - Q = 1$. It follows that

$$\mathbb{P}_{H,\Xi}\{\mathcal{E}(H,\Xi)\} \le \frac{1}{2} + \frac{1}{2}\mathbb{P}\{R_Q \le a^2\}. \tag{61}$$

Suppose $a^2 < 1$. Markov's inequality and (58) give

$$\mathbb{P}\{R_Q \le a^2\} = \mathbb{P}\{\|p_Q\|^2 \ge 1 - a^2\} \le \frac{Q}{n(1-a^2)}. \tag{62}$$

Combining (54), (61), and (62) proves

$$Q \ge (1-2\alpha)n\left(1 - \frac{n\eta^2}{4D_\varepsilon^2}\right). \tag{63}$$

If $Q \ge n$, this inequality holds automatically because its right side is at most n. Hence it holds for every query cap Q whenever $n\eta^2 < 4D_\varepsilon^2$.

For $n = 1$ and $Q < n$, we have $Q = 0$: the initial prior is the two-point sphere $\{-e_1, e_1\}$, $R_0 = 1$, and the same antipodal argument applies without any conditioning step. Thus the proof covers $n = 1$ as well. The prior was chosen independently of the algorithm's seed. Consequently, violating (63) implies failure of the target guarantee on at least one fixed input for the randomized procedure as a whole.

B.4 Constants and the precision scale

For $0 < \varepsilon \le 1/10$,

$$\delta v = \sqrt{\frac{13\varepsilon}{50}} \le \frac{1}{5}, \qquad c - \delta v \ge \frac{3}{5}. \tag{64}$$

Substitution into (50) gives

$$D_\varepsilon^2 = \frac{4\delta^2 v^2 (c - \delta v)^2}{T^2} \ge \frac{117}{2048}\varepsilon. \tag{65}$$

Choose

$$n = \min\left\{m, \left\lfloor \frac{117\varepsilon}{1024\eta^2} \right\rfloor\right\}. \tag{66}$$

When $n \ge 1$, this choice ensures $a^2 \le (512/117)n\eta^2/\varepsilon \le 1/2$. Equation (63) therefore yields

$$Q \ge \frac{1-2\alpha}{2}\, n. \tag{67}$$

Since Q is an integer, the right side may be rounded up. For fixed $\alpha < 1/2$ and widths sufficiently small relative to $\sqrt{\varepsilon}$, this is $Q = \Omega(\min\{m, \varepsilon/\eta^2\})$. If (66) gives $n = 0$, this specialization asserts no positive lower bound.

Uniform vector recovery from adaptive randomized linear information is studied by Kunsch & Wnuk (2026). Here the monotone map f_ε converts quality precision into coordinate precision, and the construction places that recovery obstruction inside a globally valid selection history. The bound concerns the additional worst-case query cap at that history, with all earlier information retained and exact norm and selected quality supplied after selection.

C GEOMETRIC INTERPRETATION AND A WORKED AUDIT

C.1 THE DIMENSION NEEDED FOR ROBUST SELECTION

Let P_W be a rank-q orthogonal projector and $M = \sum_j u_j u_j^\top$. In an orthonormal eigenbasis of M, write ω_ℓ for the diagonal entries of P_W. They lie in $[0, 1]$ and sum to q, so

$$\sum_j \|P_W u_j\|^2 = \operatorname{tr}(P_W M) = \sum_\ell \lambda_\ell \omega_\ell \le \sum_{\ell \le q} \lambda_\ell.$$

Subtracting from $\operatorname{tr} M = K$ gives the left inequality in (6). The right inequality follows from $d_k = 0$ for the anchored choice and $d_j \le \varepsilon$ for the other $K - 1$ candidates. The argument applies to each realized measured subspace, including one chosen adaptively. For orthogonal candidates, the nonzero eigenvalues are all one, giving $q \ge K - (K - 1)\varepsilon$. A missing anchor adds at most one dimension; redundant query rows still count toward the query total. The trace constraint is necessary only: it does not replace the individual residual constraints or the measured-score comparison in Proposition 2.

C.2 PROTOCOL, CALIBRATION, AND RETAINED OBSERVATIONS

We use the family in (20), fixed before the observations. The selector queries e_0, e_* and returns u_0 if $x > 0$ and $|bz| < x/10$, where $b = \sqrt{211/600}$. Otherwise it measures the remaining 64 coordinate axes and selects exactly. Appendix B proves that the open branch is robustly ε-optimal. At its displayed responses $(x, z) = (1, 0)$, $E_0 = A = 1$, $a_j = 4/5$ and $d_j = 1/120$ for $j > 0$. For these candidates,

$$d_j < \frac{1}{20}, \qquad (1 + \varepsilon)(\varepsilon - d_j) = \frac{7}{160} > \frac{4}{125} = \varepsilon a_j^2.$$

Thus the full stopping predicate holds strictly. The selected direction is already measured, so there is no anchor query.

Let $W_0 = \operatorname{span}(e_0, e_*)$ and let $r = g - P_{W_0} g$. After selection, draw $z_1, \dots, z_{16}$ independently from $N(0, P_{W_0^\perp})$. Each scalar $y_\ell = \langle g, z_\ell \rangle$ costs one query. For a fixed input with $R_0 = \|r\|^2 > 0$,

$$S = \sum_{\ell=1}^{16} y_\ell^2, \qquad S/R_0 \sim \chi^2_{16}.$$

Consequently $[S/q_{.975}, S/q_{.025}]$ covers R_0 with probability 0.95. Since the selected quality is $A/(E_0 + R_0)$, applying this decreasing map gives

$$\left[\frac{A}{E_0 + S/q_{.025}}, \frac{A}{E_0 + S/q_{.975}}\right].$$

For the displayed responses, $A = E_0 = 1$, yielding (21). If $R_0 = 0$, all responses are zero and the interval reduces to the exact selected quality. This is standard Gaussian calibration. Its probability guarantee refers to the random protocol for each fixed input, not to a newly conditioned guarantee after the realized probe rows are displayed.

For release, define the final measured space as

$$W = \operatorname{span}(e_0, e_*, z_1, \dots, z_{16}).$$

The full row matrix and scalar response vector determine $p = P_W g$. All candidate projections and their interval hulls are recomputed using this final space and the full response vector. The original interval I_0 remains unchanged. On its original coverage event, the true input belongs to $C(H, I_0)$, so all candidate hulls cover on that same event. Adding the calibration rows also preserves the already established robust selection condition. An incompatible interval is flagged using the rules of Theorem 5. The displayed branch uses two selection and 16 calibration queries, 18 in total, with no further query for release. The final rank is 18 almost surely and the remaining dimension is 48.

C.3 Reproducible illustration and equal-width comparison

The three fixed inputs are

$$g_\theta = e_0 + \sqrt{1/\theta - 1}\, e_1, \qquad \theta \in \{0.02, 0.50, 0.98\}.$$

They all produce the selection responses $(1, 0)$. Input values are used only to generate exact scalar responses and to check the released intervals afterwards. They are not arguments to the inference routine. Before any probe draw we fixed these inputs, the candidate family, $\varepsilon = 0.05$, $\alpha = 0.05$, 16 probes, and the displayed alternative u_1. We generated one 16×64 standard Gaussian matrix using NumPy's PCG64 generator with seed 20260922 and reused that matrix for the three illustrations, for a total of 54 scalar responses. All outcomes are retained. The archive includes the realized rows and responses, so the arithmetic can be replayed without drawing again.

The intervals in Figure 2(a) are individual protocol outputs. The three inputs share the same probes and are not independent replications for estimating coverage. Reported decimals are floating-point evaluations of the exact formulas.

Figure 2(b) answers a distinct controlled question. It evaluates the existing uniform bound $G_\varepsilon(L, L+w)$ with $\varepsilon = w = 0.05$ and $0 \le L \le 0.95$. These supplied intervals have the same width; only their position varies. The curve is not obtained by modifying the raw calibrated intervals in panel (a), and it is not a claim that every bound value is attained by that particular measured history.