PROTECTED UNBIASED SPARSIFICATION: A SHARP LOEWNER CURVE

Anonymous authors
Paper under double-blind review

ABSTRACT

Many learning systems replace a probability vector $p \in \Delta_k$—token attention, data weights, expert gates—by a random vector Q supported on at most $m \ll k$ coordinates. Such a replacement is trustworthy only if it is unbiased, preserves prescribed linear statistics $AQ = Ap$ on every realization (class proportions, group masses, control variates), and remains reliable for downstream responses that are not known when Q is drawn. We give a complete answer to this problem in Loewner order. Writing $d = \operatorname{rank}(AC_pA^\top) \le k-2$ for the effective number of protected directions, $C_p = \operatorname{diag}(p) - pp^\top$, and $C_{p|A}$ for the covariance of a categorical draw that remains after regressing on the protected statistics, we construct, for every budget $d+1 \le m \le k$, an unbiased exactly protected m-sparse Q with

$$\operatorname{Cov}(Q) \preceq \frac{k-m}{(m-d)(k-d-1)}\, C_{p|A},$$

and we prove that the coefficient cannot be improved uniformly over instances at any budget. The result lifts a concurrent unprotected coefficient to arbitrary pathwise protection through a new residual-covariance bridge: the protected problem on k coordinates with budget m behaves exactly like the unprotected problem on $k-d$ coordinates with budget $m-d$. For rational inputs, an exact-rational Las Vegas sampler draws Q in expected polynomial time and attains the curve within a factor $1+\varepsilon$. Finally, the universal guarantee is essentially the strongest constructible one: optimizing a law for a single revealed quadratic query W already encodes Minimum Uncut exactly, so no polynomial-time approximation scheme outputs an explicitly represented optimal law unless $\mathsf{P} = \mathsf{NP}$. A small numerical study confirms that the constructed law tracks the curve and that the minimax step is necessary.

1 INTRODUCTION

A recurring primitive in large-scale learning is to replace a dense probability vector by a sparse one. Attention over a long context is restricted to a few retained keys (Zhang et al., 2023), a softmax gate over experts is truncated to its top entries (Shazeer et al., 2017), and a training set is reweighted onto a small coreset (Mirzasoleiman et al., 2020; Killamsetty et al., 2021). Deterministic truncation is biased. The standard remedy is randomization: draw a random probability vector Q with $\mathbb{E}Q = p$ and $|\operatorname{supp}(Q)| \le m$, so that the error averages out. Two further requirements make the problem interesting. First, some linear statistics of p must be preserved on *every* realization, not merely in expectation—class proportions in a coreset, per-head or per-block attention mass, group totals in a survey, control variates in a Monte Carlo estimator. These constraints $AQ = Ap$ define an affine slice of the simplex. Second, the sparsification is chosen before the downstream use is known: the same sparse weights are consumed by a gradient, a loss, a logit, or an entire optimization run that is revealed only later. A guarantee that depends on the response is therefore of limited value; what is needed is a single *response-oblivious* certificate.

The right object for such a certificate is a matrix, not a scalar. If the future response is linear with coefficient matrix G, its error covariance is $G\operatorname{Cov}(Q)G^\top$, and a bound $\operatorname{Cov}(Q) \preceq cB$ in Loewner order controls *every* G simultaneously. Exact protection also fixes the natural comparator. A single categorical draw e_I, $I \sim p$, has covariance $C_p = \operatorname{diag}(p) - pp^\top$; once the protected statistics Ae_I

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.

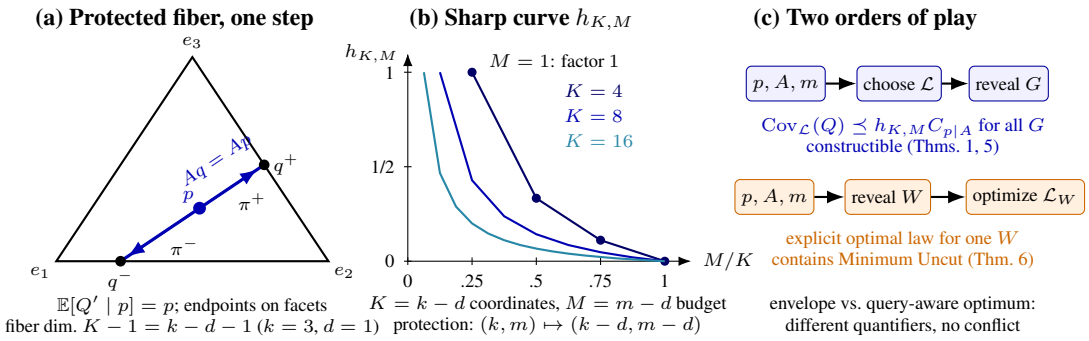


Figure 1: **Residual geometry, sharp curve, and order of play.** (a) Exact pathwise protection $Aq = Ap$ cuts the simplex to a fiber of dimension $K - 1 = k - d - 1$; the law of Theorem 1 is a martingale that walks from p to lower-dimensional faces of this fiber, dropping one coordinate per step. (b) The worst-instance response-oblivious factor $h_{K,M} = (K - M)/(M(K - 1))$ as a function of the free budget $M = m - d$, for three residual cardinalities K. (c) The positive result fixes a law before any response G is revealed and bounds all responses at once; the hardness result reveals a quadratic query W first and asks for the best explicit law, which already contains Minimum Uncut.

are known, the only part that can still fluctuate is the residual after linear regression,

$$C_{p|A} = C_p - C_p A^\top (A C_p A^\top)^\dagger A C_p.$$

The question we answer is: *what fraction of $C_{p|A}$ must an unbiased, exactly protected, m-sparse law retain, uniformly over all responses?*

Answer. Let $d = \mathrm{rank}(AC_pA^\top)$ count the protected directions that actually bind on the simplex (redundant rows and the implicit constraint $\mathbf{1}^\top Q = 1$ do not count), and assume $d \le k - 2$. We show that the answer depends on (k, m, d) only through the residual cardinality $K = k - d$ and the free budget $M = m - d$, and equals

$$h_{K,M} = \frac{K - M}{M(K - 1)}.$$

Every protected direction costs one support position and one dimension: protected sparsification on k coordinates with budget m behaves exactly like unprotected sparsification on $k - d$ coordinates with budget $m - d$, measured against $C_{p|A}$. The curve equals one at the minimum feasible budget $m = d + 1$, decays like $1/M$, and vanishes at $m = k$.

Example (class-balanced reweighting). Take k training points, c classes, and let the rows of A be the class indicators. Then $d = c - 1$ (the last class is implied by $\mathbf{1}^\top Q = 1$), a feasible coreset needs at least c points, and with m points the covariance of any unbiased class-balanced weighting can be pushed down to $h_{k-c+1,m-c+1} C_{p|A}$ but not lower. With $k = 10^4$, $c = 10$, the factor at $m = 100$ is 0.0109 against 0.0099 unprotected, but at $m = 20$ it is 0.091 versus 0.050: protection doubles the residual covariance. Section 7 turns such factors into excess-risk bounds for strongly convex objectives.

Contributions.

1. **Sharp protected Loewner curve (Theorem 1).** For every $d \le k - 2$ and $d + 1 \le m \le k$ there is a finitely supported, unbiased, pathwise-protected, m-sparse law with $\mathrm{Cov}(Q) \preceq h_{k-d,m-d} C_{p|A}$, and the coefficient is the smallest that holds uniformly over instances of effective rank d. The unprotected coefficient $h_{k,m}$ appears in concurrent work (Zhang, 2026); the protected lift is not a change of variables, because the comparator $C_{q|A}$ is a nonlinear function of the current state q. The new ingredient is a *residual covariance bridge* (Lemma 3), $\mathrm{Cov}(P) + \mathbb{E}\, C_{P|A} \preceq C_{p|A}$ for any unbiased protected P, which linearizes the recurrence exactly.
2. **Exact-rational implicit sampler (Theorem 5).** For rational (p, A) and rational $0 < \varepsilon \le 1$, a Las Vegas algorithm outputs an exactly protected, exactly unbiased rational draw with covariance

factor at most $(1+\varepsilon)h_{k-d,m-d}$ in expected time $\mathrm{poly}(k, a, L + \mathrm{enc}(\varepsilon))$, where L is the input encoding length. The algorithm samples a single root-to-leaf path of the martingale; it never lists the terminal law, which may have exponentially many atoms.

3. **Query-aware boundary (Theorem 6).** One universal law attains the sharp response-oblivious envelope, yet optimizing a law for one revealed quadratic form W encodes Cubic Minimum Uncut exactly, so no polynomial-time approximation scheme outputs an explicitly represented optimal law unless $\mathsf{P} = \mathsf{NP}$.

A numerical study (Section 8) confirms that the constructed law tracks the curve on a nontrivial protected instance and that the minimax step inside the construction is necessary: a naive radial step is unbiased and protected but violates the one-step bound by up to a factor 4.5. The two orders of play in Figure 1(c) do not conflict: a sharp robust envelope is constructible before the response is known, while the instance-optimal law for a revealed query is intractable to represent. Complete proofs are in the appendices.

2 Related work

Table 1 (Appendix A) records theorem by theorem what the statements closest to Theorem 1 prove; the paragraphs below place them in context.

Balanced sampling and dependent rounding. Unequal-probability sampling without replacement goes back to Horvitz & Thompson (1952). The cube method constructs samples with prescribed inclusion probabilities while maintaining exact or near-exact balance on auxiliary totals (Deville & Tillé, 1998; 2004; 2005; Tillé & Rivest, 2025). In theoretical computer science the same mechanism appears as dependent or pipage rounding (Srinivasan, 2001; Gandhi et al., 2006) and its sub-isotropic generalizations (Bansal, 2019): a fractional point is walked to a vertex by a martingale that preserves selected linear constraints exactly and controls correlations. These works established pathwise balance, unbiased inclusion, and flight to a low-dimensional face; they do not ask for the exact matrix-valued profile, uniformly over all support budgets, of the covariance that must remain relative to $C_{p|A}$.

Sparse representations of measures. Tchakaloff-type arguments give finite positive representations of prescribed moments (Bayer & Teichmann, 2006), and approximate Carathéodory theorems give deterministic sparse convex combinations with sharp ℓ_2 radius (Mirrokni et al., 2017). Tinarrage (2026) proves a sharp equal-weight version (barycenters of t-tuples of simplex vertices at the sharp radius $R_{d,t}$) and, under a coordinate cap $p_i \le 1/r$, an unbiased equal-weight r-subset with a scalar second-moment bound. Our atoms carry arbitrary nonnegative weights, satisfy pathwise affine constraints, and are controlled in Loewner order rather than by a scalar. Efficient unbiased sparsification (Barnes et al., 2024) characterizes optimal laws for several separable objectives; those objective-specific results are complementary to a response-oblivious matrix bound.

Protected sparse measures. Haverbeck et al. (2026) is the closest protected precursor. Their Proposition A.3 constructs an unbiased random reweighting whose barycenter error lies pathwise in $U^{\perp}$ for a chosen r-dimensional protected subspace and obeys $\mathbb{E}[\eta \otimes \eta] \preceq \frac{20}{9b}(I - \Pi_U)\Sigma(I - \Pi_U)$ with support at most $b + 2r + 2$; the construction then drives risk guarantees and practical KV-cache rounding algorithms. We do not claim to introduce protected reweighting or operator-valued covariance control. The differences are the objects and the sharpness: arbitrary $AQ = Ap$ with effective rank $d = \mathrm{rank}(AC_pA^{\top})$, the exact minimum support $d + 1$, the Schur-complement comparator $C_{p|A}$, and the exact saturating coefficient $h_{k-d,m-d}$ at every budget. Their projected covariance and $C_{p|A}$ should not be identified without an additional argument.

Concurrent unprotected sparsification. Zhang (2026) proves, for arbitrary positive weights on s points and every $1 \le t \le s$, a t-sparse convex combination with squared mean error at most $\frac{s-t}{t(s-1)}V_q$ (their Theorem 6.1 and Lemma 6.2), sharp for the uniformly weighted regular simplex, and notes that one boundary law serves all positive-semidefinite quadratic tests simultaneously. Their one-step bound rests on a simplex–ball interpolation inequality (the full-dimensional Grace–Danielsson inequality); Lemma 2 obtains the same one-step coefficient from an AM–HM inequality over radial

responses. We treat the unprotected coefficient and the boundary-law mechanism as concurrent prior. Our theorem-level delta is the lift to arbitrary pathwise protection via Lemma 3, the shift $(k, m) \mapsto (k-d, m-d)$ against the comparator $C_{p|A}$, the sharpness analysis in the protected regime, and the rational bit-model realization.

Query-aware design and learning subsets. Sparse standard quadratic optimization exposes the difficulty of optimizing one quadratic objective over a sparse simplex (Bomze et al., 2025); our fixed-barycenter law problem differs because every atom must be sparse and feasible while the mixture must reproduce p. In learning, CRAIG and Grad-Match relate weighted gradient subsets to optimization stability (Mirzasoleiman et al., 2020; Killamsetty et al., 2021); we use strong convexity only to interpret the protected covariance coefficient and do not propose a new subset-selection method.

3 Protected affine-slice geometry

Let $\Delta_k = \{q \in \mathbb{R}^k_+ : \mathbf{1}^\top q = 1\}$, fix $p \in \Delta_k^\circ$ (zero coordinates of p can be deleted first), and let $A \in \mathbb{R}^{a\times k}$. A random probability vector Q is *admissible at budget* m if

$$AQ = Ap \text{ a.s.}, \qquad \mathbb{E}Q = p, \qquad |\operatorname{supp}(Q)| \le m \text{ a.s.} \tag{1}$$

Rows of A that are redundant given $\mathbf{1}^\top q = 1$ do not restrict the fiber and must not be charged to the support budget. The effective number of protected directions is

$$d = \operatorname{rank}(AC_pA^\top), \qquad C_p = \operatorname{diag}(p) - pp^\top, \tag{2}$$

which equals $\operatorname{rank}[\mathbf{1}^\top; A] - 1$ (Appendix B). For example, protecting all c class totals gives $d = c-1$.

Residual comparator. The covariance of a categorical draw e_I, $I \sim p$, after linear regression on the protected statistics Ae_I, is

$$C_{p|A} = C_p - C_pA^\top(AC_pA^\top)^\dagger AC_p. \tag{3}$$

It is positive semidefinite of rank $k-1-d$, and its range is the tangent space $\ker[\mathbf{1}^\top; A]$ of the protected fiber $\{q \in \Delta_k : Aq = Ap\}$. Every admissible law therefore satisfies $\operatorname{range}(\operatorname{Cov}(Q)) \subseteq \operatorname{range}(C_{p|A})$, and $C_{p|A}$ is the largest covariance compatible with the constraints in the sense of Lemma 3 below.

The curve. For integers $K \ge 2$ and $1 \le M \le K$, define

$$h_{K,M} = \frac{K-M}{M(K-1)}. \tag{4}$$

Throughout, $K = k-d$ and $M = m-d$. The endpoint $M = 1$ is the minimum generally feasible support $m = d+1$ (a vertex of the fiber has at most $d+1$ nonzero coordinates), and $M = K$ is $m = k$, where $Q = p$ has zero covariance. If $d = k-1$ the fiber is the singleton $\{p\}$, feasibility forces $m = k$, and there is nothing to prove; we therefore assume $d \le k-2$.

Order of play. The law of Q is chosen after (p, A, m) are known but before any downstream response is revealed. A Loewner bound $\operatorname{Cov}(Q) \preceq c\,C_{p|A}$ is equivalent to $\mathbb{E}\|G(Q-p)\|_2^2 \le c\,\operatorname{tr}(GC_{p|A}G^\top)$ for every matrix G, so one matrix inequality certifies all linear responses at once; this quantifier order separates the universal guarantee of Section 4 from the fixed-query problem of Section 6.

4 The sharp protected Loewner curve

Theorem 1 (Sharp protected sparsification curve)**.** *Assume $p \in \Delta_k^\circ$, let $d = \operatorname{rank}(AC_pA^\top) \le k-2$, and take $d+1 \le m \le k$. There is a finitely supported admissible law satisfying*

$$\operatorname{Cov}(Q) \preceq h_{k-d,m-d}\,C_{p|A} = \frac{k-m}{(m-d)(k-d-1)}\,C_{p|A}. \tag{5}$$

For every feasible triple (k, d, m), the coefficient in Eq. (5) is the supremum, over all instances (p, A) of effective protected rank d, of the smallest constant c for which some admissible law satisfies $\operatorname{Cov}(Q) \preceq c\,C_{p|A}$.

4.1 ONE STEP: ISOTROPIC BOUNDARY CONTRACTION

Lemma 2 (Isotropic boundary contraction). *Let $p_i > 0$, $\sum_i p_i a_i = 0$, and $\sum_i p_i a_i a_i^\top = I_r$. For the polytope $\mathcal{K} = \{x \in \mathbb{R}^r : 1 + a_i^\top x \geq 0 \; \forall i\}$, there is a finitely supported random $X \in \partial\mathcal{K}$ such that*

$$\mathbb{E}X = 0, \qquad \mathrm{Cov}(X) \preceq \frac{1}{r^2} I_r. \tag{6}$$

Proof sketch. Fix a positive definite test matrix H. Shoot a ray from the origin in each direction $-H^{-1}a_i$ until it first meets $\partial\mathcal{K}$ at $x_i = -\tau_i H^{-1} a_i$, and give x_i weight $\pi_i \propto p_i/\tau_i$ (Figure 3 in Appendix B). Isotropy makes this law centered, and the facet inequality at x_i bounds its H-cost by

$$\mathrm{tr}(H \,\mathrm{Cov}\, X) \leq \frac{1}{\mathrm{tr}(H^{-1})} \leq \frac{\mathrm{tr}(H)}{r^2}, \tag{7}$$

the last step by the arithmetic–harmonic mean inequality. The set of centered laws on $\partial\mathcal{K}$ is convex and compact, the payoff $\mathrm{tr}(H \,\mathrm{Cov}\, X)$ is bilinear, and Sion's minimax theorem (Sion, 1958) over PSD H of unit trace produces one law satisfying Eq. (6) in every direction. Carathéodory compression of the moment vector makes its support finite. Appendix B gives the full computation. □

To apply the lemma at a positive state q with support size s, let $d_q = \mathrm{rank}(AC_qA^\top) \leq d$ and $r = s - 1 - d_q$. Choose an orthonormal basis $U \in \mathbb{R}^{s\times r}$ of $\ker(\sqrt{q}^\top) \cap \ker(A\,\mathrm{diag}(\sqrt{q}))$, write $u_i^\top$ for row i of U, and set $a_i = u_i/\sqrt{q_i}$. Then $\sum_i q_i a_i = 0$, $\sum_i q_i a_i a_i^\top = I_r$, every point of the protected fiber through q has the form

$$q(x)_i = q_i(1 + a_i^\top x), \qquad x \in \mathcal{K}, \tag{8}$$

and

$$C_{q|A} = \mathrm{diag}(\sqrt{q})\, UU^\top \,\mathrm{diag}(\sqrt{q}). \tag{9}$$

Hence Lemma 2 yields a mean-preserving protected move $q \mapsto Q'$ to the boundary of the fiber that drops at least one support coordinate and obeys

$$\mathrm{Cov}(Q' \mid q) \preceq \frac{1}{r^2}\, C_{q|A}. \tag{10}$$

4.2 THE BRIDGE: MAKING THE COMPARATOR TELESCOPE

The obstacle to iterating Eq. (10) is that the comparator $C_{q|A}$ is a nonlinear function of the state q; one cannot simply take expectations. The following inequality resolves this exactly and is the main new ingredient relative to the unprotected setting.

Lemma 3 (Residual covariance bridge). *If a random $P \in \Delta_k$ satisfies $\mathbb{E}P = p$ and $AP = Ap$ almost surely, then*

$$\mathrm{Cov}(P) + \mathbb{E}\, C_{P|A} \preceq C_{p|A}. \tag{11}$$

Proof sketch. Conditionally on P, draw a categorical index $I \sim P$; marginally $I \sim p$. Regress $e_I - p$ on $A(e_I - p)$ with the population coefficient $L_p = C_pA^\top(AC_pA^\top)^\dagger$; the residual has covariance exactly $C_{p|A}$. Pathwise protection makes the conditional mean of the residual given P equal to $P - p$, so the law of total covariance splits off $\mathrm{Cov}(P)$. Conditionally on $P = q$, the population coefficient L_p is a suboptimal regressor for the q-specific problem, and least-squares optimality lower-bounds the remaining conditional covariance by $C_{q|A}$. Details are in Appendix B. □

4.3 ASSEMBLING THE RECURRENCE

Proof of Theorem 1. Start with $P_k = p$. For $t = k, k-1, \ldots, m+1$, leave every state of support at most $t-1$ unchanged and apply Eq. (10) to every state of support exactly t. Since $d_q \leq d$, the tangent dimension at such a state is at least $t - d - 1$, so with $\alpha_t = (t - d - 1)^{-2}$ every branch satisfies $\mathbb{E}[P_{t-1} \mid P_t] = P_t$ and $\mathrm{Cov}(P_{t-1} \mid P_t) \preceq \alpha_t C_{P_t|A}$. Let $B_t = \mathrm{Cov}(P_t)$. The law of total covariance and Lemma 3 give

$$B_{t-1} \preceq \alpha_t \mathbb{E} C_{P_t|A} + B_t \preceq \alpha_t C_{p|A} + (1 - \alpha_t) B_t. \tag{12}$$

Starting from $B_k = 0$, the scalar recurrence $\eta_{t-1} = \alpha_t + (1 - \alpha_t)\eta_t$ has the closed form

$$\eta_t = \frac{k - t}{(t - d)(k - d - 1)} \tag{13}$$

(Appendix B), so $B_m \preceq h_{k-d,m-d} C_{p|A}$. The state P_m has support at most m, exact mean p, and exact protection. Every one-step law is finite, hence so is the terminal law.

Sharpness. For $d = 0$ and uniform p, any unbiased m-sparse Q has $\mathbb{E}\|Q - p\|_2^2 \geq 1/m - 1/k$ while $\operatorname{tr} C_p = 1 - 1/k$, so $\operatorname{Cov}(Q) \preceq c\, C_p$ forces $c \geq h_{k,m}$; the uniform m-subset law attains it. For $d > 0$, place total mass δ on d protected anchor coordinates ($A = [I_d\ 0]$) and spread $1 - \delta$ uniformly on the remaining $K = k - d$ coordinates. Every realization must spend d support positions on the anchors, and on the free block the problem is the unprotected uniform problem with budget M; a short computation (Appendix B) gives $c \geq (1 - \delta)h_{K,M}$, and $\delta \downarrow 0$ matches the upper bound. □

Remark 4 (Worst-instance, not pointwise). The supremum in Theorem 1 is over instances; for a fixed (p, A) the optimal constant can be strictly smaller. With $k = 3$, uniform p, $A = (0, 1, 3)$, and $m = 2$, the fiber is a segment with endpoints $(5/9, 0, 4/9)$ and $(0, 5/6, 1/6)$; the unique unbiased endpoint law has $\operatorname{Cov}(Q) = \frac{7}{9} C_{p|A}$, whereas $h_{2,1} = 1$ (Appendix B). Characterizing the instance-optimal constant is open; Section 6 shows that the analogous query-aware optimization is already hard.

5 A RATIONAL IMPLICIT SAMPLER

Theorem 1 is an existence theorem over the reals; its one-step law is a minimax point. For rational input, the same curve has a Turing-model realization up to any prescribed multiplicative accuracy, with exact mean and exact protection.

Theorem 5 (Exact-rational random-path sampler)**.** *Let $p \in \Delta_k^\circ \cap \mathbb{Q}^k$, $A \in \mathbb{Q}^{a \times k}$, and let L be the binary encoding length of (p, A, m, k). Write $\operatorname{enc}(\varepsilon)$ for the binary length of the reduced rational tolerance $0 < \varepsilon \leq 1$, and set $d = \operatorname{rank}(AC_pA^\top)$ and $d + 1 \leq m \leq k$. If $d = k - 1$ then $m = k$ and Algorithm 1 returns p. Otherwise it outputs a rational Q satisfying Eq.* (1) *exactly and*

$$\operatorname{Cov}(Q) \preceq (1 + \varepsilon)\, h_{k-d,m-d}\, C_{p|A}. \tag{14}$$

One draw takes expected bit time $\operatorname{poly}(k, a, L + \operatorname{enc}(\varepsilon))$, and every realized output has polynomial encoding length. In particular, for $\varepsilon = 2^{-b}$ the running time is $\operatorname{poly}(k, a, L + b)$.

Radial oracle. At a rational state q with support S, choose a rational basis T of $\ker[\mathbf{1}^\top; A]$ restricted to S, so that the fiber is $\{q + Tz\}$, and set $G = T^\top \operatorname{diag}(q)^{-1} T$, for which $C_{q|A} = TG^{-1}T^\top$. A rational test matrix $H \succ 0$ has an exact radial best response: for each nonzero row $b_i^\top$ of T,

$$u_i = H^{-1}b_i, \quad \lambda_i = \max_{j \in S} \frac{b_j^\top u_i}{q_j}, \quad z_i = -\frac{u_i}{\lambda_i}, \quad \pi_i = \frac{\lambda_i}{\sum_{\ell : b_\ell \neq 0} \lambda_\ell}. \tag{15}$$

Then $q + Tz_i$ is a rational boundary point of the fiber, $\sum_i \pi_i z_i = 0$, and the response covariance $V_H = \sum_i \pi_i z_i z_i^\top$ obeys $\langle H, V_H \rangle \leq 1/\operatorname{tr}(H^{-1}G) \leq \operatorname{tr}(HG^{-1})/r^2$ (Lemma 8); this is the rational form of the ray construction behind Lemma 2.

Two ledgers. Two devices prevent hidden exponential cost. (i) *Height of the minimax.* The one-step minimax is solved by a rational central-cut ellipsoid method with column generation (Grötschel et al., 1988) (Lemma 9); fraction-free PSD separation, explicit inner and outer radii, and a fixed-gap cut rule keep every oracle query at polynomial height, and a finite rational SDP extracts weights with factor $1 + \theta$. (ii) *Height of the state.* Denominators of q could otherwise grow geometrically along the path. Each reset re-expresses the state as a convex combination of basic feasible vertices of its face and rounds the coefficients once on a mesh $1/N$ by systematic sampling (Lemma 10); the realized height is polynomial in the face description and $\log N$, independent of the incoming denominator, and the reset is conditionally unbiased with added covariance at most $\beta\, C_{p|A}$ (Lemma 11). With $\theta = \varepsilon/8$ and total reset error $\varepsilon h_{k-d,m-d}/4$ the recurrence behind Eq. (14) closes (Appendix C).

Algorithm 1 Implicit protected random path (one draw)

Require: Rational (p, A), budget m, accuracy $\varepsilon \in (0, 1]$

Compute $d = \operatorname{rank}(AC_pA^\top)$, $K = k - d$, $M = m - d$.
if $m = k$ **then**
  return p
end if
if $M = 1$ **then**
  Decompose p into basic feasible vertices of the fiber (Lemma 10); sample one vertex exactly with fair bits (Lemma 7);
  return it.
end if
$R \leftarrow k - m$, $\theta \leftarrow \varepsilon/8$, $\beta \leftarrow \varepsilon h_{K,M}/(4R)$.
Compute a rational basis T_0 of $\ker[\mathbf{1}^\top; A]$ and a rational $\underline{\lambda} > 0$ with $C_{p|A} \succeq \underline{\lambda} P_{\mathrm{range}(T_0)}$ (Lemma 11).
$N \leftarrow \lceil (k+1)/(\beta\underline{\lambda}) \rceil$, $q \leftarrow p$.
while $|\operatorname{supp}(q)| > m$ **do**
  Minimax step: run the rational ellipsoid/column-generation routine of Lemma 9 with the radial oracle Eq. (15) to obtain a $(1+\theta)$-valid one-step mixture of boundary laws.
  Sample the mixture index and then the radial endpoint exactly with fair bits; set q to the endpoint (one coordinate becomes zero).
  Reset: decompose q into at most k basic feasible vertices of its face and replace it by the conditionally unbiased mesh-$1/N$ systematic reset of Lemma 10.
end while
return q

Scope. Mean and protection are exact and only the covariance factor carries the slack $1 + \varepsilon$. The theorem does not cover $\varepsilon = 0$, real inputs, or worst-case random-bit counts; it does not list the terminal law, which may have exponentially many leaves; and it makes no claim about practical speed.

6 The query-aware computational boundary

The universal law of Theorem 1 is chosen before the response is known. One may ask whether a better law is available once a specific quadratic response $W \succeq 0$ is revealed. Define the query-aware value

$$\mathrm{OPT}(p, A, m; W) = \inf_{\mathcal{L}} \mathbb{E}_{\mathcal{L}}(Q - p)^\top W (Q - p), \tag{16}$$

the infimum over admissible finite laws. The universal result has the quantifier order

$$\exists \mathcal{L}^{\mathrm{obl}}\ \forall W \succeq 0: \quad \operatorname{tr}(W \operatorname{Cov}_{\mathcal{L}^{\mathrm{obl}}}(Q)) \le h_{k-d,m-d} \operatorname{tr}(W C_{p|A}), \tag{17}$$

whereas Eq. (16) reveals W first and permits a different law for each W. Eq. (17) is a sharp W-uniform absolute bound; it does not assert that $\mathcal{L}^{\mathrm{obl}}$ approximates $\mathrm{OPT}(p, A, m; W)$ within any finite ratio, and indeed a low-rank W can have $\mathrm{OPT} = 0$ while every fixed response-oblivious law has positive cost. The following theorem shows that the query-aware problem is genuinely harder.

Theorem 6 (Exact Minimum-Uncut embedding)**.** *Given a graph $G = (V, E)$ on n vertices, one can construct in polynomial time an instance with $k = 2n$, uniform p, pair-mass protection A of effective rank $d = n - 1$, budget $m = n = d + 1$, and an integer matrix $W_G \succeq 0$ such that*

$$\mathrm{OPT}(p, A, n; W_G) = \frac{4 \operatorname{MinUncut}(G)}{n^2}. \tag{18}$$

For cubic G, W_G has $O(n)$ nonzeros and $\|W_G\|_{\mathrm{op}} \le 12$. Consequently, unless $\mathsf{P} = \mathsf{NP}$, there is no polynomial-time approximation scheme for the fixed-W problem whose output is an explicitly represented admissible law.

Proof. Index the $2n$ coordinates by $(i, +), (i, -)$, set $p = \mathbf{1}/(2n)$, and let the rows of A be the first $n-1$ pair sums $q_{(i,+)} + q_{(i,-)}$; the simplex equation supplies the last, so $d = n - 1$ and every feasible

atom gives each pair mass $1/n$. Because a feasible atom has support at most n, it selects exactly one coordinate per pair, with mass $1/n$. The protected sparse fiber is therefore $\{q_z : z \in \{-1,+1\}^n\}$ with $Dq_z = z/n$, where D is the pair-difference matrix, and

$$\tfrac{1}{2}q_z + \tfrac{1}{2}q_{-z} = p. \tag{19}$$

Let $\mathcal{S}_G = \mathrm{Deg}(G) + \mathrm{Adj}(G)$ be the signless Laplacian and set $W_G = D^\top \mathcal{S}_G D$. Since $Dp = 0$,

$$(q_z - p)^\top W_G (q_z - p) = \frac{1}{n^2} z^\top \mathcal{S}_G z = \frac{1}{n^2} \sum_{\{i,j\}\in E} (z_i + z_j)^2 = \frac{4\,\mathrm{uncut}_G(z)}{n^2}, \tag{20}$$

where $\mathrm{uncut}_G(z)$ counts edges with $z_i = z_j$. Every atom costs at least the right side of Eq. (18), and mixing a minimum-uncut z with $-z$ as in Eq. (19) is a feasible law attaining it. For cubic G, $\|D\|_{\mathrm{op}}^2 = 2$ and $\|\mathcal{S}_G\|_{\mathrm{op}} \le 6$.

Minimum Uncut is APX-hard on cubic graphs, by an L-reduction from cubic Max-Cut (Cilibrasi et al., 2007, Obs. 1), itself APX-hard (Alimonti & Kann, 2000). An explicit-law algorithm outputs a rational $\sum_t \lambda_t \delta_{q_{z_t}}$ of polynomial total encoding length with exact pathwise feasibility and exact barycenter. If its cost is within ratio ρ of OPT, scanning the polynomial-size support and returning the cheapest atom gives

$$\min_t \mathrm{uncut}_G(z_t) \le \sum_t \lambda_t\, \mathrm{uncut}_G(z_t) \le \rho\ \mathrm{MinUncut}(G).$$

Conversely every cut yields the two-atom law Eq. (19). Approximation ratios therefore transfer exactly in both directions; bipartite inputs (zero optimum) are recognized separately. □

Output interfaces. The hardness statement concerns an explicit feasible-law output, natural whenever every deployable sparse buffer must be listed for audit; it is strictly stronger than sampling access, and excluding an implicit-sampler PTAS would need a randomized-complexity assumption. Table 2 in Appendix D records what is and is not established, including a positive-definite pair-protected variant with polynomial condition number and an unprotected ($A = 0$) family with near-identity W for which exact threshold comparison and $1/k^2$-additive evaluation are NP-hard via Minimum Bisection (Garey et al., 1976). Constant-factor hardness at $A = 0$, or under a uniformly bounded condition number, remains open.

7 A LEARNING INTERPRETATION AND SCOPE LIMITS

Exact meaning for quadratic tasks. After an admissible law $\mathcal{L}$ is fixed, let a future task choose columns $V = [v_1, \dots, v_k]$ and losses $f_i^V(\theta) = \frac{\mu}{2}\|\theta - v_i\|_2^2$, so that the minimizer of $F_w^V = \sum_i w_i f_i^V$ is $\theta_w = Vw$. Writing $B_{\mathcal{L}} = \mathrm{Cov}_{\mathcal{L}}(Q)$,

$$\mathbb{E}_{\mathcal{L}}\big[F_p^V(\theta_Q) - F_p^V(\theta_p)\big] = \frac{\mu}{2}\,\mathrm{tr}(VB_{\mathcal{L}}V^\top). \tag{21}$$

Normalizing by the residual task dispersion $\frac{\mu}{2}\,\mathrm{tr}(VC_{p|A}V^\top)$, PSD duality and Theorem 1 give

$$\sup_{p,A:\,\mathrm{rank}(AC_pA^\top)=d}\ \inf_{\mathcal{L}}\ \sup_{V:\,\mathrm{tr}(VC_{p|A}V^\top)>0} \frac{\mathrm{tr}(VB_{\mathcal{L}}V^\top)}{\mathrm{tr}(VC_{p|A}V^\top)} = h_{k-d,m-d}. \tag{22}$$

The law sees (p, A) but not the task; the task sees the law but not the realized draw. Eq. (22) is an operational restatement of the matrix theorem for the worst future quadratic task.

Strongly convex ERM on the sparse weights. Let $F_w(\theta) = \sum_i w_i f_i(\theta)$, suppose every realized weighted objective F_q is μ-strongly convex, and let F_p be β-smooth. With $G_* = [\nabla f_1(\theta_p), \dots, \nabla f_k(\theta_p)]$,

$$\mathbb{E}\big[F_p(\theta_Q) - F_p(\theta_p)\big] \le \frac{\beta\, h_{k-d,m-d}}{2\mu^2}\ \mathrm{tr}(G_* C_{p|A} G_*^\top) \tag{23}$$

(Appendix E, which also covers a fixed draw reused along full-batch gradient descent). In the class-balanced example of Section 1, $C_{p|A}$ is the within-class gradient covariance, so exact class balance removes the between-class gradient dispersion from the excess risk entirely while the within-class part is reduced by exactly $h_{k-c+1,m-c+1}$.

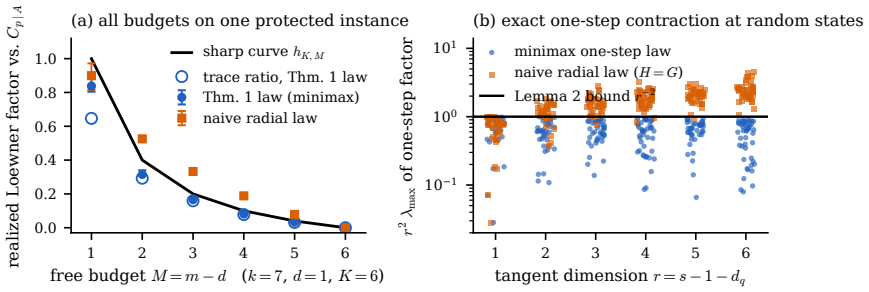


Figure 2: **Realized covariance against the sharp curve.** (a) $k = 7$, non-uniform p, one binding constraint ($d = 1$, $K = 6$); 6000 paths per budget; largest generalized eigenvalue of the estimated $\mathrm{Cov}(Q)$ relative to $C_{p|A}$ (filled, 90% bootstrap intervals) and trace ratio (hollow). (b) Exact one-step factors $r^2 \lambda_{\max}\big(C_{q|A}^{+1/2}\,\mathrm{Cov}(Q' \mid q)\,C_{q|A}^{+1/2}\big)$ at 240 random states with support 3–8 and one constraint.

Scope limits. Eq. (23) needs the displayed global curvature assumptions and does not extend automatically to nonconvex or PL-only losses, adaptive optimizers, or autoregressive deployment; we make no empirical accuracy, speed, or memory claim.

8 Numerical Illustration

This section checks that the construction of Theorem 1 runs on a nontrivial protected instance, that its realized covariance tracks the curve, and that the minimax step is what makes it do so. The one-step law is computed in floating point by matrix multiplicative weights (Arora & Kale, 2016) over the radial best responses Eq. (15) (200 rounds, uniform average of the queried responses). We compare with a *naive* step using the single radial response at $H = G$; both steps are unbiased and exactly protected and differ only in the Loewner constant. The script is in the supplementary material.

Figure 2(a) shows the realized factor of the constructed law at $0.84, 0.32, 0.17, 0.081, 0.033$ for $m = 2, \ldots, 6$ against $h_{6,M} = 1, 0.4, 0.2, 0.1, 0.04$, and zero at $m = k$; the naive law gives $0.90, 0.53, 0.33, 0.19, 0.079$, above the curve at every intermediate budget. That the constructed law sits below the curve is consistent with Remark 4. Figure 2(b) isolates the one-step lemma: the minimax law never exceeds the bound of Lemma 2 (maximum 0.999), while at anisotropic states the single radial response exceeds it by up to a factor 4.5; this is why the existence proof needs Sion's theorem and the sampler needs Lemma 9 rather than a single oracle call. In every draw, $AQ = Ap$ and the support bound held to floating-point precision. Appendix G repeats the experiment at $k = 32$, $d = 3$.

9 Conclusion

Exact linear protection of a sparsified probability vector has a precise price: in the residual geometry of $K = k - d$ coordinates and free budget $M = m - d$, the optimal worst-instance response-oblivious covariance, measured against the regression residual $C_{p|A}$, contracts by exactly $(K - M)/(M(K - 1))$, one at the minimum feasible support and zero at full support. The residual covariance bridge behind this curve also yields an exact-rational Las Vegas sampler attaining it within $1 + \varepsilon$, whereas optimizing an explicit law for a single revealed quadratic query already contains Minimum Uncut. Open directions: the instance-optimal constant for a fixed (p, A), which lies strictly below the curve on some instances, and approximate implementations of the minimax step at the scale of attention or gating vectors.

ETHICS STATEMENT

This is a theoretical paper on randomized sparsification of probability vectors. It uses no human-subject data, no personal data, and no pretrained model, and we foresee no direct negative societal impact. The exact-protection guarantee could be used to enforce group-mass or class-proportion constraints in data reweighting; whether a particular set of protected statistics is appropriate is an application-level decision that the mathematics does not settle.

AI USE STATEMENT

In this work we used large language models (LLMs) for the following tasks. Writing: drafting and editing the prose of the main text and appendices, suggesting the structure of sections, and formatting references. Literature: searching for and summarizing related work, and organizing the theorem-level comparison in Appendix A; every cited statement was then checked against the cited source by the authors. Mathematics: checking algebraic steps and numerical constants in the appendix proofs, drafting the written form of proofs whose arguments were fixed by the authors, and producing an adversarial internal review of claim boundaries. Experiments and figures: writing the numerical script of Section 8 and the TikZ and matplotlib figure code, and proposing the design of the numerical checks. The LLM-generated code was run and its outputs (exact protection, support bounds, unbiasedness) were verified against the theorems by the authors. We have reviewed all AI-assisted work and take responsibility for the final content of this paper, including all text, claims, code, and figures produced with the aid of LLMs.

REPRODUCIBILITY STATEMENT

All claims are theorems with complete proofs in Appendices B to F; the main text states each result together with the assumptions under which it holds and the interfaces it does not cover (Table 2, Section 7). The numerical illustrations of Section 8 and Appendix G use a single self-contained Python script (NumPy/SciPy, fixed random seeds, CPU only; about ten minutes for Figure 2 and twenty for Figure 4) that is included in the supplementary material together with the LaTeX source; the instances (p, A) are constructed in the script and each figure is regenerated by one command.

REFERENCES

Paola Alimonti and Viggo Kann. Some APX-completeness results for cubic graphs. *Theoretical Computer Science*, 237(1–2):123–134, 2000. doi: 10.1016/S0304-3975(98)00158-3.

Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. *Journal of the ACM*, 63(2):12:1–12:35, 2016. doi: 10.1145/2837020.

Nikhil Bansal. On a generalization of iterated and randomized rounding. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1125–1135. Association for Computing Machinery, 2019. doi: 10.1145/3313276.3316313. URL https://doi.org/10.1145/3313276.3316313.

Leighton Barnes, Timothy Chow, Emma Cohen, Keith Frankston, Benjamin Howard, Fred Kochman, Daniel Scheinerman, and Jeffrey VanderKam. Efficient unbiased sparsification. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pp. 1854–1859. IEEE, 2024. doi: 10.1109/ISIT57864.2024.10619383. URL https://doi.org/10.1109/ISIT57864.2024.10619383.

Christian Bayer and Josef Teichmann. The proof of Tchakaloff's theorem. *Proceedings of the American Mathematical Society*, 134(10):3035–3040, 2006. doi: 10.1090/S0002-9939-06-08249-9. URL https://doi.org/10.1090/S0002-9939-06-08249-9.

Immanuel M. Bomze, Bo Peng, Yuzhou Qiu, and E. Alper Yıldırım. On tractable convex relaxations of standard quadratic optimization problems under sparsity constraints. *Journal of Optimization Theory and Applications*, 204(3):38, 2025. doi: 10.1007/s10957-024-02593-1. URL https://doi.org/10.1007/s10957-024-02593-1.

Rudi Cilibrasi, Leo van Iersel, Steven Kelk, and John Tromp. The complexity of the single individual SNP haplotyping problem. *Algorithmica*, 49(1):13–36, 2007. doi: 10.1007/s00453-007-0029-z. URL https://doi.org/10.1007/s00453-007-0029-z.

Jean-Claude Deville and Yves Tillé. Unequal probability sampling without replacement through a splitting method. *Biometrika*, 85(1):89–101, 1998. doi: 10.1093/biomet/85.1.89. URL https://doi.org/10.1093/biomet/85.1.89.

Jean-Claude Deville and Yves Tillé. Efficient balanced sampling: The cube method. *Biometrika*, 91(4):893–912, 2004. doi: 10.1093/biomet/91.4.893. URL https://doi.org/10.1093/biomet/91.4.893.

Jean-Claude Deville and Yves Tillé. Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128(2):569–591, 2005. doi: 10.1016/j.jspi.2003.11.011. URL https://doi.org/10.1016/j.jspi.2003.11.011.

Rajiv Gandhi, Samir Khuller, Srinivasan Parthasarathy, and Aravind Srinivasan. Dependent rounding and its applications to approximation algorithms. *Journal of the ACM*, 53(3):324–360, 2006. doi: 10.1145/1147954.1147956.

Michael R. Garey, David S. Johnson, and Larry Stockmeyer. Some simplified NP-complete graph problems. *Theoretical Computer Science*, 1(3):237–267, 1976. doi: 10.1016/0304-3975(76)90059-1.

Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, Heidelberg, 1988. doi: 10.1007/978-3-642-97881-4. URL https://doi.org/10.1007/978-3-642-97881-4.

Lukas Haverbeck, Carmen Amo Alonso, Andres Felipe Posada-Moreno, Sebastian Trimpe, and Marco Pavone. The risk of KV cache compression. *arXiv preprint arXiv:2607.01520*, 2026. doi: 10.48550/arXiv.2607.01520. URL https://arxiv.org/abs/2607.01520.

Daniel G. Horvitz and Donovan J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952. doi: 10.1080/01621459.1952.10483446.

Krishnateja Killamsetty, Durga S, Ganesh Ramakrishnan, Abir De, and Rishabh Iyer. GRAD-MATCH: Gradient matching based data subset selection for efficient deep model training. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5464–5474. PMLR, 2021. URL https://proceedings.mlr.press/v139/killamsetty21a.html.

Vahab Mirrokni, Renato Paes Leme, Adrian Vladu, and Sam Chiu-wai Wong. Tight bounds for approximate Carathéodory and beyond. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 2440–2448. PMLR, 2017.

Baharan Mirzasoleiman, Jeff Bilmes, and Jure Leskovec. Coresets for data-efficient training of machine learning models. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6950–6960. PMLR, 2020. URL https://proceedings.mlr.press/v119/mirzasoleiman20a.html.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017.

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

Aravind Srinivasan. Distributions on level-sets with applications to approximation algorithms. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 588–597. IEEE, 2001. doi: 10.1109/SFCS.2001.959935.

Yves Tillé and Louis-Paul Rivest. Existence and applications of finite-population samples that are exactly balanced. *Biometrika*, 112(4):asaf058, 2025. doi: 10.1093/biomet/asaf058. URL https://doi.org/10.1093/biomet/asaf058.

Raphaël Tinarrage. Sharp approximate Carathéodory theorem and application to iterated Delaunay refinement. *arXiv preprint arXiv:2606.25854*, 2026. doi: 10.48550/arXiv.2606.25854. URL https://arxiv.org/abs/2606.25854.

Guangjian Zhang. Exact recovery thresholds for weighted data selection in vector-valued linear regression. *arXiv preprint arXiv:2608.30254*, 2026. doi: 10.48550/arXiv.2608.30254. URL https://arxiv.org/abs/2608.30254.

Zhenyu Zhang, Ying Sheng, Tianyi Zhou, Tianlong Chen, Lianmin Zheng, Ruisi Cai, Zhao Song, Yuandong Tian, Christopher Ré, Clark Barrett, Zhangyang Wang, and Beidi Chen. H_2O: Heavy-hitter oracle for efficient generative inference of large language models. In *Advances in Neural Information Processing Systems 36*, 2023.

A THEOREM-LEVEL COMPARISON WITH THE CLOSEST RESULTS

Table 1 records what the statements closest to Theorem 1 prove, in the terms of Section 2. The citations were checked against the arXiv versions current at submission time: Proposition A.3 of Haverbeck et al. (2026), Theorem 3 and Lemma 4 of Tinarrage (2026), and Theorem 6.1 and Lemma 6.2 of Zhang (2026).

Table 1: Theorem-level comparison with the closest protected and concurrent results. Entries report what the cited statements prove; an omission is not an impossibility claim.

Statement	Output and protection	Guarantee	Budgets and sharpness	Construction
Haverbeck et al. (2026), Prop. A.3	Unbiased reweighting; chosen U-component of the barycenter fixed pathwise	$\mathbb{E}[\eta \otimes \eta] \preceq \frac{20}{9b}(I - \Pi_U)\Sigma(I - \Pi_U)$	Support $\leq b + 2r + 2$; $1/b$ rate, constant not sharp	Constrained randomized rounding
Tinarrage (2026), Thm. 3 / Lem. 4	Equal-weight tuple; under a coordinate cap, an unbiased equal-weight subset	Sharp deterministic radius $R_{d,t}$; capped pivotal second moment (scalar)	All tuple counts; unbiased law only under $p_i \leq 1/r$	Geometric; pivotal sampling
Zhang (2026), Thm. 6.1 / Lem. 6.2	Arbitrary-weight sparse convex combination; no protection	Mean-error coefficient $h_{s,t}$; one boundary law for all PSD tests	Every $1 \leq t \leq s$; sharp on the regular simplex	Existential boundary martingale; no bit-model statement
This paper	Random probability vector, $\mathbb{E}Q = p$, arbitrary $AQ = Ap$ pathwise	$\mathrm{Cov}(Q) \preceq h_{k-d,m-d} C_{p\|A}$ in Loewner order	Every $d + 1 \leq m \leq k$; worst-instance sharp at every budget; zero at $m = k$	Finite law; exact-rational Las Vegas path sampler; APX boundary

B GEOMETRIC AND COVARIANCE PROOFS

B.1 TANGENT FACTORIZATION

Fix a positive state q on a support of size s. Put $u = \sqrt{q}$, $P_0 = I - uu^\top$, and $R = P_0 \operatorname{diag}(\sqrt{q}) A^\top$. Then $R^\top R = AC_qA^\top$. If U is an orthonormal basis for $\ker(u^\top) \cap \ker(A \operatorname{diag}(\sqrt{q}))$, orthogonal projection gives

$$UU^\top = P_0 - R(R^\top R)^\dagger R^\top.$$

Multiplication on both sides by $\operatorname{diag}(\sqrt{q})$ yields

$$\operatorname{diag}(\sqrt{q}) UU^\top \operatorname{diag}(\sqrt{q}) = C_q - C_qA^\top (AC_qA^\top)^\dagger AC_q = C_{q|A}.$$

Since $q > 0$ on its support, $\operatorname{range}(C_q) = \ker(\mathbf{1}^\top)$ there, so $\operatorname{rank}(AC_qA^\top) = \dim A(\ker \mathbf{1}^\top) = \operatorname{rank}[\mathbf{1}^\top; A] - 1$; this is the identity $d = \operatorname{rank}[\mathbf{1}^\top; A] - 1$ used in Section 3. A rational full-column-rank basis T of the common kernel $\ker[\mathbf{1}^\top; A]$ gives the coordinate-free rational identity

$$C_{q|A} = T\big(T^\top \operatorname{diag}(q)^{-1} T\big)^{-1} T^\top.$$

B.2 PROOF OF THE ISOTROPIC BOUNDARY LEMMA

Figure 3 illustrates the construction. The isotropy assumptions imply that zero is in the interior of $\mathcal{K}$ and that $\mathcal{K}$ is bounded. Indeed, on every nonzero ray the values $a_i^\top u$ have mean zero and positive second moment, so some value is negative and the ray meets a facet.

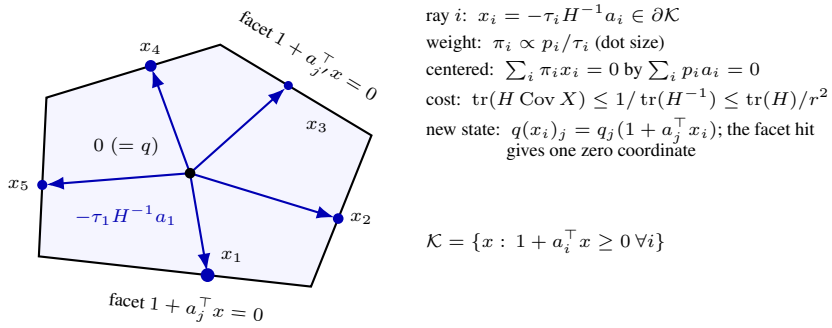


Figure 3: **The radial best response behind Lemma 2.** In tangent coordinates the protected fiber through q is the polytope $\mathcal{K}$ with facets $1 + a_i^\top x = 0$, one per support coordinate, and q sits at the origin. For a test matrix $H \succ 0$, a ray is shot from the origin in direction $-H^{-1}a_i$ to its first boundary point x_i, and the endpoint receives weight proportional to p_i/τ_i (inverse ray length). Isotropy makes the law centered and gives $\mathrm{tr}(H\,\mathrm{Cov}\,X) \le 1/\,\mathrm{tr}(H^{-1})$; a minimax over H produces one law valid in every direction. Every endpoint lies on a facet, so one support coordinate is dropped.

Fix $H \succ 0$. For every nonzero a_i, move from zero along $-H^{-1}a_i$ to the first boundary point

$$x_i = -\tau_i H^{-1} a_i, \qquad \tau_i > 0.$$

The i-th facet inequality at x_i gives $\tau_i a_i^\top H^{-1} a_i \le 1$. Define

$$Z = \sum_{a_i \neq 0} \frac{p_i}{\tau_i}, \qquad \pi_i = \frac{p_i/\tau_i}{Z}.$$

Zero normals may be omitted because they contribute to neither isotropic moment. The boundary law is centered:

$$\sum_i \pi_i x_i = -\frac{H^{-1}}{Z} \sum_i p_i a_i = 0.$$

Its cost obeys

$$\sum_i \pi_i x_i^\top H x_i = \frac{1}{Z} \sum_i p_i \tau_i a_i^\top H^{-1} a_i \le \frac{1}{Z}.$$

Moreover,

$$Z \ge \sum_i p_i a_i^\top H^{-1} a_i = \mathrm{tr}(H^{-1}).$$

The arithmetic–harmonic mean inequality therefore gives

$$\mathrm{tr}(H\,\mathrm{Cov}(X)) \le \frac{1}{\mathrm{tr}(H^{-1})} \le \frac{\mathrm{tr}(H)}{r^2}.$$

Let $\mathfrak{L}$ be the set of probability laws on the compact set $\partial\mathcal{K}$ with mean zero; it is convex and weakly compact. Let $\mathfrak{H} = \{H \succeq 0 : \mathrm{tr}(H) = 1\}$. For $\nu \in \mathfrak{L}$ the map $H \mapsto \mathrm{tr}(H\,\mathrm{Cov}_\nu\, X)$ is linear, and for fixed H the map $\nu \mapsto \mathrm{tr}(H\,\mathrm{Cov}_\nu\, X) = \int x^\top H x\, d\nu(x)$ is linear and weakly continuous. Sion's minimax theorem (Sion, 1958) therefore gives

$$\min_{\nu \in \mathfrak{L}} \max_{H \in \mathfrak{H}} \mathrm{tr}(H\,\mathrm{Cov}_\nu(X)) = \max_{H \in \mathfrak{H}} \min_{\nu \in \mathfrak{L}} \mathrm{tr}(H\,\mathrm{Cov}_\nu(X)).$$

For $H \succ 0$ the radial law shows the inner minimum is at most $\mathrm{tr}(H)/r^2 = r^{-2}$; for singular $H \in \mathfrak{H}$, apply the bound to $H_\epsilon = (H + \epsilon I)/(1 + r\epsilon)$, extract a weak limit of the radial laws as $\epsilon \downarrow 0$, and use continuity. Hence the right side is at most r^{-2}. The left side is $\min_\nu \lambda_{\max}(\mathrm{Cov}_\nu\, X)$, so a minimax law satisfies $\mathrm{Cov}\, X \preceq r^{-2} I_r$. Finally, Carathéodory's theorem applied to the moment map $x \mapsto (1, x, xx^\top)$ replaces the minimax law by a law on at most $1 + r + r(r+1)/2$ boundary points with the same mean and second moment.

B.3 Proof of the residual bridge

Conditionally on P, draw $I \sim P$. Marginally $I \sim p$. Set

$$X = e_I - p, \quad Y = AX, \quad L_p = C_p A^\top (AC_p A^\top)^\dagger, \quad R = X - L_p Y.$$

Then $\mathrm{Cov}(R) = C_{p|A}$. Since $AP = Ap$ pathwise, $\mathbb{E}[R \mid P] = P - p$. For $P = q$, the centered conditional residual is $(I - L_p A)(e_I - q)$. With $L_q = C_q A^\top (AC_q A^\top)^\dagger$, completion of squares gives

$$(I - LA)C_q(I - LA)^\top = C_{q|A} + (L - L_q)(AC_q A^\top)(L - L_q)^\top.$$

Thus the conditional covariance using L_p dominates $C_{q|A}$. The total-covariance identity now yields

$$C_{p|A} = \mathrm{Cov}(R) \succeq \mathrm{Cov}(P) + \mathbb{E}C_{P|A}.$$

B.4 Recurrence identity and sharpness

For $K = k - d$ and $M = t - d$, the update $\eta_{t-1} = \alpha_t + (1 - \alpha_t)\eta_t$ with $\alpha_t = (t - d - 1)^{-2}$ is equivalent to

$$1 - \eta_{t-1} = \left(1 - \frac{1}{(M-1)^2}\right)(1 - \eta_t).$$

Multiplying these factors and using

$$\prod_{j=M}^{K-1} \left(1 - \frac{1}{j^2}\right) = \frac{(M-1)K}{M(K-1)}$$

gives $\eta_m = 1 - (M-1)K/(M(K-1)) = h_{K,M}$.

Unprotected lower bound. For $d = 0$ and uniform $p = \mathbf{1}/k$, any unbiased m-sparse Q has $\|Q\|_2^2 \geq 1/m$ pathwise, so $\mathrm{tr}\,\mathrm{Cov}(Q) = \mathbb{E}\|Q\|_2^2 - 1/k \geq 1/m - 1/k$, while $\mathrm{tr}\, C_p = 1 - 1/k$. If $\mathrm{Cov}(Q) \preceq c\, C_p$ then taking traces gives $c \geq (1/m - 1/k)/(1 - 1/k) = h_{k,m}$. The uniform m-subset law has $\mathrm{Cov}(Q) = h_{k,m} C_p$ exactly, by symmetry and the trace identity.

Protected lower bound. For $d > 0$, take

$$p_i = \delta/d \quad (i \leq d), \qquad p_i = (1 - \delta)/K \quad (i > d), \qquad A = [I_d \;\; 0],$$

so that $AC_p A^\top = \mathrm{diag}(p_{1:d}) - p_{1:d} p_{1:d}^\top$ has rank d because $\sum_{i \leq d} p_i = \delta < 1$. Every admissible Q has $Q_i = p_i > 0$ for $i \leq d$ pathwise, so d support positions are consumed and the free block $Q_F = (Q_i)_{i>d}$ is M-sparse with $\mathbf{1}^\top Q_F = 1 - \delta$. Then $\widetilde{Q} = Q_F/(1 - \delta)$ is an unbiased M-sparse probability vector on K coordinates with uniform mean, and $\mathrm{Cov}(Q)$ is $(1 - \delta)^2\, \mathrm{Cov}(\widetilde{Q})$ embedded in the free block. On the other side, the regression of e_I on the d anchor indicators equals the conditional expectation given the event $\{I \leq d\}$ and, on it, the anchor index; hence $C_{p|A} = \mathbb{E}\,\mathrm{Cov}(e_I \mid \text{that information}) = (1 - \delta)\, C_{\mathrm{unif,K}}$ embedded in the free block, with $C_{\mathrm{unif,K}} = (I - \mathbf{1}\mathbf{1}^\top/K)/K$. If $\mathrm{Cov}(Q) \preceq c\, C_{p|A}$, taking traces on the free block and using the unprotected bound for $\widetilde{Q}$ gives $(1 - \delta)^2 h_{K,M}\, \mathrm{tr}\, C_{\mathrm{unif,K}} \leq c\,(1 - \delta)\, \mathrm{tr}\, C_{\mathrm{unif,K}}$, that is $c \geq (1 - \delta) h_{K,M}$. Letting $\delta \downarrow 0$ shows that the supremum over instances of effective rank d is at least $h_{K,M}$, which matches the upper bound of Theorem 1.

The theorem is not a pointwise equality. With

$$k = 3, \quad p = (1/3, 1/3, 1/3), \quad A = (0, 1, 3), \quad m = 2,$$

the two fiber endpoints are $q^{(1)} = (5/9, 0, 4/9)$ and $q^{(2)} = (0, 5/6, 1/6)$. The only centered endpoint law uses probabilities $(3/5, 2/5)$. For $w = (2, -3, 1)^\top$,

$$\mathrm{Cov}(Q) = \frac{1}{54} ww^\top, \qquad C_{p|A} = \frac{1}{42} ww^\top, \qquad \mathrm{Cov}(Q) = \frac{7}{9} C_{p|A},$$

whereas $h_{2,1} = 1$.

C RATIONAL SAMPLER DETAILS

This appendix proves the bit-model claims in Theorem 5. All arithmetic is exact over the rationals. Randomness consists only of fair bits; expected polynomial time therefore does not mean a worst-case bound on the number of random bits.

C.1 EXACT CATEGORICAL CHOICES AND RADIAL RESPONSES

Lemma 7 (Exact rational categorical draw). *For rational probabilities $\rho_j = a_j/b_j \geq 0$, $b_j > 0$, with $\sum_{j=1}^{J} \rho_j = 1$, there is a fair-bit Las Vegas algorithm that returns j with probability exactly ρ_j. Its expected bit and arithmetic cost is polynomial in the aggregate encoding length of the distribution.*

Proof. Take the unreduced common denominator

$$D = \prod_{j=1}^{J} b_j, \qquad A_j = a_j D/b_j.$$

Then $\sum_j A_j = D$ and $\log_2 D \leq \sum_j \log_2 b_j$. For $D > 1$, draw $\ell = \lceil \log_2 D \rceil$ bits as a uniform integer $U \in \{0, \ldots, 2^\ell - 1\}$, reject when $U \geq D$, and otherwise return the interval containing U among the cumulative A_j's. Acceptance has probability $D/2^\ell > 1/2$. Thus fewer than two trials are needed in expectation, and the accepted output has the prescribed law exactly. The case $D = 1$ is deterministic. □

Lemma 8 (Rational radial response). *Let $q \in \mathbb{Q}^s_{>0}$, let $T \in \mathbb{Q}^{s\times r}$ have full column rank with $\mathbf{1}^\top T = 0$ and $AT = 0$, and write $b_i^\top$ for row i of T. Put*

$$G = T^\top \operatorname{diag}(q)^{-1} T.$$

For rational $H \succ 0$, omit zero rows and define

$$u_i = H^{-1} b_i, \qquad \lambda_i = \max_{j \in \operatorname{supp}(q)} \frac{b_j^\top u_i}{q_j}, \qquad z_i = -u_i/\lambda_i, \qquad \pi_i = \frac{\lambda_i}{\sum_{h: b_h \neq 0} \lambda_h}.$$

Then $q + Tz_i$ is a rational boundary point of the same protected fiber, $\sum_i \pi_i z_i = 0$, and, for $V_H = \sum_i \pi_i z_i z_i^\top$,

$$\langle H, V_H \rangle \leq \frac{1}{\operatorname{tr}(H^{-1}G)} \leq \frac{\operatorname{tr}(HG^{-1})}{r^2}. \tag{24}$$

The law and all its probabilities have encoding length polynomial in $\operatorname{bits}(q, T, H)$ and are computable in rational polynomial time.

Proof. The $j = i$ term shows $\lambda_i > 0$. Maximality of λ_i gives $q_j + b_j^\top z_i \geq 0$ for every j, with equality for a maximizing coordinate. Moreover,

$$\sum_i \pi_i z_i = -\frac{H^{-1}}{\sum_h \lambda_h} \sum_i b_i = 0.$$

Writing $Z = \sum_i \lambda_i$, the same maximality gives

$$\langle H, V_H \rangle = \frac{1}{Z} \sum_i \frac{b_i^\top H^{-1} b_i}{\lambda_i} \leq \frac{1}{Z}, \qquad Z \geq \sum_i \frac{b_i^\top H^{-1} b_i}{q_i} = \operatorname{tr}(H^{-1}G).$$

The second inequality in Eq. (24) is the arithmetic–harmonic mean inequality applied to $H^{1/2} G^{-1} H^{1/2}$. Fraction-free elimination computes $H^{-1} b_i$, and exact cross multiplication finds every maximum. Determinant bounds for rational linear systems give polynomial output height. Finally, Lemma 7 samples the law exactly. □

C.2 A RATIONAL SPECTRAHEDRAL MINIMAX LEMMA

Lemma 9 (Rational one-step minimax)**.** *Let $r \geq 1$, $G \in \mathbb{Q}^{r\times r}$ be symmetric positive definite, and $0 < \theta \leq 1$ be rational. Set*

$$c = \frac{1}{r^2}, \qquad \delta = \frac{\theta}{4r}, \qquad \gamma = \frac{\theta c}{32},$$
$$\mathcal{A}_G = \{X \in \mathbb{S}^r : \langle G, X\rangle = 1\}, \qquad \mathcal{K}_\delta = \{X \in \mathcal{A}_G : X \succeq \delta G^{-1}\}.$$

Suppose a rational query $X \in \mathcal{K}_\delta$ returns, in time and height polynomial in the static input and $\mathrm{bits}(X)$, a rational $V_X \succeq 0$ such that

$$\langle X, GV_XG\rangle \leq c. \tag{25}$$

Then a deterministic rational Turing algorithm makes polynomially many queries and returns rational weights α_j and queried responses such that

$$\bar{V} = \sum_j \alpha_j V_{X_j} \preceq (1+\theta)cG^{-1}. \tag{26}$$

The number of calls, runtime, and output height are polynomial in the static input length and $\mathrm{bits}(\theta)$.

Proof. The case $r = 1$ has $\mathcal{K}_\delta = \{G^{-1}\}$, so one oracle call gives $V_X \preceq G^{-1}$. Assume $r \geq 2$ and let

$$D = \frac{r(r+1)}{2} - 1, \quad U = \max_i \sum_j |G_{ij}|, \quad \ell = \frac{\det G}{U^{r-1}}, \quad X_0 = \frac{1}{r}G^{-1}.$$

The positive rationals U, ℓ obey $\lambda_{\max}(G) \leq U$ and $\lambda_{\min}(G) \geq \ell$. In the D-dimensional affine space $\mathcal{A}_G$, with relative Frobenius balls,

$$B_{\mathcal{A}_G}(X_0, \rho_0) \subseteq \mathcal{K}_\delta \subseteq B_{\mathcal{A}_G}(X_0, R_0), \qquad \rho_0 = \frac{1/r - \delta}{2U}, \quad R_0 = \frac{2}{\ell}. \tag{27}$$

Indeed, a tangent perturbation Z satisfies $\langle G, Z\rangle = 0$; the first inclusion follows from

$$X_0 + Z - \delta G^{-1} \succeq ((1/r - \delta)/U - \|Z\|_F)\, I,$$

and the second follows by conjugating X with $G^{1/2}$, obtaining a positive semidefinite matrix of trace one. A rational basis for $\{Z : \langle G, Z\rangle = 0\}$ is obtained by eliminating any nonzero coordinate of the vectorization of G. Its Gram matrix and inverse have polynomial height, so Eq. (27) yields rational coordinate radii with polynomial logarithmic ratio.

For rational $X \in \mathcal{A}_G$, fraction-free $LDL^\top$/Bareiss elimination either certifies $M = X - \delta G^{-1} \succeq 0$ or returns a rational v with $v^\top Mv < 0$. Besides a negative pivot, a zero diagonal with nonzero off-diagonal $a = M_{ij}$ and $b = M_{jj} \geq 0$ is exposed by $v = e_i - ae_j/(b+1)$, for which $v^\top Mv = -a^2(b+2)/(b+1)^2 < 0$; Schur-complement witnesses lift by rational back substitution. The latter gives the exact separating inequality

$$\langle vv^\top, Y\rangle \geq \delta v^\top G^{-1} v, \qquad Y \in \mathcal{K}_\delta.$$

All pivots and lifted negative-curvature vectors are quotients of input minors and therefore have polynomial height.

Write $S_X = GV_XG \succeq 0$. From Eq. (25) and $X \succeq \delta G^{-1}$,

$$\mathrm{tr}(GV_X) \leq c/\delta, \qquad \|S_X\|_F \leq \mathsf{L}, \qquad \mathsf{L} = \frac{Uc}{\delta}. \tag{28}$$

Run the rational outward-rounded central-cut ellipsoid method (Grötschel et al., 1988) in $\mathcal{A}_G$. Besides PSD cuts, retain each oracle column with its fixed-gap constraint

$$\mathcal{H}_j = \{X \in \mathcal{A}_G : \langle X, S_j\rangle \geq c + 2\gamma\}. \tag{29}$$

At a center satisfying the spectrahedron and all retained constraints, query the oracle and add its new column; use only the central relaxation $\langle Y - X, S_j\rangle \geq 0$ for the ellipsoid update. Before adding it, test exactly whether $S_j \in \mathrm{span}\{G\}$. In that special case its payoff is constant on $\mathcal{A}_G$, is at most c at the query, and the single column is already valid. Otherwise its relative cut normal is nonzero.

For the volume certificate define

$$a_0 = \frac{1/r - \delta}{U}, \qquad \eta_0 = \frac{\gamma}{2(c+4\gamma)}, \qquad \rho_* = \min\left\{\frac{\eta_0 a_0}{2}, \frac{\gamma}{2\mathsf{L}}\right\}. \tag{30}$$

If some $X^* \in \mathcal{K}_\delta$ met every retained column at level $c + 4\gamma$, then $\widehat{X} = (1-\eta_0)X^* + \eta_0 X_0$ would have PSD slack at least $\eta_0 a_0 I$ and payoff at least $c + 7\gamma/2$. Hence the relative ρ_*-ball about $\widehat{X}$ would satisfy all constraints in Eq. (29). After

$$O\left(D^2 \log \frac{R_0}{\rho_*}\right)$$

rounded central cuts, the containing ellipsoid has smaller volume than that ball, a contradiction. Consequently the polynomially many retained columns satisfy

$$\max_{X \in \mathcal{K}_\delta} \min_j \langle X, S_j \rangle \le c + 4\gamma. \tag{31}$$

The rounded central-cut ellipsoid method of Grötschel et al. (1988, Thm. 3.2.1), run in the D-dimensional coordinates of $\mathcal{A}_G$ with outer radius R_0 and volume target $\operatorname{vol} B_{\mathcal{A}_G}(\cdot, \rho_*)$, fixes its iteration count $N_{\mathrm{ell}} = O(D^2 \log(R_0/\rho_*))$ and its truncation precision $O(N_{\mathrm{ell}})$ binary digits from these two rationals alone; both are polynomial in the input length because $\log(R_0/\rho_*)$ is polynomial in $\mathrm{bits}(G, \theta)$. Thus stored centers and shape matrices have one static polynomial height bound; the raw height of a processed cut affects runtime, not the precision of subsequent centers.

Finite minimax turns Eq. (31) into a rational SDP. For $S(\alpha) = \sum_j \alpha_j S_j$ and $\tau = 1 - r\delta$,

$$\max_{X \in \mathcal{K}_\delta} \langle X, S(\alpha) \rangle = \delta \langle G^{-1}, S(\alpha) \rangle + \tau \lambda_{\max}(G^{-1/2} S(\alpha) G^{-1/2}),$$

so it suffices to solve

$$\begin{aligned} \min_{\alpha, \mu} \quad & \delta \langle G^{-1}, S(\alpha) \rangle + \tau \mu \\ \text{s.t.} \quad & \alpha \ge 0, \quad \mathbf{1}^\top \alpha = 1, \quad \mu G - S(\alpha) \succeq 0. \end{aligned} \tag{32}$$

If N columns were retained, strict Slater data are $\alpha_j^0 = 1/N$, $\mu^0 = 2\mathsf{L}/\ell$; the LMI slack is at least $\mathsf{L}I$. An optimum has $\mu \le \mathsf{L}/\ell$, so adding $0 \le \mu \le 3\mathsf{L}/\ell$ makes the feasible set bounded.

For completeness, an explicit inner radius for the target with objective at most $c + 5\gamma$ is obtained by mixing an optimum with this Slater point by

$$\zeta = \min\left\{\frac{1}{4}, \frac{\gamma}{4(1 + c + 4\gamma + 3\mathsf{L}/\ell)}\right\},$$

and taking

$$\sigma = \min\left\{\frac{\zeta}{2N}, \frac{\zeta \mathsf{L}}{2(U + N\mathsf{L})}, \frac{\gamma}{4(1+cN)}, \frac{\zeta \mathsf{L}}{\ell}, \frac{\mathsf{L}}{2\ell}\right\}. \tag{33}$$

The five terms preserve, respectively, positive weights, half the LMI slack, the objective slack, and the lower and upper bounds on μ. After eliminating one simplex coordinate, a safe inner radius is $\rho_{\mathrm{SDP}} = \sigma/N$, while an outer radius is $R_{\mathrm{SDP}} = 2 + 3\mathsf{L}/\ell$. Rational weak optimization with exact scalar and $LDL^\top$ separation therefore returns an exactly feasible rational (α, μ) of polynomial height and objective at most $c + 5\gamma$.

Finally put $W = G^{1/2} \bar{V} G^{1/2}$. The SDP objective bound implies

$$(1 - r\delta)\lambda_{\max}(W) \le c + 5\gamma,$$

and hence

$$\lambda_{\max}(W) \le \frac{1 + 5\theta/32}{1 - \theta/4}\, c \le (1+\theta)c.$$

This is exactly Eq. (26). Every parameter and radius displayed above is rational and has polynomial binary length, which also proves the claimed Turing bounds. □

To instantiate Lemma 9 at a rational fiber state, use $H = GXG$ in Lemma 8; the floor $X \succeq \delta G^{-1}$ is what makes $H \succ 0$ there. Since $\operatorname{tr}(HG^{-1}) = \langle G, X \rangle = 1$,

$$\langle X, G V_H G \rangle = \langle H, V_H \rangle \le 1/r^2.$$

The radial height bound supplies the oracle-height hypothesis. Mixing the queried radial laws preserves the exact conditional mean and protected fiber, and every outcome loses a support coordinate.

C.3 A BOUNDED-HEIGHT UNBIASED RESET

Lemma 10 (Rational face reset)**.** *On a fixed support S, let*

$$F_S = \{x \in \mathbb{R}_+^S : \mathbf{1}^\top x = 1,\ Ax = Ap\}.$$

Given $x \in F_S \cap \mathbb{Q}^S$, one can compute in rational polynomial time

$$x = \sum_{j=1}^{J} \lambda_j v^j, \qquad J \le |S|,$$

where the v^j are basic feasible vertices of F_S and $\operatorname{supp}(v^j) \subseteq \operatorname{supp}(x)$. For any integer $N \ge 1$, there is an exact rational random reset $\widetilde{x}$ satisfying

$$\mathbb{E}[\widetilde{x} \mid x] = x, \qquad \operatorname{supp}(\widetilde{x}) \subseteq \operatorname{supp}(x), \qquad \|\widetilde{x} - x\|_2 \le J/N \le k/N. \tag{34}$$

Every realized reset has coordinate height at most

$$B_{\mathrm{reset}} \le \log N + JB_{\mathrm{vert}} + O(J \log J),$$

where $B_{\mathrm{vert}} = \mathrm{poly}(k, a, L)$ depends only on the original rational face description, not on the height of x.

Proof. Work in the smallest face containing x. Choose a basic feasible vertex v, extend the ray from v through x to the next boundary, and recurse on the lower-dimensional face. At most $|S| - 1$ dimension drops give at most $|S|$ vertices. Exact ratio tests implement every step. More explicitly, at one drop write the current point as y, and let coordinate i be the first coordinate hit by the ray. The new boundary point has coordinates

$$y'_j = \frac{v_i y_j - y_i v_j}{v_i - y_i}.$$

Writing $v = V/D_v$ and $y = Y/D_y$ with shared denominators, the factor $D_v D_y$ cancels before reduction and all y'_j share denominator $V_i D_y - Y_i D_v$. Thus $B(y') \le B(y) + B(v) + O(1)$: one drop adds only polynomially many bits, and the at-most-$|S| - 1$ drops do not create a repeated-squaring height recurrence. After redundant equality rows are removed, Cramer's rule applied to $[\mathbf{1}^\top; A]v = [1; Ap]$ gives the uniform bound B_{vert}.

Put $a_j = \lfloor N\lambda_j \rfloor$, $f_j = N\lambda_j - a_j$, $F_0 = 0$, and $F_j = \sum_{\ell \le j} f_\ell$. For $u \in [0, 1)$, define

$$\xi_j(u) = \lfloor F_j + u \rfloor - \lfloor F_{j-1} + u \rfloor.$$

Here

$$h_0 := F_J = \sum_j f_j = N - \sum_j a_j \in \mathbb{Z}, \qquad \xi_j(u) \in \{0, 1\}, \qquad \sum_j \xi_j(u) = h_0,$$

and $\int_0^1 \xi_j(u)\, du = f_j$. Consequently the coefficients $(a_j + \xi_j)/N$ are nonnegative and sum to one for every realized u; hence $\widetilde{x}$ below remains in F_S. This vector changes only at the rational breakpoints $(-F_j) \bmod 1$, so Lemma 7 samples one of its at most J constant intervals by exact rational length. Since $\mathbb{E}\xi_j = f_j$, the random point

$$\widetilde{x} = \sum_j \frac{a_j + \xi_j}{N} v^j$$

is conditionally unbiased. Also $\sum_j |a_j + \xi_j - N\lambda_j| \le J$; since $\|v^j\|_2 \le \|v^j\|_1 = 1$, this proves Eq. (34). Its realized coefficients have denominator N; incoming decomposition denominators occur only in the categorical probabilities. This gives the stated height bound. □

C.4 SPECTRAL AND COVARIANCE LEDGERS

Lemma 11 (Computable residual spectral floor)**.** *Let $r_0 \ge 1$, and let $T_0 \in \mathbb{Q}^{k \times r_0}$ be a full-column-rank basis of the global balanced tangent space, and put*

$$R_T = T_0^\top T_0, \qquad G_0 = T_0^\top \operatorname{diag}(p)^{-1} T_0, \qquad C_0 = T_0 G_0^{-1} T_0^\top = C_{p|A}.$$

With

$$\Lambda_T = \max_i \sum_j |(R_T)_{ij}|, \qquad \Lambda_G = \max_i \sum_j |(G_0)_{ij}|,$$

define

$$\underline{\lambda} = \min\left\{1, \frac{\det R_T}{\Lambda_T^{r_0-1}\Lambda_G}\right\}.$$

Then $C_0 \succeq \underline{\lambda} P_{\mathrm{range}(T_0)}$, and $\log(1/\underline{\lambda})$ is polynomial in the rational input length. If $\widetilde{x}$ is the reset of Lemma 10 with

$$N = \left\lceil \frac{k+1}{\beta\underline{\lambda}} \right\rceil, \qquad 0 < \beta \le 1,$$

then

$$\mathrm{Cov}(\widetilde{x} \mid x) \preceq \beta C_0. \tag{35}$$

Proof. The nonzero eigenvalues of C_0 are those of $G_0^{-1/2} R_T G_0^{-1/2}$. Gershgorin's bound and the determinant identity give

$$\lambda_{\min}^{+}(C_0) \ge \frac{\lambda_{\min}(R_T)}{\lambda_{\max}(G_0)} \ge \frac{\det R_T}{\Lambda_T^{r_0-1}\Lambda_G}.$$

Clearing rational denominators makes the nonzero determinant an integer divided by a denominator of polynomial bit length, proving the height claim. Reset errors lie in $\mathrm{range}(T_0)$, so

$$\mathrm{Cov}(\widetilde{x} \mid x) \preceq \frac{k^2}{N^2} P_{\mathrm{range}(T_0)} \preceq \frac{k^2}{N^2\underline{\lambda}} C_0 \preceq \beta C_0.$$

□

Proof of Theorem 5. If $m = k$, return p before dividing by $R = k - m$. If $K = 1$, this is the only feasible case. If $M = 1$, use the vertex decomposition in Lemma 10 without rounding and sample one basic feasible vertex. Since

$$\mathrm{rank}\begin{bmatrix}\mathbf{1}^\top \\ A\end{bmatrix} = d + 1,$$

every such vertex has support at most $d+1 = m$; the bridge Lemma 3 gives $\mathrm{Cov}(Q) \preceq C_0 = h_{K,1} C_0$.

Now assume $m < k$ and $M \ge 2$. Set

$$R = k - m, \qquad \theta = \varepsilon/8, \qquad \beta = \frac{\varepsilon h_{K,M}}{4R}.$$

Here $K > M \ge 2$, so the global tangent rank is $r_0 = K - 1 \ge 2$. Choose the global reset denominator from Lemma 11. At every active state, instantiate Lemma 9 with the rational radial oracle, sample its two categorical levels using Lemma 7, and then apply the reset. A radial outcome drops support and a reset never enlarges it, so every path has at most R active stages. At support $t > m$, with effective protected rank $d_t \le d$, the tangent dimension obeys

$$r_t = t - d_t - 1 \ge t - d - 1 \ge M \ge 2.$$

Let $j = t - d - 1$. The one-step coefficient is at most $\alpha_j = (1 + \theta)/j^2$. Total covariance and Lemma 3 and Eq. (35) give the scalar ledger

$$\eta' \le \alpha_j + (1 - \alpha_j)\eta + \beta.$$

Skipped support levels can be padded by identity transitions. Without resets, iteration from $j = K-1$ to M gives

$$\widehat{h} = 1 - \prod_{j=M}^{K-1}\left(1 - \frac{1+\theta}{j^2}\right), \qquad h_{K,M} = 1 - \prod_{j=M}^{K-1}\left(1 - \frac{1}{j^2}\right).$$

Telescoping the difference and using $\sum_{j=M}^{K-1} j^{-2} \le 2h_{K,M}$ yields

$$\widehat{h} - h_{K,M} \le 2\theta h_{K,M} = \frac{\varepsilon}{4} h_{K,M}.$$

The at most R resets add at most $R\beta C_0 = \varepsilon h_{K,M} C_0/4$. Therefore

$$\mathrm{Cov}(Q) \preceq (1+\varepsilon/2) h_{K,M} C_0 \preceq (1+\varepsilon) h_{K,M} C_0.$$

After every reset the realized state has encoding length $\mathrm{poly}(k, a, L + \mathrm{enc}(\varepsilon))$, independently of the incoming denominator. Lemma 9 then gives one uniform polynomial height bound for the next transition. There are at most k stages, and every categorical draw has expected polynomial cost by Lemma 7. Thus the total expected runtime and every realized output length have the bounds stated in Theorem 5. The finite terminal law may still have exponentially many leaves.

D Additional fixed-query complexity proofs

Table 2 summarizes what Theorem 6 and this appendix establish for the fixed-W problem, by output interface and instance family.

Table 2: Established fixed-W complexity by output interface. "No PTAS"/"no FPTAS" entries assume $\mathsf{P} \neq \mathsf{NP}$; "open" marks a claim outside the present reductions. Columns: pair-protected PSD W_G (Theorem 6); pair-protected positive definite $W_{G,\eta}$ with polynomial condition number (Appendix D.1); and unprotected $A = 0$ with near-identity W (Lemma 12).

Output interface	Pair-prot., PSD	Pair-prot., PD, poly-cond.	$A = 0$, near-identity PD
Exact threshold decision	NP-hard	NP-hard (known offset)	NP-hard
Additive value estimate	NP-hard below $2/n^2$	NP-hard (known offset)	NP-hard below $1/k^2$
Explicit-law approximation	no PTAS	no PTAS	no FPTAS; PTAS open
Value estimate, fixed ratio	open	open	open
Black-box sampler	needs randomized assumpt.	open	open

D.1 Positive-definite protected variant

Use the pair-sum and pair-difference matrices B, D from Theorem 6 and let

$$W_{G,\eta} = B^\top B + D^\top (\mathcal{S}_G + \eta I) D, \qquad \eta = n^{-4}.$$

For every x,

$$x^\top W_{G,\eta} x = \|Bx\|_2^2 + (Dx)^\top (\mathcal{S}_G + \eta I)(Dx),$$

which is zero only if $Bx = Dx = 0$, hence $x = 0$. The spectrum is

$$\{2\}_{i=1}^n \cup \{2(\lambda_i(\mathcal{S}_G) + \eta)\}_{i=1}^n.$$

For cubic G, $\kappa_2(W_{G,\eta}) \le (6+\eta)/\eta = \mathrm{poly}(n)$. On the protected Boolean fiber, $B(q_z - p) = 0$ and $\|Dq_z\|_2^2 = 1/n$, so

$$\mathrm{OPT}(p, A, n; W_{G,\eta}) = \frac{\eta}{n} + \frac{4\,\mathrm{MinUncut}(G)}{n^2}. \tag{36}$$

If a feasible explicit law has average uncut count $\bar{U}$ and ratio r, then

$$\bar{U} \le r U(G) + (r-1)\frac{\eta n}{4}.$$

Bipartite inputs can be recognized separately. On nonbipartite inputs $U(G) \ge 1$, and the additive ratio loss for $\eta = n^{-4}$ is at most $(r-1)/(4n^3)$. A PTAS for this polynomially conditioned positive-definite explicit-law family would therefore give a PTAS for Cubic Minimum Uncut. The offset construction is not ratio-lossless and does not have a uniformly bounded condition number.

D.2 Direct unprotected additive hardness

Lemma 12 (Direct unprotected additive gap)**.** *Let G be a simple unweighted graph on an even number $k \ge 4$ of vertices. Set*

$$m = k/2, \qquad p = \mathbf{1}/k, \qquad \Lambda = 8k^3, \qquad W = \Lambda I + \mathcal{L}_G.$$

Then $W \succ 0$,

$$\kappa_2(W) \le 1 + \frac{2(k-1)}{\Lambda} < 1 + \frac{1}{4k^2},$$

and, writing $b(G)$ for the minimum-bisection value,

$$\frac{\Lambda}{k} + \frac{4b(G)}{k^2} - \frac{1}{2k^2} \le \mathrm{OPT}(p, 0, k/2; W) \le \frac{\Lambda}{k} + \frac{4b(G)}{k^2}. \tag{37}$$

Consequently, exact threshold comparison and deterministic $1/k^2$-additive evaluation are NP-hard on this family. Moreover, there is no deterministic FPTAS for a certified explicit feasible law, or for a numerical value under the standard relative-error guarantee, unless $\mathsf{P} = \mathsf{NP}$.

Proof. For the upper bound, let I be a minimum bisection and define

$$u_I = \frac{1}{m}\mathbf{1}_I, \qquad u_{I^c} = \frac{1}{m}\mathbf{1}_{I^c}.$$

The law assigning probability $1/2$ to each vector is feasible and has mean p. For $x_I = u_I - p$,

$$\|x_I\|_2^2 = \frac{1}{k}, \qquad x_I^\top \mathcal{L}_G x_I = \frac{4|E(I, I^c)|}{k^2},$$

and $u_{I^c} - p = -x_I$. This proves the upper bound.

For the lower bound, fix a feasible atom q and put $s = |\operatorname{supp}(q)| \le m$. If $s < m$, then

$$\|q - p\|_2^2 = \|q\|_2^2 - \frac{1}{k} \ge \frac{1}{s} - \frac{1}{k} \ge \frac{1}{k} + \frac{1}{m(m-1)}.$$

Since $\mathcal{L}_G \succeq 0$,

$$(q-p)^\top W (q-p) \ge \frac{\Lambda}{k} + \frac{\Lambda}{m(m-1)}.$$

Here

$$\frac{\Lambda}{m(m-1)} = \frac{32k^2}{k-2} > 1 \ge \frac{4b(G)}{k^2},$$

so this is stronger than the desired lower bound.

It remains to consider $s = m$. Let $I = \operatorname{supp}(q)$, and write

$$u_I = \frac{1}{m}\mathbf{1}_I, \qquad x_I = u_I - p, \qquad \Delta = q - u_I.$$

Then Δ is supported on I, satisfies $\mathbf{1}^\top \Delta = 0$, and is orthogonal to x_I. Therefore $\|q - p\|_2^2 = 1/k + \|\Delta\|_2^2$. Completing the square and using $\mathcal{L}_G \succeq 0$ gives

$$\begin{aligned}(q-p)^\top W (q-p) &= \frac{\Lambda}{k} + x_I^\top \mathcal{L}_G x_I + \Lambda\|\Delta\|_2^2 + 2\Delta^\top \mathcal{L}_G x_I + \Delta^\top \mathcal{L}_G \Delta \\ &\ge \frac{\Lambda}{k} + \frac{4|E(I, I^c)|}{k^2} - \frac{\|\mathcal{L}_G x_I\|_2^2}{\Lambda}.\end{aligned}$$

Because I is a bisection, $|E(I, I^c)| \ge b(G)$. Moreover,

$$\frac{\|\mathcal{L}_G x_I\|_2^2}{\Lambda} \le \frac{\|\mathcal{L}_G\|_{\mathrm{op}}^2 \|x_I\|_2^2}{\Lambda} \le \frac{4(k-1)^2/k}{8k^3} \le \frac{1}{2k^2}.$$

Thus every feasible atom obeys the lower bound in Eq. (37); averaging preserves it.

For the complexity consequences, reduce an integer-threshold Minimum Bisection instance (G, C). Define

$$a_C = \frac{\Lambda}{k} + \frac{4C}{k^2}, \qquad c_C = \frac{\Lambda}{k} + \frac{4C + 7/2}{k^2}, \qquad t_C = a_C + \frac{7}{4k^2}.$$

Since $b(G)$ is integral,

$$b(G) \le C \Longrightarrow \mathrm{OPT} \le a_C < t_C, \qquad b(G) \ge C + 1 \Longrightarrow \mathrm{OPT} \ge c_C > t_C.$$

This proves exact-threshold hardness. If $|\widehat{v} - \mathrm{OPT}| \le 1/k^2$, then the two cases remain separated by t_C, proving additive hardness.

Finally, $b(G) \le k^2/4$ gives $\mathrm{OPT} \le 8k^2 + 1$. An FPTAS called with $\varepsilon_{\mathrm{F}} = 1/(9k^4)$ would have absolute error at most

$$\varepsilon_{\mathrm{F}}\,\mathrm{OPT} \le \frac{8k^2+1}{9k^4} \le \frac{1}{k^2},$$

while $1/\varepsilon_{\mathrm{F}}$ is polynomial in k. This contradicts the additive-hardness consequence. □

This lemma does not prove a constant-factor barrier, a no-PTAS result, hardness for diagonal W, or randomized hardness under the same assumption.

D.3 Output interfaces

An explicit feasible-law algorithm returns rational $\sum_t \lambda_t \delta_{q^t}$ of polynomial total encoding length with all pathwise constraints and $\sum_t \lambda_t q^t = p$ exact. A certified feasible upper value accompanied by such a law has the same extraction power. A bare number, threshold decision, and implicit sampler are different outputs. For the pair-protected instance, an implicit sampler PTAS would give a randomized PTAS for Minimum Uncut by independent repetition and selection of the cheapest sampled cut, but excluding it requires an explicit randomized-complexity assumption rather than only $\mathsf{P} \neq \mathsf{NP}$.

E Fixed-deployment learning consequences

Let $B = \mathrm{Cov}(Q) \preceq hC_{p|A}$ and $F_w(\theta) = \sum_i w_i f_i(\theta)$. Suppose every realized F_q is μ-strongly convex, F_p is β-smooth, and $\theta_w = \arg\min_\theta F_w(\theta)$. At θ_p, write $G_* = [\nabla f_1(\theta_p), \dots, \nabla f_k(\theta_p)]$. Since $G_* p = 0$, $\nabla F_q(\theta_p) = G_*(q-p)$. Strong monotonicity and Cauchy–Schwarz give

$$\|\theta_q - \theta_p\|_2 \le \mu^{-1}\|G_*(q-p)\|_2.$$

Squaring and averaging yields

$$\mathbb{E}\|\theta_Q - \theta_p\|_2^2 \le \frac{1}{\mu^2}\,\mathrm{tr}(G_* B G_*^\top) \le \frac{h}{\mu^2}\,\mathrm{tr}(G_* C_{p|A} G_*^\top).$$

Smoothness of F_p gives Eq. (23).

For a fixed draw reused along deterministic full-batch gradient descent, assume $\mu I \preceq \nabla^2 F_q(\theta) \preceq \beta I$ for all relevant (q,θ), use common initialization and $0 < \eta < 2/\beta$, and put $\rho = \max\{|1-\eta\mu|, |1-\eta\beta|\} < 1$. Along the deterministic full-data path, let $G_t = [\nabla f_1(\theta_t^p), \dots, \nabla f_k(\theta_t^p)]$. Contractivity of the gradient map gives

$$\|\theta_{t+1}^q - \theta_{t+1}^p\|_2 \le \rho\|\theta_t^q - \theta_t^p\|_2 + \eta\|G_t(q-p)\|_2.$$

Unrolling and applying Minkowski in $L_2(Q)$ yields

$$\left(\mathbb{E}\|\theta_T^Q - \theta_T^p\|_2^2\right)^{1/2} \le \eta\sqrt{h}\sum_{t<T}\rho^{T-1-t}\sqrt{\mathrm{tr}(G_t C_{p|A} G_t^\top)}.$$

This argument relies on uniform strong convexity and deterministic full-batch updates. A PL condition alone does not control which minimizer a perturbed objective selects, and generic nonconvex extensions require additional basin or error-bound assumptions.

For the quadratic family in Eq. (21), $F_p^V(\theta_Q) - F_p^V(\theta_p) = \frac{\mu}{2}\|V(Q-p)\|_2^2$. Maximization over V gives the largest generalized eigenvalue of B relative to $C_{p|A}$ on $\mathrm{range}(C_{p|A})$. Therefore the upper and sharpness parts of Theorem 1 transfer verbatim to Eq. (22). If the denominator is zero, exact protection forces the numerator to be zero as well.

F An exact conic lift

For $A = 0$, let $\mathcal{D}_{k,m} = \{q \in \Delta_k : \|q\|_0 \le m\}$ and define the sparse completely-positive cone

$$\mathcal{CP}_k^{(m)} = \mathrm{cone}\{zz^\top : z \ge 0,\ \|z\|_0 \le m\}.$$

The fixed-W value has the exact lift

$$\mathrm{OPT}(p, 0, m; W) + p^\top W p = \min\{\langle W, X\rangle : X\mathbf{1} = p,\ X \in \mathcal{CP}_k^{(m)}\}. \tag{38}$$

Indeed, a law maps to $X = \mathbb{E}[QQ^\top]$. Conversely, write $X = \sum_t z^t (z^t)^\top$, put $s_t = \mathbf{1}^\top z^t$, $q^t = z^t/s_t$ for nonzero factors, and $\lambda_t = s_t^2$. Then $X\mathbf{1} = p$ gives

$$\sum_t \lambda_t q^t = p, \qquad \sum_t \lambda_t = \mathbf{1}^\top X \mathbf{1} = \mathbf{1}^\top p = 1,$$

so the factors normalize into an admissible law supported on $\mathcal{D}_{k,m}$. Its second moment is X, proving Eq. (38).

This is a nonnegative-factor-width reformulation, not a new cone or a tractability claim. The support restriction remains encoded in the generators of $\mathcal{CP}_k^{(m)}$; the hardness results in Section 6 and Appendix D show why the lift alone does not yield a polynomial-time method for general (m, W).

G Numerical illustration on a larger instance

Figure 4 and Table 3 repeat the path experiment of Section 8 on an instance with $k = 32$ coordinates and $d = 3$ protected totals ($K = 29$): p is a Dirichlet draw with non-uniform concentration (smallest coordinate $\approx 10^{-3}$) and the three rows of A have independent entries in $\{0, 1, 2\}$, so the protected fiber is a 28-dimensional polytope. For each method 1000 paths are run from $m = k$ down to $m = d + 1$ with 150 multiplicative-weights rounds per step (about 17 CPU minutes in total), and the largest generalized eigenvalue of the estimated covariance relative to $C_{p|A}$ is reported at eight budgets together with the trace ratio $\mathrm{tr}\,\mathrm{Cov}(Q)/\,\mathrm{tr}\,C_{p|A}$.

Two features of the estimates should be kept in mind. The largest eigenvalue of a 28-dimensional sample covariance from 1000 draws is biased upward, so the reported $\lambda_{\max}$ values overstate the true factor and the bootstrap percentile interval can lie above the point estimate; the trace ratio is unbiased. Both estimates are conservative for the statement being checked. With that caveat, the law of Theorem 1 has point estimate below the curve $h_{29,M}$ at every reported budget (for instance 0.102 against 0.112 at $m = 10$, 0.0237 against 0.0290 at $m = 19$, 0.0008 against 0.0013 at $m = 31$) and its trace ratio is below the curve by a factor of 1.5–3. The naive single-radial-response law is below the curve only at the two smallest budgets, where the one-step lemma is nearly isotropic, and exceeds it at every budget $m \geq 7$, by a factor growing from 1.6 at $m = 7$ to about 7 at $m = 31$; its trace ratio also exceeds the curve from $m = 10$ on. The picture is the same as on the small instance of Figure 2: the minimax step is what keeps the construction on the curve, and a single radial response does not.

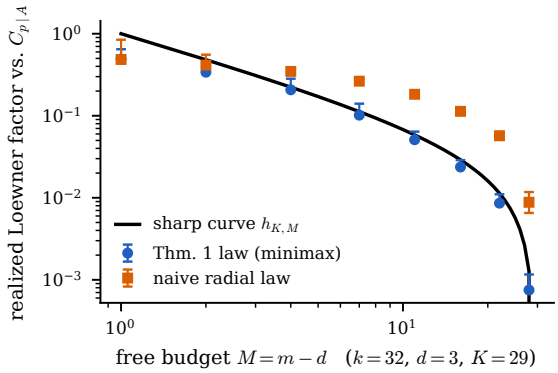


Figure 4: **Realized Loewner factor on the larger instance** ($k = 32$, $d = 3$, $K = 29$; 1000 paths per method; log–log axes). Markers are the largest generalized eigenvalue of the estimated covariance relative to $C_{p|A}$; bars are the hull of the point estimate and its 90% bootstrap percentile interval.

Table 3: Realized factors on the larger instance against the curve $h_{29,M}$: largest generalized eigenvalue relative to $C_{p|A}$ with 90% bootstrap percentile interval, and trace ratio.

m	M	$h_{K,M}$	Thm. 1 law: $\lambda_{\max}$ [90% boot.]	naive law: $\lambda_{\max}$ [90% boot.]	Thm. 1 trace	naive trace
4	1	1.0000	0.4889 [0.4805, 0.6468]	0.4852 [0.4937, 0.8451]	0.3146	0.3207
5	2	0.4821	0.3395 [0.3502, 0.4229]	0.4146 [0.4118, 0.5548]	0.2342	0.2699
7	4	0.2232	0.2071 [0.2065, 0.2829]	0.3473 [0.3385, 0.3747]	0.1391	0.2042
10	7	0.1122	0.1017 [0.1109, 0.1403]	0.2639 [0.2545, 0.2799]	0.0766	0.1396
14	11	0.0584	0.0510 [0.0555, 0.0641]	0.1827 [0.1756, 0.1943]	0.0389	0.0868
19	16	0.0290	0.0237 [0.0256, 0.0288]	0.1132 [0.1082, 0.1212]	0.0175	0.0474
25	22	0.0114	0.0086 [0.0088, 0.0110]	0.0570 [0.0529, 0.0635]	0.0060	0.0203
31	28	0.0013	0.0008 [0.0008, 0.0012]	0.0088 [0.0065, 0.0117]	0.0004	0.0026