How Much Must a Ridge Learner Remember to Delete Data?

Anonymous authors
Paper under double-blind review

Abstract

An offline ridge learner retains a finite-bit summary; a decoder receives deleted records and approximates the retained-data optimum without accessing the retained records. How do regularization, accuracy, and deletion size determine the necessary memory? For bounded grid inputs, $n = \Theta(d)$, and a fixed deletion fraction, the worst-case memory is $d \min\{d, (\lambda\varepsilon^{1/3})^{-1}\}$ bits within one factor $\log d$, for sufficiently small additive or relative error ε. When $\lambda = 1/4$ and the deletion budget ρ vanishes, the additive law becomes $d \min\{d, \rho/\sqrt{\varepsilon}, (\rho/\varepsilon)^{1/3}\}$ within a precision logarithm. For bounded noiseless teachers, a joint rank law is matching when ρ/λ is bounded. A determinantal response envelope extends its upper bound to every budget with an $\exp(O(\rho/\lambda))$ overhead. A routing construction proves that an $\exp(\Omega(\rho/\lambda))$ overhead is necessary along explicit families, despite a common teacher, original optimizer, and legal requests. Thus the law persists within polylogarithmic factors for $\rho = O(\lambda \log\log d)$, while extensions beyond that scale can fail. The lower bounds allow pointwise randomized success; deterministic summaries attain the upper bounds. Residual importance sampling gives a polynomial bit encoder for a general-label DPP bound. Controlled and frozen-feature experiments assess accuracy and storage. These guarantees concern optimization utility, not distributional certification of unlearning.

1 Introduction

A fitted model solves the original learning problem; a deletion mechanism must also solve the problems obtained by removing records. Once the training data are unavailable, the fitted parameters alone may no longer contain the information needed for these updates. Ridge regression with fixed features isolates this information requirement: both the original fit and each retraining target have explicit, unique solutions.

Exact deletion subtracts the deleted records' Gram matrix and feature–label vector from the stored statistics, using quadratically many numerical entries in the feature dimension d. This approach also supports federated ridge heads on frozen foundation models (Quan et al., 2026). Approximate deletion raises a different question: how much of the Gram matrix must be stored before the request is known? Strong regularization suppresses weak directions, whereas tighter accuracy requirements make them relevant again. We quantify the resulting memory tradeoff.

We study central summaries. The encoder processes the entire dataset offline; the decoder receives the summary and the records to delete, with no retained-data access or user-held auxiliary ticket. Success means small additive excess in the retained data's normalized ridge objective. Randomized success is required for each fixed dataset and request, rather than simultaneously for all requests.

Our main result gives lower and upper bit bounds differing by $O(\log d)$, with joint dependence

$$d \min\left\{d, \frac{1}{\lambda\varepsilon^{1/3}}\right\}. \tag{1}$$

The lower bound holds even when every original optimum is exactly zero. Its deletion requests use public anchor records common to all encoded datasets, so the request itself cannot communicate the hidden matrix. The construction remains hard after rounding to a common input grid whose precision is independent of λ and ε.

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.

The upper bound uses a low-rank spectral representation with an isotropic correction. Its error is confined to the omitted eigenspace, where every retained subset has little feature energy. Ridge stationarity then gives a cubic *relative* objective bound, adapting automatically to the retained task's optimum loss. Our contribution is the legal-deletion storage law: we combine established sketching and communication tools with a tail-supported error bound and a common-request lower-bound construction.

Smaller deletion budgets produce two regimes: a square-root accuracy dependence for concentrated requests and a cube-root dependence for spread requests. At fixed regularization, both are matched by lower bounds with a public noiseless teacher and identical original teacher-seeded moments. When regularization also vanishes, the joint law is matching for bounded ρ/λ. Determinantal reweighting extends the upper bound to every budget with an exponential overhead in this ratio. A forest-routing construction proves that exponential dependence is necessary on explicit families, while preserving a common public teacher, original optimizer, and legal requests.

For individual requests, legality supplies additional information: the unknown retained Gram matrix must remain positive semidefinite. A Schur complement converts this constraint into a covariance interval, yielding a stable midpoint update and sharper error bounds at unchanged numerical storage. Experiments isolate task fit by fixing the feature covariance, label multiset, and requests. A separate noiseless construction shows why reproducing ridge shrinkage still requires information beyond a public bounded teacher.

2 Setting and main result

A dataset S is a multiset of n records $(x, y) \in \mathbb{R}^d \times \mathbb{R}$, with $\|x\| \le 1$ and $|y| \le 1$. The feature representation and regularization parameter $0 < \lambda \le 1$ are public and fixed independently of S. For any nonempty multiset T, define

$$F_T(w) = \frac{1}{|T|} \sum_{(x,y)\in T} (x^\top w - y)^2 + \lambda \|w\|^2, \qquad w_T = \arg\min_{w\in\mathbb{R}^d} F_T(w). \tag{2}$$

Thus deleting records changes both the data average and the unnormalized regularization term $|T|\lambda I$ in the normal equations. Throughout, $G_T = \sum_T xx^\top$, $b_T = \sum_T xy$, and $Q_T = G_T/|T| + \lambda I$.

An encoder receives all of S and retains a state $M \in \{0,1\}^B$, including any length metadata. A decoder receives M and a legal submultiset $U \le S$, including its features, labels, and multiplicities, and returns $\widehat{w}_U$. It cannot access $S - U$. For every fixed admissible pair (S, U), it must satisfy

$$\Pr\big[F_{S-U}(\widehat{w}_U) - F_{S-U}(w_{S-U}) \le \varepsilon\big] \ge \frac{2}{3}, \qquad |U| \le n/3. \tag{3}$$

Arbitrary data-independent public randomness is allowed. Every persistent data-dependent bit is charged to B, including separately stored parameters, training trajectories, per-record information, and retained private seeds. Fresh decoder randomness is allowed. Offline computation, temporary decoding workspace, and terminal output length are unrestricted. Each request is evaluated from the original state. State-changing sequences and adaptive multi-query guarantees are outside this model.

For the bit theorem, fix the common dyadic grid with

$$p_{\mathrm{in}}(d) = \lceil \log_2(2^{22}d) \rceil, \qquad x_i \in 2^{-p_{\mathrm{in}}(d)}\mathbb{Z}, \qquad y \in 2^{-p_{\mathrm{in}}(d)}\mathbb{Z} \cap [-1, 1]. \tag{4}$$

The constant provides the slack required by the rounding argument. The fixed-fraction lower bound needs only labels in $\{-1, 0, 1\}$, while the upper bound also holds for any prescribed common $O(\log d)$-bit grid. The public parameters λ, ε are not encoded in the summary. Let $B^*(d, \lambda, \varepsilon)$ be the least worst-case state length for datasets with $d \le n \le 3d$ and the grid in (4). Define B^*_{rel} by replacing the success event in (3) with

$$F_{S-U}(\widehat{w}_U) \le (1 + \varepsilon) F_{S-U}(w_{S-U}). \tag{5}$$

This multiplicative formulation remains defined when the optimum is zero.

Theorem 1 (Central memory for approximate ridge deletion)**.** *There are universal constants $c, C, \varepsilon_0 > 0$ such that, for every sufficiently large even d, $0 < \lambda \le 1$, and $0 < \varepsilon \le \varepsilon_0$,*

$$cd \min\{d, (\lambda\varepsilon^{1/3})^{-1}\} \le B^*(d, \lambda, \varepsilon) \le B^*_{\mathrm{rel}}(d, \lambda, \varepsilon) \le Cd \min\{d, (\lambda\varepsilon^{1/3})^{-1}\} \log d. \tag{6}$$

The lower bound holds on families with $w_S = 0$ and a common collection of legal deletion requests. The relative upper bound is deterministic and holds for every legal request to each encoded dataset. Since $F_T(w_T) \le F_T(0) \le 1$, relative success implies additive success.

Embedding a neighboring even dimension covers odd dimensions. The theorem identifies all three parameter dependencies within a precision logarithm, rather than an exact Θ bit law. The universal threshold ε_0 is not numerically optimized. Equation (6) treats a fixed deletion fraction; Section 5 treats vanishing budgets. Exact Gram storage covers rank saturation at d, including arbitrarily small λ or ε.

The precision logarithm is necessary at sufficiently high accuracy: Proposition 15 matches $\Theta(d^2 \log d)$ bits on an explicit subregime of the same grid.

3 A SPECTRAL SUMMARY WITH CUBIC ERROR

Let $\sigma_1 \ge \cdots \ge \sigma_d$ be the eigenvalues of G_S and V_ℓ its first $\ell < d$ orthonormal eigenvectors. Put

$$\delta = \sigma_{\ell+1}, \qquad K = \mathrm{diag}_{i \le \ell} \sqrt{\sigma_i - \delta}\, V_\ell^\top, \qquad \widehat{G} = K^\top K + \delta I. \tag{7}$$

The encoder retains K, δ, b_S, n. Given U, let $m = n - |U|$ and return

$$\widehat{w}_U = [K^\top K - G_U + (m\lambda + \delta)I]^{-1}(b_S - b_U). \tag{8}$$

This representation keeps the leading eigenvalues unchanged and raises every remaining eigenvalue to δ. With $\Pi = I - V_\ell V_\ell^\top$, its error is

$$E_0 = \widehat{G} - G_S = \Pi(\delta I - G_S)\Pi, \qquad 0 \preceq E_0 \preceq \delta\Pi. \tag{9}$$

Subtracting a legal G_U leaves $G_{S-U} + E_0$. Consequently the matrix in (8) is positive definite for every ℓ and every legal request. No assumption $\delta < m\lambda$ is needed.

Proposition 2 (Tail-supported perturbation)**.** *For bounded records as in* (2)*, the decoder* (8) *satisfies, simultaneously for all $U \subseteq S$ with $m > 0$,*

$$F_{S-U}(\widehat{w}_U) - F_{S-U}(w_{S-U}) \le F_{S-U}(w_{S-U}) \left(\frac{\delta}{m\lambda}\right)^3 \le \left(\frac{\delta}{m\lambda}\right)^3. \tag{10}$$

Proof. Write $T = S - U$, $Q = Q_T$, $w = w_T$, and $P = E_0/m$. Since $G_T \preceq G_S$, the retained feature matrix X obeys $\|X\Pi\|_{\mathrm{op}} \le \sqrt{\delta}$. For the residual $e = y - Xw$, stationarity gives $\lambda w = X^\top e/m$, hence

$$\|\Pi w\|^2 \le \frac{\delta}{m\lambda^2} \frac{\|e\|^2}{m} \le \frac{\delta}{m\lambda^2} F_T(w_T). \tag{11}$$

The exact quadratic excess is $\left\|Q^{1/2}(\widehat{w} - w)\right\|^2$, where $\widehat{w} - w = -(Q+P)^{-1}Pw$. Because $P \succeq 0$,

$$(Q+P)^{-1}Q(Q+P)^{-1} \preceq Q^{-1} \preceq \lambda^{-1} I. \tag{12}$$

Combining (9)–(12) bounds the excess by $\lambda^{-1}(\delta/m)^2 \|\Pi w\|^2$, proving (10). □

The proof allows deletion to rotate the retained covariance relative to V_ℓ: it uses positive-semidefinite ordering below G_S, not simultaneous diagonalization. The trace bound $\mathrm{tr}\, G_S \le n$ gives $\delta \le n/(\ell+1)$. For $m \ge \beta_0 n$, rank $O((\beta_0 \lambda \varepsilon^{1/3})^{-1})$ suffices.

The spectral shape sharpens this bound: replace the numerator δ^3 by $\kappa_\ell = \max_{i>\ell} \sigma_i(\delta - \sigma_i)^2$ (Lemma 9). If $G_S = \delta P$ for any orthogonal projector P, taking $\ell = 0$ gives $\kappa_0 = 0$: storing b_S, δ, n answers every legal deletion exactly, without storing P, even at high rank. Corollary 10 gives an $O(d \log d)$-bit implementation. The worst-case separation therefore requires geometry that changes legal deletion responses, not merely an unknown high-rank subspace.

For a finite-bit summary, round the factor and raise the floor to preserve the one-sided covariance error. In the branch $\ell < d$, the relevant λ, ε are bounded below by inverse polynomials in d, so $O(\log d)$ bits per entry suffice. Otherwise, store G_S, b_S, n exactly. Exact storage of b_S preserves relative utility even when the optimum is zero. Appendix A gives the rounding and complete bit count; Corollary 11 proves the relative guarantee.

Dependence on task fit. With $L = F_S(w_S)$ and retained fraction β, one has $F_T(w_T) \le L/\beta$. Thus the relative bound also gives additive rank $O((L/\varepsilon)^{1/3}/\lambda)$ at fixed β. Label scaling gives matching loss-dependent bounds (Appendix S). A norm-one noiseless teacher gives $L \le \lambda$, explaining the improved regularization dependence below.

4 Why legal deletion requires matrix information

The lower bound encodes an operator-norm-bounded sign matrix $A \in \{\pm 1/\sqrt{p}\}^{p \times r}$ in zero-label records. Public pairs of oppositely labeled anchors cancel in the original feature–label vector, making $w_S = 0$. A sign query $z \in \{\pm 1\}^r$ deletes one sign of each anchor. This changes the right-hand side of the retained normal equations while preserving a hidden dependence on A. Every query is a legal deletion from every dataset in the family.

For $\lambda \ge 1/(2d)$, the private zero-label records contain columns of $(I, tA)^\top$, with A having $p = d - r$ rows and small public t. Oppositely labeled public anchors cancel the original right-hand side. Deleting one anchor sign per column makes the retained bottom coordinates proportional to $A(I + \eta_A A^\top A)^{-1} z$, with $\eta_A = O(t^2)$. This is uniformly close to a scaled matrix–vector response. Its objective signal is at least $c/(1 + r\lambda)^3$; choosing r proportional to $\min\{d, (\lambda \varepsilon^{1/3})^{-1}\}$ makes the signal larger than the allowed error. For smaller λ, a separate square construction supplies $\Omega(d^2)$ bits.

Approximate responses to random z yield an estimator of A. The sign source has entropy at least $pr - 1$ bits, and predicting its entries with a bounded average error fraction requires $\Omega(pr)$ bits. Constant repetition converts the pointwise $2/3$ success guarantee into the required mean-square estimate, without assuming simultaneous success on all queries. Appendices B–D give the construction, amplification rule, and uniform fixed-grid perturbation bounds.

5 Deletion budgets and task-dependent summaries

A small request either concentrates on a few informative directions or spreads over many. To quantify this distinction, fix $\lambda = 1/4$ and let $0 < \rho \le 1/8$ vary. Use the same input grid, with public sample sizes $4d/\rho \le n \le 8d/\rho$, and require additive excess at most ε for every fixed legal request of size at most ρn. Denote the least worst-case state length by B^*_ρ.

Theorem 3 (Memory as the deletion fraction vanishes)**.** *For sufficiently large d and $0 < \varepsilon \le c_0 \rho^2$, define*

$$R_* = \min\left\{ d, \frac{\rho}{\sqrt{\varepsilon}}, \left(\frac{\rho}{\varepsilon}\right)^{1/3} \right\}. \tag{13}$$

Then $cdR_* \le B^*_\rho(d, \varepsilon) \le CdR_* \log(2 + d/\rho)$. The upper bound is deterministic for all bounded labels. The lower bound holds even for a public unit noiseless teacher, label variance exactly $1/4$ before and after deletion, a common nonzero original optimizer, and common legal requests, with pointwise randomized success.

For the upper bound, center the spectral floor at $\theta = w_S$. For $\ell < d$, let $\nu = \sigma_{\ell+1}(G_S/n)$. The optimizer movement obeys both $\|w_T - \theta\| = O(\rho)$ and $\|\Pi(w_T - \theta)\| = O(\sqrt{\rho\nu})$. The second bound uses legality and the energy of the deleted residuals. Thus the centered excess is $O(\min\{\rho^2\nu^2, \rho\nu^3\})$, and $\nu \le 1/(\ell + 1)$ gives (13); full rank is exact. Upward covariance rounding and ordinary predictor rounding give the finite-bit implementation, including empty requests (Appendix F).

The matching construction gives each common anchor packet mass proportional to $\min\{\rho, 1/r\}$. A request deletes one packet, or about ρr packets, respectively. Their responses carry hidden-matrix signal at scales ρ^2/r^2 and ρ/r^3. Uniform sparse sign queries retain enough information

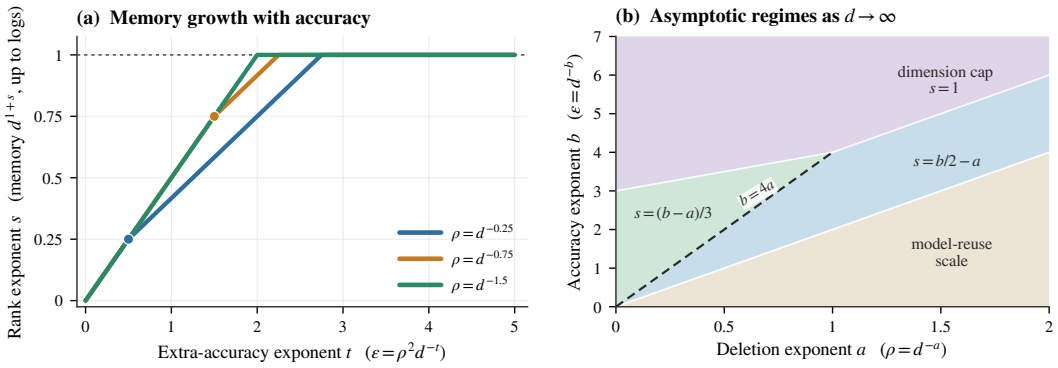

Figure 1: Asymptotic exponents in Theorem 3. Set $\rho = d^{-a}$ and $\varepsilon = d^{-b}$ for fixed $a > 0$, $b > 2a$, and let $d \to \infty$. Memory is d^{1+s} up to logarithms, with $s = \min\{1, b/2 - a, (b-a)/3\}$. Dots and dashes mark $b = 4a$ before saturation. Model reuse suffices when $b < 2a$. The diagrams describe asymptotic exponents; finite-dimensional thresholds also depend on the proof constants.

to force $\Omega(dr)$ bits. The two ranks cross at $\varepsilon = \rho^4$ before saturation. All teacher-seeded moments $(G_S/n)^j v$ are identical across the hard family: the deletion requests activate information absent from those original-task responses.

Varying regularization and deletion together. For a unit noiseless teacher, task-dependent summaries improve the regularization dependence. The next result uses $p_J = \lceil \log_2(2^{30} d/\lambda) \rceil$ fractional input bits and public sample sizes $4d/\rho \le n \le 20d/\rho$.

Theorem 4 (A joint law with a deletion-budget correction)**.** *For sufficiently large d, $0 < \lambda \le 1/4$, $0 < \rho \le 1/8$, and $0 < \varepsilon \le c_0 \min\{\rho^2/\lambda, \rho\lambda\}$, put*

$$R_J = \min\left\{ d, \frac{\rho}{\lambda^{3/2}\sqrt{\varepsilon}}, \left(\frac{\rho}{\lambda^2 \varepsilon}\right)^{1/3} \right\}. \tag{14}$$

On the stated grid and sample-size interval, bounded data with a noiseless teacher of norm at most one have minimax additive summary length

$$cdR_J \le B^*_{\text{joint}} \le Cd \min\{d, e^{8\rho/(21\lambda)} R_J\} \log(2 + d/(\lambda\rho)). \tag{15}$$

For every fixed M, the bounds match within the displayed logarithm when $\rho \le M\lambda$, with a constant depending on M. The lower bound retains a public unit teacher, a common original optimizer, common legal requests, and pointwise randomized success. The upper bound is deterministic.

For the upper bound, represent the original ridge model as the mean of the minimum-norm fit Z to a determinantal row subset. Its covariance $\mathcal{V}$ has trace at most $1/4$. Deleting an α fraction changes the sampling density to f_U; a determinant identity gives $\mathbb{E}(f_U - 1)^2 \le e^{\alpha/((1-\alpha)\lambda)} - 1$. Cauchy–Schwarz therefore bounds every movement $\Delta_U \Delta_U^\top$ by this scalar times $\mathcal{V}$. Preserve the covariance action on the leading r directions of both $\Sigma = G_S/n$ and $\mathcal{V}$. The completion $\widehat{\Sigma} = \Sigma + \Pi(\delta I - \Sigma)\Pi$, with $\delta = \sigma_{r+1}(\Sigma)$, has a one-sided error and an $O(dr)$-entry representation. Its excess is $O((e^{\rho/((1-\rho)\lambda)} - 1)/(\lambda r^3))$. Finite rounding and the centered-floor square branch prove (15) (Appendix H); Appendix G proves the information bound. The upper bound uses the records to compute Z, without knowing the teacher.

The cubic DPP branch also has a randomized encoder using $O(r \log(d/\delta))$ draws, with probability $1 - \delta$ of accuracy on all requests. For rational inputs, Theorem 34 gives polynomial bit complexity; Proposition 33 controls sampling and arithmetic errors. These guarantees reproduce ridge shrinkage: returning the teacher has excess at most λ. Joint lower-bound label variance is $\Theta(\lambda)$, whereas Theorem 3 keeps it constant.

Exponential dependence is necessary. Theorem 4 preserves dR_J within fixed polylogarithmic factors when $\rho = O(\lambda \log\log d)$. The next result proves an obstruction beyond this scale.

Theorem 5 (An exponential lower-bound correction)**.** *For any sufficiently small fixed $\kappa > 0$, there are sequences $d \to \infty$, $\lambda, \rho \to 0$, $\lambda d \to \infty$, and $\varepsilon = \kappa\rho\lambda$ in the accuracy range of Theorem 4, such that*

$$B^*_{\text{joint}} \geq cd^2 \geq c' e^{c_1 \rho/\lambda} dR_J. \tag{16}$$

Here $\rho/\lambda = o(\log d)$ and $\rho/(\lambda \log\log d)$ may diverge arbitrarily slowly. For every fixed $a \geq 0$, $dR_J[\log(2 + d/(\lambda\rho))]^a = o(d^2)$. The family has a common public unit teacher, a common public nonzero original optimizer, common legal requests, and label variance $\Theta(\lambda)$ before and after deletion. Inputs lie exactly on the prescribed grid, and the bound allows pointwise randomized success.

Appendix P constructs a forest whose binary routes end in stars. Deleting sibling edges sends a root residual to a selected leaf; hidden zero-label attachments reveal a sign-matrix column. Blockwise column pairing cancels untouched trees and keeps the original model common. Schur elimination controls relative readout error, and binary entropy forces quadratic bits. The lower bound establishes the exponential obstruction on an explicit parameter family. Matching the exponential constants and determining the full joint minimax formula remain open.

Encoding and limits of surrogate objectives. For arbitrary bounded labels, let $r_S = \|y - X\theta\|^2/n$. The DPP bound extends to rank $O(1 + (\Gamma_\rho r_S/(\lambda^2\varepsilon))^{1/3})$, capped at d, where $\Gamma_\rho = e^{\rho/((1-\rho)\lambda)} - 1$. Rare ill-conditioned subsets prevent uniform covariance concentration under ordinary sampling. Residual importance sampling bounds each contribution: $q(J) = \lambda p(J)\|z_J - \theta\|^2/r_S$ on p-positive subsets, with extra mass contributing zero matrices. Contributions are bounded by $(r_S/\lambda)I$, giving polynomial bit encoding at this rank for rational inputs (Theorem 40). Shorting the spectral-floor residual reduces a rank-m covariance/response construction to at most $2dm+d+2$ entries while preserving the cubic bound (Appendix K.2). Response-weighted covariance optimization can be $\Omega(1/\lambda)$-suboptimal at $d = 4$. Even exact reference-utility minimizers can be $\Omega(r)$-suboptimal on actual requests at $\lambda = 1$ with vanishing curvature variation (Theorem 51). Within the reference problem, an oblique correction gives a sharp condition-dependent approximation without commutation; an inflated floor bounds its factor by $9/8$. The approximation applies to the relaxed correction class. For listed requests, we also obtain a polynomial-bit actual-loss approximation under stable curvature, using more correction columns and rational weights (Theorem 71). Signed factors recover exact covariance directions for legal decoding at unchanged storage (Appendices K.3.1–K.3.5).

Simple alternatives. Exact statistics use $O(d^2 \log d)$ bits in Theorem 1's regime. With the joint input grid and sample sizes, this becomes $O(d^2 \log(2 + d/(\lambda\rho)))$. A globally shifted truncation yields rank $O((\lambda\sqrt{\varepsilon})^{-1})$ through a full-space perturbation bound. At loose tolerances, model reuse suffices; its stored predictor counts toward memory. Appendix E gives these guarantees and a matching real-input worst-case error for reuse.

6 Assessing an answer without retained data

Request-specific information sharpens the global cubic bound. A centered summary stores $K, \delta, \theta = w_S, n$, replacing b_S by the original model at unchanged numerical storage. With $\widehat{Q} = (\widehat{G} - G_U)/m + \lambda I$, compute

$$g = \frac{b_U - (G_U + |U|\lambda I)\theta}{m}, \qquad \widehat{w}_{\text{c}} = \theta - \widehat{Q}^{-1} g. \tag{17}$$

Original stationarity implies $g = Q_T(\theta - w_T)$, so the update minimizes $F_T(w) + (w-\theta)^\top P(w-\theta)$, where $P = E_0/m \succeq 0$. Writing $\mathcal{E}_T(w) = F_T(w) - F_T(w_T)$ and $\tau = \delta/m$, we obtain

$$\mathcal{E}_T(\widehat{w}_{\text{c}}) \leq \left(\frac{\tau}{\lambda + \tau}\right)^2 \mathcal{E}_T(\theta), \tag{18}$$

$$\mathcal{E}_T(\widehat{w}_{\text{c}}) \leq C_U := \min\left\{ \frac{\tau^2}{\lambda} \|\Pi(\widehat{w}_{\text{c}} - \theta)\|^2, \left(\frac{\tau}{\lambda + \tau}\right)^2 \frac{\|g\|^2}{\lambda} \right\}. \tag{19}$$

Here Π can be computed as the orthogonal complement of the factor's row space; zero-weight tied eigenvectors may be included in this complement. The summary and request determine every quantity in C_U, without the retained optimum, retained loss, or original full Gram matrix. An empty request returns the original model with bound zero.

Appendix T proves these quadratic contraction and residual bounds and accounts for reference quantization and solve residuals. The displayed formulas use exact arithmetic and an exact original optimizer. The experiments evaluate their floating-point counterparts without rigorous roundoff enclosures.

Extracting information from legality. In the stored leading/tail basis, write $G_S = \mathrm{diag}(A, C)$ and $G_U = (D_{ij})_{i,j=1}^2$, with known A and unknown $0 \preceq C \preceq \delta I$. The constraint $G_S - G_U \succeq 0$ yields

$$L_U = D_{22} + D_{21}(A - D_{11})^\dagger D_{12} \preceq C \preceq \delta I, \qquad W_U = \delta I - L_U \succeq 0. \tag{20}$$

The generalized Schur criterion handles singular leading blocks using the range condition implied by legality. If Z spans the omitted subspace, the legal midpoint uses $H_U = \widehat{Q} - ZW_U Z^\top/(2m) \succeq \lambda I$ and returns $\theta - H_U^{-1} g$. Appendix U gives its additive certificate and optimal interval contraction. The optimality domain is the outer interval of quadratics, a relaxation of those realizable by compatible bounded records. The interval also sharpens flooring's certificate; fixed midpoint with PSD repair has its own residual bound. All use the same numerical storage. Their exact-statistics guarantees and precision corrections are separate from the main bit theorem's rounded-floor construction. For DPP completions, legality similarly sharpens the density multiplier using only stored covariance columns and deleted features (Appendix L).

7 Experiments

Decoding uses only the numerical summary and deleted records; a separate evaluator accesses retained data. We compare exact statistics, model reuse, truncation, flooring, and centered variants. Stable baselines replace the omitted spectrum by zero or its fixed midpoint and repair the retained Gram to be PSD. These baselines and legal midpoint use flooring's payload and no coefficients selected from test outcomes. Appendix W gives definitions and an $11{,}340$-evaluation diagnostic varying geometry and task fit.

Changing the task at fixed geometry. A $55{,}680$-evaluation control fixes 128 paired-antipodal features in 64 dimensions. Balanced flips change a norm-one teacher's labels while preserving covariance, the label multiset, and common deletion masks (Appendix Y). At $\lambda = 0.01$ and 25% deletion, flip rates $0, 9.375\%, 25\%, 50\%$ give median training objectives $0.00976, 0.16178, 0.35284, 0.45966$. The minimum tested ranks meeting additive error 10^{-3} on the common request bank are $0, 8, 16, 16$ for flooring, $0, 8, 16, 8$ for fixed midpoint, and $0, 8, 8, 8$ for legal midpoint. Training fit alone therefore fails to order deletion difficulty. Legal midpoint's certificate instead requires ranks $0, 8, 16, 16$. Relative measurements and normalized posterior bounds use the evaluator's retained optimum. They are oracle diagnostics rather than decoder-available relative certificates.

Storage and error on frozen image features. Public ImageNet-pretrained ResNet-18 features support CIFAR-10 cat/dog and airplane/truck tasks (He et al., 2016; Krizhevsky, 2009). Features are individually normalized in 512 dimensions. Three seeds draw balanced 1000-record training and 500-record test subsets. Three regularizers, ranks $8, 32, 128$, deletion fractions $5\%, 25\%$, and random or largest signed-residual requests give 216 configurations and 1944 method evaluations. Hyperparameters are fixed independently of test accuracy. Appendix X gives the full protocol and comparisons.

At rank 32, summaries use 132 KiB versus 1030 KiB for exact statistics. Table 1 compares actual errors and request-based bounds; the corresponding centered-floor uniform bounds, including the $4/27$ spectral-shape improvement, are 10.6 and 6.77. At the predeclared tolerance 0.01, flooring's refined posterior accepts $80/108$ random and $61/108$ signed-residual configurations, versus 21 of each for its uniform bound. Fixed midpoint accepts $78/61$, and legal midpoint $82/64$. Actual legal-midpoint excess meets the tolerance in $102/92$ cases. Every configuration accepted by a bound also meets the tolerance in the floating-point evaluation.

Table 1: Medians over three seeds at $\lambda = 0.001$, rank 32, and 25% deletion. Both posteriors use the same request-conditioned covariance interval. Bounds are evaluated in floating point without rigorous roundoff enclosures.

		Centered floor		Legal midpoint	
Task	Request	Excess	Posterior	Excess	Posterior
Cat/dog	Random	0.00318	0.00949	0.00157	0.00559
Cat/dog	Signed residual	0.00796	0.0232	0.00393	0.0138
Airplane/truck	Random	0.00113	0.00276	0.000534	0.00156
Airplane/truck	Signed residual	0.00268	0.00760	0.00125	0.00423

Legal midpoint improves actual error and the posterior over centered flooring in all 216 configurations. Against fixed midpoint, it improves actual error in 123: $31/108$ at 5% deletion and $92/108$ at 25%. Its posterior is smaller in all 216: tighter certificates and smaller actual errors are distinct outcomes. No Schur fallback occurs on this grid; Appendix W describes numerical failure handling. At the table's λ and tolerance 0.01, the worst-case rank scale exceeds d; the observed savings come from favorable instance and request structure, outside the minimax compression branch.

Computable envelopes at equal storage. Operator and DPP comparisons appear in Appendices Z–Z.2. An 8352-evaluation signed-completion grid finds that Nyström floor beats importance union in all 1392 pairs but loses 1283 to legal midpoint. All six signed methods lose every class-label comparison to legal midpoint. A further 5568-evaluation reference-utility grid retains this gap: ordinary utility floor loses $1318/1392$ strong-baseline pairs, and inflation loses $1384/1392$ pairs to its own ordinary floor. Near-optimal reference objectives do not translate into deletion-accuracy gains on these grids (Appendices Z.3–Z.4). Both grids report zero numerical failures and zero observed posterior violations; rigorous roundoff certification was not performed.

8 Related work

Memory for deletion. Cherapanamjeri et al. (2025) study space for realizability testing and, in full-version Appendix G, exact 0–1 ERM, separating learning from deletion memory and central from user-held ticket storage. Our model likewise distinguishes learning memory from deletion memory, but quantifies approximate ridge deletion. Requests carry only deleted records, and all persistent data-dependent state is charged centrally. The general eluder-dimension question remains open.

Utility and certified unlearning. Sekhari et al. (2021) combine distributional unlearning, generalization, and deletion capacity using stored second-order information. Qiao et al. (2025) use training trajectories for certified updates; Allouah et al. (2025) analyze utility and computation for in- and out-of-distribution deletion. We instead measure retained ridge excess with no retained-record access. Distributional indistinguishability is a separate requirement; results allowing retained-data access use a different memory model.

Izzo et al. (2021) study approximate ridge deletion, including single-batch and frozen-last-layer settings; their projective residual update uses the original n-by-n hat matrix and emphasizes time. Tang et al. (2025) support continual deletion of analytic heads using a full matrix state. We characterize the compressed memory required in the single-request setting.

Xie et al. (2026) use retained subsamples for minimax statistical estimation, and Block et al. (2025) use retained gradients for minimum-complexity interpolation. Source-free certification can use surrogate data and statistical-distance assumptions (Basaran et al., 2025). Our target is finite-bit approximation of a fixed retained objective.

Ridge statistics and probing. Quan et al. (2026) use additive Gram and feature–label statistics for exact deletion of federated ridge heads. Quan & Montana (2026) show that identical heads can hide different statistics, and count exact full-head responses to algebraic probes for identification. Our lower bound counts bits for randomized approximate legal-deletion answers, using requests shared across all hidden instances.

Sketching and matrix approximation. Wang et al. (2018) analyze ridge sketches from optimization and statistical perspectives. Luo et al. (2019, Theorem 1) solve spectral approximation by a low-rank matrix plus an isotropic correction, and Shi & Phillips (2021, Theorem 5) give ridge coefficient guarantees using Robust Frequent Directions. Our representation and perturbation analysis build on these results. The one-sided floor confines error to the omitted eigenspace and preserves positive definiteness after legal subtraction. This yields a cubic bound even when the retained covariance does not commute with the original sketch, extending the calculation available for commuting, undeleted ridge systems.

Kacham & Woodruff (2022, Theorem 5.5) lower-bound non-dilating oblivious ridge sketches for a specified dual decoder that accesses the original design matrix. Our lower bound covers arbitrary deletion summaries rather than this decoder class. It builds on communication reductions for offline encodings, including Clarkson & Woodruff (2009, Theorem 2.9, full version); the additional constraints here are legal common requests, approximate responses, and finite-bit data.

Residual-weighted inverse-Hessian moments appear in ridge subsampling (Chen & Zhang, 2022, Theorem 1) and sketch-error covariance (Narahara et al., 2026, Proposition 1); the latter estimates coefficient error from compressed rows asymptotically. Our completion uses these moments to select exactly preserved matrix action, yielding a nonasymptotic legal-deletion memory guarantee (Proposition 24).

Using established generalized Lyapunov methods (Breiten & Ringh, 2018), Appendix M constructs a positive matrix iteration for legal-deletion responses. We bound the unfinished series and convert the stored factors into a utility guarantee.

We use determinantal representations of ridge (Dereziński et al., 2020; Dereziński & Mahoney, 2021) and inverse-minor moments (Mutný et al., 2020; Fanuel et al., 2021). Our residual importance law uses extended-L-ensemble sampling (Tremblay et al., 2023); residual-augmented volume changes of measure also appear in Rhee (2026). Comparing deletion densities yields our centered, bounded covariance and finite-bit deletion guarantee (Appendices H–K.1).

Appendix U applies the generalized Schur complement (Carlson et al., 1974) to request-conditioned covariance bounds. Its interval contraction is related to matrix midranges (Mostajeran et al., 2020). These completion and midpoint principles yield request-conditioned bounds by using the information revealed by legality, including sharper certificates for the unchanged floor.

9 Discussion and limitations

The storage laws separate the information needed to fit a model from that needed to answer future deletions. They apply to the specified sample sizes, grids, and feature representations; privately learned features count toward storage. Open questions include the remaining precision gaps, matching exponential constants, and the full joint law. An implicit SDP avoids request enumeration for frozen utility, with a curvature remainder for actual loss (Theorem 77). Actual zero loss is decidable in $2^{O(p)}\,\mathrm{poly}(n, d, L)$ time at source rank p (Appendix K.3.9); singleton certificates give conditional positive-loss bounds, but even optimized weights can be arbitrarily loose on finite-bit inputs (Appendices K.3.11–K.3.13). Exact worst-request evaluation remains NP-hard with vanishing curvature variation (Theorem 53). Repeated features admit an endpoint bank (Theorem 64); source leverage gives adaptive bounds (Appendix K.3.4). The guarantee throughout is retained-objective utility; statistical unrecoverability and distributional equivalence to retraining require separate guarantees.

AI USE AND VERIFICATION

GPT-6-astra was used for problem formulation, proof development, literature research, experimental design, code, data generation, and drafting. GPT-6 Pro contributed the optimized spectral-certificate separation and finite-bit constructions, which received independent AI proof audits. Verification currently consists of computational checks and independent AI proof audits. No human review of the mathematical claims is represented as having been completed. The synthetic experiments evaluate numerical consequences and failure modes; they do not validate the asymptotic lower bound or establish novelty.

REFERENCES

Youssef Allouah, Joshua Kazdan, Rachid Guerraoui, and Sanmi Koyejo. The utility and complexity of in- and out-of-distribution machine unlearning. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/07bc722f08f096e6ea7ee99349ff0a86-Abstract-Conference.html.

W. N. Anderson, Jr. and G. E. Trapp. Shorted operators. II. *SIAM Journal on Applied Mathematics*, 28(1):60–71, 1975. doi: 10.1137/0128007.

Erwin H. Bareiss. Sylvester's identity and multistep integer-preserving gaussian elimination. *Mathematics of Computation*, 22(103):565–578, 1968. doi: 10.1090/S0025-5718-1968-0226829-0.

Simon Barthelmé, Nicolas Tremblay, and Pierre-Olivier Amblard. A faster sampler for discrete determinantal point processes. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206, pp. 5582–5592, 2023. URL https://proceedings.mlr.press/v206/barthelme23a.html.

Umit Yigit Basaran, Sk Miraj Ahmed, Amit Roy-Chowdhury, and Basak Guler. A certified unlearning approach without access to source data. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 3071–3090, 2025. URL https://proceedings.mlr.press/v267/basaran25a.html.

Jacob L. Block, Aryan Mokhtari, and Sanjay Shakkottai. Machine unlearning under overparameterization. In *Advances in Neural Information Processing Systems*, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/9d7fe07d48d198449d0dbefe86263fdb-Paper-Conference.pdf.

Tobias Breiten and Emil Ringh. Residual-based iterations for the generalized Lyapunov equation. *arXiv preprint arXiv:1807.10715*, 2018. URL https://arxiv.org/abs/1807.10715.

Jop Briët, Fernando Mário de Oliveira Filho, and Frank Vallentin. The positive semidefinite Grothendieck problem with rank constraint. In *Automata, Languages and Programming*, volume 6198 of *Lecture Notes in Computer Science*, pp. 31–42. Springer, 2010. doi: 10.1007/978-3-642-14165-2_4. URL https://homepages.cwi.nl/~jop/psdgroth.pdf.

Alberto Bucci, Yuji Nakatsukasa, and Taejun Park. Numerical stability of the Nyström method. *arXiv preprint arXiv:2511.15583*, 2025. URL https://arxiv.org/abs/2511.15583.

David Carlson, Emilie Haynsworth, and Thomas Markham. A generalization of the Schur complement by means of the Moore–Penrose inverse. *SIAM Journal on Applied Mathematics*, 26(1): 169–175, 1974. doi: 10.1137/0126013. URL https://epubs.siam.org/doi/abs/10.1137/0126013.

Yunlu Chen and Nan Zhang. Optimal subsampling for large sample ridge regression. *arXiv preprint arXiv:2204.04776*, 2022. URL https://arxiv.org/abs/2204.04776.

Yeshwanth Cherapanamjeri, Sumegha Garg, Nived Rajaraman, Ayush Sekhari, and Abhishek Shetty. The space complexity of learning-unlearning algorithms (extended abstract). In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 1185–1193. PMLR, 2025. URL https://proceedings.mlr.press/v291/cherapanamjeri25b.html. Full version: arXiv:2506.13048.

Kenneth L. Clarkson and David P. Woodruff. Numerical linear algebra in the streaming model. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, 2009. URL https://kenclarkson.org/one_pass/p.pdf.

Etienne de Klerk and Frank Vallentin. On the Turing model complexity of interior point methods for semidefinite programming. *SIAM Journal on Optimization*, 26(3):1944–1961, 2016. doi: 10.1137/15M103114X. URL https://arxiv.org/abs/1507.03549.

Alberto Del Pia. Rational Jacobi rotations and the complexity of approximating mixed integer quadratic programming. *arXiv preprint arXiv:2607.29386*, 2026. URL https://arxiv.org/abs/2607.29386.

Michał Dereziński and Michael W. Mahoney. Determinantal point processes in randomized numerical linear algebra. *Notices of the American Mathematical Society*, 68(1):34–45, 2021. doi: 10.1090/noti2202. URL https://www.stat.berkeley.edu/~mmahoney/pubs/RandNLA_in_NoticesAMS_2021.pdf.

Michał Dereziński, Feynman Liang, and Michael W. Mahoney. Exact expressions for double descent and implicit regularization via surrogate random design. In *Advances in Neural Information Processing Systems*, volume 33, pp. 5152–5164, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/37740d59bb0eb7b4493725b2e0e5289b-Abstract.html.

Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936. doi: 10.1007/BF02288367.

Michaël Fanuel, Joachim Schreurs, and Johan A. K. Suykens. Diversity sampling is an implicit regularization for kernel methods. *SIAM Journal on Mathematics of Data Science*, 3(1):280–297, 2021. doi: 10.1137/20M1320031. URL https://arxiv.org/abs/2002.08616.

Fedor V. Fomin, Daniel Lokshtanov, Fahad Panolan, and Saket Saurabh. Efficient computation of representative families with applications in parameterized and exact algorithms. *Journal of the ACM*, 63(4):29:1–29:60, 2016. doi: 10.1145/2886094. URL https://arxiv.org/abs/1304.4626v4.

Michael R. Garey, David S. Johnson, and Larry J. Stockmeyer. Some simplified NP-complete graph problems. *Theoretical Computer Science*, 1(3):237–267, 1976. doi: 10.1016/0304-3975(76)90059-1. URL https://www.sciencedirect.com/science/article/pii/0304397576900591.

Mariko Giga. Some extensions of Kantorovich type inequalities. *RIMS Kôkyûroku*, 1359:30–37, 2004. URL https://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/pdf/1359-4.pdf.

Ryan Giordano, William Stephenson, Runjing Liu, Michael Jordan, and Tamara Broderick. A Swiss Army infinitesimal jackknife. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1139–1147, 2019. URL https://proceedings.mlr.press/v89/giordano19a.html.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016. URL https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.

Mark A. Iwen and Felix Krahmer. Fast subspace approximation via greedy least-squares. *arXiv preprint arXiv:1312.1413*, 2013. URL https://arxiv.org/abs/1312.1413.

Zachary Izzo, Mary Anne Smart, Kamalika Chaudhuri, and James Zou. Approximate data deletion from machine learning models. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 2008–2016. PMLR, 2021. URL https://proceedings.mlr.press/v130/izzo21a.html.

Praneeth Kacham and David P. Woodruff. Sketching algorithms and lower bounds for ridge regression. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 10539–10556. PMLR, 2022. URL https://proceedings.mlr.press/v162/kacham22a.html.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Master's thesis, University of Toronto, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Claire Launay, Bruno Galerne, and Agnès Desolneux. Exact sampling of determinantal point processes without eigendecomposition. *Journal of Applied Probability*, 57(4):1198–1221, 2020. doi: 10.1017/jpr.2020.56. URL https://arxiv.org/abs/1802.08429.

Luo Luo, Cheng Chen, Zhihua Zhang, Wu-Jun Li, and Tong Zhang. Robust frequent directions with application in online learning. *Journal of Machine Learning Research*, 20(45):1–41, 2019. URL https://jmlr.org/papers/v20/17-773.html.

Cyrus Mostajeran, Christian Grussler, and Rodolphe Sepulchre. Geometric matrix midranges. *SIAM Journal on Matrix Analysis and Applications*, 41(3):1347–1368, 2020. doi: 10.1137/19M1273475. URL https://arxiv.org/abs/1907.04188.

Mojmír Mutný, Michał Dereziński, and Andreas Krause. Convergence analysis of block coordinate algorithms with determinantal sampling. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, volume 108, pp. 3110–3120, 2020. URL https://proceedings.mlr.press/v108/mutny20a.html.

Akito Narahara, Takayuki Kawashima, and Takafumi Kanamori. Bootstrap error estimation and sketch-size selection for sketched ridge regression. *arXiv preprint arXiv:2608.15478*, 2026. URL https://arxiv.org/abs/2608.15478.

Yurii Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization Methods and Software*, 9(1–3):141–160, 1998. doi: 10.1080/10556789808805690.

Xinbao Qiao, Meng Zhang, Ming Tang, and Ermin Wei. Hessian-free online certified unlearning. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=C3TrHWanh5.

Yijun Quan and Giovanni Montana. Client-side probing of deleted ridge statistics in federated unlearning, 2026. URL https://arxiv.org/abs/2609.04475.

Yijun Quan, Wentai Wu, and Giovanni Montana. Exact federated continual unlearning for ridge heads on frozen foundation models, 2026. URL https://arxiv.org/abs/2603.12977. Accepted to ECML-PKDD 2026.

Kihun Rhee. When is the sharp covariance envelope tight? feature-only geometry for volume-sampled least squares. *arXiv preprint arXiv:2608.26877*, 2026. URL https://arxiv.org/abs/2608.26877.

Ayush Sekhari, Jayadev Acharya, Gautam Kamath, and Ananda Theertha Suresh. Remember what you want to forget: Algorithms for machine unlearning. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/9627c45df543c816a3ddf2d8ea686a99-Abstract.html.

Benwei Shi and Jeff Phillips. A deterministic streaming sketch for ridge regression. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 586–594. PMLR, 2021. URL https://proceedings.mlr.press/v130/shi21b.html.

Jianheng Tang, Huiping Zhuang, Di Fang, Jiaxu Li, Feijiang Han, Yajiang Huang, Kejia Fan, Leye Wang, Zhanxing Zhu, Shanghang Zhang, Houbing Herbert Song, and Yunhuai Liu. ACU: Analytic continual unlearning for efficient and exact forgetting with privacy preservation, 2025. URL https://arxiv.org/abs/2505.12239v1. Original version, arXiv:2505.12239v1.

Nicolas Tremblay, Simon Barthelmé, Konstantin Usevich, and Pierre-Olivier Amblard. Extended L-ensembles: A new representation for determinantal point processes. *The Annals of Applied Probability*, 33(1):613–640, 2023. doi: 10.1214/22-AAP1824. URL https://arxiv.org/abs/2107.06345.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1–2):1–230, 2015. doi: 10.1561/2200000048. URL https://arxiv.org/abs/1501.01571.

Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996. doi: 10.1137/1038003. URL https://web.stanford.edu/~boyd/papers/pdf/semidef_prog.pdf.

Kasturi Varadarajan, S. Venkatesh, Yinyu Ye, and Jiawei Zhang. Approximating the radii of point sets. *SIAM Journal on Computing*, 36(6):1764–1776, 2007. doi: 10.1137/050627472. URL https://webhome.cs.uvic.ca/~srinivas/papers/radii.pdf.

Shusen Wang, Alex Gittens, and Michael W. Mahoney. Sketched ridge regression: Optimization perspective, statistical perspective, and model averaging. *Journal of Machine Learning Research*, 18(218):1–50, 2018. URL https://jmlr.org/papers/v18/17-313.html.

Jingyi Xie, Linjun Zhang, and Sai Li. Efficient machine unlearning with minimax optimality. *arXiv preprint arXiv:2604.05669*, 2026. URL https://arxiv.org/abs/2604.05669.

A SPECTRAL UPPER BOUND AND FINITE-BIT IMPLEMENTATION

All vector norms below are Euclidean; matrix norms with subscripts op and F are the operator and Frobenius norms. A matrix inequality denotes the positive-semidefinite order. These proofs allow any bounded real labels until the explicit bit count.

A.1 RIDGE IDENTITIES AND A PERTURBATION INEQUALITY

Lemma 6 (Ridge identities). *Let $X \in \mathbb{R}^{m\times d}$ have the retained features as rows and $y \in \mathbb{R}^m$ their labels. For $Q = X^\top X/m + \lambda I$ and $w = w_T$,*

$$w = (X^\top X + m\lambda I)^{-1}X^\top y = X^\top (XX^\top + m\lambda I)^{-1}y, \tag{21}$$

$$F_T(v) - F_T(w) = (v-w)^\top Q(v-w), \tag{22}$$

$$\|w\| \le \frac{1}{2\sqrt{\lambda}}. \tag{23}$$

Proof. Differentiating the objective gives $(X^\top X + m\lambda I)w = X^\top y$. The dual expression follows by multiplying the identity $(X^\top X + m\lambda I)X^\top = X^\top(XX^\top + m\lambda I)$ by the two inverses. Expanding $F_T(v)$ around w cancels the linear term by the normal equations and proves (22). If s is a singular value of X, the corresponding singular value of $(X^\top X + m\lambda I)^{-1}X^\top$ is $s/(s^2 + m\lambda)$, at most $1/(2\sqrt{m\lambda})$ because $(s - \sqrt{m\lambda})^2 \ge 0$. Since $\|y\| \le \sqrt{m}$, this proves (23). □

Lemma 7 (Positive perturbations). *Let $Q \succeq \lambda I$ and $P \succeq 0$. If $Qw = b$ and $(Q+P)\widehat{w} = b$, then*

$$(\widehat{w} - w)^\top Q(\widehat{w} - w) \le \frac{\|Pw\|^2}{\lambda}. \tag{24}$$

Proof. Subtracting the equations gives $\widehat{w} - w = -(Q+P)^{-1}Pw$. Set $D = Q^{-1/2}PQ^{-1/2} \succeq 0$. Then

$$(Q+P)^{-1}Q(Q+P)^{-1} = Q^{-1/2}(I+D)^{-2}Q^{-1/2} \tag{25}$$

$$\preceq Q^{-1} \preceq \lambda^{-1}I. \tag{26}$$

Substituting the difference formula proves the claim. This calculation does not assume that Q and P commute. □

A.2 WHY THE TAIL REMAINS LOW ENERGY AFTER DELETION

For the spectral construction (7), fix an orthonormal eigenbasis of G_S, with any deterministic tie-breaking rule. Let $\Pi = I - V_\ell V_\ell^\top$. Direct inspection in that eigenbasis gives

$$E_0 = K^\top K + \delta I - G_S = \Pi(\delta I - G_S)\Pi, \qquad 0 \preceq E_0 \preceq \delta\Pi. \tag{27}$$

If $T = S - U$ for a legal request, $G_T \preceq G_S$. Thus

$$\Pi G_T \Pi \preceq \Pi G_S \Pi \preceq \delta\Pi, \qquad \left\|\Pi X^\top\right\|_{\mathrm{op}} \le \sqrt{\delta}. \tag{28}$$

Lemma 8 (Relative retained-objective error). *For every nonempty retained multiset $T = S - U$, the spectral-floor decoder satisfies*

$$F_T(\widehat{w}) - F_T(w_T) \le F_T(w_T)\left(\frac{\delta}{m\lambda}\right)^3. \tag{29}$$

The inequality holds simultaneously for all legal deletions, including those for which $F_T(w_T) = 0$.

Proof. Put $f_T = F_T(w_T)$ and $e_T = y - Xw_T$. Stationarity gives $\lambda w_T = X^\top e_T/m$. Consequently,

$$\|\Pi w_T\|^2 \le \frac{\delta}{m^2\lambda^2}\|e_T\|^2 \le \frac{\delta}{m\lambda^2}f_T. \tag{30}$$

The decoder uses $Q_T + E_0/m$ in normalized form. Applying Lemma 7 and using $E_0 = E_0\Pi$ yields

$$F_T(\widehat{w}) - f_T \le \frac{1}{\lambda m^2} \left\| E_0 \Pi w_T \right\|^2 \le \frac{\delta^2}{\lambda m^2} \frac{\delta f_T}{m\lambda^2}. \tag{31}$$

This proves the claim without dividing by f_T. If $f_T = 0$, positivity of λ forces $w_T = 0$ and every retained label to be zero, so the decoder's right-hand side is zero and its output is exact. □

Since $F_T(w_T) \le F_T(0) \le 1$, equation (30) also gives

$$\|\Pi w_T\| \le \frac{\sqrt{\delta}}{\sqrt{m}\lambda}. \tag{32}$$

Lemma 8 therefore proves the additive bound in Proposition 2, including $\delta = 0$. The argument is deterministic for each input and uses no union bound over requests.

A.3 Spectral shape and an exact structured case

Lemma 9 (Dependence on the omitted spectrum)**.** *For the spectral-floor summary, define*

$$\kappa_\ell = \max_{i>\ell} \sigma_i(\delta - \sigma_i)^2.$$

For every nonempty $T = S - U$,

$$F_T(\widehat{w}) - F_T(w_T) \le \frac{\left\| y_T - X_T w_T \right\|^2}{m} \frac{\kappa_\ell}{m^3\lambda^3} \le F_T(w_T) \frac{\kappa_\ell}{m^3\lambda^3}. \tag{33}$$

Moreover, $\kappa_\ell \le 4\delta^3/27$. If, for some $0 \le \zeta \le 1$, every omitted eigenvalue belongs to $\{0\} \cup [(1-\zeta)\delta, \delta]$, then

$$\kappa_\ell \le \zeta^2\delta^3, \qquad F_T(\widehat{w}) - F_T(w_T) \le \zeta^2 F_T(w_T) \left(\frac{\delta}{m\lambda}\right)^3. \tag{34}$$

In particular, at $\zeta = 0$ the unquantized decoder is exact for every legal deletion.

Proof. The error E_0 commutes with G_S, and its eigenvalue on an omitted eigenvector is $\delta - \sigma_i$. Legal deletion gives

$$E_0 G_T E_0 \preceq E_0 G_S E_0 \preceq \kappa_\ell I.$$

Writing $e_T = y_T - X_T w_T$, stationarity therefore yields

$$\|E_0 w_T\|^2 = \frac{\left\| E_0 X_T^\top e_T \right\|^2}{m^2\lambda^2} \le \frac{\kappa_\ell \left\| e_T \right\|^2}{m^2\lambda^2}.$$

Lemma 7 bounds the excess by $\|E_0 w_T\|^2 / (m^2\lambda)$, proving the displayed risk bound. Maximizing $s(\delta - s)^2$ for $0 \le s \le \delta$ gives $4\delta^3/27$. Under the nearly flat spectrum condition, a positive omitted eigenvalue satisfies $\sigma_i \le \delta$ and $\delta - \sigma_i \le \zeta\delta$, so its contribution is at most $\zeta^2\delta^3$; a zero eigenvalue contributes zero. □

Corollary 10 (Exact deletion for covariance isotropic on its support)**.** *Suppose $G_S = \delta P$, where $\delta > 0$ and P is an orthogonal projector that may depend on the data. Storing only b_S, δ, n allows exact answers to every legal deletion with $m = n - |U| > 0$, through*

$$w_{S-U} = [(\delta + m\lambda)I - G_U]^{-1}(b_S - b_U). \tag{35}$$

On the grid (4) *with $d \le n \le 3d$, these statistics and their encoding metadata require $O(d \log d)$ bits. The projector need not be public or stored. If $G_S = 0$, every deletion optimum is zero.*

Proof. Take the factor K to have zero rows, so that $\widehat{G} = \delta I$. Every original feature lies in $\operatorname{range} P$: for $v \in \ker P$, the identity $\sum_S (x^\top v)^2 = v^\top G_S v = 0$ forces every summand to vanish. Thus b_T and all retained features lie in this range. The decoder's normal matrix is

$$(\delta + m\lambda)I - G_U = G_T + m\lambda I + \delta(I - P) \succeq m\lambda I.$$

The true optimum w_T lies in $\operatorname{range} P$, so the added term $\delta(I - P)$ annihilates it. The two systems therefore have the same unique solution. This also follows from Lemma 9 with $\ell = 0$.

For exact finite storage, let $p = p_{\text{in}}(d)$ and $r = \operatorname{rank} G_S \geq 1$. The integer $a = 2^{2p} \operatorname{tr} G_S$ satisfies $0 \leq a \leq n2^{2p}$, and

$$\delta = \frac{\operatorname{tr} G_S}{r} = \frac{a}{r2^{2p}}.$$

Store the two integers a, r to represent this scalar exactly. They require $O(p + \log n + \log d)$ bits. Each coordinate of b_S has denominator dividing 2^{2p} and magnitude at most n, and n is explicitly stored. Since $p = O(\log d)$ and $n = \Theta(d)$, the total is $O(d \log d)$ bits. If $G_S = 0$, all features are zero and a flag selects the zero output. □

This corollary is a structural consequence of the ridge normal equations. It identifies a class where the support can have high rank while its orientation contributes no additional deletion state. A data-dependent whitening map is outside the public fixed representation assumed here; any retained whitening or inverse-transform information is charged to the state. A general nonorthogonal coordinate change also changes the form of the original ridge penalty.

A.4 Rounding the summary

Fix $\beta_0 = 2/3$ and $0 < \varepsilon \leq \varepsilon_0 \leq 1$. Let

$$L = \left\lceil \frac{4^{1/3}}{\beta_0 \lambda \varepsilon^{1/3}} \right\rceil. \tag{36}$$

If $L \geq d$, store G_S, b_S, n exactly and a flag indicating this branch. The decoder computes $(G_S - G_U + m\lambda I)^{-1}(b_S - b_U)$, so its excess is zero in exact arithmetic.

Otherwise choose $\ell = L < d$ and form K, δ as above. Choose a public dyadic number $\eta > 0$ as the largest power of two not exceeding $\beta_0 n\lambda\sqrt{\varepsilon}/6$. In particular,

$$\frac{\beta_0 n\lambda\sqrt{\varepsilon}}{12} < \eta \leq \frac{\beta_0 n\lambda\sqrt{\varepsilon}}{6}. \tag{37}$$

Here "public" means computable from n and the public parameters; n itself is explicitly stored. Round each entry of K to the nearest multiple of a dyadic step h_q satisfying

$$h_q \leq \min\left\{ \frac{\eta}{2\sqrt{n\ell d}}, \sqrt{\frac{2\eta}{\ell d}} \right\}. \tag{38}$$

The largest power of two below the right-hand side is a definite choice. Let K_q be the rounded factor. Since $\|K\|_F \leq \sqrt{n}$ and $\|K_q - K\|_F \leq h_q\sqrt{\ell d}/2$,

$$\left\| K_q^\top K_q - K^\top K \right\|_{\text{op}} \leq 2 \|K\|_{\text{op}} \|K_q - K\|_{\text{op}} + \|K_q - K\|_{\text{op}}^2 \tag{39}$$

$$\leq h_q\sqrt{n\ell d} + h_q^2 \ell d/4 \leq \eta. \tag{40}$$

Encode the integer $j = \lceil \delta/\eta \rceil + 1$ and use the upward-rounded floor $\delta_q = j\eta$. Then

$$\eta \leq \delta_q - \delta < 2\eta, \qquad 0 \preceq R := K_q^\top K_q + \delta_q I - (K^\top K + \delta I) \preceq 3\eta I. \tag{41}$$

The stored state consists of the branch flag, n, ℓ, j, the entries of K_q, and exact b_S. The decoder replaces K, δ by K_q, δ_q in (8).

For any legal deletion, the total Gram error is $E_0 + R \succeq 0$, so the quantized system is also positive definite. Applying Lemma 7, the triangle inequality, (32), and (23) gives

$$\sqrt{F_T(\widehat{w}_q) - F_T(w_T)} \leq \frac{\|E_0 w_T\| + \|R w_T\|}{m\sqrt{\lambda}} \tag{42}$$

$$\leq \left(\frac{\delta}{m\lambda}\right)^{3/2} + \frac{3\eta}{2m\lambda}. \tag{43}$$

The same quantized summary preserves the relative guarantee. Write $f_T = F_T(w_T)$. Replacing the last two norm bounds by (30) and $\|w_T\|^2 \le f_T/\lambda$ gives

$$\sqrt{F_T(\widehat{w}_q) - f_T} \le \sqrt{f_T}\left[\left(\frac{\delta}{m\lambda}\right)^{3/2} + \frac{3\eta}{m\lambda}\right]. \tag{44}$$

As $\operatorname{tr} G_S \le n$, $\delta \le n/(\ell+1)$. Equation (36) makes the first bracketed term at most $\sqrt{\varepsilon}/2$. Equations (37) and $m \ge \beta_0 n$ give the same bound for the second. Thus, simultaneously for every legal deletion,

$$F_T(\widehat{w}_q) \le (1+\varepsilon) F_T(w_T). \tag{45}$$

If $f_T = 0$, exact storage of b_S ensures the retained right-hand side is zero and the quantized decoder again returns zero. No positive lower bound on f_T, estimate of its value, or extra label precision is used. Since $f_T \le 1$, this proves the additive guarantee as well; equation (43) remains a valid absolute estimate.

A.5 Counting bits for every parameter regime

For the prescribed common input grid, every entry of G_S and b_S is an integer multiple of $2^{-2p_{\mathrm{in}}}$ and has magnitude at most n. Since $d \le n \le 3d$ and $p_{\mathrm{in}} = O(\log d)$, these entries each take $O(\log d)$ bits, including their signs. Exact storage therefore uses $O(d^2 \log d)$ bits. The ternary labels in the lower-bound families need only denominator $2^{p_{\mathrm{in}}}$ for b_S, but that restriction is not used in the upper bound.

In the low-rank branch, $L < d$ implies

$$\lambda \varepsilon^{1/3} > \frac{4^{1/3}}{\beta_0 d}. \tag{46}$$

Because $\lambda \le 1$ and $\varepsilon \le 1$, both $\lambda \ge c_1/d$ and $\varepsilon \ge c_2/d^3$ hold for universal positive constants. Consequently the lower endpoint in (37) is bounded below by an inverse polynomial in d, while $\eta \le n/3 = O(d)$. The right-hand side of (38) is also bounded below by an inverse polynomial; selecting its largest admissible dyadic step requires only $O(\log d)$ fractional bits. Every factor entry has magnitude at most $\sqrt{n} + h_q/2$, requiring $O(\log d)$ integer bits as well. Finally $j \le n/\eta + 2$ is polynomially bounded, and n, ℓ, j each take $O(\log d)$ bits. Fixed-length encodings at these known bounds include all metadata and use

$$O\big((d\ell + d)\log d\big) = O\Big(d \min\{d, (\lambda\varepsilon^{1/3})^{-1}\} \log d\Big) \tag{47}$$

bits. In the exact branch, $L \ge d$ implies $(\lambda\varepsilon^{1/3})^{-1} = \Omega(d)$, so the same final expression covers exact storage. This argument does not encode λ inside a real-valued statistic and does not count an infinite-precision real as one word.

Corollary 11 (Memory for relative optimization accuracy)**.** *Let $B^*_{\mathrm{rel}}(d, \lambda, \varepsilon)$ be the least worst-case state length in the model defining B^*, on the same grid* (4)*, when the success event in* (3) *is replaced by $F_{S-U}(\widehat{w}_U) \le (1+\varepsilon) F_{S-U}(w_{S-U})$. Under the dimension, regularization, and accuracy conditions of Theorem 1, there are universal positive constants c, C such that*

$$cd \min\{d, (\lambda\varepsilon^{1/3})^{-1}\} \le B^*_{\mathrm{rel}}(d, \lambda, \varepsilon) \le Cd \min\{d, (\lambda\varepsilon^{1/3})^{-1}\} \log d. \tag{48}$$

The lower bound retains common legal requests and $w_S = 0$, with pointwise randomized success. The upper bound is deterministic and simultaneous over legal requests.

Proof. Equation (45) and the preceding bit count prove the upper bound. Every successful relative answer has additive excess at most $\varepsilon F_{S-U}(w_{S-U}) \le \varepsilon$, because bounded labels give $F_{S-U}(w_{S-U}) \le 1$. A relative summary is therefore an additive summary with the same state length and success probability. The lower bound, including its hard families and quantifiers, follows from Theorem 1. □

This guarantee concerns the optimal regularized empirical objective. The summary needs no promise on its size: exact storage of the response vector keeps the quantized decoder linear in the labels. At fixed features and public parameters, multiplying all labels by s multiplies its output by s and its excess by s^2, whenever the scaled labels remain on the prescribed grid.

A.6 SHARPER SPECTRAL ERROR UNDER EXACT LINEAR LABELS

Suppose $y = Xv$ for a vector with $\|v\| \le 1$. The same dual formula now yields

$$\|\Pi w_T\| \le \left\|\Pi X^\top\right\|_{\mathrm{op}} \left\|(XX^\top + m\lambda I)^{-1}X\right\|_{\mathrm{op}} \|v\| \le \frac{\sqrt{\delta}}{2\sqrt{m\lambda}}. \tag{49}$$

The last inequality follows by maximizing $s/(s^2 + m\lambda)$ over singular values s of X. Substitution into Lemma 7 gives

$$F_T(\widehat{w}) - F_T(w_T) \le \frac{1}{\lambda}\left(\frac{\delta}{m}\right)^2 \|\Pi w_T\|^2 \le \frac{1}{4\lambda^2}\left(\frac{\delta}{m}\right)^3. \tag{50}$$

Also $w_T = (G_T + m\lambda I)^{-1}G_T v$ has norm at most one, because its action on each eigenvector multiplies by a number in $[0, 1]$.

For the quantized factor and upward floor constructed in Appendix A.4, this improved norm bound gives

$$\sqrt{F_T(\widehat{w}_q) - F_T(w_T)} \le \frac{(\delta/m)^{3/2}}{2\lambda} + \frac{3\eta}{m\sqrt{\lambda}}. \tag{51}$$

Choose a rank satisfying $\ell + 1 \ge 1/(\beta_0(\lambda^2\varepsilon)^{1/3})$ and a dyadic $\eta \le \beta_0 n\sqrt{\lambda\varepsilon}/6$ within a factor two of this value. Each term is then at most $\sqrt{\varepsilon}/2$. If the selected rank reaches d, use exact statistics. Otherwise $\lambda^2\varepsilon \ge c/d^3$; for $\lambda, \varepsilon \le 1$ this implies $\lambda \ge c_1 d^{-3/2}$ and $\varepsilon \ge c_2 d^{-3}$. The same factor-rounding estimate consequently uses $O(\log d)$ bits per entry.

For a common input grid with $O(\log d + \log(1/\lambda))$ bits and $n = \Theta(d)$, exact entries of G_S, b_S take that many bits as well. Combining branches gives the upper bound

$$O\Big(d \min\{d, (\lambda^2\varepsilon)^{-1/3}\}(\log d + \log(1/\lambda))\Big). \tag{52}$$

In particular the final logarithm is $O(\log d)$ when $\lambda \ge d^{-O(1)}$. At fixed ρ, $n = \Theta(d/\rho)$ is also $\Theta(d)$, so this statement applies to the noiseless lower-bound families. Constants can depend on the fixed retained fraction.

B INFORMATION AND POINTWISE AMPLIFICATION

All entropies in this section are measured in bits. Let $h_2(q) = -q\log_2 q - (1-q)\log_2(1-q)$ be binary entropy.

Lemma 12 (A large bounded sign family)**.** *For integers $p \ge r \ge 1$, define*

$$\Gamma_{p,r} = \{A \in \{\pm 1/\sqrt{p}\}^{p\times r} : \|A\|_{\mathrm{op}} \le 8\}. \tag{53}$$

Then $|\Gamma_{p,r}| \ge 2^{pr-1}$. Each member has $\|A\|_F^2 = r$.

Proof. Take $1/4$-nets of the unit spheres in $\mathbb{R}^p$ and $\mathbb{R}^r$, with sizes at most 9^p and 9^r. Approximating a maximizing left and right unit vector by net points changes their bilinear form by at most half the operator norm, so any matrix Z satisfies

$$\|Z\|_{\mathrm{op}} \le 2 \max_{u,v \text{ in the nets}} |u^\top Z v|. \tag{54}$$

For an independent Rademacher matrix Z and fixed unit u, v, the sum defining $u^\top Z v$ has squared coefficient sum one. Its exponential-moment bound gives $\Pr[|u^\top Z v| > s] \le 2e^{-s^2/2}$. Taking $s = 4\sqrt{p}$ and a union bound yields

$$\Pr[\|Z\|_{\mathrm{op}} > 8\sqrt{p}] \le 2 \cdot 9^{p+r} e^{-8p} \le 2e^{(2\log 9 - 8)p} < \frac{1}{2}. \tag{55}$$

At least half the sign matrices qualify after division by $\sqrt{p}$. The Frobenius identity follows by summing the pr squared entries $1/p$. □

Lemma 13 (Random sign responses force information)**.** *Let A be uniform on $\Gamma_{p,r}$, let R be independent public randomness, and let $M \in \{0,1\}^b$ be a message that depends on A, R, with any length metadata included in these b bits. Let z be uniform on $\{\pm 1\}^r$, independent of A, M, R. If an estimator $g(M, R, z)$, possibly using fresh randomness, satisfies for a public $a > 0$*

$$\mathbb{E}\,\|g(M,R,z) - aAz\|^2 \le a^2 r/4, \tag{56}$$

then

$$b \ge [1 - h_2(1/4)]pr - 1. \tag{57}$$

The same conclusion holds with target $-aAz$.

Proof. Fresh decoder coins are conditionally independent of A given (M, R, z). Taking their conditional expectation can only reduce squared error. Conditional expectation of the target therefore has no larger error than g. As z is independent of (A, M, R), this conditional expectation is $a\bar{A}z$, where $\bar{A} = \mathbb{E}[A \mid M, R]$. Averaging over z and using $\mathbb{E}[zz^\top] = I_r$ gives

$$a^2\mathbb{E}\left\|A - \bar{A}\right\|_F^2 = a^2\mathbb{E}\left\|(A - \bar{A})z\right\|^2 \le \mathbb{E}\,\|g - aAz\|^2 \le a^2 r/4. \tag{58}$$

Predict each sign of A by the corresponding sign of $\bar{A}$, choosing a fixed sign at zero. A wrong prediction costs at least $1/p$ in squared entrywise error. The expected number of wrong signs is thus at most $pr/4$.

Write the pr signs as a binary vector V, its predicted vector as $\widehat{V}(M, R)$, and the coordinatewise error indicators as E_i. For $q_i = \Pr[E_i = 1]$, conditional subadditivity and conditioning reducing entropy imply

$$H(V \mid M, R) \le \sum_{i=1}^{pr} H(V_i \mid M, R) = \sum_{i=1}^{pr} H(E_i \mid M, R) \tag{59}$$

$$\le \sum_{i=1}^{pr} h_2(q_i) \le pr\, h_2\left(\frac{1}{pr}\sum_i q_i\right) \le pr\, h_2(1/4). \tag{60}$$

The last step uses $\sum_i q_i/(pr) \le 1/4$ and monotonicity of binary entropy on $[0, 1/2]$. Lemma 12 gives $H(A) = H(V) \ge pr - 1$. Hence

$$I(A; M, R) \ge [1 - h_2(1/4)]pr - 1. \tag{61}$$

Since R is independent of A and M has at most 2^b possible values, $I(A; M, R) = I(A; M \mid R) \le H(M \mid R) \le b$. No coordinate independence after conditioning on $\Gamma_{p,r}$ or on M is assumed. Changing g to $-g$ handles the negative target. □

Lemma 14 (Pointwise success to bounded mean-square error)**.** *Suppose that for each fixed input and request, an output v_j satisfies $\Pr[\|v_j - v_*\| \le R_0] \ge 2/3$. Make $L = 181$ independent copies of the encoding and decoding procedure. For each candidate v_i, compute the median of the L distances $\{\|v_i - v_j\| : 1 \le j \le L\}$, including its distance to itself. Choose a candidate with minimum such median. With probability at least $1 - 1/8192$, this candidate lies within $3R_0$ of v_*. If $\|v_*\| \le R_*$, projecting the candidate onto the ball of radius R_* gives an estimator g with*

$$\mathbb{E}\,\|g - v_*\|^2 \le 9R_0^2 + 4R_*^2/8192. \tag{62}$$

The total persistent state is at most LB bits if one copy uses B bits.

Proof. For each fixed input and request, the success indicators of the independent copies have total expectation at least $2L/3$. The bounded-sum exponential estimate bounds the probability of at most $L/2$ successes by $\exp(-L/18) < 1/8192$. On the complementary event, at least 91 candidates are within R_0 of v_*. Each such candidate has median distance at most $2R_0$ because it is within that distance of all the other successful candidates, including itself. The selected candidate also has at least 91 candidates within distance $2R_0$. This set intersects the set of successful candidates, proving distance at most $3R_0$ to v_* by the triangle inequality. Projection onto a closed Euclidean ball containing v_* does not increase distance to v_*. After projection the error is always at most $2R_*$, proving (62). Independence can use independent public tapes as well as independent private and decoder coins; the public tapes remain independent of the data. The argument is pointwise before averaging over any hard input or query distribution. □

C THE SQUARE CONSTRUCTION FOR SMALL REGULARIZATION

Let $d = 2r$, $0 < \lambda \le 1/(2d)$, $n = 4r$, and $m = 3r$. Fix $t = 1/128$, $a = 1/4$, and $A \in \Gamma_{r,r}$. Let P_{top} project onto the first r coordinates and define

$$\mathcal{K}_A = \begin{pmatrix} I + t^2 A^\top A & tA^\top \\ tA & I \end{pmatrix}, \qquad G_A = \mathcal{K}_A - \tfrac{1}{8} P_{\mathrm{top}} - m\lambda I. \tag{63}$$

The off-diagonal perturbation in $\mathcal{K}_A - I$ has norm at most $t \|A\|_{\mathrm{op}} \le 1/16$, while the top diagonal perturbation is positive semidefinite. Thus $\mathcal{K}_A \succeq (15/16)I$. Since $m\lambda \le 3/4$, $G_A \succeq I/16$. Its diagonal entries are $7/8 + t^2 - m\lambda$ in the top block and $1 - m\lambda$ in the bottom block, all at most one. Also $\left\| G_A^{1/2} \right\|_{\mathrm{op}} \le 2$.

Take the $2r$ rows of the symmetric square root $G_A^{1/2}$ as zero-label records. The squared norm of each row equals the corresponding diagonal entry of G_A, so every feature is bounded by one. Add the $2r$ common records $(ae_j, +1), (ae_j, -1)$, $j \le r$. Their contributions to b_S cancel, and all other labels are zero. Therefore $w_S = 0$ for every A.

For a sign vector $z \in \{\pm 1\}^r$, delete the anchor with label $-z_j$ at each coordinate j. The deletion size is $r = n/4$. Every such request is legal for every A, and the retained normal equations are

$$H_A w_{A,z} = b_z, \qquad H_A = \mathcal{K}_A - \tfrac{1}{16} P_{\mathrm{top}}, \qquad b_z = \tfrac{1}{4} \begin{pmatrix} z \\ 0 \end{pmatrix}. \tag{64}$$

We will use the conservative lower bound $H_A \succeq (13/16)I$. The bottom equation gives $w_{\mathrm{bot}} = -tAw_{\mathrm{top}}$. Substitution into the top equation cancels $t^2 A^\top A$, leaving a Schur complement $(15/16)I$. Hence

$$(w_{A,z})_{\mathrm{top}} = \tfrac{4}{15} z, \qquad (w_{A,z})_{\mathrm{bot}} = -c_s A z, \qquad c_s = \tfrac{4t}{15}. \tag{65}$$

In particular $\|w_{A,z}\| \le \sqrt{r}$ and $\|(w_{A,z})_{\mathrm{bot}}\| \le 8c_s\sqrt{r}$.

C.1 A COMMON GRID WITHOUT LOSING THE SQUARE SIGNAL

Let $q = 2^{-p_{\mathrm{in}}(d)}$ be the grid step in (4) and $\tau = c_s/1920$. Then $q \le \tau/d$. Round each entry of $X_A = G_A^{1/2}$ toward zero to a multiple of q, obtaining $\widetilde{X}_A$. This operation does not increase a row norm. Also

$$\left\| \widetilde{X}_A - X_A \right\|_{\mathrm{op}} \le \left\| \widetilde{X}_A - X_A \right\|_F \le dq \le \tau. \tag{66}$$

The change to the private Gram matrix is

$$J = \widetilde{X}_A^\top \widetilde{X}_A - X_A^\top X_A, \qquad \|J\|_{\mathrm{op}} \le 4\tau + \tau^2 \le 5\tau. \tag{67}$$

All common anchors already lie on this grid. The rounded retained normal matrix is $\widetilde{H}_A = H_A + J \succeq (3/4)I$. Let $\widetilde{w}_{A,z}$ solve its normal equations. The resolvent identity yields

$$\widetilde{w}_{A,z} - w_{A,z} = -\widetilde{H}_A^{-1} J w_{A,z}, \tag{68}$$

$$\|\widetilde{w}_{A,z} - w_{A,z}\| \le \tfrac{4}{3} \|J\|_{\mathrm{op}} \sqrt{r} \le 10\tau\sqrt{r} = \frac{c_s\sqrt{r}}{192}. \tag{69}$$

The labels of all modified records remain zero, so the rounded original optimizer is still exactly zero. The deletion multiset itself is unchanged and common to every rounded dataset.

For an algorithm satisfying (3) on the rounded dataset, the exact quadratic identity and $m = 3r$ imply that a successful output obeys

$$\|\widehat{w} - \widetilde{w}_{A,z}\|^2 \le 4r\varepsilon \le 5r\varepsilon. \tag{70}$$

Apply Lemma 14 to the bottom output vectors, and project the selected bottom vector onto the ball of radius $9c_s\sqrt{r}$. Equations (65) and (69) show that this ball contains the rounded bottom target. Averaging also over independent uniform A and z gives

$$\mathbb{E}\, \|g - (\widetilde{w}_{A,z})_{\mathrm{bot}}\|^2 \le 45r\varepsilon + \frac{324c_s^2 r}{8192} < \frac{c_s^2 r}{16} \tag{71}$$

provided $\varepsilon \le c_s^2/2880$. Finally,

$$\mathbb{E}\,\|g + c_s A z\|^2 \le 2\mathbb{E}\,\|g - (\widetilde{w}_{A,z})_{\mathrm{bot}}\|^2 + 2\mathbb{E}\,\|(\widetilde{w}_{A,z} - w_{A,z})_{\mathrm{bot}}\|^2 \tag{72}$$

$$\le c_s^2 r\left(\frac{1}{8} + \frac{2}{192^2}\right) < \frac{c_s^2 r}{4}. \tag{73}$$

Lemma 13 applied to the concatenated 181 summaries gives

$$181B \ge [1 - h_2(1/4)]r^2 - 1 = \Omega(d^2). \tag{74}$$

Because $\lambda \le 1/(2d)$ and $\varepsilon \le 1$, this is the desired saturated branch of Theorem 1.

D The rectangular construction and accuracy-dependent rank

We now treat $\lambda \ge 1/(2d)$. Fix $t = 1/128$, $u = (1+t^2)^{-1}$, and define the universal constants

$$a_0 = t^2/2^{17}, \qquad \kappa = (a_0/64)^{1/3}, \qquad \varepsilon_0 = \min\{\kappa^3, c_s^2/2880\}. \tag{75}$$

For $0 < \varepsilon \le \varepsilon_0$, put

$$r = \min\left\{d/2, \left\lfloor \frac{\kappa}{\lambda \varepsilon^{1/3}} \right\rfloor\right\}, \qquad p = d - r, \qquad h = \lceil p/r \rceil, \qquad \mu = 2r\lambda. \tag{76}$$

As $\varepsilon^{1/3} \le \kappa$ and $\lambda \le 1$, the expression inside the floor is at least one, so $r \ge 1$. Using $\lfloor x \rfloor \ge x/2$ for $x \ge 1$,

$$r \ge \frac{\kappa}{2}\min\{d, (\lambda\varepsilon^{1/3})^{-1}\}, \qquad p \ge d/2. \tag{77}$$

If the dimension cap is active, $\mu = d\lambda \ge 1/2$. Otherwise $r \ge 1/(2\lambda)$, so $\mu \ge 1$. In all cases

$$\frac{1}{2} \le \mu \le 2\kappa\varepsilon^{-1/3}. \tag{78}$$

D.1 Dataset and exact response for every $\mu \ge 1/2$

For $A \in \Gamma_{p,r}$, let

$$X_A = \sqrt{u}\begin{pmatrix} I_r \\ tA \end{pmatrix} \in \mathbb{R}^{d\times r}. \tag{79}$$

Each column has norm exactly one, since each column of A has squared norm one. Form the dataset from h copies of every column of X_A with label zero and h copies of each anchor $(e_j, +1), (e_j, -1)$ for $j \le r$. Its size is $n = 3hr$, where

$$\frac{3d}{2} \le 3p \le n \le 3d. \tag{80}$$

Again $b_S = 0$ and $w_S = 0$. On query $z \in \{\pm 1\}^r$, delete all h copies of the anchor with label $-z_j$ for each j. The request is independent of A, legal on the entire family, and has size $hr = n/3$. There remain $m = 2hr$ records.

Divide the retained normal equations by h. With $D = u + 1 + \mu$, they become

$$H_0 w = \begin{pmatrix} z \\ 0 \end{pmatrix}, \qquad H_0 = \begin{pmatrix} DI_r & utA^\top \\ utA & \mu I_p + ut^2 AA^\top \end{pmatrix} \succeq \mu I. \tag{81}$$

Let v and b denote the top and bottom blocks of w. The top equation says $v = D^{-1}(z - utA^\top b)$. Substitution into the bottom equation gives

$$\left[\mu I_p + \frac{(1+\mu)ut^2}{D} AA^\top\right] b = -\frac{ut}{D} Az. \tag{82}$$

The identity $(I + \eta AA^\top)^{-1}A = A(I + \eta A^\top A)^{-1}$ therefore yields

$$b = -c_\mu f(A) z, \quad c_\mu = \frac{ut}{\mu D}, \quad f(A) = A(I + \eta_A A^\top A)^{-1}, \quad \eta_A = \frac{(1+\mu)ut^2}{\mu D}. \tag{83}$$

For every $\mu \ge 1/2$, $0 \le \eta_A \le 4t^2$. Moreover $\|f(A)\|_{\mathrm{op}} \le \|A\|_{\mathrm{op}} \le 8$, and

$$\|f(A) - A\|_F^2 = \left\|-\eta_A A(A^\top A)(I + \eta_A A^\top A)^{-1}\right\|_F^2 \tag{84}$$

$$\le \eta_A^2 \|A\|_{\mathrm{op}}^4 \|A\|_F^2 \le 16t^4\, 8^4 r = \frac{r}{4096}. \tag{85}$$

This estimate is uniform for arbitrarily large μ; an upper bound on μ by an absolute constant is not required.

D.2 THE RESPONSE SIGNAL DETERMINES THE RANK

Since $\lambda r = \mu/2$, define

$$E_r = \lambda r c_\mu^2 = \frac{u^2 t^2}{2\mu(u+1+\mu)^2} \geq \frac{t^2}{32(1+\mu)^3}. \tag{86}$$

For the inequality, use $u \geq 1/2$, $u + 1 + \mu \leq 2(1+\mu)$, and $\mu \leq 1 + \mu$. Because $\kappa\varepsilon^{-1/3} \geq 1$, equation (78) gives $1 + \mu \leq 3\kappa\varepsilon^{-1/3}$. The definition of κ implies

$$\varepsilon \leq \frac{a_0}{(1+\mu)^3} \leq \frac{E_r}{4096}, \tag{87}$$

where the first inequality follows from $a_0/(27\kappa^3) = 64/27 > 1$, and the second from $a_0 = t^2/(32 \cdot 4096)$. Thus the target excess remains below the hidden matrix's response scale at the rank in (76).

D.3 UNIFORM INPUT ROUNDING FOR ALL REGULARIZATION SCALES

Use the same grid step $q = 2^{-p_{\mathrm{in}}(d)}$. It satisfies $q \leq t/(23040d)$. Round every entry of X_A toward zero to obtain $\widetilde{X}_A$. Every feature norm remains at most one. Since $\|X_A\|_{\mathrm{op}} \leq 2$,

$$\left\|X_A - \widetilde{X}_A\right\|_{\mathrm{op}} \leq dq, \qquad J = \widetilde{X}_A\widetilde{X}_A^\top - X_A X_A^\top, \qquad \|J\|_{\mathrm{op}} \leq 5dq \leq t/4608. \tag{88}$$

After dividing the normal equations by h, the rounded system has matrix $H_0 + J$ and unchanged right-hand side $(z, 0)$. In particular $H_0 + J \succeq \mu I/2$. Let w and $\widetilde{w}$ be the ideal and rounded optima. Since $\|w\| \leq \sqrt{r}/\mu$, the resolvent identity gives

$$\|\widetilde{w} - w\| \leq \frac{2\,\|J\|_{\mathrm{op}}\sqrt{r}}{\mu^2}. \tag{89}$$

Dividing by $c_\mu\sqrt{r}$ and using $u \geq 1/2$, $D \leq 2 + \mu$, and $\mu \geq 1/2$ shows

$$\frac{\|\widetilde{w} - w\|}{c_\mu\sqrt{r}} \leq \frac{2D\,\|J\|_{\mathrm{op}}}{ut\mu} \leq \frac{36\,\|J\|_{\mathrm{op}}}{t} \leq \frac{1}{128}. \tag{90}$$

Only zero-label records changed. Hence every rounded original model is still zero, all requests are unchanged, and all input coordinates share the same $O(\log d)$-bit grid independently of λ, ε.

The rounded normalized retained matrix is at least $\lambda I/2$ by the displayed system bound; the ridge form in fact gives λI. Thus a successful deletion output satisfies the conservative estimate

$$\|\widehat{w} - \widetilde{w}\|^2 \leq 2\varepsilon/\lambda. \tag{91}$$

Apply Lemma 14 to its bottom block and project onto the ball of radius $9c_\mu\sqrt{r}$. This ball contains $\widetilde{w}_{\mathrm{bot}}$ by (83), $\|f(A)\|_{\mathrm{op}} \leq 8$, and (90). Equations (62) and (87) give

$$\mathbb{E}\,\|g - \widetilde{w}_{\mathrm{bot}}\|^2 \leq \frac{18\varepsilon}{\lambda} + \frac{324c_\mu^2 r}{8192} \tag{92}$$

$$\leq c_\mu^2 r\left(\frac{18}{4096} + \frac{324}{8192}\right) < \frac{c_\mu^2 r}{16}. \tag{93}$$

For each fixed A, $\mathbb{E}_z\,\|(f(A) - A)z\|^2 = \|f(A) - A\|_F^2$. Decomposing the error to $-c_\mu Az$ into the algorithmic error, rounding bias, and response bias, and applying $\|a + b + c\|^2 \leq 3(\|a\|^2 + \|b\|^2 + \|c\|^2)$, yields

$$\mathbb{E}\,\|g + c_\mu Az\|^2 \leq 3\mathbb{E}\,\|g - \widetilde{w}_{\mathrm{bot}}\|^2 + 3\mathbb{E}\,\|\widetilde{w}_{\mathrm{bot}} - w_{\mathrm{bot}}\|^2 + 3c_\mu^2\mathbb{E}\,\|(f(A) - A)z\|^2 \tag{94}$$

$$\leq c_\mu^2 r\left(\frac{3}{16} + \frac{3}{128^2} + \frac{3}{4096}\right) < \frac{c_\mu^2 r}{4}. \tag{95}$$

Lemma 13 now implies

$$181B \geq [1 - h_2(1/4)]pr - 1. \tag{96}$$

Together with (77) and $p \geq d/2$, this proves the lower bound in Theorem 1 for $\lambda \geq 1/(2d)$. Appendix C handles the complementary range. Both constructions have $n \in [d, 3d]$. The proof applies to any encoder, including one with unrestricted offline access: its only information channel to the decoder is the charged message and the common request. At no step is simultaneous randomized correctness on all requests assumed.

D.4 THE PRECISION LOGARITHM IS NECESSARY AT SUFFICIENTLY HIGH ACCURACY

The logarithmic gap in Theorem 1 cannot be removed uniformly by improving its upper bound: on the same input grid, sufficiently accurate deletion can require quadratically many $\log d$-bit quantities. The following result concerns a restricted high-accuracy regime; it does not settle the intermediate regime.

Proposition 15 (Precision saturation on the fixed input grid)**.** *There is a universal $c > 0$ such that, for every sufficiently large even d, $1/(2d) \le \lambda \le 1$, and*

$$0 < \varepsilon \le \frac{2^{-30}}{d^4\lambda(1+d\lambda)^2}, \tag{97}$$

one has

$$cd^2 \log d \le B^*(d,\lambda,\varepsilon) \le B^*_{\mathrm{rel}}(d,\lambda,\varepsilon) \le Cd^2 \log d \tag{98}$$

for a universal C. The lower bound has $w_S = 0$, common legal deletions of exactly one third of the records, and pointwise randomized success, on exactly the grid (4)*.*

Proof. Let p be the largest power of four with $p \le d/2$, and put $K = \sqrt{p}$. Thus $d/8 < p \le d/2$, and K is an even integer for sufficiently large d. Consider matrices in $\mathbb{R}^{p\times p}$ whose entries belong to the K-element alphabet

$$\mathcal{A}_p = \left\{ \frac{-(K-1)}{p}, \frac{-(K-3)}{p}, \dots, \frac{K-1}{p} \right\}.$$

Let Γ be the subfamily with operator norm at most eight. For an independent uniform matrix in this alphabet, every entry has mean zero and absolute value at most $1/\sqrt{p}$. Hoeffding's exponential-moment bound and the same two $1/4$-nets used in Lemma 12 give

$$\Pr[\|A\|_{\mathrm{op}} > 8] \le 2 \cdot 9^{2p} e^{-8p} < \tfrac{1}{2}.$$

Consequently

$$|\Gamma| \ge K^{p^2}/2, \qquad \log_2 |\Gamma| \ge p^2 \log_2 K - 1. \tag{99}$$

Each column has squared norm at most one. Distinct matrices have Frobenius distance at least $2/p$.

Use $2p$ active coordinates and pad with zero coordinates to dimension d. Set $a = 1/2$, $t = 1/128$, and $h = \lceil d/(3p) \rceil$. For every $j \le p$, include h copies of the private zero-label record

$$\left(\begin{pmatrix} ae_j \\ tA_{:j} \end{pmatrix}, 0 \right)$$

and h copies of each public anchor $(e_j, +1), (e_j, -1)$. The private feature norms are at most $\sqrt{a^2+t^2} < 1$. Their nonzero coordinates have denominator dividing $128p$, a power of two at most $64d$, and therefore lie exactly on the grid (4); no input rounding is used. The sample count is $n = 3hp$, with $d \le n < d + 3p \le 5d/2$, and $b_S = 0$ for every source matrix. On query $z \in \{\pm 1\}^p$, delete all h anchors labeled $-z_j$ at coordinate j. This common legal request leaves $m = 2hp$ records.

Write $\mu = 2p\lambda$, $D = 1 + a^2 + \mu$, and

$$c_\mu = \frac{at}{\mu D}, \qquad \eta = \frac{t^2(1+\mu)}{\mu D}, \qquad f(A) = A(I + \eta A^\top A)^{-1}.$$

Dividing the retained normal equations by h and eliminating the top block, as in (81)– (83), gives the exact bottom response

$$v_A(z) = -c_\mu f(A) z. \tag{100}$$

Here $\mu \ge p/d > 1/8$, so $\eta \le 8t^2$ and $64\eta \le 1/32$. In particular f is uniformly separated on Γ. To see this, set $C_A = (I + \eta A^\top A)^{-1}$. The identities

$$f(A) - f(B) = (A-B)C_A + B(C_A - C_B), \qquad C_A - C_B = \eta C_A (B^\top B - A^\top A) C_B$$

and $\|A\|_{\mathrm{op}}, \|B\|_{\mathrm{op}} \le 8$ imply

$$\|f(A) - f(B)\|_F \ge \left(\frac{1}{1+64\eta} - 128\eta \right) \|A-B\|_F \ge \tfrac{1}{2} \|A-B\|_F \ge \frac{1}{p} \quad (A \ne B). \tag{101}$$

For any fixed matrix J and uniform independent sign vector z,

$$\mathbb{E}_z \|Jz\|^2 = \|J\|_F^2, \qquad \mathbb{E}_z \|Jz\|^4 \le 3\|J\|_F^4.$$

The second inequality follows by expanding the fourth moment and retaining only paired sign indices. The elementary second-moment inequality applied to $\|Jz\|^2$ therefore yields

$$\Pr_z\left[\|Jz\| \ge \|J\|_F/\sqrt{2}\right] \ge 1/12.$$

Together with (101), this proves that every pair $A \ne B$ has

$$\Pr_z\left[\|v_A(z) - v_B(z)\| \ge \frac{c_\mu}{\sqrt{2}p}\right] \ge 1/12. \tag{102}$$

Suppose first that

$$\varepsilon \le \frac{\lambda c_\mu^2}{576p^2}. \tag{103}$$

A successful deletion output has bottom error at most $\sqrt{\varepsilon/\lambda}$, since $Q_T \succeq \lambda I$. Apply the candidate-selection rule of Lemma 14 to 181 independent copies, without its final projection. The resulting output g has, for every fixed A, z, probability at least $1-\delta_0$, where $\delta_0 = 1/8192$, of lying within $r_0 = c_\mu/(8p)$ of $v_A(z)$. Let M be the concatenated state and R the concatenated independent public tapes; M has at most $181B$ bits. Fresh decoder randomness, denoted by D_0, is independent of the source conditional on (M, R, z).

For each candidate source A and fixed (M, R), define the score

$$s_A(M,R) = \Pr_{z,D_0}\left[\|g(M,R,z,D_0) - v_A(z)\| \le r_0\right].$$

This is a definition of an information-theoretic estimator from the state, not an assumption that one randomized state answers all queries correctly. Because $2r_0 < c_\mu/(\sqrt{2}p)$, (102) gives $s_A(M,R) + s_B(M,R) \le 2 - 1/12$ for every distinct pair. Thus at most one candidate has score greater than $23/24$. The amplified pointwise guarantee, averaged over independent uniform A, z, gives $\mathbb{E}[1 - s_A(M,R)] \le \delta_0$. By Markov's inequality, the true source has score greater than $23/24$ with probability at least $1 - 24\delta_0$. Selecting that unique candidate, if it exists, therefore recovers A from (M,R) with error probability at most $q_0 = 24/8192$. Only the original finite state is used in defining these scores; the model imposes no bound on decoding computation.

The elementary finite-alphabet entropy bound for this source estimator is

$$H(A \mid M, R) \le h_2(q_0) + q_0 \log_2 |\Gamma|.$$

Since R is independent of the uniform source, it follows that

$$181B \ge I(A; M, R) \ge (1-q_0)\log_2|\Gamma| - h_2(q_0) = \Omega(p^2 \log K) = \Omega(d^2 \log d).$$

The score argument is why the amplification count is constant even though the alphabet has $\Theta(\sqrt{d})$ levels.

Finally, $at = 1/256$ and $\mu = 2p\lambda$ show that the right-hand side of (103) equals

$$\frac{1}{576 \cdot 256^2 \cdot 4p^4 \lambda D^2}.$$

Using $p \le d/2$ and $D = 5/4 + 2p\lambda \le (5/4)(1+d\lambda)$, its denominator is at most $14{,}745{,}600\, d^4\lambda(1+d\lambda)^2$. Since $2^{30} > 14{,}745{,}600$, (97) implies (103). Exact storage of G_S, b_S, n proves the matching upper bound on the fixed grid, and relative success implies additive success. □

E Comparison baselines

E.1 An isotropically shifted truncation

Suppose C and $\Delta \ge 0$ satisfy $0 \preceq G_S - C^\top C \preceq \Delta I$. The decoder using

$$[C^\top C - G_U + (m\lambda + \Delta)I]^{-1}(b_S - b_U) \tag{104}$$

has normalized error $P = (C^\top C + \Delta I - G_S)/m$ between zero and τI, where $\tau = \Delta/m$. Lemmas 6 and 7 imply excess at most $\tau^2/(4\lambda^2)$. There is also a bound linear in τ: optimality for $F_T(v) + v^\top P v$ implies

$$F_T(\widehat{w}) - F_T(w_T) \le w_T^\top P w_T - \widehat{w}^\top P \widehat{w} \le \tau \|w_T\|^2 \le \tau/(4\lambda). \tag{105}$$

Hence the excess is at most $\min\{\Delta/(4m\lambda), \Delta^2/(4m^2\lambda^2)\}$. Taking C to be the ordinary leading-rank factor and $\Delta = \sigma_{\ell+1} \le n/(\ell+1)$ gives the stated $O((\lambda\sqrt{\varepsilon})^{-1})$ numerical-rank baseline at a fixed retained fraction. Its error generally acts on the leading subspace as well, so this full-space estimate alone does not prove the cubic bound.

E.2 Reusing the original model

For $\alpha = |U|/n$ and $\beta = 1 - \alpha$, the universal reuse bound is

$$F_{S-U}(w_S) - F_{S-U}(w_{S-U}) \le \min\left\{\frac{1}{\beta}, \frac{\alpha^2(1+\sqrt{\lambda})^4}{4\beta^2\lambda^2}\right\}. \tag{106}$$

For $\alpha = |U|/n > 0$ and $\beta = 1 - \alpha$, let $Q_U = G_U/|U| + \lambda I$ and $\bar{b}_U = b_U/|U|$. Splitting the original normalized normal equations into retained and deleted records yields

$$Q_T(w_S - w_T) = \frac{\alpha}{\beta}(\bar{b}_U - Q_U w_S). \tag{107}$$

The feature and label bounds give $\|\bar{b}_U\| \le 1$ and $\|G_U/|U|\|_{\mathrm{op}} \le 1$. Consequently

$$\|\bar{b}_U - Q_U w_S\| \le 1 + \frac{1+\lambda}{2\sqrt{\lambda}} = \frac{(1+\sqrt{\lambda})^2}{2\sqrt{\lambda}}. \tag{108}$$

Using the quadratic excess identity and $Q_T^{-1} \preceq \lambda^{-1} I$ proves the second bound in (106). For the first, nonnegativity of the loss and regularizer implies $\beta F_T(w_S) \le F_S(w_S) \le F_S(0) \le 1$, and $F_T(w_T) \ge 0$. At $\alpha = 0$, the excess is exactly zero.

For a matching order with real-valued features, use one coordinate: take every retained record to be $(\sqrt{\lambda}, 1)$ and every deleted record to be $(1, -1)$, with the specified multiplicity ratio. These records obey the bounds for $0 < \lambda \le 1$. Direct solution gives

$$w_T = \frac{1}{2\sqrt{\lambda}}, \qquad w_S = \frac{(1-\alpha)\sqrt{\lambda} - \alpha}{(2-\alpha)\lambda + \alpha}, \tag{109}$$

and therefore

$$F_T(w_S) - F_T(w_T) = \frac{\alpha^2(1+\sqrt{\lambda})^4}{2((2-\alpha)\lambda + \alpha)^2}. \tag{110}$$

This is at least $\min\{\alpha^2/\lambda^2, 1\}/18$. Together with the upper bound and $\beta \ge 1 - \rho_0$, it proves the claimed Θ_{ρ_0} order for fixed $\alpha \le \rho_0 < 1$ and feasible multiplicity ratios. Because $\sqrt{\lambda}$ need not lie on the input grid, this exact lower example is not a separate grid-uniform lower theorem. No such restriction is needed for the universal upper bound (106).

E.3 Noiseless data with a bounded teacher

Proposition 16 (Storing a bounded interpolator)**.** *Suppose every record in S satisfies $y = x^\top v$ for some vector v with $\|v\| \le R$. Then storing any such interpolator and returning it after any legal deletion gives retained excess at most λR^2. If $R > 0$ and $\varepsilon > \lambda R^2$, put*

$$\delta_v = \frac{\sqrt{\varepsilon} - \sqrt{\lambda} R}{\sqrt{1+\lambda}}. \tag{111}$$

For public R, a quantized interpolator gives excess at most ε using

$$O\left(d \log\left(1 + \frac{R\sqrt{d}}{\delta_v}\right)\right) \tag{112}$$

bits when $\delta_v < R$. If $\delta_v \ge R$, the zero output suffices. For fixed R and $\varepsilon \ge 2\lambda R^2$, the storage is $O_R(d \log(2 + d/\varepsilon))$ bits, with no additional deletion statistics.

Proof. The encoder can choose a minimum-norm interpolator of S, which has norm at most R. It interpolates every subset $T \le S$, so $F_T(v) = \lambda \|v\|^2 \le \lambda R^2$ and the excess is no larger. The stored predictor itself is charged to the summary.

For a rounded vector v_q satisfying $\|v_q - v\| \le \delta_v$, stack the normalized prediction error and regularization term. The triangle inequality gives

$$\sqrt{F_T(v_q)} \le \sqrt{\lambda}\,\|v\| + \left\| \begin{pmatrix} X_T/\sqrt{|T|} \\ \sqrt{\lambda} I \end{pmatrix} (v_q - v) \right\| \le \sqrt{\lambda} R + \sqrt{1+\lambda}\,\delta_v = \sqrt{\varepsilon}. \tag{113}$$

The second inequality uses $\|X_T\|_{\mathrm{op}}^2 / |T| \le 1$. Round coordinates with step $2\delta_v/\sqrt{d}$, using fixed tie-breaking. The norm error is at most δ_v, and the bounded coordinate range $[-R, R]$ requires the stated number of bits; the scale is determined by the public parameters. If $\delta_v \ge R$, choosing $v_q = 0$ already satisfies the same error bound. If $R = 0$, all labels vanish and the zero vector is exact. Finally, $\varepsilon \ge 2\lambda R^2$ and $\lambda \le 1$ imply $\delta_v \ge (1 - 1/\sqrt{2})\sqrt{\varepsilon}/\sqrt{2}$, giving the last bit bound. □

F MEMORY BOUNDS THAT TRACK THE DELETION FRACTION

Fix $\lambda = 1/4$ and allow the deletion budget ρ to vanish. Centering the spectral floor at the original optimizer improves the dependence on ρ over the uncentered cubic bound. The upper bound holds for arbitrary bounded labels, including those inconsistent with every linear predictor.

Theorem 17 (A finite-bit upper bound for small deletion fractions)**.** *Let $d, n \ge 1$, $0 < \rho \le 1/8$, $0 < \varepsilon \le 1$, and $\lambda = 1/4$. Suppose that every record in a dataset S satisfies $\|x\|_2 \le 1$ and $|y| \le 1$. Define*

$$R_{\rho,\varepsilon} = \min\left\{ \frac{\rho}{\sqrt{\varepsilon}}, \left(\frac{\rho}{\varepsilon}\right)^{1/3} \right\}, \qquad \ell = \min\{d, \lceil 128 R_{\rho,\varepsilon} \rceil\}.$$

There is a deterministic offline summary using

$$O(d(\ell+1)\log(2 + d/\varepsilon) + \log(2+n)) \tag{114}$$

bits that, for every legal deletion multiset $U \subseteq S$ with $|U| \le \rho n$, returns an answer satisfying

$$F_{S-U}(\widehat{w}_U) - F_{S-U}(w_{S-U}) \le \varepsilon.$$

The decoder receives only this summary and the deleted records, including their labels. All persistent data-dependent state is included in (114)*; n may either be public or be stored in the indicated metadata. The dimension, accuracy, regularization, and deletion budget are public. The statement allows arbitrary bounded real input records and does not require a finite input grid. Temporary computation is unrestricted.*

Proof. We use normalized original moments

$$\Sigma = \frac{G_S}{n}, \qquad b = \frac{b_S}{n}, \qquad \theta = (\Sigma + \lambda I)^{-1} b = w_S.$$

Write $\sigma_1 \ge \cdots \ge \sigma_d \ge 0$ for the eigenvalues of Σ. If $\ell < d$, put $\delta = \sigma_{\ell+1}$; if $\ell = d$, put $\delta = 0$. Choose orthonormal eigenvectors v_i and let the ith row of $K \in \mathbb{R}^{\ell \times d}$ be $\sqrt{\sigma_i - \delta}\, v_i^\top$. Then

$$\widehat{\Sigma}_0 = K^\top K + \delta I = \Sigma + E_0, \qquad 0 \preceq E_0 \preceq \delta \Pi, \qquad E_0 = \Pi E_0 \Pi, \tag{115}$$

where Π projects onto the orthogonal complement of the chosen leading ℓ eigenvectors. For $\ell = d$, take $\Pi = 0$. Since $\operatorname{tr}\Sigma \le 1$, we have $\|K\|_F \le 1$ and, when $\ell < d$,

$$\delta \le \frac{1}{\ell + 1}. \tag{116}$$

The projector is used only in the analysis and is not stored or computed by the decoder.

Movement of the optimizer. Fix a legal query and set

$$\alpha = \frac{|U|}{n}, \qquad \beta = 1 - \alpha, \qquad C = \frac{G_U}{n}, \qquad f = \frac{b_U}{n}, \qquad T = S - U.$$

Thus $\beta \geq 7/8$, $0 \preceq C \preceq \Sigma$, and $\|C\|_{\mathrm{op}} \leq \alpha$. Define

$$Q = \Sigma - C + \beta\lambda I, \qquad \Delta = w_T - \theta.$$

The retained normal matrix is Q/β, and $Q \succeq \beta\lambda I$.

For any nonempty bounded dataset A, its ridge optimizer obeys $\|w_A\|_2 \leq 1/(2\sqrt{\lambda}) = 1$. Indeed, if $s^2 = w_A^\top (G_A/|A|) w_A$, its normal equation and Cauchy–Schwarz give

$$s^2 + \lambda \|w_A\|_2^2 = w_A^\top b_A/|A| \leq s, \qquad \lambda \|w_A\|_2^2 \leq s - s^2 \leq 1/4.$$

In particular, $\|\theta\|_2, \|w_T\|_2 \leq 1$. Let X_U have the deleted feature vectors as rows, and put $r_{T,U} = y_U - X_U w_T$. Each coordinate of $r_{T,U}$ has magnitude at most two. Subtracting the original and retained normal equations gives the exact identity

$$(\Sigma + \beta\lambda I)\Delta = \alpha\lambda\theta - \frac{1}{n} X_U^\top r_{T,U}. \tag{117}$$

Consequently,

$$\|\Delta\|_2 \leq \frac{\alpha(\lambda + 2)}{\beta\lambda} = \frac{9\alpha}{\beta}. \tag{118}$$

There is a second bound in the omitted subspace. Original stationarity gives $\lambda\theta = X_S^\top (y_S - X_S\theta)/n$, while $\|y_S - X_S\theta\|_2^2/n \leq F_S(\theta) \leq 1$. Since $\Pi\Sigma\Pi \preceq \delta\Pi$, this implies $\|\Pi\theta\|_2 \leq \sqrt{\delta}/\lambda$. Legality also implies $\Pi C \Pi \preceq \Pi\Sigma\Pi \preceq \delta\Pi$, so

$$\left\| \frac{1}{n} \Pi X_U^\top r_{T,U} \right\|_2 \leq \sqrt{\delta}\, \frac{\|r_{T,U}\|_2}{\sqrt{n}} \leq 2\sqrt{\alpha\delta}.$$

The matrix $\Sigma + \beta\lambda I$ commutes with Π. Applying Π to (117) therefore yields

$$\|\Pi\Delta\|_2 \leq \frac{\alpha\sqrt{\delta} + 2\sqrt{\alpha\delta}}{\beta\lambda} \leq \frac{12\sqrt{\alpha\delta}}{\beta}. \tag{119}$$

These statements include the empty query, for which $\Delta = 0$.

Centered flooring with a positive rounding error. For the moment suppose θ is exact, and let $\widehat{\Sigma} = \widehat{\Sigma}_0 + R$ with $0 \preceq R \preceq 3\eta I$, where $\eta > 0$ will be chosen below. Put

$$H = \widehat{\Sigma} - C + \beta\lambda I = Q + E_0 + R \succeq \beta\lambda I,$$

and consider the centered answer

$$\overline{w} = \theta + H^{-1}\big[(C + \alpha\lambda I)\theta - f\big]. \tag{120}$$

Its error satisfies $H(\overline{w} - w_T) = -(E_0 + R)\Delta$. The matrices need not commute: because $H \succeq Q \succ 0$,

$$H^{-1} Q H^{-1} \preceq H^{-1} \preceq Q^{-1} \preceq (\beta\lambda)^{-1} I.$$

The exact quadratic excess and the triangle inequality give

$$\begin{aligned} \sqrt{F_T(\overline{w}) - F_T(w_T)} &\leq \frac{\|E_0\Delta\|_2 + \|R\Delta\|_2}{\beta\sqrt{\lambda}} \\ &\leq 32 \min\{\alpha\delta, \sqrt{\alpha}\,\delta^{3/2}\} + 9\eta. \end{aligned} \tag{121}$$

For the first term, use (115) with both (118) and (119); their respective coefficients are $18/\beta^2$ and $24/\beta^2$, both at most 32. For the second, the coefficient is $54\alpha/\beta^2 \leq 9$, using $\alpha \leq 1/8$. In particular, before rounding the centered floor has excess at most

$$1024 \min\{\alpha^2\delta^2, \alpha\delta^3\}. \tag{122}$$

By the choice of ℓ, the first term on the right side of (121) is at most $\sqrt{\varepsilon}/4$. To check this, if $\ell = d$ then $\delta = 0$. Otherwise, $\delta \leq 1/(\ell+1)$ and $\alpha \leq \rho$. If $R_{\rho,\varepsilon} = \rho/\sqrt{\varepsilon}$, use the first term in the minimum and $\ell \geq 128\rho/\sqrt{\varepsilon}$. If $R_{\rho,\varepsilon} = (\rho/\varepsilon)^{1/3}$, use the second term and $\ell \geq 128(\rho/\varepsilon)^{1/3}$; its coefficient is $32/128^{3/2} \leq 1/4$.

Finite precision and the actual decoder. Choose the public dyadic number η to be the largest power of two not exceeding $\sqrt{\varepsilon}/100$. Thus

$$\frac{\sqrt{\varepsilon}}{200} < \eta \le \frac{\sqrt{\varepsilon}}{100} \le \frac{1}{100}.$$

Let h be the largest power of two not exceeding $\eta/(2\sqrt{\ell d})$, and round each entry of K to the nearest multiple of h, giving K_q. Then

$$\|K_q - K\|_F \le \frac{h\sqrt{\ell d}}{2} \le \frac{\eta}{4}, \qquad \|K_q^\top K_q - K^\top K\|_{\mathrm{op}} \le \frac{\eta}{2} + \frac{\eta^2}{16} \le \eta.$$

Store the integer $j = \lceil \delta/\eta \rceil + 1$ and set $\delta_q = j\eta$. Since $\eta \le \delta_q - \delta < 2\eta$, the error

$$R = K_q^\top K_q + \delta_q I - (K^\top K + \delta I) \quad \text{satisfies} \quad 0 \preceq R \preceq 3\eta I. \tag{123}$$

This also applies when $\ell = d$ and $\delta = 0$.

Choose a public dyadic coordinate step at most $\sqrt{\varepsilon}/(50\sqrt{d})$, within a factor two of that value, and round θ coordinatewise to obtain θ_q. Then

$$\|\theta_q - \theta\|_2 \le \sqrt{\varepsilon}/100. \tag{124}$$

The actual stored state is $(n, \ell, K_q, j, \theta_q)$, and the decoder computes the query moments and returns

$$\begin{aligned} H &= K_q^\top K_q + \delta_q I - C + \beta\lambda I, \\ \widehat{w}_U &= \theta_q + H^{-1}\big[(C + \alpha\lambda I)\theta_q - f\big]. \end{aligned} \tag{125}$$

All matrices and vectors on the right are available from the summary, public parameters, and deleted records. Equation (123) guarantees $H \succeq \beta\lambda I$.

Replacing θ by θ_q in (120) changes its output by

$$[I + H^{-1}(C + \alpha\lambda I)](\theta_q - \theta).$$

The operator norm of the bracket is at most

$$1 + \frac{\alpha(1+\lambda)}{\beta\lambda} = 1 + \frac{5\alpha}{\beta} \le \frac{12}{7} < 2.$$

The retained curvature satisfies $\|Q/\beta\|_{\mathrm{op}} \le 1 + \lambda = 5/4$. Thus reference rounding adds at most $3\|\theta_q - \theta\|_2$ to the square root of the excess. Combining this with (121) gives

$$\sqrt{F_T(\widehat{w}_U) - F_T(w_T)} \le \left(\frac{1}{4} + \frac{9}{100} + \frac{3}{100}\right)\sqrt{\varepsilon} < \sqrt{\varepsilon}.$$

The guarantee is deterministic for every legal query.

Finally, every entry of K and θ has magnitude at most one. Their quantized entries consequently take $O(\log(2 + d/\varepsilon))$ bits each: the reciprocal factor step is $O(\sqrt{\ell d}/\sqrt{\varepsilon})$, and the reciprocal reference step is $O(\sqrt{d}/\sqrt{\varepsilon})$. The integer j takes $O(\log(2 + 1/\varepsilon))$ bits. Counts and any encoding metadata take $O(\log(2 + d) + \log(2 + n))$ bits; the quantization steps are public functions of the parameters and ℓ. These bounds charge every persistent entry and prove (114). □

Corollary 18 (Upper bound on a finite input grid)**.** *Under the hypotheses of Theorem 17, suppose in addition that $0 < \varepsilon \le \rho^2$, $n = \Theta(d/\rho)$, and all feature and label coordinates belong to a public dyadic grid with $p_{\mathrm{in}} = O(\log(2 + d))$ fractional bits. There is a deterministic summary of*

$$O\left(d \min\left\{d, \frac{\rho}{\sqrt{\varepsilon}}, \left(\frac{\rho}{\varepsilon}\right)^{1/3}\right\} \log(2 + d/\rho)\right) \tag{126}$$

bits with the same deletion guarantee and access model.

Proof. If the chosen rank is d, store G_S, b_S, n exactly and use the exact ridge decoder. An entry of either moment has denominator dividing $2^{2p_{\mathrm{in}}}$ and magnitude at most n, so this branch costs $O(d^2[p_{\mathrm{in}} + \log(2 + n)])$ bits, independently of ε. If $\ell < d$, then $128R_{\rho,\varepsilon} < d$. At least one of

$$\varepsilon > (128\rho/d)^2, \qquad \varepsilon > 128^3\rho/d^3$$

therefore holds. In either case $\log(2 + 1/\varepsilon) = O(\log(2 + d/\rho))$, so Theorem 17 has the stated precision cost. The condition $\varepsilon \le \rho^2$ ensures $R_{\rho,\varepsilon} \ge 1$, absorbing the extra reference vector and the rounding of the rank into the asserted order. The two branches can be distinguished by one stored bit. □

F.1 A LOWER BOUND FROM COMMON SPARSE DELETION PACKETS

Theorem 19 (A matching lower bound as the deletion fraction vanishes)**.** *There are absolute constants $c, c_0 > 0$ and $d_0 \geq 17$ such that, for $d \geq d_0$, $\lambda = 1/4$, $0 < \rho \leq 1/8$, and $0 < \varepsilon \leq c_0\rho^2$, any summary satisfying the pointwise randomized deletion guarantee requires*

$$B \geq cd \min\left\{d, \frac{\rho}{\sqrt{\varepsilon}}, \left(\frac{\rho}{\varepsilon}\right)^{1/3}\right\} \tag{127}$$

bits on a finite family with one public sample size $n = \Theta(d/\rho)$. The family has the following properties. Every feature has norm at most one, every label belongs to $\{-1/2, 1/2\}$, and every record satisfies $y = x^\top e_0$ for the same public unit teacher. All coordinates lie on the fixed input grid in (4)*. Original and queried retained datasets are centered and have label variance exactly $1/4$. Every original ridge optimizer is the same public vector $w_S = e_0/2$. The requests are common legal deletions, each of fraction between $\rho/64$ and $\rho/32$. All persistent data-dependent bits are charged, and the decoder has only the state and the deleted records. The probability of success need only be $2/3$ for each fixed dataset and request.*

Proof. We first give the construction and its information bound for an arbitrary power of four $r \leq (d-1)/4$. Let p be the largest power of four at most $d-1-r$. Then $p \geq r$ and $p = \Theta(d)$. Use one signal coordinate, r probe coordinates, and p output coordinates, padding unused coordinates with zero. Let J embed the probe coordinates into the nuisance space $\mathbb{R}^{r+p}$, and put $P = JJ^\top$. The source matrix A ranges over $\Gamma_{p,r}$ from Lemma 12, so

$$A_{ij} \in \{-1/\sqrt{p}, 1/\sqrt{p}\}, \qquad \|A\|_{\mathrm{op}} \leq 8 \leq C_0 := 16, \qquad |\Gamma_{p,r}| \geq 2^{pr-1}. \tag{128}$$

The more generous constant C_0 simplifies estimates below.

Integer sample counts and common packets. Define the public integers and masses

$$\begin{aligned} &L = \lceil 16/\rho \rceil, \qquad N = \max\{L, 32r\}, \qquad k = \lfloor N/r \rfloor - 1, \\ &\omega = \frac{1}{2N}, \qquad q = \lfloor \rho N/16 \rfloor, \qquad \alpha = q\omega, \qquad \beta = 1 - \alpha, \\ &M = \left\lceil \frac{d}{\rho N} \right\rceil, \qquad n = 4NM. \end{aligned} \tag{129}$$

If $N = L$, then $q = 1$; if $N = 32r$, then $q = \lfloor 2\rho r \rfloor$. Hence $1 \leq q \leq r$ in both cases. Since $\rho N/16 \geq 1$, the inequality $\lfloor t \rfloor \geq t/2$ for $t \geq 1$ gives

$$\frac{\rho}{64} \leq \alpha \leq \frac{\rho}{32}, \qquad \omega, \alpha \leq \frac{1}{256}, \qquad \beta \geq \frac{255}{256}. \tag{130}$$

Also $N \geq 32r$, so the total hidden mass

$$q_h := \frac{kr}{N} \quad \text{satisfies} \quad \frac{15}{16} \leq q_h \leq 1. \tag{131}$$

Indeed, $kr > N - 2r$ and $kr \leq N - r$.

Set $a = b = v = 1/2$ and $\tau = 2^{-10}$. For a nuisance vector z, let the effective packet $\mathcal{P}(z)$ consist of the two actual records

$$((a, z), a), \qquad ((-a, -z), -a).$$

For each $j \leq r$, include M copies of each public packet $\mathcal{P}(ve_j)$ and $\mathcal{P}(-ve_j)$, and kM copies of each hidden packet $\mathcal{P}(z_j)$ and $\mathcal{P}(-z_j)$, where

$$z_j = b \begin{pmatrix} e_j \\ \tau A_{:j} \end{pmatrix}.$$

Finally include $2M[N - r(1+k)]$ copies of $\mathcal{P}(0)$. The bracket is a nonnegative integer smaller than r, and these counts give exactly $4NM$ actual records. Every hidden column has squared nuisance norm $b^2(1+\tau^2)$, so all actual feature norms are strictly less than one. Mirror pairs give zero feature and label means, exact labels for teacher e_0, and label variance $1/4$.

Write $V = (I_r, \tau A^\top)^\top$. The normalized original moments are

$$\Sigma_S = \mathrm{diag}(a^2, H_A), \qquad \frac{b_S}{n} = a^2 e_0, \qquad H_A = 2\omega v^2 P + \frac{q_h b^2}{r} V V^\top. \tag{132}$$

Since $a^2 = \lambda = 1/4$, the original optimizer is $e_0/2$, independently of A.

A query is a vector $z \in \{0, -1, 1\}^r$ with exactly q nonzero entries. For its support I, delete M copies of $\mathcal{P}(v z_j e_j)$ for every $j \in I$. These are common public packets, and deletion removes exactly $2Mq = \alpha n$ records. Retained records still occur in mirror pairs and still have labels of magnitude $1/2$.

A uniform approximation to the matrix response. Let P_I be the coordinate projector onto I in $\mathbb{R}^r$, and define

$$K_A = H_A + \beta\lambda I, \qquad K_{A,I} = K_A - \omega v^2 J P_I J^\top, \qquad t_z = \omega v J z.$$

Both normal matrices are at least $\beta\lambda I$: the deletion removes only one of the two common anchor packets in a selected coordinate. If the retained optimizer is $w_T = (u, h_T)$, its normal equations give

$$(a^2 + \lambda)u - \frac{a}{\beta} t_z^\top h_T = a^2, \qquad K_{A,I} h_T = a(u-1) t_z. \tag{133}$$

Put $s_z = t_z^\top K_{A,I}^{-1} t_z$. Then

$$a_z := -2(u-1) = \frac{1}{1 - s_z/(2\beta)}, \qquad 0 \le s_z \le \frac{\omega\alpha}{\beta}, \qquad 0 \le a_z - 1 \le \frac{1}{130049}. \tag{134}$$

The middle inequality uses $\|z\|_2^2 = q$ and $v^2 = \lambda$. For the last, $s_z/(2\beta) \le 1/(2 \cdot 255^2)$ by (130).

For the baseline matrix K_A, introduce public coefficients

$$\mu = \beta\lambda r, \qquad c_* = 2\omega v^2 r, \qquad h_* = q_h b^2, \qquad D = \mu + c_* + h_*.$$

Then $r K_A = \mu I + c_* P + h_* V V^\top$. Block elimination, as in (424), yields

$$\begin{aligned} \Pi_{\mathrm{out}} K_A^{-1} J &= -r c_\mu f(A), \qquad c_\mu = \frac{h_* \tau}{D\mu}, \\ f(A) &= A(I + \kappa A^\top A)^{-1}, \qquad \kappa = \frac{(\mu + c_*) h_* \tau^2}{D\mu}. \end{aligned} \tag{135}$$

Here Π_{out} selects the p output coordinates. The bounds

$$\mu \ge \frac{255 r}{1024}, \qquad h_* \le \frac{1}{4}, \qquad 0 \le \kappa \le 2\tau^2$$

imply

$$\|f(A)\|_{\mathrm{op}} \le C_0, \qquad \|f(A) - A\|_{\mathrm{op}} \le \kappa C_0^3 \le \frac{1}{128}. \tag{136}$$

The selected coordinates change the baseline inverse by a controlled factor. Define

$$T_A = J^\top K_A^{-1} J, \qquad L_{A,I} = (I - \omega v^2 P_I T_A)^{-1}.$$

Since $\|T_A\|_{\mathrm{op}} \le 1/(\beta\lambda)$,

$$\|\omega v^2 P_I T_A\|_{\mathrm{op}} \le \frac{1}{255}, \qquad \|L_{A,I}\|_{\mathrm{op}} \le \frac{255}{254}, \qquad \|L_{A,I} - I\|_{\mathrm{op}} \le \frac{1}{254}. \tag{137}$$

Direct multiplication proves $K_{A,I}^{-1} J = K_A^{-1} J L_{A,I}$; this identity does not assume commutativity. Combining it with (133)–(135) gives the exact output response

$$\Pi_{\mathrm{out}} h_T = \gamma a_z f(A) L_{A,I} z, \qquad \gamma = \frac{\omega r c_\mu}{8} > 0. \tag{138}$$

Every factor in γ is public. By (134), (136), and (137), for every source and query,

$$\begin{aligned} \left\| \frac{\Pi_{\mathrm{out}} h_T}{\gamma} - Az \right\|_2 &\le \left(\frac{16 \cdot 255}{130049 \cdot 254} + \frac{16}{254} + \frac{1}{128} \right) \sqrt{q} \\ &< \frac{\sqrt{q}}{8}. \end{aligned} \tag{139}$$

In particular, no estimate of the scalar coordinate u is needed to read out the hidden matrix approximately.

The two signal regimes. We have $c_* \le 1/128$, $h_* \ge 15/64$, and $\mu \le r/4$. Thus $D \le 65r/128$ and

$$c_\mu \ge \frac{24\tau}{13r^2} \ge \frac{\tau}{r^2}, \qquad \gamma \ge \frac{\tau\omega}{8r}.$$

If $N = L$, then $q = 1$ and $\omega \ge \rho/33$. If $N = 32r$, then $\omega = 1/(64r)$ and $q \ge \rho r$, since $2\rho r \ge 1$. Consequently, in both cases,

$$\omega^2 q \ge \frac{1}{4096}\min\{\rho^2, \rho/r\}, \qquad \lambda\gamma^2 q \ge 2^{-40}\min\{\rho^2/r^2, \rho/r^3\}. \tag{140}$$

The single-packet queries give the first regime. When r is larger, each packet has mass proportional to $1/r$, and deleting q packets with q proportional to ρr gives the second regime.

Pointwise success forces information. Assume for now that

$$\varepsilon \le \frac{\lambda\gamma^2 q}{1024}. \tag{141}$$

Run 225 independent copies of the encoder, using independent public tapes and private coins, and retain their states. For each fixed query, run their decoders independently and use the majority medoid in the Euclidean metric with good-output radius $\sqrt{\varepsilon/\lambda}$. The true retained curvature is at least λI, so every successful output lies within this radius of the optimum. A majority of good outputs yields a selected output within three times this radius. Hoeffding's inequality bounds failure by $\exp(-225/18) < 2^{-18}$. This is a pointwise guarantee for each fixed source and query; it makes no assertion about simultaneous success.

Divide the selected output block by γ and project it onto the Euclidean ball of radius $C_0\sqrt{q}$, obtaining Y. This ball contains Az for every query. On the good event, projection and (139) give

$$\|Y - Az\|_2^2 \le \frac{18\varepsilon}{\lambda\gamma^2} + \frac{q}{32}.$$

Projection bounds the squared error on failure by $4C_0^2 q$. Therefore (141) implies, for each fixed A, z,

$$\mathbb{E}\|Y - Az\|_2^2 \le \left(\frac{18}{1024} + \frac{1}{32} + \frac{1024}{2^{18}}\right) q < \frac{q}{8}. \tag{142}$$

Choose the source uniformly from $\Gamma_{p,r}$, and choose a query independently by taking a uniform q-element subset and independent uniform signs on that subset. Then

$$\mathbb{E}[zz^\top] = \frac{q}{r} I_r.$$

Let M be the concatenated amplified state and R its independent public tapes. Fresh decoder coins are independent of the source conditional on $(\mathsf{M}, \mathsf{R}, z)$. Thus the conditional mean $\overline{A}z$, where $\overline{A} = \mathbb{E}[A \mid \mathsf{M}, \mathsf{R}]$, is the mean-square optimal predictor of Az from those variables. Averaging (142) gives

$$\frac{q}{r}\mathbb{E}\|A - \overline{A}\|_F^2 = \mathbb{E}\|(A - \overline{A})z\|_2^2 \le \mathbb{E}\|Y - Az\|_2^2 < \frac{q}{8}. \tag{143}$$

Predicting each sign of A by the sign of the corresponding entry of $\overline{A}$ therefore makes fewer than $pr/8$ errors in expectation: an incorrect sign costs at least $1/p$ in squared error. The entropy argument of Lemma 13, with $1/8$ in place of $1/4$, now yields

$$225B \ge [1 - h_2(1/8)]pr - 1 = \Omega(pr). \tag{144}$$

Every retained parameter and private seed is part of the charged state. If variable-length strings of at most B bits are used, append a terminal one and pad each amplified copy to $B+1$ bits; the resulting additive constant does not change the lower bound.

Rank choice, sample size, and the fixed grid. By (140), condition (141) follows from

$$\varepsilon \le 2^{-50}\min\{\rho^2/r^2, \rho/r^3\}.$$

Choose r as the largest power of four not exceeding

$$\min\left\{\frac{d-1}{4}, c_1\frac{\rho}{\sqrt{\varepsilon}}, c_1\left(\frac{\rho}{\varepsilon}\right)^{1/3}\right\}, \tag{145}$$

where the absolute constant $c_1 > 0$ is sufficiently small that $c_1^2, c_1^3 \le 2^{-50}$. For sufficiently small absolute c_0, the assumption $\varepsilon \le c_0\rho^2$ and sufficiently large d ensure that the displayed minimum is at least one. Taking the largest power of four loses at most a factor four, so $pr = \Omega(d \min\{d, \rho/\sqrt{\varepsilon}, (\rho/\varepsilon)^{1/3}\})$. This proves (127).

For $d \ge 17$, the sample count in (129) obeys

$$\frac{\rho N}{d} \le \max\left\{\frac{16+\rho}{d}, 8\rho\right\} \le 1, \qquad \frac{4d}{\rho} \le n < \frac{8d}{\rho}.$$

The first bound uses $r \le (d-1)/4$; the second follows from $n = 4N\lceil d/(\rho N)\rceil < 4d/\rho + 4N$. It is public and independent of the source matrix. All hidden coordinates are dyadic: p is a power of four, and their denominators divide $2^{11}\sqrt{p}$. Thus the fixed grid in (4) contains every feature and label exactly. There is no input rounding and no input-precision dependence on ρ or ε. The common original model, common requests, and mirror pairing are preserved by the integer duplication. The information argument uses only the charged state and common query, so applies to an arbitrary offline encoder under the stated access and randomness model. □

Corollary 20 (The storage law at fixed regularization)**.** *Fix $\lambda = 1/4$ and the grid in* (4)*. Let the permitted public sample sizes range over $4d/\rho \le n \le 8d/\rho$. For $d \ge d_0$, $0 < \rho \le 1/8$, and $0 < \varepsilon \le c_0\rho^2$, the least worst-case state length under the pointwise randomized guarantee satisfies*

$$cdR_* \le B_\rho^*(d,\varepsilon) \le CdR_* \log(2 + d/\rho),$$
$$R_* = \min\left\{d, \frac{\rho}{\sqrt{\varepsilon}}, \left(\frac{\rho}{\varepsilon}\right)^{1/3}\right\}. \tag{146}$$

The same bounds hold if the data are promised to have a public unit noiseless teacher, or merely to admit a teacher of norm at most one. The upper bound is deterministic and also holds without a noiseless promise. The lower bound can additionally require a common public original optimizer and common legal requests.

Proof. Combine Corollary 18 with Theorem 19. The latter family has the public unit teacher e_0, so belongs to each stated noiseless class; the former upper bound holds for all bounded datasets. Its precision factor includes $\log(1/\rho)$, as required when the permitted sample count grows with $1/\rho$. □

Corollary 21 (The teacher's covariance iterates can all be public)**.** *On the hard family in Theorem 19, for every integer $j \ge 0$,*

$$\Sigma_S^j e_0 = 4^{-j} e_0.$$

Thus the entire sequence of covariance iterates of the public teacher, as well as the original optimizer, is identical across the family. Its teacher Krylov space has dimension one, while deletion still requires the state length in (127)*. These iterates alone carry no information about the hidden source.*

Proof. Equation (132) gives $\Sigma_S e_0 = e_0/4$. Iterate this identity and apply Theorem 19. □

Before the dimension cap is reached, the two ranks cross at $\varepsilon = \rho^4$: the rank is proportional to $\rho/\sqrt{\varepsilon}$ above that scale and to $(\rho/\varepsilon)^{1/3}$ below it. This improves the previously available rare-anchor lower bound at fixed regularization and settles the vanishing-fraction question left open by Proposition 117. That proposition remains consistent with (146): its parameters $d = \Theta(\rho^{-2})$ and $\varepsilon = \Theta(\rho^7)$ already place the new rank at the dimension cap.

F.2 Partial bounds when regularization also varies

The matching result above fixes $\lambda = 1/4$. The same analysis gives a partial upper bound when both λ and ρ vary, under a bounded-teacher promise. We state the bound separately because neither its optimality nor the corresponding joint parameter law is established.

Proposition 22 (A joint-parameter upper bound). *Suppose* $0 < \lambda \le 1/4$, $0 < \rho \le 1/8$, $0 < \varepsilon \le 1$, *and the bounded records satisfy* $y = x^\top v$ *for some* $\|v\|_2 \le 1$. *Put*

$$R_{\lambda,\rho,\varepsilon} = \min\left\{\frac{\rho}{\lambda^{3/2}\sqrt{\varepsilon}}, \left(\frac{\min\{\rho,\lambda\}}{\lambda^3\varepsilon}\right)^{1/3}\right\}, \qquad \ell = \min\{d, \lceil 128 R_{\lambda,\rho,\varepsilon}\rceil\}.$$

There is a deterministic summary using

$$O(d(\ell+1)\log(2+d/(\lambda\varepsilon)) + \log(2+n)) \tag{147}$$

bits that guarantees retained excess at most ε *for every legal deletion of at most* ρn *records. The access model is that of Theorem 17; in particular the teacher need not be known or stored, and the inputs need not lie on a finite grid.*

Proof. Use the normalized notation, rank-ℓ floor, and projector from the proof of Theorem 17. For a noiseless dataset A, both w_A and $v - w_A$ have norm at most one: their spectral multipliers are respectively $s/(s+\lambda)$ and $\lambda/(s+\lambda)$. Write $u_T = v - w_T$. Identity (117) becomes

$$(\Sigma + \beta\lambda I)\Delta = \alpha\lambda\theta - C u_T.$$

Since $\|\theta\|_2, \|u_T\|_2 \le 1$, it implies

$$\|\Delta\|_2 \le \frac{2\alpha}{\beta\lambda}.$$

Also $\|\Pi\theta\|_2 \le \delta/\lambda$, because $\theta = \Sigma(\Sigma+\lambda I)^{-1}v$. The domination $\Pi C\Pi \preceq \delta\Pi$ gives

$$\|\Pi C u_T\|_2^2 \le \delta u_T^\top C u_T \le \alpha\delta.$$

Consequently, using $\alpha\delta \le 1$,

$$\|\Pi\Delta\|_2 \le \frac{\alpha\delta + \sqrt{\alpha\delta}}{\beta\lambda} \le \frac{2\sqrt{\alpha\delta}}{\beta\lambda}.$$

The noncommutative perturbation argument in (121), before rounding, therefore bounds the centered-floor excess by

$$\frac{4}{\beta^4}\min\left\{\frac{\alpha^2\delta^2}{\lambda^3}, \frac{\alpha\delta^3}{\lambda^3}\right\}. \tag{148}$$

The uncentered floor, by the source-condition bound (50), has excess at most

$$\frac{\delta^3}{4\beta^3\lambda^2}. \tag{149}$$

Thus the public parameters select between three rank requirements:

$$\frac{\rho}{\lambda^{3/2}\sqrt{\varepsilon}}, \qquad \left(\frac{\rho}{\lambda^3\varepsilon}\right)^{1/3}, \qquad \left(\frac{1}{\lambda^2\varepsilon}\right)^{1/3}.$$

Choose the centered decoder if either of the first two attains their minimum, and the uncentered decoder otherwise, with a fixed public rule for ties. Their minimum is $R_{\lambda,\rho,\varepsilon}$. Since $\beta \ge 7/8$ and $\delta \le 1/(\ell+1)$ when $\ell < d$, the selected ideal decoder has excess square root at most $\sqrt{\varepsilon}/4$. If $\ell = d$, its ideal error is zero.

For a finite representation, repeat the factor and upward-scalar rounding in (123), now with a public dyadic $\eta \in (\sqrt{\lambda\varepsilon}/200, \sqrt{\lambda\varepsilon}/100]$. This gives $\widehat{\Sigma} = \widehat{\Sigma}_0 + R$ with $0 \preceq R \preceq 3\eta I$. For the centered decoder with exact reference, $\|\Delta\|_2 \le 2$ bounds the additional excess square root by

$$\frac{6\eta}{\beta\sqrt{\lambda}} \le \frac{7\eta}{\sqrt{\lambda}}.$$

Round its stored reference to satisfy $\|\theta_q - \theta\|_2 \le \lambda\sqrt{\varepsilon}/100$. In (125), this error is multiplied by $I + H^{-1}(C + \alpha\lambda I)$, whose operator norm is at most

$$1 + \frac{\alpha(1+\lambda)}{\beta\lambda} \le 1 + \frac{5}{28\lambda} \le \frac{1}{\lambda}.$$

Since the retained curvature has operator norm at most $5/4$, reference rounding adds at most $2\|\theta_q - \theta\|_2/\lambda$ to the excess square root. The total is less than $\sqrt{\varepsilon}$.

For the uncentered branch store a quantized normalized response b_q instead, with $\|b_q - b\|_2 \le \sqrt{\lambda\varepsilon}/100$, and return

$$[\widehat{\Sigma} - C + \beta\lambda I]^{-1}(b_q - f).$$

The covariance rounding contributes at most $3\eta/(\beta\sqrt{\lambda})$, using $\|w_T\|_2 \le 1$. The response rounding contributes at most $\|b_q - b\|_2/(\beta\sqrt{\lambda})$ by the same quadratic perturbation bound. Again the total excess square root is less than $\sqrt{\varepsilon}$.

Both branches store a rounded $\ell \times d$ factor, the upward-floor integer, one d-vector, n, and encoding metadata. The branch choice is public, or may be recorded in one bit. Factor steps are of order $\sqrt{\lambda\varepsilon}/\sqrt{\ell d}$; vector coordinate steps are of order $\lambda\sqrt{\varepsilon}/\sqrt{d}$ in the centered branch and $\sqrt{\lambda\varepsilon}/\sqrt{d}$ in the uncentered branch. All exact entries being rounded have magnitude at most one. This proves (147), with every persistent entry charged. □

The next example shows that the factor λ^{-3} in the deletion-sensitive cubic estimate (148) cannot uniformly be replaced by λ^{-2} for these floor decoders. It does not establish that the joint-parameter upper bound is optimal among all summaries.

Proposition 23 (A joint-parameter obstruction for flooring)**.** *For $\lambda = 4^{-j}$ and sufficiently large integers j, set*

$$\alpha = \lambda^2, \qquad t = \lambda^3, \qquad q = \frac{1}{4t}, \qquad d = q + 2, \qquad n = \frac{2}{\alpha t} = \Theta(d/\alpha).$$

There is a centered dataset with a public unit teacher, public records on a dyadic grid of $O(\log d)$ bits per coordinate, and a legal deletion of fraction α, for which every unquantized ordinary or centered spectral floor of rank $1 \le \ell \le d-2$ has excess

$$\left(\frac{1}{256} + o(1)\right)\frac{\alpha t^3}{\lambda^3}. \tag{150}$$

Rank zero has at least this excess, while rank $d-1$ is exact. The original label variance is λ and the retained label variance is $\Theta(\lambda)$. In particular, $\varepsilon = \alpha t^3/(8192\lambda^3) = \lambda^8/8192$ requires rank $d-1$ within these decoder families, although

$$\left(\frac{\alpha}{\lambda^2\varepsilon}\right)^{1/3} = \Theta(\lambda^{-8/3}) = o(d). \tag{151}$$

Proof. Use coordinates e_0, q dummy directions, and e_b, and the teacher e_0. A population of mass p at x means equal mass $p/2$ at the two labeled records (x, x_0) and $(-x, -x_0)$. Include populations of mass α at each of

$$x_+ = \tfrac{1}{2}e_0 + \sqrt{t/(4\alpha)}\,e_b, \qquad x_- = \tfrac{1}{2}e_0 - \sqrt{t/(4\alpha)}\,e_b;$$

a population of mass $\lambda - \alpha/2$ at e_0; mass t at the unit vector in each dummy direction; and the remaining mass $3/4 - \lambda - 3\alpha/2$ at zero. All masses are nonnegative and sum to one. The features and labels are bounded, and mirror pairing gives zero means. Directly,

$$\Sigma = \operatorname{diag}(\lambda, tI_q, t/2), \qquad b = \lambda e_0, \qquad \theta = w_S = e_0/2.$$

Delete the entire x_+ population. This is a public legal request of fraction α. The original and retained label variances are λ and $(\lambda - \alpha/4)/(1-\alpha)$, respectively.

For $1 \le \ell \le q$, the normalized spectral floor is t and its error is $E = (t/2)e_b e_b^\top$. It annihilates θ, so the centered and ordinary answers agree. Put $\beta = 1 - \alpha$ and, on coordinates (e_0, e_b), write the retained normal matrix normalized by n as

$$Q = \begin{pmatrix} A & -c \\ -c & \beta\lambda + t/4 \end{pmatrix}, \qquad A = 2\lambda - \alpha\lambda - \alpha/4, \qquad c = \sqrt{\alpha t}/4.$$

Let $D_0 = \det Q$ and $D_1 = D_0 + At/2$. The noiseless identity $w_T = e_0 - \beta\lambda Q^{-1}e_0$ gives its e_b coordinate $h_T = -\beta\lambda c/D_0$. More generally, for a covariance error $ee_be_b^\top$ with $e \ge 0$,

$$(Q + ee_be_b^\top)^{-1}e_b = \frac{(c, A)^\top}{D_0 + Ae}, \qquad (c, A)Q(c, A)^\top = AD_0.$$

The decoder error is $-eh_T(Q+ee_be_b^\top)^{-1}e_b$, and true retained curvature is Q/β. Its excess is exactly

$$\frac{e^2\beta\lambda^2c^2A}{D_0(D_0+Ae)^2}. \tag{152}$$

Taking $e = t/2$ and using $A \sim 2\lambda$, $D_0 \sim D_1 \sim 2\lambda^2$, $\beta \to 1$, and $c^2 = \alpha t/16$ proves (150). This expression is increasing in e. At rank zero the floor is λ, with $e = \lambda - t/2 \geq t/2$ on e_b; its error still annihilates θ, and the dummy block has zero response and is decoupled. At rank $q+1$ the floor is $t/2$, which makes the approximate covariance exact.

For finite multiplicities, start with $n_0 = 2/t$ records. Each sign of either anchor has α/t copies; each sign of e_0 has $(\lambda - \alpha/2)/t$ copies; and each dummy sign has one copy. All are integers because $\lambda = 4^{-j}$, $\alpha = \lambda^2$, and $t = \lambda^3$. The remaining zero count is also a nonnegative integer. Duplicate every record $1/\alpha$ times, giving $n = 2/(\alpha t)$, and delete every copy of the selected anchor. Normalized statistics and deletion fraction are unchanged. Since $\sqrt{t/(4\alpha)} = \sqrt{\lambda}/2$ is dyadic, all coordinates use $O(\log d)$ bits. Finally, (150) exceeds the stated ε for sufficiently large j, proving the rank requirement and (151). □

Every record in Proposition 23 is public, so the example supplies no information-theoretic memory lower bound. Its vanishing label variance also differs from the constant variance of the hard family at $\lambda = 1/4$. A jointly matching law for varying λ and ρ remains open.

F.3 A SUMMARY USING RESIDUAL SECOND MOMENTS

The preceding obstruction concerns a floor selected from the covariance spectrum alone. A second moment matrix, computed using the original ridge residuals, identifies directions that are sensitive to deletion. Preserving its leading inverse-response directions gives the following stronger upper bound in an explicit joint-parameter regime.

Proposition 24 (A task-dependent upper bound)**.** *Suppose* $0 < \lambda \leq 1/4$, $0 < \rho \leq \lambda^{3/2}$, $0 < \varepsilon \leq 1$, *and the bounded records satisfy* $y = x^\top v$ *for some* $\|v\|_2 \leq 1$. *Put*

$$R_{\mathrm{task}} = \min\left\{\frac{\rho}{\lambda^{3/2}\sqrt{\varepsilon}}, \left(\frac{\rho}{\lambda^2\varepsilon}\right)^{1/3}\right\}, \qquad m = \min\{d, \lceil 128R_{\mathrm{task}}\rceil\}.$$

A deterministic offline summary of

$$O(d(m+1)\log(2+d/(\lambda\varepsilon)) + \log(2+n)) \tag{153}$$

bits guarantees retained excess at most ε *for every legal deletion of at most* ρn *records. Its decoder uses only this state and the deleted records. The teacher need not be known, and the input records may be arbitrary bounded reals.*

Proof. If the first term attains the minimum defining R_{task}, use the centered-floor branch and its finite representation from Proposition 22. Bound (148) already gives the required square rank. We construct the second branch below.

Write $A = \Sigma + \lambda I$, $\theta = A^{-1}b$, and form the residual second moment and its inverse-response second moment

$$\mathcal{R} = \frac{1}{n}\sum_{(x,y)\in S}(y - x^\top\theta)^2xx^\top, \qquad B = A^{-1}\mathcal{R}A^{-1}. \tag{154}$$

Both are computed during encoding. They require no knowledge of the teacher. For the analysis set $u = v - \theta$; then

$$\operatorname{tr}\mathcal{R} \leq \frac{1}{n}\sum_S(x^\top u)^2 = u^\top\Sigma u \leq \frac{\lambda}{4}, \qquad \operatorname{tr}B \leq \frac{1}{4\lambda}. \tag{155}$$

The scalar inequality behind the first bound is $s\lambda^2/(s+\lambda)^2 \leq \lambda/4$ for $s \geq 0$.

Let $\mathcal{S}$ be the span of leading m eigenvectors of Σ, leading m eigenvectors of B, and the single vector $A^{-1}\theta$. Choose an orthonormal basis $V \in \mathbb{R}^{d\times k}$ of this span; then $k \leq \min\{d, 2m+1\}$. Put $P = VV^\top$, $\Pi = I - P$, and let $\delta = \sigma_{m+1}(\Sigma)$ if $m < d$, with $\delta = 0$ if $m = d$. Define

$$\widehat{\Sigma} = \Sigma + E, \qquad E = \Pi(\delta I - \Sigma)\Pi. \tag{156}$$

Because $\mathcal{S}$ contains the leading covariance eigenspace, $\Pi\Sigma\Pi \preceq \delta\Pi$. Thus $0 \preceq E \preceq \delta\Pi$ and $E\mathcal{S} = 0$. This completion admits a short representation. Set $Z = \Sigma V$ and $J = V^\top \Sigma V$. Then

$$\widehat{\Sigma} = \delta I + VZ^\top + ZV^\top - V(J + \delta I)V^\top. \tag{157}$$

The correction to δI need not be positive semidefinite; the full matrix is positive semidefinite and dominates Σ by (156). Its representation uses $O(dk + k^2) = O(d(m+1))$ numerical entries.

Fix a legal query and retain the notation $C, f, \alpha, \beta, Q, \Delta$ from Theorem 17. Its original-residual gradient is

$$g_U = f - C\theta = \frac{1}{n}\sum_U (y - x^\top\theta)x.$$

For every vector z, Cauchy–Schwarz on the deleted records gives $(z^\top g_U)^2 \le \alpha z^\top \mathcal{R} z$. Hence

$$g_U g_U^\top \preceq \alpha\mathcal{R}, \qquad (A^{-1}g_U)(A^{-1}g_U)^\top \preceq \alpha B. \tag{158}$$

Let Π_B project off the selected leading m eigenvectors of B. Since E annihilates that leading subspace,

$$\|EA^{-1}g_U\|_2 \le \delta\|\Pi_B A^{-1} g_U\|_2 \le \delta\sqrt{\frac{\alpha}{4\lambda(m+1)}}. \tag{159}$$

Here the norm squared of the projected vector is the largest eigenvalue of its rank-one outer product, which is at most $\alpha\lambda_{m+1}(B) \le \alpha \operatorname{tr} B/(m+1)$. This argument also holds with a zero left side when $m = d$.

It remains to control the response to the deleted covariance. The exact normal equations imply

$$A\Delta = (C + \alpha\lambda I)\Delta + \alpha\lambda\theta - g_U.$$

Since $EA^{-1}\theta = 0$, multiplication by EA^{-1} gives

$$E\Delta = -EA^{-1}g_U + EA^{-1}C\Delta + \alpha\lambda EA^{-1}\Delta. \tag{160}$$

Let Π_0 project off the leading m covariance eigenvectors. Although E need not commute with Σ, one has $E = E\Pi_0$, and Π_0 commutes with A^{-1}. Legality therefore gives

$$\|EA^{-1}C\Delta\|_2 \le \frac{\delta}{\lambda}\|\Pi_0 C\Delta\|_2 \le \frac{\delta^{3/2}}{\lambda}\sqrt{\Delta^\top C\Delta} \le \frac{2\alpha^{3/2}\delta^{3/2}}{\beta\lambda^2}.$$

The last step uses $\|C\|_{\mathrm{op}} \le \alpha$ and the already proved bound $\|\Delta\|_2 \le 2\alpha/(\beta\lambda)$. Likewise the omitted-subspace bound from the proof of Proposition 22 gives

$$\alpha\lambda\|EA^{-1}\Delta\|_2 \le \alpha\delta\|\Pi_0\Delta\|_2 \le \frac{2\alpha^{3/2}\delta^{3/2}}{\beta\lambda}.$$

Use the centered decoder

$$\overline{w} = \theta + [\widehat{\Sigma} - C + \beta\lambda I]^{-1}[(C + \alpha\lambda I)\theta - f].$$

Since $E \succeq 0$, its excess square root is at most $\|E\Delta\|_2/(\beta\sqrt{\lambda})$. Combining the preceding bounds, $\delta \le 1/(m+1)$, and $\alpha \le \rho \le \lambda^{3/2}$ yields

$$\begin{aligned}\sqrt{F_T(\overline{w}) - F_T(w_T)} &\le \frac{\sqrt{\alpha}}{\lambda(m+1)^{3/2}}\left[\frac{1}{2\beta} + \frac{2(1+\lambda)}{\beta^2}\frac{\alpha}{\lambda^{3/2}}\right] \\ &\le \frac{4\sqrt{\rho}}{\lambda(m+1)^{3/2}}.\end{aligned} \tag{161}$$

The square root is zero if $m = d$. Otherwise the selected cubic rank makes it at most $\sqrt{\varepsilon}/4$.

For completeness, the representation can be quantized without assuming that a rounded basis stays orthonormal. Choose a public dyadic $\eta \in (\sqrt{\lambda\varepsilon}/200, \sqrt{\lambda\varepsilon}/100]$ and a public dyadic step h within a factor two below $\eta/(16d)$. Round every entry of V, Z, J, and δ to that step, preserving the symmetry of J, and call the results V_q, Z_q, J_q, δ_q. Let

$$\widetilde{\Sigma} = \delta_q I + V_q Z_q^\top + Z_q V_q^\top - V_q(J_q + \delta_q I)V_q^\top.$$

All exact matrix factors have operator norm at most one, and $\|J + \delta I\|_{\mathrm{op}} \le 2$. Every factor's operator error is at most $dh/2$, and the scalar error is at most $h/2$. Expanding the two bilinear terms and the term with three factors consequently gives

$$\|\widetilde{\Sigma} - \widehat{\Sigma}\|_{\mathrm{op}} \le 8dh \le \eta/2.$$

For example, writing $\gamma = dh \le 1$, the two bilinear errors total at most $5\gamma/2$, the term with three factors contributes at most $19\gamma/4$, and the scalar term at most $\gamma/2$; their sum is less than 8γ. Thus the decoder uses $\widehat{\Sigma}_q = \widetilde{\Sigma} + \eta I$, for which

$$0 \preceq \widehat{\Sigma}_q - \widehat{\Sigma} \preceq 2\eta I.$$

Round the stored reference so that $\|\theta_q - \theta\|_2 \le \lambda\sqrt{\varepsilon}/100$, and replace $\widehat{\Sigma}, \theta$ by $\widehat{\Sigma}_q, \theta_q$ in the centered decoder. Covariance rounding adds at most $4\eta/(\beta\sqrt{\lambda}) \le 5\eta/\sqrt{\lambda}$ to the excess square root, using $\|\Delta\|_2 \le 2$. Reference rounding adds at most $2\|\theta_q - \theta\|_2/\lambda$, exactly as in Proposition 22. The total is less than $\sqrt{\varepsilon}$.

The state is $(n, k, V_q, Z_q, J_q, \delta_q, \theta_q)$, together with a branch bit if needed. The unrounded entries have magnitude at most one, the reciprocal factor step is $O(d/\sqrt{\lambda\varepsilon})$, and reference coordinate steps are of order $\lambda\sqrt{\varepsilon}/\sqrt{d}$. Including all metadata gives (153). Neither the full matrices $\mathcal{R}, B, \Sigma$ nor their unretained eigenvectors are persistent. □

Corollary 25 (The task-dependent completion resolves the floor example)**.** *On the family in Proposition 23, the exact covariance completion* (156) *with $m = 2$ satisfies $\widehat{\Sigma} = \Sigma$ and has a representation with $k = 3$. Its entries can be stored exactly in $O(d \log d)$ bits. Thus this representation gives exact deletion answers on that family, including the query for which ordinary and centered flooring require rank $d - 1$ at the stated tolerance.*

Proof. There $\theta = e_0/2$, and direct evaluation of (154) gives

$$\mathcal{R} = \mathrm{diag}(\lambda/4 - 3\alpha/32, 0_q, t/32),$$

so the two positive eigenvalues of B lie on e_0, e_b. The leading two covariance eigenvectors can be chosen as e_0 and one dummy coordinate. The vector $A^{-1}\theta$ is a multiple of e_0. Thus $\mathcal{S}$ is the span of e_0, e_b, and that dummy coordinate, while $\delta = t$. On its orthogonal complement the covariance equals tI, making $E = \Pi(tI - \Sigma)\Pi = 0$. Choose the indicated coordinate basis for V. Then V, Z, J, δ and θ are dyadic, with $O(\log d)$ bits per entry, so their exact representation costs $O(d \log d)$ bits. Exact covariance and reference give the exact retained minimizer for every legal deletion. □

This corollary separates two covariance representations on the specified family. Because its records are public, it is not an information-theoretic memory lower bound. Proposition 24 is restricted to $\rho \le \lambda^{3/2}$; it does not establish a matching law over the entire range of jointly varying regularization and deletion fraction.

F.4 A RESPONSE ENVELOPE FOR LARGER DELETION BUDGETS

The preceding summary preserves selected directions of the first inverse response and bounds the remaining feedback. A positive linear operator can instead control all orders of that response before any query is known. This gives a second parameter regime for the same storage rate.

Proposition 26 (An upper bound from a response envelope)**.** *Under the bounded-teacher and access assumptions of Proposition 24, replace its deletion-budget restriction by*

$$0 < \rho \le \frac{\lambda}{2(1+\lambda)}. \tag{162}$$

For $0 < \lambda \le 1/4$ and $0 < \varepsilon \le 1$, its rank $m = \min\{d, \lceil 128 R_{\mathrm{task}} \rceil\}$ and finite-bit bound (153) *still suffice for deterministic excess at most ε on every legal deletion of at most ρn records. All persistent state is charged, and neither the teacher nor retained records are supplied to the decoder.*

Proof. The centered-floor square branch is unchanged. We construct the cubic branch, using $A = \Sigma + \lambda I$ and $\theta = A^{-1}b$ as before. All expectations in this proof are empirical averages over the full original dataset, with normalization $1/n$.

For original residuals $r_i = y_i - x_i^\top \theta$, define

$$q_i = r_i x_i - \lambda\theta, \qquad R_0 = \mathbb{E}[q_i q_i^\top], \qquad B_0 = A^{-1} R_0 A^{-1}.$$

Original stationarity gives $\mathbb{E}[r_i x_i] = \lambda\theta$, so

$$\mathbb{E}q_i = 0, \qquad R_0 = \mathcal{R} - \lambda^2 \theta\theta^\top \succeq 0, \qquad \operatorname{tr} R_0 \le \lambda/4,$$

by (155). For each original record put $D_i = x_i x_i^\top + \lambda I$, and define the positive linear operator on symmetric matrices

$$\mathcal{T}(M) = A^{-1}\mathbb{E}[D_i M D_i] A^{-1}. \tag{163}$$

It is positive because it sends positive semidefinite matrices to positive semidefinite matrices.

We first bound an operator series using the positive functional $\phi(M) = \operatorname{tr}(AM)$. Let $\kappa = (1+\lambda)/\lambda$. Since $0 \preceq D_i \preceq (1+\lambda) I \preceq \kappa A$, one has $D_i A^{-1} D_i \preceq \kappa D_i$. Also $\mathbb{E}D_i = A$. For every $M \succeq 0$ this gives

$$\phi(\mathcal{T}(M)) = \operatorname{tr}\left(M\mathbb{E}[D_i A^{-1} D_i]\right) \le \kappa\phi(M), \qquad \phi(B_0) = \operatorname{tr}(R_0 A^{-1}) \le 1/4. \tag{164}$$

Define the data-dependent matrix

$$B_* = 3\sum_{j=0}^{\infty} (3\rho/2)^j \mathcal{T}^j(B_0). \tag{165}$$

Its terms are positive semidefinite. Hypothesis (162) implies $\rho\kappa \le 1/2$; therefore the weighted traces of its positive terms are summable, so the series converges, and

$$\operatorname{tr} B_* \le \frac{3}{4\lambda} \sum_{j=0}^{\infty} (3\rho\kappa/2)^j \le \frac{3}{\lambda}. \tag{166}$$

It may equivalently be obtained by solving the finite-dimensional linear matrix equation

$$B_* - (3\rho/2)\mathcal{T}(B_*) = 3B_0.$$

Indeed, the convergent series telescopes to that equation. Convergence also holds on arbitrary symmetric inputs by decomposing them into their positive and negative parts and applying (164); hence the corresponding linear operator is invertible. Its computation is permitted temporary work. Neither the operator nor the full matrix B_* is stored in the summary.

We now show that B_* controls every deletion response. Fix a legal query and define, with normalization by the original sample size,

$$D_U = \frac{1}{n}\sum_{i\in U} D_i = C + \alpha\lambda I, \qquad \widetilde{g}_U = \frac{1}{n}\sum_{i\in U} q_i = f - C\theta - \alpha\lambda\theta.$$

The retained normal equations give

$$\Delta = w_T - \theta = -(A - D_U)^{-1}\widetilde{g}_U.$$

Cauchy–Schwarz over the deleted records implies $\widetilde{g}_U \widetilde{g}_U^\top \preceq \alpha R_0$. For every $M \succeq 0$, the same averaging argument gives

$$D_U M D_U \preceq \alpha \frac{1}{n}\sum_{i\in U} D_i M D_i \preceq \alpha\, \mathbb{E}[D_i M D_i]. \tag{167}$$

For example, the first inequality follows by applying convexity of $\|M^{1/2} z\|_2^2$ to the average of the vectors $D_i z$ for each fixed z.

Since $\|A^{-1} D_U\|_{\mathrm{op}} \le \alpha\kappa \le 1/2$, write the convergent Neumann series

$$\Delta = \sum_{j=0}^{\infty} \Delta_j, \qquad \Delta_j = (A^{-1} D_U)^j (-A^{-1}\widetilde{g}_U).$$

The initial outer product is bounded by αB_0. Induction using (167) proves

$$\Delta_j \Delta_j^\top \preceq \alpha^{j+1} \mathcal{T}^j(B_0).$$

Weighted Cauchy–Schwarz with weights $(2/3)^j$, whose sum is three, then yields

$$\Delta\Delta^\top \preceq 3 \sum_{j=0}^{\infty} (3/2)^j \Delta_j \Delta_j^\top \preceq \alpha B_*. \tag{168}$$

All comparisons can first be made for finite partial sums and then passed to their limits. The empty query has $\Delta = 0$ and satisfies the same conclusion. Importantly, B_* depends on the original dataset and the public budget ρ, not on the query.

Let $\mathcal{S}$ span the leading m eigenvectors of Σ and the leading m eigenvectors of B_*. Its dimension is at most $2m$. Use exactly the covariance completion (156) and representation (157) for this subspace. If $m < d$, then $0 \preceq E \preceq \delta\Pi$ with $\delta \le 1/(m+1)$, and E annihilates the leading B_* subspace. The rank-one projection consequence of (168) and (166) gives

$$\|E\Delta\|_2^2 \le \delta^2 \alpha \lambda_{m+1}(B_*) \le \frac{3\alpha}{\lambda(m+1)^3}.$$

For the centered decoder this proves

$$F_T(\overline{w}) - F_T(w_T) \le \frac{3\alpha}{\beta^2 \lambda^2 (m+1)^3}. \tag{169}$$

At $m = d$ the ideal covariance is exact. In either case the chosen cubic rank leaves excess square root at most $\sqrt{\varepsilon}/4$.

Finally, use precisely the finite encoding of $V, Z, J, \delta, \theta, n$ from Proposition 24: entrywise rounding of the covariance representation at step at most $\eta/(16d)$, followed by $+\eta I$, with η of order $\sqrt{\lambda\varepsilon}$; and reference error at most $\lambda\sqrt{\varepsilon}/100$. The proof there depends only on $0 \preceq E \preceq \delta I$, $\|\Delta\|_2 \le 2$, and $\beta \ge 7/8$, all of which still hold. The covariance and reference rounding therefore preserve excess at most ε and the complete bit count (153). No query-dependent or additional unbounded-precision state is retained. □

The two task-dependent upper bounds apply on the union of $\rho \le \lambda^{3/2}$ and $\rho \le \lambda/[2(1+\lambda)]$. Neither condition contains the other throughout $0 < \lambda \le 1/4$. The response-envelope construction is an offline finite-bit existence result with unrestricted temporary computation. The decoder reference implements the simpler first inverse-response summary; a separate small-dimensional diagnostic checks the response envelope.

G A JOINT LOWER BOUND WITH A BOUNDED NOISELESS TEACHER

The packet construction tracks both regularization and the deletion fraction. Scaling its hidden covariance at small λr avoids a separate low-regularization construction. Together with the task-dependent upper bound, it gives a matching law in an explicit joint-parameter regime. For this appendix, use the public dyadic input grid with

$$p_{\mathrm{J}}(d, \lambda) = \left\lceil \log_2 \left(\frac{2^{30} d}{\lambda} \right) \right\rceil \tag{170}$$

fractional bits for every feature and label coordinate. Its precision is $O(\log d + \log(1/\lambda))$, independently of ρ and ε.

Theorem 27 (A joint packet lower bound)**.** *There are universal constants $c, c_0 > 0$ and $d_0 \ge 17$ such that, for $d \ge d_0$, $0 < \lambda \le 1/4$, $0 < \rho \le 1/8$, and*

$$0 < \varepsilon \le c_0 \min\{\rho^2/\lambda, \rho\lambda\}, \tag{171}$$

there is a finite family with one public sample size $4d/\rho \le n < 20d/\rho$ on which every summary satisfying the pointwise randomized deletion guarantee requires

$$B \ge cd \min\left\{ d, \frac{\rho}{\lambda^{3/2}\sqrt{\varepsilon}}, \left(\frac{\rho}{\lambda^2 \varepsilon} \right)^{1/3} \right\} \tag{172}$$

persistent bits. All inputs lie exactly on the grid in (170)*, satisfy* $\|x\|_2 \le 1$, $|y| \le 1/2$*, and have the same public unit teacher* $y = x^\top e_0$*. Every original optimizer is a common public nonzero vector. Original and queried retained datasets are centered and have label variance* $\Theta(\lambda)$*, with universal constants. The requests are common legal deletions of a fraction in* $[\rho/64, \rho/32]$.

The encoder may be an arbitrary randomized offline procedure. Every persistent data-dependent bit is charged; the decoder has only the summary and the deleted records. The success probability need only be $2/3$ *for each fixed dataset and request, including arbitrary independent public randomness and fresh decoder coins.*

Proof. Fix a power of four $r \le (d-1)/4$, to be selected below. Let p be the largest power of four at most $d-1-r$, so $p \ge r$ and $p = \Theta(d)$. Use one signal coordinate, r probe coordinates, and p output coordinates, padding any unused coordinates with zero. Let $J : \mathbb{R}^r \to \mathbb{R}^{r+p}$ be the probe embedding, and put $P = JJ^\top$. The hidden source is $A \in \Gamma_{p,r}$ from Lemma 12. In particular,

$$A_{ij} \in \{-1/\sqrt{p}, 1/\sqrt{p}\}, \qquad \|A\|_{\mathrm{op}} \le 8 \le C_0 := 16, \qquad |\Gamma_{p,r}| \ge 2^{pr-1}.$$

Packet masses and integer counts. Define the public quantities

$$\begin{aligned} &\Lambda = \max\{\lambda, \rho\}, \qquad N = \max\{\lceil 16/\rho \rceil, \lceil 16r/\Lambda \rceil\}, \qquad k = \lfloor N/r \rfloor - 1, \\ &\omega = \frac{1}{2N}, \qquad q = \lfloor \rho N/16 \rfloor, \qquad \alpha = q\omega, \qquad \beta = 1 - \alpha, \\ &t = r\omega, \qquad q_h = \frac{kr}{N}, \qquad M = \left\lceil \frac{d}{\rho N} \right\rceil, \qquad n = 4NM. \end{aligned} \tag{173}$$

Since $\Lambda \ge \rho$ and $r \ge 1$, $N \le \lceil 16r/\rho \rceil$. Together with $N \ge 16/\rho$, this implies $1 \le q \le r$. The same floor estimate used in (130) gives

$$\frac{\rho}{64} \le \alpha \le \frac{\rho}{32}, \qquad \omega, \alpha \le \frac{1}{256}, \qquad \beta \ge \frac{255}{256}, \qquad \frac{31}{32} \le q_h \le 1. \tag{174}$$

For the last inequality, $kr > N - 2r$ and $r/N \le \Lambda/16 \le 1/64$. If

$$T_0 = \min\{\rho r, \Lambda\}, \quad \text{then} \quad \frac{T_0}{34} \le t \le \frac{T_0}{32}. \tag{175}$$

Indeed, with $N_0 = \max\{1/\rho, r/\Lambda\}$ we have $16N_0 \le N \le 17N_0$ and $r/N_0 = T_0$.

Dyadic amplitudes and the common original model. For $s > 0$, let $\mathcal{D}(s)$ be the largest power of two not exceeding s, so $s/2 < \mathcal{D}(s) \le s$. Put

$$\begin{aligned} &\nu = \min\{1, \lambda r\}, \qquad \tau = 2^{-10}, \qquad a_h = \mathcal{D}(\sqrt{\lambda}/2), \qquad b = \mathcal{D}(\sqrt{\nu}/4), \\ &a_a = \mathcal{D}\left(\min\left\{\frac{1}{2}, \sqrt{\frac{\lambda}{8t}}\right\}\right), \qquad v = \mathcal{D}\left(\min\left\{\frac{1}{2}, \sqrt{\frac{\lambda}{1024\omega}}\right\}\right). \end{aligned} \tag{176}$$

The useful bounds are

$$\begin{aligned} \frac{\lambda}{16} \le a_h^2 \le \frac{\lambda}{4}, \qquad \frac{\nu}{64} \le b^2 \le \frac{\nu}{16}, \qquad ta_a^2 \le \frac{\lambda}{8}, \qquad \omega v^2 \le \frac{\lambda}{1024}, \\ a_a^2 \ge \frac{1}{32}\min\{1, \lambda/t\}, \qquad v^2 \ge \frac{1}{4096}\min\{1, \lambda/\omega\}. \end{aligned} \tag{177}$$

For a signal amplitude a and nuisance vector z, define

$$\mathcal{P}_a(z) = \{((a, z), a), ((-a, -z), -a)\}.$$

For every $j \le r$, include M copies of each public packet $\mathcal{P}_{a_a}(ve_j)$ and $\mathcal{P}_{a_a}(-ve_j)$. Also include kM copies of each hidden packet $\mathcal{P}_{a_h}(z_j)$ and $\mathcal{P}_{a_h}(-z_j)$, where

$$z_j = b\begin{pmatrix} e_j \\ \tau A_{:j} \end{pmatrix}.$$

Finally include $2M[N - r(1+k)]$ copies of $\mathcal{P}_{a_h}(0)$. The bracket is an integer in $[0, r)$, and these counts give exactly $4NM$ actual records. Anchor features have squared norm at most $1/2$. A hidden feature has squared norm $a_h^2 + b^2(1+\tau^2) < 1$, since every source column has norm one. Every actual packet is centered and exactly labeled by the public unit teacher e_0.

Writing $V = (I_r, \tau A^\top)^\top$, the normalized original moments are

$$\begin{aligned} \Sigma_S &= \mathrm{diag}(s_0, H_A), \qquad b_S/n = s_0 e_0, \\ s_0 &= (1-2t)a_h^2 + 2ta_a^2, \qquad H_A = 2\omega v^2 P + \frac{q_h b^2}{r} VV^\top. \end{aligned} \tag{178}$$

Equations (174) and (177) give

$$\frac{63\lambda}{1024} \le s_0 \le \frac{\lambda}{2}, \qquad w_S = \frac{s_0}{s_0+\lambda} e_0. \tag{179}$$

All quantities in the signal block are public and independent of A.

For a query $z \in \{0, -1, 1\}^r$ with exactly q nonzero entries, delete M copies of $\mathcal{P}_{a_a}(vz_j e_j)$ for every nonzero coordinate. This is a common legal deletion of $2Mq = \alpha n$ records. Mirror pairing remains intact. Its retained signal second moment, normalized by the original sample size, is

$$s' = s_0 - \alpha a_a^2 = (1-2t)a_h^2 + (2t-\alpha)a_a^2.$$

Since $\alpha \le t$, both s_0 and s'/β lie between positive universal multiples of λ. These are precisely the original and retained label variances.

An inverse response valid at every value of λr. Let I be the support of the query, and let P_I be its coordinate projector in $\mathbb{R}^r$. Define

$$K_A = H_A + \beta\lambda I, \qquad K_{A,I} = K_A - \omega v^2 J P_I J^\top, \qquad t_z = \omega v J z, \qquad D_s = s' + \beta\lambda.$$

Legality gives $K_A, K_{A,I} \succeq \beta\lambda I$. The retained optimizer $w_T = (u, h_T)$ obeys

$$D_s u - a_a t_z^\top h_T = s', \qquad K_{A,I} h_T = a_a(u-1)t_z. \tag{180}$$

Consequently,

$$\begin{aligned} u - 1 &= -\kappa_0 a_z, \qquad \kappa_0 = \frac{\beta\lambda}{D_s} \in [1/2, 1], \qquad a_z = \left(1 - \frac{a_a^2 t_z^\top K_{A,I}^{-1} t_z}{D_s}\right)^{-1}, \\ 0 &\le a_z - 1 \le \frac{1}{8000}. \end{aligned} \tag{181}$$

For the last bound, $D_s \ge \beta\lambda$ and (177) imply

$$\frac{a_a^2 t_z^\top K_{A,I}^{-1} t_z}{D_s} \le \frac{\alpha a_a^2\, \omega v^2}{\beta^2\lambda^2} \le \frac{1}{8192\beta^2}.$$

Thus $a_z - 1 \le 1/(8192\beta^2 - 1) < 1/8000$. The coefficient κ_0 is public because the query size is fixed.

Introduce the public coefficients

$$\mu = \beta\lambda r, \qquad c_* = 2\omega v^2 r, \qquad h_* = q_h b^2, \qquad D = \mu + c_* + h_*.$$

Then $rK_A = \mu I + c_* P + h_* VV^\top$. Block elimination gives

$$\begin{aligned} \Pi_{\mathrm{out}} K_A^{-1} J &= -r c_\mu f(A), \qquad c_\mu = \frac{h_*\tau}{D\mu}, \qquad f(A) = A(I + \kappa A^\top A)^{-1}, \\ \kappa &= \frac{(\mu + c_*)h_*\tau^2}{D\mu}. \end{aligned} \tag{182}$$

Here Π_{out} selects the p output coordinates. Unlike the unscaled hidden covariance, the choice of b makes all these bounds uniform even when λr tends to zero:

$$\begin{aligned} c_* &\le \frac{\mu}{512\beta}, \qquad \frac{31\nu}{2048} \le h_* \le \frac{\nu}{16} \le \frac{\mu}{16\beta}, \qquad D \le \frac{9}{8}\mu, \\ c_\mu &\ge \frac{\tau\nu}{128\lambda^2 r^2}, \qquad 0 \le \kappa \le \frac{\tau^2}{16\beta}, \qquad \|f(A)\|_{\mathrm{op}} \le C_0, \qquad \|f(A) - A\|_{\mathrm{op}} \le \kappa C_0^3 \le \frac{1}{128}. \end{aligned} \tag{183}$$

Set

$$T_A = J^\top K_A^{-1} J, \qquad L_{A,I} = (I - \omega v^2 P_I T_A)^{-1}.$$

The bound $\omega v^2 \le \lambda/1024$ yields

$$\|\omega v^2 P_I T_A\|_{\mathrm{op}} \le \frac{1}{1020}, \qquad \|L_{A,I}\|_{\mathrm{op}} \le \frac{1020}{1019}, \qquad \|L_{A,I} - I\|_{\mathrm{op}} \le \frac{1}{1019}. \tag{184}$$

As before, direct multiplication gives $K_{A,I}^{-1} J = K_A^{-1} J L_{A,I}$. The exact bottom response is therefore

$$\Pi_{\mathrm{out}} h_T = \gamma a_z f(A) L_{A,I} z, \qquad \gamma = a_a \omega v \kappa_0 r c_\mu > 0. \tag{185}$$

Its scale γ is public. Combining (181), (183), and (184) proves, for every source and query,

$$\left\| \frac{\Pi_{\mathrm{out}} h_T}{\gamma} - Az \right\|_2 \le \left(\frac{16 \cdot 1020}{8000 \cdot 1019} + \frac{16}{1019} + \frac{1}{128} \right) \sqrt{q} < \frac{\sqrt{q}}{8}. \tag{186}$$

The signal and its parameter dependence. Equations (181) and (183) give $\gamma \ge \tau a_a \omega v \nu / (256 \lambda^2 r)$. Since $\nu^2/(\lambda^2 r^2) = \min\{1, (\lambda r)^{-2}\} \ge (1 + \lambda r)^{-2}$, it follows that

$$\begin{aligned} \lambda \gamma^2 q &\ge c \frac{a_a^2 v^2 \omega^2 q}{\lambda (1 + \lambda r)^2} \\ &\ge c \frac{\rho t}{\lambda r (1 + \lambda r)^2} \min\{1, \lambda/t\} \min\{1, \lambda r/t\}. \end{aligned} \tag{187}$$

All constants here and below are universal; the fixed τ is absorbed into them. The second step uses $\omega^2 q = \alpha \omega$ and the lower bounds in (174) and (177).

Replacing t on the right by T_0 costs at most a factor 34: $t \ge T_0/34$, while both minimum factors can only decrease when t is increased to T_0. If $\rho \le \lambda$, then $T_0 = \min\{\rho r, \lambda\}$, and

$$\frac{\rho T_0}{\lambda r} \min\{1, \lambda/T_0\} \min\{1, \lambda r/T_0\} = \min\{\rho^2/\lambda, \rho/r\}.$$

If $\rho \ge \lambda$, then $T_0 = \rho$, and the same expression equals $\min\{\rho/r, \lambda\}$. Thus, in every parameter regime,

$$\lambda \gamma^2 q \ge c_s \frac{\min\{\rho^2/\lambda, \rho/r, \lambda\}}{(1 + \lambda r)^2} \tag{188}$$

for a universal $c_s > 0$.

Pointwise randomized answers and information. Suppose

$$\varepsilon \le \lambda \gamma^2 q / 1024. \tag{189}$$

Run 225 independent encoders with independent public tapes and private coins. At any fixed query, the majority-medoid rule in Theorem 19 selects an answer within $3\sqrt{\varepsilon/\lambda}$ of the true optimum except with probability at most 2^{-18}. This uses only the retained curvature bound $Q_T \succeq \lambda I$, so applies unchanged for the present value of λ.

Divide its output block by γ and project onto the ball of radius $C_0\sqrt{q}$, obtaining Y. Equation (186) and the failure projection bound give

$$\mathbb{E}\|Y - Az\|_2^2 \le \frac{18\varepsilon}{\lambda \gamma^2} + \frac{q}{32} + \frac{1024 q}{2^{18}} < \frac{q}{8}.$$

This bound holds for every fixed source and query, averaged over all encoding, public, and decoder randomness. Choose the source uniformly and independently choose a uniform q-element support with independent uniform signs. Then $\mathbb{E}[zz^\top] = (q/r) I$. Writing M, R for the amplified state and public tapes, the conditional mean $\overline{A} = \mathbb{E}[A \mid \mathsf{M}, \mathsf{R}]$ therefore satisfies

$$\frac{q}{r} \mathbb{E}\|A - \overline{A}\|_F^2 \le \mathbb{E}\|Y - Az\|_2^2 < \frac{q}{8}.$$

As in (143), this is the conditional mean-square prediction inequality, using the independence of the query and fresh decoder coins. Sign prediction followed by the entropy argument of Lemma 13 now gives

$$225 B \ge [1 - h_2(1/8)] pr - 1 = \Omega(pr). \tag{190}$$

Length metadata and retained private seeds are included in B. For variable-length strings of at most B bits, padding to $B + 1$ bits changes only an additive constant. No simultaneous-success condition or access to retained records is used.

Choosing the rank in both regularization regimes. Choose r as the largest power of four at most

$$\min\left\{\frac{d-1}{4}, c_1\frac{\rho}{\lambda^{3/2}\sqrt{\varepsilon}}, c_1\left(\frac{\rho}{\lambda^2\varepsilon}\right)^{1/3}\right\}, \tag{191}$$

for a sufficiently small universal $c_1 > 0$. The accuracy assumption gives

$$\frac{\rho}{\lambda^{3/2}\sqrt{\varepsilon}} \geq \frac{1}{\sqrt{c_0}\lambda}, \qquad \left(\frac{\rho}{\lambda^2\varepsilon}\right)^{1/3} \geq \frac{1}{c_0^{1/3}\lambda}.$$

Thus, for small enough c_0 and large enough d_0, the minimum in (191) is at least one.

If $\lambda r \geq 1$, then $\rho/r \leq \rho\lambda < \lambda$ and (188) gives

$$\lambda\gamma^2 q \geq \frac{c_s}{4}\min\left\{\frac{\rho^2}{\lambda^3 r^2}, \frac{\rho}{\lambda^2 r^3}\right\}.$$

Choose c_1 with $c_1^2, c_1^3 \leq c_s/4096$; the rank choice then implies (189). If $\lambda r \leq 1$, then $\rho/r \geq \rho\lambda$ and (188) instead gives

$$\lambda\gamma^2 q \geq \frac{c_s}{4}\min\{\rho^2/\lambda, \rho\lambda\}.$$

Taking $c_0 \leq c_s/4096$ proves the same local accuracy condition. The largest-power choice loses at most a factor four, and $p = \Theta(d)$, so (190) proves (172). This includes the dimension-saturated case $\lambda d \leq 1$.

Sample interval and exact input precision. For $d \geq 17$ and $r \leq (d-1)/4$, the count in (173) satisfies

$$\frac{\rho N}{d} \leq \max\left\{\frac{16+\rho}{d}, \frac{16r\rho}{d\Lambda} + \frac{\rho}{d}\right\} < 4.$$

Consequently $4d/\rho \leq n < 4d/\rho + 4N < 20d/\rho$. This sample size and all multiplicities are public and independent of A. Every amplitude in (176) is a dyadic power. Their reciprocals are at most a universal constant times $1/\sqrt{\lambda}$: this follows from (177), $\nu \geq \lambda$, and $t, \omega \leq 1$. A hidden output coordinate has the form $\pm b\tau/\sqrt{p}$, whose denominator is a power of two at most $2^{13}\sqrt{d/\lambda}$. The simpler signal and probe coordinates also fit on (170). Thus all inputs lie exactly on this single finite grid. No rounding of inputs or dependence of their precision on ρ or ε is needed. □

Corollary 28 (A matching joint law in a small-deletion regime)**.** *Let $B_{\mathrm{J}}^*(d, \lambda, \rho, \varepsilon)$ be the least worst-case state length for bounded datasets on the grid in* (170)*, with any permitted public integer sample size $4d/\rho \leq n \leq 20d/\rho$, and labels $y = x^\top v$ for some $\|v\|_2 \leq 1$. The decoder must succeed with probability at least $2/3$ for each fixed dataset and legal deletion of at most ρn records. Then, for $d \geq d_0$, $0 < \lambda \leq 1/4$, $0 < \rho \leq \max\{\lambda^{3/2}, \lambda/[2(1+\lambda)]\}$, and* (171)*,*

$$\begin{aligned} cdR_{\mathrm{J}} \leq B_{\mathrm{J}}^*(d, \lambda, \rho, \varepsilon) \leq CdR_{\mathrm{J}}\log\left(2 + \frac{d}{\lambda\rho}\right), \\ R_{\mathrm{J}} = \min\left\{d, \frac{\rho}{\lambda^{3/2}\sqrt{\varepsilon}}, \left(\frac{\rho}{\lambda^2\varepsilon}\right)^{1/3}\right\}. \end{aligned} \tag{192}$$

The same bounds hold if the unit teacher is public. The lower bound can additionally require a common public original optimizer and common legal requests; the upper bound is deterministic for every legal request to every admissible dataset.

Proof. The lower bound follows from Theorem 27, whose family has one admissible public sample size and the public teacher e_0. For the upper bound use Proposition 24 when $\rho \leq \lambda^{3/2}$, and Proposition 26 when $\rho \leq \lambda/[2(1+\lambda)]$. Their union is the stated range, and both give the same rank and precision cost. In the rank-d branch, store G_S, b_S, n exactly. An entry has denominator dividing $2^{2p_{\mathrm{J}}}$ and magnitude at most n, so the cost is

$$O(d^2[p_{\mathrm{J}} + \log(2+n)]) = O\left(d^2\log\left(2 + \frac{d}{\lambda\rho}\right)\right).$$

If the proposition's rank is smaller than d, at least one of

$$\varepsilon > \frac{128^2\rho^2}{\lambda^3 d^2}, \qquad \varepsilon > \frac{128^3\rho}{\lambda^2 d^3}$$

holds. Hence its factor $\log(2 + d/(\lambda\varepsilon))$ is $O(\log(2 + d/(\lambda\rho)))$. The accuracy condition ensures $R_{\mathrm{J}} \geq 1$, and the rounded rank and possible exact branch therefore cost $O(dR_{\mathrm{J}})$ numerical entries up to universal constants. The sample count contributes only the displayed precision factor. Every persistent entry and any branch metadata are charged. The upper proposition requires only existence of a bounded teacher, so applies in both stated teacher models. □

The lower theorem allows all $\rho \leq 1/8$. The matching statement in Corollary 28 uses the union of the two proved task-dependent upper-bound regimes.

H A DETERMINANTAL RESPONSE ENVELOPE FOR ARBITRARY DELETION BUDGETS

A determinantal second moment controls deletion responses for arbitrary budgets, without a convergence condition on the positive-operator series. We use the representation of ridge regression as the mean of a minimum-norm fit under an L-ensemble determinantal point process Dereziński et al. (2020); Dereziński & Mahoney (2021). A direct projection identity and a second-moment bound on the change of measure yield a storage guarantee with exponential budget dependence.

Theorem 29 (A universal determinantal upper bound)**.** *Suppose* $0 < \lambda \leq 1/4$, $0 < \rho \leq 1/8$, $0 < \varepsilon \leq 1$, *and the records satisfy* $\|x_i\|_2 \leq 1$ *and* $y_i = x_i^\top v$ *for some* $\|v\|_2 \leq 1$. *Define*

$$\begin{aligned} \Gamma_\rho &= \exp\left(\frac{\rho}{(1-\rho)\lambda}\right) - 1, \\ R_{\mathrm{D}} &= \min\left\{\frac{\rho}{\lambda^{3/2}\sqrt{\varepsilon}}, \left(\frac{\Gamma_\rho}{\lambda\varepsilon}\right)^{1/3}\right\}, \qquad m = \min\{d, \lceil 128 R_{\mathrm{D}}\rceil\}. \end{aligned} \tag{193}$$

A deterministic offline summary of

$$O\left(d(m+1)\log\left(2 + \frac{d}{\lambda\varepsilon}\right) + \log(2+n)\right) \tag{194}$$

bits guarantees retained excess at most ε on every legal deletion of at most ρn records. The decoder receives only the summary and the deleted records. The teacher need not be known, and the records may be arbitrary bounded reals. The encoding computation is unrestricted. A randomized polynomial bit-operation construction for rational inputs in the cubic DPP branch is given separately in Theorem 34.

Proof. Write $X \in \mathbb{R}^{n\times d}$ for the matrix with rows $x_i^\top$, $\Sigma = X^\top X/n$, and $\theta = (\Sigma + \lambda I)^{-1}\Sigma v$. The first branch of (193) uses the centered-floor construction of Proposition 22. Its square-rank bound does not require a relation between ρ and λ. We prove the second branch.

A common random projection and its variance. For an index subset $J \subseteq [n]$, let P_J be the orthogonal projection onto the span of its feature vectors, with $P_\varnothing = 0$. Consider the finite probability distribution

$$p(J) = \frac{\det(X_J X_J^\top/(n\lambda))}{\mathcal{Z}}, \qquad \mathcal{Z} = \det(I + \Sigma/\lambda). \tag{195}$$

The determinant of the empty matrix is one. The principal-minor expansion and Sylvester's determinant identity show that these masses sum to one. Subsets with linearly dependent rows have zero mass.

The projection identity is

$$\mathbb{E}_p P_J = K := \Sigma(\Sigma + \lambda I)^{-1}. \tag{196}$$

To prove it, let M be any symmetric matrix and replace each Gram matrix in the partition sum by $X_J(I + sM)X_J^\top/(n\lambda)$ for sufficiently small real s. For a subset of positive mass, the derivative

of its determinant at $s = 0$ is its original determinant times $\operatorname{tr}(P_J M)$. The determinants for the remaining subsets stay zero, since $I + sM \succ 0$ preserves their rank. The partition sum equals

$$\det\left(I + (I + sM)\Sigma/\lambda\right),$$

whose derivative at zero is $\mathcal{Z}\operatorname{tr}(KM)$. Dividing by $\mathcal{Z}$ and varying M proves (196).

Put $z_J = P_J v$ and define the common covariance

$$\mathcal{V} = \mathbb{E}_p[(z_J - \theta)(z_J - \theta)^\top]. \tag{197}$$

Equation (196) gives $\mathbb{E}_p z_J = Kv = \theta$. Since $P_J^2 = P_J$,

$$\operatorname{tr}\mathcal{V} = v^\top K(I - K)v \le \frac{1}{4}. \tag{198}$$

This uses $0 \preceq K \preceq I$ and $\|v\| \le 1$. The matrix $\mathcal{V}$ is computable from the original records without the teacher: on every subset of positive mass,

$$z_J = X_J^\top (X_J X_J^\top)^{-1} y_J.$$

Thus the finite sum (197) is one permitted encoding procedure. Neither its subsets nor their fitted vectors will be stored.

The deletion response as a change of measure. Fix a legal deletion U and write

$$\alpha = |U|/n, \qquad \beta = 1 - \alpha, \qquad C = \frac{1}{n}\sum_{i \in U} x_i x_i^\top, \qquad Q = \frac{\Sigma - C}{\lambda}.$$

The retained normalized ridge optimizer is

$$w_{S-U} = (\Sigma - C + \beta\lambda I)^{-1}(\Sigma - C)v.$$

Relative to (195), the density

$$f_U(J) = \frac{\mathcal{Z}}{\det(I + Q/\beta)}\,\beta^{-|J|}\mathbf{1}_{\{J \cap U = \varnothing\}} \tag{199}$$

defines the L-ensemble of the retained rows at regularization $n\beta\lambda$. Consequently the same projection identity gives

$$\mathbb{E}_p f_U = 1, \qquad \mathbb{E}_p[f_U z_J] = w_{S-U}, \qquad \Delta_U := w_{S-U} - \theta = \mathbb{E}_p[(f_U - 1)(z_J - \theta)].$$

Its exact density second moment is another partition ratio:

$$\mathbb{E}_p f_U^2 = \frac{\det(I + \Sigma/\lambda)\det(I + Q/\beta^2)}{\det(I + Q/\beta)^2}. \tag{200}$$

We bound its logarithm using

$$\log \mathbb{E}_p f_U^2 = \log\frac{\det(I + \Sigma/\lambda)}{\det(I + Q)} + \sum_j \log\frac{(1 + q_j)(1 + q_j/\beta^2)}{(1 + q_j/\beta)^2},$$

where the q_j are the eigenvalues of Q. For the first term, integration along the positive-semidefinite increment C/λ gives an upper bound $\operatorname{tr} C/\lambda \le \alpha/\lambda$. For each $q \ge 0$, the exact scalar identity

$$\frac{(1 + q)(1 + q/\beta^2)}{(1 + q/\beta)^2} = 1 + \frac{\alpha^2 q}{(\beta + q)^2}$$

and $\log(1 + x) \le x$ give an upper bound $\alpha^2 q/\beta^2$ on the corresponding logarithm. Bounded retained features imply $\operatorname{tr} Q \le \beta/\lambda$. Therefore

$$\log \mathbb{E}_p f_U^2 \le \frac{\alpha}{\lambda} + \frac{\alpha^2}{\beta^2}\operatorname{tr} Q \le \frac{\alpha}{\lambda} + \frac{\alpha^2}{\beta\lambda} = \frac{\alpha}{\beta\lambda}. \tag{201}$$

Cauchy–Schwarz, tested against any vector, now yields

$$\Delta_U \Delta_U^\top \preceq \mathbb{E}_p[(f_U - 1)^2]\,\mathcal{V} \preceq \left(\exp\left(\frac{\alpha}{(1 - \alpha)\lambda}\right) - 1\right)\mathcal{V}. \tag{202}$$

The function $a \mapsto \exp(a/((1 - a)\lambda)) - 1$ is convex on $[0, 1)$ and vanishes at zero. Hence its value at $0 \le \alpha \le \rho$ is at most $\alpha\Gamma_\rho/\rho$. We have obtained the common response envelope

$$\Delta_U \Delta_U^\top \preceq \alpha\mathcal{B}_\rho, \qquad \mathcal{B}_\rho = \frac{\Gamma_\rho}{\rho}\mathcal{V}, \qquad \operatorname{tr}\mathcal{B}_\rho \le \frac{\Gamma_\rho}{4\rho}. \tag{203}$$

The empty query satisfies this with zero response. The covariance $\mathcal{V}$ depends only on the original dataset and λ; only the scalar multiplying it depends on the public budget.

Covariance completion and finite encoding. Let $\mathcal{S}$ span the leading m eigenvectors of Σ and the leading m eigenvectors of $\mathcal{V}$, and let V be an orthonormal basis of this span. Its dimension is at most $2m$. Use the completion (156) and its representation (157). If $m < d$, its error $E = \widehat{\Sigma} - \Sigma$ satisfies $0 \preceq E \preceq \delta I$, with $\delta \le 1/(m+1)$, and annihilates the selected $\mathcal{V}$ eigenspace. Projecting (203) off this eigenspace gives

$$\|E\Delta_U\|^2 \le \delta^2 \alpha \lambda_{m+1}(\mathcal{B}_\rho) \le \frac{\alpha \Gamma_\rho}{4\rho(m+1)^3}.$$

The centered decoder from Proposition 24 therefore obeys

$$F_{S-U}(\overline{w}) - F_{S-U}(w_{S-U}) \le \frac{\Gamma_\rho}{4(1-\rho)^2\lambda(m+1)^3}. \tag{204}$$

The ideal error is zero when $m = d$. Otherwise the cubic choice in (193) makes its square root at most $\sqrt{\varepsilon}/4$.

For the finite encoding, retain exactly the covariance factors, reference, and metadata from Proposition 24. In its notation, round V, Z, J, δ at a public dyadic step at most $\eta/(16d)$, where $\eta \in (\sqrt{\lambda\varepsilon}/200, \sqrt{\lambda\varepsilon}/100]$, and add ηI to the reconstructed covariance. Writing $\widehat{\Sigma}_q$ for the rounded reconstruction, the perturbation relative to the ideal completion satisfies $0 \preceq \widehat{\Sigma}_q - \widehat{\Sigma} \preceq 2\eta I$. Round the reference so that $\|\theta_q - \theta\| \le \lambda\sqrt{\varepsilon}/100$. The same perturbation calculation adds at most $5\eta/\sqrt{\lambda} + 2\|\theta_q - \theta\|/\lambda$ to the excess square root: it uses only $E \succeq 0$, $\|\Delta_U\| \le 2$, and $\beta \ge 7/8$, all valid here. The total excess is at most ε. All exact covariance factors and the reference have bounded entries, so their number and precision give (194). All persistent state is charged. The probability distribution, full covariance $\mathcal{V}$, and subset enumeration are temporary encoding objects, not inputs to the decoder. □

Corollary 30 (An exponential comparison with the joint rank)**.** *Under the sample-size, input-grid, and accuracy assumptions in Corollary 28, allow every $0 < \rho \le 1/8$. There are universal constants $c, C > 0$ such that*

$$\begin{aligned} cdR_{\mathrm{J}} &\le B^*_{\mathrm{J}}(d, \lambda, \rho, \varepsilon) \\ &\le Cd \min\left\{d, \exp\left(\frac{8\rho}{21\lambda}\right) R_{\mathrm{J}}\right\} \log\left(2 + \frac{d}{\lambda\rho}\right). \end{aligned} \tag{205}$$

In particular, for every fixed $M > 0$, the joint rank law is matching up to the displayed logarithm throughout $\rho \le M\lambda$, with the upper constant allowed to depend on M. If $\rho \le M\lambda \log\log(3+d)$, the upper and lower bounds differ by at most a fixed power of the displayed logarithm.

Proof. The lower bound is Theorem 27. For the upper bound, $e^x - 1 \le xe^x$ gives

$$\left(\frac{\Gamma_\rho}{\lambda\varepsilon}\right)^{1/3} \le (1-\rho)^{-1/3} \exp\left(\frac{\rho}{3(1-\rho)\lambda}\right) \left(\frac{\rho}{\lambda^2\varepsilon}\right)^{1/3}.$$

Since $1 - \rho \ge 7/8$, the rank in Theorem 29 is bounded by a universal constant times $\min\{d, \exp(8\rho/(21\lambda))R_{\mathrm{J}}\}$. The accuracy assumption ensures $R_{\mathrm{J}} \ge 1$, absorbing the reference vector and rank rounding.

As in the proof of Corollary 28, use exact storage of the original sufficient statistics when $m = d$. Its cost is $O(d^2 \log(2 + d/(\lambda\rho)))$ on the prescribed input grid. If $m < d$, then at least one of

$$\varepsilon > \frac{128^2\rho^2}{\lambda^3 d^2}, \qquad \varepsilon > \frac{128^3\Gamma_\rho}{\lambda d^3} \ge \frac{128^3\rho}{\lambda^2 d^3}$$

holds. Thus the precision logarithm in (194) is $O(\log(2 + d/(\lambda\rho)))$, proving (205). The two stated budget consequences follow by bounding the exponential factor by a constant or by $\log(3+d)^{8M/21}$, respectively. The upper bound applies whether or not the teacher is public. □

I CONSTRUCTING THE DETERMINANTAL SUMMARY BY SAMPLING

We approximate the covariance in Appendix H with polynomially many independent determinantal samples. Regularized covariance concentration preserves the spectral tail at the required scale. The randomized encoder succeeds simultaneously for every legal deletion of a fixed dataset, without enumerating requests or the determinantal distribution.

Proposition 31 (A sampled determinantal covariance)**.** *Use the distribution p, optimizer θ, and covariance $\mathcal{V}$ from* (195)–(197)*. Let $1 \le m < d$ and $0 < \delta < 1$. Draw independent $J_1, \dots, J_s \sim p$, where*

$$s = \left\lceil 12(m+1)\log\frac{2d}{\delta} \right\rceil, \qquad z_j = X_{J_j}^{\dagger} y_{J_j}, \qquad \widehat{\mathcal{V}} = \frac{1}{s}\sum_{j=1}^{s}(z_j-\theta)(z_j-\theta)^{\top}. \tag{206}$$

The empty subset gives $z_j = 0$. With probability at least $1-\delta$,

$$\mathcal{V} \preceq 2\widehat{\mathcal{V}} + \frac{1}{m+1}I. \tag{207}$$

Moreover $\operatorname{tr}\widehat{\mathcal{V}} \le 1$ deterministically. If Π is the orthogonal projector off the leading m eigenvectors of $\widehat{\mathcal{V}}$, the same event gives

$$\lambda_{\max}(\Pi\mathcal{V}\Pi) \le \frac{3}{m+1}. \tag{208}$$

The procedure uses the original labels and exact original ridge optimizer, without requiring the realizing teacher.

Proof. Write $z = P_J v$ and $u = z - \theta$. Orthogonality of P_J gives

$$\|P_J v - v/2\|^2 = \|v\|^2/4.$$

All z therefore lie on the sphere of radius $\|v\|/2$ centered at $v/2$. Since $\theta = \mathbb{E}z$ lies in its closed ball, $\|u\| \le \|v\| \le 1$. This also proves the deterministic trace bound in the proposition.

Set $r = m+1$, $A = \mathcal{V} + r^{-1}I$, and

$$Y_j = A^{-1/2}(z_j-\theta)(z_j-\theta)^{\top}A^{-1/2}, \qquad M = \mathbb{E}Y_j = A^{-1/2}\mathcal{V}A^{-1/2}.$$

Then $0 \preceq Y_j \preceq rI$ and $0 \preceq M \preceq I$. Consequently

$$\|Y_j - M\|_{\mathrm{op}} \le r, \qquad \mathbb{E}[(Y_j-M)^2] = \mathbb{E}[Y_j^2] - M^2 \preceq rM \preceq rI.$$

The self-adjoint matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1) therefore yields

$$\Pr\left\{ \left\| \sum_{j=1}^{s}(Y_j - M) \right\|_{\mathrm{op}} > s/2 \right\} \le 2d\exp\left(-\frac{(s/2)^2}{2sr + rs/3}\right) = 2d\exp\left(-\frac{3s}{28r}\right) \le \delta.$$

On the complementary event, $-A/2 \preceq \widehat{\mathcal{V}} - \mathcal{V} \preceq A/2$. Its lower inequality rearranges to (207). The trace bound implies $\lambda_{m+1}(\widehat{\mathcal{V}}) \le 1/(m+1)$, and projecting (207) proves (208). The unknown matrix A is used only in the proof, not in the algorithm. □

Corollary 32 (Polynomial arithmetic encoding with simultaneous utility)**.** *Under the hypotheses of Theorem 29, for any $0 < \delta < 1$, its rank and finite-bit storage bounds are attained by a randomized encoder that, for each fixed original dataset, succeeds with probability at least $1-\delta$ simultaneously on every legal query. For the determinantal branch with $m < d$, it uses the number of samples in* (206)*. With exact sampling and the usual arithmetic operation counts for dense spectral linear algebra, a standard determinantal sampler gives a polynomial encoding procedure; one conservative bound is*

$$O\big((s+1)(nd^2 + d^3)\big). \tag{209}$$

The sampling history, sampled fits, and random seed are temporary and are not decoder inputs. The final stored factors and reference are charged exactly as in Theorem 29.

Proof. For the determinantal branch, replace the leading m eigenspace of $\mathcal{V}$ in Appendix H by that of $\widehat{\mathcal{V}}$. Combine it with the leading m eigenspace of Σ, and use the same covariance completion and centered decoder. On the event (207), the covariance floor has size at most $1/(m+1)$ and the omitted response covariance has size at most $3/(m+1)$. Equation (202) then gives, simultaneously for all legal U,

$$F_{S-U}(\overline{w}_U) - F_{S-U}(w_{S-U}) \le \frac{3\Gamma_\rho}{(1-\rho)^2\lambda(m+1)^3}. \tag{210}$$

The event concerns only the original dataset and the encoder's samples. No union bound over queries or simultaneous accuracy of individual random fits is needed. The rank constant 128 in (193) still makes the ideal excess square root at most $\sqrt{\varepsilon}/4$. The same covariance-factor and reference rounding in Theorem 29 therefore yields the same finite-bit utility bound. If the square-rank branch is smaller, use its deterministic encoder; if $m = d$, store the full covariance and reference without sampling.

For the arithmetic statement, the L-ensemble kernel is $XX^\top/(n\lambda)$ and has rank at most $\min\{n, d\}$. Standard exact spectral sampling first independently selects its eigenvectors with the corresponding Bernoulli probabilities, then samples the associated projection DPP. The low-rank factorization can be formed in $O(nd^2 + d^3)$ arithmetic operations. Conditional on a selected rank $q \le d$, the standard optimized projection step costs $O(nq^2)$; these low-rank exact-sampling costs are described by Barthelmé et al. (2023). The factorization is reusable across draws. Each sampled fit costs at most $O(d^3)$, and its contribution to the empirical covariance costs $O(d^2)$. Forming the original moments, computing θ, and the final spectral decompositions cost $O(nd^2 + d^3)$. Summing these bounds gives (209). □

Arithmetic and implementation. The probability statement assumes independent exact DPP draws and exact encoding-time linear algebra before the prescribed final rounding. It establishes polynomial arithmetic complexity and a finite-bit persistent summary. Appendix J additionally proves polynomial bit complexity for rational inputs in the cubic DPP branch, and propagates sampling-law, solve, and spectral errors without an eigengap assumption. Poorly conditioned sampled Gram matrices remain a numerical concern. Appendix Z.1 reports floating-point diagnostics, with no rigorous enclosure of these error tolerances.

J FINITE ARITHMETIC FOR THE DETERMINANTAL CONSTRUCTION

We bound the effects of errors in the sampling law, fitted vectors, covariance formation, and selected subspace. These tolerances specify the enclosures required for a certified implementation. For rational inputs, an exact conditional sampler and rational approximate diagonalization give a separate polynomial bit-operation construction.

Proposition 33 (Perturbed sampling and an omitted-spectrum certificate)**.** *Under the hypotheses of Theorem 29, let $1 \le m < d$, $r = m + 1$, and $0 < \delta < 1$. Let q be a distribution on subsets of $[n]$ satisfying $\|q - p\|_{\mathrm{TV}} \le \zeta$, where total variation is $\sup_A |q(A) - p(A)|$. Draw $s \ge \lceil 12r \log(2d/\delta) \rceil$ independent subsets from q. All error tolerances below are nonnegative. For every subset, including a linearly dependent one, define the exact target $z_J = X_J^\dagger y_J$, with $z_\varnothing = 0$.*

Suppose that on one event of probability at least $1 - \nu$, the computed vectors and reference obey

$$\max_{j \le s} \|\widetilde{z}_j - z_{J_j}\| \le \eta, \qquad \|\widetilde{\theta} - \theta\| \le t, \qquad \left\| M - \frac{1}{s}\sum_{j=1}^{s} (\widetilde{z}_j - \widetilde{\theta})(\widetilde{z}_j - \widetilde{\theta})^\top \right\|_{\mathrm{op}} \le \xi,$$

where M is symmetric. The computed fits and reference may first be projected onto the closed unit ball; this cannot increase the stated errors. Set

$$e = \eta + t, \qquad \ell = \frac{1}{r} + \zeta + 4e + 2e^2 + 2\xi. \tag{211}$$

With probability at least $1 - \delta - \nu$,

$$\mathcal{V} \preceq 2M + \ell I. \tag{212}$$

On this same event, every orthogonal projector Π *with certified bounds*

$$0 \le a \le 1, \quad b \ge 0, \qquad \lambda_{\max}(\Pi\Sigma\Pi) \le a, \qquad \lambda_{\max}(\Pi M \Pi) \le b \tag{213}$$

satisfies

$$\lambda_{\max}(\Pi\mathcal{V}\Pi) \le \tau_V := \frac{1}{r} + \zeta + 2\left(\sqrt{b+\xi} + e\right)^2, \tag{214}$$

as well as the bound $2b + \ell$ *from* (212)*. Use the covariance completion* $\widehat{\Sigma} = \Sigma + \Pi(aI - \Sigma)\Pi$*. Allow a final covariance perturbation* $0 \preceq \widehat{\Sigma}_q - \widehat{\Sigma} \preceq \omega I$*, a stored reference with error at most* t_θ*, and a decoder output error of Euclidean norm at most* t_w *relative to the centered decoder formed from these stored quantities and the deleted records. Then, simultaneously for every legal* U*,*

$$\sqrt{F_{S-U}(\widetilde{w}_U) - F_{S-U}(w_{S-U})} \le \frac{a\sqrt{\Gamma_\rho \tau_V}}{(1-\rho)\sqrt{\lambda}} + \frac{2\omega}{(1-\rho)\sqrt{\lambda}} + \frac{2t_\theta}{\lambda} + \sqrt{1+\lambda}\, t_w. \tag{215}$$

Neither a spectral gap nor closeness to a particular eigenbasis is required.

Proof. Realizability gives $z_J = P_J v$ for every subset, not only those with positive p-mass. Therefore $\|z_J\| \le 1$ and $\|z_J - v/2\| = \|v\|/2$. Since θ belongs to the closed ball with this center and radius, $u_J = z_J - \theta$ satisfies $\|u_J\| \le 1$. Also $\|\theta\| \le 1$, proving the assertion about projection onto the unit ball by nonexpansiveness of Euclidean projection. This clipping leaves the exact targets unchanged.

Put $\mathcal{V}_q = \mathbb{E}_q u_J u_J^\top$. For each unit vector h, the function $(h^\top u_J)^2$ lies in $[0,1]$. The total-variation bound consequently gives $\mathcal{V} \preceq \mathcal{V}_q + \zeta I$. The Bernstein argument in Proposition 31, reapplied with $A_q = \mathcal{V}_q + r^{-1} I$, gives exactly the same constants: it only uses independent draws and $\|u_J\| \le 1$, and does not require $\mathbb{E}_q u_J = 0$. Thus, except on an event of probability δ,

$$\mathcal{V}_q \preceq \frac{2}{s}\sum_{j=1}^{s} u_{J_j} u_{J_j}^\top + \frac{1}{r} I.$$

On the computational accuracy event, $\|\widetilde{z}_j - \widetilde{\theta} - u_{J_j}\| \le e$. Expanding the outer products shows that their operator difference is at most $2e + e^2$. Combining these bounds with the formation error ξ proves (212). A union bound with the one computational event suffices; no independence from that event is used.

For the sharper omitted bound, form matrices with columns $u_{J_j}/\sqrt{s}$ and $(\widetilde{z}_j - \widetilde{\theta})/\sqrt{s}$, denoted T and $\widetilde{T}$. The formation certificate gives $\|\Pi\widetilde{T}\|_{\mathrm{op}} \le \sqrt{b+\xi}$, whereas $\|T - \widetilde{T}\|_{\mathrm{op}} \le \|T - \widetilde{T}\|_{\mathrm{F}} \le e$. The operator triangle inequality and the concentration bound for $TT^\top$ prove (214). Writing $E = \Pi(aI - \Sigma)\Pi$, the first certificate gives $0 \preceq E \preceq a\Pi$. The response envelope (202) gives

$$\|E\Delta_U\|^2 \le a^2 \Gamma_\rho \tau_V$$

for every legal query. The centered-decoder calculation of Proposition 102 bounds its excess square root by $\|E\Delta_U\|/((1-\rho)\sqrt{\lambda})$. The final covariance perturbation adds at most $2\omega/((1-\rho)\sqrt{\lambda})$ since $\|\Delta_U\| \le 2$. Reference error adds at most

$$\sqrt{1+\lambda}\left(1 + \frac{\rho(1+\lambda)}{(1-\rho)\lambda}\right) t_\theta \le \frac{2t_\theta}{\lambda},$$

using the retained curvature lower bound, as in Proposition 24. Finally, the retained ridge Hessian has operator norm at most $1+\lambda$, so output error adds at most $\sqrt{1+\lambda}\, t_w$. The triangle inequality in this Hessian norm proves (215). □

The distributional error enters as a spectral floor ζ, rather than a failure probability multiplied by s. In particular, attained bounds $a, b = O(1/r)$ and tolerances $\zeta, \xi = O(1/r)$, $e = O(1/\sqrt{r})$ preserve the cubic ideal risk by (214). If instead one only has a coupling to an independent exact sample sequence that agrees with the computed subset sequence with probability at least $1 - \kappa$, the same proof gives (212) without the ζ term, at failure probability at most $\delta + \nu + \kappa$. Independent per-draw couplings with mismatch probability at most ζ give $\kappa \le s\zeta$.

The certificates concern the original dataset and encoding computation. They never require the retained records at decoding time. If the decoder reports (215), its upper-bound scalars must be public or rounded upward and included in the stored state. The matrices used to establish the certificates and the sampling history remain temporary. In particular, an unverified invariant-space residual does not supply (213): one needs upper bounds on the omitted spectra and a valid orthogonal projector.

Theorem 34 (A polynomial bit-operation determinantal encoder)**.** *Assume the bounded realizable input records and the public parameters $\lambda, \rho, \varepsilon, \delta$ are rational, with $\lambda, \rho, \varepsilon$ in the ranges of Theorem 29 and $0 < \delta < 1$. Let B bound their individual rational encoding lengths. For any $1 \le m < d$, a randomized encoder using*

$$s = 12(m+1)\left\lceil \log_2 \frac{4d}{\delta} \right\rceil \tag{216}$$

exact DPP samples has polynomial expected bit-operation complexity in n, d, B. With probability at least $1 - \delta/2$, it constructs an ideal covariance completion whose centered decoder satisfies, simultaneously for every legal U,

$$F_{S-U}(\overline{w}_U) - F_{S-U}(w_{S-U}) \le \frac{63\Gamma_\rho}{(1-\rho)^2\lambda(m+1)^3}. \tag{217}$$

The sampling stage can be truncated at a polynomial bit-operation budget with additional failure probability at most $\delta/2$. For $m = d$, or a supplied rank $m < d$ satisfying

$$m + 1 \ge 128\left(\frac{\Gamma_\rho}{\lambda\varepsilon}\right)^{1/3}, \tag{218}$$

final rounding gives an encoder with the storage bound (194)*, polynomial bit-operation complexity, and retained excess at most ε on all legal queries with probability at least $1 - \delta$. Such a rank can be chosen in polynomial bit time with $m = \min\{d, O(1 + (\Gamma_\rho/(\lambda\varepsilon))^{1/3})\}$. The decoder can also produce a dyadic output in polynomial bit time. The teacher is not needed by either procedure.*

Proof. **Rational conditional sampling.** Set $L = XX^\top/(n\lambda)$. Let A contain the indices already required to be selected, and let R contain the undecided indices; excluded indices appear in neither set. The unnormalized mass of this partial decision is

$$W(A, R) = \sum_{T \subseteq R} \det L_{A \cup T} = \det \begin{pmatrix} L_A & L_{A,R} \\ L_{R,A} & I_R + L_R \end{pmatrix}. \tag{219}$$

The equality follows by expanding the determinant in the identity entries indexed by R. For $i \in R$, select i with probability

$$\frac{W(A \cup \{i\}, R \setminus \{i\})}{W(A, R)}, \tag{220}$$

and remove it from R in either case. The complementary numerator is $W(A, R \setminus \{i\})$, and the two sum to the denominator. Along a positive-probability path the denominator is positive. Thus all probabilities are well-defined rationals, and the chain rule proves that the final subset has law p. This is a direct conditional sampling construction; non-spectral sequential DPP samplers are also developed by Launay et al. (2020). No eigenvalue or eigenvector oracle is used here.

All matrices in (219) have order at most n. Clearing the input denominators gives integer matrices with polynomial entry length. Hadamard's bound gives polynomial bit length for every minor. Determinants, their rational ratios, and the linear solves below can therefore be computed in polynomial bit time by integer-preserving elimination (Bareiss, 1968); pivoting handles zero minors. Recomputing the two determinants at each decision is sufficient for this existence result, regardless of possible update optimizations.

For a conditional probability $a/b \in (0, 1)$ in lowest terms, draw $k = \lceil \log_2 b \rceil$ unbiased bits, interpret them as an integer, and reject if that integer is at least b. An accepted integer is uniform on $\{0, \dots, b-1\}$; select the index if it is below a. Each attempt succeeds with probability at least $1/2$, independently of its predecessors. Probabilities zero and one need no random bits. This yields independent exact samples using fresh bits, with fewer than two attempts in expectation per nontrivial decision. Capping each decision at $\lceil \log_2(2ns/\delta) \rceil$ attempts adds probability at most $\delta/2$ of any

abort across the ns decisions. A fixed failure flag can be returned on abort. All other work has a deterministic polynomial bit bound, so this also proves the claimed total time guarantee.

Exact rational fits and a subspace without a gap assumption. The original moments and ridge reference are rational and computable by linear solves. Every nonempty subset of positive p-mass has independent rows, so its fit is exactly $X_J^\top (X_J X_J^\top)^{-1} y_J$. The same determinant bounds give polynomial rational bit length for these fits, even if a sampled Gram matrix is poorly conditioned. Form the exact empirical covariance $\widehat{\mathcal{V}}$ of Proposition 31. Its entries also have polynomial bit length, since s is polynomial in d and B.

Put $r = m + 1$ and $H = \Sigma + \widehat{\mathcal{V}}$. Then $H \succeq 0$ and $\operatorname{tr} H \le 2$. Rational approximate diagonalization (Del Pia, 2026, Theorem 2) computes an exactly orthogonal rational matrix Q in polynomial bit time such that

$$Q^\top H Q = D + N, \qquad D = \operatorname{diag}(Q^\top H Q), \qquad \|N\|_{\mathrm{F}} \le 1/r.$$

Order the diagonal entries of D decreasingly, let V consist of the first m corresponding columns of Q, and put $\Pi = I - VV^\top$. Since the nonnegative diagonal entries of D sum to at most two,

$$\lambda_{\max}(\Pi H \Pi) \le D_{m+1,m+1} + \|N\|_{\mathrm{op}} \le \frac{3}{r}.$$

Both Σ and $\widehat{\mathcal{V}}$ are dominated by H. Consequently $a = \min\{1, 3/r\}$ is a valid covariance floor and $b = 3/r$ is an empirical tail bound. This uses a single selected subspace of dimension m and requires no separation of eigenvalues. Proposition 31, applied with failure probability $\delta/2$, gives a true omitted covariance bound $7/r$. The completion error obeys $0 \preceq E \preceq a\Pi$; its response bound is therefore $a^2(7/r)\Gamma_\rho$, proving (217).

For completeness, rank selection need not compute the exact ceiling of a transcendental number. Put $x = \rho/((1-\rho)\lambda)$. If $x > 3\lceil \log_2(2d) \rceil$, then $\Gamma_\rho > d^3$ and use $m = d$. Otherwise $x = O(\log(2d))$. Set $K = \max\{1, \lceil 2x \rceil\}$, $S_K = \sum_{j=1}^{K} x^j/j!$, and $\overline{\Gamma} = S_K + 2x^{K+1}/(K+1)!$. Successive omitted terms have ratio at most $1/2$, so $\Gamma_\rho \le \overline{\Gamma} \le 2\Gamma_\rho$. These are rational computations with polynomial bit lengths. Choose $m = \min\{d, \lceil 128(\overline{\Gamma}/(\lambda\varepsilon))^{1/3} \rceil\}$ using rational cube comparisons. This proves the rank-selection assertion with only a constant-factor increase in rank.

Persistent rounding and query arithmetic. Set $Z = \Sigma V$ and $J = V^\top \Sigma V$. The completion has representation (157) with floor a. These factors are rational, and their operator norms and the floor are at most one. Apply the dyadic rounding and positive scalar padding from Theorem 29. For $m < d$ satisfying (218), the ideal excess square root is at most $\sqrt{\varepsilon}/4$, since $63/((7/8)^2 128^3) < 1/16$. The same rounding adds at most $7\sqrt{\varepsilon}/100$. The public grid sizes are determined by comparisons of rational squares, so irrational square roots need not be represented. All retained entries have the original precision and storage bound. For $m = d$, use the full rational covariance and reference and the same final rounding, with zero ideal completion error.

Only $(n, m, V_q, Z_q, J_q, a_q, \theta_q)$ and the possible failure flag are persistent. Exact decoder arithmetic on these dyadic factors and the rational deleted records consists of rational matrix operations and one positive-definite solve, again of polynomial bit complexity. If a dyadic answer is required, round that solution to Euclidean error at most $\sqrt{\varepsilon}/100$. The last term of (215) is then less than $2\sqrt{\varepsilon}/100$, leaving the total below $\sqrt{\varepsilon}$. The simultaneous event and the sampling cutoff give overall success probability at least $1 - \delta$. □

The bit-complexity result uses exact rational determinants, solves, and approximate diagonalization; its polynomial bound is asymptotic rather than a claim of practical speed. For floating-point implementations, Proposition 33 requires the stated bounds on sampling, solve, formation, subspace, and decoder errors. Those encoding-time requirements are separate from the precision of the stored summary.

K Determinantal envelopes for arbitrary bounded labels

For arbitrary labels, the original ridge residual controls the determinantal covariance trace. The change-of-measure argument therefore extends beyond the unit-teacher model to labels inconsistent

with every linear predictor. The mean identity is established in Dereziński & Mahoney (2021, Theorem 3). The second-moment bound below applies the expected embedded-pseudoinverse inequality for positive semidefinite matrices in Mutný et al. (2020, Theorem 1 and Remark 1); the positive definite inverse identity also appears in Fanuel et al. (2021, Theorem 2.1). Our direct determinant proof fixes the normalization and accounts for zero-weight subsets in these established matrix-moment identities.

Lemma 35 (A residual bound on the determinantal covariance)**.** *Let $X \in \mathbb{R}^{n\times d}$, $y \in \mathbb{R}^n$, and $\lambda > 0$. Set $t = n\lambda$, $\Sigma = X^\top X/n$, $\theta = (X^\top X + tI)^{-1}X^\top y$, and*

$$p(J) = \frac{\det(X_J X_J^\top/t)}{\mathcal{Z}}, \qquad \mathcal{Z} = \det(I + XX^\top/t), \qquad z_J = X_J^\dagger y_J,$$

with $z_\varnothing = 0$. Define

$$r_S(\theta) = \frac{1}{n}\|y - X\theta\|_2^2, \qquad \mathcal{V} = \mathbb{E}_p[(z_J - \theta)(z_J - \theta)^\top].$$

Then

$$\mathbb{E}_p z_J = \theta, \qquad \operatorname{tr}\mathcal{V} \le \frac{r_S(\theta)}{\lambda}. \tag{221}$$

Equality in the trace bound holds whenever $y \in \operatorname{range} X$. No boundedness or rank assumption is needed for this lemma.

Proof. For completeness, fix $h \in \mathbb{R}^d$ and replace $X_J X_J^\top$ in the principal-minor partition sum by $X_J(X_J + syh^\top)^\top$. A positive-weight summand has derivative at zero equal to its original weight times $h^\top z_J$. Every zero-weight summand remains zero, because its left factor X_J has deficient row rank. The partition sum equals

$$\det\big(I + (X + syh^\top)^\top X/t\big),$$

whose derivative at zero is $\mathcal{Z}h^\top\theta$. Dividing by $\mathcal{Z}$ and varying h proves the mean identity.

For the second moment, differentiate instead

$$D(s) = \det\big(I + (XX^\top + syy^\top)/t\big) = \sum_{J\subseteq[n]} \det\big((X_J X_J^\top + syy_J^\top)/t\big).$$

For every positive-weight subset, the derivative is

$$\det(X_J X_J^\top/t)\, y_J^\top (X_J X_J^\top)^{-1} y_J = \mathcal{Z}p(J)\|z_J\|_2^2.$$

A zero-weight summand has nonnegative right derivative at zero: its determinant vanishes at zero and is nonnegative for $s \ge 0$. The finite sum is polynomial, so these derivatives exist and can be summed. Consequently

$$\begin{aligned}\mathbb{E}_p\|z_J\|_2^2 \le \frac{D'(0)}{\mathcal{Z}} &= \frac{1}{t} y^\top (I + XX^\top/t)^{-1} y \\ &= y^\top (XX^\top + tI)^{-1} y = \|\theta\|_2^2 + \frac{1}{t}\|y - X\theta\|_2^2.\end{aligned}$$

The last identity follows from $y - X\theta = t(XX^\top + tI)^{-1}y$ and $\theta = X^\top(XX^\top + tI)^{-1}y$. Subtracting $\|\mathbb{E}_p z_J\|_2^2$ proves (221). If $y = Xv$, every perturbed subset Gram matrix has the form $X_J(I + svv^\top)X_J^\top$; zero-weight determinants remain zero. All inequalities in the second-moment calculation are then equalities. □

For $y = Xv$, writing $K = \Sigma(\Sigma + \lambda I)^{-1}$ gives $r_S(\theta)/\lambda = v^\top K(I - K)v$. Thus the trace bound recovers (198) when $\|v\|_2 \le 1$.

Theorem 36 (A finite-bit determinantal bound with general labels)**.** *Suppose $0 < \lambda \le 1/4$, $0 < \rho \le 1/8$, $0 < \varepsilon \le 1$, and every record satisfies $\|x_i\|_2 \le 1$ and $|y_i| \le 1$. Define*

$$\Gamma_\rho = \exp\left(\frac{\rho}{(1-\rho)\lambda}\right) - 1, \qquad m = \min\left\{d, \left\lceil 128\left(\frac{\Gamma_\rho r_S(\theta)}{\lambda^2\varepsilon}\right)^{1/3}\right\rceil\right\}.$$

A deterministic offline summary of

$$O\left(d(m+1)\log\left(2+\frac{d}{\lambda\varepsilon}\right)+\log(2+n)\right) \tag{222}$$

bits guarantees retained excess at most ε simultaneously for every legal deletion of at most ρn records. The decoder uses only the summary and the deleted records. The inputs may be arbitrary bounded reals, and encoding computation is unrestricted.

Proof. Use the distribution and covariance in Lemma 35. For a query U, write $\alpha = |U|/n$, $\beta = 1-\alpha$, $C = X_U^\top X_U/n$, and $f = X_U^\top y_U/n$. With $Q = (\Sigma - C)/\lambda$, the density (199) depends only on features. Applying the general mean identity to the retained rows, at scale $n\beta\lambda$, gives

$$\mathbb{E}_p[f_U z_J] = w_{S-U}, \qquad \Delta_U = w_{S-U} - \theta = \mathbb{E}_p[(f_U - 1)(z_J - \theta)].$$

The density calculation (200)– (201) uses only bounded features. Thus

$$\Delta_U \Delta_U^\top \preceq \frac{\alpha\Gamma_\rho}{\rho}\mathcal{V}, \qquad \operatorname{tr}\mathcal{V} \le \frac{r_S(\theta)}{\lambda}. \tag{223}$$

Let $\mathcal{S}$ span the leading m eigenvectors of both Σ and $\mathcal{V}$, and use the covariance completion (156) with its representation (157). Its basis has at most $2m$ columns. For $m < d$, the completion error E satisfies $0 \preceq E \preceq \delta I$, where $\delta \le 1/(m+1)$, and annihilates the chosen $\mathcal{V}$ eigenspace. Therefore

$$\|E\Delta_U\|_2^2 \le \frac{\alpha\Gamma_\rho}{\rho}\,\delta^2\lambda_{m+1}(\mathcal{V}) \le \frac{\alpha\Gamma_\rho r_S(\theta)}{\rho\lambda(m+1)^3}.$$

The centered decoder with exact reference is

$$\overline{w}_U = \theta + [\widehat{\Sigma} - C + \beta\lambda I]^{-1}[(C+\alpha\lambda I)\theta - f].$$

The same quadratic excess calculation as in (121) gives the ideal guarantee

$$F_{S-U}(\overline{w}_U) - F_{S-U}(w_{S-U}) \le \frac{\Gamma_\rho r_S(\theta)}{(1-\rho)^2\lambda^2(m+1)^3}. \tag{224}$$

At $m = d$ the ideal covariance is exact. The chosen rank makes the ideal excess square root at most $\sqrt{\varepsilon}/4$.

It remains to adjust the rounding precision for general labels. Every nonempty bounded dataset A has $\|w_A\|_2 \le 1/(2\sqrt{\lambda})$: if $a^2 = \|X_A w_A\|_2^2/|A|$, stationarity and Cauchy–Schwarz give

$$a^2 + \lambda\|w_A\|_2^2 = w_A^\top X_A^\top y_A/|A| \le a, \qquad \lambda\|w_A\|_2^2 \le a - a^2 \le 1/4.$$

Hence $\|\Delta_U\|_2 \le 1/\sqrt{\lambda}$. Choose a public dyadic $\eta \in (\lambda\sqrt{\varepsilon}/200, \lambda\sqrt{\varepsilon}/100]$. Round the factors V, Z, J, δ in (157) at a dyadic entry step within a factor two below $\eta/(16d)$, preserving the symmetry of J, and add ηI to the reconstructed covariance. The factor calculation in Proposition 24 gives

$$0 \preceq R := \widehat{\Sigma}_q - \widehat{\Sigma} \preceq 2\eta I.$$

For exact reference the resulting extra excess square root is at most

$$\frac{\|R\Delta_U\|_2}{\beta\sqrt{\lambda}} \le \frac{2\eta}{\beta\lambda} \le \frac{3\eta}{\lambda}.$$

Round the reference to satisfy $\|\theta_q - \theta\|_2 \le \lambda\sqrt{\varepsilon}/100$. Its error is multiplied by $I + H^{-1}(C+\alpha\lambda I)$, where $H = \widehat{\Sigma}_q - C + \beta\lambda I \succeq \beta\lambda I$. As in Proposition 22, this adds at most $2\|\theta_q - \theta\|_2/\lambda$ to the excess square root. The total is at most $(1/4 + 3/100 + 2/100)\sqrt{\varepsilon} < \sqrt{\varepsilon}$.

The stored state is $(n, k, V_q, Z_q, J_q, \delta_q, \theta_q)$, where $k \le \min\{d, 2m\}$ is the basis dimension. Covariance factors have entries of magnitude at most one, the reference entries have magnitude at most $1/(2\sqrt{\lambda})$, and all reciprocal rounding steps are polynomial in $d, 1/\lambda, 1/\varepsilon$. This proves (222). The residual value is used only to choose the encoded rank; neither it nor the full determinantal covariance needs to be stored. □

Since $r_S(\theta) \le F_S(\theta) \le 1$, the theorem also gives a uniform bounded-label upper bound. If $\rho \le M\lambda$ for a fixed $M > 0$, then $\Gamma_\rho = O_M(\rho/\lambda)$, and the sufficient rank is

$$m = O_M\left(\min\left\{d, 1 + \left(\frac{\rho r_S(\theta)}{\lambda^3 \varepsilon}\right)^{1/3}\right\}\right).$$

This is an upper bound; no matching general-label lower bound is asserted.

Proposition 37 (Bounded labels do not ensure covariance concentration)**.** *The sample-size guarantee in Proposition 31 does not extend uniformly to arbitrary bounded labels. Even for $n = d = 2$ and fixed $\lambda > 0$, the centered determinantal fit can have an arbitrarily large fourth moment, while its covariance stays bounded away from zero and any fixed number of draws misses that covariance with probability tending to one.*

Proof. Take $x_1 = x_2 = ae_1$ and $(y_1, y_2) = (1, -1)$, where $0 < a \le 1$. These bounded labels are inconsistent, and $\theta = 0$. The only positive-weight subsets are the empty set and the two singletons, with

$$p(\varnothing) = \frac{\lambda}{\lambda + a^2}, \qquad p(\{1\}) = p(\{2\}) = \frac{a^2}{2(\lambda + a^2)}, \qquad z_{\{1\}} = \frac{e_1}{a}, \quad z_{\{2\}} = -\frac{e_1}{a}.$$

Consequently

$$\mathcal{V} = \frac{e_1 e_1^\top}{\lambda + a^2}, \qquad \mathbb{E}_p \|z_J - \theta\|_2^4 = \frac{1}{a^2(\lambda + a^2)}.$$

With s independent draws, the empirical covariance in (206) is zero on the event that every draw is empty. This event has probability $[\lambda/(\lambda + a^2)]^s$, which tends to one as $a \downarrow 0$ for every fixed s. For $0 < \lambda \le 1/4$ and $m = 1$, the desired comparison $\mathcal{V} \preceq 2\widehat{\mathcal{V}} + I/2$ fails on this event for all $0 < a \le 1$. Thus a sample count depending only on n, d, λ, m and the failure probability cannot give that comparison uniformly over bounded records. □

The deterministic envelope therefore extends to general bounded labels, whereas the ordinary-DPP sampled encoder in Appendix I uses the bounded-teacher condition. A finite second moment alone does not provide its covariance concentration step. The different importance law below supplies bounded matrix contributions and a polynomial bit encoder for rational bounded-label inputs.

K.1 Residual importance sampling for arbitrary labels

The obstruction in Proposition 37 concerns ordinary draws from p. A label-dependent importance law gives bounded covariance contributions for arbitrary labels. Its normalizer is exactly the residual bound in Lemma 35. The construction combines a rank-one determinant derivative with a bordered Gram determinant; both terms are necessary because a centered fit has a component orthogonal to the selected row span.

Lemma 38 (A residual-normalized importance law)**.** *Use $X, y, t, \theta, p, \mathcal{Z}$, and $\mathcal{V}$ from Lemma 35, and set*

$$e = y - X\theta, \qquad R = \frac{\|e\|^2}{t} = \frac{r_S(\theta)}{\lambda}, \qquad L = \frac{XX^\top}{t}.$$

Suppose $R > 0$. Introduce one additional index $*$ and define

$$X^+ = \begin{pmatrix} X \\ \theta^\top \end{pmatrix}, \qquad L^+ = X^+(X^+)^\top / t.$$

For every $J \subseteq [n]$, including the empty set, put

$$a_J = \det(L_J + e_J e_J^\top / t) - \det L_J, \qquad b_J = t \det L^+_{J \cup \{*\}},$$
$$q(J) = \frac{a_J + b_J}{\mathcal{Z} R}. \tag{225}$$

Then $a_J, b_J \ge 0$, q is a probability distribution, and

$$q(J) = \frac{p(J)\|z_J - \theta\|^2}{R} \quad \textit{whenever } p(J) > 0. \tag{226}$$

Define a symmetric random matrix for $J \sim q$ by

$$H_J = \begin{cases} R\dfrac{(z_J-\theta)(z_J-\theta)^\top}{\|z_J-\theta\|^2}, & p(J)>0 \text{ and } z_J \neq \theta, \\ 0, & \text{otherwise}. \end{cases} \tag{227}$$

It satisfies

$$\mathbb{E}_q H_J = \mathcal{V}, \qquad 0 \preceq H_J \preceq RI, \qquad \operatorname{tr} H_J \leq R. \tag{228}$$

If $R = 0$, then $y = 0$, and every retained ridge optimizer is zero.

Proof. The determinant of $L_J + se_J e_J^\top/t$ is affine in s. Both its value at zero and its right derivative are nonnegative: for an invertible L_J this follows from the determinant lemma, and for a singular L_J it follows from nonnegativity for $s \geq 0$. Thus a_J is that nonnegative derivative. The bordered Gram matrix is positive semidefinite, so $b_J \geq 0$ as well.

For $p(J) > 0$, write $P_J = X_J^\dagger X_J$. The two terms in

$$z_J - \theta = X_J^\dagger e_J - (I - P_J)\theta$$

are orthogonal. The determinant lemma and the Schur complement give

$$a_J = \det L_J \|X_J^\dagger e_J\|^2, \qquad b_J = \det L_J \|(I-P_J)\theta\|^2.$$

These identities also hold for $J = \varnothing$, with $a_\varnothing = 0$ and $b_\varnothing = \|\theta\|^2$. If $p(J) = 0$, the rows of X_J are dependent, so the bordered Gram determinant vanishes and $b_J = 0$; a_J can remain positive.

Set $K = X(X^\top X + tI)^{-1}X^\top$ and $K_d = X^\top X(X^\top X + tI)^{-1}$. Summing the rank-one derivative over principal minors gives

$$\frac{1}{\mathcal{Z}}\sum_J a_J = e^\top (XX^\top + tI)^{-1} e = \frac{1}{t} e^\top (I-K)e.$$

For the other term, expanding in the identity entries indexed by the original rows and taking the Schur complement yields

$$\begin{aligned} \sum_J b_J &= t \det \begin{pmatrix} I + XX^\top/t & X\theta/t \\ \theta^\top X^\top/t & \|\theta\|^2/t \end{pmatrix} \\ &= \mathcal{Z}\,\theta^\top (I - K_d)\theta. \end{aligned}$$

Ridge stationarity gives $X^\top e = t\theta$, and hence

$$\theta^\top (I - K_d)\theta = \frac{1}{t} e^\top K e.$$

Consequently $\sum_J (a_J + b_J) = \mathcal{Z}R$. The orthogonal decomposition proves (226); multiplying it by (227) and summing proves unbiasedness. Each nonzero H_J is R times a rank-one orthogonal projector, which proves the remaining bounds. Finally $R = 0$ implies $e = 0$; stationarity then gives $\theta = 0$ and $y = X\theta + e = 0$. □

The extra q-mass on subsets with $p(J) = 0$ contributes the zero matrix. Its total mass is $1 - \operatorname{tr}\mathcal{V}/R$. Thus the proposal need not condition on row independence or reject these subsets. In particular, small original DPP probabilities do not create large importance-weighted covariance contributions.

For the duplicate-row example in Proposition 37, $\theta = 0$ and $R = 1/\lambda$. The importance law is

$$q(\varnothing) = 0, \qquad q(\{1\}) = q(\{2\}) = \frac{\lambda}{2(\lambda + a^2)}, \qquad q(\{1,2\}) = \frac{a^2}{\lambda + a^2}.$$

Each singleton contributes $e_1 e_1^\top/\lambda$, and the dependent pair contributes zero. As $a \downarrow 0$, the informative singleton draws therefore have total probability tending to one.

Proposition 39 (Concentration of the importance covariance)**.** *Suppose $R > 0$, $1 \le m < d$, and $0 < \delta < 1$. For independent $J_1, \dots, J_s \sim q$, let*

$$s \ge \left\lceil 12(m+1)\log\frac{2d}{\delta} \right\rceil, \qquad \widehat{\mathcal{V}}_q = \frac{1}{s}\sum_{j=1}^{s} H_{J_j}.$$

With probability at least $1-\delta$,

$$\mathcal{V} \preceq 2\widehat{\mathcal{V}}_q + \frac{R}{m+1} I. \tag{229}$$

Also $\operatorname{tr}\widehat{\mathcal{V}}_q \le R$ deterministically. If Π projects off its leading m eigenvectors, then on the same event

$$\lambda_{\max}(\Pi\mathcal{V}\Pi) \le \frac{3R}{m+1}.$$

Proof. Apply the Bernstein proof of Proposition 31 to the matrices H_J/R. That proof uses only a positive semidefinite summand bounded by I and its expectation; the summand need not be the outer product of an unweighted centered fit. Its conclusion is $\mathcal{V}/R \preceq 2\widehat{\mathcal{V}}_q/R + I/(m+1)$. The deterministic trace bound and the eigenvalue tail argument are unchanged after this normalization. □

Theorem 40 (Polynomial bit encoding for arbitrary bounded labels)**.** *Assume the records satisfy $\|x_i\| \le 1$ and $|y_i| \le 1$. Assume these records and the public parameters $\lambda, \rho, \varepsilon, \delta$ are rational, with $\lambda, \rho, \varepsilon$ in the ranges of Theorem 36 and $0 < \delta < 1$. Let B bound their individual rational encoding lengths, and set $R = r_S(\theta)/\lambda$. For $R > 0$ and any supplied $1 \le m < d$, there is a randomized encoder using*

$$s = 12(m+1)\left\lceil \log_2 \frac{4d}{\delta} \right\rceil \tag{230}$$

exact samples from (225)*, with polynomial expected bit-operation complexity in n, d, B. With probability at least $1-\delta/2$, it constructs an ideal covariance completion whose centered decoder obeys, simultaneously for every legal U,*

$$F_{S-U}(\overline{w}_U) - F_{S-U}(w_{S-U}) \le \frac{63\Gamma_\rho R}{(1-\rho)^2\lambda(m+1)^3}. \tag{231}$$

The sampling can be truncated at a polynomial bit-operation budget with additional failure probability at most $\delta/2$. For $m = d$, or for a supplied rank satisfying

$$m+1 \ge 128\left(\frac{\Gamma_\rho R}{\lambda\varepsilon}\right)^{1/3}, \tag{232}$$

final rounding gives retained excess at most ε simultaneously for all legal deletions, with probability at least $1-\delta$, using the storage bound (222) *and polynomial bit-operation complexity. A sufficient rank can be selected in polynomial bit time with*

$$m = \min\left\{ d, O\left(1 + \left(\frac{\Gamma_\rho r_S(\theta)}{\lambda^2\varepsilon}\right)^{1/3}\right)\right\}.$$

The decoder can produce a dyadic answer in polynomial bit time. If $R = 0$, a constant-size zero-output flag suffices.

Proof. **Conditional sampling by rational determinants.** The original reference, residual, and R are rational and computable by a positive-definite linear solve. For disjoint sets $A, T \subseteq [n]$, let A contain the indices required to be included and T the undecided indices, and put $G = A \cup T$. Excluded indices appear in neither set. Let D be diagonal on G, with zeros on A and ones on T, and extend it by a zero in coordinate $*$ to obtain D^+. The unnormalized mass of this partial decision is

$$\begin{aligned} W_q(A,T) &= \sum_{J:\, A\subseteq J\subseteq G} (a_J + b_J) \\ &= \det(D + L_G + e_G e_G^\top/t) - \det(D + L_G) + t\det(D^+ + L^+_{G\cup\{*\}}). \end{aligned} \tag{233}$$

To see the equality, expand each determinant in the diagonal identity entries on T. In the last determinant the zero diagonal entry at $*$ forces its inclusion. This is the same principal-minor expansion as (219), and remains valid for singular matrices and empty sets.

For each $i \in T$, include it with conditional probability

$$\frac{W_q(A \cup \{i\}, T \setminus \{i\})}{W_q(A, T)}.$$

The excluded and included masses sum to the denominator, which is positive along every positive-probability path. At the initial state $W_q(\varnothing, [n]) = \mathcal{Z}R$, so the final subset has law q. All matrices in (233) are rational and have order at most $n+1$. Their determinants and differences have polynomial rational bit length by denominator clearing and Hadamard's bound. Exact integer-preserving elimination (Bareiss, 1968) therefore computes all conditional probabilities in polynomial bit time. Recomputing the determinants at each decision suffices; no eigenvalue computation or conditioning on a rare acceptance event is required.

Implement each rational Bernoulli probability by the uniform-integer rejection procedure in the proof of Theorem 34. Its expected number of attempts is less than two. Fresh bits give independent exact subsets; capping each of the ns decisions at $\lceil \log_2(2ns/\delta) \rceil$ attempts adds failure probability at most $\delta/2$ and gives a deterministic polynomial bit budget.

Exact covariance and an approximate rational subspace. For a sampled dependent subset, set $H_J = 0$ without solving its Gram system. For an independent nonempty subset, compute $z_J = X_J^\top (X_J X_J^\top)^{-1} y_J$ by exact rational elimination; for the empty subset use $z_J = 0$. Formula (227) uses rational outer products and a rational squared norm, so it requires no square roots. Every entry has polynomial bit length, even when a Gram matrix is poorly conditioned. The same holds for the exact average $\widehat{\mathcal{V}}_q$, since s is polynomial in d and B.

Put $r = m+1$ and $H = \Sigma + \widehat{\mathcal{V}}_q/R$. Then $H \succeq 0$ and $\operatorname{tr} H \le 2$. Use rational approximate diagonalization (Del Pia, 2026, Theorem 2), exactly as in Theorem 34, to obtain an exactly orthogonal rational matrix Q such that the off-diagonal part of $Q^\top H Q$ has Frobenius norm at most $1/r$. Choose the m columns with largest diagonal entries, form V from these columns, and let $\Pi = I - VV^\top$. Their nonnegative diagonal sum is at most two, giving

$$\lambda_{\max}(\Pi \Sigma \Pi) \le a := \min\{1, 3/r\}, \qquad \lambda_{\max}(\Pi \widehat{\mathcal{V}}_q \Pi) \le 3R/r.$$

Proposition 39, with failure probability $\delta/2$, implies $\lambda_{\max}(\Pi \mathcal{V} \Pi) \le 7R/r$. The covariance completion $\widehat{\Sigma} = \Sigma + \Pi(aI - \Sigma)\Pi$ therefore has error $0 \preceq E \preceq a\Pi$. On this single sampling event, the general-label envelope (223) gives

$$\|E\Delta_U\|^2 \le \frac{63\Gamma_\rho R}{r^3}$$

for every legal U. The centered-decoder excess bound is this quantity divided by $(1-\rho)^2\lambda$, proving (231). No eigengap or union bound over deletion requests is used.

Choosing the residual-dependent rank. Let $c = R/(\lambda\varepsilon) > 0$ and $x = \rho/((1-\rho)\lambda)$. Both are rationals of polynomial bit length. Let B_c be the binary encoding length of the positive denominator of c in lowest terms, so $c \ge 2^{-B_c}$, and set

$$T_0 = 3\lceil \log_2(2d) \rceil + B_c + 2.$$

If $x > T_0$, then $(e^x - 1)c > d^3$, so $m = d$ is within the asserted rank bound. Otherwise x is bounded by a polynomial in the input length. Set $k = \max\{1, \lceil 2x \rceil\}$ and

$$S_k = \sum_{j=1}^{k} x^j/j!, \qquad \overline{\Gamma} = S_k + \frac{2x^{k+1}}{(k+1)!}.$$

Successive omitted Taylor terms have ratio at most $1/2$, which gives $\Gamma_\rho \le \overline{\Gamma} \le 2\Gamma_\rho$. Choose $m = \min\{d, \lceil 128(\overline{\Gamma}c)^{1/3} \rceil\}$ by rational cube comparisons. All computations have polynomial bit complexity. The denominator-dependent cutoff is needed here: $\Gamma_\rho > d^3$ alone does not justify full rank when R is small.

Rounding, storage, and decoding. Set $Z = \Sigma V$ and $J = V^\top \Sigma V$, and use the completion representation (157) with floor a. The factors and floor have operator norms at most one. For a rank satisfying (232), the ideal excess square root is at most $\sqrt{\varepsilon}/4$, since $63/((7/8)^2 128^3) < 1/16$. Apply the general-label factor and reference rounding from the proof of Theorem 36: choose $\eta \in (\lambda\sqrt{\varepsilon}/200, \lambda\sqrt{\varepsilon}/100]$, use covariance padding at most $2\eta I$, and round the reference within $\lambda\sqrt{\varepsilon}/100$ in Euclidean norm. These steps add at most $5\sqrt{\varepsilon}/100$ to the excess square root. The dyadic grid sizes can be selected by rational square comparisons. For $m = d$, use the full rational covariance and reference with the same rounding and zero ideal completion error.

Only $(n, m, V_q, Z_q, J_q, a_q, \theta_q)$ and a possible failure flag are persistent. In particular, e, R, q, the sampled subsets, and the sample matrices are temporary. The general-label factor bounds give exactly (222). The decoder uses rational matrix operations and a positive-definite solve on the stored state and deleted records. Rounding its output within $\sqrt{\varepsilon}/100$ adds less than $2\sqrt{\varepsilon}/100$ to the excess square root. Thus the final excess is below ε, with overall success probability at least $1 - \delta$. The case $R = 0$ follows from Lemma 38. □

The determinant identities and conditional rational-sampling method build on the standard DPP machinery already cited in Appendices K and J. A residual-augmented volume change of measure also appears in Rhee (2026, Theorem 2) for fixed-size, overdetermined ordinary least squares, where the augmented law is used only in the analysis. Squared-norm importance weighting and the bounded-matrix concentration step are standard. The additional construction here combines the derivative and bordered determinant into a computable centered importance law for variable-size minimum-norm fits. Its bounded matrix contributions preserve the residual-dependent cubic deletion rank for arbitrary bounded labels. The polynomial bit-operation construction uses exact rational determinants and solves. Its arithmetic guarantees are distinct from the floating-point evaluations of ordinary DPP sampling in Appendix Z.1 and spectral importance sampling in Appendix Z.2. These grids provide numerical comparisons, without rigorous arithmetic certification or an advantage over the strongest centered spectral comparators.

K.1.1 A FAST SPECTRAL SAMPLER FOR THE RESIDUAL IMPORTANCE LAW

The rational determinant sampler above establishes polynomial bit complexity. The same importance law also admits a low-rank spectral implementation. Its derivative component is a partial-projection DPP, a special case of the extended L-ensembles of Tremblay et al. (2023); its bordered component is an ordinary L-ensemble after a feature projection. Thus the standard Bernoulli-eigenvector and projection-DPP sampling steps suffice.

Proposition 41 (A two-component spectral representation)**.** *Use the notation of Lemma 38, with $R > 0$, and put*

$$u = \frac{e}{\|e\|}, \qquad L_a = (I - uu^\top)L(I - uu^\top), \qquad K_a = uu^\top + L_a(I + L_a)^{-1}.$$

If $\theta \neq 0$, also put

$$L_b = \frac{1}{t} X \left(I - \frac{\theta\theta^\top}{\|\theta\|^2} \right) X^\top.$$

Define

$$\alpha = \frac{e^\top (I - K)e}{\|e\|^2}, \qquad \beta = \frac{e^\top K e}{\|e\|^2} = 1 - \alpha, \qquad K = L(I + L)^{-1}. \tag{234}$$

Then $0 < \alpha \leq 1$, $\beta \geq 0$, and a draw from q can be obtained as follows: with probability α, draw a DPP with marginal kernel K_a; with probability β, draw the L-ensemble with likelihood kernel L_b. If $\theta = 0$, then $\beta = 0$ and the second component is omitted. More precisely, the component probabilities are

$$\Pr_a(J) = \frac{a_J}{\mathcal{Z} R \alpha}, \qquad \Pr_b(J) = \frac{b_J}{\mathcal{Z} R \beta}, \tag{235}$$

where the second identity is used only when $\beta > 0$. These identities hold for singular X, dependent selected rows, and the empty subset.

Proof. Let $v = e/\sqrt{t}$, $C = I + L$, and $c = v^\top C^{-1} v > 0$. For $s > 0$, the L-ensemble with likelihood kernel $L + svv^\top$ has probability

$$\frac{\det L_J + s a_J}{\mathcal{Z}(1 + sc)}.$$

The numerator identity holds even for singular L_J, because the determinant is affine in a rank-one update. As $s \to \infty$, its probability converges to $a_J/(\mathcal{Z}c)$ for every subset J. The Sherman–Morrison identity gives the limiting marginal kernel

$$K + C^{-1} v v^\top C^{-1}/c. \tag{236}$$

Continuity of all principal minors shows that this is the marginal kernel of the limiting law.

Choose an orthonormal complement Q_e to u. In the orthogonal coordinates $[u, Q_e]$, the inverse of $C + svv^\top$ tends to the block matrix with a zero first row and column and lower block $(I + Q_e^\top L Q_e)^{-1}$. Thus (236) equals K_a. In particular, its eigenvalue-one direction is u. The normalizer from Lemma 38 gives $c = R\alpha$, proving the first identity in (235).

For $\theta \neq 0$, taking the Schur complement of the additional diagonal entry $\|\theta\|^2/t$ gives, without requiring L_J to be invertible,

$$b_J = \|\theta\|^2 \det(L_b)_J.$$

Summing and using ridge stationarity $X^\top e = t\theta$ yields

$$\|\theta\|^2 \det(I + L_b) = \mathcal{Z}\, \theta^\top (I - K_d)\theta = \mathcal{Z} R \beta.$$

This proves the second component law. If $\theta = 0$, stationarity gives $X^\top e = 0$, hence $\beta = 0$ and every $b_J = 0$. Finally, $I - K \succ 0$ and $K \succeq 0$ establish the asserted signs, and the component mixture has probability $(a_J + b_J)/(\mathcal{Z}R)$. □

Implementation and arithmetic cost. No $n \times n$ kernel is needed. Let

$$Y_a = Q_e^\top X, \qquad Y_b = X Q_\theta,$$

where Q_θ is an orthonormal complement to $\theta/\|\theta\|$ when $\theta \neq 0$. If $Y_a = U_a S_a V_a^\top$ is a thin singular-value decomposition, the first component always retains u and independently retains each column of $Q_e U_a$ with probability $s_{a,j}^2/(s_{a,j}^2 + t)$. It then draws a projection DPP from the retained orthonormal columns together. The second component uses the left singular vectors of Y_b, retained independently with probabilities $s_{b,j}^2/(s_{b,j}^2 + t)$, followed by the same projection sampler. Zero singular values have zero retention probability. The mandatory direction in the first component is an eigenvector, not a mandatory data index.

Householder reflections implement both complements without explicitly forming Q_e or an $n \times n$ matrix. Reflect u to a coordinate vector, apply the reflection to X, and omit that coordinate row before the first singular-value decomposition. Lift its left singular vectors by restoring a zero row and reflecting back. For the second component, reflect $\theta/\|\theta\|$ to a coordinate vector in feature space and omit the corresponding feature column of the reflected X. Empty complements, including $n = 1$ or $d = 1$, cause no difficulty. This construction explicitly separates the forced direction from the other spectral columns.

The mixture weights also have a formula that avoids subtracting nearly equal scalars. Put $w = X^\top u$ and

$$\eta = w^\top (tI + Y_a^\top Y_a)^{-1} w, \qquad \alpha = \frac{1}{1 + \eta}, \qquad \beta = \frac{\eta}{1 + \eta}. \tag{237}$$

Indeed, $X^\top X = Y_a^\top Y_a + ww^\top$, and the determinant lemma gives

$$\frac{\det(I + L_a)}{\det(I + L)} = \frac{1}{1 + \eta}.$$

In the coordinates $[u, Q_e]$, the rank-one determinant derivative also gives $\sum_J a_J = R \det(I + L_a)$, proving (237). Its regularized solve can use the same singular-value decomposition, including any right null space directions; it requires no subtraction of a projected squared norm from the original squared norm.

Preparing the ridge reference, both thin decompositions, and the mixture weights costs $O(nd^2 + d^3)$ arithmetic operations and $O(nd + d^2)$ working storage. For a draw of cardinality k, the projection step costs $O(nk^2)$ operations; spectral selection adds $O(d)$ operations. Here $k \le \min\{n, d+1\}$ for the first component and $k \le \min\{n, d-1\}$ for the second. Computing the rank and, when appropriate, the minimum-norm fit and H_J adds at most $O(d^3)$. Thus each covariance draw costs $O(nk^2 + d^3)$ after preparation. Dependent subsets sampled by the first component contribute $H_J = 0$ as prescribed in (227); they are retained in the average. The second component can produce the empty subset. When $R = 0$, the zero-output case of Lemma 38 applies and no draw is needed.

Corollary 42 (Low-rank exact rational conditionals)**.** *Under the rational-input assumptions of Theorem 40, the importance law also has an exact rational sampler with $O(nd^2 + d^3)$ preprocessing operations and $O(n(d+1)^2)$ rational arithmetic operations per subset. It uses $O(nd + d^2)$ working storage, has polynomial expected bit complexity, and admits a deterministic polynomial bit budget with the same additional failure allowance as that theorem.*

Proof. Write $E = e^\top e > 0$ and

$$A_e = X - \frac{e(e^\top X)}{E}, \qquad F_a = [e, A_e], \qquad G_a = \operatorname{diag}\left(E^{-1}, (A_e^\top A_e + tI)^{-1}\right).$$

All these matrices are rational, and $K_a = F_a G_a F_a^\top$. For the second component, when present, use

$$F_b = X\left(I - \frac{\theta\theta^\top}{\theta^\top\theta}\right), \qquad G_b = (F_b^\top F_b + tI)^{-1}.$$

Its marginal kernel is $K_b = F_b G_b F_b^\top$. The mixture weights can also be computed without square roots:

$$\eta = \frac{(X^\top e)^\top (A_e^\top A_e + tI)^{-1}(X^\top e)}{E}, \qquad \alpha = \frac{1}{1+\eta}, \qquad \beta = \frac{\eta}{1+\eta}.$$

These are the weights in (237) because $A_e^\top A_e = Y_a^\top Y_a$.

After drawing the component, fix its factor F and initialize $G = G_a$ or G_b. Process its rows in order. If f_i is the transpose of row i of F, compute $v_i = Gf_i$ and $p_i = f_i^\top v_i$. Include index i with probability p_i, using the update

$$G \leftarrow \begin{cases} G - v_i v_i^\top / p_i, & i \text{ included}, \\ G + v_i v_i^\top / (1 - p_i), & i \text{ excluded}. \end{cases} \tag{238}$$

Only the chosen update is evaluated. In particular, $p_i = 0$ forces exclusion and $p_i = 1$ forces inclusion, so no zero denominator is used. The DPP inclusion and exclusion Schur complements show inductively that the conditional kernel on the remaining rows is $F_{\mathrm{rem}} G F_{\mathrm{rem}}^\top$. Consequently every encountered p_i lies in $[0, 1]$, and the final subset has the chosen component law. Each update costs $O((d+1)^2)$ operations.

For completeness, the intermediate rational lengths remain polynomial even though many updates are composed. Fix the original component kernel $K_c = F G_0 F^\top$. After a set P of rows has been processed, let D_- be diagonal on P, with a one exactly at an excluded row, and write $N_- = \operatorname{tr} D_-$. The probability of these partial decisions is

$$(-1)^{N_-} \det\left((K_c)_P - D_-\right).$$

Thus $B_P = (K_c)_P - D_-$ is invertible along every realized path of positive probability. Repeated Schur complements give

$$G = G_0 - G_0 F_P^\top B_P^{-1} F_P G_0.$$

This identity is a bit-length bound; the implementation need not form either K_c or B_P. Their entries have polynomial rational length, so determinant and cofactor bounds for matrices of order at most n bound every entry of G and every next probability by a polynomial rational length. Reducing fractions during the updates therefore gives polynomial bit complexity.

There are at most $(n+1)s$ rational Bernoulli calls for s subsets, including their mixture choices. Use the uniform-integer procedure from Theorem 34, with fresh bits for each call. Its expected number of attempts is less than two. Capping each call at

$$\left\lceil \log_2 \frac{2(n+1)s}{\delta} \right\rceil$$

attempts adds failure probability at most $\delta/2$ and gives a deterministic polynomial bit budget. The previously proved covariance and encoding conclusions then apply to the exact draws. □

Sampler arithmetic and covariance estimation. The spectral construction has law q in exact arithmetic. These operation counts are arithmetic counts, with spectral decompositions and sampling from the resulting probabilities available; they do not replace the rational bit-complexity proof in Theorem 40. The low-rank conditional corollary provides a separate rational alternative with a sharper arithmetic bound. A stable floating-point implementation must handle roundoff, numerical rank decisions, conditional probabilities near zero or one, and very small mixture or spectral probabilities. Failed draws must remain in the reported outcomes rather than being replaced by successful ones.

The representation specializes established extended-L-ensemble and projection-DPP sampling machinery to the residual importance law. The resulting samples estimate $\mathcal{V}$ through the bounded contributions H_J. The importance law does not assert $\mathbb{E}_q z_J = \theta$, so an unweighted average of the sampled fits is not the covariance estimator in Proposition 39.

K.1.2 Perturbations of normalized importance matrices

Bounded importance matrices yield an arithmetic perturbation bound for arbitrary bounded labels. Certification requires control of both the sampling law and the normalized matrix contributions, beyond a backward-residual check for the least-squares solve.

Theorem 43 (A normalized importance covariance certificate)**.** *Under the bounded-record and parameter hypotheses of Theorem 36, suppose $R > 0$, let $1 \le m < d$, and put $r = m + 1$. Define, for every subset, $B_J = H_J/R$ using (227); thus $0 \preceq B_J \preceq I$, including $B_J = 0$ on original-p-zero subsets. Let $\widetilde{q}$ satisfy $\|\widetilde{q} - q\|_{\mathrm{TV}} \le \zeta$, and draw*

$$s \ge \left\lceil 12r \log \frac{2d}{\delta} \right\rceil$$

independent subsets $J_1, \dots, J_s$ from $\widetilde{q}$, where $0 < \delta < 1$.

Suppose that on one event of probability at least $1 - \nu$, the computed symmetric matrices $\widetilde{B}_j, M$ satisfy

$$\begin{aligned} 0 \preceq \widetilde{B}_j \preceq I, \qquad & \frac{1}{s}\sum_{j=1}^{s} \|\widetilde{B}_j - B_{J_j}\|_{\mathrm{op}} \le \kappa, \\ & \left\| M - \frac{1}{s}\sum_{j=1}^{s} \widetilde{B}_j \right\|_{\mathrm{op}} \le \xi, \end{aligned} \tag{239}$$

where ν, ζ, κ, ξ are nonnegative. Then, with probability at least $1 - \delta - \nu$,

$$\mathcal{V} \preceq R(2M + \ell I), \qquad \ell = \frac{1}{r} + \zeta + 2\kappa + 2\xi. \tag{240}$$

On the same event, let $\overline{R} \ge R$ be a certified bound and let Π be any orthogonal projector with certified bounds

$$0 \le a \le 1, \quad b \ge 0, \qquad \lambda_{\max}(\Pi\Sigma\Pi) \le a, \qquad \lambda_{\max}(\Pi M \Pi) \le b.$$

Then

$$\lambda_{\max}(\Pi\mathcal{V}\Pi) \le \tau_V := \overline{R}(2b + \ell). \tag{241}$$

Use $\widehat{\Sigma} = \Sigma + \Pi(aI - \Sigma)\Pi$ and allow a final covariance perturbation $0 \preceq \widehat{\Sigma}_q - \widehat{\Sigma} \preceq \omega I$, stored-reference error at most t_θ, and decoder output error at most t_w relative to the centered decoder formed from these stored quantities. Simultaneously for every legal deletion U,

$$\begin{aligned} \sqrt{F_{S-U}(\widetilde{w}_U) - F_{S-U}(w_{S-U})} \le{}& \frac{a\sqrt{\Gamma_\rho \overline{R}(2b + \ell)}}{(1-\rho)\sqrt{\lambda}} + \frac{\omega}{(1-\rho)\lambda} \\ & + \frac{2t_\theta}{\lambda} + \sqrt{1+\lambda}\, t_w. \end{aligned} \tag{242}$$

If a computed residual scale $\widetilde{R} > 0$ has the certificate $|\widetilde{R}/R - 1| \le \chi < 1$, one may take $\overline{R} = \widetilde{R}/(1-\chi)$. Thus scale uncertainty inflates the first excess-square-root term computed with $\widetilde{R}$ by at most $(1-\chi)^{-1/2}$.

Proof. Put $A = \mathbb{E}_{\widetilde{q}} B_J$. For each unit vector h, $h^\top B_J h \in [0,1]$, so total variation gives $\mathcal{V}/R \preceq A + \zeta I$. The bounded-matrix Bernstein argument from Proposition 39 applies to B_J under $\widetilde{q}$ as well. Except on an event of probability δ, it gives

$$A \preceq \frac{2}{s}\sum_{j=1}^{s} B_{J_j} + \frac{1}{r} I.$$

On the computational event, $s^{-1}\sum_j B_{J_j} \preceq M + (\kappa + \xi) I$. Combining these statements and taking a union bound proves (240). The computational event need not be independent of the samples. Compression by Π and $R \le \overline{R}$ prove the tail bound.

The completion error $E = \Pi(aI - \Sigma)\Pi$ obeys $0 \preceq E \preceq a\Pi$. The general-label response envelope (223) therefore gives $\|E\Delta_U\|^2 \le a^2 \Gamma_\rho \tau_V$ for every legal U. The centered-decoder calculation bounds its ideal excess square root by $\|E\Delta_U\|/((1-\rho)\sqrt{\lambda})$. Bounded labels give $\|\Delta_U\| \le 1/\sqrt{\lambda}$, as proved in Theorem 36; hence covariance perturbation adds at most $\omega/((1-\rho)\lambda)$. The reference and output terms follow exactly as in Proposition 33. The triangle inequality in the retained Hessian norm proves (242). Finally, $\widetilde{R} \ge (1-\chi)R$ proves the scale assertion. □

Direction errors and scale normalization. For a subset known to have $p(J) > 0$, put $u = z_J - \theta$ and define $P(u) = uu^\top/\|u\|^2$ for $u \ne 0$, with $P(0) = 0$. If a computed $\widetilde{u} \ne 0$ has a certified forward error $\|u - \widetilde{u}\| \le \eta_J$, then

$$\|P(u) - P(\widetilde{u})\|_{\mathrm{op}} \le \min\left\{1, \frac{\eta_J}{\|\widetilde{u}\|}\right\}. \tag{243}$$

For $u \ne 0$, the projector difference is the sine of the angle between the two one-dimensional spans; this is $\mathrm{dist}(\widetilde{u}, \mathrm{span}\, u)/\|\widetilde{u}\| \le \eta_J/\|\widetilde{u}\|$. If $u = 0$, then $\eta_J \ge \|\widetilde{u}\|$, and the bound remains true. For $\widetilde{u} = 0$ one can use the universal bound one unless $u = 0$ is known exactly. Correctly identified original-p-zero subsets contribute zero with zero error. Averaging these bounds supplies κ in (239), without a factor involving the possibly unbounded raw fit norm. The bound still requires forward-error enclosures and correct support classification; a small interpolation residual alone proves neither.

For example, when $A = X_J$ has full row rank, suppose one has certified $\sigma_{\min}(A) \ge \sigma_* > 0$, $\|A\widetilde{z} - y_J\| \le r_J$, $\|(I - A^\dagger A)\widetilde{z}\| \le h_J$, and $\|\widetilde{\theta} - \theta\| \le \eta_\theta$. Then, for $\widetilde{u} = \widetilde{z} - \widetilde{\theta}$, one may use

$$\eta_J = \sqrt{(r_J/\sigma_*)^2 + h_J^2} + \eta_\theta.$$

Indeed, the row-span part of $\widetilde{z} - z_J$ equals $A^\dagger(A\widetilde{z} - y_J)$, and its orthogonal part equals $(I - A^\dagger A)\widetilde{z}$. This also identifies the necessary minimum-norm control and singular-value lower bound in a residual-based certificate; uncertified floating-point estimates of these quantities remain diagnostics.

The exact orthogonal decomposition

$$u = X_J^\dagger e_J - (I - X_J^\dagger X_J)\theta$$

can avoid subtracting the full fit and reference. Where the relevant vector is nonzero, normalizing e to $e/\|e\|$ or θ to $\theta/\|\theta\|$ before an augmented-feature calculation removes its arbitrary magnitude from that direction calculation, provided the corresponding scalar weight is retained. Stable scaled norm evaluation likewise avoids squaring a raw vector of extreme magnitude. These algebraic scalings do not establish a total-variation certificate, a forward-error enclosure, or exact rank. In particular, a numerically tiny nonzero direction cannot be declared zero solely because it falls below an absolute threshold.

A discontinuity at dependent original rows. The distinction between a mathematical dummy sample and numerical rank uncertainty is essential. Let $n = d = 2$, fix $t = 2\lambda > 0$ and $0 < c < 1$, and for sufficiently small $\gamma \geq 0$ set

$$X_\gamma = \begin{pmatrix} c & 0 \\ c & \gamma \end{pmatrix}, \qquad y = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

These records are bounded. As $\gamma \downarrow 0$, $\theta_\gamma \to 0$, $R_\gamma \to 2/t$, and $\mathcal{Z}_\gamma \to 1 + 2c^2/t$. For $\gamma > 0$, the pair $J = \{1, 2\}$ has independent rows and

$$z_J = \begin{pmatrix} 1/c \\ -2/\gamma \end{pmatrix}, \qquad p_\gamma(J) = \frac{c^2\gamma^2}{t^2\mathcal{Z}_\gamma}.$$

Consequently

$$\begin{aligned} q_\gamma(J) &\longrightarrow \frac{2c^2}{t + 2c^2} > 0, \qquad B_J \longrightarrow e_2 e_2^\top, \\ (\mathcal{V}_\gamma)_{22} &\longrightarrow \frac{4c^2}{t(t + 2c^2)} > 0. \end{aligned} \tag{244}$$

The last limit follows because all other subsets' centered second coordinates tend to zero. At $\gamma = 0$, however, the pair is dependent and its importance contribution is exactly zero; every positive-p fit lies in $\operatorname{span}(e_1)$, so $(\mathcal{V}_0)_{22} = 0$. Thus arbitrarily small feature perturbations can change a positive-probability importance contribution by order R. Treating every numerically rank-deficient pair as a dummy can lose a constant covariance component. A conservative implementation must establish dependence exactly, obtain adequate enclosures for an independent fit, or report numerical failure. The exact rational construction already supplies one way to distinguish these cases. This discontinuity concerns the auxiliary sampling covariance; the ridge optimizer itself remains continuous because $t > 0$.

An optional conservative contribution. For a sampled subset whose support or direction is unresolved, the matrix I is a valid upper bound on its unknown normalized target B_J. Replacing that contribution by I therefore permits a one-sided version of the theorem: it suffices that

$$\frac{1}{s}\sum_j B_{J_j} \preceq \frac{1}{s}\sum_j \widetilde{B}_j + \kappa I$$

on the computational event, and the same proof applies. Such a replacement is retained as an uncertain contribution, rather than discarded or counted as a mathematical dummy. It incurs no one-sided error κ for that draw. If f of the s draws are replaced by I and the other computed contributions are rank-one projectors or zero, their exact arithmetic average has trace at most

$$1 + (d - 1)\frac{f}{s}.$$

The original deterministic trace bound one no longer follows. However, the uncertainty is isotropic, so a sharper bound avoids multiplying its effect by the dimension. Write $\eta = f/s$ and let M_0 be the average of the resolved contributions, using zero for the unresolved draws. Then $M = M_0 + \eta I$ and $\operatorname{tr} M_0 \leq 1 - \eta$. For exact draws and exact resolved contributions, the same single concentration event gives

$$\mathcal{V}/R \preceq 2M_0 + (2\eta + 1/r)I.$$

Preserve the leading m directions of both Σ and M_0. Their common orthogonal complement has covariance tail at most $1/r$ and response tail at most

$$R\left[\frac{2(1 - \eta)}{r} + 2\eta + \frac{1}{r}\right] \leq R(3/r + 2\eta).$$

Consequently the ideal centered decoder satisfies, on that event,

$$F_{S-U}(\overline{w}_U) - F_{S-U}(w_{S-U}) \leq \frac{\Gamma_\rho R}{(1 - \rho)^2\lambda}\left(\frac{3}{r^3} + \frac{2\eta}{r^2}\right) \quad \text{for all legal } U. \tag{245}$$

Thus an unresolved fraction $O(1/r)$ preserves the cubic rank rate without a dimension factor. The uncertainty flag and η may depend on the samples: the concentration event concerns the original

exact contributions, and replacing any of them by upper bounds preserves the inequality. Adding ηI also leaves the selected eigenspaces of M_0 or $\Sigma + M_0$ unchanged. Other sampling and matrix errors enter through the additional terms in (240).

A reported omitted-spectrum bound b still yields (242). Using this fallback in a certified implementation requires the stated law, formation, scale, subspace, and decoder enclosures.

K.2 A SMALLER SIGNED REPRESENTATION BY SHORTING THE FLOOR RESIDUAL

Shorting the positive covariance residual reduces persistent storage while preserving the response-envelope argument. The construction uses the classical generalized Schur complement and shorted operator (Carlson et al., 1974; Anderson & Trapp, 1975). The gain is a smaller representation for a response-aware deletion decoder; the asymptotic minimax rate is unchanged, and utility comparisons with other summaries remain separate.

Proposition 44 (Shorting a covariance upper bound)**.** *Let $\Sigma \succeq 0$, let V have orthonormal columns, and write $\Pi = I - VV^\top$. Suppose*

$$A_0 = aI + FF^\top \succeq \Sigma, \qquad a \geq 0, \qquad \mathrm{range}(F) \subseteq \mathrm{range}(V).$$

Set $G = A_0 - \Sigma$, $R = G^{1/2}V$, and

$$N = GV(V^\top GV)^\dagger V^\top G = G^{1/2}P_R G^{1/2} = LL^\top, \qquad \Sigma_{\mathrm{s}} = A_0 - LL^\top, \tag{246}$$

where P_R projects onto $\mathrm{range}(R)$ and L has $s = \mathrm{rank}(R)$ columns. Then

$$\begin{aligned} &E_{\mathrm{s}} := \Sigma_{\mathrm{s}} - \Sigma \succeq 0, && E_{\mathrm{s}}V = 0, \\ &E_{\mathrm{s}} \preceq \Pi G \Pi = \Pi(aI - \Sigma)\Pi \preceq a\Pi. \end{aligned} \tag{247}$$

Moreover, E_{s} is the largest matrix E in Loewner order such that $0 \preceq E \preceq G$ and $EV = 0$.

Proof. The identity $R(R^\top R)^\dagger R^\top = P_R$ proves the two formulas for N. In particular, $0 \preceq N \preceq G$ and $NV = GV$, so $E_{\mathrm{s}} = G - N \succeq 0$ and $E_{\mathrm{s}}V = 0$. Consequently $E_{\mathrm{s}} = \Pi E_{\mathrm{s}} \Pi \preceq \Pi G \Pi$. The assumption on F gives the equality and last inequality in (247).

For the extremal assertion, decompose G along $\mathrm{range}(V) \oplus \mathrm{range}(V)^\perp$ as $\left[\begin{smallmatrix} A & B^\top \\ B & C \end{smallmatrix}\right]$. Positivity gives $\mathrm{range}(B^\top) \subseteq \mathrm{range}(A)$. The corresponding blocks of E_{s} are zero except for $C - BA^\dagger B^\top$. Any E in the stated class has only a lower-right block D, and the generalized Schur criterion for $G - E \succeq 0$ gives $D \preceq C - BA^\dagger B^\top$. This includes singular A. □

Corollary 45 (Two signed factors for the determinantal envelope)**.** *Use the subspace selected in Theorem 29. Let U consist of the leading m covariance eigenvectors, with eigenvalues $\sigma_1, \ldots, \sigma_m$. If $T_{\mathcal{V}}$ contains the selected leading response eigenvectors, let W be an orthonormal basis of $\mathrm{range}((I - UU^\top)T_{\mathcal{V}})$. Put $r = \mathrm{rank}(W)$, so $k = m + r \leq \min\{d, 2m\}$. For $m < d$ set $a = \sigma_{m+1}$, and set $a = 0$ for $m = d$. Define*

$$F = U\,\mathrm{diag}\big(\sqrt{\sigma_i - a}\big), \qquad A_0 = aI + FF^\top, \qquad G = A_0 - \Sigma.$$

The zero columns of F may be omitted; let $p \leq m$ denote its remaining column count. Applying Proposition 44 with $V = [U, W]$ gives

$$\Sigma_{\mathrm{s}} = aI + FF^\top - LL^\top, \qquad s = \mathrm{rank}(G^{1/2}W) \leq r, \qquad F^\top L = 0. \tag{248}$$

The persistent arrays (F, L, a, θ, n) contain exactly

$$d(p+s) + d + 2 \;\leq\; dk + d + 2 \;\leq\; 2dm + d + 2 \tag{249}$$

numerical entries, excluding the $O(\log(2+d))$ bits encoding their dimensions. They implement the same centered decoder formula with Σ_{s} replacing $\widehat{\Sigma}$. Its ideal excess obeys exactly the bound in (204)*.*

Proof. Spectral flooring gives $0 \preceq G \preceq aI$ and $GU = 0$. Thus $G^{1/2}V$ has the same range as $G^{1/2}W$, and the negative factor can be chosen with $s \leq r$ columns in $U^\perp$. The errors of the two completions satisfy

$$0 \preceq E_{\mathrm{s}} \preceq E_{\mathrm{old}} := \Pi(aI - \Sigma)\Pi \preceq a\Pi.$$

In particular E_{s} annihilates the selected response eigenspace and has operator norm at most a. These are precisely the properties used to bound $\|E\Delta_U\|$ in the DPP proof. Its response envelope and $a \le 1/(m+1)$ therefore give the unchanged cubic risk bound. At $m = d$, $G = 0$ and the completion is exact. □

The new covariance generally differs from the old completion. The Loewner comparison of their errors does not order their actual retained excesses: the retained Hessian and these errors need not commute. Also, the smaller state does not retain W or the original covariance action on W. Query certificates that require those quantities cannot be imported without charging the necessary additional state.

Finite encoding. In Corollary 45, bounded features imply $\|F\|_{\mathrm{op}} \le 1$ and $\|L\|_{\mathrm{op}} \le \sqrt{a} \le 1$. Their orthogonality gives $\|[F, L]\|_{\mathrm{op}} \le 1$, and $p + s \le d$. Let $\eta \in (\sqrt{\lambda\varepsilon}/200, \sqrt{\lambda\varepsilon}/100]$ be public and dyadic. Round all entries of F, L, a at a dyadic step $h \le \eta/(4d)$, and write $H = [F, L]$, $H_q = [F_q, L_q]$ and $J = \mathrm{diag}(I_p, -I_s)$. With $\gamma = dh/2$,

$$\|H_q - H\|_{\mathrm{op}} \le \gamma, \qquad \|a_q I + H_q J H_q^\top - \Sigma_{\mathrm{s}}\|_{\mathrm{op}} \le 2\gamma + \gamma^2 + h/2 \le 2dh \le \eta/2.$$

The decoder uses $a_q I + H_q J H_q^\top + \eta I$. Its perturbation relative to Σ_{s} is positive semidefinite and at most $2\eta I$. Rounding θ to error at most $\lambda\sqrt{\varepsilon}/100$ therefore has exactly the covariance and reference perturbation costs already proved in Theorem 29. The same total excess guarantee holds. No rounded orthogonality is assumed.

More explicitly, if $h = 2^{-b}$ and the reference coordinate step is 2^{-b_θ}, one may take

$$b = \lceil \log_2(4d/\eta) \rceil, \qquad b_\theta = \left\lceil \log_2 \frac{50\sqrt{d}}{\lambda\sqrt{\varepsilon}} \right\rceil.$$

A fixed-width signed encoding then uses at most

$$[d(p+s)+1](b+2) + d(b_\theta + 2) + O(\log(2+d) + \log(2+n))$$

bits. This charges the floor, reference, sample size, dimensions, and both factors. The construction is a finite-bit existence statement even when the exact factors are irrational or their defining positive eigenvalues are arbitrarily small.

Controlling a numerical rank truncation. Pseudoinversion in Nyström approximation is sensitive to numerical rank; see Bucci et al. (2025). The bound below controls its effect on the deleted response in this construction. In the spectral-floor corollary with $m < d$, let T consist of the selected leading response eigenvectors, and let Q_0 have orthonormal columns approximating the range of $G^{1/2}T$. If $\|(I - Q_0 Q_0^\top) G^{1/2} T\|_{\mathrm{op}} \le \tau$, replace L by $G^{1/2} Q_0$. Its error E_τ remains positive semidefinite and at most $G \preceq aI$, while $\|E_\tau T\|_{\mathrm{op}} \le \sqrt{a}\,\tau$. Splitting a deletion response along T and $T^\perp$ gives the explicit ideal bound

$$\sqrt{F_{S-U}(\overline{w}) - F_{S-U}(w_{S-U})} \le \frac{a\sqrt{\Gamma_\rho \lambda_{m+1}(\mathcal{V})} + 2\sqrt{a}\,\tau}{(1-\rho)\sqrt{\lambda}}.$$

For an exact singular value decomposition, discarding singular values of $G^{1/2}T$ at most τ satisfies this condition. Thus a truncation tolerance can be budgeted explicitly, rather than treated as exact rank determination. The Nyström matrix can otherwise be discontinuous at rank changes: for $G = \mathrm{diag}(1, 0)$ and $W = (u, \sqrt{1-u^2})^\top$, it equals G for every $u \neq 0$ but vanishes at $u = 0$. A floating implementation additionally needs enclosures for its square-root, orthogonality, and reconstruction errors to turn this calculation into a certificate.

A rational construction without a covariance eigengap. The signed form also preserves the polynomial bit-operation claim in Theorem 34. Reuse its exact empirical response covariance M, put $H = \Sigma + M$, and use its exactly orthogonal rational approximate diagonalization

$$Q^\top H Q = D + N_0, \qquad D = \mathrm{diag}(Q^\top H Q), \qquad \|N_0\|_{\mathrm{op}} \le 1/(m+1), \qquad \mathrm{tr}\, H \le 2.$$

Order D decreasingly, take V to be the first m columns of Q, and set $t = D_{m+1,m+1}$, $\nu = 1/(m+1)$ and

$$a = t + \nu, \qquad FF^\top = V(D_{1:m,1:m} - tI)V^\top, \qquad A_0 = aI + FF^\top.$$

Then $A_0 \succeq H \succeq \Sigma$: in the Q basis the diagonal part of $A_0 - H$ is at least νI, while its off-diagonal part has operator norm at most ν. Apply Proposition 44 with this V. It gives $0 \preceq E_{\mathrm{s}} \preceq a\Pi$, where $a \le 3/(m+1)$. The original sampled argument supplies the same true response tail $7/(m+1)$. Consequently the existing ideal bound $63\Gamma_\rho/((1-\rho)^2\lambda(m+1)^3)$ and the existing constant 128 in rank selection remain valid.

All of A_0, G, V and $N = GV(V^\top GV)^\dagger V^\top G$ are rational with polynomial encoding lengths. Exact elimination finds the rank $s \le m$ and independent principal columns of the positive semidefinite core $V^\top GV$. A rational $LDL^\top$ factorization of the resulting positive definite principal block expresses a factor of N using rational arithmetic and square roots of positive rationals. These square roots and the corresponding factor of $FF^\top$ can be approximated to any specified dyadic entry accuracy in polynomial bit time. Small positive pivots affect intermediate precision, but have polynomial bit length; no numerical pivot cutoff is part of this existence construction.

Here the two factors need not be orthogonal, but $\|F\|_{\mathrm{op}} \le \sqrt{2}$, $\|L\|_{\mathrm{op}} \le \sqrt{5/2}$, and $\|[F, L]\|_{\mathrm{op}} < 3$. There are at most $2d$ columns. Therefore the original factor grid $h \le \eta/(16d)$ yields covariance operator error at most $8dh \le \eta/2$, followed by the same positive padding and reference rounding. The arrays contain $d(m+s) + d + 2 \le 2dm + d + 2$ numerical entries and obey the original bit bound. The full-rank branch can use the original full covariance encoding. This establishes a rational construction of the signed summary; it does not certify the conditioning decisions of a float64 implementation.

Why this changes the completion. An exact rewrite of the old completion can already save storage. With the basis $[U, W]$ above, whose U block is covariance invariant, define

$$T = \Sigma W - \tfrac{1}{2}W(W^\top \Sigma W + aI).$$

Then $\widehat{\Sigma} = aI + FF^\top + WT^\top + TW^\top$, requiring $d(p+2r) + d + 2$ numerical entries. In general a signed factorization of that same matrix needs more columns than (248).

This obstruction occurs for a realizable DPP example. Let $n = d = 3$, $\lambda = 1/8$, and take records with features $e_1, e_2/2, (e_2 + e_3)/2$ and teacher $v = e_2$. Direct enumeration gives

$$\Sigma = \tfrac{1}{3} \oplus \tfrac{1}{12}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathcal{V} = 0 \oplus \tfrac{1}{961}\begin{bmatrix} 147 & -3 \\ -3 & 57 \end{bmatrix}.$$

For $m = 1$, $U = e_1$, $a = (3 + \sqrt{5})/24$, and the leading response direction W lies in $U^\perp$ but is not a covariance eigenvector. Writing G in the orthonormal tail basis $[W, W_\perp]$ as $\left[\begin{smallmatrix} g & b \\ b & c \end{smallmatrix}\right]$, one has $g > 0$ and $b \neq 0$. The tail block of $\widehat{\Sigma} - aI$ is $\left[\begin{smallmatrix} -g & -b \\ -b & 0 \end{smallmatrix}\right]$ and has rank two. The head contributes a third nonzero eigenvalue, although the selected union has dimension two. Thus a two-column signed factorization with the same scalar floor a cannot reproduce this old completion. In the shorted construction G has rank one and $G^{1/2}W \neq 0$, so $N = G$ and $\Sigma_{\mathrm{s}} = \Sigma$ with two signed columns. This example separates representation costs; it leaves the ordering of realized utility unspecified.

K.3 Selecting a signed correction in the response metric

We select signed corrections by weighting the response envelope with the actual covariance gap. This gives a rank-optimal envelope criterion and a bound on retained excess. The criterion optimizes the envelope, rather than pointwise error relative to spectral or midpoint decoders.

Theorem 46 (Optimal correction of the envelope-weighted gap)**.** *Let $\Sigma \succeq 0$, let $A_0 \succeq \Sigma$, and put $G = A_0 - \Sigma \succeq 0$. Let $B \succeq 0$ and suppose every legal request has response $\Delta_U = w_{S-U} - \theta$ satisfying*

$$\Delta_U \Delta_U^\top \preceq c_U B.$$

Here B is one common encoding-time envelope; $c_U \ge 0$ may depend on the request. Define

$$K = G^{1/2} B G^{1/2}.$$

For an integer $0 \le r \le d$, let H contain orthonormal leading eigenvectors for the largest $\min\{r, \operatorname{rank} K\}$ positive eigenvalues of K, and set

$$L = G^{1/2}H, \qquad N = LL^\top, \qquad \widehat{\Sigma} = A_0 - N, \qquad E = G - N. \tag{250}$$

Write $\mu_{r+1} = \lambda_{r+1}(K)$, with zero understood past the dimension or positive rank. Then

$$0 \preceq E \preceq G, \qquad \lambda_{\max}(B^{1/2}EB^{1/2}) = \mu_{r+1}, \qquad \Delta_U^\top E \Delta_U \le c_U \mu_{r+1}. \tag{251}$$

Moreover the eigenvalue in this display is optimal among all admissible rank-r negative corrections:

$$\min_{\substack{0 \preceq N' \preceq G \\ \operatorname{rank} N' \le r}} \lambda_{\max}\Big(B^{1/2}(G - N')B^{1/2}\Big) = \mu_{r+1}. \tag{252}$$

Proof. Put $P = HH^\top$. The factorization $E = G^{1/2}(I - P)G^{1/2}$ proves the positive orders. The nonzero eigenvalues of $B^{1/2}EB^{1/2}$ are those of $(I - P)K(I - P)$. Its largest eigenvalue is μ_{r+1} by the choice of H, including singular B or G.

The rank-one order $\Delta_U \Delta_U^\top \preceq c_U B$ is equivalent to $\Delta_U = B^{1/2}z$ for some $\|z\|^2 \le c_U$. This follows on the range of B by congruence with $B^{\dagger/2}$; the same order puts Δ_U in that range. The energy bound follows.

For optimality put $D = B^{1/2}GB^{1/2}$, whose nonzero spectrum equals that of K. Every admissible correction gives $D' = B^{1/2}N'B^{1/2} \succeq 0$ of rank at most r. If $\mu_{r+1} = 0$, nonnegativity gives the desired lower bound. Otherwise $r < d$. The span of the leading $r + 1$ eigenvectors of D intersects $\ker D'$ in a nonzero vector. Testing $D - D'$ on a unit vector in this intersection gives $\lambda_{\max}(D - D') \ge \lambda_{r+1}(D) = \mu_{r+1}$. The constructed correction attains this lower bound. □

Proposition 47 (Transfer through the exact ridge response)**.** *In Theorem 46, assume $G \preceq aI$. For a request of fraction α put $\beta = 1 - \alpha > 0$, $C = n^{-1}\sum_U xx^\top$, and $R = \Sigma - C + \beta\lambda I$. The centered decoder uses the true retained half-gradient at θ and covariance $\widehat{\Sigma}$. Its retained excess satisfies*

$$F_{S-U}(\widehat{w}_U) - F_{S-U}(w_{S-U}) \le \frac{c_U}{\beta}\,\psi\!\left(\frac{a}{\beta\lambda}\right)\mu_{r+1} \le \frac{c_U \mu_{r+1}}{4\beta},$$
$$\psi(t) = \begin{cases} t/(1+t)^2, & 0 \le t \le 1, \\ 1/4, & t \ge 1. \end{cases} \tag{253}$$

The true gradient is available from $\theta = w_S$ and the request in exact arithmetic. No retained records are required.

Proof. Let $J = R + E$. Legality gives $R \succeq \beta\lambda I$. The centered error is $-J^{-1}E\Delta_U$, and hence its exact retained excess equals

$$\frac{1}{\beta}\Delta_U^\top E J^{-1} R J^{-1} E \Delta_U.$$

Set $T = E^{1/2}J^{-1}E^{1/2}$. The identity $R = J - E$ gives

$$EJ^{-1}RJ^{-1}E = E^{1/2}(T - T^2)E^{1/2}.$$

Moreover

$$0 \preceq T \preceq E^{1/2}(\beta\lambda I + E)^{-1}E^{1/2} \preceq \frac{a}{\beta\lambda + a} I.$$

The maximum of $s - s^2$ on this last spectral interval is $\psi(a/(\beta\lambda))$, and is at most $1/4$. Thus the exact excess is at most $\beta^{-1}\psi(a/(\beta\lambda))\Delta_U^\top E \Delta_U$. Apply (251). □

Comparison with the unweighted response rule. For a spectral floor, let U be its leading covariance subspace, $P_U^\perp = I - UU^\top$, and $0 \preceq G \preceq aP_U^\perp$. Then

$$\mu_{r+1} \le a\,\lambda_{r+1}(P_U^\perp B P_U^\perp). \tag{254}$$

Indeed, $B^{1/2}GB^{1/2} \preceq aB^{1/2}P_U^\perp B^{1/2}$, and the nonzero spectra can be transferred to the corresponding compressed matrices. Also $\psi(t) \le t$. Consequently (253) is no larger than the earlier

envelope certificate $c_U a^2 \lambda_{r+1}(P_U^{\perp} B P_U^{\perp})/(\beta^2 \lambda)$. The stronger bound accounts for response directions on which the spectral floor is already accurate. It can be arbitrarily sharper than this bound as an abstract matrix statement. This comparison does not order actual excesses of different decoders and does not establish that the gain is attained by a realizable ridge dataset.

For example, if a measured response envelope has the correction $B \preceq B_{\mathrm{emp}} + \eta I$, one may construct the eigenspace using $G^{1/2} B_{\mathrm{emp}} G^{1/2}$. On the event of that order, its residual criterion is at most the corresponding empirical $\mu_{r+1} + \eta a$, since $E \preceq aI$. The assertion requires a certified covariance event; it does not follow from a small observed sample tail alone.

Corollary 48 (Importance sampling with the smaller signed state)**.** *Assume $\lambda > 0$, $0 \le \rho < 1$, $0 < \delta < 1$, $\|x_i\| \le 1$, $|y_i| \le 1$, and $R = r_S/\lambda > 0$. Let M be the empirical importance covariance, normalized so that $\operatorname{tr} M \le R$. Fix $0 \le m < d$, put $r = m + 1$, and suppose the common event*

$$\mathcal{V} \preceq 2M + (R/r) I$$

holds. Use the leading m covariance directions for the spectral floor, and at most m negative-factor columns selected by Theorem 46 with $B = 2M/R + I/r$. With $\beta_0 = 1 - \rho$, every legal deletion of fraction at most $\rho < 1$ satisfies

$$\begin{aligned} F_{S-U}(\widehat{w}_U) - F_{S-U}(w_{S-U}) &\le \frac{3\Gamma_\rho R}{\beta_0 r^2} \psi\!\left(\frac{1}{\beta_0 \lambda r}\right) \\ &\le \min\left\{ \frac{3\Gamma_\rho R}{4\beta_0 r^2}, \frac{3\Gamma_\rho R}{\beta_0^2 \lambda r^3} \right\}. \end{aligned} \tag{255}$$

The state contains at most $2dm + d + 2$ numerical entries. The importance concentration theorem supplies the assumed common event with probability at least $1 - \delta$ when $s \ge \lceil 12 r \log(2d/\delta) \rceil$.

Proof. The floor is $a = \sigma_{m+1}(\Sigma) \le 1/r$. Since $\operatorname{tr}(M/R) \le 1$, eigenvalue interlacing gives $\lambda_{m+1}(B) \le 3/r$, and (254) yields $\mu_{m+1} \le 3a/r \le 3/r^2$. The deletion envelope gives $\Delta_U \Delta_U^{\top} \preceq \Gamma_\rho \mathcal{V} \preceq \Gamma_\rho R B$ simultaneously. Apply Proposition 47 and use that ψ is nondecreasing and $\psi(t) \le \min\{t, 1/4\}$. The state count follows from the two factor ranks. For $m \ge 1$ the concentration proposition applies directly. For $m = 0$, the moment bound $\mathcal{V} \preceq RI$ makes the assumed event deterministic, since $M \succeq 0$ and $r = 1$. □

The zero-residual case $R = 0$ has $y = 0$ by ridge stationarity, so the zero predictor answers every deletion exactly. Selecting a full covariance head is also exact. The corollary preserves the earlier cubic rate with smaller factors and gives a sharper certificate before its cubic regime; it does not claim a new minimax rate. Choosing a zero covariance head and allocating all columns to the negative factor remains valid in Theorem 46, but no $a \le 1/(m+1)$ decay with the number of negative columns follows for that choice.

Storage and finite precision. For $A_0 = aI + FF^{\top}$ with p positive-factor columns, the persistent numerical state remains (F, L, a, θ, n), containing at most $d(p + r) + d + 2$ entries. G, B, K, and H are temporary encoder objects. There is no need to store $G^{\dagger/2} H$, the response basis whose shorted gap gives (250). For a spectral floor, $F^{\top} L = 0$ and $\|F\|_{\mathrm{op}}, \|L\|_{\mathrm{op}} \le 1$ under bounded features. Thus signed factor rounding followed by positive padding is stable. In this spectral-floor setting, for labels $|y_i| \le 1$ and $0 < \lambda \le 1$, suppose the ideal uniform bound is at most $\varepsilon/4$, with $0 < \varepsilon \le 1$. Put $\beta_0 = 1 - \rho$ and choose a public dyadic $\eta \le \beta_0 \lambda \sqrt{\varepsilon}/64$ within a factor two. Rounding the factors and floor at entry step at most $\eta/(8d)$, then adding ηI, produces covariance $\widehat{\Sigma} + K_0$ with $0 \preceq K_0 \preceq 2\eta I$. Round θ to Euclidean error at most $\beta_0 \sqrt{\lambda \varepsilon}/64$. Bounded labels give $\|\Delta_U\| \le 1/\sqrt{\lambda}$ and the ideal guarantee gives $\|\widehat{w}_U - \theta\| \le 3/(2\sqrt{\lambda})$. The resolvent identity therefore bounds the excess square root added by covariance rounding by $3\eta/(\beta\lambda)$. The reference perturbation adds at most $3\|\theta_q - \theta\|/(\beta\sqrt{\lambda})$, since $\|\widehat{\Sigma} + \lambda I + K_0\|_{\mathrm{op}} \le 3$. Their sum with $\sqrt{\varepsilon}/2$ is smaller than $\sqrt{\varepsilon}$. The resulting state uses

$$O\!\left(d(p + r + 1) \log\!\left(2 + \frac{d}{(1-\rho)\lambda\sqrt{\varepsilon}} \right) + \log(2 + n) \right)$$

bits. An approximate eigenspace needs a certified bound on the unselected response-gap spectrum; a small eigenvector residual alone does not certify that it is the leading eigenspace. The optimality claim is in exact arithmetic, independent of such an eigenspace's computation cost.

Relation to spectral approximation and empirical comparison. The rank-optimality proof is an elementary spectral approximation argument in the tradition of Eckart & Young (1936), and the shorted matrix identity is classical (Anderson & Trapp, 1975). The deletion-specific content is the response-weighted choice in the compact signed summary together with (253). The determinantal envelope's exponential scalar remains in the resulting bound. Its practical claim must be evaluated at the same persistent scalar budget and with the same response samples as the unweighted signed correction, and against the centered legal spectral midpoint. Failure to improve actual retained excess in cases where (254) is much sharper would show that optimizing this envelope remains a poor proxy for the requests, despite the valid certificate.

K.3.1 Exact response loss and a limitation of envelope selection

The envelope objective in Theorem 46 is rank-optimal for its stated quadratic form. The exact retained loss has a different geometry. The following reformulation isolates that difference, and a bounded ridge family shows that it can produce an unbounded gap in worst-case utility, even with an exact common response envelope.

Proposition 49 (Exact utility at a fixed retained curvature)**.** *Fix $R \succ 0$, $G, B \succeq 0$, $\beta > 0$, $c \geq 0$, and an integer $0 \leq r \leq d$. For an admissible correction $0 \preceq N \preceq G$ with $\operatorname{rank} N \leq r$, put*

$$E = G - N, \qquad T = R^{-1/2} E R^{-1/2}, \qquad Q = T(I+T)^{-1},$$

and define $T_0 = R^{-1/2} G R^{-1/2}$, $Q_0 = T_0(I+T_0)^{-1}$, and $D = R^{1/2} B R^{1/2}$. The map from N to Q is a bijection onto

$$0 \preceq Q \preceq Q_0, \qquad \operatorname{rank}(Q_0 - Q) \leq r.$$

Moreover, the exact worst loss over the response ellipsoid is

$$\sup_{\Delta\Delta^\top \preceq cB} \frac{1}{\beta} \Delta^\top E (R+E)^{-1} R (R+E)^{-1} E \Delta = \frac{c}{\beta} \|Q D^{1/2}\|_{\mathrm{op}}^2. \tag{256}$$

If R, G, B commute, write their eigenvalues in a common eigenbasis as R_i, g_i, b_i, and arrange the scores

$$s_i = \frac{b_i R_i g_i^2}{(R_i + g_i)^2}$$

in decreasing order $s_{(1)} \geq \cdots \geq s_{(d)}$. The minimum of (256) *over all admissible corrections is $c s_{(r+1)}/\beta$, with zero understood past d. It is attained by correcting the r coordinates with largest scores.*

Proof. The identity $Q = I - (I+T)^{-1}$ preserves positive order, and its inverse is $T = (I-Q)^{-1} - I$. This inverse is defined throughout $0 \preceq Q \preceq Q_0 \prec I$ and also preserves positive order. The resolvent identity

$$Q_0 - Q = (I+T)^{-1}(T_0 - T)(I+T_0)^{-1}$$

preserves the rank of $T_0 - T$, which equals the rank of N. This proves the bijection. Direct congruence gives

$$E(R+E)^{-1} R (R+E)^{-1} E = R^{1/2} Q^2 R^{1/2}.$$

Maximizing this quadratic form over the ellipsoid proves (256), including singular B.

In the commuting case, $Q_0 D^{1/2}$ is diagonal and has squared singular values s_i. Every feasible Q subtracts from this matrix the rank-at-most-r matrix $(Q_0 - Q)D^{1/2}$, so the spectral approximation lower bound is $s_{(r+1)}$. Setting the selected diagonal entries of Q to zero attains it and maps back to the stated correction of G. □

This is a resolvent reformulation followed, in the commuting case, by the classical spectral approximation argument. It is not a new low-rank approximation principle. In a deletion problem $R = \Sigma - C + \beta\lambda I$ varies with the request, so this fixed-R result alone does not supply an encoder rule for all legal requests. In particular, its scores contain a curvature factor that is absent from the envelope scores $b_i g_i$.

Proposition 50 (Envelope optimality has no uniform utility factor)**.** *There are bounded ridge datasets in fixed dimension $d = 4$, with $\rho = \lambda = 1/n$, an exact one-dimensional spectral head, and an exact common response envelope, for which every rank-one minimizer of the envelope objective has worst-case retained excess at least $n/128$ times that of another admissible rank-one correction. Both summaries have one positive and one negative factor column. The comparison includes every legal deletion from the dataset. For the same family, if the response ellipsoid is paired with every $0 \preceq C \preceq \Sigma$ satisfying $\operatorname{tr} C \le \rho$ at $\beta = 1 - \rho$, the corresponding worst-loss ratio is greater than $n/96$.*

Proof. Let $n \ge 256$ be divisible by 16, and put $\lambda = \rho = 1/n$. Use two copies of e_1 with labels $+\eta_n, -\eta_n$; $n/8$ copies of e_2, half with each label $+1, -1$; $n/4$ copies of e_3 and $3n/8$ copies of e_4, all with label zero; and $n/4 - 2$ zero records with label zero. Define

$$R_1 = \frac{2n-1}{n^2}, \qquad R_2 = \frac{n^2-8}{8n^2}, \qquad \eta_n = \frac{R_1}{R_2} = \frac{8(2n-1)}{n^2-8}.$$

All feature norms and label magnitudes are at most one, since $\eta_n \le 1$ for $n \ge 16$. The normalized moments satisfy

$$\Sigma = \operatorname{diag}(2/n, 1/8, 1/4, 3/8), \qquad b = 0, \qquad \theta = w_S = 0.$$

Use the exact head e_4, the floor $a = 1/4$, and the factor $F = \sqrt{1/8}\, e_4$. Their gap is

$$G = \operatorname{diag}(g_1, g_2, 0, 0), \qquad g_1 = \frac{1}{4} - \frac{2}{n} > g_2 = \frac{1}{8}.$$

Every nonempty legal request deletes one record, so $\beta = (n-1)/n$. A deletion of a nonzero-labelled record in coordinate i has retained curvature R_i in that coordinate and response $-y_i e_i/(nR_i)$. Thus all nonzero responses are $\pm h_n e_1$ and $\pm h_n e_2$, where $h_n = 1/(nR_2)$. All other responses, including the empty request, are zero. Hence

$$B = h_n^2 \operatorname{diag}(1, 1, 0, 0), \qquad \Delta_U \Delta_U^\top \preceq B \quad \text{for every legal } U. \tag{257}$$

This B has minimum trace among all simultaneous envelopes with right-hand scalar one: the two coordinate responses force the first two diagonal entries to be at least h_n^2.

The response-gap rule corrects coordinate one, giving $N_1 = g_1 e_1 e_1^\top$. In fact every minimizer of the envelope objective must leave coordinate two uncorrected. To see this, all admissible corrections are supported on the first two coordinates. The optimal objective is $h_n^2 g_2$, so its residual must satisfy $E \preceq g_2 I$ on this support. Consequently

$$N \succeq \operatorname{diag}(g_1 - g_2, 0).$$

A positive semidefinite rank-one N satisfying this order has range $\operatorname{span}(e_1)$, and therefore is $N = t e_1 e_1^\top$ with $g_1 - g_2 \le t \le g_1$.

For the saturated rule N_1, the coordinate-one requests are exact. For the competing correction $N_2 = g_2 e_2 e_2^\top$, the coordinate-two requests are exact. The common denominator for the uncorrected coordinate is

$$j = R_1 + g_1 = R_2 + g_2 = \frac{1}{4} - \frac{1}{n^2}.$$

Their exact worst excesses over all legal requests are therefore

$$\mathcal{R}_S(N_1) = \frac{h_n^2 R_2 g_2^2}{\beta j^2}, \qquad \mathcal{R}_S(N_2) = \frac{h_n^2 R_1 g_1^2}{\beta j^2}.$$

Every envelope minimizer has excess at least $\mathcal{R}_S(N_1)$ because its coordinate-two loss is unchanged. Finally,

$$\frac{\mathcal{R}_S(N_1)}{\mathcal{R}_S(N_2)} = \frac{n^2(n^2-8)}{32(2n-1)(n-8)^2} \ge \frac{n}{128}, \qquad \frac{\mathcal{R}_S(N_1)}{\mathcal{R}_S(N_2)} \sim \frac{n}{64}. \tag{258}$$

For the inequality use $R_2 \ge 1/16$, $R_1 \le 2/n$, $g_1 \le 1/4$, and $g_2 = 1/8$.

For the last assertion, let $\mathcal{U}_\rho(N)$ denote the supremum in (256), also maximized over the displayed constraints on C, with $R = \Sigma - C + \beta\lambda I$ and $c = 1$. For a rank-one residual $E = ge_je_j^\top$, the resolvent identity reduces its loss to

$$\frac{\Delta_j^2}{\beta}\frac{g^2\kappa_j}{(\kappa_j+g)^2}, \qquad \kappa_j = \frac{1}{e_j^\top R^{-1}e_j}.$$

Put $t = \beta\lambda$. Since $C \preceq \rho I$ and $\Sigma + (t-\rho)I \succ 0$ for this family,

$$\sigma_j + t - \rho \le \kappa_j \le \sigma_j + t.$$

Every intermediate value is attained by $C = ze_je_j^\top$, $0 \le z \le \rho$, for $j = 1, 2$. In coordinate two this interval contains g_2, where the scalar loss is maximal. In coordinate one the entire interval lies below g_1, so its maximum occurs at $C = 0$. Consequently

$$\mathcal{U}_\rho(N_1) = \frac{h_n^2}{32\beta}, \qquad \mathcal{U}_\rho(N_2) = \frac{h_n^2}{\beta}\frac{(n-8)^2(3n-1)}{(n^2+4n-4)^2} < \frac{3h_n^2}{\beta n}.$$

The ratio is greater than $n/96$. Every envelope minimizer has at least the first value: choose $C = te_2e_2^\top$ and $\Delta = h_ne_2$. □

The gap is $\Omega(1/\lambda)$ with $\rho/\lambda = 1$ and fixed dimension. It concerns the envelope selection criterion, including all of its minimizers; it is not a lower bound for arbitrary summaries or for the optimal robust utility. It also does not order requests pointwise: the competing correction loses exactness on coordinate one. Indeed any admissible rank-one correction that preserves exactness on that response must equal N_1. The relaxed comparison treats C and Δ independently and does not assert that every such pair is realizable; the preceding actual-deletion comparison requires no such relaxation.

Both constructions store (F, L, a, θ, n) with the same 14 numerical entries; only L changes. The envelope is an encoder object and needs no persistent storage. The separation also survives a common finite precision. Choose a public dyadic $\zeta_n \le n^{-3}$ within a factor two, round the factors at entry step $\zeta_n/(8d)$, and use the positive padding from the signed-factor rounding argument. Store $\theta = 0$ exactly. Both ideal worst losses are below $1/4$, so the existing resolvent bound gives a uniform perturbation of their excess square roots of at most

$$\delta_n \le \frac{3\zeta_n}{\beta\lambda} < \frac{4}{n^2}.$$

Write $\mathcal{R}_{\mathrm{good}} = \mathcal{R}_S(N_2)$ and $\mathcal{R}_{\mathrm{bad}} = \mathcal{R}_S(N_1)$. Since $h_n \ge 8/n$, $R_1 \ge 1/n$, $g_1 \ge 1/8$, $j \le 1/4$, and $\beta \le 1$, one has $\mathcal{R}_{\mathrm{good}} \ge 16/n^3$. Thus $\delta_n \le \sqrt{\mathcal{R}_{\mathrm{good}}}/16$ for $n \ge 256$; the same bound holds relative to $\sqrt{\mathcal{R}_{\mathrm{bad}}}$. The rounded states consequently satisfy

$$\frac{\mathcal{R}_{\mathrm{bad,q}}}{\mathcal{R}_{\mathrm{good,q}}} \ge \left(\frac{15}{17}\right)^2 \frac{n}{128} \ge \frac{n}{192}.$$

The common precision uses $O(\log n)$ bits for these fixed-dimensional states. This is an existence statement at a specified common precision, not a comparison for a fixed floating-point format. There is no sampling approximation, empirical covariance event, or response-tail estimation in this obstruction.

K.3.2 Exact reference utility versus paired legal requests

The preceding obstruction concerns the response-gap objective. There is also a limitation of optimizing the exact reference utility over a common ellipsoid: the ellipsoid can change which correction is best for the actual response set, even when retained curvature varies arbitrarily little. The underlying loss from replacing a point set by an ellipsoid is known in subspace approximation. In particular, Iwen & Krahmer (2013, Section 1.1) discuss a square-root ambient-dimension loss in distance, hence a linear loss in squared distance, for an ellipsoid-based subspace approximation. The result below realizes such a loss with bounded ridge records and legal deletions, and covers every exact reference-utility minimizer, including nonsaturated corrections.

Theorem 51 (Exact reference utility need not approximate actual requests)**.** *For every integer $r \geq 1$, there is a bounded ridge dataset with $d = r+2$, $n = 4(r+1)$, $\lambda = 1$, and deletion budget $\rho = 1/n$ with the following properties. Its zero-head spectral floor has gap $G \succeq 0$, and its legal responses have a unique minimum-trace common envelope B, with right-hand scalar one. For*

$$\mathcal{C}_r(G) = \{N : 0 \preceq N \preceq G,\ \operatorname{rank} N \leq r\}, \qquad R_+ = \Sigma + I,$$

define the exact reference objective and actual worst-deletion loss by

$$\ell_c(N) = \lambda_{\max}\left[B^{1/2} W_{cR_+}(G-N) B^{1/2}\right], \qquad c > 0,$$
$$\mathcal{R}_S(N) = \max_{\substack{U \subseteq S \\ |U| \leq 1}} \frac{1}{\beta_U} \Delta_U^\top W_{R_U}(G-N)\Delta_U.$$

There is a correction $N_{\mathrm{bal}} \in \mathcal{C}_r(G)$ such that, for every $c > 0$ and every minimizer N_c of ℓ_c,

$$\mathcal{R}_S(N_c) \geq \frac{r+1}{4}\,\mathcal{R}_S(N_{\mathrm{bal}}) \geq \frac{r+1}{4} \min_{N \in \mathcal{C}_r(G)} \mathcal{R}_S(N) > 0. \tag{259}$$

Every legal request satisfies

$$(1-\eta_n)R_+ \preceq R_U \preceq R_+, \qquad \eta_n = \frac{2}{n+2} \longrightarrow 0. \tag{260}$$

The positive reference resolvent has condition number one for every $c > 0$. Both corrections use the same state format, with r negative factor columns, no positive factor columns, and $dr + d + 2$ numerical entries.

Proof. Put $p = r+1 \geq 2$, so $d = p+1$ and $n = 4p$. For each $1 \leq i \leq p$, use two records with feature e_i and labels $+a_i, -a_i$, where $a_1 = 1/2$ and $a_i = 1$ for $i \geq 2$. Add $2p$ copies of e_{p+1} with label zero. Every feature norm and label magnitude is at most one, and

$$\Sigma = \operatorname{diag}\left(\frac{2}{n} I_p, \frac{1}{2}\right), \qquad b = 0, \qquad \theta = w_S = 0. \tag{261}$$

The zero-head spectral floor is $A_0 = \delta I$ with $\delta = \sigma_1(\Sigma) = 1/2$. Thus

$$G = A_0 - \Sigma = g I_p \oplus 0, \qquad g = \frac{1}{2} - \frac{2}{n} = \frac{p-1}{2p} \in [1/4, 1/2).$$

This is the admissible zero-head branch; the correction budget r is allocated entirely to the negative factor.

Every nonempty legal request deletes one record, so $\beta = 1 - 1/n$. Deleting a labelled record in coordinate $i \leq p$ gives

$$(R_U)_{ii} = \frac{2}{n} - \frac{1}{n} + \beta = 1, \qquad \Delta_U = -\frac{y_i}{n} e_i.$$

Deleting any of the zero-labelled records gives zero response, as does the empty request. Consequently

$$B = \frac{1}{n^2} \operatorname{diag}(1/4, 1, \ldots, 1, 0) \tag{262}$$

is a common envelope for every legal request. Any common envelope M must have $M_{ii} \geq a_i^2/n^2$ for $i \leq p$, and $M_{p+1,p+1} \geq 0$. Therefore $\operatorname{tr} M \geq \operatorname{tr} B$. Equality forces these diagonal equalities. For each $i \leq p$, the positive semidefinite matrix $M - (a_i^2/n^2) e_i e_i^\top$ then has a zero ith diagonal entry, so its ith row and column vanish. Together with the last diagonal entry zero, this proves $M = B$ and establishes uniqueness.

All feasible corrections are supported on the first p coordinates. On this support let

$$s = 1 + \frac{2}{n}, \qquad R_c = csI_p, \qquad Q = E(csI_p + E)^{-1}, \qquad E = gI_p - N, \qquad q = \frac{g}{cs+g}.$$

The positive reference resolvent is $Q_0 = qI_p$, and Proposition 49 gives $\ell_c(N) = cs\lambda_{\max}(QBQ)$, with B restricted to this support. Since $\operatorname{rank} N \le p-1$, choose a unit $u \in \ker N$ on the support. Then $Eu = gu$ and $Qu = qu$, whence

$$\ell_c(N) \ge cs\, u^\top QBQu = csq^2 u^\top Bu \ge \frac{csq^2}{4n^2}. \tag{263}$$

The feasible correction $N_{\mathrm{diag}} = g\sum_{i=2}^{p} e_i e_i^\top$ attains this lower bound. For any minimizer, equality in the last inequality forces $u = \pm e_1$, because $1/(4n^2)$ is the unique smallest eigenvalue of B on the active support. In particular every minimizer obeys $N_c e_1 = 0$ and $(G - N_c)e_1 = ge_1$. This argument imposes no saturation assumption on N_c. It also shows that every minimizer has rank $p-1$ on the support.

For a coordinate-one deletion, both the true retained curvature and the error preserve e_1. The exact excess therefore gives

$$\mathcal{R}_S(N_c) \ge L_{\mathrm{bad}} := \frac{g^2}{4\beta n^2(1+g)^2}. \tag{264}$$

For the competing correction put

$$u_{\mathrm{bal}} = p^{-1/2}\sum_{i=1}^{p} e_i, \qquad N_{\mathrm{bal}} = g(I_p - u_{\mathrm{bal}}u_{\mathrm{bal}}^\top) \oplus 0.$$

It has rank $p-1$ and leaves $E_{\mathrm{bal}} = gu_{\mathrm{bal}}u_{\mathrm{bal}}^\top$. For any labelled deletion, the retained curvature on the first p coordinates is $(1+1/n)I_p - e_i e_i^\top/n$. Its effective curvature in direction u_{bal} is the same for every i:

$$\kappa = \frac{1}{u_{\mathrm{bal}}^\top R_U^{-1} u_{\mathrm{bal}}} = \frac{1+1/n}{1+1/(np)} \in [1, 1+1/n]. \tag{265}$$

For any unit u, the rank-one resolvent identity gives

$$W_R(guu^\top) = \frac{g^2\kappa_u}{(\kappa_u+g)^2}uu^\top, \qquad \kappa_u = (u^\top R^{-1}u)^{-1}.$$

It follows that the worst loss of the balanced correction over all legal requests is exactly

$$\mathcal{R}_S(N_{\mathrm{bal}}) = L_{\mathrm{bal}} := \frac{g^2\kappa}{\beta n^2 p(\kappa+g)^2} \le \frac{g^2}{\beta n^2 p(1+g)^2}. \tag{266}$$

The inequality holds because $\kappa \ge 1 > g$ and $t/(t+g)^2$ decreases for $t > g$. Equations (264)–(266) prove the stated separation; the exact comparison is

$$\frac{L_{\mathrm{bad}}}{L_{\mathrm{bal}}} = \frac{p}{4}\frac{(\kappa+g)^2}{\kappa(1+g)^2} \ge \frac{p}{4}.$$

The minimum actual loss is positive. Indeed, zero loss for every labelled deletion would force $Ee_i = 0$ for all $i \le p$, hence $N = gI_p$ on the active support, which has rank p. The feasible set is compact and the finite maximum of the continuous request losses attains its minimum, so the absence of a zero-loss feasible correction proves strict positivity.

For a singleton deletion with feature x, write $R_U = R_+ - (xx^\top + I)/n$. All matrices here are diagonal. The largest eigenvalue of $R_+^{-1/2}(xx^\top + I)R_+^{-1/2}/n$ is $2/(n+2)$ for an active-coordinate deletion and $4/(3n)$ for a dummy-coordinate deletion. Since $n \ge 8$, their maximum is $2/(n+2)$. This proves (260); the empty request is immediate. The condition number of $Q_0 = qI_p$ is one.

Finally, factor either correction as $N = LL^\top$ with $p-1 = r$ columns and store (L, δ, θ, n). This uses $dr + d + 2 = p^2 + p + 2$ numerical entries. The matrices B, R_+, G and the vector u_{bal} are encoding-time objects; the balanced direction is represented through L and needs no separate stored basis. Both encoders may use the same original records and the same exact envelope. No additional retained-data access or response oracle is supplied to the competing decoder. □

The curvature comparison can thus approach equality while every exact reference minimizer has an $\Omega(r)$ gap on actual requests. The balanced reference scale in Theorem 55 gives a product-utility

factor $(1-\eta_n)^{-1} = 1 + 2/n$ on this family, and the theorem above covers that scale as well as R_+ itself. The oblique reference construction also attains its exact reference optimum here, because the positive resolvent has condition number one. Improving its optimization accuracy does not address this example.

This separation identifies distortion from the common-ellipsoid enlargement of the discrete response set. It does not isolate a loss caused only by associating responses with different curvatures; freezing the active curvatures at a common scalar already gives the same order of separation. The construction uses general bounded labels, not a noiseless teacher. It is a comparison within the stated covariance-correction class and centered one-sided decoder, not an information lower bound for arbitrary summaries or an ordering of legal midpoint decoders.

Corollary 52 (The reference-to-actual gap survives common rounding)**.** *For the dataset in Theorem 51, every ideal reference minimizer and the balanced competitor can be rounded at one common public precision, with the same $dr + d + 2$ numerical entries and $O(\log n)$ bits per entry, so that their rounded centered decoders satisfy*

$$\frac{\mathcal{R}_{S,q}(N_c)}{\mathcal{R}_{S,q}(N_{\mathrm{bal}})} \geq \frac{r+1}{6}. \tag{267}$$

Here reference optimality is asserted before rounding. The corollary does not characterize minimizers of a quantized reference objective.

Proof. Use $p = r + 1$, $n = 4p$, and the factors from the preceding proof. For both states, $\|L\|_{\mathrm{op}} \leq \sqrt{g} \leq 1$. Choose a public dyadic $\zeta_n \leq 1/(64n^3)$ within a factor two, and round every factor entry to its nearest multiple of a common dyadic step $h \leq \zeta_n/(8d)$ within a factor two. Store $\theta = 0$ and n exactly, and replace the floor $1/2$ by $1/2 + \zeta_n$. If L_q is the rounded factor, then

$$\|L_q - L\|_{\mathrm{op}} \leq \|L_q - L\|_{\mathrm{F}} \leq \frac{dh}{2} \leq \frac{\zeta_n}{16}, \qquad \|L_q L_q^\top - LL^\top\|_{\mathrm{op}} \leq \frac{\zeta_n}{4}.$$

Consequently the rounded upper covariance is the ideal upper covariance plus $K_0 = \zeta_n I - (L_q L_q^\top - LL^\top)$, with $0 \preceq K_0 \preceq 2\zeta_n I$. Both rounded states remain valid covariance upper summaries.

For completeness, every ideal request loss is at most $1/(\beta n^2) < 1/4$. This follows from $W_R(E) \preceq R$ and the displayed coordinate responses. In the signed-factor rounding bound of Appendix K.3, the reference-rounding term is therefore zero and the covariance term changes the excess square root, uniformly over legal requests, by at most

$$\delta_n \leq \frac{3\zeta_n}{\beta\lambda} < 4\zeta_n \leq \frac{1}{16n^3},$$

using $\lambda = 1$ and $\beta \geq 7/8$. The empty response remains exact. The same bound applies to the square root of the maximum request loss, by the triangle inequality for the finite maximum. The formulas above also give

$$L_{\mathrm{bad}} \geq \frac{1}{144n^2}, \qquad L_{\mathrm{bal}} \geq \frac{1}{64n^2 p}.$$

For the second bound use $g \geq 1/4$, $\kappa \geq 1$, $\kappa + g < 2$, and $\beta \leq 1$; the first uses $g \geq 1/4$ and $1 + g \leq 3/2$. Since $n = 4p \geq 8$, these lower bounds imply $\delta_n \leq \sqrt{L_{\mathrm{bad}}}/16$ and $\delta_n \leq \sqrt{L_{\mathrm{bal}}}/16$. Hence

$$\frac{\mathcal{R}_{S,q}(N_c)}{\mathcal{R}_{S,q}(N_{\mathrm{bal}})} \geq \left(\frac{15}{17}\right)^2 \frac{L_{\mathrm{bad}}}{L_{\mathrm{bal}}} \geq \left(\frac{15}{17}\right)^2 \frac{p}{4} > \frac{p}{6}.$$

All source features, labels, and λ lie on a fixed rational grid. Rounded factors and the padded floor have bounded magnitude and require $O(\log n)$ bits per entry, since $d = p + 1 = O(n)$. The count, the common precision, and the state format are the same for both summaries. This is a finite-bit existence comparison for the rounded ideal states, rather than a statement about a fixed floating-point format or exact minimization on the factor grid. □

A computational obstruction to an exact request oracle. The finite-bank results leave a separate question: can an encoder find a worst legal request without enumerating the subsets? The following reduction from unweighted Max-Cut (Garey et al., 1976) shows that exact separation is already hard for a fixed decoder, even when all retained curvatures approach one common matrix.

Theorem 53 (Hardness of exact worst-request evaluation)**.** *For bounded rational ridge datasets with $\lambda = 1$, deciding whether the worst retained excess over deletions of at most k records exceeds a given rational threshold is NP-hard, even for the zero-head spectral-floor decoder $A_0 = \sigma_1(\Sigma)I = \frac{1}{2}I$, zero correction, and original optimizer $\theta = 0$. The instances can lie on a common dyadic input grid with $O(\log n)$ bits per coordinate and satisfy*

$$(1-\eta_n)(\Sigma + I) \preceq R_U \preceq \Sigma + I \quad \textit{for all } |U| \le k, \qquad \eta_n \longrightarrow 0.$$

The reduction also makes evaluation to a specified inverse-polynomial additive accuracy, or finding a worst request, NP-hard. The decoder's persistent state is only $(a, \theta, n) = (1/2, 0, n)$.

Proof. Let a simple unweighted graph have $m \ge 2$ vertices and $d-1$ edges, and let $B \in \{-1, 0, 1\}^{(d-1)\times m}$ be an oriented incidence matrix. Choose a dyadic $c \in [1/(20m), 1/(10m)]$, and set

$$n = 2m^4, \qquad k = m, \qquad (x_i, y_i) = ((cB_{:i}, 0), 1) \quad (1 \le i \le m).$$

Add m^4 copies of $(e_d, 0)$ and $m^4 - m$ copies of $(0, 0)$. Each graph feature has norm at most $c\sqrt{m-1} < 1$, and all features have norm at most one. Since $B\mathbf{1} = 0$, the original cross moment vanishes and $\theta = 0$. The graph Laplacian bound $\|B^\top B\|_{\mathrm{op}} \le 2(m-1)$ gives

$$\Sigma = \Sigma_g \oplus \tfrac{1}{2}, \qquad 0 \preceq \Sigma_g = \frac{c^2}{n} BB^\top \preceq \epsilon_m I, \qquad \epsilon_m = \frac{1}{100m^5}. \tag{268}$$

Consequently $\sigma_1(\Sigma) = 1/2$. The zero-head spectral floor is exactly $A_0 = \frac{1}{2}I$, with $E = (\frac{1}{2}I - \Sigma_g) \oplus 0 \succeq 0$.

Fix a legal request U, and put

$$J = U \cap \{1, \dots, m\}, \qquad h_U = \sum_{i \in U} y_i (x_i)_{1:d-1} = cB\mathbf{1}_J, \qquad \|h_U\|^2 = c^2 \operatorname{cut}(J).$$

Write $\beta = 1 - |U|/n$ and $\beta_0 = 1 - m/n \ge 15/16$. All source and deletion covariance matrices are block diagonal in the graph and last coordinates. Deleted $(e_d, 0)$ records affect β and the last curvature block but contribute no label cross moment. Both the true and decoded optimizers therefore have last coordinate zero. Let C_g be the graph block of C_U. On this block, the retained curvature and decoder's normal matrix are, respectively,

$$R = \beta I + \Sigma_g - C_g, \qquad T = (\beta + 1/2)I - C_g, \qquad M = R^{-1} - T^{-1}.$$

Their graph coordinates are $-R^{-1}h_U/n$ and $-T^{-1}h_U/n$. Their full retained excess is therefore

$$L(U) = \frac{1}{\beta n^2} h_U^\top MRMh_U. \tag{269}$$

By legality and (268),

$$\beta I \preceq R \preceq (\beta + \epsilon_m)I, \qquad (\beta + 1/2 - \epsilon_m)I \preceq T \preceq (\beta + 1/2)I.$$

Inverse order consequently gives

$$M_{\mathrm{lo}} I \preceq M \preceq M_0 I, \qquad M_0 = \frac{1}{2\beta(\beta + 1/2)},$$
$$M_{\mathrm{lo}} = \frac{1}{\beta + \epsilon_m} - \frac{1}{\beta + 1/2 - \epsilon_m} \ge (1 - 5\epsilon_m) M_0.$$

For the last step, dividing by M_0 gives

$$\frac{M_{\mathrm{lo}}}{M_0} = \frac{1 - 4\epsilon_m}{1 + \epsilon_m(1/2 - \epsilon_m)/[\beta(\beta + 1/2)]} \ge 1 - 5\epsilon_m,$$

using $0 \le (1/2 - \epsilon_m)/[\beta(\beta + 1/2)] \le 1$. Applying the bounds on R to Mh_U in (269) now yields

$$(1 - 10\epsilon_m) C_\beta \operatorname{cut}(J) \le L(U) \le (1 + 2\epsilon_m) C_\beta \operatorname{cut}(J),$$
$$C_\beta = \frac{c^2}{4\beta^2(\beta + 1/2)^2 n^2}. \tag{270}$$

Here $(1-5\epsilon_m)^2 \geq 1-10\epsilon_m$ and $1+\epsilon_m/\beta \leq 1+2\epsilon_m$.

Let $H_{\max}$ be the graph's maximum cut value, and put $C_* = C_{\beta_0}$. Every graph-vertex subset can be supplemented with zero records to obtain a legal request of exactly $k = m$ records. Conversely, every request has $\beta \geq \beta_0$, and $C_\beta \leq C_*$. Thus

$$(1-10\epsilon_m)H_{\max} \leq \frac{\max_{|U|\leq k} L(U)}{C_*} \leq (1+2\epsilon_m)H_{\max}.$$

Both bounds differ from $H_{\max}$ by less than $1/4$, since $H_{\max} \leq m^2/2$ and $10\epsilon_m m^2/2 = 1/(20m^3) < 1/4$. For an integer target K, it follows that

$$H_{\max} \geq K \quad \Longleftrightarrow \quad \max_{|U|\leq k} L(U) > C_*(K-1/2).$$

This proves the threshold-evaluation reduction. Its gap also permits additive evaluation error at most $C_*/8$, which is inverse polynomial in m, hence in the explicit input size. Given a worst request, its rational loss can be evaluated by rational matrix solves in polynomial time. The construction has polynomial size, uses one dyadic grid of $O(\log m)$-bit coordinates, and its threshold has polynomial bit length.

Finally, bounded features give $C_U \preceq (|U|/n)I$. For $R_+ = \Sigma + I$ and every legal request,

$$0 \preceq R_+ - R_U = C_U + (|U|/n)I \preceq \frac{2m}{n} I = m^{-3} I \preceq m^{-3} R_+.$$

Taking $\eta_n = m^{-3}$ proves the claimed vanishing curvature variation. The fixed upper summary stores no graph, request bank, or retained statistics: its state is $(1/2, 0, n)$. □

The reduction uses the source's actual zero-head spectral flooring rule. The theorem concerns evaluating or separating requests for this fixed decoder. It does not prove hardness of selecting an optimal encoder, and its inverse-polynomial decision gap does not rule out a constant-factor approximate request oracle or an encoder that avoids exact separation.

K.3.3 Exact utility under stable retained curvature

The obstruction in Proposition 50 concerns the envelope objective. Optimizing the exact utility at an encoding-time reference curvature has a different guarantee: a common Loewner comparison of retained curvatures transfers its optimum to the product relaxation of the response set. The comparison below allows noncommuting matrices and singular response envelopes.

Lemma 54 (Sharp curvature comparison for exact loss)**.** *For $R \succ 0$ and $E \succeq 0$, write*

$$W_R(E) = E(R+E)^{-1}R(R+E)^{-1}E.$$

If $R_0 \succ 0$ and $0 < a \leq b$ satisfy $aR_0 \preceq R \preceq bR_0$, then, with $\kappa = \max\{b, a^{-1}\}$,

$$\kappa^{-1} W_{R_0}(E) \preceq W_R(E) \preceq \kappa W_{R_0}(E). \tag{271}$$

Both constants are sharp given only the displayed curvature interval.

Proof. First suppose $E \succ 0$. Direct multiplication gives the identity

$$W_R(E)^{-1} = E^{-1}RE^{-1} + 2E^{-1} + R^{-1}. \tag{272}$$

The curvature assumption implies $aE^{-1}R_0E^{-1} \preceq E^{-1}RE^{-1} \preceq bE^{-1}R_0E^{-1}$ and $b^{-1}R_0^{-1} \preceq R^{-1} \preceq a^{-1}R_0^{-1}$. Since $\kappa \geq 1$, termwise comparison in (272) yields

$$\kappa^{-1} W_{R_0}(E)^{-1} \preceq W_R(E)^{-1} \preceq \kappa W_{R_0}(E)^{-1}.$$

Inverting proves (271). For singular E, apply this result to $E + \epsilon I$. The resolvent identity gives

$$\|(R+E+\epsilon I)^{-1} - (R+E)^{-1}\|_{\mathrm{op}} \leq \frac{\epsilon}{\lambda_{\min}(R)^2},$$

and the same bound holds with R_0 in place of R. Thus $W_R(E+\epsilon I) \to W_R(E)$ and $W_{R_0}(E+\epsilon I) \to W_{R_0}(E)$ in operator norm. Taking limits preserves both positive orders.

Sharpness already holds for scalars. With $R_0 = 1$, $R = s > 0$, and $E = e > 0$, the ratio is

$$\frac{W_s(e)}{W_1(e)} = \frac{s(1+e)^2}{(s+e)^2} \longrightarrow \begin{cases} s^{-1}, & e \downarrow 0, \\ s, & e \to \infty. \end{cases}$$

Choosing $s = a$ or $s = b$ approaches both κ and κ^{-1}. □

Theorem 55 (Reference-curvature selection for product utility)**.** *Fix $G, B \succeq 0$, $R_0 \succ 0$, and an integer rank budget $0 \le r \le d$. Let $\mathcal{J}$ be a nonempty family of pairs (R_j, w_j) with $w_j \ge 0$, $\sup_j w_j < \infty$, and $aR_0 \preceq R_j \preceq bR_0$ for every j, where $0 < a \le b$. For $0 \preceq N \preceq G$ with $\operatorname{rank} N \le r$, define*

$$\ell_0(N) = \lambda_{\max}\Big(B^{1/2} W_{R_0}(G-N) B^{1/2}\Big),$$
$$\mathcal{U}(N) = \sup_{j \in \mathcal{J}} w_j \lambda_{\max}\Big(B^{1/2} W_{R_j}(G-N) B^{1/2}\Big).$$

If N_0 minimizes ℓ_0, then

$$\mathcal{U}(N_0) \le \kappa^2 \min_{\substack{0 \preceq N \preceq G \\ \operatorname{rank} N \le r}} \mathcal{U}(N), \qquad \kappa = \max\{b, a^{-1}\}. \tag{273}$$

In particular, if $(1-\eta)R_0 \preceq R_j \preceq (1+\eta)R_0$ for $0 \le \eta < 1$, the factor is $(1-\eta)^{-2}$. For retained loss over the product ellipsoids $\Delta\Delta^\top \preceq c_j B$, take $w_j = c_j/\beta_j$.

Proof. Congruence by $B^{1/2}$ and the variational characterization of the largest eigenvalue transfer Lemma 54 to every summand. Writing $w_* = \sup_j w_j$ gives, for every admissible N,

$$\kappa^{-1} w_* \ell_0(N) \le \mathcal{U}(N) \le \kappa w_* \ell_0(N).$$

If $w_* = 0$, every objective is zero. Otherwise these inequalities and the optimality of N_0 prove (273). The admissible set is compact, since positive order bounds it and the rank constraint is closed. The reference objective is continuous there, and the robust objective is a bounded supremum of continuous functions, hence lower semicontinuous. Both minima therefore exist. The last assertion follows from the ellipsoid identity (256). □

Balancing the reference scale. If the available comparison is $aR_+ \preceq R_j \preceq bR_+$, choose $R_\star = \sqrt{ab}\, R_+$ as the reference in the theorem. Then $\sqrt{a/b}\, R_\star \preceq R_j \preceq \sqrt{b/a}\, R_\star$, so the product-utility approximation factor is b/a. This follows by balancing the two scalar comparison constants; it does not require a new matrix optimization argument.

An encoding-time condition for bounded ridge data. Let $\|x_i\| \le 1$, $\lambda > 0$, and $0 \le \rho < 1$. With $R_0 = \Sigma + \lambda I$, every legal request satisfies

$$R_U = R_0 - C - \alpha\lambda I, \qquad 0 \preceq C \preceq \rho I, \qquad 0 \le \alpha \le \rho.$$

Consequently, the computable sufficient condition

$$\eta = \frac{\rho(1+\lambda)}{\lambda_{\min}(\Sigma) + \lambda} < 1 \quad \Longrightarrow \quad (1-\eta)R_0 \preceq R_U \preceq R_0 \tag{274}$$

gives the factor $(1-\eta)^{-2}$ in Theorem 55. Using the scaled reference $R_\star = \sqrt{1-\eta}\,(\Sigma + \lambda I)$ instead improves this factor to $(1-\eta)^{-1}$. Indeed, $C + \alpha\lambda I \preceq \rho(1+\lambda) I \preceq \eta R_0$. This also applies to a product relaxation containing every $0 \preceq C \preceq \Sigma$ with $\operatorname{tr} C \le \rho$ and every $0 \le \alpha \le \rho$, since these constraints imply $C \preceq \rho I$. Without a lower spectral bound on Σ, the sufficient condition $\rho(1+\lambda) < \lambda$ still applies. A verified tighter common bound on $R_0^{-1/2}(C + \alpha\lambda I)R_0^{-1/2}$ can replace the displayed η.

Selection and storage. Proposition 49 supplies the fixed-R_0 objective. If R_0, G, B commute, the minimizer is explicit: correct the r coordinates with largest scores

$$s_i^{(0)} = \frac{b_i (R_0)_i g_i^2}{((R_0)_i + g_i)^2}.$$

In the noncommuting case the theorem compares an exact minimizer of the stated reference problem; it does not assert an efficient algorithm for finding that minimizer. Any such correction has a factorization $N_0 = LL^\top$ with at most r columns. For $A_0 = aI + FF^\top$ with p columns, the persistent state remains (F, L, a, θ, n), with at most $d(p + r) + d + 2$ numerical entries. The reference curvature and the optimization matrices are encoding-time objects. The existing signed-factor rounding analysis applies under its spectral-floor or stated factor-norm assumptions; the relative comparison above is stated in exact arithmetic.

Application to the envelope obstruction. For the bounded family in Proposition 50, $R_+ = \Sigma + \lambda I$ commutes with G and B, and

$$\eta = \frac{n+1}{3n} < 1, \qquad (1-\eta)^{-1} = \frac{3n}{2n-1} < 1.51 \quad (n \geq 256).$$

The two positive scores s_i^+ at R_+ have the common squared denominator $(1/4 + 1/n)^2$ and satisfy

$$\frac{s_1^+}{s_2^+} = \frac{96(n-8)^2}{n^2(n+8)} \leq \frac{96}{n} < 1.$$

The scaled reference $R_\star = cR_+$, with $c = \sqrt{1-\eta} \leq 1$, also selects coordinate two. To see this, put $j_0 = 1/4 + 1/n$. Its score denominators before squaring are $c(R_+)_i + g_i = cj_0 + (1-c)g_i$. Since $g_1 > g_2$, the scaled score ratio is at most s_1^+/s_2^+. Thus either reference-utility selection corrects coordinate two, exactly the competing correction N_2 in that proposition. It uses the same 14 numerical entries, and the scaled reference gives the displayed factor below 1.51 for the product utility, while every envelope minimizer has the displayed unbounded gap in actual worst-deletion loss. In general, a product-ellipsoid guarantee alone is not a relative approximation theorem for the actual responses paired with their curvatures; it compares the specified relaxation.

K.3.4 Source leverage certificates for curvature and response

A certificate computed from the source features can remain informative when the minimum-eigenvalue condition (274) fails. The ridge leverage scores and their interpretation as DPP marginals are classical (Dereziński & Mahoney, 2021, Theorem 10). The result below combines their deletion-budget sum with a union bound and the existing density identity; it does not introduce a new leverage or concentration principle.

Theorem 56 (Common certificates from a leverage budget)**.** *Suppose $\lambda > 0$, $\|x_i\| \leq 1$, and $|y_i| \leq 1$. Let $0 \leq k < n$ be an integer deletion budget and put*

$$R_+ = \Sigma + \lambda I, \qquad q = k/n,$$
$$h_i = \frac{1}{n} x_i^\top R_+^{-1} x_i, \qquad d_{\text{eff}} = \sum_{i=1}^n h_i = \operatorname{tr}\left[\Sigma(\Sigma + \lambda I)^{-1}\right].$$

Let L_k be the sum of the k largest h_i, with $L_0 = 0$, and assume $L_k < 1$. Define

$$a_{\text{lev}} = (1-q)(1-L_k) > 0,$$
$$\Gamma_{\text{lev}} = \frac{\exp\left(q^2 d_{\text{eff}}/(1-q)^2\right)}{1-L_k} - 1. \tag{275}$$

For every legal deletion U with $|U| \leq k$, writing $\alpha = |U|/n$, $\beta = 1 - \alpha$, and $R_U = \Sigma - C + \beta\lambda I$, one has simultaneously

$$a_{\text{lev}} R_+ \preceq R_U \preceq R_+,$$
$$\Delta_U \Delta_U^\top \preceq \Gamma_{\text{lev}} \mathcal{V}, \qquad \Delta_U = w_{S-U} - w_S, \tag{276}$$

where $\mathcal{V}$ is the general-label DPP covariance in Lemma 35. These are deterministic encoding-time certificates. They allow singular Σ.

Proof. For a fixed legal request put $A_U = R_+^{-1/2} C R_+^{-1/2} \succeq 0$. Its trace is $\sum_{i\in U} h_i \le L_k$, so $C \preceq L_k R_+$. Legality gives

$$R_U = \beta(R_+ - C) + \alpha(\Sigma - C) \succeq \beta(1 - L_k)R_+ \succeq a_{\mathrm{lev}} R_+.$$

The upper bound follows from $R_+ - R_U = C + \alpha\lambda I \succeq 0$.

Use the exact density split (346), now with $H = \Sigma - C \succeq 0$ and eigenvalues z_j of H/λ. Its first term is

$$\log \frac{\det(\lambda I + H + C)}{\det(\lambda I + H)} = -\log\det(I - A_U) \le -\log(1 - L_k).$$

Indeed, if $a_j \ge 0$ are the eigenvalues of A_U, then $\sum_j a_j \le L_k < 1$ and $\prod_j (1 - a_j) \ge 1 - \sum_j a_j \ge 1 - L_k$. Equivalently, $\det(I - A_U)$ is the source DPP probability of avoiding U, and the same inequality is the union bound using the marginal probabilities h_i.

For the normalization term, $\beta + z \ge \beta(1 + z)$ gives

$$\frac{z}{(\beta + z)^2} \le \frac{1}{\beta^2} \frac{z}{1 + z} \quad (z \ge 0).$$

Moreover, inverse order and $H(H + \lambda I)^{-1} = I - \lambda(H + \lambda I)^{-1}$ imply

$$\sum_j \frac{z_j}{1 + z_j} = \mathrm{tr}\big[H(H + \lambda I)^{-1}\big] \le d_{\mathrm{eff}}.$$

Thus $\log(1 + t) \le t$ yields

$$\log \mathbb{E}_p f_U^2 \le -\log(1 - L_k) + \frac{\alpha^2 d_{\mathrm{eff}}}{\beta^2} \le \log(1 + \Gamma_{\mathrm{lev}}).$$

The general-label change-of-measure identity and Cauchy–Schwarz give the response order in (276). □

An informative regime and the scope of the product family. If $d_{\mathrm{eff}} > 0$, define the ridge coherence $\mu = n \max_i h_i / d_{\mathrm{eff}} \ge 1$. Then

$$L_k \le q\mu d_{\mathrm{eff}}, \qquad a_{\mathrm{lev}} \ge (1 - q)(1 - q\mu d_{\mathrm{eff}}) \quad \text{when } q\mu d_{\mathrm{eff}} < 1.$$

Thus low effective dimension and bounded coherence can certify a constant curvature factor even for singular covariance and small λ. Dependence on λ remains through the leverage scores and effective dimension. More precisely, for fixed $0 \le q_0, t_0 < 1$, the conditions $q \le q_0$ and $q\mu d_{\mathrm{eff}} \le t_0$ give $\Gamma_{\mathrm{lev}} = O_{q_0,t_0}(q\mu d_{\mathrm{eff}} + q^2 d_{\mathrm{eff}})$. If $d_{\mathrm{eff}} = 0$, all features vanish and every retained optimizer is zero, regardless of the labels. If $k = 0$, the theorem has $a_{\mathrm{lev}} = 1$ and $\Gamma_{\mathrm{lev}} = 0$.

The curvature certificate concerns the actual legal subset curvatures. It therefore applies to the product family obtained by pairing those curvatures with an independent common response ellipsoid. It does not certify every matrix in the larger relaxation $0 \preceq C \preceq \Sigma$, $\mathrm{tr}\, C \le k/n$. Within the inflated-base correction class of Corollary 65, the balanced reference $R_{\mathrm{ref}} = \sqrt{a_{\mathrm{lev}}} R_+$ consequently gives product approximation factor $9/(8a_{\mathrm{lev}})$; the unscaled reference $R_{\mathrm{ref}} = R_+$ gives $9/(8a_{\mathrm{lev}}^2)$. Both comparisons use the same persistent factor budget.

Combining the common envelope with importance sampling. Bounded features also permit the universal common multiplier, so put

$$\overline{\Gamma} = \min\left\{\Gamma_{\mathrm{lev}}, \exp\left(\frac{q}{(1 - q)\lambda}\right) - 1\right\}.$$

Then $\Delta_U \Delta_U^\top \preceq \overline{\Gamma}\mathcal{V}$ simultaneously. No factor α/q is asserted for this bound: the sum of the largest j leverage scores need not be at most $(j/k)L_k$ when $j < k$. For $R = r_S(w_S)/\lambda > 0$, let M be the sampled importance covariance and set $B_m = 2M/R + I/(m+1)$. On the common event $\mathcal{V} \preceq RB_m$ supplied by Proposition 39, all legal responses lie in the common ellipsoid $\Delta\Delta^\top \preceq \overline{\Gamma} R B_m$. The uniform envelope bounds may therefore use $\overline{\Gamma}$ in place of Γ_ρ.

In particular, use the inflated-floor encoder with rank $r = m < d$ and $R_{\mathrm{ref}} = \sqrt{a_{\mathrm{lev}}} R_+$, and denote its fixed correction by N_{enc}. Let $\mathcal{U}_{\mathrm{sub}}$ be its product utility over the actual legal subset curvatures, paired with this last common ellipsoid. The shorted comparator used for (304) gives

$$\sup_{|U|\le k} \left[F_{S-U}(\widehat{w}_U) - F_{S-U}(w_{S-U})\right] \le \mathcal{U}_{\mathrm{sub}}(N_{\mathrm{enc}}) \le \frac{27\overline{\Gamma}R}{2a_{\mathrm{lev}}(1-q)^2\lambda(m+1)^3}. \tag{277}$$

Indeed, its squared floor bound is $4/(m+1)^2$ and $\lambda_{m+1}(B_m) \le 3/(m+1)$, giving comparator product loss at most $12\overline{\Gamma}R/[(1-q)^2\lambda(m+1)^3]$; multiply by $9/(8a_{\mathrm{lev}})$. The existing $m = 0$, zero-residual, and exact-covariance branches apply as before. The leverage scores and these scalar certificates are temporary encoding-time objects; neither they nor a list of legal requests is needed in the persistent state.

A bounded full-rank family with a constant certificate. Let $d = n \ge 4$ be a power of two, and let $H \in \{-1, 1\}^{d\times d}$ be a Hadamard matrix with $H^\top H = dI$. Define

$$x_i = \left(\frac{H_{i1}}{2}, \frac{H_{i2}}{2d}, \dots, \frac{H_{id}}{2d}\right), \qquad y_i = (x_i)_1, \qquad \lambda = 1/d, \qquad k = d/4.$$

The teacher is e_1, every feature norm is below one, and all coordinates and labels are dyadic numbers with $O(\log d)$ bits. The covariance has full rank, with

$$\Sigma = \mathrm{diag}\left(\frac{1}{4}, \frac{1}{4d^2}, \dots, \frac{1}{4d^2}\right), \qquad d_{\mathrm{eff}} = \frac{d}{d+4} + \frac{d-1}{4d+1} < \frac{5}{4}.$$

All leverage scores equal d_{eff}/d, so $\mu = 1$ and $L_k = d_{\mathrm{eff}}/4 < 5/16$. Therefore

$$a_{\mathrm{lev}} > \frac{33}{64}, \qquad \Gamma_{\mathrm{lev}} < \frac{16}{11} e^{5/36} - 1 < 0.68.$$

The balanced and unscaled inflated reference rules have product factors below $24/11$ and $512/121$, respectively. For the same family, the minimum-eigenvalue condition has $\eta = d(d+1)/(4d+1) > 1$, while the universal envelope multiplier is $\exp(d/3) - 1$ and $q/\lambda = d/4 \to \infty$. Thus a full algebraic rank does not preclude a constant common certificate when the effective dimension and coherence stay bounded. The Hadamard construction is standard; this example compares the certificates and does not assert a new minimax rate or empirical dominance of the resulting decoder.

Corollary 57 (A memory bound from effective dimension)**.** *Under Theorem 56, put $s = d_{\mathrm{eff}}$, $\beta_0 = 1 - q$, and $r_S = r_S(w_S)$, and use $\overline{\Gamma}$ defined above. Suppose $s > 0$, $r_S > 0$, and m is an integer satisfying $0 \le m < d$ and $j = m + 1 > s$. On the common importance-covariance event $\mathcal{V} \preceq RB_m$, where $R = r_S/\lambda$, the centered floor decoders with m covariance-head directions and at most m response-correction directions satisfy*

$$\begin{aligned} \mathcal{R}_{\mathrm{ord}} &\le \frac{3\overline{\Gamma} r_S s^2}{\beta_0^2 j(j-s)^2}, \\ \mathcal{R}_{\mathrm{infl}} &\le \frac{27\overline{\Gamma} r_S s^2}{2a_{\mathrm{lev}}\beta_0^2 j(j-s)^2}. \end{aligned} \tag{278}$$

Here $\mathcal{R}$ denotes the worst retained excess over all legal deletions of at most k records. The ordinary rule uses the floor $\sigma_{m+1}(\Sigma)$ and either the response-gap selection of Theorem 46 or a shorted correction of the leading m eigenvectors of B_m. The inflated rule uses Corollary 65 with reference $\sqrt{a_{\mathrm{lev}}} R_+$. The legal-midpoint variants have their separate interval guarantees; the reference-utility comparison used here concerns the floor decoders.

In particular, $j \ge 2s$ gives

$$\mathcal{R}_{\mathrm{infl}} \le \frac{54\overline{\Gamma} r_S s^2}{a_{\mathrm{lev}}\beta_0^2 j^3}. \tag{279}$$

For an ideal target $\varepsilon > 0$, either construction therefore admits a sufficient persistent numerical state of size

$$O\left(d \min\left\{d, 1 + s + \left(\frac{\overline{\Gamma} r_S s^2}{a_{\mathrm{lev}}\beta_0^2 \varepsilon}\right)^{1/3}\right\}\right). \tag{280}$$

The rank can be chosen from the source data before drawing the importance samples. The guarantee then holds on the existing concentration event for that chosen rank.

Proof. Put $\tau = \sigma_j(\Sigma)$. Since the leading j eigenvalues are at least τ,

$$s = \sum_i \frac{\sigma_i}{\sigma_i + \lambda} \geq \frac{j\tau}{\tau + \lambda}, \qquad \tau \leq \frac{\lambda s}{j - s}.$$

For the ordinary response-gap rule, $\lambda_{m+1}(G^{1/2} B_m G^{1/2}) \leq 3\tau/j$. Proposition 47, using $\psi(t) \leq t$ and the common scalar $\overline{\Gamma} R$, gives

$$\mathcal{R}_{\mathrm{ord}} \leq \frac{3\overline{\Gamma} R \tau^2}{\beta_0^2 \lambda j} \leq \frac{3\overline{\Gamma} r_S s^2}{\beta_0^2 j (j - s)^2}.$$

The shorted rule has the same bound because its residual annihilates the leading m response directions and is bounded by τI. This estimate also holds over the common product ellipsoids. Inflating the floor to 2τ multiplies this comparator bound by four. The product comparison multiplies it by $9/(8a_{\mathrm{lev}})$, proving the second line of (278); it also bounds the actual worst retained excess. For $j \geq 2s$, use $j - s \geq j/2$.

For example, choose

$$m = \min\left\{ d, \left\lceil 2s + \left(\frac{54\overline{\Gamma} r_S s^2}{a_{\mathrm{lev}} \beta_0^2 \varepsilon} \right)^{1/3} \right\rceil \right\}.$$

If $m < d$, then $j = m + 1 \geq 2s$ and (279) is at most ε. If $m = d$, store the exact covariance instead. The signed representations have at most $2m$ factor columns, which proves (280). If $\tau = 0$, the exact-head branch is already exact. All quantities choosing m precede the importance draws, so no additional union bound over ranks is needed. □

The cases $s = 0$ and $r_S = 0$ use the exact branches already described. For the ordinary response-gap or shorted rule, the cubic coefficient is 12 in place of $54/a_{\mathrm{lev}}$, so a_{lev} can be omitted from the cubic rank term in (280). The rank bound has no separate inverse power of λ: its dependence on regularization is through $s, \overline{\Gamma}, a_{\mathrm{lev}}$ and the source residual. This is a sufficient bound controlled by source geometry; it is not an instance-optimality statement or a matching lower bound. For finite-bit existence, take a fixed ideal error margin and apply the existing signed-factor and reference rounding bounds under their stated parameter assumptions. The inflated factors have operator norms at most $\sqrt{2}$ for bounded features, so only absolute constants in that rounding argument change. The precision cost still contains its logarithmic dependence on $1/\lambda$. This does not assert polynomial bit-operation complexity at the adaptive rank in (280).

Worst-case tightness within a constant-certificate class. At $\lambda = 1$, every bounded-feature dataset has $s \leq \operatorname{tr} \Sigma \leq 1$ and $h_i \leq 1/n$. For $k = \lfloor n/3 \rfloor$, it follows that

$$L_k \leq \frac{1}{3}, \qquad a_{\mathrm{lev}} \geq \frac{4}{9}, \qquad \Gamma_{\mathrm{lev}} \leq \frac{3}{2} e^{1/4} - 1 < 1.$$

The lower-bound families of Theorem 1 at $\lambda = 1$ therefore already lie in this class and require $\Omega(d \min\{d, \varepsilon^{-1/3}\})$ bits on that theorem's grid and accuracy range, allowing pointwise randomized success. Since $r_S \leq 1$, Corollary 57 gives the same d, ε dependence in numerical entries. In the unsaturated branch $\varepsilon^{-1/3} < d$, its precision logarithm is $O(\log d)$; exact grid statistics cover the saturated branch. The upper and lower bounds therefore match within a precision logarithm in this constant-certificate class. The separate minimax dependence on effective dimension, coherence, and leverage mass remains uncharacterized here.

K.3.5 A CONSTRUCTIVE UTILITY ENCODER WITH A CONDITIONED RESOLVENT

The reference problem admits a spectral construction whose approximation factor depends on the condition number of its positive resolvent. Its ingredients are a generalized Nyström correction, the spectral approximation lower bound, and the Kantorovich inequality (Anderson & Trapp, 1975; Eckart & Young, 1936; Giga, 2004). These tools give an approximation guarantee for the constrained utility problem, where saturation need not be optimal. An inflated spectral floor makes the factor at most $9/8$, while preserving the number of stored factor columns.

Theorem 58 (Oblique correction for exact reference utility)**.** *Let $0 \preceq Q_0 \prec I$, $D \succeq 0$, and let $0 \le r \le d$ be an integer. Define*

$$\ell_* = \min_{\substack{0 \preceq Q \preceq Q_0 \\ \operatorname{rank}(Q_0 - Q) \le r}} \|QD^{1/2}\|_{\mathrm{op}}^2.$$

If $Q_0 = 0$, the problem is exact with $Q = 0$. Otherwise let S have orthonormal columns spanning $\operatorname{range}(Q_0)$, and put

$$A = S^\top Q_0 S \succ 0, \qquad D_S = S^\top D S, \qquad X = A D_S^{1/2}, \qquad k = \operatorname{rank}(Q_0).$$

Let $s = \min\{r, \operatorname{rank}(X)\}$ and choose U to contain s leading left singular vectors of X. For $s > 0$ define

$$K_U = U(U^\top A^{-1} U)^{-1} U^\top, \qquad Q_U = S(A - K_U)S^\top. \tag{281}$$

For $s = 0$, set $K_U = 0$ and $Q_U = Q_0$. Then Q_U is feasible. Complete U to an orthonormal basis $[U, U_\perp]$ of $\mathbb{R}^k$ and set

$$H = (U^\top A^{-1} U)^{-1} U^\top A^{-1} U_\perp, \qquad \gamma_U^2 = 1 + \|H\|_{\mathrm{op}}^2,$$

with $H = 0$ if $s = 0$ or $s = k$. Writing

$$\kappa_A = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}, \qquad \gamma_A = \frac{\kappa_A + 1}{2\sqrt{\kappa_A}},$$

the following computable certificates and comparison hold:

$$\begin{aligned} \sigma_{r+1}(X)^2 &\le \ell_*, \\ \|Q_U D^{1/2}\|_{\mathrm{op}}^2 &\le \gamma_U^2 \sigma_{r+1}(X)^2 \le \gamma_U^2 \ell_* \le \gamma_A^2 \ell_*. \end{aligned} \tag{282}$$

Singular values past k are zero. In particular, $\ell_* = 0$ if and only if $\operatorname{rank}(X) \le r$, and the construction attains zero in this case. No separation between successive singular values is required.

Proof. Every feasible Q is supported on $\operatorname{range}(Q_0)$, and, writing $Q = S\widetilde{Q}S^\top$, one has

$$\|QD^{1/2}\|_{\mathrm{op}}^2 = \lambda_{\max}(\widetilde{Q} D_S \widetilde{Q}).$$

The correction in (281) satisfies

$$K_U = A^{1/2} P_{A^{-1/2}U} A^{1/2},$$

where $P_{A^{-1/2}U}$ is the orthogonal projector onto the displayed column space. Thus $0 \preceq K_U \preceq A$ and $\operatorname{rank}(K_U) = s \le r$. Equivalently, it is the generalized Nyström correction of A using the columns $A^{-1}U$.

Write $M = A^{-1}$ and use blocks in the basis $[U, U_\perp]$. Since

$$(A - K_U) D_S^{1/2} = \big(I - U(U^\top M U)^{-1} U^\top M\big) X,$$

the exact block identity is

$$[U, U_\perp]^\top (A - K_U) D_S^{1/2} = \begin{bmatrix} -H \\ I \end{bmatrix} U_\perp^\top X. \tag{283}$$

The chosen singular subspace gives $\|U_\perp^\top X\|_{\mathrm{op}} = \sigma_{s+1}(X) = \sigma_{r+1}(X)$, with the stated zero convention. Moreover, $\|[-H^\top, I]^\top\|_{\mathrm{op}}^2 = 1 + \|H\|_{\mathrm{op}}^2$. Every feasible residual subtracts a matrix of rank at most r from X, so the spectral approximation lower bound is $\sigma_{r+1}(X)^2$. This proves the first two comparisons and the assertion about zero loss.

For the condition-number bound, let $a_M = \lambda_{\min}(M)$ and $b_M = \lambda_{\max}(M)$. Functional calculus gives

$$M^2 \preceq (a_M + b_M) M - a_M b_M I.$$

Its upper-left block and congruence by M_{11}^{-1} imply

$$\begin{aligned} HH^\top &\preceq (a_M + b_M) M_{11}^{-1} - I - a_M b_M M_{11}^{-2} \\ &\preceq \frac{(b_M - a_M)^2}{4 a_M b_M} I. \end{aligned}$$

The last step maximizes $(a_M + b_M)/t - 1 - a_M b_M / t^2$ for $t > 0$; its maximum occurs at $t = 2a_M b_M/(a_M + b_M)$. Since $b_M/a_M = \kappa_A$, this gives $\gamma_U^2 \le (\kappa_A + 1)^2/(4\kappa_A)$. The empty-block cases satisfy this bound directly. □

Sharpness of the condition-number factor. The factor γ_A^2 is sharp for this construction, even when the leading singular vector is unique. Fix $\kappa > 1$, choose $0 < h < 1$, and set

$$M = \frac{1}{\kappa+1}\begin{bmatrix} 2\kappa & \sqrt{\kappa}(\kappa-1) \\ \sqrt{\kappa}(\kappa-1) & \kappa^2+1 \end{bmatrix}, \qquad A = hM^{-1}.$$

The eigenvalues of M are 1 and κ, so $0 \prec A \prec I$ and $\kappa_A = \kappa$. For $r = 1$ and $\epsilon > 0$, let

$$D_\epsilon = A^{-1}\,\mathrm{diag}(1+\epsilon, 1)A^{-1}.$$

Then $AD_\epsilon A = \mathrm{diag}(1+\epsilon, 1)$, so the selected subspace is uniquely $U = e_1$. Identity (283) gives loss exactly $1 + (\kappa-1)^2/(4\kappa) = \gamma_A^2$. The optimal loss is at least 1 by the spectral lower bound and at most $1+\epsilon$ by the feasible choice $Q = A$. Thus the ratio between the constructed loss and the optimum tends to γ_A^2 as $\epsilon \downarrow 0$.

Recovering the covariance correction. For the reference problem in Proposition 49, use $R = R_{\mathrm{ref}}$ and $D = R_{\mathrm{ref}}^{1/2} B R_{\mathrm{ref}}^{1/2}$ in the theorem. The construction also supplies the factor to store; there is no need to factor a dense difference of resolvents. For $s > 0$ put $V = SU$, $M_0 = I - Q_0$, and

$$\begin{aligned} C_U &= U^\top\big[A^{-1} + (I-A)^{-1}\big]U, \\ L_G &= R_{\mathrm{ref}}^{1/2} M_0^{-1} V C_U^{-1/2}, \qquad N_U = L_G L_G^\top. \end{aligned} \tag{284}$$

For $s = 0$ the factor is empty. These formulas give $0 \preceq N_U \preceq G$ and $\mathrm{rank}(N_U) = s$. Indeed, the Woodbury identity gives

$$(I-Q_0)^{-1} - (I-Q_U)^{-1} = M_0^{-1} V C_U^{-1} V^\top M_0^{-1}.$$

Congruence by $R_{\mathrm{ref}}^{1/2}$ and the inverse resolvent map identify the right-hand side with $G - E_U = N_U$. Consequently (282) is a constructive approximation guarantee for the original rank-constrained covariance correction, not only for its resolvent coordinates.

When R_{ref} and G commute, the factor has a simpler form. On the positive support of G, let

$$Z = \mathrm{orth}\Big([Q_0(I-Q_0)]^{-1/2} V\Big), \qquad L_G = G^{1/2} Z. \tag{285}$$

Here orth returns any orthonormal basis of the column space, and the inverse square root is restricted to that support. To verify the formula, the Gram matrix of the displayed columns is C_U, while simultaneous diagonalization gives

$$G^{1/2}[Q_0(I-Q_0)]^{-1/2} V = R_{\mathrm{ref}}^{1/2}(I-Q_0)^{-1} V.$$

Thus (285) gives the same N_U as (284), up to a rotation of its factor columns. In a spectral base this uses diagonal scaling and a thin orthogonalization.

Theorem 59 (Spectral bands with an augmented correction rank)**.** *Use Q_0, D, S, A, D_S, X, k and the rank-r optimum ℓ_* from Theorem 58, with integer $0 \le r \le d$. If $Q_0 = 0$ or $r = 0$, the corresponding exact or uncorrected rule applies. If $\mathrm{rank}(X) \le r$, that theorem attains zero loss with correction rank at most r. Otherwise, partition the positive spectrum of A into L nonempty bands of condition number at most two, where*

$$L \le \max\{1, \lceil \log_2 \kappa_A \rceil\}.$$

Let P_j be their orthogonal spectral projectors. Choose U to contain r leading left singular vectors of X, and let U_j have orthonormal columns spanning $\mathrm{range}(P_j U)$. Set $r_j = \mathrm{rank}(P_j U)$ and define

$$\begin{aligned} K_j &= U_j (U_j^\top A^{-1} U_j)^{-1} U_j^\top, \\ K_{\mathrm{band}} &= \sum_{j=1}^{L} K_j, \qquad Q_{\mathrm{band}} = S(A - K_{\mathrm{band}})S^\top, \end{aligned} \tag{286}$$

with $K_j = 0$ when $r_j = 0$. Then

$$\begin{aligned} 0 &\preceq Q_{\mathrm{band}} \preceq Q_0, \\ \mathrm{rank}(Q_0 - Q_{\mathrm{band}}) &= \sum_j r_j \le \min\{k, rL\}, \\ \|Q_{\mathrm{band}} D^{1/2}\|_{\mathrm{op}}^2 &\le \frac{9}{8}\sigma_{r+1}(X)^2 \le \frac{9}{8}\ell_*. \end{aligned} \tag{287}$$

No commutation assumption is imposed on D. In covariance coordinates this leaves the original base and gap G unchanged, and increases only the permitted correction rank.

Proof. All operations take place on the positive support of Q_0; thus singular Q_0 causes no inverse outside that support. Scan the sorted positive eigenvalues of A. Start each band at the smallest unassigned eigenvalue a and assign all remaining eigenvalues in $[a, 2a]$ to it. Successive band minima increase by more than a factor of two, giving the stated band count, including the single-band case. The scan creates at most k nonempty bands and does not iterate across empty spectral intervals. Each P_j commutes with A and A^{-1}. The projector factorization in Theorem 58, applied within the band, gives

$$0 \preceq K_j \preceq P_j A P_j, \qquad \operatorname{rank}(K_j) = r_j.$$

The bands are orthogonal, so summing proves the positive orders and the rank assertion.

Write A_j for the restriction of A to the jth band, identify K_j with its restriction there, and consider the block operator

$$T = I - K_{\mathrm{band}} A^{-1} = \bigoplus_j (I_j - K_j A_j^{-1}).$$

The Kantorovich bound used in Theorem 58 (Giga, 2004) applies to every U_j, irrespective of its relation to a singular subspace inside the band. Since $\kappa(A_j) \le 2$, it gives

$$\|I_j - K_j A_j^{-1}\|_{\mathrm{op}}^2 \le \frac{9}{8}, \qquad \|T\|_{\mathrm{op}}^2 \le \frac{9}{8}.$$

The empty correction block is the identity, and a fully corrected block is zero, so both endpoints obey this bound. Moreover, $K_j A_j^{-1} U_j = U_j$, hence $T P_j U = 0$ and $TU = 0$. Consequently

$$(A - K_{\mathrm{band}}) D_S^{1/2} = TX = T(I - UU^\top)X.$$

Its squared norm is at most $(9/8)\sigma_{r+1}(X)^2$. The positive-support identity and the rank-r spectral lower bound from Theorem 58 prove (287). The operator norm of a block diagonal matrix is the maximum block norm, so there is no factor depending on the number of bands in the loss bound.

Finally, $W = [U_1, \dots, U_L]$ has orthonormal columns and $W^\top A^{-1} W$ is block diagonal. Thus

$$K_{\mathrm{band}} = W(W^\top A^{-1} W)^{-1} W^\top.$$

Formula (284), with W in place of U, gives $0 \preceq N_{\mathrm{band}} \preceq G$ of the same rank. The base covariance is unchanged. □

A cutoff for small resolvent eigenvalues. The band count can instead be controlled after a cutoff $t > 0$. Let $P_{\le t}$ project onto the eigenvalues of A at most t. Leave that block uncorrected and apply the band construction to the remaining spectrum, still using projections of the same global singular subspace U. Denote its residual by Q_t. If $0 < t < \|A\|_{\mathrm{op}}$, the number of retained bands satisfies

$$L_t \le \max\left\{1, \left\lceil \log_2 \frac{\|A\|_{\mathrm{op}}}{t} \right\rceil \right\}, \qquad \operatorname{rank}(Q_0 - Q_t) \le \min\{k, rL_t\}.$$

For $t \ge \|A\|_{\mathrm{op}}$ there are no retained bands and the correction is zero. Orthogonality of the low and high output blocks gives the additive guarantee

$$\begin{aligned} \|Q_t D^{1/2}\|_{\mathrm{op}}^2 &\le \frac{9}{8}\sigma_{r+1}(X)^2 + \|P_{\le t} X\|_{\mathrm{op}}^2 \\ &\le \frac{9}{8}\ell_* + t^2 \|D_S\|_{\mathrm{op}} \le \frac{9}{8}\ell_* + t^2 \|D\|_{\mathrm{op}}. \end{aligned} \tag{288}$$

Indeed, the low output equals $P_{\le t} X$, while the high output is its block operator applied to $(I - UU^\top)X$. The high operator has squared norm at most $9/8$ and $\|P_{\le t} A\|_{\mathrm{op}} \le t$; summing the two squared output bounds proves the display. If $\sigma_{r+1}(X) > 0$, then $\|D_S\|_{\mathrm{op}} > 0$. For any $\zeta > 0$, choosing

$$t = \frac{\sqrt{\zeta}\,\sigma_{r+1}(X)}{\sqrt{\|D_S\|_{\mathrm{op}}}} \quad \Longrightarrow \quad \|Q_t D^{1/2}\|_{\mathrm{op}}^2 \le (9/8 + \zeta)\ell_* \tag{289}$$

removes dependence on the smallest positive eigenvalue from the retained band count. Zero optimum uses the exact rank-r branch of Theorem 58.

The construction combines spectral bands with the preceding oblique projection and Kantorovich estimate. It compares against the original rank-r reference optimum while charging the additional factor columns. It makes no finite-bit efficiency claim for exact spectral partitioning or exact subspace selection. Comparisons of actual deletion loss still require the separate curvature and response transfer hypotheses.

A finite bank of actual responses. For an explicitly available request bank, the encoder can work with the response points themselves. The greedy least-squares strategy of repeatedly fitting a large fraction of the remaining points is classical (Iwen & Krahmer, 2013). Combining it with the preceding oblique correction gives the following comparison, with additional factor columns and without enlarging the points to an ellipsoid.

Theorem 60 (Finite-point reference utility with augmented rank)**.** *Let $0 \preceq Q_0 \prec I$, let $z_1, \dots, z_M \in \mathbb{R}^d$ be a finite list, and let $0 \le r \le d$ be an integer. Define*

$$f_r^* = \min_{\substack{0 \preceq Q \preceq Q_0 \\ \operatorname{rank}(Q_0 - Q) \le r}} f(Q), \qquad f(Q) = \max_{1 \le i \le M} \|Q z_i\|^2.$$

For an empty list set $f = 0$ and use the zero correction. For $k = \operatorname{rank}(Q_0) > 0$, let S have orthonormal columns spanning its positive support, and put

$$A = S^\top Q_0 S \succ 0, \qquad X_i = A S^\top z_i, \qquad X = [X_1, \dots, X_M].$$

The cases $k = 0$ and $r = 0$ use the exact zero residual and the uncorrected residual, respectively. If $\operatorname{rank}(X) \le r$, a single oblique correction with subspace $\operatorname{range}(X)$ attains zero loss with correction rank at most r.

Otherwise start with all indices remaining and an empty subspace. At each round, for the remaining index set I, take a leading left singular subspace V_I of $X_I = [X_i]_{i \in I}$ of dimension $\min\{r, \operatorname{rank}(X_I)\}$. Add V_I to the accumulated subspace and remove the $\lceil |I|/2 \rceil$ points having the smallest values of $\|(I - P_{V_I}) X_i\|^2$, with arbitrary tie breaking. Let W have orthonormal columns spanning the final accumulated subspace. The number of rounds is at most

$$J = \lfloor \log_2 M \rfloor + 1, \qquad \dim \operatorname{range}(W) \le \min\{k, rJ\}.$$

Partition the positive spectrum of A into L nonempty bands as in Theorem 59, so that $L \le \max\{1, \lceil \log_2 \kappa_A \rceil\}$. Let $U_j = \operatorname{orth}(P_j W)$ and define

$$\begin{aligned} K_{\mathrm{fin}} &= \sum_{j=1}^{L} U_j (U_j^\top A^{-1} U_j)^{-1} U_j^\top, \\ Q_{\mathrm{fin}} &= S(A - K_{\mathrm{fin}}) S^\top, \end{aligned} \tag{290}$$

where empty summands are zero. Then

$$\begin{aligned} 0 \preceq Q_{\mathrm{fin}} &\preceq Q_0, \\ \operatorname{rank}(Q_0 - Q_{\mathrm{fin}}) &= \sum_j \operatorname{rank}(P_j W) \le \min\{k, rJL\}, \\ f(Q_{\mathrm{fin}}) &\le \frac{9}{4} f_r^*. \end{aligned} \tag{291}$$

No alignment between the points and the spectral bands is assumed.

Proof. Consider the nonempty, positive-support case. The feasible set is compact, so a minimizer Q_* exists. It is supported on $\operatorname{range}(S)$; write $\widetilde{Q}_* = S^\top Q_* S$. For any remaining set I, the matrix with columns $(A - \widetilde{Q}_*) S^\top z_i$, $i \in I$, has rank at most r and approximates X_I with every squared column error at most f_r^*. The Frobenius optimality of the leading singular subspace therefore gives

$$\sum_{i \in I} \|(I - P_{V_I}) X_i\|^2 \le |I| f_r^*.$$

Every one of the best $\lceil |I|/2 \rceil$ points has squared residual at most $2 f_r^*$: a larger residual at that order statistic would make the displayed sum exceed $|I| f_r^*$. Its fitted subspace is contained in the final range of W. Consequently every original point satisfies

$$\|(I - W W^\top) X_i\|^2 \le 2 f_r^*. \tag{292}$$

The number of remaining points becomes $\lfloor |I|/2 \rfloor$ at each round, proving the asserted round and dimension bounds. The construction uses only residual ordering, not the unknown value of f_r^*.

The band proof in Theorem 59 applies to the spaces $\mathrm{range}(P_j W)$ and gives feasibility, the rank formula, and

$$T = I - K_{\mathrm{fin}} A^{-1}, \qquad \|T\|_{\mathrm{op}}^2 \le \frac{9}{8}, \qquad TW = 0.$$

Thus $Q_{\mathrm{fin}} z_i = ST(I - WW^\top)X_i$. Combining this identity with (292) proves (291). The block operator uses its maximum block norm, so neither the number of rounds nor the number of bands multiplies the loss. All inverses are restricted to the positive support.

Finally, $f_r^* = 0$ implies $X = (A - \widetilde{Q}_*)[S^\top z_i]_{i=1}^M$ and hence $\mathrm{rank}(X) \le r$. Conversely, the global oblique correction with subspace $\mathrm{range}(X)$ annihilates every X_i and has that rank. This proves the zero-loss branch; the empty list, zero support, and zero rank cases are immediate. □

In the nonzero-loss branch, the small-eigenvalue cutoff has a pointwise version. Set $z_{\max} = \max_i \|S^\top z_i\|$, leave the eigenvalues of A at most t uncorrected, and apply the construction to the retained bands using the same W. Orthogonality of the low and high output blocks gives

$$f(Q_t) \le \frac{9}{4} f_r^* + t^2 z_{\max}^2. \tag{293}$$

The correction rank is at most $\min\{k, rJL_t\}$, with L_t as in the preceding cutoff construction. Every rank-r comparator also gives a rank-r approximation to X, so $\sigma_{r+1}(X)^2/M \le f_r^*$. When $\sigma_{r+1}(X) > 0$, one has $z_{\max} > 0$; for any $\zeta > 0$, the choice

$$t = \frac{\sqrt{\zeta}\,\sigma_{r+1}(X)}{\sqrt{M}\,z_{\max}} \quad \Longrightarrow \quad f(Q_t) \le (9/4 + \zeta) f_r^*$$

removes the smallest positive eigenvalue from the retained band count. A zero singular-value lower bound is exactly the zero-loss branch already treated above.

Removing the number of requests from the correction rank. The semidefinite relaxation for point-set subspace approximation is classical (Varadarajan et al., 2007, Section 2, equation (2)). Its elementary eigenvalue-threshold rounding gives a constant-factor approximation with additional subspace dimensions. Combining that rounding with the feasible oblique band correction removes the factor J from the stored correction rank; the SDP and its threshold rounding are not new subspace-approximation methods.

Theorem 61 (Semidefinite rounding for a finite response bank)**.** *Use the notation and endpoint rules of Theorem 60. In its remaining case* $1 \le r < \mathrm{rank}(X)$*, define*

$$\begin{aligned} t_{\mathrm{sdp}} = \min_{P,t} \quad & t \\ \textit{subject to} \quad & 0 \preceq P \preceq I_k, \qquad \mathrm{tr}\, P \le r, \\ & X_i^\top (I_k - P) X_i \le t, \qquad 1 \le i \le M. \end{aligned} \tag{294}$$

Let $\delta \ge 0$*, and take any feasible pair* (P, t) *with* $t \le t_{\mathrm{sdp}} + \delta$*. Let* W *have orthonormal columns spanning the eigenspaces of* P *with eigenvalues at least* $1/2$*. Apply* (290) *to this* W*, using the same* L *spectral bands of* A*, and denote the correction and residual by* K_{sdp} *and* Q_{sdp}*, respectively. Then*

$$\begin{aligned} 0 \preceq Q_{\mathrm{sdp}} \preceq Q_0, \\ \mathrm{rank}(Q_0 - Q_{\mathrm{sdp}}) = \sum_j \mathrm{rank}(P_j W) \le \min\{k, 2rL\}, \\ f(Q_{\mathrm{sdp}}) \le \frac{9}{4}(f_r^* + \delta). \end{aligned} \tag{295}$$

Thus exact optimization gives the same $9/4$ *loss factor with correction rank at most* $2rL$*, independently of the bank size* M*. The bank size still enters the encoding problem through its* M *constraints.*

Proof. The feasible set for P is compact. The objective after eliminating t is a finite maximum of continuous functions, so an SDP minimizer exists. Let $\widetilde{Q}_* = S^\top Q_* S$ be a minimizer for f_r^*, and let

Π_* project orthogonally onto $\mathrm{range}(A - \widetilde{Q}_*)$. This projector satisfies $0 \preceq \Pi_* \preceq I_k$ and $\operatorname{tr}\Pi_* \le r$. Since $(A - \widetilde{Q}_*)S^\top z_i$ belongs to its range,

$$X_i^\top (I_k - \Pi_*) X_i = \|(I_k - \Pi_*)X_i\|^2 \le \|\widetilde{Q}_* S^\top z_i\|^2 \le f_r^*.$$

Consequently $t_{\mathrm{sdp}} \le f_r^*$.

If q eigenvalues of P are at least $1/2$, then $q/2 \le \operatorname{tr} P \le r$. Moreover, scalar comparison in an eigenbasis of P gives the positive order

$$I_k - WW^\top \preceq 2(I_k - P).$$

It follows that

$$\|(I_k - WW^\top)X_i\|^2 \le 2X_i^\top (I_k - P)X_i \le 2t \le 2(f_r^* + \delta) \quad \text{for every } i.$$

The band proof in Theorem 59 applies to this subspace without any alignment assumption. It gives the positive orders, the displayed rank formula, and

$$T = I_k - K_{\mathrm{sdp}} A^{-1}, \qquad TW = 0, \qquad \|T\|_{\mathrm{op}}^2 \le \frac{9}{8}.$$

Therefore $Q_{\mathrm{sdp}} z_i = ST(I_k - WW^\top)X_i$. The preceding pointwise estimate proves the loss bound. The zero-loss and endpoint constructions are those already proved in Theorem 60. □

More generally, selecting eigenvalues strictly greater than $\tau \in (0,1)$ gives

$$\dim W \le \left\lceil \frac{r}{\tau} \right\rceil - 1, \qquad f(Q_{\mathrm{sdp}}) \le \frac{9}{8(1-\tau)}(f_r^* + \delta),$$

with correction rank at most $\min\{k, L(\lceil r/\tau \rceil - 1)\}$. In particular, the strict threshold at $1/2$ improves $2r$ to $2r-1$ with the same loss bound. These exact spectral statements include eigenvalues equal to the threshold in the specified way.

The additive optimization accuracy is explicit. In the nonzero-loss branch, the computable lower bound $\mu = \sigma_{r+1}(X)^2/M > 0$ satisfies $\mu \le f_r^*$. For any $\varepsilon > 0$, choosing $\delta \le (4\varepsilon/9)\mu$ yields $f(Q_{\mathrm{sdp}}) \le (9/4 + \varepsilon) f_r^*$. The same low-eigenvalue cutoff as above gives

$$f(Q_{\mathrm{sdp},t_0}) \le \frac{9}{4}(f_r^* + \delta) + t_0^2 z_{\max}^2, \qquad \mathrm{rank}(Q_0 - Q_{\mathrm{sdp},t_0}) \le \min\{k, 2rL_{t_0}\}.$$

The theorem assumes an exactly feasible SDP point and the stated spectral subspaces; δ measures objective accuracy only. The following proposition replaces exact SDP feasibility and exact threshold eigenspaces by explicit residual certificates. It retains the exact spectral bands of A; subsequent factor and reference rounding must still be charged through the existing absolute loss-perturbation bounds.

Proposition 62 (Certified residuals for SDP threshold rounding)**.** *Use the notation of Theorem 61, with $M \ge 1$ and $1 \le r < k$, and put $B^2 = \max_i \|X_i\|^2$. Let $\widehat{P}$ be symmetric, $t \ge 0$, and $\eta, \eta_{\mathrm{tr}}, \eta_c \ge 0$ satisfy*

$$-\eta I_k \preceq \widehat{P} \preceq (1+\eta) I_k, \qquad \operatorname{tr}\widehat{P} \le r + \eta_{\mathrm{tr}},$$
$$X_i^\top (I_k - \widehat{P}) X_i \le t + \eta_c \quad (1 \le i \le M).$$

Define

$$u = \eta_{\mathrm{tr}} + k\eta, \qquad d_0 = 1 + 2\eta + u/r, \qquad P_f = (\widehat{P} + \eta I_k)/d_0.$$

Suppose $V \in \mathbb{R}^{k\times k}$ satisfies $V^\top V = I_k$, Λ is diagonal, and $\omega \ge 0$ satisfies

$$\|P_f - V\Lambda V^\top\|_{\mathrm{op}} \le \omega.$$

Let W contain the columns V_j for which $\Lambda_{jj} \ge 1/2 + \omega$. Apply (290) *to this W using the same L exact spectral bands of A, and denote the resulting correction and residual by K_{res} and Q_{res}. Then*

$$0 \preceq Q_{\mathrm{res}} \preceq Q_0, \qquad \mathrm{rank}(Q_0 - Q_{\mathrm{res}}) \le \min\{k, 2rL\},$$
$$f(Q_{\mathrm{res}}) \le \frac{9}{4}\left(\frac{t + \eta_c + (\eta + u/r)B^2}{d_0} + 2\omega B^2\right). \tag{296}$$

In particular, if $t \le t_{\mathrm{sdp}} + \delta$ for $\delta \ge 0$, then

$$f(Q_{\mathrm{res}}) \le \frac{9}{4}(f_r^* + E_{\mathrm{sdp}}), \qquad E_{\mathrm{sdp}} = \delta + \eta_c + (\eta + u/r + 2\omega)B^2.$$

No separation between the eigenvalues of P_f and $1/2$ is assumed.

Proof. The spectral bounds and the trace bound imply $0 \preceq P_f \preceq I_k$ and $\operatorname{tr} P_f \le r$. Moreover,

$$I_k - P_f = \frac{I_k - \widehat{P} + (\eta + u/r)I_k}{d_0}, \qquad X_i^\top (I_k - P_f) X_i \le \frac{t + \eta_c + (\eta + u/r)B^2}{d_0}.$$

Write $H = V\Lambda V^\top$ and $\Pi = WW^\top$. Every selected column satisfies $V_j^\top P_f V_j \ge \Lambda_{jj} - \omega \ge 1/2$. Since $P_f \succeq 0$, summing over selected columns gives $\dim W \le 2 \operatorname{tr} P_f \le 2r$. Also $P_f \preceq I_k$ implies $\Lambda_{jj} \le 1 + \omega$. Scalar comparison in the basis V therefore yields

$$I_k - \Pi \preceq 2(I_k - H + \omega I_k) \preceq 2(I_k - P_f + 2\omega I_k).$$

Consequently,

$$\|(I_k - \Pi)X_i\|^2 \le 2\left[X_i^\top (I_k - P_f) X_i + 2\omega \|X_i\|^2\right].$$

The exact band construction has positive orders as stated, correction rank at most $L \dim W$, and $T = I_k - K_{\mathrm{res}} A^{-1}$ with $TW = 0$ and $\|T\|_{\mathrm{op}}^2 \le 9/8$. Thus $Q_{\mathrm{res}} z_i = ST(I_k - \Pi)X_i$, which proves (296). Finally, $d_0 \ge 1$, $t \ge 0$, and $t_{\mathrm{sdp}} \le f_r^*$ give the simplified bound. □

When the displayed input matrices, points, and tolerance bounds are rational, the residual certificates can be checked by exact rational arithmetic: check $V^\top V = I_k$, the scalar inequalities, and the positive semidefiniteness of the required rational matrices, including $\omega I_k \pm (P_f - V\Lambda V^\top)$. The threshold comparison is then a rational comparison. These certificates do not establish how the encoder obtains the exact spectral bands of A or rounds the covariance factor and reference. Appendix K.3.6 supplies a complete rational construction, with the extra weights and precision in that theorem charged explicitly. An empty bank uses the existing endpoint convention rather than this proposition.

Corollary 63 (Actual loss on a finite deletion bank)**.** *Fix $\lambda > 0$, an original covariance base $A_0 \succeq \Sigma$, its gap $G = A_0 - \Sigma$, and an integer $0 \le r \le d$. Let $\mathcal{J} = \{U_1, \dots, U_M\}$ be a finite bank of actual legal requests, and set*

$$\beta_i = 1 - |U_i|/n > 0, \qquad R_i = \Sigma - C_i + \beta_i \lambda I, \qquad \Delta_i = w_{S - U_i} - w_S.$$

Suppose $R_+ \succ 0$ and $0 < a \le b$ satisfy $aR_+ \preceq R_i \preceq bR_+$ for every request in the bank. For the centered decoder at $\theta = w_S$, define its exact worst loss on the bank by

$$\mathcal{R}_{\mathcal{J}}(N) = \max_i \frac{1}{\beta_i} \Delta_i^\top W_{R_i}(G - N)\Delta_i, \qquad 0 \preceq N \preceq G.$$

Use the balanced reference and transformed response points

$$R_{\mathrm{ref}} = \sqrt{ab}\, R_+, \qquad z_i = \frac{R_{\mathrm{ref}}^{1/2} \Delta_i}{\sqrt{\beta_i}}. \tag{297}$$

Let $H_0 = R_{\mathrm{ref}}^{-1/2} G R_{\mathrm{ref}}^{-1/2}$ and $Q_0 = H_0(I + H_0)^{-1}$. Apply Theorem 60 and recover the covariance factor using (284) *with the concatenated band subspaces. The resulting N_{fin} satisfies*

$$\begin{aligned} 0 \preceq N_{\mathrm{fin}} \preceq G, \qquad & \operatorname{rank}(N_{\mathrm{fin}}) \le \min\{k, rJL\}, \\ \mathcal{R}_{\mathcal{J}}(N_{\mathrm{fin}}) \le \frac{9b}{4a} & \min_{\substack{0 \preceq N \preceq G \\ \operatorname{rank} N \le r}} \mathcal{R}_{\mathcal{J}}(N), \end{aligned} \tag{298}$$

where $k = \operatorname{rank}(G)$ and J, L have the definitions above. For an empty bank or zero support, use the harmless endpoint conventions $J = L = 1$. The endpoint rules of that theorem apply, with zero loss by convention for an empty bank. The comparison uses the same original base and actual requests on both sides.

Proof. For $E = G - N$, set $H = R_{\mathrm{ref}}^{-1/2} E R_{\mathrm{ref}}^{-1/2}$ and $Q = H(I + H)^{-1}$. The exact resolvent identity in Proposition 49 gives

$$\frac{1}{\beta_i} \Delta_i^\top W_{R_{\mathrm{ref}}}(E) \Delta_i = \|Q z_i\|^2.$$

That proposition identifies the feasible covariance corrections with $0 \preceq Q \preceq Q_0$ and preserves their correction ranks. The band Gram matrix is block diagonal, so the covariance-factor conversion applies exactly as in Theorem 59. Let $\kappa = \sqrt{b/a}$. The balanced curvature interval and Lemma 54 give, for every feasible N, the pointwise comparison

$$\kappa^{-1} f(Q) \le \mathcal{R}_{\mathcal{J}}(N) \le \kappa f(Q).$$

Apply the upper comparison to the construction, its $9/4$ reference guarantee, and the lower comparison to a rank-r actual minimizer. This gives $(9/4)\kappa^2 = 9b/(4a)$. The finite maximum is continuous on the compact feasible set, so this minimizer exists. No common response ellipsoid is used. □

Alternatively, applying Theorem 61 to the same transformed response bank gives a covariance correction N_{sdp} with

$$0 \preceq N_{\mathrm{sdp}} \preceq G, \qquad \operatorname{rank}(N_{\mathrm{sdp}}) \le \min\{k, 2rL\},$$
$$\mathcal{R}_{\mathcal{J}}(N_{\mathrm{sdp}}) \le \frac{9b}{4a} \min_{\substack{0 \preceq N \preceq G \\ \operatorname{rank} N \le r}} \mathcal{R}_{\mathcal{J}}(N) + \frac{9}{4}\sqrt{\frac{b}{a}}\,\delta.$$

This follows from the same two pointwise curvature comparisons. The endpoint and zero-loss branches use their exact constructions. For $\delta = 0$, the relative factor is unchanged and the correction needs at most $d \min\{k, 2rL\}$ numerical entries, with $2r - 1$ in place of $2r$ under the strict-threshold convention above.

The response bank is needed only during encoding. The persistent signed state stores the original base factors, the new covariance factor, θ, and n; the additional factor uses at most $d \min\{k, rJL\}$ numerical entries. It stores neither the bank nor request identifiers. For all singleton requests, the encoder can compute the n responses from the source records by ridge solves; the empty request has zero response and may be omitted. The greedy construction has polynomial cost in n, d in a real-arithmetic model with spectral operations. For a general explicitly listed bank its corresponding cost is polynomial in n, d and its listed size; enumerating all legal subsets can have exponential size. This finite-bank construction is separate from the sampled-response encoders evaluated in the performance grid. Under the spectral-floor or stated factor-norm assumptions, the existing signed-factor and reference-rounding bounds yield absolute loss perturbations. They do not preserve the displayed relative factor unconditionally. Appendix K.3.6 instead gives a polynomial-bit construction for rational inputs and a listed bank, with a weighted factor format and certified arithmetic.

An exact bank from finitely many feature types. Repeated features allow the request bank to be generated without enumerating all deletion subsets, even when their labels differ.

Theorem 64 (Endpoint requests for repeated features)**.** *Suppose the original dataset has t distinct feature vectors $v_1, \dots, v_t$, with respective multiplicities $n_1, \dots, n_t$. Labels may be arbitrary bounded reals. Fix $\lambda > 0$ and an integer $0 \le k < n$. Within each feature group, sort the labels, breaking ties in a fixed order. For every integer count vector*

$$c = (c_1, \dots, c_t), \qquad 0 \le c_j \le n_j, \qquad \sum_j c_j \le k,$$

form requests by independently deleting either the c_j smallest or the c_j largest labels in group j. Let $\mathcal{B}$ be the resulting bank, with duplicate requests removed. Then, for every $E \succeq 0$,

$$|\mathcal{B}| \le 2^t \binom{k+t}{t},$$
$$\max_{|U| \le k} \frac{1}{\beta_U} \Delta_U^\top W_{R_U}(E) \Delta_U = \max_{U \in \mathcal{B}} \frac{1}{\beta_U} \Delta_U^\top W_{R_U}(E) \Delta_U. \tag{299}$$

Here $\beta_U = 1 - |U|/n$, $R_U = \Sigma - C_U + \beta_U \lambda I$, and $\Delta_U = w_{S-U} - w_S$. In particular, the same bank is exact simultaneously for all admissible covariance corrections to any fixed original base.

Proof. Fix a count vector c, put $\theta = w_S$, and let Y_j be the sum of the labels deleted from group j. The values of β_U and $C_U = n^{-1}\sum_j c_j v_j v_j^\top$, hence of R_U, depend only on c. The retained normal equations give

$$nR_U\Delta_U = \sum_j c_j(v_j v_j^\top \theta + \lambda\theta) - \sum_j v_j Y_j.$$

Thus, for fixed c and E, the loss in (299) is a convex quadratic function of $Y = (Y_1, \dots, Y_t)$, since $W_{R_U}(E) \succeq 0$. The smallest and largest feasible values of Y_j are the sums of the c_j smallest and largest labels in that group. Every feasible Y lies in the box with these endpoints, and every vertex of this box is attained by a legal request in the bank. Convexity bounds the value at any point of the box by its largest vertex value. This proves equality of the maxima for each c and then over all counts. The argument applies to every E using the same endpoints. There are at most $\binom{k+t}{t}$ nonnegative count vectors with sum at most k, and at most 2^t endpoint choices for each. □

Sorting the labels and storing their prefix sums temporarily lets the encoder compute each endpoint response from its group counts and label sums. Including feature grouping and the original moment calculation, the construction takes $\mathrm{poly}(n,d) + 2^t\binom{k+t}{t}\,\mathrm{poly}(d,t)$ real-arithmetic operations; it need not list the individual deleted records for every endpoint. This is polynomial for fixed t, with an exponent depending on t, rather than a fixed-parameter tractable bound in t. Features $x_i = \pm v_j$ can also be grouped by replacing their labels with the corresponding signed labels, which preserves their covariance and cross-moment contributions.

Applying Corollary 63 and its SDP variant to $\mathcal{B}$ therefore compares against the actual worst loss over all deletions of at most k records. Under that corollary's curvature interval, exact SDP optimization gives factor $9b/(4a)$ with correction rank at most $\min\{\mathrm{rank}(G), 2rL\}$. The additive optimization and rounding qualifications remain those stated above. The sorting, prefix sums, counts, and response bank are encoding-time objects; the persistent state contains only the original base factors, the correction factor, θ, and n. This is a reduction in request enumeration, not a new memory law: when t is small, exact covariance itself has a representation with $O(dt)$ numerical entries.

Corollary 65 (An inflated floor with a $9/8$ utility factor)**.** *Let $\Sigma \succeq 0$ have eigenpairs (σ_i, v_i) ordered by decreasing eigenvalue, let $0 \le m < d$ be an integer, and suppose $\tau = \sigma_{m+1} > 0$. Set $\delta = 2\tau$ and define*

$$\begin{aligned} F_+ &= \big[\sqrt{\sigma_i - \delta}\, v_i : i \le m,\ \sigma_i > \delta\big], \\ L_{\mathrm{h}} &= \big[\sqrt{\delta - \sigma_i}\, v_i : i \le m,\ \sigma_i < \delta\big], \\ A_\delta &= \delta I + F_+F_+^\top - L_{\mathrm{h}}L_{\mathrm{h}}^\top, \qquad G_\delta = A_\delta - \Sigma. \end{aligned} \tag{300}$$

Fix $B \succeq 0$, $\lambda > 0$, and an integer $0 \le r \le d$. For every reference $R_{\mathrm{ref}} = c(\Sigma + \lambda I)$ with $0 < c \le 1$, Theorem 58 and (285) *construct a correction N_U whose reference utility is at most $9/8$ times the minimum over*

$$\mathcal{C}_{\delta,r} = \{N : 0 \preceq N \preceq G_\delta, \quad \mathrm{rank}(N) \le r\}. \tag{301}$$

No commutation assumption is imposed on B. The resulting covariance has the signed representation

$$\widehat{\Sigma}_U = \delta I + F_+F_+^\top - LL^\top, \qquad L = [L_{\mathrm{h}}, L_G],$$

with at most $m+r$ factor columns in total. Thus the state $(F_+, L, \delta, \theta, n)$ uses at most $d(m+r)+d+2$ numerical entries, apart from dimension metadata.

Suppose additionally that a family of retained curvatures satisfies

$$(1-\eta)(\Sigma + \lambda I) \preceq R_j \preceq \Sigma + \lambda I, \qquad 0 \le \eta < 1.$$

For the product utility $\mathcal{U}$ in Theorem 55, with gap G_δ and bounded nonnegative weights, choose $c = \sqrt{1-\eta}$. Then

$$\mathcal{U}(N_U) \le \frac{9}{8(1-\eta)} \min_{N \in \mathcal{C}_{\delta,r}} \mathcal{U}(N). \tag{302}$$

This comparison is within the inflated-base class $\mathcal{C}_{\delta,r}$; it does not compare against the original spectral-floor optimum or against all summaries at the same storage.

Proof. The head of A_δ equals the head of Σ exactly. Its tail gap eigenvalues satisfy

$$g_i = \delta - \sigma_i \in [\delta/2, \delta], \qquad i > m,$$

so $G_\delta \succeq 0$ and its positive support is the covariance tail. The reference resolvent on this support has eigenvalues

$$q_i = \frac{\delta - \sigma_i}{\delta + c\lambda - (1-c)\sigma_i}.$$

This expression decreases with σ_i. Therefore

$$\kappa_A \le \frac{\delta(1+c) + 2c\lambda}{\delta + c\lambda} = 1 + \frac{c(\delta+\lambda)}{\delta + c\lambda} \le 2.$$

Theorem 58 gives $\gamma_A^2 \le 9/8$. The positive and negative head factors have disjoint nonzero columns indexed by at most m head directions; zero head weights are omitted. The correction factor is supported on the tail and has at most r columns. This proves the storage claim.

For $c = \sqrt{1-\eta}$ the curvature comparison relative to R_{ref} has constants c and c^{-1}. The proof of Theorem 55, applied to a $9/8$-approximate reference minimizer, multiplies its factor by $c^{-2} = (1-\eta)^{-1}$ and proves (302). □

If $m = d$ or $\sigma_{m+1} = 0$, the head already represents Σ exactly: take scalar floor zero and store only its positive spectral factors. No response correction is needed. For bounded ridge data, the sufficient condition (274) supplies the curvature hypothesis of the corollary.

The cubic comparison under stable curvature. The preceding relative guarantee also preserves the cubic dependence on the factor budget under its curvature hypothesis. First suppose $\|x_i\| \le 1$, $y_i = \langle w^*, x_i\rangle$ with $\|w^*\| \le 1$, and $0 < \rho < 1$. For the common DPP response envelope (203), take $r = m$ (or correct the entire tail when it has fewer than m dimensions). Shorting G_δ on the leading m response eigenvectors gives a member N_{sh} of $\mathcal{C}_{\delta,m}$ whose residual E_{sh} annihilates those vectors and satisfies $0 \preceq E_{\mathrm{sh}} \preceq \delta I$. Since $\operatorname{tr}\Sigma \le 1$, one has $\delta \le 2/(m+1)$. For every $\Delta\Delta^\top \preceq \alpha\mathcal{B}_\rho$,

$$\|E_{\mathrm{sh}}\Delta\|^2 \le \delta^2 \alpha \lambda_{m+1}(\mathcal{B}_\rho) \le \frac{\alpha\Gamma_\rho}{\rho(m+1)^3}.$$

For the ridge product relaxation with $0 \preceq C \preceq \Sigma$, $\operatorname{tr} C \le \rho$, $0 \le \alpha \le \rho < 1$, $\beta = 1-\alpha$, and $R = \Sigma - C + \beta\lambda I$, one has $R \succeq \beta\lambda I$. Consequently its exact loss is at most $\|E_{\mathrm{sh}}\Delta\|^2/(\beta^2\lambda)$, and (302) gives

$$\mathcal{U}(N_U) \le \frac{9\Gamma_\rho}{8(1-\eta)(1-\rho)^2\lambda(m+1)^3}. \tag{303}$$

Here the curvature hypothesis must hold over the same product family. There is no claim of an unconditional cubic guarantee for reference-utility selection when that hypothesis fails.

For arbitrary bounded labels, there is a corresponding sampled statement using the envelope actually available to the encoder. Let $R = r_S(\theta)/\lambda > 0$, let $M = \widehat{\mathcal{V}}_q$ be the importance covariance in Proposition 39, and put

$$B_m = \frac{2M}{R} + \frac{I}{m+1}.$$

The deterministic bound $\operatorname{tr} M \le R$ gives $\lambda_{m+1}(B_m) \le 3/(m+1)$. On the event $\mathcal{V} \preceq RB_m$, supplied by that proposition, the general-label response bound (223) places all legal responses in the product ellipsoids

$$\Delta\Delta^\top \preceq \frac{\alpha\Gamma_\rho R}{\rho} B_m.$$

Run the inflated-floor encoder with $B = B_m$ and $r = m$. Shorting on the leading m eigenvectors of B_m gives a comparator with product loss at most $12\Gamma_\rho R/[(1-\rho)^2\lambda(m+1)^3]$. Under the same stable-curvature hypothesis, the resulting product utility therefore satisfies

$$\mathcal{U}(N_U) \le \frac{27\Gamma_\rho R}{2(1-\eta)(1-\rho)^2\lambda(m+1)^3}. \tag{304}$$

On the same event, this absolute bound holds simultaneously for every legal deletion in the specified family. The constant 12 is the product of the squared floor inflation $\delta^2 \le 4/(m+1)^2$ and the response tail bound $3/(m+1)$. For $m = 0$, the envelope event is deterministic because $\mathcal{V} \preceq RI \preceq RB_0$. If $R = 0$, Lemma 38 gives zero labels and zero responses; the decoder is exact without forming B_m.

Computation and numerical qualification. The encoder forms AD_SA, computes a leading eigenspace, and constructs the factor by (284) or (285). It uses a bounded number of spectral decompositions, positive-definite solves, and thin orthogonalizations on matrices of order at most d; standard dense implementations have $O(d^3)$ arithmetic cost at prescribed working precision, once the envelope and reference matrices are supplied. This does not include the cost of obtaining B. The exact inequalities above are linear-algebra statements, not a polynomial bit-complexity claim for exact eigenspaces or exact support determination.

There is also a residual certificate for an approximate subspace. For any orthonormal U on the positive support, the correction remains feasible in exact arithmetic and (283) gives

$$\|Q_U D^{1/2}\|_{\mathrm{op}} \le \sqrt{1 + \|H\|_{\mathrm{op}}^2}\, \|(I - UU^\top)X\|_{\mathrm{op}}.$$

Thus a verified residual bound $\|(I - UU^\top)X\|_{\mathrm{op}} \le \sigma_{r+1}(X) + \epsilon_{\mathrm{sv}}$ yields an additive $\gamma_U \epsilon_{\mathrm{sv}}$ term in the utility square root, without an eigengap assumption. Factor and reference rounding can then be treated by the existing signed-factor perturbation bounds, under their stated floor or factor-norm assumptions. Those bounds control absolute error. Preserving a relative factor after rounding additionally requires that the resulting error be small relative to a positive utility lower bound; zero optimum is covered exactly by the algebraic construction above. For $\|x_i\| \le 1$, the inflated factors satisfy $\|F_+\|_{\mathrm{op}} \le 1$, $\|L\|_{\mathrm{op}} \le \sqrt{2}$, and $F_+^\top L = 0$, hence $\|[F_+, L]\|_{\mathrm{op}} \le \sqrt{2}$ with at most d columns after clamping the correction rank to the tail dimension; these supply the stated factor-norm bounds for signed-factor rounding.

Recovering an exact head from the inflated signed state. The inflated construction also identifies additional exact covariance directions at decoding time. Their recovery uses the spectrum of the combined negative factor; the state need not retain its head–tail partition.

Proposition 66 (Exact directions from the negative factor)**.** *Use the exact inflated state of Corollary 65, with $\delta > 0$, and put $\tau = \delta/2$ and*

$$\Sigma_{\mathrm{hi}} = \delta I + F_+ F_+^\top - LL^\top.$$

Every eigenvector v of $LL^\top$ with eigenvalue $0 < \nu \le \tau$ satisfies

$$\Sigma v = (\delta - \nu)v, \qquad (\Sigma_{\mathrm{hi}} - \Sigma)v = 0. \tag{305}$$

Choose any collection of orthonormal eigenvectors with these eigenvalues, and let V have orthonormal columns spanning their union with $\mathrm{range}(F_+)$*. Then, with $A_V = V^\top \Sigma_{\mathrm{hi}} V$, one has $\Sigma V = V A_V$. In particular, the eligible threshold modes can be omitted, as can any other eligible eigenvectors.*

Let Z complete V to an orthogonal basis. For a legal deletion covariance $0 \preceq C \preceq \Sigma$, write $C_{11} = V^\top C V$, $C_{12} = V^\top C Z$, $C_{21} = C_{12}^\top$, and $C_{22} = Z^\top C Z$. Then the state and deleted features determine

$$\begin{aligned} J &= A_V - C_{11}, \qquad D_{\mathrm{hi}} = Z^\top \Sigma_{\mathrm{hi}} Z, \\ D_{\mathrm{lo}} &= C_{22} + C_{21} J^\dagger C_{12}, \qquad W = D_{\mathrm{hi}} - D_{\mathrm{lo}}, \\ \Sigma_{\mathrm{lo}} &= V A_V V^\top + Z D_{\mathrm{lo}} Z^\top, \end{aligned} \tag{306}$$

and these matrices satisfy

$$J \succeq 0, \quad (I - JJ^\dagger)C_{12} = 0, \quad W \succeq 0, \qquad \Sigma_{\mathrm{lo}} - C \succeq 0, \qquad \Sigma_{\mathrm{lo}} \preceq \Sigma \preceq \Sigma_{\mathrm{hi}}.$$

The centered legal midpoint and posterior bounds of Corollary 98 therefore apply with this V, A_V, Z, W at the same persistent storage.

Proof. The head and tail column spaces of $L = [L_{\mathrm{h}}, G_\delta^{1/2} Z_G]$ are orthogonal, where Z_G has orthonormal columns on the tail by (285). The nonzero head eigenvalues of $LL^\top$ are $\delta - \sigma_i \le \tau$, because $\sigma_i \ge \tau$ for $i \le m$. They have the claimed covariance eigenvalues and zero residual. The nonzero eigenvalues of the tail correction are those of $Z_G^\top G_\delta Z_G \succeq \tau I$. If $Z_G^\top G_\delta Z_G z = \tau z$, positivity of $G_\delta - \tau I$ on the tail gives

$$(G_\delta - \tau I) Z_G z = 0.$$

Thus $u = Z_G z$ satisfies $G_\delta u = \tau u$ and $G_\delta^{1/2} Z_G Z_G^\top G_\delta^{1/2} u = \tau u$. Its residual vanishes and $\Sigma u = (\delta - \tau)u$. These are all tail modes at eigenvalue τ; head and tail directions mixed within that eigenspace have the same covariance eigenvalue. This proves (305). The positive factor is already an exact covariance head and is orthogonal to L, so the claimed invariance of V follows.

Write $D = Z^\top \Sigma Z$. In the basis $[V, Z]$, legality gives

$$\begin{bmatrix} J & -C_{12} \\ -C_{21} & D - C_{22} \end{bmatrix} \succeq 0.$$

The generalized Schur criterion gives the range condition and $D \succeq D_{\mathrm{lo}}$; the covariance upper bound gives $D \preceq D_{\mathrm{hi}}$. The same criterion at $D = D_{\mathrm{lo}}$ gives $\Sigma_{\mathrm{lo}} - C \succeq 0$. This is the interval argument of Proposition 97 with the enlarged exact head. □

The decoder must know that the multiplier-two construction is the publicly specified method. A public method identifier suffices; no data-dependent head partition, threshold, or additional factor is stored. Head directions with zero factor weight can remain in Z. The case $\delta = 0$ uses the exact-covariance branch described above. Using only $\mathrm{range}(F_+)$ gives another valid Schur interval for both floor multipliers one and two, and for the ordinary shorted construction with its exact positive head. Either head rule can be specified in advance as the decoder geometry. These are exact-arithmetic interval statements. Omitting uncertain threshold modes is compatible with the theorem, but rounding of factors, eigenvectors, range conditions, and solves still requires verified enclosures before a numerical implementation can claim a rigorous interval or posterior certificate.

K.3.6 Finite-bit encoding for a listed deletion bank

Theorem 71 gives a polynomial-bit approximation for a listed bank, including exact zero loss when the comparison optimum is zero. Its weighted rational factor format charges one additional scalar per correction column. The following perturbation and representation results preserve the original covariance base and rank-r comparison class while providing its arithmetic guarantees.

Lemma 67 (Certified rational bands and perturbed response points)**.** *Use Q_0, S, A, X_i, f, and f_r^* from Theorem 60, with $M \geq 1$, $k = \mathrm{rank}(Q_0) > 0$, and $1 \leq r \leq k$. Put*

$$\zeta_i = S^\top z_i, \qquad Z_{\max} = \max_i \|\zeta_i\|.$$

Let $x_i \in \mathbb{Q}^k$ satisfy $\|x_i - X_i\| \leq \xi$ for $\xi \geq 0$. Write t_x^* for the optimum of (294) *with x_i in place of X_i. Suppose W has rational orthonormal columns and, for $E_P \geq 0$, satisfies*

$$\dim \mathrm{range}(W) \leq 2r, \qquad \max_i \|(I_k - WW^\top)x_i\|^2 \leq 2(t_x^* + E_P).$$

For $1 \leq r < k$, the rounding step of Proposition 62, applied to the points x_i, provides these conditions with

$$B_x^2 = \max_i \|x_i\|^2, \qquad u = \eta_{\mathrm{tr}} + k\eta,$$

$$E_P = \delta + \eta_c + (\eta + u/r + 2\omega)B_x^2,$$

provided its residual certificates hold and $t \leq t_x^* + \delta$. For $r = k$, one may take $W = I_k$ and $E_P = 0$.

Suppose an orthogonal $O \in \mathbb{Q}^{k\times k}$, a rational diagonal matrix $\mathrm{diag}(d)$, and a rational $\eta_A \geq 0$ satisfy

$$\|A - O\,\mathrm{diag}(d)O^\top\|_{\mathrm{op}} \leq \eta_A, \qquad \min_\ell d_\ell > \eta_A.$$

Define

$$\bar{A} = O\,\mathrm{diag}(d - \eta_A \mathbf{1})O^\top.$$

Partition its diagonal values into $\bar{L}$ nonempty bands of condition number at most two by the greedy scan in Theorem 59, and let P_j be their orthogonal projectors. For each j, let B_j be any full-column-rank rational matrix spanning $\mathrm{range}(P_j W)$, and set

$$K_j = B_j(B_j^\top \bar{A}^{-1} B_j)^{-1} B_j^\top,$$

$$K = \sum_{j=1}^{\bar{L}} K_j, \qquad Q_{\mathrm{safe}} = S(A - K)S^\top,$$

where empty summands are zero. Then

$$\begin{aligned} 0 \preceq Q_{\mathrm{safe}} \preceq Q_0, \\ \operatorname{rank}(Q_0 - Q_{\mathrm{safe}}) = \sum_j \operatorname{rank}(P_j W) \le \min\{k, 2r\bar{L}\}, \\ \sqrt{f(Q_{\mathrm{safe}})} \le \frac{3}{2}\sqrt{f_r^* + E_P} + \left(\frac{3}{2} + \sqrt{\frac{9}{8}}\right)\xi + 2\sqrt{\frac{9}{8}}\,\eta_A Z_{\max}. \end{aligned} \tag{307}$$

If $\eta_A \le \lambda_{\min}(A)/4$, then

$$\bar{L} \le \min\left\{k, \lceil \log_2(2\kappa_A) \rceil\right\}.$$

The loss bound includes $f_r^* = 0$ and requires no singular-value separation or exact identification of a zero-loss branch.

Proof. The diagonalization certificate implies

$$0 \prec \bar{A} \preceq A, \qquad 0 \preceq A - \bar{A} \preceq 2\eta_A I_k.$$

The P_j are rational because O and the diagonal band indicators are rational. Thus each $P_j W$ has a rational column-space basis, and all matrices K_j are rational. The generalized Nyström identity in Theorem 58, applied within each band of $\bar{A}$, gives

$$0 \preceq K_j \preceq P_j \bar{A} P_j, \qquad \operatorname{rank}(K_j) = \operatorname{rank}(P_j W).$$

Their ranges lie in orthogonal bands. Summing gives $0 \preceq K \preceq \bar{A} \preceq A$ and the asserted rank formula, and therefore proves feasibility of Q_{safe}.

Set $c = \sqrt{9/8}$ and

$$T = I_k - K\bar{A}^{-1}.$$

This operator is block diagonal on the bands of $\bar{A}$. To bound both complementary operators, choose an orthonormal basis U_j of $\operatorname{range}(B_j)$ within band j, and complete it by $U_{j,\perp}$. Writing $\bar{A}_j$ for the restriction to that band and

$$H_j = (U_j^\top \bar{A}_j^{-1} U_j)^{-1} U_j^\top \bar{A}_j^{-1} U_{j,\perp},$$

the two blocks in this basis are

$$K_j \bar{A}_j^{-1} = \begin{bmatrix} I & H_j \\ 0 & 0 \end{bmatrix}, \qquad I_j - K_j \bar{A}_j^{-1} = \begin{bmatrix} 0 & -H_j \\ 0 & I \end{bmatrix}.$$

When both subspaces are nonempty, each squared operator norm is $1 + \|H_j\|_{\mathrm{op}}^2$. The Kantorovich calculation in Theorem 58 bounds it by $9/8$ because $\kappa(\bar{A}_j) \le 2$. Empty and fully corrected blocks have operator norms zero or one and obey the same bound. Moreover, $K_j \bar{A}^{-1}$ acts as the identity on $\operatorname{range}(P_j W)$. Consequently

$$TW = 0, \qquad \|T\|_{\mathrm{op}} \le c, \qquad \|I_k - T\|_{\mathrm{op}} \le c.$$

Put $\Pi = WW^\top$. The point approximation and the assumed rounding certificate imply

$$\|(I_k - \Pi)X_i\| \le \sqrt{2(t_x^* + E_P)} + \xi.$$

The original matrix A need not commute with the new bands. Instead, the exact identity

$$(A - K)\zeta_i = TX_i + (I_k - T)(A - \bar{A})\zeta_i$$

and $TX_i = T(I_k - \Pi)X_i$ give

$$\sqrt{f(Q_{\mathrm{safe}})} \le c\sqrt{2(t_x^* + E_P)} + c\xi + 2c\eta_A Z_{\max}.$$

Let Q_* attain the original optimum f_r^*, and let Π_* project onto $\operatorname{range}(A - S^\top Q_* S)$. This projector has rank at most r. For every i,

$$\begin{aligned} \|(I_k - \Pi_*)x_i\| &\le \|(I_k - \Pi_*)X_i\| + \xi \\ &\le \|S^\top Q_* S\zeta_i\| + \xi \le \sqrt{f_r^*} + \xi. \end{aligned}$$

Thus Π_* is feasible in the SDP with objective at most $(\sqrt{f_r^*}+\xi)^2$, and $t_x^* \le (\sqrt{f_r^*}+\xi)^2$. For nonnegative f, E, ξ, one has

$$\sqrt{(\sqrt{f}+\xi)^2+E} \le \sqrt{f+E}+\xi.$$

Substitution and $c\sqrt{2}=3/2$ prove (307).

Finally, if $\eta_A \le \lambda_{\min}(A)/4$, then

$$\lambda_{\min}(\bar{A}) \ge \tfrac{1}{2}\lambda_{\min}(A), \qquad \lambda_{\max}(\bar{A}) \le \lambda_{\max}(A),$$

so $\kappa(\bar{A}) \le 2\kappa_A$. The greedy scan has at most k nonempty bands and at most $\max\{1, \lceil \log_2 \kappa(\bar{A}) \rceil\}$ bands. Since $\kappa_A \ge 1$, this proves the displayed count. □

Proposition 68 (Rational coordinates on a singular covariance support)**.** *Let $R \succ 0$ and $G \succeq 0$ be rational symmetric matrices, with $k = \operatorname{rank}(G) > 0$. In the finite deletion bank, take rational Δ_i and rational $\beta_i > 0$, and set*

$$H_0 = R^{-1/2} G R^{-1/2}, \qquad Q_0 = H_0(I+H_0)^{-1}, \qquad z_i = R^{1/2}\Delta_i/\sqrt{\beta_i}.$$

Choose any full-column-rank rational $V \in \mathbb{Q}^{d\times k}$ spanning $\operatorname{range}(G)$, and define the rational positive-definite matrices

$$H = (V^\top V)^{-1} V^\top G V (V^\top V)^{-1},$$
$$J = V^\top R^{-1} V, \qquad D = H^{-1} + J.$$

Then $G = VHV^\top$. The conceptual orthonormal support basis and its resolvent coordinates can be written as

$$S = R^{-1/2} V J^{-1/2},$$
$$A = S^\top Q_0 S = J^{1/2} D^{-1} J^{1/2},$$
$$v_i = V^\top \Delta_i,$$
$$\zeta_i = S^\top z_i = J^{-1/2} v_i/\sqrt{\beta_i},$$
$$X_i = A\zeta_i = J^{1/2} D^{-1} v_i/\sqrt{\beta_i}.$$

In particular,

$$\|\zeta_i\|^2 = \frac{v_i^\top J^{-1} v_i}{\beta_i}, \qquad \|X_i\|^2 = \frac{v_i^\top D^{-1} J D^{-1} v_i}{\beta_i}$$

are rational.

For the construction of Lemma 67, let B concatenate its nonempty band bases B_j, and put $s = \operatorname{rank}(B)$. If $s > 0$, define

$$F = B^\top\left[\bar{A}^{-1} + (I_k - A)^{-1}\right]B,$$
$$C = (J^{-1/2} + HJ^{1/2})BF^{-1/2},$$
$$L = VC, \qquad N = LL^\top.$$

Then $F \succ 0$, $0 \preceq N \preceq G$, and $\operatorname{rank}(N) = s = \operatorname{rank}(K)$. For $s = 0$, take the empty factor and $N = 0$. This N is the covariance correction corresponding to Q_{safe} under the original reference R and original gap G, so it leaves the original covariance base unchanged. The matrix S is used only to identify the coordinates: the displayed formulas for A, ζ_i, X_i, and C use matrix functions of the reduced matrices and do not require an exact algebraic representation of S. Approximating C and then multiplying by the rational V preserves the exact support $\operatorname{range}(G)$.

Proof. The orthogonal projector onto $\operatorname{range}(G)$ is $V(V^\top V)^{-1}V^\top$. Applying it on both sides of G proves $G = VHV^\top$; positivity of G on its range proves $H \succ 0$. Also $J \succ 0$ and $S^\top S = I_k$. The columns of S span $\operatorname{range}(H_0) = \operatorname{range}(Q_0)$, and

$$S^\top H_0 S = J^{1/2} H J^{1/2}.$$

Taking the resolvent on this support gives

$$A = \left[(J^{1/2}HJ^{1/2})^{-1} + I_k\right]^{-1}$$
$$= J^{1/2}(H^{-1}+J)^{-1}J^{1/2}.$$

Cancellation of $R^{-1/2}R^{1/2}$ gives the expression for ζ_i, and multiplication by A gives that for X_i. Taking squared Euclidean norms proves the two rational formulas.

The band ranges are orthogonal and invariant under $\bar{A}^{-1}$. Consequently B has full column rank, $B^\top \bar{A}^{-1} B$ is block diagonal, and

$$K = B(B^\top \bar{A}^{-1} B)^{-1} B^\top.$$

Write $M_0 = I_d - Q_0$ and $\mathcal{B} = SB$. Since $M_0^{-1} S = S(I_k - A)^{-1}$, the Woodbury identity gives

$$\begin{aligned} &M_0^{-1} - (M_0 + SKS^\top)^{-1} \\ &\quad = M_0^{-1}\mathcal{B}\big[B^\top \bar{A}^{-1} B + \mathcal{B}^\top M_0^{-1}\mathcal{B}\big]^{-1}\mathcal{B}^\top M_0^{-1} \\ &\quad = M_0^{-1} S B F^{-1} B^\top S^\top M_0^{-1}. \end{aligned}$$

Both terms defining F are positive definite on the columns of B, so $F \succ 0$. The first term in its definition uses $\bar{A}$, whereas the second uses the original A because M_0 is the original reference resolvent complement.

On the support, $(I_k - A)^{-1} = I_k + J^{1/2} H J^{1/2}$. Hence

$$\begin{aligned} R^{1/2} M_0^{-1} S &= V J^{-1/2} (I_k - A)^{-1} \\ &= V(J^{-1/2} + H J^{1/2}). \end{aligned}$$

Congruence of the Woodbury identity by $R^{1/2}$ therefore identifies its right-hand side with $LL^\top$ as defined above. Since $I_d - Q_{\mathrm{safe}} = M_0 + SKS^\top$, this is precisely the inverse-resolvent covariance correction. Feasibility of Q_{safe} gives

$$0 \preceq M_0^{-1} - (I_d - Q_{\mathrm{safe}})^{-1} \preceq M_0^{-1} - I_d = H_0.$$

Congruence proves $0 \preceq N \preceq G$. The matrices $R^{1/2}$ and M_0^{-1} are invertible, SB has full column rank, and $F \succ 0$, so $\operatorname{rank}(N) = s$. Finally, the expression $L = VC$ proves the support assertion even when S is irrational. No inverse of a singular covariance matrix is required. □

Factor quantization inside the original support. An entrywise ambient rounding can leave the support of a singular G. The next construction rounds coordinates in an exact rational basis of that support and then contracts the factor. It retains the original covariance base without adding a scalar padding term.

Lemma 69 (Support-preserving rational factor quantization)**.** *Let $G \in \mathbb{Q}^{d\times d}$ be positive semidefinite and let $N = LL^\top \preceq G$, where L has $q = \operatorname{rank}(N)$ columns. If $G = 0$ or $q = 0$, use the empty factor. Otherwise let P_G be its support projector, and let rational $\nu, g > 0$ certify*

$$G \succeq \nu P_G, \qquad \|G\|_{\mathrm{op}} \le g.$$

Suppose $\widehat{L} \in \mathbb{Q}^{d\times q}$ satisfies $\operatorname{range}(\widehat{L}) \subseteq \operatorname{range}(G)$ and $\|\widehat{L} - L\|_F \le h$. Choose rational $e \ge 2\sqrt{g}\,h + h^2$ and set

$$s = (1 + e/\nu)^{-1}, \qquad L_q = s\widehat{L}, \qquad N_q = L_q L_q^\top. \tag{308}$$

Then

$$0 \preceq N_q \preceq G, \qquad \operatorname{rank}(N_q) \le q, \qquad \|N_q - N\|_{\mathrm{op}} \le e + 2eg/\nu. \tag{309}$$

Such a rational $\widehat{L}$ exists at every positive accuracy: take an exact rational basis V_G for $\operatorname{range}(G)$, write $L = V_G T$, and round only T. The basis and the unrounded coordinates are used only during encoding.

Proof. The order $LL^\top \preceq G$ implies $\operatorname{range}(L) \subseteq \operatorname{range}(G)$ and $\|L\|_{\mathrm{op}} \le \sqrt{g}$. Expanding the two Gram matrices therefore gives

$$\|\widehat{L}\widehat{L}^\top - N\|_{\mathrm{op}} \le 2\|L\|_{\mathrm{op}}\|\widehat{L} - L\|_{\mathrm{op}} + \|\widehat{L} - L\|_{\mathrm{op}}^2 \le e.$$

This uses the operator norm of L, so there is no extra factor depending on its number of columns. Both Gram matrices have the same ambient support restriction. Consequently

$$\widehat{L}\widehat{L}^\top \preceq N + eP_G \preceq G + eP_G \preceq (1 + e/\nu)G, \qquad N_q \preceq sG \preceq G.$$

The number of columns remains q. Finally,

$$\|N_q - N\|_{\mathrm{op}} \le s^2 e + (1 - s^2)\|N\|_{\mathrm{op}} \le e + 2eg/\nu,$$

using $1 - (1 + x)^{-2} \le 2x$ for $x \ge 0$. Rounding the coefficients in a fixed full-column-rank V_G approximates L arbitrarily well while preserving the support exactly. The thin-factor assumption is what turns preservation of the column count into nonincrease of the matrix rank; rounding redundant dependent columns need not preserve their dependence. □

Proposition 70 (Actual loss after factor and reference rounding)**.** *Let $\mathcal{J}$ be a nonempty finite bank of legal requests with $\beta_U \ge \beta_0 > 0$, and suppose the source labels satisfy $|y_j| \le 1$. Fix the original $A_0 \succeq \Sigma$ and $G = A_0 - \Sigma$, and put $\theta = w_S$, $R_U = \Sigma - C_U + \beta_U \lambda I$, and $\Delta_U = w_{S-U} - \theta$. For any $0 \preceq N \preceq G$ and reference ϑ, define*

$$\begin{aligned} H_U(N) &= R_U + G - N, \\ w_U(N, \vartheta) &= \vartheta + H_U(N)^{-1}[(C_U + \alpha_U \lambda I)\vartheta - f_U], \\ \mathcal{R}_{\mathcal{J}}(N, \vartheta) &= \max_{U \in \mathcal{J}} [F_{S-U}(w_U(N, \vartheta)) - F_{S-U}(w_{S-U})]. \end{aligned}$$

Thus $\mathcal{R}_{\mathcal{J}}(N, \theta)$ is the loss already defined in Corollary 63. For feasible N, N_q, let $e_N = \|N_q - N\|_{\mathrm{op}}$ and $e_\theta = \|\theta_q - \theta\|$. Then

$$\begin{gathered} \sqrt{\mathcal{R}_{\mathcal{J}}(N_q, \theta_q)} \le \sqrt{\mathcal{R}_{\mathcal{J}}(N, \theta)} + \frac{e_N v_* + (\|A_0\|_{\mathrm{op}} + \lambda) e_\theta}{\beta_0 \sqrt{\lambda}}, \\ v_* := \max_{U \in \mathcal{J}} \|H_U(N)^{-1} R_U \Delta_U\| \le \frac{1}{\sqrt{\beta_0 \lambda}}. \end{gathered} \tag{310}$$

Moreover, $\mathcal{R}_{\mathcal{J}}(N, \theta) \le 1/\beta_0$. In particular, for $0 < \varepsilon \le 1$, retaining θ exactly and requiring

$$e_N \le t_N := \min\{1, \beta_0^2 \lambda \varepsilon / 4\} \quad \Longrightarrow \quad \mathcal{R}_{\mathcal{J}}(N_q, \theta) \le \mathcal{R}_{\mathcal{J}}(N, \theta) + \varepsilon. \tag{311}$$

The rank and comparator class of the ideal construction are unchanged. If its optimum is zero, this finite-precision conclusion is additive; it does not assert exact zero loss for rational factors.

Proof. Fix a request, suppress its subscript, and write $H = H_U(N)$ and $H_q = H_U(N_q)$. The source normal equations give $(C + \alpha\lambda I)\theta - f = R\Delta$. The resolvent identity and the affine dependence on the reference yield

$$\begin{aligned} w_U(N_q, \theta_q) - w_U(N, \theta) = H_q^{-1} \big[&(N_q - N) H^{-1} R\Delta \\ &+ (A_0 - N_q + \lambda I)(\theta_q - \theta) \big]. \end{aligned}$$

In particular, the reference multiplier contains $A_0 - N_q$. Feasibility gives $H_q \succeq R \succeq \beta\lambda I$ and $0 \preceq A_0 - N_q \preceq A_0$. Hence

$$H_q^{-1} R H_q^{-1} \preceq H_q^{-1} \preceq (\beta\lambda)^{-1} I, \qquad \|\beta^{-1/2} R^{1/2} H_q^{-1}\|_{\mathrm{op}} \le \frac{1}{\beta\sqrt{\lambda}}.$$

The excess square root is $\|\beta^{-1/2} R^{1/2}(w - w_{S-U})\|$. Its triangle inequality proves the first bound with the displayed v_*. To bound that quantity, first observe that

$$\|R^{1/2}\Delta\|^2 = \beta[F_{S-U}(\theta) - F_{S-U}(w_{S-U})] \le \beta F_{S-U}(\theta) \le F_S(\theta) \le F_S(0) \le 1.$$

Also $H^{-1} R H^{-1} \preceq (\beta\lambda)^{-1} I$, so $\|H^{-1} R\Delta\| \le \|R^{1/2}\Delta\| / \sqrt{\beta\lambda}$. This proves the uniform bound on v_* without a feature norm or a commutation assumption.

For the ideal-loss bound, put $T = R^{-1/2}(G - N)R^{-1/2}$. Then

$$R^{1/2}(w_U(N, \theta) - w_{S-U}) = -T(I + T)^{-1} R^{1/2}\Delta, \qquad 0 \preceq T(I + T)^{-1} \preceq I.$$

Thus the loss is at most $1/\beta$. Under (311), the added square-root error is at most $t_N / (\beta_0^{3/2}\lambda) \le \sqrt{\beta_0}\varepsilon/4$. Squaring and using the ideal-loss bound adds at most $\varepsilon/2 + \beta_0 \varepsilon^2/16 < \varepsilon$. Taking the maximum over requests proves the assertions. □

Explicit factor bit lengths and persistent storage. Here is a conservative bound which also accounts for a poorly conditioned support basis. Let $B_G \geq 1$ be the total binary encoding length of the rational entries of G, let D_G be the product of their denominators, and put

$$m_G = 2B_G + \lceil \log_2 \max\{2, d\} \rceil, \qquad M_G = 2^{m_G}, \qquad k = \operatorname{rank}(G) > 0.$$

The matrix $D_G G$ is integral, $D_G \leq 2^{B_G}$, and $\|D_G G\|_{\mathrm{op}} \leq M_G$. Its nonzero eigenvalues have a positive integral product, namely the absolute value of its lowest nonzero characteristic coefficient. Therefore the rational choices

$$g = M_G, \qquad \nu = (D_G M_G^{k-1})^{-1}$$

are valid certificates in Lemma 69. Choose V_G to be k independent columns of $D_G G$ by exact elimination. Then $\|V_G\|_{\mathrm{op}} \leq M_G$, and $\det(V_G^\top V_G)$ is a positive integer. Bounding its other eigenvalues by M_G^2 gives

$$\sigma_{\min}(V_G) \geq M_G^{-(k-1)}, \qquad T = V_G^\dagger L, \qquad |T_{ij}| \leq M_G^k.$$

For a rational target $0 < t_N \leq 1$, set

$$u = \left\lceil \log_2 \frac{3M_G(1 + 2M_G/\nu)}{t_N} \right\rceil, \quad h = 2^{-u},$$
$$p = u + m_G + \lceil \log_2 d \rceil, \qquad e = 2M_G h + h^2.$$

Take dyadic $\widehat{T}$ on the 2^{-p} grid with entry errors at most 2^{-p}, and use $\widehat{L} = V_G \widehat{T}$ in (308). Since $q \leq k \leq d$,

$$\|\widehat{L} - L\|_F \leq M_G \sqrt{kq}\, 2^{-p} \leq h, \qquad e + 2eM_G/\nu \leq t_N.$$

All displayed choices use rational arithmetic and comparisons with powers of two. No spectral gap of G is estimated numerically.

For clarity, both the numerator and denominator of each stored entry are charged. Write

$$s = \frac{2^{2u}}{Z}, \qquad Z = 2^{2u} + D_G M_G^{k-1}(2M_G 2^u + 1).$$

The chosen precision gives $e/\nu \leq 1/2$, so $Z < 2^{2u+1}$. Each entry of $\widehat{L}$ has denominator 2^p and magnitude at most $\sqrt{M_G} + h \leq 2M_G$. Thus its scaled entry in $L_q = s\widehat{L}$ has total signed numerator and denominator length at most

$$6u + 3m_G + 2\lceil \log_2 d \rceil + O(1) = O(kB_G + k \log(2 + d) + \log(1/t_N)),$$

with the usual length headers adding only logarithmic overhead. Here one may use $u \leq B_G + (k + 1)m_G + \log_2(1/t_N) + 5$. The potentially large coefficients of T affect intermediate precision and arithmetic; the stored entries are the bounded entries of $sV_G\widehat{T}$. In particular, the displayed bound does not depend on the response-bank size M once G and the target t_N are fixed. Exact elimination, rational multiplication, and this quantization have polynomial bit cost given certified approximations to the coefficients of T at the requested accuracy. This conditional postprocessing assertion does not by itself prove polynomial computation of an arbitrary ideal factor.

For rational source data and rational $\lambda > 0$, the reference $\theta = (\Sigma + \lambda I)^{-1} b$ is rational and may be retained exactly in d entries of polynomial source-input bit length. Taking $\beta_0 = 1/n$ in (311) permits $t_N = \min\{1, \lambda\varepsilon/(4n^2)\}$. Consequently the factor uses dq rational entries with per-entry bit length

$$O\big(kB_G + k \log(2 + d) + \log(2 + n) + \log^+(1/\lambda) + \log(1/\varepsilon)\big), \qquad \log^+ x = \max\{0, \log x\}.$$

Alternatively, the reference can be rounded on a dyadic grid; Proposition 70 charges its norm error explicitly. The persistent state consists of the original base representation, L_q, the reference, n, and dimension metadata. It retains neither V_G, the source records, the response bank, nor request identifiers. The storage cost of the original base must still be included; rationality of its matrix does not by itself provide a short rational square factor.

A complete rational encoder, including zero optimum. An additional scalar weight per correction column permits exact rational representation of a zero-loss correction. This avoids requiring a rational square root of every positive rational pivot. The resulting representation is slightly larger than the unweighted factor format used in the performance experiments.

Theorem 71 (Polynomial-bit approximation on a listed request bank)**.** *Let the bounded source records and the public $0 < \lambda \le 1$ be rational. Suppose a rational covariance base is supplied as*

$$A_0 = aI + F\,\mathrm{diag}(c)F^\top \succeq \Sigma, \qquad F \in \mathbb{Q}^{d\times p}, \quad c \in \mathbb{Q}^p, \quad a \in \mathbb{Q},$$

where the entries of c may have either sign. Let $\mathcal{J}$ be an explicitly listed finite bank of legal requests with nonempty retained sets. Suppose rational $R \succ 0$ and $\kappa \ge 1$ satisfy

$$\kappa^{-1}R \preceq R_U \preceq \kappa R \quad (U \in \mathcal{J}).$$

Fix an integer $0 \le r \le d$ and rational $0 < \gamma \le 1$. Put $G = A_0 - \Sigma$, $k = \mathrm{rank}(G)$, and, when $k > 0$, let A be the positive-support resolvent in Proposition 68. Define

$$L_A = \lceil \log_2(2\kappa_A) \rceil, \qquad \kappa_A = \lambda_{\max}(A)/\lambda_{\min}(A),$$

with $L_A = 1$ if $k = 0$. There is a deterministic algorithm, of polynomial bit complexity in the listed rational input length and $\mathrm{size}(\gamma)$, that outputs $Z \in \mathbb{Q}^{d\times q}$ and positive rational weights $\omega_1, \dots, \omega_q$ such that

$$\begin{aligned} N_{\mathrm{out}} &= Z\,\mathrm{diag}(\omega)Z^\top, \qquad 0 \preceq N_{\mathrm{out}} \preceq G, \\ q &\le \min\{k, 2rL_A\}, \\ \mathcal{R}_{\mathcal{J}}(N_{\mathrm{out}}, \theta) &\le \left(\frac{9}{4}\kappa^2 + \gamma\right) \min_{\substack{0\preceq N\preceq G \\ \mathrm{rank}(N)\le r}} \mathcal{R}_{\mathcal{J}}(N, \theta), \qquad \theta = w_S. \end{aligned} \tag{312}$$

If the comparison optimum is zero, the output has zero loss exactly and $q \le r$. The persistent state is $(a, F, c, Z, \omega, \theta, n)$, with the original base and the exact rational reference retained. It uses

$$d(p + q) + p + q + d + 2$$

rational numerical entries, together with dimension and length metadata. Every entry has polynomial bit length. The reference R, source data, request bank, support bases, and SDP iterates are used only during encoding. The decoder uses the stored state and the deleted records, and can compute its answer by rational arithmetic in polynomial bit time.

Proof. The source moments, each retained optimizer, and θ are rational and are obtained by positive-definite linear solves. Let $\beta_0 = 1/n$, which is valid for every request. Empty banks, $G = 0$, and $r = 0$ use the empty correction; these cases satisfy the claimed comparison directly. Every rank decision performed by the algorithm concerns an explicitly rational matrix and uses exact elimination.

An additive construction. We first prove, for $0 < \varepsilon \le 1$, the same rank bound with

$$\mathcal{R}_{\mathcal{J}}(N_{\mathrm{out}}, \theta) \le \frac{9}{4}\kappa^2\,\mathrm{OPT}_r + \varepsilon, \qquad \mathrm{OPT}_r = \min_{\substack{0\preceq N\preceq G \\ \mathrm{rank}(N)\le r}} \mathcal{R}_{\mathcal{J}}(N, \theta),$$

using unweighted rational factors. If $r \ge k$, approximate a thin factor of G by Lemma 69. An exact rational $LDL^\top$ factorization of its positive support requires only scalar square roots to obtain such a factor to any prescribed accuracy. Proposition 70 then gives loss at most ε with at most k columns. We may therefore assume $1 \le r < k$ for this additive construction.

Use the rational matrices V, H, J, D in Proposition 68. Clearing denominators, bounding integer matrix norms, and using their nonzero characteristic coefficients provide positive rational lower bounds on all positive eigenvalues needed here. Their logarithms have polynomial input length, as in the explicit bound for G above. In particular, a rational $a_A > 0$ with $A \succeq a_A I$ is obtained from

$$\lambda_{\min}(A) \ge \lambda_{\min}(J)/\|D\|_{\mathrm{op}}.$$

The exact algebraic matrix S is never represented.

The matrix functions in these coordinates can be approximated with certified rational errors in polynomial bit time. To see the required primitive, let a rational positive-definite matrix B have a known

positive rational lower spectral bound. Apply rational approximate diagonalization (Del Pia, 2026, Theorem 2) to B at a tolerance smaller than half this bound, and approximate the scalar square roots of the resulting positive diagonal entries by rational bisection. For positive-definite B, C, the equation

$$B^{1/2}(B^{1/2} - C^{1/2}) + (B^{1/2} - C^{1/2})C^{1/2} = B - C$$

bounds the Frobenius error of the square roots by the Frobenius error of $B - C$ divided by $\sqrt{\lambda_{\min}(B)} + \sqrt{\lambda_{\min}(C)}$. Rational inverse perturbation bounds give inverse square roots. Thus only polynomially many precision bits are required. The same procedure applies to a matrix supplied by certified rational approximations, once a positive lower spectral bound is known. Products and rational inverses in the displayed coordinate formulas have explicit norm bounds, so their errors can be propagated by the triangle inequality and the inverse identity.

Choose a rational $Z_0 \geq \max\{1, Z_{\max}\}$ using the exact rational values of $\|\zeta_i\|^2$ in Proposition 68. Set

$$\varepsilon_{\mathrm{ref}} = \frac{\varepsilon}{2\kappa}, \qquad \tau = \min\left\{1, \frac{\varepsilon_{\mathrm{ref}}}{3Z_0 + 1}\right\}, \qquad \xi = \frac{\tau}{12}, \qquad \eta_A = \min\left\{\frac{a_A}{4}, \frac{\tau}{12Z_0}\right\}.$$

Compute rational points x_i within ξ of X_i, and a rational symmetric approximation to A within $\eta_A/2$ in operator norm. Rational approximate diagonalization of the latter, with off-diagonal Frobenius norm at most $\eta_A/2$, gives the O, d certificate of Lemma 67. In particular $\min_j d_j > \eta_A$. All diagonal band comparisons for $\bar{A}$ are rational comparisons.

A polynomial-bit SDP step. Put $B_x^2 = \max_i \|x_i\|^2$ and add the harmless bounds $0 \leq t \leq 4(B_x^2+1)$ to the point-set SDP. They do not change its optimum, since $P = 0$ attains a value at most B_x^2. The feasible region in (P, t) contains a ball, in the product Frobenius and Euclidean norm, of rational radius

$$\rho = \frac{r}{8k(B_x^2 + 1)} \quad \text{about} \quad P_0 = \frac{r}{2k} I_k, \qquad t_0 = 2B_x^2 + 1.$$

Indeed, a perturbation of this size preserves both matrix orders, uses less than the trace slack $r/2$, and changes each scalar point constraint by at most $(B_x^2+1)\rho$. The entire feasible region is contained in a ball about this point of radius $4k + 8(B_x^2 + 1)$. Introduce the scalar slacks and the two matrix blocks $P, I - P$ in one block-diagonal positive-semidefinite matrix. The resulting affine-slice SDP has rational data and the same inner-ball bound: the norm of its perturbation dominates the norm of $(\Delta P, \Delta t)$. Its outer radius can be multiplied by the rational bound $\Gamma_{\mathrm{emb}} = 4 + k + M(1 + B_x^2)$ on this affine embedding's operator norm. The logarithm of this factor has polynomial input length. The bounded-interior SDP complexity theorem (de Klerk & Vallentin, 2016, Theorem 1.1) therefore computes a rational *feasible* P, t with $t \leq t_x^* + \delta$ in polynomial bit time for any positive rational δ.

Take

$$\delta = \frac{\tau^2}{72}, \qquad \omega = \frac{\tau^2}{144(1 + B_x^2)}.$$

Rational approximate diagonalization of P to residual ω and the safe threshold $1/2 + \omega$ give the subspace W of Proposition 62. Here the SDP feasibility residuals are zero, so

$$E_P = \delta + 2\omega B_x^2 \leq \frac{\tau^2}{36}.$$

The proof does not interpret a floating-point solver status as a feasibility or objective certificate.

Preserving the original comparison through arithmetic. Construct the rational band bases and their shorted correction as in Lemma 67. Its square-root bound, $\sqrt{9/8} < 3/2$, and $\sqrt{f_r^*} \leq Z_0$ imply

$$\sqrt{f(Q_{\mathrm{safe}})} \leq \frac{3}{2}\sqrt{f_r^*} + e_0, \qquad e_0 \leq \frac{3}{2}\sqrt{E_P} + 3\xi + 3\eta_A Z_0 \leq \frac{3}{4}\tau.$$

Since $\tau \leq 1$,

$$f(Q_{\mathrm{safe}}) \leq \frac{9}{4} f_r^* + 3Z_0\tau + \tau^2 \leq \frac{9}{4} f_r^* + \varepsilon_{\mathrm{ref}}.$$

The actual-to-reference curvature comparison from Corollary 63 now gives

$$\mathcal{R}_{\mathcal{J}}(N, \theta) \leq \frac{9}{4}\kappa^2 \,\mathrm{OPT}_r + \varepsilon/2.$$

Here $N \preceq G$ is the ideal covariance correction obtained from the mixed Gram formula, and its column count is at most $\min\{k, 2rL_A\}$.

To compute its factor, form the rational bases B_j by exact elimination and use the coefficient formula $L = VC$ in Proposition 68. The only further nonrational operations are certified square roots and inverse square roots. The mixed Gram has a positive lower bound from

$$B^\top[\bar{A}^{-1} + (I - A)^{-1}]B \succeq B^\top B.$$

The matrix B is rational and has independent columns, so this lower bound has polynomial logarithmic size in the already computed data. One can also use $(I - A)^{-1} = I + J^{1/2}HJ^{1/2}$, avoiding an uncertain subtraction or inverse near the upper resolvent endpoint. After the bases are fixed, refine these matrix-function approximations to the precision required for their coefficient factor. Poor conditioning of these bases affects temporary precision; it does not change the bases or create a circular precision requirement. There are only a fixed number of these approximation stages, each polynomial in the bit lengths of the preceding stage and its requested accuracy.

Apply Lemma 69 with

$$t_N = \min\{1, \beta_0^2\lambda\varepsilon/8\}.$$

The explicit integer support basis above can be used for V from the outset, or obtained by an exact rational change of coordinates. Obtain coefficient errors at most half the chosen dyadic grid step before rounding; the final coefficient error is then at most one step. Proposition 70, with budget $\varepsilon/2$, proves the asserted additive guarantee. This finishes its polynomial-bit construction, including singular G and zero ideal reference loss. No exact zero output is required from this unweighted additive step.

Detecting and representing zero optimum. Form $v_i = V^\top\Delta_i$ exactly, and let $q_0 = \operatorname{rank}[v_1, \dots, v_M]$. The coordinate identities give

$$\operatorname{rank}[X_i] = q_0 = \operatorname{rank}[G\Delta_i].$$

Actual loss at a request is zero exactly when $(G - N)\Delta_i = 0$. Necessity follows from the invertible retained normal matrices in the exact error identity. Hence a rank-r zero-loss correction requires $q_0 \le r$. Conversely, suppose $q_0 \le r$ and select q_0 response columns $Y = [\Delta_{i_1}, \dots, \Delta_{i_{q_0}}]$ with independent $V^\top Y$. If $q_0 = 0$, use $N = 0$. Otherwise define

$$B_0 = GY, \qquad C_0 = Y^\top GY \succ 0, \qquad N_0 = B_0C_0^{-1}B_0^\top.$$

The shorting identity gives $0 \preceq N_0 \preceq G$ and $(G - N_0)Y = 0$. Every remaining response is a combination of the selected columns modulo $\ker G$, so $(G - N_0)\Delta_i = 0$ for the entire bank. Thus $\mathrm{OPT}_r = 0$ if and only if $q_0 \le r$. An exact rational factorization

$$C_0 = L_0 \operatorname{diag}(p_j) L_0^\top, \qquad p_j > 0, \quad Z = B_0 L_0^{-\top}, \quad \omega_j = 1/p_j$$

represents $N_0 = Z \operatorname{diag}(\omega) Z^\top$ in $dq_0 + q_0$ rational entries. Its actual loss is zero with the exact stored θ. These weights remove the rational-square-root obstruction of an unweighted format.

A positive lower bound without optimization. If $q_0 > r$, select $q_0 \le k$ independent columns v_i and put

$$T_0 = D^{-1}\left(\sum_{i \text{ selected}} \frac{v_i v_i^\top}{\beta_i}\right) D^{-1}.$$

This is a rational positive-semidefinite matrix of rank q_0. Obtain positive rational lower bounds ν_J and ν_T on $\lambda_{\min}(J)$ and the smallest positive eigenvalue of T_0 by the integer characteristic-coefficient argument used above. Since

$$X_{\mathrm{sel}}X_{\mathrm{sel}}^\top = J^{1/2}T_0J^{1/2}, \qquad \lambda_{\min}^+(J^{1/2}T_0J^{1/2}) \ge \nu_J\nu_T,$$

every rank-r comparison residual has worst point loss at least its squared spectral residual divided by q_0. It follows that

$$f_r^* \ge \frac{\nu_J\nu_T}{k}, \qquad \mathrm{OPT}_r \ge \mu := \frac{\nu_J\nu_T}{\kappa k} > 0.$$

The nonzero spectra in the display agree with those of $T_0^{1/2}JT_0^{1/2}$ on its positive support, which proves the eigenvalue lower bound. The number μ has polynomial encoding length and $\log^+(1/\mu)$

is polynomially bounded. Use the additive algorithm with $\varepsilon = \min\{1, \gamma\mu\}$. Its error is at most $\gamma \operatorname{OPT}_r$, proving (312). Set all output weights to one in this nonzero branch.

Persistent state and decoding. The original base arrays are retained exactly and charged. The correction uses $dq + q$ numerical entries; the reference uses d. The zero branch uses rational linear algebra only. In the other branch, the preceding support-quantization bound charges both the numerator and denominator of every output entry. The lower bound μ uses at most k actual requests, rather than an M-column singular-value bound. Each such response has bit length bounded in terms of the source, base, reference, and public parameters, independently of how many other requests are listed. Thus the persistent precision can be bounded by a polynomial in those lengths and $\operatorname{size}(\gamma)$, with no additional bank-size factor; the running time still depends polynomially on the full listed bank. Neither the selected request identifiers nor R, V, H, J, D, B or the SDP solution is stored.

Given the deleted records, the decoder forms their moments and uses $A_0 - N_{\mathrm{out}}$ in the centered update. Its normal matrix dominates $R_U \succeq \beta_U \lambda I$, so it is positive definite. All stored and request quantities are rational, and exact linear solves produce a rational answer of polynomial bit length. This establishes the decoder claim and completes the proof. □

Corollary 72 (Rational balancing of the curvature comparison)**.** *Suppose rational $R_+ \succ 0$ and rational $0 < a \le b$ satisfy $aR_+ \preceq R_U \preceq bR_+$ on a listed bank, under the other rational-input assumptions of Theorem 71. For any rational $0 < \zeta \le 1$, its weighted rational construction attains*

$$\mathcal{R}_{\mathcal{J}}(N_{\mathrm{out}}, \theta) \le \left(\frac{9b}{4a} + \zeta\right) \operatorname{OPT}_r$$

in polynomial bit time, with the theorem's rank bound for the chosen rational reference and exact loss zero when $\operatorname{OPT}_r = 0$.

Proof. Set $\eta = \zeta a/(27b)$. Rational bisection chooses $\sqrt{ab} \le c \le (1+\eta)\sqrt{ab}$ in polynomial bit time. Take $R = cR_+$ and the rational $\kappa = \max\{c/a, b/c\} \ge 1$. Then $\kappa \le (1+\eta)\sqrt{b/a}$, so

$$\frac{9}{4}\kappa^2 \le \frac{9b}{4a} + \frac{\zeta}{4},$$

using $(1+\eta)^2 - 1 \le 3\eta$. Apply the theorem with $\gamma = \zeta/2$. □

The theorem gives a polynomial-bit approximation within the original covariance-correction class, at the cost of additional correction columns and up to one additional scalar per column. Its comparator is this class, not all summaries of equal bit length, and its precision bound is weaker than the main memory law's per-entry logarithm. A listed bank may still be exponentially large for a general deletion budget. The endpoint bank of Theorem 64 supplies a polynomial list for fixed feature-type count. The weighted rational format and the certified arithmetic in this theorem have not been evaluated by the sampled-response performance grid.

K.3.7 A SINGLE SDP FOR ALL CENTERED DELETION SUMS

The positive-semidefinite Grothendieck relaxation and its Gaussian rounding inequality are classical (Nesterov, 1998; Briët et al., 2010). We incorporate the deletion budget through a centered vector-sum comparison and minimize the dual relaxation jointly with the response subspace. The encoder solves a single SDP of size polynomial in the number of records, without deletion-subset enumeration or an adaptive worst-request oracle. We first analyze the frozen-reference utility defined below; transfer to actual retained loss requires a separate bound.

Lemma 73 (Centered sums and the Grothendieck upper bound)**.** *Let $u_1, \dots, u_n \in \mathbb{R}^d$ satisfy $\sum_i u_i = 0$, where $n \ge 2$, and let $M \succeq 0$. Write $U = [u_1, \dots, u_n]$ and*

$$F_{\mathrm{all}}(M) = \max_{I \subseteq [n]} \Big\| \sum_{i \in I} u_i \Big\|_M^2,$$

$$\mathfrak{G}(M) = \frac{1}{4} \max_{T \succeq 0,\ \operatorname{diag} T = \mathbf{1}} \langle U^\top M U, T \rangle$$

$$= \frac{1}{4} \min_y \left\{ \mathbf{1}^\top y : \operatorname{Diag}(y) - U^\top M U \succeq 0 \right\}.$$

For an integer $1 \le K \le n-1$, *put*

$$m_0 = \lfloor n/2 \rfloor, \quad K_0 = \min\{K, m_0\}, \quad \beta_j = 1 - j/n,$$
$$c_K = (m_0/K_0)^2, \qquad D_K = \max\{\beta_1, \beta_{K_0} c_K\}.$$

Define the budgeted unweighted and weighted maxima by

$$F_K(M) = \max_{|I|\le K} \Big\| \sum_{i\in I} u_i \Big\|_M^2, \qquad F_K^{\mathrm{w}}(M) = \max_{|I|\le K} \frac{1}{\beta_{|I|}} \Big\| \sum_{i\in I} u_i \Big\|_M^2.$$

Then

$$\begin{aligned} F_{\mathrm{all}}(M) &\le \mathfrak{G}(M) \le \frac{\pi}{2} F_{\mathrm{all}}(M), \\ F_K(M) &\le \mathfrak{G}(M) \le \frac{\pi}{2} c_K F_K(M), \\ F_K^{\mathrm{w}}(M) &\le \frac{\mathfrak{G}(M)}{\beta_K}, \\ \mathfrak{G}(M) &\le \frac{\pi}{2} D_K F_K^{\mathrm{w}}(M). \end{aligned} \tag{313}$$

The constants c_K *in* $F_{\mathrm{all}} \le c_K F_K$ *and, for* $K \le m_0$, D_K *in* $F_{\mathrm{all}} \le D_K F_K^{\mathrm{w}}$ *are sharp over centered vector lists.*

Proof. Centering gives the exact identity

$$F_{\mathrm{all}}(M) = \frac{1}{4} \max_{s\in\{-1,1\}^n} s^\top U^\top M U s.$$

The correlation-matrix SDP is strictly feasible at $T = I_n$; its dual is strictly feasible for a sufficiently large constant vector y. This proves the displayed dual equality. The positive-semidefinite Grothendieck inequality gives the first line of (313). For completeness, Gaussian signs from a Gram representation of T have expected correlation $(2/\pi)\arcsin(T)$, entrywise. The positive coefficients of the power series of $\arcsin(t) - t$ and the Schur product theorem imply $\arcsin(T) - T \succeq 0$. Its inner product with $U^\top M U \succeq 0$ is nonnegative, giving the required $2/\pi$ rounding bound.

Complementation leaves the norm of a sum unchanged. Therefore an unrestricted maximizer has a representative S with $s = |S| \le m_0$. If $s > K_0$, put $v = \sum_{i\in S} u_i$ and choose the K_0 indices in S with largest scores $u_i^\top M v$. For their sum v_0, averaging and Cauchy–Schwarz give

$$v_0^\top M v \ge \frac{K_0}{s} \|v\|_M^2, \qquad \|v_0\|_M^2 \ge \left(\frac{K_0}{s}\right)^2 \|v\|_M^2.$$

The assertion is immediate if $\|v\|_M = 0$. If $s \le K_0$, use S itself. This proves $F_{\mathrm{all}} \le c_K F_K$ without a partition or a ceiling loss. For the weighted version, $s \ge K_0$ gives

$$F_{\mathrm{all}} \le \beta_{K_0} (s/K_0)^2 F_K^{\mathrm{w}} \le \beta_{K_0} c_K F_K^{\mathrm{w}},$$

whereas $1 \le s < K_0$ gives $F_{\mathrm{all}} \le \beta_s F_K^{\mathrm{w}} \le \beta_1 F_K^{\mathrm{w}}$. Finally $F_K^{\mathrm{w}} \le F_{\mathrm{all}}/\beta_K$.

Sharpness already occurs in one dimension. Take m_0 copies of 1 and $n - m_0$ copies of $-m_0/(n-m_0)$. This gives the ratio c_K in the unweighted comparison and, when $K \le m_0$, the ratio $\beta_K (m_0/K)^2$ in the weighted comparison. The centered list $1, -1/(n-1), \dots, -1/(n-1)$ gives weighted ratio β_1 for $K \le m_0$. These attain the larger of the two branches defining D_K. □

The truncation in this proof is deterministic once a sign vector has been chosen. Combining it with Gaussian sign rounding gives an expected unweighted approximation ratio $(2/\pi)(K_0/m_0)^2$. The joint encoder below uses the SDP upper bound directly, so it need not sample signs or implement a randomized request oracle.

Theorem 74 (One SDP for weighted frozen-reference utility)**.** *Let* $0 \preceq Q_0 \prec I_d$, *and let* $a_1, \dots, a_n \in \mathbb{R}^d$ *satisfy* $\sum_i a_i = 0$. *Fix* $n \ge 2$, $1 \le K \le n-1$, *and integer* $0 \le r \le d$. *Use* $m_0, K_0, \beta_j, c_K, D_K$ *from Lemma 73, and define*

$$\mathcal{F}_K(Q) = \max_{|I|\le K} \frac{\|Q \sum_{i\in I} a_i\|^2}{\beta_{|I|}},$$
$$f_r^* = \min_{\substack{0\preceq Q\preceq Q_0 \\ \mathrm{rank}(Q_0 - Q)\le r}} \mathcal{F}_K(Q).$$

For $k = \operatorname{rank}(Q_0) > 0$, let S have orthonormal columns spanning its support, and put

$$A = S^\top Q_0 S \succ 0, \qquad X_i = AS^\top a_i, \qquad X = [X_1, \dots, X_n].$$

If $Q_0 = 0$, use $Q = 0$. If $r = 0$, use $Q = Q_0$. If $\operatorname{rank}(X) \le r$, the global oblique correction on $\operatorname{range}(X)$ gives zero loss and correction rank at most r. This includes $r \ge k$. Otherwise consider the single SDP

$$\begin{aligned} \Phi_* = \min_{P,y} \quad & \frac{1}{4}\mathbf{1}^\top y \\ \textit{subject to} \quad & 0 \preceq P \preceq I_k, \qquad \operatorname{tr} P \le r, \\ & \operatorname{Diag}(y) - X^\top (I_k - P) X \succeq 0. \end{aligned} \tag{314}$$

Let (P, y) be feasible with objective at most $\Phi_* + \delta$, where $\delta \ge 0$, and let W contain orthonormal eigenvectors spanning the eigenspaces of P with eigenvalues at least $1/2$. Partition the positive spectrum of A into L nonempty bands of condition number at most two, with $L \le \max\{1, \lceil \log_2 \kappa_A \rceil\}$. For their projectors P_j, put $U_j = \operatorname{orth}(P_j W)$ and define

$$K_{\mathrm{imp}} = \sum_j U_j (U_j^\top A^{-1} U_j)^{-1} U_j^\top, \qquad Q_{\mathrm{imp}} = S(A - K_{\mathrm{imp}})S^\top.$$

Empty summands are zero. Then

$$\begin{aligned} & 0 \preceq Q_{\mathrm{imp}} \preceq Q_0, \qquad \operatorname{rank}(Q_0 - Q_{\mathrm{imp}}) \le \min\{k, 2rL\}, \\ & \mathcal{F}_K(Q_{\mathrm{imp}}) \le C_K f_r^* + \frac{9\delta}{4\beta_K}, \qquad C_K = \frac{9\pi D_K}{8\beta_K}. \end{aligned} \tag{315}$$

For $K \le m_0$, this constant simplifies to

$$C_K = \frac{9\pi}{8} \max\left\{ \frac{n-1}{n-K}, \left(\frac{m_0}{K}\right)^2 \right\}.$$

For the unweighted utility, the same SDP and construction give factor $(9\pi/8)c_K$, with additive term $9\delta/4$. The rank-r comparator is the original covariance-correction class under the inverse resolvent map.

Proof. The compact comparison class has a minimizer Q_*. Let Π_* project onto $\operatorname{range}(A - S^\top Q_* S)$. It has rank at most r. For every subset I,

$$\Big\| (I_k - \Pi_*) \sum_{i \in I} X_i \Big\| \le \Big\| S^\top Q_* S \sum_{i \in I} S^\top a_i \Big\|.$$

Lemma 73, with metric $I_k - \Pi_*$ and centered columns X_i, therefore implies

$$\Phi_* \le \frac{\pi}{2} D_K f_r^*.$$

This application minimizes the Grothendieck dual jointly with P; it does not interchange a maximum and a minimum.

The threshold projector $\Pi = WW^\top$ satisfies $\dim W \le 2 \operatorname{tr} P \le 2r$ and $I_k - \Pi \preceq 2(I_k - P)$. Thus for every I,

$$\Big\| (I_k - \Pi) \sum_{i \in I} X_i \Big\|^2 \le 2 \Big(\sum_{i \in I} X_i \Big)^\top (I_k - P) \Big(\sum_{i \in I} X_i \Big) \le \frac{1}{2} \mathbf{1}^\top y.$$

The last inequality uses the unrestricted-subset upper bound in Lemma 73. The band argument of Theorem 59 gives feasibility, the rank assertion, and

$$T = I_k - K_{\mathrm{imp}} A^{-1}, \qquad TW = 0, \qquad \|T\|_{\mathrm{op}}^2 \le 9/8.$$

Consequently $Q_{\mathrm{imp}} \sum_{i \in I} a_i = ST(I_k - \Pi) \sum_{i \in I} X_i$. Divide the squared bound by $\beta_{|I|} \ge \beta_K$ to obtain $\mathcal{F}_K(Q_{\mathrm{imp}}) \le 9(\Phi_* + \delta)/(4\beta_K)$. The unweighted version uses the unweighted comparison in the lemma and omits this division.

Every singleton is legal. Zero optimum therefore implies $X = (A - S^\top Q_* S)[S^\top a_i]_{i=1}^n$ and hence $\operatorname{rank}(X) \le r$. Conversely, the global oblique correction on $\operatorname{range}(X)$ annihilates all singleton residuals, and therefore all their sums. This proves the zero and endpoint assertions. □

A bounded rational SDP and approximate subspaces. For rational centered $X \in \mathbb{Q}^{k\times n}$ and $1 \le r \le k$, the SDP in (314) has a quantitative interior. Put $b_X = \|X\|_F^2$ and add $0 \le y_i \le Y_{\max} := 2n(b_X+1)$. This does not change the optimum: $y_i \ge 0$ follows already from the diagonal of the main matrix inequality, while the feasible constant vector $y = (b_X+1)\mathbf{1}$ bounds the optimal sum by $n(b_X+1)$. A bounded sublevel set therefore contains an optimum. The center

$$P_0 = \frac{r}{2k} I_k, \qquad y_0 = (b_X+1)\mathbf{1}$$

has main matrix slack at least I_n. In the product Frobenius and Euclidean norm it contains a feasible ball of radius

$$\rho = \frac{r}{8k(1+b_X)}.$$

Indeed, the main matrix perturbation has operator norm at most $(1+b_X)\rho$, the trace perturbation is at most $\sqrt{k}\rho$, and the remaining matrix and scalar margins are larger than these perturbations. An outer radius about this center is $R_0 = 4k + 4n^2(b_X+1)$.

Embed the constraints into the affine slice of the block matrix

$$\mathrm{Diag}_{\mathrm{block}}\big(P,\, I_k - P,\, r - \operatorname{tr} P,\, \mathrm{Diag}(y),\, \mathrm{Diag}(Y_{\max}\mathbf{1} - y),\, \mathrm{Diag}(y) - X^\top (I_k - P)X\big)\,.$$

Its perturbation norm dominates the product norm and is at most $\Gamma = 4 + k + b_X$ times that norm. Thus ρ and ΓR_0 are valid inner and outer radii in the affine-slice Frobenius norm. The bounded-interior complexity theorem (de Klerk & Vallentin, 2016, Theorem 1.1) computes an exactly rational feasible pair of objective at most $\Phi_* + \delta$ in polynomial bit time for positive rational δ.

Rational approximate diagonalization of its P, followed by the threshold $1/2 + \omega$, gives an exactly rational orthonormal W with at most $2r$ columns and

$$\|(I_k - WW^\top)x\|^2 \le 2\big[x^\top (I_k - P)x + 2\omega\|x\|^2\big] \quad (x \in \mathbb{R}^k). \tag{316}$$

This is Proposition 62 with zero SDP feasibility residuals. In particular, rational data and a prescribed precision suffice; no separation at the threshold is required.

Corollary 75 (Rational encoding of weighted frozen utility)**.** *Let the source, its ridge parameter, and the supplied original weighted rational base be as in Theorem 71. Let $R \succ 0$ be rational, and let rational $g_1, \dots, g_n \in \mathbb{Q}^d$ satisfy $\sum_i g_i = 0$. Use $G = A_0 - \Sigma$, $1 \le K \le n-1$, and set*

$$\Delta_I^0 = R^{-1}\frac{1}{n}\sum_{i\in I} g_i, \qquad a_i = R^{-1/2} g_i / n.$$

For $0 \preceq N \preceq G$, define

$$\mathcal{F}_K(N) = \max_{|I|\le K} \frac{\|R^{1/2}(R+G-N)^{-1}(G-N)\Delta_I^0\|^2}{\beta_{|I|}}.$$

This equals $\mathcal{F}_K(Q_N)$ under the original reference resolvent map. For integer $0 \le r \le d$ and rational $0 < \gamma \le 1$, there is a deterministic algorithm of polynomial bit complexity in the rational input length and $\mathrm{size}(\gamma)$ which outputs

$$N_{\mathrm{out}} = Z\,\mathrm{diag}(\omega)Z^\top, \qquad Z \in \mathbb{Q}^{d\times q}, \quad \omega_j \in \mathbb{Q}_{>0},$$

such that

$$\begin{aligned} 0 \preceq N_{\mathrm{out}} \preceq G, \qquad & q \le \min\{k, 2rL_A\}, \\ \mathcal{F}_K(N_{\mathrm{out}}) \le (C_K + \gamma) \min_{\substack{0\preceq N\preceq G \\ \mathrm{rank}\, N \le r}} & \mathcal{F}_K(N), \end{aligned} \tag{317}$$

where $k = \mathrm{rank}(G)$, $L_A = \lceil \log_2(2\kappa_A)\rceil$ for $k > 0$, and $L_A = 1$ otherwise. The constant C_K is the one in (315)*. Zero optimum is attained exactly with $q \le r$. The original weighted base, correction columns and weights, exact rational $\theta = w_S$, and n use $d(p+q) + p + q + d + 2$ rational entries, plus dimension and length metadata. Their bit lengths are polynomial. The algorithm forms no subset bank and stores neither the g_i nor any SDP or support coordinates. This is a frozen-utility guarantee.*

Proof. Only singleton source columns are needed. For $\Delta_i^0 = R^{-1}g_i/n$, use the rational support coordinates V, H, J, D from Proposition 68, with the column normalization $\beta_i = 1$. Thus

$$\zeta_i = J^{-1/2}V^\top \Delta_i^0, \qquad X_i = J^{1/2}D^{-1}V^\top \Delta_i^0.$$

Their sums are zero. The denominator-clearing spectral bounds and the rational square-root procedure in the proof of Theorem 71 supply certified polynomial-bit approximations to these quantities and to A, including rational lower bounds $A \succeq a_A I$ and $R \succeq a_R I$ with $a_A, a_R > 0$. All inverse and square-root operations take place on known positive-definite supports.

The empty-support and $r = 0$ endpoints use the empty correction. The zero test below handles $\operatorname{rank}(X) \le r$. For the additive construction assume $1 \le r < \operatorname{rank}(X) \le k$. First construct an additive approximation, of error at most $0 < \varepsilon \le 1$. Write $\beta = \beta_K$ and choose rational

$$\begin{aligned}
T_R &= 1 + K_0^2 \max_i \frac{g_i^\top R^{-1} g_i}{n^2}, \\
Z_0 &\ge \max\{1, K_0 \max_i \|\zeta_i\|\}, \\
M_0 &= 1 + 5D_K/\beta, \qquad U_0 = 1 + T_R/\beta, \\
\tau &= \min\left\{1, \frac{\varepsilon}{2(2M_0U_0 + 1)}\right\}, \\
\chi &= \frac{\beta\tau}{12(M_0 + 2)}, \qquad \eta_A = \min\{a_A/4, \beta\tau/(12Z_0)\}.
\end{aligned}$$

The squared norms defining Z_0 are rational by the coordinate identity in that proposition, so a suitable rational upper bound is computed by rational comparisons. Centering and complementation give $\|\sum_{i\in I} a_i\|^2 \le T_R$ and $\|\sum_{i\in I}\zeta_i\| \le Z_0$ for every legal I. In particular every feasible ideal correction has $\sqrt{\mathcal{F}_K(N)} \le \sqrt{T_R/\beta} \le U_0$. Also $M_0 \ge \sqrt{C_K}$.

Approximate each X_i rationally to error at most $\chi/(2K_0)$ and subtract the mean of the approximate columns. The resulting rational columns x_i have sum exactly zero and individual errors at most χ/K_0. For all legal I, complementation then bounds the error of their sum by χ. This exact centering is needed for the Grothendieck identity. Construct the conservative rational $\bar{A} \preceq A$ and its rational bands as in Lemma 67, with error parameter η_A. Their number is at most L_A.

Apply the bounded rational SDP above to $x = [x_1, \dots, x_n]$. Put $B = K_0^2 \max_i \|x_i\|^2$ and choose

$$\delta = \frac{\beta^2\tau^2}{72}, \qquad \omega = \frac{\beta^2\tau^2}{144(1 + B)}, \qquad E = \delta + 2\omega B \le \frac{\beta^2\tau^2}{36}.$$

Let f_r^* now denote the original weighted frozen optimum. Its comparator projector in the theorem has weighted residual at most $\sqrt{f_r^*} + \chi/\sqrt{\beta}$ on the approximate columns. Therefore the approximate-column SDP optimum obeys

$$\Phi_x \le \frac{\pi}{2} D_K (\sqrt{f_r^*} + \chi/\sqrt{\beta})^2.$$

The safe threshold (316) and the conservative band proof, with $c = \sqrt{9/8} < 3/2$, give

$$\sqrt{\mathcal{F}_K(Q_{\mathrm{safe}})} \le \frac{1}{\sqrt{\beta}} \left[\frac{3}{2}\sqrt{\Phi_x + E} + c\chi + 2c\eta_A Z_0\right] \le \sqrt{C_K}\sqrt{f_r^*} + \tau.$$

Indeed the three additional terms are bounded by $(M_0 + 2)\chi/\beta$, $3\sqrt{E}/(2\beta)$, and $3\eta_A Z_0/\beta$, whose sum is at most τ. The choice of τ consequently gives

$$\mathcal{F}_K(Q_{\mathrm{safe}}) \le C_K f_r^* + \varepsilon/2.$$

The factor of its ideal covariance correction is computed by the mixed Gram formula in Proposition 68. Freeze the rational bands and column bases before refining matrix-function precision. Their mixed Gram is bounded below by $B_{\mathrm{bas}}^\top B_{\mathrm{bas}} \succ 0$, where B_{bas} is their full-column-rank rational concatenation. Thus the additional working precision has polynomial logarithmic size, as in the preceding rational encoder.

To charge final quantization directly in the frozen metric, let $H = R+G-N$ and $H_q = R+G-N_q$ for two feasible corrections. The resolvent identity and $H, H_q \succeq R \succeq a_R I$ imply

$$\sqrt{\mathcal{F}_K(N_q)} \le \sqrt{\mathcal{F}_K(N)} + \frac{\|N_q - N\|_{\mathrm{op}}\sqrt{T_R}}{\sqrt{\beta}\, a_R}.$$

For example, write the decoder displacement as $H^{-1}R\Delta_I^0$; its norm is at most $\sqrt{T_R/a_R}$, and the weighted norm of $R^{1/2}H_q^{-1}$ is at most $1/\sqrt{\beta a_R}$. Use Lemma 69 with

$$t_N = \min\{1, \varepsilon\beta a_R/(8T_R)\}.$$

Since the ideal frozen loss is at most T_R/β, squaring the preceding perturbation bound adds less than $\varepsilon/2$. This completes the additive construction with the original base and at most $\min\{k, 2rL_A\}$ columns. All its output weights may be set to one.

For exact zero detection, form the rational columns $v_i = V^\top \Delta_i^0$. Their rank equals $\operatorname{rank} X$. If this rank q_0 is at most r, select independent columns $Y = [\Delta_i^0]$ as in the zero branch of Theorem 71 and use

$$N_0 = GY(Y^\top GY)^{-1}Y^\top G.$$

It annihilates all frozen singleton residuals and hence all subset residuals. Rational $LDL^\top$ factorization of $Y^\top GY$ gives its exact weighted representation with q_0 columns. The case $q_0 = 0$ uses the empty correction. Conversely, zero weighted loss forces all singleton residuals to vanish and therefore requires $q_0 \le r$.

If $q_0 > r$, select $q_0 \le k$ independent v_i and form the rational matrix

$$T_0 = D^{-1}\Big(\sum_{i \text{ selected}} v_i v_i^\top\Big)D^{-1}.$$

Rational lower bounds $\nu_J, \nu_T > 0$ for $\lambda_{\min}(J)$ and $\lambda^+_{\min}(T_0)$ give

$$f_r^* \ge \mu := \nu_J\nu_T/k > 0.$$

Indeed the selected-column Gram is $J^{1/2}T_0J^{1/2}$, whose smallest positive eigenvalue is at least $\nu_J\nu_T$. Every rank-r comparison leaves spectral residual at least this large, and its largest squared column residual is at least the spectral residual divided by $q_0 \le k$. Singleton weighting only increases this lower bound. The characteristic-coefficient argument bounds the encoding length of μ polynomially. Use the additive construction with $\varepsilon = \min\{1, \gamma\mu\}$ to obtain (317).

The empty-support and $r = 0$ endpoints use the empty correction; $r \ge k$ is covered by exact zero detection. All rank decisions and all SDP inputs are rational. Each stage has polynomial bit cost, and the number of matrix-function refinement stages is fixed. The support-quantization bounds charge both numerator and denominator of each stored entry. Retaining the original base and exact rational reference adds exactly the entries stated in the corollary. There is no enumerated subset bank in either the computation or the persistent state. □

K.3.8 Transferring the implicit approximation to deletion loss

The implicit SDP optimizes a frozen-curvature response objective. Building on finite-sample analyses of linear approximations to reweighted estimators (Giordano et al., 2019), we bound its difference from actual deletion loss, including noncommuting covariance corrections. The guarantee contains an explicit remainder: small curvature variation alone is insufficient for a relative comparison of the two objectives.

Write

$$R_+ = \Sigma + \lambda I, \qquad \theta = w_S, \qquad g_i = (x_i x_i^\top + \lambda I)\theta - x_i y_i.$$

The normal equations give $\sum_i g_i = 0$. For a request U, put

$$\alpha_U = |U|/n, \quad \beta_U = 1 - \alpha_U, \quad g_U = \frac{1}{n}\sum_{i\in U} g_i, \quad D_U = C_U + \alpha_U\lambda I, \quad R_U = R_+ - D_U.$$

Then $\Delta_U = R_U^{-1}g_U$. For a fixed feasible $E = G - N \succeq 0$, let

$$H = R_+^{-1/2}ER_+^{-1/2}, \qquad Q = H(I+H)^{-1}, \qquad h_U = R_+^{-1/2}g_U.$$

The actual and frozen point losses are, respectively,

$$\mathcal{L}_U(N) = \frac{1}{\beta_U}\Delta_U^\top W_{R_U}(E)\Delta_U, \qquad \mathcal{F}_U(N) = \frac{\|Qh_U\|^2}{\beta_U}.$$

Thus the columns $a_i = R_+^{-1/2} g_i/n$ in Corollary 75 sum to zero, and its weighted objective is $\max_{|U|\le K} \mathcal{F}_U(N)$.

Lemma 76 (A sharp uniform freezing remainder)**.** *Suppose*

$$0 \preceq R_+^{-1/2} D_U R_+^{-1/2} \preceq \eta_U I, \qquad 0 \le \eta_U < 1.$$

For every $E \succeq 0$, including singular E, put $q_E = \|Q\|_{\mathrm{op}}$. Then

$$\begin{aligned} \left|\sqrt{\mathcal{L}_U(N)} - \sqrt{\mathcal{F}_U(N)}\right| &\le \big((1-\eta_U)^{-3/2} - 1\big) q_E \frac{\|h_U\|}{\sqrt{\beta_U}}, \\ |\mathcal{L}_U(N) - \mathcal{F}_U(N)| &\le \big((1-\eta_U)^{-3} - 1\big) q_E^2 \frac{\|h_U\|^2}{\beta_U}. \end{aligned} \tag{318}$$

The two coefficients are sharp as uniform matrix bounds with only the displayed curvature and norm information.

Proof. Set $K_U = R_+^{-1/2} D_U R_+^{-1/2}$, $M = I - K_U$, $m = 1 - \eta_U$, and $B_0 = I - Q = (I+H)^{-1}$. The difference between the retained optimizer and the centered decoder gives

$$\beta_U \mathcal{L}_U(N) = \left\| M^{1/2}\left[M^{-1} - (M+H)^{-1}\right] h_U \right\|^2.$$

Since $\|K_U B_0\|_{\mathrm{op}} \le \eta_U < 1$, write $J_U = (I - K_U B_0)^{-1}$. Direct multiplication, preserving the order of the factors, gives

$$(M+H)^{-1} = B_0 J_U, \qquad M^{-1} - (M+H)^{-1} = M^{-1} Q J_U.$$

Hence the actual amplitude is $Y_U h_U$ with $Y_U = M^{-1/2} Q J_U$. We have

$$\|J_U\|_{\mathrm{op}} \le m^{-1}, \quad \|J_U - I\|_{\mathrm{op}} \le \eta_U/m, \quad \|M^{-1/2} - I\|_{\mathrm{op}} \le m^{-1/2} - 1.$$

Using

$$Y_U - Q = (M^{-1/2} - I)Q + M^{-1/2} Q (J_U - I)$$

therefore gives

$$\|Y_U - Q\|_{\mathrm{op}} \le q_E\big(m^{-1/2} - 1 + \eta_U m^{-3/2}\big) = q_E(m^{-3/2} - 1), \qquad \|Y_U\|_{\mathrm{op}} \le q_E m^{-3/2}.$$

The reverse triangle inequality proves the first bound in (318); multiplying it by the sum of the two amplitudes proves the second. No step interchanges K_U and Q.

For sharpness take scalars $K_U = \eta_U$, $H = t > 0$, and $h_U \neq 0$. Then

$$\frac{Y_U}{Q} = \frac{1+t}{\sqrt{m}(m+t)} \longrightarrow m^{-3/2} \qquad (t \downarrow 0).$$

This approaches equality in both coefficients. The sharpness claim concerns the matrix perturbation bound; it does not assert a ridge lower bound for every choice of its parameters. □

Computing a common remainder from the source. For $1 \le K < n$, let $\rho = K/n$ and $\beta_K = 1 - \rho$. Suppose $(1-\eta)R_+ \preceq R_U \preceq R_+$ for every $|U| \le K$, with $0 \le \eta < 1$. Such an η can be supplied by (274) or Theorem 56. Define

$$V_g = \frac{1}{n}\sum_{i=1}^n R_+^{-1/2} g_i g_i^\top R_+^{-1/2}, \qquad \tau_K = \text{sum of the } K \text{ largest values of } \|R_+^{-1/2} g_i/n\|.$$

A source-only bound is

$$B_K^2 = \min\left\{ \frac{\tau_K^2}{\beta_K},\ \rho\,\mathrm{tr}(V_g),\ \frac{\eta F_S(\theta)}{\beta_K} \right\}, \qquad \frac{\|h_U\|^2}{\beta_U} \le B_K^2 \quad (|U| \le K). \tag{319}$$

The first bound is the triangle inequality. For the second, put $v_i = R_+^{-1/2} g_i$. Since $\sum_i v_i = 0$,

$$h_U = \frac{1}{n} \sum_i (\mathbf{1}_{\{i \in U\}} - \alpha_U) v_i, \qquad h_U h_U^\top \preceq \alpha_U \beta_U V_g$$

by Cauchy–Schwarz. For the third, write $e_i = x_i^\top \theta - y_i$. Matrix Cauchy–Schwarz applied to $g_i = [x_i, \sqrt{\lambda} I](e_i, \sqrt{\lambda}\theta)^\top$ gives

$$g_U g_U^\top \preceq \left[\frac{1}{n} \sum_{i \in U} (e_i^2 + \lambda \|\theta\|^2) \right] D_U.$$

The scalar in brackets is at most $F_S(\theta)$, proving the last bound. In particular, $\mathrm{tr}(V_g) = n^{-1} \sum_i g_i^\top R_+^{-1} g_i$ is rational for rational input. All three bounds can be evaluated to certified rational upper precision without generating any request.

Theorem 77 (Implicit encoding with an actual-loss remainder)**.** *Under the rational-input and covariance-base assumptions of Corollary 75, use $R = R_+$ and assume the preceding curvature bound with $\eta < 1$. Let C_K be the weighted approximation constant in that corollary, and fix rational $0 < \gamma \le 1$. Define*

$$\begin{aligned} Q_0 &= H_0 (I + H_0)^{-1}, \qquad H_0 = R_+^{-1/2} G R_+^{-1/2}, \\ \delta_K &= \big((1-\eta)^{-3/2} - 1 \big) \|Q_0\|_{\mathrm{op}} B_K, \\ A_K &= \sqrt{C_K + \gamma}, \\ \mathrm{OPT}_{r,K} &= \min_{\substack{0 \preceq N \preceq G \\ \mathrm{rank}(N) \le r}} \max_{|U| \le K} \mathcal{L}_U(N). \end{aligned}$$

The corollary's deterministic polynomial-bit encoder, which does not enumerate deletion subsets, outputs a feasible weighted rational correction satisfying

$$\sqrt{\max_{|U| \le K} \mathcal{L}_U(N_{\mathrm{out}})} \le A_K \sqrt{\mathrm{OPT}_{r,K}} + (A_K + 1) \delta_K. \tag{320}$$

It has the same correction-rank and persistent-state bounds as that corollary, and the decoder receives only the state and deleted records. In particular, the reference matrix and the n-column SDP are encoding objects.

Proof. For every feasible correction $Q \preceq Q_0$. Maximizing Lemma 76 over the finite family of all legal requests therefore shows

$$\left| \sqrt{\max_U \mathcal{L}_U(N)} - \sqrt{\max_U \mathcal{F}_U(N)} \right| \le \delta_K.$$

The actual minimum exists by compactness of the feasible rank-constrained set and continuity of the finite maximum. Let N_* attain it and let $\mathrm{OPT}_{r,K}^{\mathrm{fr}} = \min_N \max_U \mathcal{F}_U(N)$, where the minimum is over the same feasible rank-r corrections. Then

$$\sqrt{\mathrm{OPT}_{r,K}^{\mathrm{fr}}} \le \sqrt{\max_U \mathcal{F}_U(N_*)} \le \sqrt{\mathrm{OPT}_{r,K}} + \delta_K.$$

The frozen approximation of Corollary 75 gives

$$\sqrt{\max_U \mathcal{L}_U(N_{\mathrm{out}})} \le \sqrt{C_K + \gamma} \sqrt{\mathrm{OPT}_{r,K}^{\mathrm{fr}}} + \delta_K,$$

which proves the claim. Its arithmetic and storage statements are unchanged by this comparison; η and B_K certify the guarantee and are not required in the stored state. □

For a fixed deletion fraction bounded away from zero and one, C_K is bounded independently of the source size and dimension. The remainder is explicit and tends to zero with the curvature variation when the other quantities are bounded. Equation (320) is not a pure multiplicative approximation to actual loss when $\mathrm{OPT}_{r,K}$ is small. It also does not implement a finely accurate worst-request oracle or contradict Theorem 53.

Proposition 78 (Frozen exactness need not imply deletion exactness)**.** *There are bounded rational ridge sources in dimension two, a rational covariance base, and a fixed feasible residual for which every frozen deletion loss with $|U| \le 2$ is zero but some actual deletion loss is positive. The common curvature variation tends to zero as the source size tends to infinity.*

Proof. Let $n \ge 8$ and $\lambda = 1/4$. Take the six records

$$((1/2,1/2),0), \quad ((0,1/2),1), \quad ((0,1/2),-1),$$
$$((1/2,-1/2),0), \quad ((1/2,0),0), \quad ((1/2,0),0),$$

and $n-6$ zero records. Then $\Sigma = I/n$, $b = \theta = 0$, and $R_+ = (1/4 + 1/n)I$. Set

$$A_0 = \Sigma + \tfrac{1}{4} e_1 e_1^\top, \qquad G = E = \tfrac{1}{4} e_1 e_1^\top, \qquad N = 0.$$

All g_i lie in $\mathrm{span}(e_2)$, whereas Q is supported on $\mathrm{span}(e_1)$. Hence every frozen loss is zero, including requests of size greater than two. Since $\|x_i\|^2 \le 1/2$, all $|U| \le 2$ satisfy

$$0 \preceq R_+^{-1/2} D_U R_+^{-1/2} \preceq \frac{6}{n+4} I, \qquad \frac{6}{n+4} \le \frac{1}{2}.$$

Delete the first two records. Writing $\beta = 1 - 2/n$ and $d_n = n^2 + n - 1$ gives

$$R_U = \frac{1}{4n} \begin{bmatrix} n+1 & -1 \\ -1 & n \end{bmatrix}, \qquad g_U = -\frac{e_2}{2n}, \qquad \Delta_U = -\frac{2}{d_n} \begin{bmatrix} 1 \\ n+1 \end{bmatrix}.$$

Thus $E\Delta_U = -e_1/(2d_n) \ne 0$. Substitution in the actual loss yields the strictly positive rational value

$$\mathcal{L}_U(0) = \frac{n^2}{\beta\, d_n (2n^2 + n - 1)^2}.$$

This compares the two objectives for a fixed residual. In particular, it is not by itself a lower bound on the approximation ratio of a rank-augmented encoder. □

Proposition 79 (The frozen optimum need not preserve an exact rank budget)**.** *There are bounded dyadic ridge sources in dimension three, with $\lambda = 1$, the source's exact zero-head spectral floor, and deletion budget one, such that*

$$\mathrm{OPT}_{1,1} = 0, \qquad \mathrm{OPT}_{1,1}^{\mathrm{fr}} > 0.$$

Their common relative curvature variation tends to zero. Thus no finite purely multiplicative upper bound on the optimal frozen loss in terms of the optimal actual loss follows from curvature stability alone.

Proof. Take $n = 8q$ with integer $q \ge 1$. Choose the dyadic power of two

$$\epsilon = 2^{-\lceil \log_2(19n-16) \rceil}, \qquad t = \frac{(19n-16)\epsilon}{16}.$$

Then $1/[2(19n-16)] \le \epsilon \le 1/(19n-16)$. The source consists of q copies of the following eight-record pattern. Its features are

$$e_1, \quad e_2, \quad v_+ = (e_1+e_2)/2, \quad v_- = (e_1-e_2)/2, \quad e_3, e_3, e_3, e_3.$$

The respective labels are $9t - \epsilon/2$, 0, $(t+\epsilon)/2$, $(t+\epsilon)/2$, and four zeros. Here $1/32 \le t \le 1/16$, so all records are bounded, and every coordinate has $O(\log n)$ bits. Direct calculation gives

$$\Sigma = \mathrm{diag}(3/16, 3/16, 1/2), \qquad R_+ = \mathrm{diag}(19/16, 19/16, 3/2), \qquad b = (19t/16)e_1.$$

Consequently $\theta = te_1$. The exact zero-head spectral floor is $A_0 = I/2$, with $G = \mathrm{diag}(5/16, 5/16, 0)$.

Within each pattern, the unnormalized gradient vectors are

$$g_{e_1} = (-7t + \epsilon/2)e_1,$$
$$g_{v_\pm} = (t - \epsilon/4)e_1 \mp (\epsilon/4)e_2,$$
$$g_{e_2} = g_{e_3} = te_1.$$

Their sum is zero. Put $h = 19/16 - 1/n$, so $nh\epsilon = t$. For every singleton request other than an e_1 record,

$$nR_{\{i\}}\,\epsilon e_1 = te_1 - \epsilon x_i x_{i1} = g_i.$$

Thus $\Delta_{\{i\}} = \epsilon e_1$ for these requests. Deleting an e_1 record gives

$$\Delta_{\{i\}} = \frac{-7t + \epsilon/2}{n(h - 1/n)} e_1.$$

The empty request has zero response. Every actual response is therefore in $\mathrm{span}(e_1)$, and the rank-one correction $N_* = (5/16)e_1 e_1^\top$ makes every actual loss zero.

The frozen singleton responses $R_+^{-1} g_i/n$, on the other hand, span the first two coordinate directions: the v_+ and v_- columns have nonzero opposite second components, and the e_2 column has a nonzero first component. For any feasible correction, zero frozen loss on a vector z is equivalent to $(G - N)z = 0$. Hence zero on all these columns would require N to agree with G on its two-dimensional positive support, forcing $\mathrm{rank}(N) \geq 2$. Compactness and continuity then imply $\mathrm{OPT}_{1,1}^{\mathrm{fr}} > 0$.

Finally, all singleton curvature changes satisfy

$$\left\| R_+^{-1/2} \frac{x_i x_i^\top + I}{n} R_+^{-1/2} \right\|_{\mathrm{op}} \leq \frac{32}{19n} \longrightarrow 0.$$

The maximum is attained for $x_i = e_1$ or e_2; the mixed and e_3 features give $24/(19n)$ and $4/(3n)$, respectively. □

This last obstruction compares optima at the same rank. The implicit algorithm is allowed additional correction columns; for the displayed source, two columns can recover the full gap exactly. Accordingly, neither zero-level obstruction establishes a failure ratio for that algorithm. They explain why its actual guarantee explicitly retains the curvature remainder.

The implicit SDP, its weighted rational format, and its source remainder have not been evaluated by the performance grid. A polynomial-time construction and uniform utility bounds do not establish a practical accuracy or runtime advantage.

A relative certificate for a specified correction. Curvature can also be compared in a metric adapted to a fixed residual. This gives a purely relative bound when curvature does not move a direction out of the residual's kernel.

Lemma 80 (Kernel-preserving curvature transfer)**.** *Fix a feasible residual $E = G - N$ and its Q as above. Put*

$$K_U = R_+^{-1/2} D_U R_+^{-1/2}, \qquad B_Q = I - Q, \qquad A_{U,Q} = Q K_U B_Q Q^\dagger.$$

Suppose $0 \preceq K_U \preceq \eta I$ with $\eta < 1$, $K_U \ker Q \subseteq \ker Q$, and $\nu_{U,Q} = \|A_{U,Q}\|_{\mathrm{op}} < 1$. Then

$$\frac{\mathcal{F}_U(N)}{(1 + \nu_{U,Q})^2} \leq \mathcal{L}_U(N) \leq \frac{\mathcal{F}_U(N)}{(1 - \eta)(1 - \nu_{U,Q})^2}.$$

Under a common bound $K_U \preceq \eta I$ with $\eta < 1$, the remaining hypotheses hold for all $|U| \leq K$ if each $K_i = R_+^{-1/2}(x_i x_i^\top + \lambda I) R_+^{-1/2}/n$ preserves $\ker Q$ and that the sum of the K largest numbers

$$\|Q K_i (I - Q) Q^\dagger\|_{\mathrm{op}}$$

is less than one. This sum upper-bounds every $\nu_{U,Q}$. These are source checks for the specified correction and require no request enumeration. The common curvature bound may be supplied by (274) *or Theorem 56.*

Proof. Let $P = Q^\dagger Q$ be the projection onto $\mathrm{range}\, Q$. The invariance assumption and $B_Q(I - P) = I - P$ imply $Q K_U B_Q (I - P) = 0$, and hence $Q K_U B_Q = A_{U,Q} Q$. For $J_U = (I - K_U B_Q)^{-1}$ this gives

$$(I - A_{U,Q})Q = Q(I - K_U B_Q), \qquad Q J_U = (I - A_{U,Q})^{-1} Q.$$

The actual amplitude from Lemma 76 is $(I - K_U)^{-1/2} Q J_U h_U$. The inverse factor on the right has operator norm at most $(1-\nu_{U,Q})^{-1}$ and smallest singular value at least $(1+\nu_{U,Q})^{-1}$. The left factor has singular values between one and $(1-\eta)^{-1/2}$. Squaring the two norm bounds and dividing by β_U proves the result. Invariance is preserved by summing the K_i; the triangle inequality proves the top-K sufficient bound. The case $Q = 0$ uses $Q^\dagger = 0$ and both losses are zero. □

For rational E and source data, the kernel test itself is exact: if the columns of Z are a rational basis of $\ker E$, it is equivalent to

$$E R_+^{-1} \frac{x_i x_i^\top + \lambda I}{n} Z = 0 \qquad (1 \le i \le n).$$

Indeed $\ker Q = R_+^{1/2} \ker E$. Certified norm upper bounds can replace the norms in the strict sufficient inequality. Checking this lemma only for an encoder's output does not compare that output to the unrestricted rank-r actual optimum: the actual optimal comparator would need its own relative certificate. Common diagonal curvature does not justify restricting that comparator to diagonal corrections; the non-diagonal correction in Theorem 51 illustrates this distinction. In that family with $p = r + 1 > 4$, every diagonal rank-r correction leaves some active coordinate uncorrected and therefore has loss at least L_{bad} from (264). The non-diagonal feasible correction has loss $L_{\mathrm{bal}} \le 4L_{\mathrm{bad}}/p < L_{\mathrm{bad}}$ by (266). Thus even diagonal source curvatures can have a strictly better non-diagonal correction.

K.3.9 An exact implicit actual-response span

We compute the span of actual responses without listing the request bank, resolving the zero-loss branch. Positive-loss worst-request optimization is a separate problem. Throughout this subsection $\lambda > 0$, $0 \le K < n$, and

$$R_U = \Sigma + \lambda I - \frac{1}{n} \sum_{i \in U} (x_i x_i^\top + \lambda I), \qquad \Delta_U = R_U^{-1} \frac{1}{n} \sum_{i \in U} g_i, \qquad g_i = x_i (x_i^\top \theta - y_i) + \lambda \theta.$$

The source normal equations determine $\theta = w_S$ and give $\sum_i g_i = 0$. Also write $R_+ = \Sigma + \lambda I$ and $D_U = n^{-1} \sum_{i \in U} (x_i x_i^\top + \lambda I)$, so $R_U = R_+ - D_U$. Write $\mathcal{S}_K = \mathrm{span}\{\Delta_U : |U| \le K\}$.

Theorem 81 (Finite deletion order and an exact span algorithm)**.** *Set $t = \min\{K, d\}$. Then*

$$\mathcal{S}_K = \mathrm{span}\{\Delta_U : |U| \le t\}.$$

For rational source records and rational λ, this space and a basis consisting of at most d actual responses can be computed deterministically in $2^{O(d)}\,\mathrm{poly}(n, L)$ bit operations, where L is the total input bit length. Each basis response has a witness request of size at most t. The algorithm does not enumerate all requests of that size. Its output has polynomial bit length in the source input. For $K = 0$ it returns the zero space.

Proof. First introduce formal variables $z_1, \dots, z_n$ and put

$$R(z) = \Sigma + \lambda I - \frac{1}{n} \sum_i z_i (x_i x_i^\top + \lambda I), \qquad p(z) = \mathrm{adj}(R(z)) \frac{1}{n} \sum_i z_i g_i.$$

Every coordinate of p has degree at most d. Multilinearizing by $z_i^a = z_i$ for $a \ge 1$ preserves its values on Boolean points and cannot increase degree. Write the resulting polynomial as $p(z) = \sum_{|T| \le d} c_T \prod_{i \in T} z_i$. Its coefficients obey

$$c_T = \sum_{V \subseteq T} (-1)^{|T| - |V|} p(\mathbf{1}_V).$$

Consequently $p(\mathbf{1}_U)$ is a linear combination of $p(\mathbf{1}_V)$ with $V \subseteq U$ and $|V| \le d$. All these sets are legal when U is legal, and

$$p(\mathbf{1}_V) = \det(R_V) \Delta_V, \qquad R_V \succeq (1 - |V|/n) \lambda I \succ 0.$$

The nonzero scalar determinants preserve spans. This proves the deletion-order assertion, but enumerating its generators alone would only give an $n^{O(d)}$ algorithm.

For the stronger algorithm, put $e_i = x_i^\top \theta - y_i$ and define the d by $d+1$ rank-one matrices

$$Z_i = \frac{1}{n} x_i [-x_i^\top, e_i], \qquad Z_U = \sum_{i \in U} Z_i.$$

For each such matrix Z, let $\Phi(Z)$ contain all square minors $\det Z[I, J]$, with $|I| = |J|$, including the empty minor 1. Its dimension is

$$D = \sum_{j=0}^{d} \binom{d}{j} \binom{d+1}{j} = \binom{2d+1}{d} = 2^{O(d)}.$$

For a rank-one update $uv^\top$, the determinant expansion gives

$$\det(Z + uv^\top)[I, J] = \det Z[I, J] + \sum_{a=1}^{j} \sum_{b=1}^{j} (-1)^{a+b} u_{i_a} v_{j_b} \det Z[I \setminus \{i_a\}, J \setminus \{j_b\}].$$

Here row and column index sets are in increasing order. Thus adding Z_i acts on the entire vector of minors by a known linear map T_i, evaluable in $O(d^2 D)$ arithmetic operations.

After scanning the first m records maintain, for each $0 \le k \le \min\{m, t\}$, a basis of

$$V_{m,k} = \operatorname{span}\{\Phi(Z_U) : U \subseteq \{1, \dots, m\},\ |U| = k\}.$$

Start with $V_{0,0} = \operatorname{span}\{\Phi(0)\}$ and all other spaces zero. The exact recurrence is

$$V_{m,0} = V_{m-1,0}, \qquad V_{m,k} = V_{m-1,k} + T_m V_{m-1,k-1} \quad (k \ge 1).$$

Use previous-layer spaces on its right-hand side. Retain an independent subset of the at most $2D$ candidate vectors, retaining each candidate's actual witness request as well. Induction proves both the recurrence and the witness property: linearity makes discarded dependent vectors unnecessary for every later update. A selected update appends the new record to its previous witness. Each space has dimension at most D.

For a fixed cardinality k, set

$$B_k = [\Sigma + (1 - k/n)\lambda I,\ (k/n)\lambda \theta].$$

Then $[R_U, g_U] = B_k + Z_U$ whenever $|U| = k$, where $g_U = n^{-1} \sum_{i \in U} g_i$. Each Cramer numerator, obtained by replacing one column of R_U by g_U, is a linear functional of $\Phi(Z_U)$. Indeed, expansion of a maximal minor of $B_k + Z_U$ by the rows and columns selected from Z_U expresses it as a sum of minors of Z_U multiplied by signed complementary minors of B_k. The column permutation used in a replacement only changes the corresponding sign. Hence a basis of $V_{n,k}$ maps to a spanning set of the Cramer numerators at cardinality k. Compute actual responses of its witness requests by rational ridge solves. Their spans agree with the numerator spans because their individual determinants are positive. Combining all $k \le t$ and extracting independent response columns gives the required basis.

There are $O(nt)$ spaces. Classical elimination on at most $2D$ vectors in $\mathbb{Q}^D$ costs $O(D^3)$ arithmetic operations per space; generating candidates costs $O(d^2 D^2)$. The final stage uses at most $(t+1)D$ ridge solves and polynomial-dimensional elimination. These costs are $2^{O(d)} \operatorname{poly}(n)$.

For completeness, retain original candidate columns, rather than normalized elimination columns, as the stored basis. Each is the minor vector of an actual sum of source rank-one matrices, so its coordinates have polynomial bit length in L, n, d. Exact source moments and θ also have polynomial bit length. Fraction-free elimination, or determinant bounds for rational elimination, bounds intermediate bit lengths by $O(D)$ times a polynomial in the input length. The number and bit cost of all operations therefore remain $2^{O(d)} \operatorname{poly}(n, L)$. Final response columns solve d-dimensional rational systems on actual requests, and hence have polynomial bit length independent of the lifted dimension. □

Corollary 82 (Intrinsic source rank)**.** *Let $p = \dim \operatorname{span}\{x_1, \dots, x_n\}$. In Theorem 81, the deletion-order threshold, the maximum number of returned basis columns, and the parameter in the exponential running time can all be replaced by p. Thus the running time is $2^{O(p)} \operatorname{poly}(n, d, L)$, and the returned witness requests have size at most $\min\{K, p\}$.*

Proof. When $p = 0$ all features, the source optimizer, and all responses are zero. Otherwise obtain a rational full-column-rank basis $C \in \mathbb{Q}^{d\times p}$ of the source feature span by exact elimination. This span and its orthogonal complement are invariant under every retained curvature. Both θ and g_U lie in the span, so $\Delta_U = C\zeta_U$, where

$$(C^\top R_U C)\zeta_U = C^\top g_U.$$

Writing $v_i = C^\top x_i$ and $H = C^\top C \succ 0$, the reduced curvature is

$$C^\top \Sigma C + (1 - |U|/n)\lambda H - \frac{1}{n}\sum_{i\in U} v_i v_i^\top.$$

The reduced right-hand side is $n^{-1}\sum_{i\in U}(v_i e_i + \lambda C^\top\theta)$. Thus the proof applies in dimension p, with a positive definite rational regularization metric H, and the fixed-cardinality augmented baseline has last column $(k/n)\lambda C^\top\theta$. Its updates remain rank one and its lift has dimension $\binom{2p+1}{p}$. Mapping the result back by C preserves spans and adds polynomial cost. No rational orthonormal basis or exact square root is required. □

Corollary 83 (Exact zero optimum without a listed bank)**.** *Let $G \succeq 0$ be a rational covariance gap for the fixed original base, and let $r \geq 0$ be an integer. Use the actual loss and $\mathrm{OPT}_{r,K}$ from Theorem 77. For its comparison class $0 \preceq N \preceq G$, $\operatorname{rank} N \leq r$, define*

$$q_K = \dim G\mathcal{S}_K.$$

Then $\mathrm{OPT}_{r,K} = 0$ if and only if $q_K \leq r$. This condition and, when it holds, a feasible rank-q_K exact correction are computable within the complexity of Theorem 81. The correction admits a polynomial-bit rational weighted factor with $dq_K + q_K$ entries.

Proof. The exact deletion-error identity in Proposition 49 gives zero actual loss precisely when $(G - N)\Delta_U = 0$. Necessity follows because N must map the response space onto $G\mathcal{S}_K$, so its rank is at least q_K. From the computed basis choose response columns Y whose images under G are independent and span $G\mathcal{S}_K$. If $q_K = 0$, take $N_0 = 0$. Otherwise put

$$C = Y^\top G Y \succ 0, \qquad N_0 = GYC^{-1}Y^\top G.$$

The projection representation $N_0 = G^{1/2}P_{\mathrm{range}(G^{1/2}Y)}G^{1/2}$ proves $0 \preceq N_0 \preceq G$ and $\operatorname{rank} N_0 = q_K$. It annihilates the selected responses in the residual. Every other response differs from their linear combination by a vector in $\ker G$, which also lies in $\ker N_0$. Thus it annihilates all responses. Rational $LDL^\top$ factorization $C = L\operatorname{diag}(p_j)L^\top$ yields

$$Z = GYL^{-\top}, \qquad \omega_j = p_j^{-1} > 0, \qquad N_0 = Z\operatorname{diag}(\omega)Z^\top.$$

All matrices in this last step have order at most d and polynomial input bit length. The original base, θ, and n are retained as in the existing weighted format; lifted features and witnesses are encoding-time objects. □

Proposition 84 (The deletion-order bound is tight for ridge sources)**.** *For every $d \geq 2$ and $n \geq 2d$, there is a bounded dyadic ridge source in dimension d, with $\lambda = 1$ and its exact zero-head spectral floor, for which*

$$G\mathcal{S}_{d-1} = \{0\}, \qquad G\mathcal{S}_d \neq \{0\}.$$

In particular $\mathcal{S}_{d-1} \neq \mathcal{S}_d$, and even zero-optimum detection cannot uniformly replace d by $d - 1$. The example has $q_{d-1} = 0$ and $q_d = 1$. With d fixed, its relative curvature variation for requests of size at most d tends to zero as n tends to infinity.

Proof. Put $a = 1/2$. For every $1 \leq j < d$ include the two zero-label records with features $a(e_j + e_{j+1})$ and $a(e_j - e_{j+1})$. Add two records with feature ae_1 and labels a and $-a$, respectively, and pad with $n - 2d$ zero records. All feature norms are at most one and all labels have magnitude at most one. The labels cancel in the original cross moment, so $\theta = 0$. Directly summing the covariance gives

$$\Sigma = \frac{4a^2}{n}I - \frac{2a^2}{n}e_d e_d^\top, \qquad A_0 = \frac{4a^2}{n}I, \qquad G = \frac{2a^2}{n}e_d e_d^\top.$$

Thus A_0 is the source's exact zero-head spectral floor. Only the two nonzero-label records contribute to g_U, and their contributions are opposite multiples of e_1. A nonzero response therefore requires deletion of exactly one of these two records. The only possible off-diagonal entries of R_U connect consecutive coordinates. The entry at edge j is nonzero only when exactly one of its two zero-label records is deleted. To connect coordinate one to coordinate d requires a deletion in each of the $d-1$ edge groups, in addition to the nonzero-label record. For $|U| < d$, a missing edge makes R_U block diagonal across a cut separating one from d. Consequently $e_d^\top \Delta_U = 0$.

Now delete the record with label a and all $d-1$ plus-edge records. For this size-d request, $g_U = -a^2 e_1/n$ and R_U is positive definite tridiagonal, with each adjacent off-diagonal entry equal to $-a^2/n$. Expansion of the corner cofactor yields

$$e_d^\top \Delta_U = -\frac{(a^2/n)^d}{\det R_U} \neq 0.$$

Hence the claimed response and projected-response spaces differ. Since G has rank one, the two projected dimensions are exactly zero and one. The rank-zero actual optimum is zero below deletion order d and positive at order d, while a rank-one correction is exact at every order. Finally,

$$\left\| R_+^{-1/2} D_U R_+^{-1/2} \right\|_{\mathrm{op}} \leq \frac{3|U|}{2n},$$

because $R_+ \succeq I$ and $\|x_i\|^2 \leq 1/2$. □

For $d = 1$, two identical nonzero features with opposite nonzero labels give $\mathcal{S}_0 = \{0\} \neq \mathcal{S}_1$, proving tightness of the response-order bound in that dimension as well. Its zero-head spectral floor has zero gap, so the projected-gap example is stated only for $d \geq 2$.

The span algorithm uses the standard principle that a linear representative basis preserves all future linear transitions; determinantal and exterior-algebra compression also underlies representative-family methods (Fomin et al., 2016). It does not assert that a small response basis preserves the maximum positive loss: both the response magnitude and the retained curvature enter that objective. The exact consequence established here is zero-loss detection and representation for the original covariance base, without a curvature remainder or a larger rank budget. If $q_K > r$, compactness of the feasible correction class and continuity of the finite maximum imply a strictly positive optimum, but no multiplicative approximation for that branch is claimed by the span computation.

A quantitative certificate for the positive branch. The selected actual responses also give a positive lower bound on the optimum. Unlike a worst-request approximation, this bound only uses their independent gap images. It can be small when the chosen responses have small magnitude or nearly dependent images.

Proposition 85 (A rational lower certificate from actual witnesses)**.** *Let $A_0 \succeq \Sigma$ be the rational original covariance base, $G = A_0 - \Sigma$, and $q = q_K > r \geq 0$, with integer r. Select q actual response columns $Y = [\Delta_{U_1}, \ldots, \Delta_{U_q}]$ whose images under G are independent, using Theorem 81. Define*

$$T = (GY)^\top GY \succ 0, \qquad a_0 = \lambda + \operatorname{tr}(A_0), \qquad \mu_{r,K} = \frac{\lambda(q-r)}{q a_0^2 \operatorname{tr}(T^{-1})} > 0.$$

Then $\mu_{r,K} \leq \mathrm{OPT}_{r,K}$. The certificate requires no curvature-stability assumption. It is a rational number of polynomial bit length, computable together with its at most $q \leq p$ legal witnesses in $2^{O(p)}\,\mathrm{poly}(n, d, L)$ bit operations, where L includes the original base's representation and p is the source feature rank from Corollary 82. No witness is added to the persistent decoder state.

Proof. For any feasible correction let $E = G - N \succeq 0$, $\beta_U = 1 - |U|/n$, and $H_U = R_U + E$. Then $R_U \succeq \beta_U \lambda I$ and

$$0 \prec H_U = A_0 - N - C_U + \beta_U \lambda I \preceq A_0 + \lambda I \preceq a_0 I, \qquad C_U = \frac{1}{n} \sum_{i \in U} x_i x_i^\top.$$

The exact loss identity consequently gives

$$\mathcal{L}_U(N) = \frac{1}{\beta_U}(E\Delta_U)^\top H_U^{-1} R_U H_U^{-1}(E\Delta_U)$$
$$\geq \lambda \|H_U^{-1} E\Delta_U\|^2 \geq \frac{\lambda}{a_0^2}\|E\Delta_U\|^2.$$

In particular the retained-fraction factor cancels; replacing it by a sample-count factor would change the normalization. The matrix NY has rank at most r. By the rank-approximation bound applied to GY, its squared Frobenius residual obeys

$$\|(G-N)Y\|_F^2 \geq \sum_{j=r+1}^{q} \lambda_j(T) \geq \frac{q-r}{\operatorname{tr}(T^{-1})},$$

where eigenvalues are in nonincreasing order. The last inequality uses $1/\operatorname{tr}(T^{-1}) \leq \lambda_{\min}(T)$. The maximum loss is at least the average over these q legal witnesses. Combining the last two displays and minimizing over the original rank-r class proves the claim. The witness responses have polynomial bit length by Theorem 81; all further inversions and products have order at most d and are rational. Standard rational elimination bounds thus give polynomial output bit length, including $\log^+(1/\mu_{r,K})$. □

Corollary 86 (A checkable relative-loss regime)**.** *Under Theorem 77, first run the exact span algorithm. If $q_K \leq r$, return its exact correction. Otherwise run the implicit encoder and compute the preceding $\mu_{r,K}$. In this positive branch, for any certified $\bar{\delta}_K \geq \delta_K$, the output satisfies*

$$\max_{|U|\leq K} \mathcal{L}_U(N_{\text{out}}) \leq \left[A_K + (A_K+1)\frac{\bar{\delta}_K}{\sqrt{\mu_{r,K}}}\right]^2 \operatorname{OPT}_{r,K}.$$

Thus a certificate $\bar{\delta}_K^2 \leq s^2\mu_{r,K}$ gives the factor $[A_K+(A_K+1)s]^2$ for any specified rational $s \geq 0$. The original rank augmentation and weighted storage of the implicit encoder remain unchanged in the positive branch.

Proof. In the positive branch, $\mu_{r,K} \leq \operatorname{OPT}_{r,K}$, so the additive term in Theorem 77 is at most $(A_K+1)\bar{\delta}_K\sqrt{\operatorname{OPT}_{r,K}/\mu_{r,K}}$. Squaring gives the claim. A rational upper enclosure for the remainder is available from the source computations in that theorem, and the displayed regime test is a rational comparison. Neither the certificate nor its witnesses need be stored after encoding. □

This is an instance-dependent certificate, not an unconditional constant-factor approximation. The ratio $\bar{\delta}_K/\sqrt{\mu_{r,K}}$ need not be bounded, even if η is small. The span pass has exponential dependence on source rank, and this corollary does not implement the complete certified implicit SDP encoder or evaluate its performance.

K.3.10 Polynomial-time positive certificates from singleton responses

Computing the full response span is unnecessary for a positive certificate. Any collection of legal requests gives a lower bound on the worst loss over all requests. In particular, singleton responses can certify the positive branch in polynomial time, even when the source rank is large. Failure of this certificate does not decide the zero branch at larger deletion budgets.

Proposition 87 (Weighted actual-witness certificates)**.** *Let $\mathcal{B} = (U_1, \ldots, U_m)$ be a nonempty bank of legal requests with $|U_j| \leq K < n$. Put $z_j = G\Delta_{U_j}$, $a_0 = \lambda + \operatorname{tr}(A_0)$, and choose scalar probability weights $w_j \geq 0$, $\sum_j w_j = 1$. Define*

$$M_w = \sum_{j=1}^{m} w_j z_j z_j^\top, \qquad \tau_r(M_w) = \sum_{j=r+1}^{d} \lambda_j(M_w),$$

with eigenvalues in nonincreasing order and an empty sum equal to zero. For every integer $r \geq 0$,

$$\operatorname{OPT}_{r,K} \geq \frac{\lambda}{a_0^2}\tau_r(M_w). \tag{321}$$

For rational data and weights, if $s = \operatorname{rank}(M_w) > r$, one can compute a rational ℓ_w satisfying

$$\tfrac{1}{2}\lambda^+_{\min}(M_w) \le \ell_w \le \lambda^+_{\min}(M_w)$$

in polynomial bit complexity, using $O(\log(s+1))$ exact PSD tests. Consequently

$$\underline{\mu}_w = \frac{\lambda(s-r)\ell_w}{a_0^2} > 0 \quad \textit{satisfies} \quad \underline{\mu}_w \le \mathrm{OPT}_{r,K}\,. \tag{322}$$

If $s \le r$, this bank yields zero as a lower bound, which is inconclusive about the full request family.

Proof. Write $Y = [\Delta_{U_j}]$ and $W = \operatorname{diag}(w_j)$. The pointwise inequality proved in Proposition 85 uses only $0 \preceq N \preceq G$, not a rank or spanning premise. For a feasible N it therefore gives

$$\begin{aligned}\max_{|U|\le K} \mathcal{L}_U(N) &\ge \sum_j w_j \mathcal{L}_{U_j}(N) \\ &\ge \frac{\lambda}{a_0^2}\|(G-N)YW^{1/2}\|_F^2 \ge \frac{\lambda}{a_0^2}\tau_r(M_w).\end{aligned}$$

The last step uses $\operatorname{rank}(NYW^{1/2}) \le r$ and the standard best low-rank approximation bound (Eckart & Young, 1936). Minimizing over the same original base and rank-r class proves (321). Only scalar nonnegative weights are used; an arbitrary PSD matrix in place of W need not be an average of legal point losses.

For the rational certificate, write a positive-weight rational LDL factorization $M_w = V\Omega V^\top$ with V full column rank, $\Omega = \operatorname{diag}(\omega_j) \succ 0$, and put

$$J = V^\top V, \quad P = VJ^{-1}V^\top, \quad a = \left(\sum_{j=1}^{s} \frac{(J^{-1})_{jj}}{\omega_j}\right)^{-1}.$$

Here P projects onto $\operatorname{range} M_w$, and the sum in parentheses equals $\operatorname{tr}(M_w^\dagger)$. Therefore

$$a \le \lambda^+_{\min}(M_w) \le sa, \qquad M_w - tP \succeq 0 \iff t \le \lambda^+_{\min}(M_w).$$

Start with $l = a$, $u = sa$. While $u > 2l$, test $M_w - \frac{1}{2}(l+u)P \succeq 0$ and retain the feasible lower half-endpoint or the upper endpoint, respectively. More explicitly, on a feasible test set $l = (l+u)/2$; otherwise set $u = (l+u)/2$. The interval always contains $\lambda^+_{\min}(M_w)$ and its width halves. After $O(\log(s+1))$ tests, return $\ell_w = l$. This proves the stated two-sided bound; the case $s = 1$ needs no test. Rational elimination and the logarithmic number of midpoint operations keep all bit lengths polynomial. Finally, $\tau_r(M_w) \ge (s-r)\lambda^+_{\min}(M_w)$ proves (322). □

The initial a alone gives the inverse-trace certificate. For an independent q-column bank with uniform weights, $\operatorname{tr}(M_w^\dagger) = q\operatorname{tr}(T^{-1})$, so this recovers Proposition 85 without needing the bank to span all actual responses. The bisection can improve that certificate by up to a dimension factor. Its factor two concerns the smallest positive eigenvalue, not the full spectral tail or the actual optimum.

Optional weight optimization. For $0 \le r \le d$, the best weighted spectral-tail certificate is the value of

$$\begin{aligned}\max_{w,t,S} \quad & \operatorname{tr}(M_w) - rt - \operatorname{tr} S \\ \text{subject to} \quad & w \ge 0, \quad \mathbf{1}^\top w = 1, \\ & S \succeq 0, \quad S + tI \succeq M_w.\end{aligned} \tag{323}$$

For fixed w, the minimum of $rt + \operatorname{tr} S$ is the sum of the largest r eigenvalues. This is the classical eigenvalue-sum SDP (Vandenberghe & Boyd, 1996, Equation (7)), applied to the trace complement. Any rational feasible triple supplies the valid rational lower bound $\lambda \max\{0, \operatorname{tr}(M_w) - rt - \operatorname{tr} S\}/a_0^2$; its PSD conditions can be checked exactly. This observation does not claim an implemented certified optimizer for (323).

Corollary 88 (A polynomial singleton route to the positive branch)**.** *For rational source data, $\lambda > 0$, and $1 \le K < n$, all singleton responses can be formed using $O(d^3 + nd^2)$ rational arithmetic operations and polynomial bit complexity. If*

$$s_1 = \operatorname{rank}[G\Delta_{\{1\}}, \dots, G\Delta_{\{n\}}] > r,$$

uniform weights in Proposition 87 give $\underline{\mu} > 0$ in polynomial bit complexity without the exponential lift. Under Theorem 77, its encoder then satisfies

$$\max_{|U| \le K} \mathcal{L}_U(N_{\mathrm{out}}) \le \left[A_K + (A_K + 1) \frac{\bar{\delta}_K}{\sqrt{\underline{\mu}}} \right]^2 \mathrm{OPT}_{r,K} .$$

Thus the certificate test $\bar{\delta}_K^2 \le s_0^2 \underline{\mu}$ gives factor $[A_K + (A_K + 1)s_0]^2$ without any exponential source-rank step. The same result holds with the largest valid certificate from several chosen banks or weights.

Proof. Put $B = \Sigma + (1 - 1/n)\lambda I$, $e_i = x_i^\top \theta - y_i$, and $g_i = x_i e_i + \lambda\theta$. Sherman–Morrison gives

$$v_i = B^{-1} g_i / n, \qquad a_i = B^{-1} x_i,$$

$$\Delta_{\{i\}} = v_i + a_i \frac{x_i^\top v_i}{n - x_i^\top a_i}.$$

The denominator is positive since $B - x_i x_i^\top / n = R_{\{i\}} \succeq (1 - 1/n)\lambda I \succ 0$. After computing θ and B^{-1}, every column costs $O(d^2)$ operations. The rational witness Gram, its support, and the preceding PSD tests have polynomial bit complexity. Uniform weights have rank s_1, so their certificate is positive when $s_1 > r$. Substitution into the existing actual-loss remainder bound proves the result, with the same rank augmentation and weighted storage. All response columns and lower certificates are discarded after encoding. □

For $K > 1$, $s_1 \le r$ is inconclusive; the chain in Proposition 84 has no projected singleton response yet a positive optimum at deletion order d. For $K = 1$ the singleton bank is complete, so the existing exact zero criterion does apply. The prototype labels every zero singleton certificate as inconclusive to avoid applying that special inference at larger budgets.

The default prototype also greedily selects $r+1$ independent columns of $[G\Delta_{\{i\}}]$: each step chooses the largest squared residual norm after orthogonal projection onto the previous columns, using rational projections and fixed index tie-breaking. When $s_1 > r$, this succeeds, and uniform weights on the selected actual requests give another valid certificate. The prototype returns the larger of it and the full-bank certificate. This selection is not claimed to optimize (323) or to approximate the best bank within a universal constant.

Proposition 89 (A weak response direction can spoil a full-basis certificate)**.** *There are bounded dyadic ridge sources with $d = 4$, $\lambda = 1$, the source's exact zero-head spectral floor, $K = 1$, and $r = 1$, for which $q_K = 3$ and the certificate of Proposition 85 satisfies $\delta_K / \sqrt{\mu_{r,K}} \to \infty$ even as the relative curvature variation tends to zero. On the same sources, two actual singleton witnesses give a lower certificate $\mu_{\mathrm{strong}} > 0$ with $\delta_K / \sqrt{\mu_{\mathrm{strong}}} \to 0$.*

Proof. Let $n = 2^m \ge 16$ and let $0 < \epsilon \le 1$ be dyadic. For $j = 1, 2, 3$, take two records with feature $e_j/2$ and labels $+c_j, -c_j$, where $(c_1, c_2, c_3) = (1, 1, \epsilon)$. The remaining $n - 6$ records have feature $e_4/2$ and label zero. All records are bounded, $b = \theta = 0$, and

$$\Sigma = \operatorname{diag}\left(\frac{1}{2n}, \frac{1}{2n}, \frac{1}{2n}, \sigma_n \right), \quad \sigma_n = \frac{n-6}{4n},$$

$$A_0 = \sigma_n I_4, \qquad G = \frac{n-8}{4n} \operatorname{diag}(1, 1, 1, 0).$$

Since $n > 8$, this is the exact source spectral floor. The nonzero singleton responses and their gap images are

$$\Delta_{j,\pm} = \mp \frac{2c_j}{4n - 3} e_j, \qquad G\Delta_{j,\pm} = \mp s_n c_j e_j, \qquad s_n = \frac{n-8}{2n(4n-3)}.$$

All other singleton responses vanish. Thus every actual response basis selects one response from each of the three active axes, and $q_K = 3$. With $a_0 = 2(n-3)/n$, its Gram and certificate are

$$T = s_n^2 \operatorname{diag}(1, 1, \epsilon^2), \qquad \mu_{\text{full}} = \frac{2s_n^2}{3a_0^2(2+\epsilon^{-2})}.$$

Selecting just one witness on each strong axis instead gives the valid inverse-trace certificate

$$\mu_{\text{strong}} = \frac{s_n^2}{4a_0^2} \le \mathrm{OPT}_{1,1}, \qquad \frac{\mu_{\text{strong}}}{\mu_{\text{full}}} = \frac{3}{8}(2+\epsilon^{-2}).$$

This bank has $r+1 = 2$ independent gap images; it need not span all actual responses. The default greedy singleton selection chooses these strong axes when $\epsilon < 1$, and its subsequent PSD bisection cannot weaken their inverse-trace certificate.

For the remainder in Theorem 77, the exact common curvature bound and the reference resolvent norm are

$$\eta_n = \frac{5}{4n+2}, \qquad \|Q_0\|_{\text{op}} = \frac{n-8}{5n-6}.$$

Put $b_n = 1 + 1/(2n)$ and $\beta_n = 1 - 1/n$. The three quantities in (319) become

$$\frac{1}{4n^2 b_n \beta_n}, \qquad \frac{2+\epsilon^2}{2n^2 b_n}, \qquad \frac{2\eta_n(2+\epsilon^2)}{n\beta_n}.$$

For $n \ge 16$ the first is smallest, so

$$\delta_{1,n} = \left[(1-\eta_n)^{-3/2} - 1\right] \frac{n-8}{5n-6} \frac{1}{2n\sqrt{b_n\beta_n}}.$$

Now set $\epsilon = n^{-3}$. All input coordinates still have $O(\log n)$ bits. Direct limits give

$$n^2\delta_{1,n} \longrightarrow \frac{3}{16}, \qquad \sqrt{\mu_{\text{full}}} \sim \frac{\epsilon}{8\sqrt{6}\,n}, \qquad \sqrt{\mu_{\text{strong}}} \sim \frac{1}{32n}.$$

Consequently

$$\frac{\delta_{1,n}}{\sqrt{\mu_{\text{full}}}} \sim \frac{3\sqrt{6}}{2} n^2 \longrightarrow \infty, \qquad \frac{\delta_{1,n}}{\sqrt{\mu_{\text{strong}}}} \sim \frac{6}{n} \longrightarrow 0.$$

The curvature bound η_n tends to zero, proving the claim. Even with n fixed, letting $\epsilon \downarrow 0$ makes the first ratio unbounded while $\mathrm{OPT}_{1,1}$ remains bounded below by the same positive μ_{strong}. □

The separation removes a weak-direction dependence from the lower certificate, rather than proving an encoder failure ratio. Budget dependence through C_K remains, and the certificate has no proved approximation ratio to the actual optimum. Strict positivity alone gives no uniform bound on $\delta_K/\sqrt{\mu}$.

K.3.11 The gap between optimized spectral witnesses and feasible corrections

Even with optimal scalar weights, spectral witnesses leave a gap to corrections constrained by the original covariance gap. We quantify this separation for singleton deletions through an eight-record example and a family exhibiting the joint dependence on rank and gap conditioning.

Throughout this subsection, $n \ge 2$, $\|x_i\| \le 1$, $|y_i| \le 1$, $\lambda = 1$, and $K = 1$. Define

$$\Sigma = \frac{1}{n}\sum_i x_i x_i^\top, \qquad b = \frac{1}{n}\sum_i x_i y_i, \qquad \theta = (\Sigma + I)^{-1} b, \qquad A_0 = \|\Sigma\|_{\text{op}} I, \qquad G = A_0 - \Sigma.$$

Thus A_0 is the source's exact zero-head spectral floor. Put

$$\beta = 1 - 1/n, \qquad a = 1 + \|\Sigma\|_{\text{op}},$$
$$R_i = \Sigma + \beta I - x_i x_i^\top/n, \quad g_i = (x_i x_i^\top + I)\theta - x_i y_i, \quad \Delta_i = R_i^{-1} g_i/n.$$

For an integer $r \ge 0$ and a feasible correction $0 \preceq N \preceq G$, $\operatorname{rank} N \le r$, write $E = G - N$, $H_i = R_i + E$, and use the actual loss

$$\mathcal{L}_i(N) = \frac{1}{\beta}(E\Delta_i)^\top H_i^{-1} R_i H_i^{-1}(E\Delta_i), \qquad \mathrm{OPT}_r = \min_N \max_i \mathcal{L}_i(N). \tag{324}$$

The empty request has zero loss, so the singleton maximum is the full $K = 1$ objective. With eigenvalues ordered nonincreasingly, set

$$z_i = G\Delta_i, \qquad M_w = \sum_i w_i z_i z_i^\top, \qquad \mu^* = \frac{1}{a^2} \max_{w \geq 0,\, \mathbf{1}^\top w = 1} \sum_{j>r} \lambda_j(M_w). \tag{325}$$

An empty spectral tail is zero. The normalization here uses the sharper operator bound $A_0 + I \preceq aI$ available for this scalar base. It does not redefine the earlier certificate's $a_0 = \lambda + \operatorname{tr}(A_0)$ or change its rational-certificate formulas. On the present sources that earlier optimized certificate equals $(a/a_0)^2\mu^*$.

The basic comparison, with no commutativity assumption, is

$$\frac{1}{a^2}\|E\Delta_i\|^2 \leq \mathcal{L}_i(N) \leq \frac{1}{\beta^2}\|E\Delta_i\|^2. \tag{326}$$

Indeed, $R_i \succeq \beta I$, while $\beta I \preceq H_i = A_0 + \beta I - N - x_i x_i^\top/n \preceq aI$. The lower bound follows from $\beta^{-1}H_i^{-1}R_iH_i^{-1} \succeq H_i^{-2} \succeq a^{-2}I$. For the upper bound, congruence of $R_i \preceq H_i$ gives $H_i^{-1}R_iH_i^{-1} \preceq H_i^{-1} \preceq \beta^{-1}I$.

Theorem 90 (Comparison with the optimized spectral witness)**.** *Always $\mu^* \leq \mathrm{OPT}_r$, and $\mu^* = 0$ if and only if $\mathrm{OPT}_r = 0$. If $G \neq 0$, write*

$$\kappa_+(G) = \frac{\lambda_{\max}(G)}{\lambda^+_{\min}(G)}.$$

In the positive branch,

$$\mathrm{OPT}_r \leq \frac{a^2}{\beta^2}(r+1)\kappa_+(G)\mu^* \leq 16(r+1)\kappa_+(G)\mu^*. \tag{327}$$

For $r = 0$, the stronger bound $\mathrm{OPT}_0 \leq (a^2/\beta^2)\mu^* \leq 16\mu^*$ has no gap-condition factor. If $G = 0$ or $r \geq \operatorname{rank} G$, both quantities vanish.

Proof. For any weights, the matrix with columns $\sqrt{w_i}N\Delta_i$ has rank at most r. The best rank-r approximation bound and (326) therefore give

$$\max_i \mathcal{L}_i(N) \geq a^{-2}\sum_i w_i\|(G-N)\Delta_i\|^2 \geq a^{-2}\sum_{j>r}\lambda_j(M_w).$$

Maximizing over w proves the lower bound.

Work on $W = \operatorname{range} G$, with identity I_W, and suppose $0 \leq r < \dim W$. The standard eigenvalue-sum variational formula and compact convex minimax give

$$\tau := a^2\mu^* = \min_{0 \preceq Q \preceq I_W,\, \operatorname{tr} Q \leq r} \max_i z_i^\top (I_W - Q) z_i. \tag{328}$$

For completeness, for fixed w diagonalizing M_w shows that $\max_Q \operatorname{tr}(QM_w)$ is the sum of its largest r eigenvalues: each diagonal entry of Q lies in $[0,1]$ and their sum is at most r, and the leading spectral projector attains the bound. The bilinear function $\sum_i w_i z_i^\top (I_W - Q) z_i$ has compact convex domains, so minimax interchanges its maximum and minimum. These are the classical eigenvalue-sum SDP and point-set subspace relaxation (Vandenberghe & Boyd, 1996; Varadarajan et al., 2007).

Let Q attain (328), and let P project onto r leading eigenvectors of Q. The $(r+1)$st eigenvalue is at most $r/(r+1)$, whence

$$I_W - P \preceq (r+1)(I_W - Q), \qquad \max_i \|(I_W - P)z_i\|^2 \leq (r+1)\tau.$$

The subspace $U = \operatorname{range} P$ need not directly define a feasible residual. Lift it instead by

$$\widetilde{P} = P_{G^{-1/2}U}, \qquad N = G^{1/2}\widetilde{P}G^{1/2}, \qquad T = G^{1/2}(I_W - \widetilde{P})G^{-1/2}.$$

All inverse powers are restricted to W, and N is extended by zero on $W^\perp$. Then $0 \preceq N \preceq G$, $\operatorname{rank} N \leq r$, T annihilates U, and $\|T\|_{\mathrm{op}} \leq \sqrt{\kappa_+(G)}$. Consequently

$$(G-N)\Delta_i = Tz_i = T(I_W - P)z_i, \qquad \max_i \|(G-N)\Delta_i\|^2 \leq \kappa_+(G)(r+1)\tau.$$

The sandwich proves the first upper bound. Bounded features imply $a \le 2$, and $n \ge 2$ implies $\beta \ge 1/2$, proving the constant 16.

If $\mu^* = 0$, uniform weights show that $\dim \operatorname{span}\{z_i\} \le r$. Choose columns Y of actual responses whose gap images form a basis. If the span is nonzero, the shorted correction

$$N = GY(Y^\top GY)^{-1}Y^\top G$$

is feasible, has that span's dimension as its rank, and satisfies $(G-N)\Delta_i = 0$ for every i. To see the last assertion, subtract from Δ_i a combination of the selected columns with the same gap image; the remainder lies in $\ker G \subseteq \ker N$. For the zero span take $N = 0$. This proves the zero equivalence. The cases $G = 0$ and $r \ge \operatorname{rank} G$ also follow by taking $N = G$. Finally, when $r = 0$ the only correction is $N = 0$, and $a^2\mu^* = \max_i \|z_i\|^2$, giving the stated stronger bound. □

Theorem 91 (An eight-record separation)**.** *For every $0 < t \le 1/2$, there is a bounded ridge source with $n = 8$, $d = 3$, and $r = 1$ such that, with $\gamma = 1/512$, $a = 9/8$, and $\beta = 7/8$,*

$$\mu^* = \frac{\gamma^2}{a^2}\frac{t^6}{1+t^2}, \qquad \frac{\gamma^2}{a^2}\frac{t^4}{(3+t)^2} \le \mathrm{OPT}_1 \le \frac{\gamma^2}{\beta^2}t^4(1+t)^2. \tag{329}$$

In particular $\mathrm{OPT}_1/\mu^* = \Theta(t^{-2})$, although $\operatorname{cond}(\Sigma) = 2$ and $\operatorname{cond}(\Sigma + I) = 18/17$. Here $\kappa_+(G) = t^{-2}$. The unstructured rank-one subspace problem already attains the relaxation value $a^2\mu^*$.

Proof. Use an orthonormal basis e_0, e_1, e_2 and let $V = \operatorname{span}\{e_1, e_2\}$. Prescribe

$$\Sigma^\circ = \operatorname{diag}(1/8, 1/8 - t^2/16, 1/16), \qquad D = \Sigma^\circ + (7/8)I = \operatorname{diag}(1, 1 - t^2/16, 15/16).$$

For $\sigma \in \{+1, -1\}$, put

$$v_\sigma = (1+\sigma t)e_1 + te_2, \qquad x_\sigma = Dv_\sigma/4, \qquad \ell_\sigma = 1 - v_\sigma^\top Dv_\sigma/128.$$

On V define

$$B = \operatorname{diag}_V(1 - t^2/2, 1/2) - 2x_+x_+^\top - 2x_-x_-^\top, \qquad f_j = B^{1/2}e_j \quad (j = 1, 2).$$

The eight records are

$$(x_+, \ell_+), (x_+, -\ell_+), (x_-, \ell_-), (x_-, -\ell_-), (e_0, 0), (f_1, 0), (f_2, 0), (0, 0).$$

Indeed, $\|x_\sigma\|^2 \le ((1+t)^2 + t^2)/16 \le 5/32$, and

$$\operatorname{tr}(2x_+x_+^\top + 2x_-x_-^\top) \le (1 + 2t^2)/4 \le 3/8.$$

Since the smaller diagonal entry of the first term of B is $1/2$, this gives $I_V/8 \preceq B \preceq I_V$. Thus $\|f_j\| \le 1$. Also $251/256 \le \ell_\sigma \le 1$. Summing the covariance gives exactly $\Sigma = \Sigma^\circ$; paired labels give $b = \theta = 0$. In particular

$$A_0 = I/8, \qquad G = \tfrac{1}{16}\operatorname{diag}(0, t^2, 1).$$

For either record with feature x_σ,

$$R_i(v_\sigma/4) = (D - x_\sigma x_\sigma^\top/8)(v_\sigma/4) = \ell_\sigma x_\sigma.$$

Hence its response is $\Delta_{\sigma,\pm} = \mp v_\sigma/32$, while the four zero-label records have zero response. On V,

$$z_{\sigma,\pm} = \mp\gamma\big(t^2(1+\sigma t), t\big).$$

Set $q = (1, -t)/\sqrt{1+t^2}$ and $p = (t, 1)/\sqrt{1+t^2}$ in V. Every nonzero response image has squared q-projection $\gamma^2 t^6/(1+t^2)$. For any weights the smaller eigenvalue of $M_w|_V$ is at most its q-Rayleigh quotient, so this bounds the optimized tail above. To attain it, place no weight on the zero responses and give the two σ groups total probabilities $(1 + \sigma\rho)/2$, where

$$\rho = -\frac{t^3}{1+t^2}.$$

The signs of the paired labels do not affect their outer products. In the (q, p) basis the off-diagonal entry then vanishes, and the diagonal entries are

$$\frac{\gamma^2 t^6}{1+t^2}, \qquad \gamma^2\left[t^2(1+t^2) - \frac{t^8}{1+t^2}\right].$$

The second is no smaller than the first. This proves the exact formula for μ^*. Projection onto the line spanned by p has maximum squared residual equal to the first entry, proving the assertion about the unstructured subspace problem.

It remains to bound the feasible residual

$$\Phi = \min_{0 \preceq N \preceq G,\, \operatorname{rank} N \le 1} \max_i \|(G - N)\Delta_i\|.$$

Every feasible N vanishes on e_0 and on V has the form

$$N|_V = \tfrac{1}{16} TQT, \qquad T = \operatorname{diag}(t, 1), \qquad Q = \begin{pmatrix} p_0 & h \\ h & q_0 \end{pmatrix}, \qquad 0 \preceq Q \preceq I_V, \quad \operatorname{rank} Q \le 1.$$

In particular $p_0 + q_0 \le 1$. The nonzero residuals, up to sign and the factor γ, are

$$f_\sigma = \left(t^2[(1-p_0)(1+\sigma t) - h], \quad t[1 - q_0 - h(1+\sigma t)]\right).$$

Let $M = \max_\sigma \|f_\sigma\|$. Their mean and the difference of their second coordinates give

$$|1 - p_0 - h| \le M/t^2, \qquad |1 - q_0 - h| \le M/t, \qquad |h| \le M/t^2.$$

Therefore

$$1 \le 2 - p_0 - q_0 \le |1 - p_0 - h| + |1 - q_0 - h| + 2|h| \le M(3/t^2 + 1/t).$$

Conversely $N = e_2 e_2^\top/16$ is feasible and leaves maximum residual $\gamma t^2(1+t)$. Thus

$$\frac{\gamma t^2}{3+t} \le \Phi \le \gamma t^2(1+t).$$

Applying (326) proves the loss bounds. The claimed condition numbers follow directly from the diagonal entries of Σ and $\Sigma + I$. □

The preceding example separates the actual optimum even from the best scalar weighting of the complete singleton bank. Since its unstructured rank-one projection attains the semidefinite value, the separation is caused by lifting that projection to $0 \preceq N \preceq G$, rather than by suboptimal weights or a fractional rank relaxation. The next construction shows that the rank and gap-condition factors can also occur together.

Proposition 92 (A product lower bound)**.** *For every integer $m \ge 1$ and every $0 < t \le 1$, there is a finite bounded ridge source in dimension $d = 2m + 1$, with $r = 2m - 1$, such that $\mu^* > 0$, $\kappa_+(G) = t^{-2}$, and*

$$\frac{\mathrm{OPT}_r}{\mu^*} \ge \frac{m(1+t^2)}{36t^2} = \frac{(r+1)(\kappa_+(G)+1)}{72}. \tag{330}$$

The construction is an exact realization by legal singleton ridge responses; no finite-bit representation claim is made for this family.

Proof. Decompose $\mathbb{R}^d = \operatorname{span}\{e_0\} \oplus X \oplus Y$, where X and Y have dimension m and fixed identified coordinates. Take a finite Euclidean $1/2$-net $\mathcal{U}$ of the unit sphere in $\mathbb{R}^m$, invariant under coordinate permutations and independent coordinate sign changes. Such a net results by closing any finite $1/2$-net under this finite group. Its uniform second moment is I_m/m: sign changes kill off-diagonal entries, permutations equalize the diagonal entries, and the trace is one.

We first analyze responses $\pm s h_{uv}$, for every $u, v \in \mathcal{U}$, where $s > 0$ and

$$h_{uv} = (u + tv, tu) \in X \oplus Y, \qquad G = g \operatorname{diag}(0, t^2 I_m, I_m), \qquad \gamma = gs > 0.$$

Their gap images, up to sign, are $z_{uv} = \gamma(t^2 u + t^3 v, tu)$. Zero responses may also be present. The isometry from $\mathbb{R}^m$ to $X \oplus Y$ given by $Jw = (w, -tw)/\sqrt{1+t^2}$ satisfies

$$J^\top z_{uv} = \frac{\gamma t^3}{\sqrt{1+t^2}} v.$$

For any weights, the smallest eigenvalue of $M_w|_{X\oplus Y}$ is at most $m^{-1}\operatorname{tr}(J^\top M_w J)$, hence at most $\gamma^2 t^6/[m(1+t^2)]$. This smallest eigenvalue is the tail at $r = 2m-1$, since M_w vanishes on e_0. Equality is attained by choosing u uniformly, taking $v = \sigma u$ with probabilities $(1+\sigma\rho)/2$, and using $\rho = -t^3/(1+t^2)$. Indeed, in the orthogonal coordinates $Jw = (w, -tw)/\sqrt{1+t^2}$ and $J_\perp w = (tw, w)/\sqrt{1+t^2}$, the off-diagonal covariance block vanishes and the diagonal blocks are

$$\frac{\gamma^2 t^6}{m(1+t^2)} I_m, \qquad \frac{\gamma^2}{m}\left[t^2(1+t^2) - \frac{t^8}{1+t^2}\right] I_m.$$

The latter is no smaller for $0 < t \le 1$, proving

$$\mu^* = \frac{\gamma^2}{a^2}\frac{t^6}{m(1+t^2)}. \tag{331}$$

Every feasible correction on $X \oplus Y$ is $N = gDQD$, where $D = \operatorname{diag}(tI_m, I_m)$, $0 \preceq Q \preceq I$, and $\operatorname{rank} Q \le 2m-1$. Write

$$F = I - Q = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}.$$

The rank restriction implies $\|F\|_{\mathrm{op}} = 1$. The normalized residual $(G-N)sh_{uv}/\gamma$ equals

$$f(u,v) = \left(t^2(A+B)u + t^3 Av, \quad t(B^\top + C)u + t^2 B^\top v\right).$$

Put $M = \max_{u,v\in\mathcal{U}} \|f(u,v)\|$. Averaging over $v, -v$ bounds the two mean blocks, and subtracting them bounds $t^2 B^\top v$. For any linear map L, the net property gives $\|L\|_{\mathrm{op}} \le 2\max_{u\in\mathcal{U}}\|Lu\|$: approximate a maximizing unit vector within $1/2$ and absorb the resulting $\|L\|_{\mathrm{op}}/2$ error. Therefore

$$\|A+B\|_{\mathrm{op}} \le 2M/t^2, \qquad \|B^\top + C\|_{\mathrm{op}} \le 2M/t, \qquad \|B\|_{\mathrm{op}} \le 2M/t^2.$$

As $t \le 1$, both $\|A\|_{\mathrm{op}}$ and $\|C\|_{\mathrm{op}}$ are at most $4M/t^2$. Splitting F into its block diagonal and off-diagonal parts gives $1 = \|F\|_{\mathrm{op}} \le 6M/t^2$. Thus every feasible correction has maximum residual at least $\gamma t^2/6$. The lower sandwich and (331) yield (330).

It remains to realize these responses with the asserted gap. Let $L = |\mathcal{U}|^2$ and set

$$n = 4L + 4d, \qquad \alpha = 1/(4d), \qquad g = \alpha/2, \qquad \Sigma = \alpha e_0 e_0^\top + (\alpha - gt^2)P_X + (\alpha - g)P_Y.$$

This positive definite matrix has operator norm α, so its zero-head gap is exactly the stated G. Put $\beta = 1 - 1/n$, $D_0 = \Sigma + \beta I$, and $c = \sqrt{\alpha}/8$. For each $h = h_{uv}$ include two records with common feature and opposite labels

$$x = cD_0 h, \qquad y = \pm\ell_h, \qquad \ell_h = 1 - \frac{c^2 h^\top D_0 h}{n}.$$

Since $\|h\|^2 \le 5$ and $\|D_0\|_{\mathrm{op}} < 2$, these records satisfy $\|x\|^2 \le 5\alpha/16$ and $0 < \ell_h \le 1$. Their unnormalized covariance is $S_{\mathrm{sig}} = 2\sum_{u,v} xx^\top$, with $\|S_{\mathrm{sig}}\|_{\mathrm{op}} \le 5L\alpha/8$. The covariance still to be supplied is

$$B_0 = n\Sigma - S_{\mathrm{sig}} \succeq 0, \qquad \operatorname{tr} B_0 \le n \operatorname{tr}\Sigma \le n/4.$$

Here positivity follows from $n\lambda_{\min}(\Sigma) \ge n\alpha/2 > 5L\alpha/8$. For each positive eigenvalue λ_j of B_0, include $\lceil\lambda_j\rceil$ zero-label records with feature $\sqrt{\lambda_j/\lceil\lambda_j\rceil}$ times its unit eigenvector. These bounded records supply precisely B_0 and number at most

$$\operatorname{tr} B_0 + d \le n/4 + d = L + 2d.$$

Together with the $2L$ signal records this is fewer than n; fill the remaining slots with zero records. The completed source has exactly covariance Σ and, by label cancellation, $b = \theta = 0$. For a signal record,

$$(D_0 - xx^\top/n)(ch) = \ell_h x, \qquad \Delta_i = \mp(c/n)h.$$

Every zero-label record has zero response. Hence the preceding analysis applies exactly with $s = c/n$, completing the realization. □

K.3.12 A GAP FOR THE PSD TRACE RELAXATION

For a finite vector bank (Δ_i), a PSD matrix G, and an integer $r \geq 0$, consider the squared residual objectives

$$J_r = \min_{0 \preceq N \preceq G,\, \operatorname{rank} N \leq r} \max_i \|(G-N)\Delta_i\|^2,$$
$$\Psi_r = \min_{0 \preceq N \preceq G,\, \operatorname{tr}(G^\dagger N) \leq r} \max_i \|(G-N)\Delta_i\|^2.$$

Here $G^\dagger$ denotes the Moore–Penrose inverse. On $W = \operatorname{range} G$, write $N = G^{1/2} Q G^{1/2}$ with $0 \preceq Q \preceq I_W$. Then $\operatorname{tr}(G^\dagger N) = \operatorname{tr} Q \leq \operatorname{rank} Q = \operatorname{rank} N$, so $\Psi_r \leq J_r$. For the ridge model of the preceding subsection, (326) gives $a^{-2}\Psi_r \leq a^{-2} J_r \leq \mathrm{OPT}_r$. The two residual objectives also have the same zero branch. Indeed, both minima are attained on compact sets. If $\Psi_r = 0$, an attaining Q fixes every $G^{1/2}\Delta_i$; its eigenvalue-one subspace has dimension at most $\operatorname{tr} Q \leq r$. Projection onto the span of these vectors therefore gives a rank-at-most-r feasible correction with zero residual. The converse follows from $\Psi_r \leq J_r$. This argument also covers $G = 0$.

Proposition 93 (Trace feasibility does not approximate rank feasibility)**.** *There is no universal constant bounding J_1/Ψ_1, even for actual singleton responses of bounded ridge sources with $n = 10$, $d = 5$, $\lambda = 1$, and the exact zero-head spectral floor. More precisely, for every real $T \geq 4$ such a source satisfies*

$$\Psi_1 > 0, \qquad \frac{J_1}{\Psi_1} \geq \frac{T^4}{34}, \qquad \frac{\mathrm{OPT}_1}{a^{-2}\Psi_1} \geq \frac{T^4}{34}.$$

This is a real-input construction, without a finite-grid claim.

Proof. First work in $\mathbb{R}^2 \oplus \mathbb{R}^2$, with

$$\overline{G} = \operatorname{diag}(T^4 I_2, T^{-2} I_2), \qquad h_i = (T^{-4} e_i, e_i) \quad (i = 1, 2).$$

Set $B = \overline{G}^{1/2}$ and

$$Q = \begin{pmatrix} pI_2 & bI_2 \\ bI_2 & qI_2 \end{pmatrix}, \qquad p = 4/T^2, \quad b = 1/T, \quad q = 1/2 - 4/T^2.$$

The scalar two-by-two block has nonnegative diagonal entries, determinant $(T^2 - 16)/T^4 \geq 0$, and trace $1/2$. Thus $0 \preceq Q \preceq I$ and $\operatorname{tr} Q = 1$. The trace-feasible correction $\overline{N} = BQB$ leaves residuals

$$(\overline{G} - \overline{N}) h_i = \left(-\frac{4}{T^2} e_i, \left[\frac{1}{2T^2} + \frac{3}{T^4} \right] e_i \right),$$

whose squared norms are at most $17/T^4$. The trace optimum is strictly positive: a zero residual would force Q to fix the two independent vectors Bh_i, requiring trace at least two. For any rank-one correction, the top two rows of $(\overline{G} - \overline{N})[h_1, h_2]$ are $I_2 - K$ with $\operatorname{rank} K \leq 1$. Since K has a unit null vector, $\|I_2 - K\|_F^2 \geq 1$. The sum of the two full squared residuals is therefore at least one, proving $J_1 \geq 1/2$ and the abstract ratio $J_1/\Psi_1 \geq T^4/34$.

To realize this example, use coordinates $\operatorname{span}\{e_0\} \oplus X \oplus Y$ with $\dim X = \dim Y = 2$, and set

$$n = 10, \quad \alpha = 1/16, \quad g = 1/32, \quad G = g \operatorname{diag}(0, I_2, T^{-6} I_2), \quad \Sigma = \alpha I - G, \quad \beta = 9/10.$$

Embed h_i as $(0, T^{-4} e_i, e_i)$, put $D_0 = \Sigma + \beta I$, and let $c = 1/8$. Include the four records with features $x_i = c D_0 h_i$ and labels $\pm \ell_i$, where

$$\ell_i = 1 - c^2 h_i^\top D_0 h_i / n.$$

Here $\|D_0\|_{\mathrm{op}} = 77/80 < 1$ and $\|h_i\|^2 \leq 2$, so $\|x_i\|^2 \leq 1/32$ and $1 - 1/320 \leq \ell_i \leq 1$. The features x_1, x_2 are orthogonal; hence $S_{\mathrm{sig}} = 2x_1 x_1^\top + 2x_2 x_2^\top$ has operator norm at most $1/16$. It follows that

$$B_{\mathrm{fill}} = n\Sigma - S_{\mathrm{sig}}, \qquad \tfrac{1}{4} I \preceq B_{\mathrm{fill}} \preceq \tfrac{5}{8} I.$$

Add the five columns of $B_{\mathrm{fill}}^{1/2}$ as zero-label features and one zero record. All features are bounded, and the ten records have exactly covariance Σ. Their labels cancel, so $b = \theta = 0$, and $\|\Sigma\|_{\mathrm{op}} = \alpha$ gives $A_0 = \alpha I$ and the prescribed gap. The identity

$$(D_0 - x_i x_i^\top / n)(c h_i) = \ell_i x_i$$

shows that the four signal responses are $\mp s h_i$ with $s = c/n = 1/80$; all other responses vanish. On $X \oplus Y$ the gap is $(g/T^4)\overline{G}$. Scaling the abstract objectives thus gives

$$J_1 \geq \frac{s^2 g^2}{2T^8}, \qquad 0 < \Psi_1 \leq \frac{17 s^2 g^2}{T^{12}}.$$

These inequalities prove the residual ratio, and the lower sandwich proves the actual-loss ratio. □

This separation does not refute a bound using both certificates. Write $\tau_r = \max_w \sum_{j>r} \lambda_j(M_w) = a^2\mu^*$ in the ridge setting. In the abstract example above, the images $\overline{G}h_i = (e_i, T^{-2}e_i)$ are orthogonal with equal squared norm $1 + T^{-4}$. Hence $\tau_1 = (1 + T^{-4})/2$, while $N = 0$ gives $J_1 \leq 1 + T^{-4} = 2\tau_1$. Conversely, the residual lower bound in Theorem 91 uses only $p_0 + q_0 \leq 1$ once the correction is whitened. It therefore also applies to the trace relaxation, giving $\Psi_1 = \Theta(t^4)$ there, whereas $\tau_1 = \Theta(t^6)$. Thus neither certificate uniformly dominates the other up to a constant factor. Whether $J_1 \leq C \max\{\tau_1, \Psi_1\}$ holds with a universal constant C remains open here.

K.3.13 Dyadic inputs and exact dimension embedding

The separation in Theorem 91 persists with finite-bit inputs. Throughout this subsection the spectral floor is $A_0 = \|\Sigma\|_{\mathrm{op}} I$. As in the preceding subsection, the optimized certificate uses the sharper normalization $a = 1 + \|\Sigma\|_{\mathrm{op}}$; the earlier normalization $a_0 = 1 + \mathrm{tr}(A_0)$ instead gives the certificate $(a/a_0)^2\mu^*$. Weights and feasible corrections are still arbitrary real numbers; only the input records are quantized. We retain singleton deletion, $n = 8$, $\lambda = 1$, and rank budget $r = 1$, and write $\alpha = 1/8$, $a = 9/8$, $\beta = 7/8$, and $\gamma = 1/512$. For the rounded instance we abbreviate $\widehat{\mu} = \widehat{\mu^*}$ and $\widehat{\mathrm{OPT}} = \widehat{\mathrm{OPT}}_1$, with analogous shorthand for the ideal and embedded instances.

Theorem 94 (Dyadic separation)**.** *For every integer $k \geq 1$, put $t = 2^{-k}$ and $h = 2^{-(4k+12)}$. There is a bounded eight-record dataset in $\mathbb{R}^3$ whose feature and label coordinates belong to $h\mathbb{Z}$, whose original fitted parameter is $\theta = 0$, and whose spectral floor is exactly $A_0 = I/8$. Its optimized weighted spectral bound $\widehat{\mu}$ and optimal worst-request loss $\widehat{\mathrm{OPT}}$ satisfy*

$$\frac{\gamma^2}{2a^2} t^6 \leq \widehat{\mu} \leq \frac{2\gamma^2}{a^2} t^6, \tag{332}$$

$$\frac{\gamma^2}{16a^2} t^4 \leq \widehat{\mathrm{OPT}} \leq \frac{4\gamma^2}{\beta^2} t^4. \tag{333}$$

Consequently,

$$\frac{1}{32} t^{-2} \leq \frac{\widehat{\mathrm{OPT}}}{\widehat{\mu}} \leq \frac{648}{49} t^{-2}. \tag{334}$$

Proof. Use coordinates (e_0, e_1, e_2) and put $V = \mathrm{span}(e_1, e_2)$. Superscript $\circ$ denotes the unrounded construction. Set

$$\begin{aligned}
\Sigma^\circ &= \mathrm{diag}(1/8, 1/8 - t^2/16, 1/16), \\
D &= \Sigma^\circ + \beta I = \mathrm{diag}(1, 1 - t^2/16, 15/16), \\
v_\sigma &= (1 + \sigma t)e_1 + t e_2, \\
x_\sigma &= D v_\sigma / 4, \\
\ell_\sigma &= 1 - v_\sigma^\top D v_\sigma / 128 \qquad (\sigma \in \{-1, 1\}).
\end{aligned}$$

Define, on V,

$$B = \mathrm{diag}_V(1 - t^2/2, 1/2) - 2x_+x_+^\top - 2x_-x_-^\top, \qquad f_j = B^{1/2} e_j \quad (j = 1, 2).$$

Round the four V coordinates of f_1, f_2 to nearest multiples of h, with any fixed tie rule, leaving their e_0 coordinates zero. Call the resulting vectors $\widehat{f}_j$. The records are

$$(x_+, \ell_+), (x_+, -\ell_+), (x_-, \ell_-), (x_-, -\ell_-), (e_0, 0), (\widehat{f}_1, 0), (\widehat{f}_2, 0), (0, 0).$$

No rounding of signal records is necessary: their nonzero feature coordinates and their labels are

$$x_{\sigma,1} = 1/4 + \sigma t/4 - t^2/64 - \sigma t^3/64, \qquad x_{\sigma,2} = 15t/64,$$
$$\ell_\sigma = 127/128 - \sigma t/64 - 15t^2/1024 + \sigma t^3/1024 + t^4/2048.$$

Their denominator exponents are at most $3k + 6$ and $4k + 11$, respectively, so they lie on the prescribed grid.

The unrounded estimates give $B \succeq I_V/8$, $\|x_\sigma\|^2 \le 5/32$, and $251/256 \le \ell_\sigma \le 1$. For completeness, $\left\|2x_+x_+^\top + 2x_-x_-^\top\right\|_{\mathrm{op}} \le \mathrm{tr}(2x_+x_+^\top + 2x_-x_-^\top) \le 3/8$, whereas $\mathrm{diag}_V(1 - t^2/2, 1/2) \succeq I_V/2$. Also $B_{22} \le 1/2$. Writing $d_1 = 1 - t^2/16 \ge 15/16$, the subtracted $(1,1)$ entry is $d_1^2(1 + t^2)/4 \ge 225/1024$; thus $B_{11} \le 799/1024 < 13/16$. It follows that $\|f_j\| \le 15/16$. Each rounding error has norm at most h, and $h \le 2^{-16}$, so all rounded features remain bounded by one.

Let $S = \widehat{\Sigma} - \Sigma^\circ$. It is supported on V and

$$\|S\|_{\mathrm{op}} \le (4h + 2h^2)/8 \le h. \tag{335}$$

Since $\widehat{\Sigma}e_0 = e_0/8$ and $\lambda_{\max}(\widehat{\Sigma}|_V) \le 1/8 - t^2/16 + h < 1/8$, the spectral floor is exactly $I/8$. Consequently

$$G^\circ = \mathrm{diag}(0, t^2, 1)/16, \qquad \widehat{G} = G^\circ - S.$$

Both have kernel $\mathrm{span}(e_0)$. The two positive eigenvalues of $\widehat{G}$ lie, respectively, within h of $t^2/16$ and $1/16$, and its smallest positive eigenvalue is at least $t^2/32$. With

$$\eta = 16h/t^2 = t^2/256 \le 1/1024, \qquad (1 - \eta)G^\circ \preceq \widehat{G} \preceq (1 + \eta)G^\circ, \tag{336}$$

the inequalities follow on V from $G^\circ \succeq (t^2/16)I_V$, and on its orthogonal complement from the common kernel. In particular, no commutation of S and G° is assumed.

Paired labels make the original fitted parameter $\theta = 0$, and deleting any zero-label record has zero response. For a signal deletion, $\widehat{R}_i = R_i^\circ + S$ and the ideal response is $\Delta_i^\circ = \mp v_\sigma/32$, according to the label sign. Thus $\max_i \|\Delta_i^\circ\| \le 1/16$, and, since $\widehat{R}_i \succeq \beta I$,

$$\widehat{\Delta}_i - \Delta_i^\circ = -\widehat{R}_i^{-1} S \Delta_i^\circ, \qquad \max_i \left\|\widehat{\Delta}_i - \Delta_i^\circ\right\| \le h/14.$$

Writing $z_i^\circ = G^\circ \Delta_i^\circ$ and $\widehat{z}_i = \widehat{G}\widehat{\Delta}_i$, we have $\left\|\widehat{G}\right\|_{\mathrm{op}} \le 1/8$ and hence

$$\max_i \|\widehat{z}_i - z_i^\circ\| \le h/14. \tag{337}$$

For weights $w_i \ge 0$ summing to one, let Z_w be the 2×8 matrix with columns $\sqrt{w_i} z_i^\circ$, in coordinates on V, and define $\widehat{Z}_w$ similarly. The optimized spectral quantity is $a^{-2} \max_w \sigma_{\min}(Z_w)^2$. Theorem 91 gives

$$\mu^\circ = \frac{\gamma^2}{a^2} \frac{t^6}{1 + t^2}.$$

Equation (337) implies $\left\|\widehat{Z}_w - Z_w\right\|_{\mathrm{op}} \le \left\|\widehat{Z}_w - Z_w\right\|_{\mathrm{F}} \le h/14$ uniformly in w. The singular-value perturbation inequality, followed by maximization over all weights, therefore gives

$$|a\sqrt{\widehat{\mu}} - a\sqrt{\mu^\circ}| \le h/14.$$

Since $h/(a\sqrt{\mu^\circ}) = t\sqrt{1 + t^2}/8 \le \sqrt{5}/32 < 1/8$, we obtain $49\mu^\circ/64 \le \widehat{\mu} \le 81\mu^\circ/64$. Using $1 \le 1 + t^2 \le 5/4$ proves (332).

To control the true loss, define the residual optimum

$$\Phi(G, \Delta) = \min_{0 \preceq N \preceq G,\ \mathrm{rank}(N) \le 1} \max_i \|(G - N)\Delta_i\|.$$

Write Φ° and $\widehat{\Phi}$ for the two instances. The ideal bounds are $\gamma t^2/(3 + t) \le \Phi^\circ \le \gamma t^2(1 + t)$. Every rounded feasible $\widehat{N}$ maps to the ideal feasible $N = \widehat{N}/(1 + \eta)$; conversely, every ideal

feasible N maps to $\widehat{N} = (1-\eta)N$. Both maps preserve positive semidefiniteness and rank, and satisfy $\left\|\widehat{N} - N\right\|_{\mathrm{op}} \le \eta \left\|G^\circ\right\|_{\mathrm{op}} = h/t^2$. For corresponding residual matrices $\widehat{E} = \widehat{G} - \widehat{N}$ and $E^\circ = G^\circ - N$, this gives

$$\left\|\widehat{E} - E^\circ\right\|_{\mathrm{op}} \le h + h/t^2, \qquad \left\|\widehat{E}\right\|_{\mathrm{op}} \le 1/8.$$

For every request the residual norm changes by at most

$$h/112 + (h + h/t^2)/16 \le h/(8t^2).$$

Applying the two maps in opposite directions proves $|\widehat{\Phi} - \Phi^\circ| \le h/(8t^2) = \gamma t^2/64$, and hence

$$\gamma t^2/4 \le \widehat{\Phi} \le 2\gamma t^2.$$

The loss sandwich (326) $a^{-2}\widehat{\Phi}^2 \le \widehat{\mathrm{OPT}} \le \beta^{-2}\widehat{\Phi}^2$ proves (333); division yields (334). The square root defining the f_j specifies the existence of the rounded inputs; no running-time claim for generating them is needed. □

Lemma 95 (Exact zero-coordinate embedding)**.** *Appending zero feature coordinates to the dyadic construction preserves both its optimized weighted spectral bound and its optimal worst-request loss exactly. This conclusion allows arbitrary real feasible corrections, including corrections coupling old and new coordinates.*

Proof. Embed into $\mathbb{R}^d$, $d \ge 4$, by $x_i^{(d)} = (\widehat{x}_i, 0)$. Then

$$\Sigma^{(d)} = \widehat{\Sigma} \oplus 0, \qquad A_0^{(d)} = \alpha I_d, \qquad G^{(d)} = \widehat{G} \oplus \alpha I_{d-3},$$
$$R_i^{(d)} = \widehat{R}_i \oplus \beta I_{d-3}, \qquad \Delta_i^{(d)} = (\widehat{\Delta}_i, 0), \qquad z_i^{(d)} = (\widehat{z}_i, 0).$$

The weighted covariance eigenvalues merely acquire zeros, and the sharper normalization $a = 1 + \left\|\Sigma^{(d)}\right\|_{\mathrm{op}} = 9/8$ is independent of the ambient dimension, so μ is unchanged. The earlier trace-normalized certificate would not be preserved exactly, since $a_0 = 1 + \mathrm{tr}(A_0^{(d)}) = 1 + d/8$ changes with d. For the true optimum, take any $0 \preceq N \preceq G^{(d)}$ with $\mathrm{rank}(N) \le r$ and write it in old/new coordinate blocks. Put

$$C = (\alpha+\beta)I - N_{22} = I - N_{22} \succeq \beta I, \qquad N_{\mathrm{eff}} = N_{11} + N_{12}C^{-1}N_{21}.$$

The matrix $\mathrm{diag}(\widehat{G}, (\alpha+\beta)I) - N$ is positive semidefinite, so its Schur complement gives $0 \preceq N_{\mathrm{eff}} \preceq \widehat{G}$. Writing $N = UU^\top$ in the same blocks shows

$$N_{\mathrm{eff}} = U_1(I + U_2^\top C^{-1} U_2)U_1^\top, \qquad \mathrm{rank}(N_{\mathrm{eff}}) \le \mathrm{rank}(N) \le r.$$

It is independent of the request because β is common to all singleton deletions.

Let $H_i^{(d)} = R_i^{(d)} + G^{(d)} - N$ and $u_i = (H_i^{(d)})^{-1} R_i^{(d)} \Delta_i^{(d)}$. Block elimination gives

$$u_{i,1} = (\widehat{R}_i + \widehat{G} - N_{\mathrm{eff}})^{-1} \widehat{R}_i \widehat{\Delta}_i.$$

Since $(H_i^{(d)})^{-1}(G^{(d)} - N)\Delta_i^{(d)} = \Delta_i^{(d)} - u_i$, and the exact loss is $\beta^{-1}(\Delta_i^{(d)} - u_i)^\top R_i^{(d)}(\Delta_i^{(d)} - u_i)$, the individual losses satisfy the exact identity

$$\mathcal{L}_i^{(d)}(N) = \widehat{\mathcal{L}}_i(N_{\mathrm{eff}}) + \left\|u_{i,2}\right\|^2.$$

Thus every ambient correction compresses to a legal three-dimensional correction with no greater loss, proving $\mathrm{OPT}^{(d)} \ge \widehat{\mathrm{OPT}}$. Embedding any original correction as $\widehat{N} \oplus 0$ gives the reverse inequality. The argument preserves any rank budget r, although the separation here uses $r = 1$. □

Taking $k = \lfloor \log_2 d \rfloor$ gives $1/d \le t < 2/d$, precision $4\lfloor \log_2 d \rfloor + 12$, and, after exact embedding,

$$\mu = \Theta(d^{-6}), \qquad \mathrm{OPT} = \Theta(d^{-4}), \qquad \mathrm{OPT}/\mu = \Theta(d^2).$$

This is an $O(\log d)$-bit input grid, but is generally finer than the particular grid in (4). That grid also supports a diverging separation, with a slower choice of k.

Corollary 96 (Separation on the prescribed input grid)**.** *For every $d \geq 16$, there is a bounded eight-record instance on the exact input grid* (4) *with*

$$\mu = \Theta(d^{-3/2}), \qquad \mathrm{OPT} = \Theta(d^{-1}), \qquad \mathrm{OPT}/\mu = \Theta(\sqrt{d}).$$

Proof. Take $k = \lfloor (\log_2 d)/4 \rfloor \geq 1$. Then $4k+12 \leq p_{\mathrm{in}}(d) = \lceil \log_2(2^{22}d) \rceil$, so the constructed grid is contained in the prescribed one, while $d^{-1/4} \leq t < 2d^{-1/4}$. Apply Theorem 94 and Lemma 95. □

These are fixed-sample geometric separations: $n = 8$ and the feature covariance has rank three in every ambient dimension. They lie outside the main memory theorem's regime $d \leq n \leq 3d$ for large d. One can spread the coordinates without changing this fact: for $d = 4^m$, $k = 2m$, a Sylvester Hadamard matrix gives the dyadic orthogonal map $Q = 2^{-m}H_d$. The rotated inputs use precision $9m + 12$, and Qe_0 has every coordinate nonzero; orthogonal invariance preserves the two optima, and the covariance rank remains three.

K.4 Request legality from the signed state

The signed-state extension requires an exact invariant covariance head. A signed correction of an exact spectral floor satisfies this requirement, including the response-metric choice in Appendix K.3. A generic positive factor mixing covariance and response directions may fail this assumption.

Proposition 97 (A signed-state covariance interval)**.** *Use normalized original moments $\Sigma = n^{-1}\sum_S xx^\top$ and $b = n^{-1}\sum_S xy$. Suppose the exact stored factors satisfy*

$$\Sigma_{\mathrm{hi}} = \delta I + FF^\top - LL^\top \succeq \Sigma, \qquad F^\top L = 0.$$

Let V be any orthonormal basis of range F *and assume the exact-head property*

$$\Sigma V = VA, \qquad A = \delta I + V^\top FF^\top V. \tag{338}$$

Let Z complete V to an orthogonal basis, and put $D_{\mathrm{hi}} = \delta I - Z^\top LL^\top Z$. For a legal request define

$$C = \frac{1}{n}\sum_U xx^\top, \qquad C_{11} = V^\top CV, \qquad C_{12} = V^\top CZ,$$
$$C_{21} = C_{12}^\top, \qquad C_{22} = Z^\top CZ, \qquad J = A - C_{11}.$$

Then $J \succeq 0$, $(I - JJ^\dagger)C_{12} = 0$, and, for

$$D_{\mathrm{lo}} = C_{22} + C_{21}J^\dagger C_{12}, \qquad W = D_{\mathrm{hi}} - D_{\mathrm{lo}}, \tag{339}$$

one has

$$0 \preceq D_{\mathrm{lo}} \preceq Z^\top \Sigma Z \preceq D_{\mathrm{hi}}, \qquad W \succeq 0.$$

Consequently

$$\Sigma_{\mathrm{lo}} = VAV^\top + ZD_{\mathrm{lo}}Z^\top, \qquad \Sigma_{\mathrm{lo}} - C \succeq 0, \qquad \Sigma_{\mathrm{lo}} \preceq \Sigma \preceq \Sigma_{\mathrm{hi}}. \tag{340}$$

All the displayed quantities are available from (F, L, δ, n) and the deleted features; Z and the full matrices are temporary decoder workspace.

Proof. The exact-head property and symmetry give $\Sigma = VAV^\top + ZDZ^\top$, where $D = Z^\top \Sigma Z \succeq 0$. The upper bound and $F^\top L = 0$ give $D \preceq D_{\mathrm{hi}}$. In the basis $[V, Z]$, legality is exactly

$$\begin{pmatrix} J & -C_{12} \\ -C_{21} & D - C_{22} \end{pmatrix} \succeq 0.$$

Its generalized Schur criterion gives the stated range condition and $D \succeq C_{22} + C_{21}J^\dagger C_{12} = D_{\mathrm{lo}}$. Conversely the same range condition shows the block matrix is positive semidefinite at $D = D_{\mathrm{lo}}$, proving $\Sigma_{\mathrm{lo}} - C \succeq 0$. □

Zero columns of F are omitted when constructing its range. In particular, head eigenvectors whose eigenvalue equals the floor δ are allowed to move into Z; the bounds still hold. If F has no nonzero columns, the formula reads $D_{\mathrm{lo}} = C$ in the full tail basis. The assumption (338) is preserved when zero-weight directions are omitted. It is not implied by $F^\top L = 0$ alone. If an encoder's declared full-rank factor dimensions determine $\Sigma_{\mathrm{hi}} = \Sigma$ identically, the decoder instead uses that exact covariance. This applies to a full covariance head or to a response-metric correction selecting every tail direction. It requires no comparison to the original covariance at decoding time.

Corollary 98 (Centered legal midpoint at the same storage)**.** *Let $\theta = w_S$ be stored exactly. For a request of fraction α and $\beta = 1 - \alpha > 0$, put $d_U = n^{-1} \sum_U xy$ and form*

$$\begin{aligned} Q_{\mathrm{lo}} &= (\Sigma_{\mathrm{lo}} - C)/\beta + \lambda I, \quad Q_{\mathrm{hi}} = (\Sigma_{\mathrm{hi}} - C)/\beta + \lambda I, \\ H &= (Q_{\mathrm{lo}} + Q_{\mathrm{hi}})/2, \qquad\quad g = [d_U - (C + \alpha\lambda I)\theta]/\beta, \\ a &= -H^{-1} g, \qquad\quad \widehat{w}_{\mathrm{lm}} = \theta + a. \end{aligned} \tag{341}$$

Then g is the true retained half-gradient at θ and $\lambda I \preceq Q_{\mathrm{lo}} \preceq Q_T \preceq Q_{\mathrm{hi}}$, where $Q_T = \Sigma_{S-U} + \lambda I$. Let

$$t = \lambda_{\max}\left[Q_{\mathrm{lo}}^{-1/2}(ZWZ^\top/\beta)Q_{\mathrm{lo}}^{-1/2}\right], \qquad \gamma = \frac{t}{2+t}, \qquad a_t = Z^\top a.$$

The retained excess obeys

$$\begin{aligned} &F_T(\widehat{w}_{\mathrm{lm}}) - F_T(w_T) \le \gamma^2 [F_T(\theta) - F_T(w_T)], \\ &F_T(\widehat{w}_{\mathrm{lm}}) - F_T(w_T) \le \min\left\{\gamma^2 g^\top Q_{\mathrm{lo}}^{-1} g,\ \frac{\|W\|_{\mathrm{op}}\, a_t^\top W a_t}{4\beta^2\lambda}\right\}. \end{aligned} \tag{342}$$

No persistent response basis, covariance action, or extra matrix is needed beyond $(F, L, \delta, \theta, n)$.

Proof. Original stationarity gives $b = (\Sigma + \lambda I)\theta$, which verifies the formula for g. The interval follows from (340). Put $P = H - Q_T$. The width is $ZWZ^\top/\beta$, so $-ZWZ^\top/(2\beta) \preceq P \preceq ZWZ^\top/(2\beta)$. The generalized midpoint argument from Proposition 131 gives contraction γ^2. Also $Q_T^{-1} \preceq Q_{\mathrm{lo}}^{-1}$. For the second posterior term, the exact residual is $Q_T(\widehat{w}_{\mathrm{lm}} - w_T) = -Pa$. The signed tail interval gives $\|Pa\|^2 \le \|W\|_{\mathrm{op}} a_t^\top W a_t/(4\beta^2)$. Using $Q_T^{-1} \preceq \lambda^{-1} I$ finishes the proof. □

The unsigned spectral floor with the same exact head has upper tail δI. Thus the signed state replaces its interval width by the smaller matrix $W = \delta I - Z^\top L L^\top Z - D_{\mathrm{lo}}$. This is an interval refinement, not an ordering of the realized errors of the two midpoint rules. If $W = 0$, the midpoint is exact. For the empty request its centered output equals θ.

These are exact-arithmetic statements. Factor rounding can alter the head subspace and its exact covariance invariance. A pseudoinverse cutoff can invalidate the range condition. Therefore the formulas above do not establish a finite-bit or floating-point midpoint guarantee. Such a claim needs verified interval enclosures that include factor, reference, range, and solve errors.

K.4.1 Smaller positive covariance error need not lower excess

The failure of pointwise monotonicity occurs on a finite realizable noiseless ridge problem. Take $d = 5$, $\lambda = 1$, $n = 400$ and

$$\Sigma = \begin{pmatrix} .10 & -.03 & -.03 \\ -.03 & .11 & 0 \\ -.03 & 0 & .18 \end{pmatrix} \oplus .2 \oplus .3.$$

This covariance is realized by the following feature counts: 25 copies each of $(.6, -.8, 0, 0, 0)$ and $(.6, 0, -.8, 0, 0)$; 22, 28, 56, 80, and 120 copies of $e_1, \ldots, e_5$, respectively; and 44 zero vectors. Every feature has norm at most one. Delete 40 zero records, so $\alpha = .1$, $\beta = .9$, and $C = 0$.

Let $\xi = .01$, let $d = (0, -2/5, 1, 0, 0)^\top$, and define

$$\theta = \frac{\xi}{\alpha}(\Sigma + \beta I)d, \qquad v = \Sigma^{-1}(\Sigma + I)\theta, \qquad y_i = x_i^\top v.$$

The nonzero coordinates of v/ξ are

$$\left(\frac{55281}{9550}, -\frac{186891}{4775}, \frac{68564}{955}\right).$$

Their scaled absolute values are less than $.06, .4, .72$, respectively, so $\|v\|^2 < .682 < 1$ and all labels are bounded. By construction $\theta = w_S$. Deleting zero records changes the normalization, giving $\Delta_U = w_T - \theta = \xi d$ exactly.

Use the exact covariance head $V = e_5$ and the additional response direction $W_0 = e_1$. The next covariance eigenvalue is exactly $\delta = .2$: the first block is strictly below $.2I$, and e_4 has eigenvalue $.2$. The positive floor factor is $F = \sqrt{.1}\, e_5$. The spectral gap has first block

$$G_3 = \begin{pmatrix} .10 & .03 & .03 \\ .03 & .09 & 0 \\ .03 & 0 & .02 \end{pmatrix},$$

which is positive definite; it is zero on e_4, e_5. The old projected completion and the shorted completion have errors

$$E_{\mathrm{old}} = 0 \oplus \begin{pmatrix} .09 & 0 \\ 0 & .02 \end{pmatrix} \oplus 0 \oplus 0, \qquad E_{\mathrm{s}} = 0 \oplus \begin{pmatrix} .081 & -.009 \\ -.009 & .011 \end{pmatrix} \oplus 0 \oplus 0.$$

Here $0 \preceq E_{\mathrm{s}} \preceq E_{\mathrm{old}}$.

Their exact centered excesses are

$$\mathcal{E}_{\mathrm{old}}/\xi^2 = 0.0015988211942\ldots, \qquad \mathcal{E}_{\mathrm{s}}/\xi^2 = 0.0018122709431\ldots.$$

The strict comparison can be checked entirely with rational arithmetic. Put $k = -14/275$, $s_{\mathrm{old}} = -80/5491$, $s_{\mathrm{s}} = -134/5401$, and

$$f(s) = \frac{10}{9}\left(\frac{326}{625}s^2 - \frac{7}{200}sk + \frac{209}{400}k^2\right).$$

The two normalized excesses equal $f(s_{\mathrm{old}})$ and $f(s_{\mathrm{s}})$. Their difference is strictly positive because

$$(s_{\mathrm{s}} - s_{\mathrm{old}})\left[\frac{326}{625}(s_{\mathrm{s}} + s_{\mathrm{old}}) - \frac{7}{200}k\right] > 0.$$

Both bracketed factors are negative. Thus the smaller covariance error incurs about 13.4% more retained excess for this request. The comparison is between two centered one-sided completions on the same selected subspace. Response-metric optimality and the ordering of midpoint rules are separate questions.

L A REQUEST-DEPENDENT DETERMINANTAL BOUND FROM THE SUMMARY

The scalar multiplier in the determinantal response envelope can be sharpened after receiving a deletion request. This calculation needs only the covariance-completion factors and the deleted features. It does not require the original covariance or the retained records.

Proposition 99 (A decoder-computable request bound)**.** *Assume $\lambda > 0$ and $\|x_i\| \le 1$, and use the common DPP covariance $\mathcal{V}$ and response identity from Appendix H. Let V have orthonormal columns, $P = VV^\top$, $\Pi = I - P$, $Z = \Sigma V$, and $J = V^\top \Sigma V$. Suppose the stored completion satisfies $\widehat{\Sigma} = \Sigma + E$ with $0 \preceq E \preceq \delta\Pi$. For a legal request U, put*

$$\alpha = |U|/n, \qquad \beta = 1 - \alpha > 0, \qquad C = \frac{1}{n}\sum_{i \in U} x_i x_i^\top, \qquad s_U = \min\{d, n - |U|\}.$$

Define the following matrices and scalars from the stored factors and the request:

$$\begin{aligned} A_U &= J - V^\top C V, \qquad B_U = Z - CV, \qquad L_U = B_U A_U^\dagger B_U^\top, \\ a_U &= \log\det\left(I + C^{1/2}(\lambda I + L_U)^{-1} C^{1/2}\right), \\ h_U &= \min\left\{\frac{\alpha^2}{\beta\lambda},\, s_U \log\left(1 + \frac{\alpha^2}{4\beta}\right)\right\}, \qquad \overline{D}_U = a_U + h_U. \end{aligned} \tag{343}$$

Here † denotes the Moore–Penrose inverse, and $L_U = 0$ if V has no columns. Then

$$\Delta_U \Delta_U^\top \preceq (e^{\overline{D}_U} - 1)\mathcal{V}, \qquad 0 \le \overline{D}_U \le \frac{\alpha}{\beta\lambda}. \tag{344}$$

In particular the empty request has multiplier zero. If a stored scalar b_Π satisfies $\lambda_{\max}(\Pi\mathcal{V}\Pi) \le b_\Pi$, the ideal centered decoder has retained excess at most

$$\frac{\delta^2 b_\Pi}{\beta^2\lambda}(e^{\overline{D}_U} - 1). \tag{345}$$

If the bound on b_Π holds on an encoding-time event, this conclusion holds on that same event simultaneously for all legal requests.

Proof. Write $R = \Sigma - C \succeq 0$. Its known columns are $RV = B_U$ and $V^\top RV = A_U \succeq 0$. If $A_U u = 0$, then $\|R^{1/2}Vu\|^2 = 0$, so $B_U u = 0$; thus the required range condition for the generalized Schur complement holds. More explicitly,

$$L_U = R^{1/2}\big[R^{1/2}V(V^\top RV)^\dagger V^\top R^{1/2}\big]R^{1/2}.$$

The bracket is an orthogonal projector, proving $0 \preceq L_U \preceq R$. In particular singular A_U causes no ambiguity in the exact statement. The same range condition shows that $\widetilde{L}_U = B_U(A_U + \lambda I)^{-1}B_U^\top \preceq L_U$; using $\widetilde{L}_U$ in a_U preserves the bounds while permitting a positive-definite solve.

For $Q = R/\lambda$, the exact logarithmic density second moment in (200) splits as

$$D_U = \log\frac{\det(\lambda I + R + C)}{\det(\lambda I + R)} + \sum_j \log\left(1 + \frac{\alpha^2 q_j}{(\beta + q_j)^2}\right), \tag{346}$$

where q_j are the eigenvalues of Q. The inverse order $(\lambda I + R)^{-1} \preceq (\lambda I + L_U)^{-1}$ and the determinant identity bound the first term by a_U. Since $L_U \succeq 0$, $a_U \le \log\det(I + C/\lambda) \le \alpha/\lambda$. For the second term, $q/(\beta+q)^2 \le 1/(4\beta)$, and at most s_U eigenvalues are nonzero. Alternatively, $\log(1+\alpha^2 q/(\beta+q)^2) \le \alpha^2 q/\beta^2$ and $\operatorname{tr} Q \le \beta/\lambda$. These give the two branches of h_U. Thus $D_U \le \overline{D}_U \le \alpha/(\beta\lambda)$. The same Cauchy–Schwarz argument as in (202) proves (344).

Projection gives $\|E\Delta_U\|^2 \le \delta^2 b_\Pi(e^{\overline{D}_U} - 1)$. The centered-decoder inequality from Proposition 102 divides this by $\beta^2\lambda$, proving (345). □

For example, if $r = \operatorname{rank} C > 0$, concavity of the scalar logarithm gives

$$e^{\overline{D}_U} - 1 \le \left(1 + \frac{\alpha}{r\lambda}\right)^r \left(1 + \frac{\alpha^2}{4\beta}\right)^{s_U} - 1. \tag{347}$$

At fixed α, d, r, this multiplier grows only polynomially in $1/\lambda$, whereas the universal multiplier is exponential in $1/\lambda$. The retained support can improve it further: if the decoder verifies $C \preceq \kappa L_U$, then $a_U \le r\log(1+\kappa)$, which is independent of λ for fixed κ. These are improvements of the scalar response bound, not a removal of the other factors in (345).

The normalization term cannot be controlled by the deleted covariance rank alone. To see this, let $d \ge 4$, $\lambda = 1/d$, and take αn zero feature vectors and $\beta n/d$ copies of each coordinate vector, with the counts integral, and set $y_i = x_i^\top e_1$ using the public unit teacher e_1. Delete precisely the zero vectors. Then $C = 0$, $Q = \beta I$, and the exact multiplier is

$$e^{D_U} - 1 = \left(1 + \frac{\alpha^2}{4\beta}\right)^d - 1.$$

It is exponential in $1/\lambda$ at fixed positive α, despite the deleted covariance having rank zero.

The proof of (345) uses only the common change-of-measure identity and its finite covariance, not the noiseless trace bound on $\mathcal{V}$. The same conditional statement therefore applies to any DPP fit model with those properties. If b_Π is stored to report a certificate, its encoding cost must be charged. The formulas here are exact-arithmetic statements; numerical matrix evaluation and a measured sample covariance alone do not establish the required event.

M COMPUTING AND CERTIFYING A RESPONSE ENVELOPE

Finite matrix iterations approximate the response envelope without assembling a linear operator on d^2 coordinates. An instance-dependent convergence test and a positive-semidefinite remainder bound yield a certificate that accommodates approximate subspaces. The envelope bounds hold for arbitrary bounded labels, with a sharper trace bound in the noiseless specialization. Numerical certification requires rigorous floating-point enclosures in addition to evaluating these formulas.

Assume $\|x_i\| \le 1$ and $|y_i| \le 1$. Write $\Sigma = n^{-1}\sum_i x_i x_i^\top$, $b = n^{-1}\sum_i x_i y_i$, $A = \Sigma + \lambda I$, $\theta = A^{-1}b$, and, for $0 < \lambda \le 1$, define

$$D_i = x_i x_i^\top + \lambda I, \qquad q_i = (y_i - x_i^\top \theta)x_i - \lambda\theta, \qquad B_0 = A^{-1}\mathbb{E}[q_i q_i^\top]A^{-1}.$$

Here and below expectations average the original records. Let

$$\begin{aligned} \mathcal{T}(M) &= A^{-1}\mathbb{E}[D_i M D_i]A^{-1}, \qquad H = \mathbb{E}[D_i A^{-1} D_i], \\ c_A &= \lambda_{\max}(A^{-1/2} H A^{-1/2}), \qquad \phi(M) = \operatorname{tr}(AM). \end{aligned} \tag{348}$$

A certified upper bound $\bar{c} \ge c_A$ may replace c_A throughout. In exact arithmetic, c_A itself is an available choice.

Proposition 100 (An instance-dependent envelope)**.** *Let $0 < \rho < 1$ and choose $t > 1$ with $s = t\rho\bar{c} < 1$. Put $a_t = t/(t-1)$ and $\gamma = t\rho$. The positive-semidefinite series*

$$B_t = a_t \sum_{j=0}^{\infty} \gamma^j \mathcal{T}^j(B_0) \tag{349}$$

converges, satisfies $B_t = a_t B_0 + \gamma\mathcal{T}(B_t)$, and obeys

$$\Delta_U \Delta_U^\top \preceq \alpha B_t, \qquad \phi(B_t) \le G := \frac{a_t \phi(B_0)}{1-s}, \qquad \operatorname{tr} B_t \le \frac{G}{\lambda} \tag{350}$$

for every legal deletion U of fraction $\alpha \le \rho$, where $\Delta_U = w_{S-U} - \theta$. Moreover, $1 \le c_A \le (1+\lambda)/\lambda$. Thus a choice of t exists whenever $\rho\bar{c} < 1$. The choice $t = (\rho\bar{c})^{-1/2}$ minimizes the displayed bound G and gives

$$s = \sqrt{\rho\bar{c}}, \qquad G = \frac{\phi(B_0)}{(1-\sqrt{\rho\bar{c}})^2}.$$

Proof. The map $\mathcal{T}$ is positive. Since $\mathbb{E}D_i = A$,

$$H - A = \mathbb{E}[(D_i - A)A^{-1}(D_i - A)] \succeq 0.$$

Also $D_i \preceq (1+\lambda)I \preceq \kappa A$, with $\kappa = (1+\lambda)/\lambda$, implies $D_i A^{-1} D_i \preceq \kappa D_i$, and hence $H \preceq \kappa A$. These inequalities give the bounds on c_A. For $M \succeq 0$,

$$\phi(\mathcal{T}(M)) = \operatorname{tr}(MH) \le \bar{c}\phi(M). \tag{351}$$

The positive terms in (349) therefore have a summable weighted trace. Since $A \succeq \lambda I$, the series converges and has the stated trace bounds. Decomposing a symmetric input into its positive and negative parts also shows that $\mathcal{L} = (I - \gamma\mathcal{T})^{-1} = \sum_{j\ge 0}\gamma^j\mathcal{T}^j$ exists on symmetric matrices and is positive.

For a legal query, let $\beta = 1 - \alpha$, $D_U = n^{-1}\sum_{i\in U} D_i$, and $g_U = n^{-1}\sum_{i\in U} q_i$. Stationarity gives $\mathbb{E}q_i = 0$ and

$$\Delta_U = -(A - D_U)^{-1} g_U, \qquad A - D_U = \beta(\Sigma_{S-U} + \lambda I) \succ 0.$$

Consequently $A^{-1}D_U$ is similar to the symmetric matrix $A^{-1/2}D_U A^{-1/2} \prec I$, which is positive semidefinite. Its Neumann series converges, without requiring a Euclidean operator norm bound on $A^{-1}D_U$. Set $\Delta_j = (A^{-1}D_U)^j(-A^{-1}g_U)$. Cauchy–Schwarz over the deleted records gives

$$g_U g_U^\top \preceq \alpha\mathbb{E}[q_i q_i^\top], \qquad D_U M D_U \preceq \alpha\mathbb{E}[D_i M D_i] \quad (M \succeq 0).$$

Induction yields $\Delta_j\Delta_j^\top \preceq \alpha^{j+1}\mathcal{T}^j(B_0)$. Applying scalar weighted Cauchy–Schwarz after testing against any vector, with weights t^{-j}, gives

$$\Delta_U\Delta_U^\top \preceq a_t \sum_{j\ge 0} t^j \Delta_j\Delta_j^\top \preceq \alpha B_t.$$

Finite partial sums justify the calculation before taking limits. The empty query has zero response. Finally, minimizing $t/[(t-1)(1-t\rho\bar{c})]$ over $1 < t < 1/(\rho\bar{c})$ gives the stated choice. □

For arbitrary bounded labels, $\phi(B_0)$ is a measured instance quantity. In particular, if $L = F_S(\theta)$, then $\mathbb{E}[q_i q_i^\top] = \mathcal{R} - \lambda^2\theta\theta^\top$, so $\phi(B_0) \le L/\lambda$. A noiseless teacher of norm at most one gives the sharper $\phi(B_0) \le 1/4$ from (155). With $t = 3/2$ and $\rho\bar{c} \le 1/2$, this recovers $G \le 3$ and the envelope in Appendix F.4. The instance-dependent test is sufficient, not necessary for useful compression or for the convergence of every individual legal response.

Proposition 101 (Finite iterates and a residual certificate)**.** *Under Proposition 100, set $S_0 = 0$ and*

$$W_1 = a_t B_0, \qquad W_{k+1} = \gamma\mathcal{T}(W_k), \qquad S_k = \sum_{j=1}^{k} W_j.$$

For $k \ge 1$, these matrices are positive semidefinite and

$$0 \preceq B_t - S_k \preceq \tau_k A^{-1}, \qquad \tau_k = \frac{s\phi(W_k)}{1-s} \le Gs^k. \tag{352}$$

One may instead use the tighter value $\phi(W_{k+1})/(1-s)$ when the next term has been evaluated.

More generally, let S be any symmetric matrix, including an inexact or rank-truncated iterate, and define

$$R = a_t B_0 + \gamma\mathcal{T}(S) - S, \qquad p_R = \operatorname{tr}[(A^{1/2}RA^{1/2})_+]. \tag{353}$$

Here K_+ replaces the negative eigenvalues of a symmetric K by zero. Then

$$B_t \preceq S + \tau A^{-1}, \qquad \tau = \frac{p_R}{1-s}. \tag{354}$$

In particular the matrix on the right is positive semidefinite. The formulas require bounds for the exact residual and contraction constant; a small floating-point residual alone is not a machine certificate.

Proof. The missing positive terms after S_k have weighted trace at most $s\phi(W_k)/(1-s)$. Every $M \succeq 0$ satisfies $M \preceq \phi(M)A^{-1}$, by applying the scalar trace bound to $A^{1/2}MA^{1/2}$. This proves (352).

For the second assertion, put $R_+^A = A^{-1/2}(A^{1/2}RA^{1/2})_+A^{-1/2}$. Then $R \preceq R_+^A$, $R_+^A \succeq 0$, and $\phi(R_+^A) = p_R$. Positivity of $\mathcal{L}$ gives

$$B_t - S = \mathcal{L}(R) \preceq \mathcal{L}(R_+^A), \qquad \phi(\mathcal{L}(R_+^A)) \le \frac{p_R}{1-s}.$$

Apply the preceding matrix trace bound to $\mathcal{L}(R_+^A)$. □

For an interval or other verified arithmetic implementation, the same argument permits an upper residual enclosure. For example, if $\widetilde{R}$ is symmetric and $\|A^{1/2}(R - \widetilde{R})A^{1/2}\|_{\mathrm{op}} \le \xi$, replace p_R by $\operatorname{tr}[(A^{1/2}\widetilde{R}A^{1/2} + \xi I)_+]$. Errors in forming $A, B_0, \mathcal{T}$, and the contraction test must also be included in such enclosures. No such enclosures are presumed for the floating-point reference implementation.

Proposition 102 (A certificate for an approximate subspace)**.** *Suppose $B_t \preceq S + \tau A^{-1}$ is supplied by Proposition 101. Let V have orthonormal columns, $\Pi = I - VV^\top$, and choose nonnegative certified bounds*

$$\delta \ge \lambda_{\max}(\Pi\Sigma\Pi), \qquad b_S \ge \lambda_{\max}(\Pi S\Pi), \qquad a_A \ge \lambda_{\max}(\Pi A^{-1}\Pi).$$

The universal choices $\delta = 1$ and $a_A = 1/\lambda$ are available. Let $\widehat{\Sigma} = \Sigma + E$, where $E = \Pi(\delta I - \Sigma)\Pi$, and use the centered decoder with this covariance. Its retained excess satisfies

$$F_{S-U}(\overline{w}) - F_{S-U}(w_{S-U}) \le \frac{\alpha\delta^2(b_S + \tau a_A)}{\beta^2\lambda} \le \frac{\rho\delta^2(b_S + \tau a_A)}{(1-\rho)^2\lambda}. \tag{355}$$

No eigengap assumption or exact leading eigenspace is required.

Proof. The covariance tail condition gives $0 \preceq E \preceq \delta\Pi$. Projection of the response envelope yields $\|\Pi\Delta_U\|^2 \le \alpha(b_S + \tau a_A)$ and hence $\|E\Delta_U\|^2 \le \alpha\delta^2(b_S + \tau a_A)$. Writing $Q = \Sigma_{S-U} + \lambda I$ and $K = \beta Q + E$, the centered decoder error is $-K^{-1}E\Delta_U$. Since $K^{-1}(\beta Q)K^{-1} \preceq K^{-1} \preceq (\beta\lambda)^{-1}I$, its excess is at most $\|E\Delta_U\|^2/(\beta^2\lambda)$. □

The subspace certificate uses upper bounds on the omitted spectra. A small invariant-subspace residual alone does not show that the selected eigenvectors correspond to the largest eigenvalues. An implementation can explicitly evaluate the projected symmetric matrices and bound their largest eigenvalues. Orthonormality, or a verified orthogonal projector obtained from a full-rank basis, must be respected.

For example, take the span of the leading $1 \le m < d$ eigenvectors of both Σ and the finite positive iterate S_k. Its dimension is at most $2m$. Since $\operatorname{tr}\Sigma \le 1$ and $\operatorname{tr} S_k \le G/\lambda$, one has $\delta \le 1/(m+1)$ and $b_S \le G/[\lambda(m+1)]$. Choosing

$$k \ge \left\lceil \frac{\log(m+1)}{-\log s} \right\rceil \tag{356}$$

makes $\tau_k/\lambda \le G/[\lambda(m+1)]$, and therefore

$$F_{S-U}(\overline{w}) - F_{S-U}(w_{S-U}) \le \frac{2\rho G}{(1-\rho)^2\lambda^2(m+1)^3}. \tag{357}$$

When $s \le 3/4$, the iteration count is $O(\log(m+1))$. Approximate subspaces with certified $\delta \le 2/(m+1)$ and $b_S \le 2G/[\lambda(m+1)]$ instead give the same rate with constant 12 in place of 2. This statement concerns attained spectral-tail bounds, and does not require closeness to a particular eigenbasis. It is unnecessary to bound the trace of the corrected matrix $S_k + (\tau_k/\lambda)I$, which could introduce an extraneous factor d.

Representation and finite precision. For k_V columns of V, set $Z = \Sigma V$ and $J = V^\top \Sigma V$. The completion has exactly the representation

$$\widehat{\Sigma} = \delta I + VZ^\top + ZV^\top - V(J + \delta I)V^\top.$$

Only $(n, k_V, V, Z, J, \delta, \theta)$ is needed by the decoder, using $O(dk_V + k_V^2 + d)$ numerical entries. The matrices S_k, B_0, A, H and the operator's representation are temporary encoding objects. If a decoder additionally reports the scalar bound in (355), its stored certificate scalars and their precision must also be charged.

For completeness, the utility guarantee admits finite-bit storage even for arbitrary bounded labels. Assume the final, uniform bound in (355) is at most $\varepsilon/4$, where $0 < \varepsilon \le 1$, and take $\delta \le 1$ and $\beta_0 = 1 - \rho$. Choose a public dyadic $\zeta \in (\beta_0\lambda\sqrt{\varepsilon}/8, \beta_0\lambda\sqrt{\varepsilon}/4]$. Round V, Z, J, δ with a public dyadic entry step h within a factor two below $\zeta/(32d)$, preserving symmetry of J. Add $\zeta I/2$ to the covariance reconstructed from those rounded factors. The expansion in Proposition 24 gives a reconstruction error at most $8dh \le \zeta/4$, so the resulting covariance is $\widehat{\Sigma} + K_0$ with $0 \preceq K_0 \preceq \zeta I$. Round the reference to Euclidean error at most $\beta_0\sqrt{\lambda\varepsilon}/16$.

To check these choices, bounded labels give $\|w_T\| \le 1/(2\sqrt{\lambda})$ for every nonempty T: the singular values of the ridge map from $y_T/\sqrt{|T|}$ to w_T are $u/(u^2+\lambda) \le 1/(2\sqrt{\lambda})$. Thus $\|\Delta_U\| \le 1/\sqrt{\lambda}$. Covariance rounding adds at most $\zeta/(\beta\lambda)$ to the excess square root. For reference error e, its extra output is $[\beta Q + E + K_0]^{-1}(A + E + K_0)e$, whose excess square root is at most $4\|e\|/(\beta\sqrt{\lambda})$, since $\|A + E + K_0\|_{\mathrm{op}} \le 4$. Each rounding contribution is at most $\sqrt{\varepsilon}/4$, so the total excess is at most ε. All factor entries have magnitude at most one before rounding; reference coordinates have magnitude at most $1/(2\sqrt{\lambda})$. Including metadata, the state therefore uses

$$O\left(d(k_V+1)\log\left(2 + \frac{d}{(1-\rho)\lambda\sqrt{\varepsilon}}\right) + \log(2+n)\right) \tag{358}$$

bits. This storage statement is distinct from the arithmetic cost of computing the factors and verifying their spectral bounds.

Matrix evaluation and cost. With X containing the original feature rows, one application is

$$\mathcal{T}(M) = A^{-1}\left[\frac{1}{n}X^\top \operatorname{diag}(x_i^\top M x_i)X + \lambda(\Sigma M + M\Sigma) + \lambda^2 M\right]A^{-1}. \tag{359}$$

The analogous formula with $M = A^{-1}$ computes H before its outer congruence. A Cholesky factorization of A supplies the solves; no d^2-by-d^2 matrix is assembled. For $A = LL^\top$, the eigenvalues needed for c_A and p_R are those of $L^{-1}HL^{-\top}$ and $L^\top RL$, respectively. Dense evaluation

uses $O(nd^2 + d^3)$ arithmetic per matrix iteration and $O(d^2)$ working entries beyond the input, with recordwise accumulation if desired. Setup and dense spectral decompositions have the same or smaller arithmetic order, so k iterations cost $O((k+1)(nd^2 + d^3))$ at this level of accounting. Computing verified spectral enclosures and controlling finite arithmetic require additional precision-dependent work; these counts are not bit-operation bounds. Keeping the input in memory adds $O(nd)$ entries. None of this temporary workspace is claimed to equal the persistent summary size.

If $M = LL^\top$ is represented at low rank, the data weights in (359) are $\|L^\top x_i\|^2$. The image can nevertheless have full rank. Low-rank iteration with truncation is therefore an additional implementation choice, whose discarded error must be covered, for example, by (354); no low-rank computational complexity claim follows merely from the short final summary.

M.1 Choosing a different positive trace weight

The choice $\phi(M) = \operatorname{tr}(AM)$ gives a convenient test, but need not give the largest admissible deletion budget. Let $\mathcal{T}^*$ denote the adjoint on symmetric matrices under the Frobenius inner product:

$$\mathcal{T}^*(W) = \mathbb{E}[D_i A^{-1} W A^{-1} D_i].$$

For any positive-definite encoding-time matrix W, define

$$\phi_W(M) = \operatorname{tr}(WM), \qquad c_W = \lambda_{\max}(W^{-1/2}\mathcal{T}^*(W)W^{-1/2}). \tag{360}$$

In particular, $c_W = c_A$ when $W = A$.

Proposition 103 (Dual weights and their spectral limit)**.** *Let $W \succ 0$, and let $\bar{c}_W \geq c_W$ be an upper bound. If $t > 1$ and $s_W = t\rho\bar{c}_W < 1$, the series* (349)*, with $a_t = t/(t-1)$ and $\gamma = t\rho$, is a response envelope and satisfies*

$$\Delta_U \Delta_U^\top \preceq \alpha B_t, \qquad \phi_W(B_t) \leq G_W := \frac{a_t \phi_W(B_0)}{1 - s_W}. \tag{361}$$

For any symmetric approximation S, use the same residual $R = a_t B_0 + \gamma\mathcal{T}(S) - S$ and put

$$\tau_W = \frac{\operatorname{tr}[(W^{1/2} R W^{1/2})_+]}{1 - s_W}. \tag{362}$$

Then $B_t \preceq S + \tau_W W^{-1}$. Consequently the completion in Proposition 102 has excess at most

$$\frac{\alpha\delta^2(b_S + \tau_W a_W)}{\beta^2\lambda}, \qquad a_W \geq \lambda_{\max}(\Pi W^{-1}\Pi). \tag{363}$$

For positive partial sums, the finite-tail formula is likewise $\tau_{W,k} = s_W\phi_W(W_k)/(1 - s_W)$, where W_k here denotes the iteration term in Proposition 101.

Let $r(\mathcal{T})$ be the spectral radius of $\mathcal{T}$ acting on the real vector space of symmetric matrices. Then

$$\inf_{W \succ 0} c_W = r(\mathcal{T}). \tag{364}$$

Thus some positive-definite weight admits this geometric-series test if and only if $\rho\, r(\mathcal{T}) < 1$. This is optimality within the stated class of positive weighted-trace tests, not a necessary condition for a useful deletion summary or for any other response envelope.

Proof. For $M \succeq 0$, the dual inequality gives

$$\phi_W(\mathcal{T}(M)) = \operatorname{tr}(M\mathcal{T}^*(W)) \leq \bar{c}_W \phi_W(M).$$

The convergence, response, and residual proofs above therefore apply with W in place of the trace weight A. The matrix trace bound now reads $M \preceq \phi_W(M)W^{-1}$ for $M \succeq 0$; in the residual argument the dominating positive part is $W^{-1/2}(W^{1/2}RW^{1/2})_+W^{-1/2}$. This proves the displayed bounds and the finite-tail assertion.

To prove (364), use the norm $\|M\|_{W,1} = \|W^{1/2}MW^{1/2}\|_*$ on symmetric matrices. Decomposing $W^{1/2}MW^{1/2}$ into positive and negative parts and using positivity of $\mathcal{T}$ shows $\|\mathcal{T}(M)\|_{W,1} \leq$

$c_W\|M\|_{W,1}$. Hence $r(\mathcal{T}) \le c_W$ for every $W \succ 0$. Conversely, for any $a > r(\mathcal{T})$ the finite-dimensional Neumann series

$$W^{(a)} = \sum_{j=0}^{\infty} a^{-j}(\mathcal{T}^*)^j(I)$$

converges and is positive definite. It satisfies $\mathcal{T}^*(W^{(a)}) = a(W^{(a)} - I) \prec aW^{(a)}$, so $c_{W^{(a)}} < a$. Let a decrease to $r(\mathcal{T})$. The spectral radius is positive: since $\mathbb{E}[A^{-1}D_i] = I$, positivity gives $\mathcal{T}(M) \succeq M$ for $M \succeq 0$, and in particular $r(\mathcal{T}) \ge 1$. The strict budget equivalence follows by choosing $a < 1/\rho$ and then $t > 1$ sufficiently close to one. □

The original noiseless bound $\phi_A(B_0) \le 1/4$ does not transfer unchanged to an arbitrary weight. If $W \preceq CA$, it gives $\phi_W(B_0) \le C/4$; otherwise the measured quantity $\phi_W(B_0)$ and the size of W^{-1} must be retained in the bound. The complete risk expression is invariant under a positive scalar rescaling of W. A better contraction constant alone need not improve the final numerical risk certificate.

A strict improvement at a quarter deletion. Consider eight bounded feature records: one each of $x_+ = (\sqrt{3}/2, 1/2)$ and $x_- = (\sqrt{3}/2, -1/2)$, and six zero feature vectors. Set $\lambda = 1/16$. Labels can be any bounded values; $y = x^\top e_1$ is also an admissible unit noiseless teacher example. Then

$$\Sigma = \mathrm{diag}(3/16, 1/16), \qquad A = \mathrm{diag}(1/4, 1/8).$$

Under the change of coordinates $N = A^{1/2}MA^{1/2}$, the operator is similar to the self-adjoint map $\mathcal{S}(N) = \mathbb{E}[C_iNC_i]$, where $C_i = A^{-1/2}D_iA^{-1/2}$. Its two active matrices and its zero-feature matrix are, respectively,

$$C_\pm = \begin{pmatrix} 13/4 & \pm\sqrt{6} \\ \pm\sqrt{6} & 5/2 \end{pmatrix}, \qquad C_0 = \mathrm{diag}(1/4, 1/2).$$

The diagonal subspace is invariant, with transfer matrix

$$F = \begin{pmatrix} 43/16 & 3/2 \\ 3/2 & 7/4 \end{pmatrix}. \tag{365}$$

The off-diagonal symmetric subspace has eigenvalue $29/8$. Therefore

$$c_A = \max\{43/16 + 3/2, 7/4 + 3/2\} = 67/16, \qquad r(\mathcal{T}) = \frac{71 + 3\sqrt{281}}{32} < 4.$$

For $W = \mathrm{diag}(1/3, 1/8)$, the whitened dual weight is $A^{-1/2}WA^{-1/2} = \mathrm{diag}(4/3, 1)$. Applying F to its diagonal entries gives the exact constant

$$c_W = \max\{43/16 + 9/8,\ 2 + 7/4\} = 61/16.$$

At $\rho = 1/4$, the original test fails because $\rho c_A = 67/64 > 1$, while the new weight passes because $\rho c_W = 61/64 < 1$. The request budget includes two records, so this is a nonempty legal-deletion regime. Also $A \preceq W \preceq (4/3)A$, so the weight comparison is bounded in this example. These exact calculations strictly improve the sufficient convergence condition. Empirical savings at a prescribed storage or error target are a separate question.

Computational and storage scope. Evaluating $\mathcal{T}^*(W)$ requires the same type of d-by-d matrix operations as (359), applied to $A^{-1}WA^{-1}$ before the outer data operation. A proposed weight can therefore be checked without assembling a d^2-by-d^2 operator. The infimum characterization specifies the optimal value without an efficient method for finding its minimizing weight. The numerical encoder uses $W = A$; optimized dual weights are analyzed only theoretically.

The decoder still needs only its covariance factors and reference. The chosen weight and its search history can remain temporary. If it reports a certificate, a suitable upper bound for the combined omitted response term must be stored and charged as before. The finite-bit utility-only construction above continues to apply whenever the uniform version of (363) meets its stated pre-rounding tolerance. The construction specializes positive-operator and weighted matrix-iteration techniques to a stored deletion summary.

N AN OBSTRUCTION TO A GLOBAL RESPONSE ENVELOPE

We construct a family on which the response envelope in Proposition 26 requires trace larger than a universal multiple of $1/\lambda$. This rules out an arbitrary-budget extension of that matrix certificate. A lower bound for arbitrary summaries requires the additional information construction below.

Proposition 104 (A growing lower bound for response-envelope trace)**.** *For every integer $h \geq 5$, put*

$$m = 2^h, \qquad E = 2m - 2, \qquad d = 2m - 1, \qquad \rho = \frac{h}{E}, \qquad \lambda = \frac{1}{16E(h+1)^2}.$$

There is a centered dataset in $\mathbb{R}^d$, with features of norm at most one and labels in $\{0, -1/2, 1/2\}$, exactly realized by one public unit teacher, for which the following holds. If a positive semidefinite matrix B satisfies

$$(w_{S-U} - w_S)(w_{S-U} - w_S)^\top \preceq \alpha B, \qquad \alpha = \frac{|U|}{|S|}, \tag{366}$$

for every legal deletion with $\alpha \leq \rho$, then

$$\lambda \operatorname{tr} B \geq \frac{(2^h - 1)^2}{64h\, 2^h (h+1)^4} \longrightarrow \infty. \tag{367}$$

All input coordinates belong to the fixed dyadic grid with step $1/2$. The sample size can additionally be chosen in $[4d/\rho, 5d/\rho]$. In particular, no universal constant C makes $\operatorname{tr} B \leq C/\lambda$ possible for every bounded-teacher dataset and every $0 < \rho \leq 1/8$.

Proof. Index the coordinates by the vertices of a complete binary tree of depth h, whose root is o. It has m leaves and E edges. For each edge from a parent p to its child c, include the mirror pair

$$\left(\frac{e_p - e_c}{2}, \frac{(e_p - e_c)^\top e_o}{2}\right), \qquad \left(-\frac{e_p - e_c}{2}, -\frac{(e_p - e_c)^\top e_o}{2}\right).$$

Every feature has norm $1/\sqrt{2}$, every label is in $\{0, -1/2, 1/2\}$, and the common teacher is $v = e_o$. The pairs center both features and labels. Initially take one copy of each pair, giving $n = 2E$. If L denotes the unweighted graph Laplacian, the normalized feature covariance is $L/(4E)$.

For each leaf z, delete the mirror pairs on the h sibling edges adjacent to the root-to-z path: at each nonleaf vertex of the path, delete the edge to the child that is outside the path. Denote this request by U_z. It deletes exactly $2h$ records, so $\alpha = h/E = \rho \leq 1/8$ for $h \geq 5$. The component containing the root in the retained graph is precisely the path of $h+1$ vertices from o to z. Every other retained component has zero teacher coordinates and hence zero ridge optimizer coordinates.

Let $w_z = w_{S-U_z}$ and let L_{path} be the Laplacian of that path, ordered from the root to z. On the path the normalized retained covariance is $L_{\mathrm{path}}/[4(E-h)]$. Consequently the residual vector $e = v - w_z$, restricted to the path, satisfies

$$e = t(L_{\mathrm{path}} + tI)^{-1} e_0, \qquad t = 4(E-h)\lambda \leq \frac{1}{4(h+1)^2}. \tag{368}$$

Its coordinates are nonnegative and sum to one. To see nonnegativity directly, a negative minimum coordinate would make the corresponding coordinate of $(L_{\mathrm{path}} + tI)e$ negative, contradicting its nonnegative right-hand side. Summing the equations gives $\sum_{i=0}^{h} e_i = 1$.

For $1 \leq i \leq h$, summing the equations over the tail of the path gives

$$e_{i-1} - e_i = t \sum_{j=i}^{h} e_j \geq 0.$$

It follows that $e_0 - e_h \leq th$. Since every coordinate is at most $e_h + th$, we obtain

$$a := e_h \geq \frac{1}{h+1} - th \geq \frac{3}{4(h+1)} \geq \frac{1}{2(h+1)}. \tag{369}$$

The path and its normalization are the same for every query, so a does not depend on the chosen leaf. The leaf coordinates of w_z are $-a$ at z and zero at every other leaf.

Every automorphism of the original tree that fixes its root preserves the original dataset and the teacher. Uniqueness of the ridge optimizer therefore implies that all leaf coordinates of w_S equal one scalar c. Let $J : \mathbb{R}^m \to \mathbb{R}^d$ embed the leaf coordinates, and let

$$\Pi = I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^\top, \qquad p_z = \Pi e_z, \qquad B_{\mathrm{leaf}} = J^\top B J.$$

Writing $\Delta_z = w_z - w_S$, the exact leaf response is

$$J^\top \Delta_z = -a e_z - c\mathbf{1}.$$

Applying (366) to Jp_z gives

$$\alpha p_z^\top B_{\mathrm{leaf}} p_z \geq a^2(1 - 1/m)^2.$$

Now $\sum_z p_z p_z^\top = \Pi$. Summing over the m leaves and using $\mathrm{tr}(\Pi B_{\mathrm{leaf}}) \leq \mathrm{tr}\, B$ yields

$$\mathrm{tr}\, B \geq \frac{a^2(m-1)^2}{\alpha m} \geq \frac{E(m-1)^2}{4hm(h+1)^2}.$$

Multiplication by λ proves (367). Its right-hand side grows as $2^h/(64h^5)$, and hence is unbounded.

Finally, duplicating every mirror pair q times leaves all normalized moments, optimizers, and deletion fractions unchanged, provided every request deletes all q copies of each selected pair. Choose

$$q = \left\lceil \frac{2d}{\rho E} \right\rceil.$$

The resulting public sample size obeys

$$\frac{4d}{\rho} \leq n = 2qE < \frac{4d}{\rho} + 2E < \frac{5d}{\rho}.$$

The features and labels already lie exactly on the claimed dyadic grid. □

The unweighted family has $\lambda d = \Theta(h^{-2})$. The next construction assigns different masses to different tree levels and shows that the envelope obstruction is not confined to this dimension-saturated range.

Proposition 105 (An envelope obstruction when the proposed rank is smaller than dimension)**.** *For every integer $h \geq 64$, put*

$$m = 2^h, \qquad d = 2m - 1, \qquad \lambda = \frac{h}{m},$$
$$c_i = \min\left\{ \frac{h^3}{m}, \frac{1}{16h\,2^i} \right\} \quad (1 \leq i \leq h), \qquad \rho = 4\sum_{i=1}^{h} c_i.$$

There is a centered, bounded dataset with the public unit teacher e_o, all input coordinates on the grid of step $1/2$, and a public sample size in $[4d/\rho, 5d/\rho]$, such that

$$2\lambda h^3 \leq \rho \leq 4\lambda h^3 \leq \frac{1}{8}, \qquad \lambda d \longrightarrow \infty.$$

If a positive semidefinite B satisfies (366) *for every legal deletion of at most $\rho|S|$ records, then, with $J = \lceil 2\log_2(32h^2) \rceil$,*

$$\lambda\, \mathrm{tr}\, B \geq \frac{e^{-10}(m-1)^2}{4h^3 m(h+1)^2(1+32h^2)^{2J}} \longrightarrow \infty. \tag{370}$$

Moreover, for any fixed $0 < \kappa \leq c_0$, where c_0 is the constant in (171)*, the choice $\varepsilon = \kappa\rho\lambda$ satisfies that accuracy condition, and the expression R_{J} in* (192) *obeys $R_{\mathrm{J}}/d \to 0$.*

Proof. Use the same complete binary tree, edge features, and teacher as in Proposition 104. For each edge whose child is at depth i, give its two mirror records total empirical mass $4c_i$, split equally between the signs. There are 2^i such edges. Their total mass is at most

$$4\sum_{i=1}^{h} 2^i c_i \le \frac{1}{4}.$$

Fill the remaining mass with the record $(0,0)$. Writing $b_{pc} = e_p - e_c$, the original normalized covariance is

$$\Sigma = \sum_{(p,c)} c_{\mathrm{depth}(c)} b_{pc} b_{pc}^\top.$$

The feature norms, exact labels, and centering are unchanged.

These are exact empirical masses. Initially take $n_0 = 16hm$ records. The multiplicity of each sign of a depth-i edge is $2n_0c_i$; according to which term attains the minimum, it equals either $32h^4$ or 2^{h-i+1}. Both are integers. The remaining zero-record count is therefore a nonnegative integer.

For each leaf z, delete all records on the sibling edges adjacent to its root-to-leaf path, as before. Every query has fraction $\alpha = 4\sum_i c_i = \rho$. For

$$1 \le i \le \lfloor h - 4\log_2 h - 4 \rfloor,$$

the minimum defining c_i is $h^3/m = \lambda h^2$. The upper endpoint is at least $h/2$ when $h \ge 64$. Consequently

$$2\lambda h^3 \le \alpha \le 4\lambda h^3 = \frac{4h^4}{2^h} \le \frac{1}{8}. \tag{371}$$

Put $\beta = 1 - \alpha$.

The root component after deletion is again a path. Let L_c be its weighted Laplacian, with edge conductance c_i at depth i. The retained normal matrix, normalized by the original sample size, is $L_c + \beta\lambda I$. Its residual vector on the path is therefore

$$e = \beta\lambda(L_c + \beta\lambda I)^{-1} e_0.$$

The maximum-principle and summation arguments used above show that $e_i \ge 0$ and $\sum_{i=0}^{h} e_i = 1$. Summing over each tail gives

$$c_i(e_{i-1} - e_i) = \beta\lambda \sum_{j=i}^{h} e_j \ge 0. \tag{372}$$

Thus the coordinates decrease along the path and $e_0 \ge 1/(h+1)$. Using $e_j \le e_i$ for $j \ge i$ in the same identity yields

$$e_{i-1} \le \left(1 + \frac{\beta\lambda(h-i+1)}{c_i}\right) e_i.$$

In particular, the common leaf amplitude $a = e_h$ satisfies

$$a \ge \frac{1}{h+1} \prod_{i=1}^{h} \left(1 + \frac{\beta\lambda(h-i+1)}{c_i}\right)^{-1}. \tag{373}$$

This product loses only $\exp(O((\log h)^2))$. To give an explicit bound, put $K = 32h^2$ and $J = \lceil 2\log_2 K \rceil$. For $j = h - i$, the ratio in (373) is at most

$$\max\left\{\frac{1}{h}, 16h^2(j+1)2^{-j}\right\} \le \max\left\{\frac{1}{h}, K2^{-j/2}\right\},$$

where $j+1 \le 2^{j/2+1}$ for every integer $j \ge 0$. Since $1 + \max\{s,t\} \le (1+s)(1+t)$ for nonnegative s, t,

$$\begin{aligned}\prod_{i=1}^{h} \left(1 + \frac{\beta\lambda(h-i+1)}{c_i}\right) &\le (1+1/h)^h \prod_{j=0}^{\infty} (1 + K2^{-j/2}) \\ &\le e(1+K)^J \exp\left(\sum_{j=J}^{\infty} K2^{-j/2}\right) < e^5(1+K)^J.\end{aligned}$$

The last step uses $K2^{-J/2} \le 1$ and $1/(1-2^{-1/2}) < 4$. It follows that

$$a \ge \frac{e^{-5}}{(h+1)(1+32h^2)^J}. \tag{374}$$

The original weighted tree is still invariant under all root-fixing tree automorphisms. Its leaf coordinates are therefore a common scalar, and the retained optimizer has leaf value $-a$ only at the selected leaf. The leaf projection and summation argument from Proposition 104 applies without change, giving

$$\lambda \operatorname{tr} B \ge \frac{\lambda a^2 (m-1)^2}{\alpha m} \ge \frac{a^2 (m-1)^2}{4h^3 m}.$$

Equation (374) proves (370). Its logarithm is $h \log 2 - O((\log h)^2)$, which tends to infinity.

For the promised sample interval, note that

$$\rho n_0 \le 64h^5 < d \qquad (h \ge 64).$$

Duplicate every record $q = \lceil 4d/(\rho n_0) \rceil$ times, and delete all copies of each selected edge record. Normalized quantities are unchanged, while

$$\frac{4d}{\rho} \le n = q n_0 < \frac{4d}{\rho} + n_0 < \frac{5d}{\rho}.$$

All actual feature and label coordinates remain in $\{0, -1/2, 1/2\}$.

Finally, $\lambda d = 2h - h/2^h \to \infty$. The lower budget bound in (371) implies $\rho \ge \lambda^2$, so $\min\{\rho^2/\lambda, \rho\lambda\} = \rho\lambda$. Thus $\varepsilon = \kappa\rho\lambda$ has the claimed accuracy. The cubic term in the minimum defining R_{J} gives

$$\frac{R_{\mathrm{J}}}{d} \le \frac{\kappa^{-1/3}}{\lambda d} \longrightarrow 0.$$

This conclusion concerns the rank expression itself; it does not assume an upper bound outside the deletion-budget regime of Corollary 28. □

Both constructions disprove a global extension of the response-envelope certificate with trace $O(1/\lambda)$, even when the teacher, original optimizer, and entire dataset are public. The weighted construction shows that the obstruction persists when the proposed joint rank is much smaller than dimension. Neither construction gives an information lower bound: a decoder can reconstruct the public dataset and its retained optimizer. Appendix O goes further: hidden leaf attachments turn related routing geometry into an information lower bound and disprove the global extension of the joint upper rate itself.

O A QUADRATIC LOWER BOUND BEYOND THE SMALL-DELETION REGIME

We encode hidden information in the weighted tree. Attaching a hidden output vector to each leaf makes the common deletion requests reveal different columns of a source matrix. The resulting lower bound applies to arbitrary summaries and rules out extending the finite-bit joint upper rate to every $0 < \rho \le 1/8$.

Theorem 106 (A separation from the proposed global joint upper rate)**.** *There is a universal constant $c > 0$ with the following property. Fix $0 < \kappa \le c_0$, where c_0 is the constant in* (171)*. For every sufficiently large even integer h, define*

$$\begin{gathered} m = 2^h, \qquad d = 4m - 1, \qquad k = \left\lfloor \frac{\sqrt{h}}{64} \right\rfloor, \qquad L = 2^k, \qquad \lambda = \frac{L}{m}, \\ c_i = \min\left\{ 64\lambda, \frac{1}{16h\,2^i} \right\} \quad (1 \le i \le h), \qquad \rho = 4\sum_{i=1}^{h} c_i, \qquad \varepsilon = \kappa\rho\lambda. \end{gathered} \tag{375}$$

There is a finite family with one public sample size $4d/\rho \le n < 5d/\rho$ on which every summary satisfying the pointwise randomized deletion guarantee requires

$$B \ge cd^2 \tag{376}$$

persistent bits. Every feature has norm at most one, every label is in $\{0, -1/2, 1/2\}$*, and all records are exactly realized by the same public unit teacher. All input coordinates lie on the grid in* (170)*. Original and queried retained datasets are centered and have label variance* $\Theta(\lambda)$*, with universal constants. Every original optimizer is the same public nonzero vector. The requests are common legal deletions, each of fraction exactly* ρ*.*

All persistent data-dependent state is charged. The decoder receives only the state and the deleted records, and its success probability need only be $2/3$ *for each fixed dataset and request. Independent public randomness and fresh decoder coins are allowed.*

These parameters satisfy $0 < \lambda \le 1/4$*,* $0 < \rho \le 1/8$*, and* (171)*. Nevertheless, the rank expression* R_{J} *in* (192) *obeys*

$$dR_{\mathrm{J}} \log\left(2 + \frac{d}{\lambda\rho}\right) = o(d^2). \tag{377}$$

Thus the upper rate in that corollary cannot hold for all deletion budgets under the same bounded-teacher access model.

Proof. Use $2m - 1$ tree coordinates indexed by the vertices of a complete binary tree of depth h, followed by m auxiliary coordinates and m output coordinates. Let o be the root, and let $J : \mathbb{R}^m \to \mathbb{R}^{2m-1}$ embed the leaf coordinates. The teacher is $v = (e_o, 0, 0)$.

The source and exact empirical construction. Since h is even, both m and $s = m/4$ are powers of four. Let the source matrix range over $W \in \Gamma_{m,s}$ from Lemma 12, and set

$$A = (W, -W, W, -W) \in \mathbb{R}^{m\times m}.$$

Then

$$W_{ij} \in \{-1/\sqrt{m}, 1/\sqrt{m}\}, \qquad \|W\|_{\mathrm{op}} \le 8, \qquad |\Gamma_{m,s}| \ge 2^{ms-1},$$

and every column of A has norm one. Also $\|A\|_{\mathrm{op}} = 2\|W\|_{\mathrm{op}} \le C_0 := 16$ and $A\mathbf{1} = 0$. Put

$$\tau = 2^{-10}, \qquad t = \frac{1}{64m}, \qquad V = \begin{pmatrix} I_m \\ \tau A \end{pmatrix} \in \mathbb{R}^{2m\times m}.$$

For each tree edge (p, c) whose child is at depth i, include the mirror pair with features $\pm(e_p - e_c, 0, 0)/2$ and their exact teacher labels. Assign that pair total empirical mass $4c_i$, equally split between its two signs. For each leaf j, include the hidden mirror pair

$$\left(\frac{1}{2}\begin{pmatrix} Je_j \\ -Ve_j \end{pmatrix}, 0\right), \qquad \left(-\frac{1}{2}\begin{pmatrix} Je_j \\ -Ve_j \end{pmatrix}, 0\right), \tag{378}$$

with total empirical mass $4t$, equally split between signs. The hidden feature has squared norm $(2 + \tau^2)/4 < 1$. Its label is zero because its root coordinate is zero.

The public tree pairs have total mass at most $4\sum_{i=1}^{h} 2^i c_i \le 1/4$, and the hidden pairs have total mass $4mt = 1/16$. Fill the remaining mass with $(0, 0)$. To realize these masses exactly, initially use

$$n_0 = 16hm.$$

The multiplicity of each sign of a depth-i tree edge is $2n_0c_i$, equal to either $2048hL$ or 2^{h-i+1}. The multiplicity of each sign of a hidden pair is $2n_0t = h/2$. All are integers, and the remaining zero-record count is a nonnegative integer. The sample size and every multiplicity are independent of A.

For a leaf z, delete all records on the sibling edges adjacent to the root-to-z path. These records are public and belong to every source dataset. Every such query deletes a fraction $\alpha = 4\sum_i c_i = \rho$. For all sufficiently large h, at least $h/2$ indices satisfy

$$i \le \lfloor h - k - \log_2 h - 10 \rfloor;$$

on those indices $c_i = 64\lambda$. Consequently

$$128\lambda h \le \rho \le 256\lambda h = \frac{256Lh}{m} \longrightarrow 0. \tag{379}$$

Write $\beta = 1 - \rho$. All original records and all retained records occur in mirror pairs, apart from zero records, so centering remains exact. For large h, $c_1 = 64\lambda$. Only the two root edges carry nonzero labels. Their original label variance is $2c_1 = 128\lambda$; every query removes one of these edges, giving retained label variance $c_1/\beta = 64\lambda/\beta$. These are universal multiples of λ because $\beta \ge 7/8$.

Exact elimination of the hidden output coordinates. Let L_z be the weighted Laplacian on the tree after the public sibling edges for query z are removed. Its surviving depth-i edges have conductance c_i. Let $w_{A,z}$ be the retained ridge optimizer, and write its residual from the teacher as $v - w_{A,z} = (u_A, g_A)$ in tree and non-tree coordinates. Define $\ell_A = J^\top u_A$. With moments normalized by the original sample size, the residual normal equations are

$$\begin{pmatrix} L_z + tJJ^\top + \beta\lambda I & -tJV^\top \\ -tVJ^\top & tVV^\top + \beta\lambda I \end{pmatrix} \begin{pmatrix} u_A \\ g_A \end{pmatrix} = \beta\lambda \begin{pmatrix} e_o \\ 0 \end{pmatrix}.$$

Set

$$r = \frac{t}{\beta\lambda + t} = \frac{1}{64\beta L + 1}, \qquad R_A = t(\beta\lambda I + tV^\top V)^{-1} = r(I + r\tau^2 A^\top A)^{-1}. \tag{380}$$

Eliminating g_A gives the exact identities

$$\begin{aligned} M_A u_A &= \beta\lambda e_o, \qquad & M_A &= L_z + \beta\lambda I + \beta\lambda J R_A J^\top \succeq \beta\lambda I, \\ g_A &= V R_A \ell_A, \qquad & \Pi_{\mathrm{out}} w_{A,z} &= -\tau A R_A \ell_A. \end{aligned} \tag{381}$$

Here Π_{out} selects the last m coordinates. The Schur identity used in the first line is

$$tI - t^2 V^\top (\beta\lambda I + tVV^\top)^{-1} V = \beta\lambda R_A.$$

For analysis, set $A = 0$ in these matrices and denote the corresponding tree residual by u_0. This public reference need not belong to the source family. Then $R_0 = rI$, so different components of the retained tree do not couple. Only the root component has a nonzero teacher coordinate, and this component is the path to z. Therefore

$$J^\top u_0 = a_t e_z \tag{382}$$

for a positive public number a_t that is the same for every query.

We bound this amplitude first. Let L_{path} be the weighted Laplacian on that path without its additional leaf potential, and put

$$e = \beta\lambda(L_{\mathrm{path}} + \beta\lambda I)^{-1} e_0, \qquad a = e_h.$$

The positivity, summation, and tail-balance arguments in (372) give $e_i \geq 0$, $\sum_i e_i = 1$, and

$$c_i(e_{i-1} - e_i) = \beta\lambda \sum_{j=i}^{h} e_j.$$

The coordinates are nonincreasing. If $e_h = 0$, the tail identities would recursively force every preceding coordinate to be zero, contradicting their sum. Thus every e_i is positive. Put $\mu = \beta\lambda$ and define, for $1 \leq i \leq h$,

$$\xi_i = \frac{\mu \sum_{j=i}^{h} e_j}{e_i}.$$

The tail equations give

$$\frac{e_{i-1}}{e_i} = 1 + \frac{\xi_i}{c_i}, \qquad \xi_h = \mu, \qquad \xi_i = \mu + \frac{c_{i+1}\xi_{i+1}}{c_{i+1} + \xi_{i+1}} \quad (i < h).$$

Since $\mu \leq \lambda$ and every $c_i \leq 64\lambda$, backward induction gives $\xi_i \leq 9\lambda$: the induction step is bounded by

$$\lambda + \frac{64\lambda \cdot 9\lambda}{64\lambda + 9\lambda} = \frac{649}{73}\lambda < 9\lambda.$$

For $j = h - i$ and $K = 144hL$, it follows that

$$\frac{e_{i-1}}{e_i} \leq 1 + \frac{9\lambda}{c_i} = \max\left\{\frac{73}{64}, 1 + K2^{-j}\right\} \leq \frac{73}{64}(1 + K2^{-j}).$$

Let $J_h = \lceil \log_2 K \rceil$. Splitting the following product at J_h gives

$$\prod_{j=0}^{\infty}(1 + K2^{-j}) \leq (1+K)^{J_h} \exp\left(\sum_{j=J_h}^{\infty} K2^{-j}\right) \leq e^2(1+K)^{J_h}.$$

The path coordinates decrease, so $e_0 \geq 1/(h+1)$. Multiplying the bounds on consecutive coordinate ratios now yields

$$a \geq \frac{e^{-2}(64/73)^h}{(h+1)(1+144hL)^{J_h}}. \tag{383}$$

The reference matrix on the root path adds only the endpoint potential $\beta\lambda r e_h e_h^\top$. The rank-one inverse identity gives

$$a_t = \frac{a}{1+\beta\lambda r e_h^\top (L_{\mathrm{path}} + \beta\lambda I)^{-1} e_h} \geq \frac{a}{1+r} \geq \frac{a}{2}. \tag{384}$$

The perturbation caused by A is relative to this leaf amplitude. Indeed, subtracting the two tree equations in (381) gives

$$u_A - u_0 = M_A^{-1}\beta\lambda J(rI - R_A)J^\top u_0.$$

Since $\|rI - R_A\|_{\mathrm{op}} \leq r^2\tau^2 C_0^2$, equations (381) and (382) imply

$$\|\ell_A - a_t e_z\|_2 \leq a_t r^2\tau^2 C_0^2. \tag{385}$$

In particular, this estimate does not divide an absolute perturbation bound by a small path amplitude.

Define the public response scale

$$\gamma = \tau r a_t > 0.$$

Equations (381) and (385) give, uniformly over every source and query,

$$\left\| \frac{-\Pi_{\mathrm{out}} w_{A,z}}{\gamma} - A_{:z} \right\|_2 \leq \tau^2 C_0^3 (r + r^2) \leq \frac{1}{128}. \tag{386}$$

For the first term in this bound, use $\|(I + r\tau^2 A^\top A)^{-1} - I\|_{\mathrm{op}} \leq r\tau^2 C_0^2$; for the second use (385). The numerical bound uses $0 < r < 1$, $C_0 = 16$, and $\tau = 2^{-10}$.

The original optimizer is common and public. For the undeleted dataset, apply the same elimination with $\beta = 1$. Write $r_\circ = t/(\lambda + t)$ and $R_{A,\circ} = r_\circ(I + r_\circ\tau^2 A^\top A)^{-1}$, and let L_{all} be the complete weighted tree Laplacian. The public reference tree residual $u_\circ$ solves

$$(L_{\mathrm{all}} + \lambda I + \lambda r_\circ J J^\top)u_\circ = \lambda e_o.$$

Root-fixing tree automorphisms preserve this equation, so $J^\top u_\circ = b\mathbf{1}$ for one public scalar b. The identity $A\mathbf{1} = 0$ implies $R_{A,\circ}\mathbf{1} = r_\circ\mathbf{1}$. Hence

$$(L_{\mathrm{all}} + \lambda I + \lambda J R_{A,\circ} J^\top)u_\circ = \lambda e_o,$$

and uniqueness shows that the original tree residual is exactly $u_\circ$ for every source. Its remaining coordinates are

$$V R_{A,\circ} J^\top u_\circ = r_\circ b \begin{pmatrix} \mathbf{1} \\ \tau A\mathbf{1} \end{pmatrix} = \begin{pmatrix} r_\circ b\mathbf{1} \\ 0 \end{pmatrix}.$$

Thus the complete original optimizer is the same public vector for every source. It is nonzero because the original cross moment Σv is nonzero on the root edges.

The response is larger than the required accuracy. Since $r \geq 1/(65L)$, the preceding amplitude bounds and (379) imply, for a universal $c_* > 0$,

$$\frac{\gamma^2}{\rho} \geq c_* \frac{m}{L^3 h(h+1)^2(73/64)^{2h}(1+144hL)^{2J_h}} \longrightarrow \infty. \tag{387}$$

Here is an explicit check of the limit. Put $q_h = \log_2(144hL) = k + \log_2 h + \log_2 144$. Then $J_h \leq q_h + 1$ and $\log_2(1+144hL) \leq q_h + 1$, so the base-two logarithm of the lower bound in (387) is at least

$$h - 2h\log_2(73/64) - 2q_h^2 - O(k + \log h) = \left(1 - 2\log_2(73/64) - \frac{1}{2048} - o(1)\right) h.$$

The limiting coefficient exceeds $3/5$. Thus, for every fixed $\kappa > 0$ and all sufficiently large h,

$$\varepsilon = \kappa\rho\lambda \leq \frac{\lambda\gamma^2}{1024}. \tag{388}$$

Pointwise randomized answers and a binary entropy bound. Let W be uniform on $\Gamma_{m,s}$, independently of all public randomness, and choose a uniform leaf $Z \in \{1, \dots, s\}$ independently of the source and all encoder randomness. The deletion records are a deterministic public function of Z. At a successful answer to this request, the retained curvature bound $\Sigma_{S-U_Z} + \lambda I \succeq \lambda I$ gives

$$\|\widehat{w} - w_{A,Z}\|_2 \le \sqrt{\varepsilon/\lambda}.$$

Consequently, by (386) and (388),

$$\left\| \frac{-\Pi_{\mathrm{out}}\widehat{w}}{\gamma} - A_{:Z} \right\|_2 \le \frac{1}{32} + \frac{1}{128} = \frac{5}{128}.$$

Predict the signs of the m entries in column Z from the signs of this scaled answer, resolving zero ties by any fixed rule. For these queries $A_{:Z} = W_{:Z}$. Each wrong sign contributes at least $1/m$ to its squared Euclidean error. Hence a successful answer makes fewer than $m/24$ sign errors. A failed answer makes at most m sign errors. The pointwise success probability $2/3$ therefore implies an average sign-error fraction at most $1/3 + 1/24 = 3/8$. No simultaneous-success guarantee is used.

Let M be the complete persistent state and R the independent public randomness. For each source sign, its optimal prediction from (M, R) has error no greater than the prediction just constructed using the corresponding public query and fresh decoder coins. The query is independent of the source and state, and fresh decoder coins carry no additional source information. Thus the average of these ms optimal bit-error probabilities is at most $3/8$. Conditional binary entropy, subadditivity, and concavity of h_2 give

$$H(W \mid \mathsf{M}, \mathsf{R}) \le ms\, h_2(3/8).$$

On the other hand, $H(W \mid \mathsf{R}) = H(W) \ge ms - 1$. For a state of at most B bits, the information bound therefore yields

$$B + 1 \ge I(W; \mathsf{M} \mid \mathsf{R}) \ge [1 - h_2(3/8)]ms - 1. \tag{389}$$

The harmless $+1$ allows arbitrary strings of length at most B; for a fixed-length state it can be omitted. Any retained original model, private seed, or other data-dependent information is part of M and is charged. The readout uses the absolute retained output, and the original optimizer is common and public in any case. Since $s = m/4$ and $d = 4m - 1$, equation (389) proves (376).

Public sample size, finite grid, and separation of rates. For all sufficiently large h,

$$\rho n_0 \le 4096 L h^2 < d.$$

Duplicate every record $q = \lceil 4d/(\rho n_0) \rceil$ times and delete all copies of each selected public edge record. This leaves every normalized quantity unchanged and gives one public sample size

$$\frac{4d}{\rho} \le n = q n_0 < \frac{4d}{\rho} + n_0 < \frac{5d}{\rho}.$$

The only coordinates beyond $0, \pm 1/2$ are the hidden output coordinates $\pm\tau/(2\sqrt{m}) = \pm 2^{-11-h/2}$. They lie exactly on (170), because its precision exceeds $h/2 + 11$ for all sufficiently large h. No precision depending on ρ or ε is added to the prescribed input grid.

Equation (379) implies $\rho \ge \lambda^2$ for large h, so $\min\{\rho^2/\lambda, \rho\lambda\} = \rho\lambda$ and the claimed accuracy condition follows from $\kappa \le c_0$. The parameters $\lambda = L/m$ and $\rho \le 256Lh/m$ both tend to zero. Finally, the cubic term in R_{J} gives

$$R_{\mathrm{J}} \le \frac{\kappa^{-1/3}}{\lambda}, \qquad \log\left(2 + \frac{d}{\lambda\rho}\right) = O(h).$$

Since $L/h \to \infty$, these bounds imply

$$dR_{\mathrm{J}} \log\left(2 + \frac{d}{\lambda\rho}\right) = O_\kappa\left(\frac{m^2 h}{L}\right) = o(d^2),$$

which proves (377). □

The theorem concerns deletion budgets outside the regimes of the proved joint upper bounds: here $\rho/\lambda = \Theta(h)$. Its quadratic bit lower bound applies to a finite family; a matching constant-precision upper bound for all inputs remains separate. For every fixed $a \ge 0$, the same sequence satisfies

$$dR_{\mathrm{J}} \left[\log\left(2 + \frac{d}{\lambda\rho}\right)\right]^a = O_{\kappa,a}\left(\frac{m^2 h^a}{L}\right) = o(d^2).$$

Thus no fixed power of this logarithmic factor absorbs the separation.

P AN EXPONENTIAL DELETION-BUDGET CORRECTION IS NECESSARY

A forest permits many independent query locations without increasing the length of a selected route. Replacing the final binary levels by one star also avoids their cumulative attenuation: the selected star edge is weak, but all of its deleted siblings together have mass only $O(\lambda)$. The construction below combines these two changes with the hidden attachments of Appendix O.

Theorem 107 (A lower bound with exponential budget dependence)**.** *There are universal constants $c, c_1 > 0$ with the following property. Fix $0 < \kappa \le c_0$, where c_0 is the constant in* (171)*. Let q be a sufficiently large even integer, and let H be any even integer satisfying*

$$k = 2\left\lfloor \frac{q}{128} \right\rfloor, \qquad H \ge q + k + 16. \tag{390}$$

Define

$$\begin{aligned} &L = 2^k, \qquad F = 2^{k+16}, \qquad b = F2^q, \qquad m = 2^H, \qquad g = m/b, \\ &d = 3m + \frac{2m}{F} - g, \qquad \lambda = \frac{L}{m}, \qquad \rho = 256\lambda q + \frac{F-1}{4m}, \qquad \varepsilon = \kappa\rho\lambda. \end{aligned} \tag{391}$$

There is a finite family, with one public sample size $4d/\rho \le n < 5d/\rho$, for which every summary satisfying the pointwise randomized deletion guarantee requires at least cd^2 persistent bits. Every feature has norm at most one, every label has absolute value at most one, and every record is exactly realized by the same public unit teacher. All input coordinates belong to the grid in (170)*. Original and queried retained datasets are centered and have label variance $\Theta(\lambda)$, with universal constants. Every original optimizer is the same public nonzero vector. The requests are common legal deletions, each of fraction exactly ρ.*

All persistent data-dependent state is charged. The decoder receives only that state and the deleted records. Success probability $2/3$ for each fixed dataset and request suffices for the lower bound; independent public randomness and fresh decoder coins are allowed. The parameters satisfy $0 < \lambda \le 1/4$, $0 < \rho \le 1/8$, and (171)*. Moreover,*

$$256q \le \frac{\rho}{\lambda} \le 512q, \qquad \frac{d}{R_{\mathrm{J}}} \ge 3\kappa^{1/3}L, \qquad L \ge \frac{1}{4}\exp\left(c_1\frac{\rho}{\lambda}\right), \tag{392}$$

where R_{J} is the rank expression in (192)*. Thus the required number of bits divided by dR_{J} is at least $c\kappa^{1/3}\exp(c_1\rho/\lambda)$, after adjusting the universal constant c.*

Proof. The integers m, g, b are powers of four by (390). Use g disjoint identical routing trees. Each consists of a complete binary tree of depth q, with F new leaves attached to each of its 2^q terminal vertices. Each routing tree has b leaves and $2^{q+1} - 1 + b$ vertices. There are therefore $m + 2m/F - g$ routing coordinates. Add m auxiliary and m output coordinates, giving the dimension in (391). Let $o_1, \dots, o_g$ denote the roots, and let J embed all m leaf coordinates into the routing coordinates. The teacher is

$$v = \left(g^{-1/2}\sum_{a=1}^{g} e_{o_a}, 0, 0\right), \qquad \|v\|_2 = 1.$$

The source and the exact empirical family. Put $s = m/4$ and use the sign family $W \in \Gamma_{m,s}$ of Lemma 12. Partition its columns into g consecutive blocks $W^{(a)}$ of $b/4$ columns, and form

$$A^{(a)} = (W^{(a)}, -W^{(a)}, W^{(a)}, -W^{(a)}), \qquad A = (A^{(1)}, \dots, A^{(g)}) \in \mathbb{R}^{m\times m}.$$

The leaf ordering agrees with these g blocks. If $\mathbf{1}_a$ denotes the indicator of the leaves of routing tree a, then

$$A\mathbf{1}_a = 0 \quad (1 \le a \le g), \qquad \|A\|_{\mathrm{op}} = 2\|W\|_{\mathrm{op}} \le C_0 := 16. \tag{393}$$

Every column of A has norm one, every entry is $\pm 1/\sqrt{m}$, and $|\Gamma_{m,s}| \ge 2^{ms-1}$. As in Appendix O, put

$$\tau = 2^{-10}, \qquad t = \frac{1}{64m}, \qquad V = \begin{pmatrix} I_m \\ \tau A \end{pmatrix}.$$

Give each binary edge conductance $c = 64\lambda$ and each star edge conductance $c_* = 1/(16m)$. For any routing edge (p, u), include the mirror features $\pm(e_p - e_u, 0, 0)/2$, with their exact teacher labels, and give this pair total empirical mass four times its conductance. For each leaf z, include the hidden mirror pair

$$\left(\frac{1}{2}\begin{pmatrix} Je_z \\ -Ve_z \end{pmatrix}, 0\right), \qquad \left(-\frac{1}{2}\begin{pmatrix} Je_z \\ -Ve_z \end{pmatrix}, 0\right),$$

with total mass $4t$. Hidden features have squared norm $(2 + \tau^2)/4 < 1$, and their exact teacher labels are zero. Public edge features have norm $1/\sqrt{2}$ and labels in $\{0, -1/(2\sqrt{g}), 1/(2\sqrt{g})\}$.

The binary, star, and hidden pairs have total masses bounded by

$$256g\lambda(2^{q+1} - 2) \le \frac{512L}{F} = \frac{1}{128}, \qquad 4mc_* = \frac{1}{4}, \qquad 4mt = \frac{1}{16},$$

respectively. Fill the remaining mass with $(0, 0)$. These are exact empirical masses: initially take $n_0 = 32m$ records. The multiplicities of each sign of a binary edge, a star edge, and a hidden pair are, respectively,

$$2n_0c = 4096L, \qquad 2n_0c_* = 4, \qquad 2n_0t = 1.$$

The remaining zero-record count is a nonnegative integer. All multiplicities and the sample size are independent of the source.

For a selected leaf z, delete the q sibling edges beside the binary part of its root-to-leaf route, together with the $F - 1$ other star edges incident to its terminal binary vertex. Delete all records on each of these edges. The deletion records are public, and every source contains them. The fraction deleted is exactly

$$\alpha = 4qc + 4(F - 1)c_* = \rho.$$

The root component of the selected tree is now a path of $q+2$ vertices. Its other components contain no teacher coordinate; all other routing trees remain intact. Write $\beta = 1 - \rho$.

Every original and retained dataset is centered, because records occur in mirror pairs apart from zero records. Only edges incident to the g roots have nonzero labels. The original label variance is $2c = 128\lambda$, and every request removes exactly one of those edges. The retained label variance is therefore

$$\frac{2c - c/g}{\beta} = \frac{128\lambda - 64\lambda/g}{\beta}, \tag{394}$$

which is a universal multiple of λ when $\beta \ge 7/8$.

Eliminating the hidden coordinates. Let L_z denote the Laplacian of the retained routing forest, with the conductances just specified, and write the residual from the teacher as $v - w_{A,z} = (u_A, g_A)$. The moments normalized by the original sample size give

$$\begin{pmatrix} L_z + tJJ^\top + \beta\lambda I & -tJV^\top \\ -tVJ^\top & tVV^\top + \beta\lambda I \end{pmatrix}\begin{pmatrix} u_A \\ g_A \end{pmatrix} = \beta\lambda\begin{pmatrix} g^{-1/2}\sum_{a=1}^{g} e_{o_a} \\ 0 \end{pmatrix}.$$

Set $\ell_A = J^\top u_A$ and

$$r = \frac{t}{\beta\lambda + t} = \frac{1}{64\beta L + 1}, \qquad R_A = t(\beta\lambda I + tV^\top V)^{-1} = r(I + r\tau^2 A^\top A)^{-1}. \tag{395}$$

The same block elimination used in (381) yields the exact equations

$$\begin{aligned} M_A u_A &= \beta\lambda g^{-1/2}\sum_{a=1}^{g} e_{o_a}, \\ M_A &= L_z + \beta\lambda I + \beta\lambda J R_A J^\top \succeq \beta\lambda I, \\ g_A &= VR_A\ell_A, \qquad \Pi_{\mathrm{out}} w_{A,z} = -\tau A R_A \ell_A. \end{aligned} \tag{396}$$

For a public reference, set $A = 0$ in these matrices and denote the corresponding routing residual by u_0. Its leaf potential is $\beta\lambda r I$, so distinct routing components do not couple. If the selected leaf

belongs to tree a, symmetry in every untouched tree and the zero teacher on the detached components give

$$\ell_0 := J^\top u_0 = \frac{a_t}{\sqrt{g}} e_z + \frac{b_\beta}{\sqrt{g}} \sum_{a' \neq a} \mathbf{1}_{a'}. \tag{397}$$

Here $a_t > 0$ is the endpoint amplitude for the selected path with a unit root impulse and the additional leaf potential $\beta\lambda r$, and b_β is a public scalar. Both scalars are independent of the source and of the selected leaf.

The selected path amplitude. Let L_{path} be the Laplacian on the selected path without its additional leaf potential, put $\mu = \beta\lambda$, and define

$$e = \mu(L_{\mathrm{path}} + \mu I)^{-1} e_0, \qquad a = e_{q+1}.$$

The path has q edges of conductance 64λ, followed by one edge of conductance c_*. The maximum principle and summation of its equations show that $e_i \geq 0$ and $\sum_i e_i = 1$. Summing over a tail gives

$$c_i(e_{i-1} - e_i) = \mu \sum_{j=i}^{q+1} e_j \quad (1 \leq i \leq q+1).$$

Thus the coordinates decrease, and $e_0 \geq 1/(q+2)$. They are strictly positive: if the last coordinate were zero, the tail equations would successively force every preceding coordinate to vanish. Define

$$\xi_i = \frac{\mu \sum_{j=i}^{q+1} e_j}{e_i}.$$

The tail identities imply

$$\frac{e_{i-1}}{e_i} = 1 + \frac{\xi_i}{c_i}, \qquad \xi_{q+1} = \mu, \qquad \xi_i = \mu + \frac{c_{i+1}\xi_{i+1}}{c_{i+1} + \xi_{i+1}}.$$

Every conductance is at most 64λ. Backward induction therefore gives $\xi_i \leq 9\lambda$, since

$$\lambda + \frac{64\lambda \cdot 9\lambda}{64\lambda + 9\lambda} = \frac{649}{73}\lambda < 9\lambda.$$

Each of the first q coordinate ratios is consequently at most $73/64$. The final ratio obeys the sharper exact bound

$$\frac{e_q}{e_{q+1}} = 1 + \frac{\mu}{c_*} = 1 + 16\beta L \leq 17L.$$

Multiplying these bounds yields

$$a \geq \frac{(64/73)^q}{17L(q+2)}. \tag{398}$$

The rank-one inverse identity for the additional endpoint potential now gives

$$a_t = \frac{a}{1 + \mu r\, e_{q+1}^\top (L_{\mathrm{path}} + \mu I)^{-1} e_{q+1}} \geq \frac{a}{1+r} \geq \frac{a}{2}.$$

Cancellation of the untouched trees and column readout. Equation (393) implies

$$(rI - R_A)\mathbf{1}_{a'} = 0, \qquad A R_A \mathbf{1}_{a'} = 0 \quad (1 \leq a' \leq g).$$

This cancels the second term of (397) in both the perturbation and the output. Subtracting the two routing equations in (396) gives

$$u_A - u_0 = M_A^{-1} \beta\lambda J (rI - R_A)\ell_0.$$

Since $\|rI - R_A\|_{\mathrm{op}} \leq r^2\tau^2 C_0^2$ and $M_A \succeq \beta\lambda I$, it follows that

$$\|\ell_A - \ell_0\|_2 \leq \frac{a_t}{\sqrt{g}} r^2 \tau^2 C_0^2. \tag{399}$$

Define the public scale

$$\gamma = \frac{\tau r a_t}{\sqrt{g}} > 0.$$

Using (397), the cancellation above, and (399) in the output equation gives

$$\left\| \frac{-\Pi_{\mathrm{out}} w_{A,z}}{\gamma} - A_{:z} \right\|_2 \leq \tau^2 C_0^3 (r + r^2) \leq \frac{1}{128}. \tag{400}$$

Indeed, the reference output contributes $\|A[(I + r\tau^2 A^\top A)^{-1} - I]e_z\|_2 \leq r\tau^2 C_0^3$, and the perturbation contributes at most $r^2\tau^2 C_0^3$. The numerical bound uses $r < 1$, $C_0 = 16$, and $\tau = 2^{-10}$.

The original optimizer is common. Apply the same elimination to the undeleted dataset, with $\beta = 1$. Let $r_\circ = t/(\lambda + t)$, let $R_{A,\circ} = r_\circ(I + r_\circ \tau^2 A^\top A)^{-1}$, and let L_{all} be the complete routing forest Laplacian. The public reference residual solves

$$(L_{\mathrm{all}} + \lambda I + \lambda r_\circ J J^\top) u_\circ = \lambda g^{-1/2} \sum_{a=1}^{g} e_{o_a}.$$

The identical trees and their root-fixing automorphisms make all leaf coordinates equal, say $J^\top u_\circ = b_\circ \mathbf{1}$. Since $A\mathbf{1} = 0$, one has $R_{A,\circ}\mathbf{1} = r_\circ \mathbf{1}$. Hence $u_\circ$ also solves the exact routing equation for every source. Uniqueness shows that the original routing residual is independent of A. Its other coordinates are

$$V R_{A,\circ} J^\top u_\circ = r_\circ b_\circ \begin{pmatrix} \mathbf{1} \\ \tau A \mathbf{1} \end{pmatrix} = \begin{pmatrix} r_\circ b_\circ \mathbf{1} \\ 0 \end{pmatrix}.$$

Thus every original optimizer is the same public vector. It is nonzero, because its original cross moment Σv is nonzero on the root edges.

Signal strength and the information bound. For $q \geq 64$, the parameter identities give

$$256\lambda q \leq \rho = \lambda \left(256q + 2^{14} - \frac{1}{4L} \right) \leq 512\lambda q. \tag{401}$$

Also $r \geq 1/(65L)$. The preceding amplitude estimates therefore imply, for a universal $c_* > 0$,

$$\frac{\gamma^2}{\rho} \geq c_* \frac{2^q (64/73)^{2q}}{L^4 q (q+2)^2} \longrightarrow \infty. \tag{402}$$

For example, before substituting $F = 2^{16} L$, the lower bound is a universal constant times $F 2^q (64/73)^{2q} / [L^5 q (q+2)^2]$; this follows by using $m/g = F 2^q$ in γ^2/ρ. Since $k \leq q/64$, the base-two logarithm of the displayed bound is at least

$$\left(1 - 2\log_2(73/64) - \frac{1}{16} \right) q - O(\log q).$$

Its linear coefficient is positive. In particular, for all sufficiently large q,

$$\varepsilon = \kappa \rho \lambda \leq \frac{\lambda \gamma^2}{1024}. \tag{403}$$

This estimate is uniform over every permitted H.

Let W be uniform on $\Gamma_{m,s}$, independently of all public randomness. Query a uniformly selected one of the s leaves corresponding to the first copy of each source column in the blocks $A^{(a)}$. These deleted records are a deterministic public function of the selected source-column index. At a successful answer, the retained curvature lower bound $\Sigma_{S-U} + \lambda I \succeq \lambda I$ gives

$$\|\widehat{w} - w_{A,z}\|_2 \leq \sqrt{\varepsilon/\lambda}.$$

Consequently, (400) and (403) show that the scaled output is within $5/128$ in Euclidean norm of the selected source column. Predict its m signs from that output, resolving zero ties by a fixed rule. Every wrong sign contributes at least $1/m$ to squared error. A successful answer therefore makes fewer than $m/24$ sign errors. A failed answer makes at most m errors, so the pointwise success probability $2/3$ bounds the average error fraction by $1/3 + 1/24 = 3/8$.

Write M for the complete persistent state and R for independent public randomness. For each fixed source-bit index, hardwire its column's public query into its predictor; the uniform query above is only an averaging device. The optimal prediction of that bit from (M, R) has error no greater than this decoder-based prediction. Fresh decoder coins carry no source information. Conditional binary entropy, subadditivity, and concavity of h_2 therefore give

$$H(W \mid \mathsf{M}, \mathsf{R}) \leq ms\, h_2(3/8), \qquad H(W \mid \mathsf{R}) \geq ms - 1.$$

If the state has at most B bits, then

$$B + 1 \geq I(W; \mathsf{M} \mid \mathsf{R}) \geq [1 - h_2(3/8)] ms - 1.$$

The $+1$ permits arbitrary strings of length at most B. This charges every retained model, private seed, and other data-dependent state. Since $s = m/4$ and $3m < d < 4m$, the required number of bits is at least cd^2, for a universal $c > 0$.

Sample size, grid, and the quantitative separation. The bound $m \geq F2^q = 2^{16}L2^q$ implies $\lambda \leq 2^{-q-16}$. Together with (401), this proves $\lambda \leq 1/4$ and $\rho \leq 1/8$ for all sufficiently large q. It also gives

$$\rho n_0 \leq 16384Lq < d.$$

Duplicate every record $a_0 = \lceil 4d/(\rho n_0) \rceil$ times and delete all copies of each selected edge record. Normalized quantities are unchanged, and the public sample size satisfies

$$\frac{4d}{\rho} \leq n = a_0 n_0 < \frac{4d}{\rho} + n_0 < \frac{5d}{\rho}.$$

The nonzero feature coordinates are $\pm 1/2$ and $\pm 2^{-11-H/2}$. The nonzero labels are $\pm 2^{-1-(H-q-k-16)/2}$. All their exponents are integers, and the precision in (170) exceeds $H/2+11$. Thus the specified input grid suffices without any additional precision depending on ρ or ε.

Since $\rho \geq \lambda^2$, one has $\min\{\rho^2/\lambda, \rho\lambda\} = \rho\lambda$. The accuracy condition follows from $\kappa \leq c_0$. Finally, the cubic term in R_{J} gives

$$R_{\mathrm{J}} \leq \frac{\kappa^{-1/3}}{\lambda}, \qquad \frac{d}{R_{\mathrm{J}}} \geq \kappa^{1/3}\lambda d \geq 3\kappa^{1/3}L.$$

As $k \geq q/64 - 2$ and $\rho/\lambda \leq 512q$,

$$L \geq \frac{1}{4}\, 2^{q/64} \geq \frac{1}{4} \exp\left(\frac{\log 2}{32768} \frac{\rho}{\lambda} \right).$$

These inequalities prove (392) and the stated quantitative lower bound. □

Corollary 108 (Separation beyond every fixed precision logarithm)**.** *Along any sequence of parameters in Theorem 107 satisfying*

$$q \longrightarrow \infty, \qquad \log H = o(q), \tag{404}$$

every fixed $a \geq 0$ obeys

$$dR_{\mathrm{J}} \left[\log\left(2 + \frac{d}{\lambda\rho} \right) \right]^a = o(d^2).$$

Such sequences can have $\rho/\lambda = o(\log d)$, with $(\rho/\lambda)/\log\log d$ diverging arbitrarily slowly.

Proof. The logarithm in the statement is $O(H)$. The rank bound above therefore gives

$$\frac{dR_{\mathrm{J}}[\log(2 + d/(\lambda\rho))]^a}{d^2} = O_{\kappa,a}\left(\frac{H^a}{L} \right) = o(1),$$

because $\log H = o(q)$ and $\log L = \Theta(q)$. The constraint $q + k + 16 \leq H$ permits $\log H \ll q \ll H$, and permits $q/\log H$ to diverge arbitrarily slowly. In this range, $\log d = \Theta(H)$, $\log\log d = \Theta(\log H)$, and $\rho/\lambda = \Theta(q)$, proving the final claim. □

At these accuracy values, the lower bound exhibits an exponential cost in ρ/λ. The scale $\rho = \Theta(\lambda \log\log d)$ is a crossover up to polylogarithmic factors, not a universal constant threshold: when ρ/λ is a fixed multiple of $\log\log d$, the exponential correction is itself a fixed power of the precision logarithm.

Q Compression from a conditioned teacher subspace

A conditioned subspace carrying the teacher permits compression under the assumptions below, even when response-envelope iterations are unstable. The condition below controls the original covariance across that subspace and its complement; individual records may mix both. No condition is imposed on the spectrum or dimension of the complementary covariance.

Theorem 109 (A robust teacher-subspace bound)**.** *Let $0 < \lambda \le 1$ and $0 < \rho < 1$. Suppose that $\|x_i\| \le 1$ and $y_i = x_i^\top v$ for a vector $\|v\| \le 1$. Write $\Sigma = n^{-1}\sum_i x_i x_i^\top$ and $b = n^{-1}\sum_i x_i y_i$. Let P be an orthogonal projector of rank $1 \le k \le d$, set $Q = I - P$, and suppose*

$$a = \lambda_{\min}(P\Sigma P|_{\operatorname{range} P}) > \rho. \qquad \eta = \|P\Sigma Q\|_{\mathrm{op}}, \qquad e^2 = (Qv)^\top \Sigma (Qv).$$

The decoder that minimizes the retained ridge objective over $\operatorname{range} P$ *needs only a basis for P and the projected original covariance and cross moment. For every deletion of fraction $\alpha \le \rho$, with $\beta = 1 - \alpha$, its output $\widetilde{w}_U$ satisfies*

$$\sqrt{F_{S-U}(\widetilde{w}_U) - F_{S-U}(w_{S-U})} \le \frac{\sqrt{\lambda}(\eta + \alpha/2)\|Pv\|}{a - \alpha + \beta\lambda} + \frac{e}{\sqrt{\beta}}. \tag{405}$$

In particular, a uniform upper bound is obtained by replacing α by ρ and β by $1 - \rho$ on the right.

The exact representation has $O(dk + k^2 + k)$ real entries. If the last uniform bound is at most $\sqrt{\varepsilon}/2$, where $0 < \varepsilon \le 1$, there is a deterministic finite-bit version with excess at most ε and storage

$$O\left(d(k+1)\log\left(2 + \frac{d}{(1-\rho)\lambda\sqrt{\varepsilon}}\right) + \log(2+n)\right). \tag{406}$$

All persistent data-dependent state, including a data-dependent basis, is charged. The decoder receives only this state and the deleted records. The teacher and the displayed structural quantities are not decoder inputs.

Proof. Fix an orthonormal basis $V \in \mathbb{R}^{d\times k}$ for P. Store V, $C_0 = V^\top \Sigma V$, $f_0 = V^\top b$, and n if it is not public. On receiving U, form

$$C_U = \frac{1}{n}\sum_{i\in U}(V^\top x_i)(V^\top x_i)^\top, \qquad f_U = \frac{1}{n}\sum_{i\in U} V^\top x_i y_i.$$

The exact answer is $\widetilde{w}_U = V(C_0 - C_U + \beta\lambda I)^{-1}(f_0 - f_U)$. This uses no retained records.

For the error calculation, work in orthonormal coordinates for P and Q, and write the retained covariance and teacher as

$$\Sigma_{S-U} = \begin{pmatrix} C & E \\ E^\top & D \end{pmatrix}, \qquad v = \begin{pmatrix} v_P \\ v_Q \end{pmatrix}.$$

The first block of the constrained answer is $u = (C + \lambda I)^{-1}(Cv_P + Ev_Q)$; its second block is zero. The first block of its normal-equation residual vanishes. Its second block is

$$g = E^\top(u - v_P) - Dv_Q = -\lambda E^\top (C + \lambda I)^{-1} v_P - Kv_Q, \qquad K = D - E^\top (C + \lambda I)^{-1} E \succeq 0. \tag{407}$$

Positivity of K follows from the Schur complement of $\Sigma_{S-U} + \operatorname{diag}(\lambda I, 0) \succeq 0$. The bottom block of $(\Sigma_{S-U} + \lambda I)^{-1}$ is $(K + \lambda I)^{-1}$. The quadratic excess identity therefore gives

$$F_{S-U}(\widetilde{w}_U) - F_{S-U}(w_{S-U}) = g^\top (K + \lambda I)^{-1} g. \tag{408}$$

The deleted covariance has operator norm at most α. Each record also satisfies $\|Px_i\|\|Qx_i\| \le \|x_i\|^2/2 \le 1/2$. Consequently

$$C \succeq \frac{a - \alpha}{\beta} I, \qquad \|E\|_{\mathrm{op}} \le \frac{\eta + \alpha/2}{\beta}.$$

The first term of (407), measured in the norm induced by $(K + \lambda I)^{-1}$, is at most

$$\frac{\sqrt{\lambda}\|E\|_{\mathrm{op}}\|v_P\|}{\lambda_{\min}(C) + \lambda} \le \frac{\sqrt{\lambda}(\eta + \alpha/2)\|Pv\|}{a - \alpha + \beta\lambda}.$$

For the second term, spectral calculus and $0 \preceq K \preceq D$ give $K(K + \lambda I)^{-1} K \preceq K \preceq D$. Also $v_Q^\top D v_Q \le e^2/\beta$, because deletion only removes positive-semidefinite covariance. Applying the triangle inequality in (408) proves the risk bound.

We give finite-precision details to distinguish the real-entry representation from the bit bound. Put $\beta_0 = 1 - \rho$. Round each entry of V to a public dyadic grid with step at most

$$h_V = \frac{\sqrt{\lambda\varepsilon}}{128\sqrt{dk}},$$

and denote the resulting matrix by V'. Its operator error is at most $\sqrt{\lambda\varepsilon}/128$, so it has full column rank. The projector P' onto its columns satisfies $\|P' - P\|_{\mathrm{op}} \le \sqrt{\lambda\varepsilon}/32$. For example, the bound follows by projecting V' onto the complement of P and using $\sigma_{\min}(V') \ge 1 - \|V' - V\|_{\mathrm{op}}$; the two projectors have the same rank. Encoder and decoder use the same deterministic orthonormalization $B = V'(V'^{\top}V')^{-1/2}$, which requires no extra persistent state.

Every constrained ridge optimizer for bounded labels has norm at most $1/(2\sqrt{\lambda})$, by applying the ridge singular-value bound to the projected features. Thus the candidate $P'\widetilde{w}_U$ differs from $\widetilde{w}_U$ by at most $\sqrt{\varepsilon}/64$. Since $\|\Sigma_{S-U} + \lambda I\|_{\mathrm{op}} \le 2$, the excess square root of this candidate increases by at most $\sqrt{2\varepsilon}/64$. The exact minimizer over $\mathrm{range}\, P'$ has no larger objective.

Compute the original moments $C'_0 = B^{\top}\Sigma B$ and $f'_0 = B^{\top}b$ at encoding time, and round their entries with a public dyadic step at most

$$h_M = \frac{\beta_0\lambda^{3/2}\sqrt{\varepsilon}}{512k},$$

preserving symmetry. Their errors E_C, e_f satisfy $\|E_C\|_{\mathrm{op}} \le kh_M$ and $\|e_f\| \le \sqrt{k}h_M$. Deleted moments are evaluated in the same basis B. The unrounded retained normal matrix in these coordinates is at least $\beta\lambda I$, and its exact solution has norm at most $1/(2\sqrt{\lambda})$. The rounded matrix remains positive definite, and the solution changes by at most

$$\frac{2}{\beta\lambda}\left(\|e_f\| + \frac{\|E_C\|_{\mathrm{op}}}{2\sqrt{\lambda}}\right) \le \frac{3\sqrt{\varepsilon}}{512}.$$

Its excess square root increases by at most another $3\sqrt{2\varepsilon}/512$. Together with the assumed $\sqrt{\varepsilon}/2$ bound, these errors are smaller than $\sqrt{\varepsilon}$. Orthonormalization and solves may be evaluated to any additional required output precision; these are temporary computations, not extra persistent information.

The stored basis and moment entries are bounded by universal constants. Choose each public dyadic step within a factor two below the displayed value. Counting their precision, $k \le d$, and the sample-size metadata gives (406). Neither the teacher nor a structural certificate is stored by this utility-only construction. □

Corollary 110 (A global regime for a conditioned signal subspace)**.** *Suppose that P is a rank-k invariant subspace of Σ, $Pv = v$, and $P\Sigma P \succeq aP$ for $a > 0$. If $\rho \le a/2$, the exact projected decoder has excess at most*

$$\frac{\lambda\rho^2}{a^2}. \tag{409}$$

In particular, for $0 < \kappa \le 1$, $\varepsilon = \kappa\rho\lambda$, and $\rho \le \min\{a/2, \kappa a^2/4\}$, the finite-bit bound (406) *applies. This budget condition is independent of λ. Also $k \le 1/a$ because $\mathrm{tr}\,\Sigma \le 1$.*

Proof. Here $\eta = e = 0$ and $\|Pv\| \le 1$. In (405), use $a - \rho \ge a/2$. The final budget restriction makes the squared bound at most $\varepsilon/4$. □

Sharpness for the projected decoder. The scaling in (409) cannot be improved uniformly for this decoder, even in two dimensions. Fix $0 < a \le 1/4$, $0 < \alpha \le a/2$, set $b_* = a - \alpha/2$ and $\beta = 1 - \alpha$, and use the teacher $v = e_1$. Let the deleted records have feature $(e_1 - e_2)/\sqrt{2}$ and total mass α. Choose the retained, originally normalized covariance to be

$$M_R = \begin{pmatrix} b_* & \alpha/2 \\ \alpha/2 & \alpha^2/(4b_*) \end{pmatrix}.$$

This rank-one matrix is realized by one feature direction of norm one with mass $\operatorname{tr} M_R$ and zero records for the remaining retained mass. Indeed, $\operatorname{tr} M_R \le a \le 1-\alpha$. The full covariance is

$$\Sigma = \operatorname{diag}\left(a, \frac{a\alpha}{2a-\alpha}\right).$$

Thus $P = e_1 e_1^\top$ is a spectral subspace containing the teacher, with eigenvalue a. Direct substitution into the Schur identity gives the exact projected-decoder excess

$$\frac{\lambda\alpha^2}{4(b_* + \beta\lambda)(b_* + \alpha^2/(4b_*) + \beta\lambda)}. \tag{410}$$

For $0 < \lambda \le a/2$, both factors in the denominator are at most $3a/2$, so this is at least $\lambda\alpha^2/(9a^2)$. All labels are exact bounded teacher labels. Mirror pairs give centered versions without changing any covariance or calculation. Rational masses give exact finite datasets after duplication. This establishes sharpness of the stated projection rule; it is not a lower bound against arbitrary summaries.

A bypass of every positive weighted-trace convergence test. The theorem is not implied by restricting the response operator to a teacher-reachable invariant subspace. Fix $0 < a \le 1/4$ and take $0 < \lambda \le 1/4$. Assign total mass a to feature e_1, total mass λ equally to $x_+ = (e_1 + e_2)/\sqrt{2}$ and $x_- = (e_1 - e_2)/\sqrt{2}$, and fill the remaining mass with zero records. Again $v = e_1$ and

$$\Sigma = \operatorname{diag}(a + \lambda/2, \lambda/2).$$

The subspace $P = e_1 e_1^\top$ satisfies the corollary with lower bound a, although the mixed records couple its teacher and nuisance coordinates. For any fixed positive $\rho \le \min\{a/2, \kappa a^2/4\}$, its finite-bit projected summary therefore meets $\varepsilon = \kappa\rho\lambda$ as λ tends to zero.

In contrast, let $\mathcal{T}$ be the operator in (348). Its whitened version $N \mapsto \mathbb{E}[C_i N C_i]$, with $C_i = (\Sigma + \lambda I)^{-1/2}(x_i x_i^\top + \lambda I)(\Sigma + \lambda I)^{-1/2}$, is self-adjoint and has the same spectrum. The Frobenius Rayleigh quotient of $e_2 e_2^\top$ gives

$$r(\mathcal{T}) \ge \frac{\mathbb{E}[(x_{i2}^2 + \lambda)^2]}{(3\lambda/2)^2} = \frac{1}{9\lambda} + \frac{8}{9}. \tag{411}$$

Thus $\rho r(\mathcal{T}) > 1$ for small λ, excluding every positive weighted-trace test in Proposition 103. Nor is the bad direction absent from the initial forcing: the two mixed records have opposite nonzero second components in q_i, while zero records have $q_i = -\lambda\theta \ne 0$ in the first coordinate. Hence $B_0 \succ 0$ and its supporting face is already the full positive-semidefinite cone. Mirror pairs and rational masses again give centered exact finite datasets. The example separates the structural decoder from the envelope convergence tests. Its conclusion is algorithmic, not a worst-case memory lower bound.

R A NOISELESS LOWER BOUND WITH A COMMON NONZERO MODEL

This appendix strengthens the lower bound to exactly realizable linear responses. The first theorem is an intermediate construction; Corollary 113 additionally gives a bounded teacher and nonvanishing label variance. The input-precision statement is different from the fixed-grid lower bound: the grid below uses $O(\log d + \log(1/\lambda))$ bits per coordinate. This is $O(\log d)$ bits when $\lambda \ge d^{-O(1)}$; smaller λ retains the additional precision dependence.

Throughout this appendix,

$$F_T(w) = \frac{1}{|T|}\sum_{(x,y)\in T} (y - x^\top w)^2 + \lambda\|w\|_2^2.$$

All persistent data-dependent state is charged to the summary; the deletion decoder receives only that summary and the deleted records. Dimensions, regularization, accuracy, and the construction's fixed parameters are public.

Theorem 111 (Noiseless, pointwise randomized lower bound)**.** *There are absolute constants $c > 0$ and d_0 with the following property. For every $d \ge d_0$, $0 < \lambda \le 1$, $0 < \rho \le 1/4$, and $0 < \varepsilon \le c\rho^2$, there is a finite family of datasets in $\mathbb{R}^d \times \{-1, 1\}$, all of one public size $n = \Theta(d/\rho)$, and a public family of deletion multisets legal for every dataset, such that:*

1. *Every record satisfies* $\|x\|_2 \le 1$ *and* $y = x_1/a$*, for a common public dyadic* $a > 0$ *satisfying*
$$\lambda/4 \le a^2 \le \min\{\lambda, 1/4\}.$$
The coordinates belong to a common dyadic grid with $O(\log d + \log(1/\lambda))$ *bits per coordinate.*

2. *Every original dataset, and every retained dataset in the query family, has zero feature mean and zero label mean. Every original ridge minimizer is the same nonzero public vector*
$$w_S = \frac{a}{a^2+\lambda} e_1.$$

3. *Every query deletes an* α *fraction of the records, where* $\rho/2 \le \alpha \le \rho$*.*

4. *If a possibly randomized offline summary of at most* B *bits permits a possibly randomized decoder to satisfy*
$$\Pr\big[F_{S\setminus U}(\widehat{w}) - F_{S\setminus U}(w_{S\setminus U}) \le \varepsilon\big] \ge \frac{2}{3}$$
for each fixed dataset S *and each fixed query* U *in the family, then*
$$B = \Omega\left(d \min\left\{d, \frac{\rho^{2/3}}{\lambda \varepsilon^{1/3}}\right\}\right).$$
The probability may include both training and decoder randomness, including shared public randomness independent of the dataset.

For fixed ρ*, this is* $\Omega(d \min\{d, 1/(\lambda\varepsilon^{1/3})\})$ *bits with* $n = \Theta(d)$*. It requires no simultaneous-success guarantee over queries.*

The public noiseless teacher is $a^{-1}e_1$, of norm $\Theta(\lambda^{-1/2})$. The common trained predictor has $x^\top w_S = y\, a^2/(a^2+\lambda)$, with the multiplier in $[1/5, 1/2]$. Thus the construction has a nonzero predictor and constant ridge shrinkage; it is not a bounded-teacher-norm statement uniform in λ. When $\lambda \le 1/4$ is a dyadic square, we can take $a = \sqrt{\lambda}$, so that $w_S = e_1/(2\sqrt{\lambda})$. Reproducing the common original model requires no instance-dependent bits, to any specified numerical accuracy.

R.1 Finite matrix families

We use the following elementary finite packings. There are absolute constants $c_0, c_1 > 0$ and an operator-norm bound $C_0 = 16$ such that:

1. For every sufficiently large power of four p, there is a family $\mathcal{Z}_p$ of symmetric matrices with zero diagonal and off-diagonal entries in $\{-1/\sqrt{p}, 1/\sqrt{p}\}$ satisfying
$$|\mathcal{Z}_p| \ge \exp(c_0 p^2), \qquad \|Z\|_{\mathrm{op}} \le C_0, \qquad \|Z - Z'\|_F \ge c_1\sqrt{p} \quad (Z \ne Z').$$

2. If $m \ge r$ are powers of four, there is a family $\mathcal{R}_{m,r} \subset \{-1/\sqrt{m}, 1/\sqrt{m}\}^{m\times r}$ with
$$|\mathcal{R}_{m,r}| \ge \exp(c_0 mr), \qquad \|R\|_{\mathrm{op}} \le C_0, \qquad \|R - R'\|_F \ge c_1\sqrt{r} \quad (R \ne R').$$
The constants may be reduced to cover the finitely many small dimensions used below.

Here is a counting proof. For a random rectangular sign matrix normalized by $\sqrt{m}$, the scalar $u^\top R v$, for fixed unit vectors u, v, has tail at most $2\exp(-mt^2/2)$. A pair of $1/4$-nets, each of size at most 9 to the relevant dimension, bounds the operator norm by twice the maximum bilinear form on the nets. Since $m \ge r$, at least half the sign matrices have operator norm at most 16. For a symmetric sign matrix normalized by $\sqrt{p}$, the quadratic form $u^\top Z u$ has tail at most $2\exp(-pt^2/4)$; one such net gives the same constant-fraction conclusion. In either case, greedily pack this constant fraction of the binary cube at Hamming distance at least one eighth of its dimension. A Hamming ball has at most $2^{h_2(1/8)N}$ points in dimension N, whereas the retained set has at least 2^{N-1} points. This gives exponentially many matrices. Hamming separation translates to the stated Frobenius separation: each changed rectangular entry contributes $4/m$, and each changed symmetric off-diagonal sign contributes $8/p$, to squared Frobenius distance.

R.2 Common samples and legal deletion packets

Let

$$L = \left\lceil \frac{1}{2\rho} \right\rceil, \qquad \alpha = \frac{1}{2L}, \qquad \beta = 1 - \alpha, \qquad q = 1 - 2\alpha.$$

Then $\rho/2 \le \alpha \le \rho$, $\alpha \le 1/4$, and $q \ge 1/2$. Take a to be the largest power of two not exceeding $\min\{\sqrt{\lambda}, 1/2\}$. It has the bounds in the theorem. Let $v = b = 1/2$.

An *effective* nuisance vector z denotes the two actual records

$$((a, z), 1), \qquad ((-a, -z), -1). \tag{412}$$

Unused coordinates are padded with zero. Every such pair is exactly noiseless for $a^{-1}e_1$, and has zero feature and label means.

For k probe coordinates, include the effective anchor vectors $+ve_j$ and $-ve_j$, once each, for every $j \le k$. Also include effective hidden vectors $+z_j$ and $-z_j$, each repeated $L - 1$ times, for every $j \le k$. Thus the base sample has exactly

$$n_0 = 4Lk$$

actual records. Write P for the orthogonal projector onto the probe coordinates and J for their isometric inclusion into the nuisance space, so that $P = JJ^\top$. The nuisance covariance is

$$H = 2C + H_{\text{hid}}, \qquad C = \frac{\alpha v^2}{k} P, \qquad H_{\text{hid}} = \frac{q}{k} \sum_{j=1}^{k} z_j z_j^\top.$$

The effective nuisance mean is zero. Consequently the normalized full statistics are

$$G_S = \operatorname{diag}(a^2, H), \qquad b_S = (a, 0),$$

which proves the common original minimizer.

For $s \in \{-1, 1\}^k$, delete the two actual records representing effective anchor $s_j ve_j$, for each $j \le k$. This deletes exactly $2k = \alpha n_0$ records. Every query consists of public anchors common to all hidden instances. It therefore communicates no hidden matrix information. The retained records still occur in actual mirror pairs, so their feature and label means remain zero.

We eventually duplicate the entire base dataset $H_* = \lceil d/k \rceil$ times and delete all H_* copies of each selected anchor record. This changes neither normalized statistics nor deletion fraction. The resulting sample size obeys

$$n = 4Lk\lceil d/k \rceil = \Theta(d/\rho).$$

All formulas below can therefore be derived on the base sample.

R.3 Exact readout from a retained minimizer

Set

$$t_s = \frac{\alpha v}{k} Js, \qquad K = H - C + \beta\lambda I.$$

For the retained minimizer $w_T = (u, h)$, the normal equations are

$$(a^2 + \lambda)u - \frac{a}{\beta} t_s^\top h = a, \qquad Kh = (au - 1)t_s. \tag{413}$$

In particular, if $r_s = au - 1$, then

$$h = r_s K^{-1} t_s = \alpha v r_s W s, \qquad W = (kK)^{-1} J. \tag{414}$$

Indeed, the retained statistics are

$$G_T = \begin{pmatrix} a^2 & -at_s^\top/\beta \\ -at_s/\beta & (H - C)/\beta \end{pmatrix}, \qquad b_T = (a, -t_s/\beta),$$

which gives (413) directly.

Let $\theta = \lambda/(a^2+\lambda)$ and $\phi = 1-\theta$. Substituting (414) into the first normal equation yields

$$r_s = -\frac{\theta}{1-\phi\, t_s^\top K^{-1} t_s/\beta}. \tag{415}$$

Since $H \succeq 2C$, we have $K \succeq C + \beta\lambda I$ and

$$t_s^\top K^{-1} t_s \le \frac{\alpha^2 v^2/k}{\alpha v^2/k + \beta\lambda} \le \alpha.$$

Now $\theta \in [1/2, 4/5]$, $\phi \le 1/2$, and $t_s^\top K^{-1} t_s/\beta \le \alpha/\beta \le 1/3$. The denominator in (415) is positive, and

$$-1 \le r_s \le -1/2. \tag{416}$$

For the lower bound on the magnitude use $|r_s| \ge \theta$; for the upper bound use $1 - \phi t_s^\top K^{-1} t_s/\beta \ge 1-\phi = \theta$.

R.4 The square construction

Suppose $\lambda(d-1) \le 1$. Let p be the largest power of four at most $d-1$, and take $k = p$, $J = I_p$, and $\eta = 1/64$. For $Z \in \mathcal{Z}_p$, choose the hidden vectors as the columns of

$$R_Z = b(I + \eta Z).$$

The operator norm of ηZ is at most $1/4$, so every hidden column has norm at most $5/8$. Thus the actual feature norms are bounded by $\sqrt{1/4 + 25/64} < 1$; anchor features also have norm at most one. The covariance and inverse response are

$$H_{\mathrm{hid}} = \frac{qb^2}{p}(I+\eta Z)^2, \tag{417}$$

$$W_Z^{-1} = pK = (\alpha v^2 + \beta\lambda p) I + qb^2 (I+\eta Z)^2. \tag{418}$$

The spectra of pK and W_Z are bounded above and below by positive absolute constants, uniformly in all parameters, because $\lambda p \le 1$ and $q \ge 1/2$. There are therefore constants $R_0, \gamma > 0$ with

$$\|W_Z\|_{\mathrm{op}} \le R_0, \qquad \|W_Z - W_{Z'}\|_F \ge \gamma\sqrt{p}. \tag{419}$$

To verify separation without a commutativity assumption, for positive definite $A, B \succeq (3/4)I$ put $D = A - B$ and observe

$$\langle D, A^2 - B^2\rangle_F = \langle D, AD + DB\rangle_F \ge \frac{3}{2}\|D\|_F^2.$$

Hence $\|A^2 - B^2\|_F \ge (3/2)\|A-B\|_F$. Also, on a uniformly spectrally bounded family, inversion preserves lower Frobenius separation, by $A - B = A(B^{-1} - A^{-1})B$. Apply these facts to (418) and the sign-matrix packing.

All hidden packets are retained. Their effective nuisance mean is zero, so their contribution alone gives

$$G_T + \lambda I \succeq \mathrm{diag}(\lambda, H_{\mathrm{hid}}/\beta) \succeq c_D\, \mathrm{diag}(\lambda, I_p/p) \tag{420}$$

for an absolute $c_D > 0$. This is the needed curvature estimate when λ is much smaller than $1/d$.

R.5 The rectangular construction and its cubic signal

Suppose $\lambda(d-1) > 1$. An absolute constant $\zeta > 0$ will be fixed below. Let r be the largest power of four not exceeding

$$\min\left\{\frac{d-1}{4}, \frac{\zeta\alpha^{2/3}}{\lambda\varepsilon^{1/3}}\right\}, \tag{421}$$

and let m be the largest power of four at most $d-1-r$. We choose the theorem's constant c sufficiently small that $\varepsilon \le \zeta^3\alpha^2$. The quantity in (421) is then at least one. Also $m \ge r$, $m = \Theta(d)$, and

$$r = \Omega\left(\min\left\{d, \frac{\alpha^{2/3}}{\lambda\varepsilon^{1/3}}\right\}\right). \tag{422}$$

Take $k = r$, with r probe and m output nuisance coordinates. For $R \in \mathcal{R}_{m,r}$, set $\tau = 1/1024$ and choose

$$z_j = b \begin{pmatrix} e_j \\ \tau R_{:j} \end{pmatrix}, \qquad j \le r.$$

Every column of R has squared norm one, so the actual feature norm is at most $\sqrt{1/4 + b^2(1+\tau^2)} < 1$.

Write

$$\mu = \beta\lambda r, \quad c_* = \alpha v^2, \quad h_* = q b^2, \quad A_* = \mu + c_* + h_*, \quad V = \begin{pmatrix} I_r \\ \tau R \end{pmatrix}.$$

The parameter choice ensures

$$\frac{3}{64} \le \mu \le \zeta \alpha^{2/3} \varepsilon^{-1/3}. \tag{423}$$

For the lower bound, the largest-power-of-four choice loses at most a factor of four. If the dimension bound in (421) is smaller, use $\lambda(d-1) > 1$; otherwise use $\zeta\alpha^{2/3}\varepsilon^{-1/3} \ge 1$.

We have

$$rK = \mu I + c_* P + h_* V V^\top.$$

Let F_R be the bottom $m \times r$ block of $W_R = (rK)^{-1} J$. Block inversion gives

$$F_R = -c_\mu f_\mu(R), \qquad c_\mu = \frac{h_*\tau}{A_*\mu}, \qquad f_\mu(R) = R(I + \kappa_\mu R^\top R)^{-1}, \qquad \kappa_\mu = \frac{(\mu + c_*)h_*\tau^2}{A_*\mu}. \tag{424}$$

For example, the bottom-right block of rK is $D_* = \mu I + h_*\tau^2 R R^\top$ and $D_*^{-1} R = R(\mu I + h_*\tau^2 R^\top R)^{-1}$. Substituting this identity into the Schur complement of D_* gives (424).

Uniformly for $\mu \ge 3/64$, we have $\kappa_\mu \le \tau^2/\mu$ and $\kappa_\mu C_0^2 \le 1/8$. Consequently

$$\|f_\mu(R)\|_{\mathrm{op}} \le C_0, \qquad \|f_\mu(R) - f_\mu(R')\|_F \ge \frac{1}{2}\|R - R'\|_F. \tag{425}$$

Indeed, put $Q_R = I + \kappa_\mu R^\top R$ and add and subtract $R' Q_R^{-1}$. The first difference has Frobenius norm at least $\|R - R'\|_F/(1 + \kappa_\mu C_0^2)$, while the second has norm at most $2\kappa_\mu C_0^2 \|R - R'\|_F$, by the inverse-difference identity. The difference of these coefficients is greater than $1/2$. Thus the matrices $f_\mu(R)$ form an $\exp(\Omega(mr))$ packing at Frobenius separation $\gamma\sqrt{r}$, with constants independent of μ.

The signal scale relevant to ridge excess is

$$E_\mu := \lambda r c_\mu^2 = \frac{h_*^2 \tau^2}{\beta\mu(\mu + c_* + h_*)^2} \ge \frac{\tau^2}{64(1+\mu)^3}. \tag{426}$$

Here $h_* \ge 1/8$, $\beta \le 1$, and $c_* + h_* \le 1$. The readout amplitude itself is $c_\mu = \Theta((1+\mu)^{-2})$; the additional curvature factor $\lambda r = \mu/\beta$ yields the cubic signal in (426).

R.6 From pointwise randomized answers to finite-bit information

We give the randomness and precision argument explicitly. Consider a finite family of matrices M_i with $\|M_i\|_{\mathrm{op}} \le R_0$ and pairwise Frobenius separation at least $\gamma\sqrt{k}$. Let R denote any public randomness, independent of the packed instance. Suppose a retained bit string Z and a randomized decoder produce vectors $Y(Z, \mathsf{R}, s)$, for $s \in \{-1, 1\}^k$, such that for every fixed i, s,

$$\mathbb{E}\|Y(Z, \mathsf{R}, s) - M_i s\|_2^2 \le \delta_0 k, \tag{427}$$

where the expectation includes the public randomness, encoding randomness, and fresh decoder coins. For sufficiently small absolute δ_0 relative to γ^2, the string must have $\Omega(\log |\{M_i\}|)$ bits.

To prove this, define the mathematical decoder

$$\widehat{M}(Z, \mathsf{R}) = \mathbb{E}_s \left[\mathbb{E}_{\mathrm{dec}}[Y(Z, \mathsf{R}, s) \mid Z, \mathsf{R}, s] s^\top \right],$$

with s uniform on the Boolean cube. The characters $s_1, \dots, s_k$ are orthonormal. Bessel's inequality, row by row, and conditional Jensen give

$$\mathbb{E}_{Z,\mathsf{R}}\|\widehat{M}(Z, \mathsf{R}) - M_i\|_F^2 \le \mathbb{E}_{Z,\mathsf{R},s,\mathrm{dec}}\|Y(Z, \mathsf{R}, s) - M_i s\|_2^2 \le \delta_0 k.$$

Nearest-neighbor decoding identifies i with probability at least $3/4$ when δ_0 is sufficiently small, by Markov's inequality. This success probability is averaged over R; no pointwise success claim for a fixed public seed is needed. For each fixed value of R, a string of at most b bits has fewer than 2^{b+1} possible values, and each value is decoded to one packed instance. Since a uniform packed instance remains uniform conditional on the independent R, the conditional probability of correct identification is at most $2^{b+1}/|\{M_i\}|$. Averaging this counting bound over R and comparing with the averaged success probability proves the claim. Fixed-length summaries give 2^b instead; this additive constant does not affect the lower bound. Unlimited computation and evaluation of the expectation are allowed in this information reduction; no additional information about the retained records is supplied.

We next obtain (427) from the assumed pointwise probability $2/3$. Run the training algorithm independently an odd constant number ℓ of times, using independent public random tapes as well as independent private coins, and retain its ℓ summaries. The public random tapes remain available to their corresponding decoders. If summaries have variable length, append a terminal one and pad each to $B+1$ bits; this encodes every string of length at most B uniquely. Thus the amplification uses at most $\ell(B+1)$ bits, and we suppress this additive constant in the asymptotic expressions below. For these readouts, discard any padded output coordinates and keep only the signal and active nuisance coordinates; this cannot increase excess because the padded feature coordinates are identically zero. For a fixed query, run the decoders independently and select a majority medoid in the public metric used below. Specifically, if good outputs lie within distance R of the true optimum, choose an output with a majority of candidates within distance $2R$, when one exists. When a majority are good such an output exists, and every selected output is within distance $3R$ of the truth, since its majority intersects the good majority. Otherwise choose any output. A Chernoff bound makes the failure probability at most any prescribed absolute constant $\delta > 0$, using constant ℓ. This amplification is valid separately for every fixed dataset and fixed query; it does not require joint correctness for all queries.

For either construction, clip the selected output's scalar residual as

$$\bar{r} = \mathrm{clip}(a\widehat{u} - 1, [-2, -1/4]). \tag{428}$$

The true residual lies in $[-1, -1/2]$ by (416), so $|\bar{r} - r_s| \le a|\widehat{u} - u|$ and $|\bar{r}| \ge 1/4$. Projection of a vector onto a Euclidean ball containing its target cannot increase its distance to that target.

In the square construction, use the public metric $D = \mathrm{diag}(\lambda, I_p/p)$ from (420). On the amplified good event,

$$a^2|\widehat{u} - u|^2 \le C\varepsilon, \qquad \|\widehat{h} - h\|_2^2 \le Cp\varepsilon.$$

Decode and clip

$$Y = \Pi_{\{\|z\|_2 \le R_0\sqrt{p}\}}\left(\frac{\widehat{h}}{\alpha v \bar{r}}\right).$$

Its target is $W_Z s$. Using $h = \alpha v r_s W_Z s$ and (428), the good-event squared error is at most $C\varepsilon p/\alpha^2$. On failure it is at most $4R_0^2 p$ because of projection. Hence

$$\mathbb{E}\|Y - W_Z s\|_2^2 \le C\frac{\varepsilon}{\alpha^2}p + 4R_0^2\delta p.$$

Choose a sufficiently small constant δ, then require $\varepsilon \le c_s\alpha^2$ for a sufficiently small absolute c_s. The information argument proves $\ell B = \Omega(p^2) = \Omega(d^2)$.

For the rectangular construction, use Euclidean distance for the medoid. Since $G_T + \lambda I \succeq \lambda I$, on the amplified good event

$$\|\widehat{w} - w_T\|_2^2 \le 9\varepsilon/\lambda, \qquad a^2|\widehat{u} - u|^2 \le 9\varepsilon,$$

where the second inequality uses $a^2 \le \lambda$. Write h_{out} for the bottom m nuisance coordinates. Crucially, decode only this block and normalize by its public amplitude c_μ:

$$Y = \Pi_{\{\|z\|_2 \le C_0\sqrt{r}\}}\left(-\frac{\widehat{h}_{\mathrm{out}}}{\alpha v c_\mu \bar{r}}\right). \tag{429}$$

Its target is $f_\mu(R)s$, because $h_{\mathrm{out}} = -\alpha v r_s c_\mu f_\mu(R)s$. By (428), on the good event

$$\|Y - f_\mu(R)s\|_2 \le \frac{4}{\alpha v c_\mu}\|\widehat{h}_{\mathrm{out}} - h_{\mathrm{out}}\|_2 + 4\|f_\mu(R)s\|_2\,|\bar{r} - r_s|,$$

$$\|Y - f_\mu(R)s\|_2^2 \le C\left(\frac{\varepsilon}{\alpha^2\lambda c_\mu^2} + r\varepsilon\right).$$

The second term is the scalar-denominator error; it carries no adverse power of μ. On failure, projection bounds squared error by $4C_0^2 r$. Therefore

$$\mathbb{E}\|Y - f_\mu(R)s\|_2^2 \le C\left(\frac{\varepsilon}{\alpha^2\lambda c_\mu^2} + r\varepsilon\right) + 4C_0^2\delta r. \tag{430}$$

The normalization and clipping in (429) keep the required amplification constant independent of μ, even when μ grows without bound.

Choose δ sufficiently small for the packing in (425). Equation (430) meets (427) whenever

$$\varepsilon \le c_a\alpha^2 E_\mu, \qquad \varepsilon \le c_b, \tag{431}$$

for absolute constants $c_a, c_b > 0$. Fix $\zeta > 0$ sufficiently small that $8\zeta^3 \le c_a\tau^2/64$. Our restriction $\varepsilon \le \zeta^3\alpha^2$ and (423) give

$$1 + \mu \le 2\zeta\alpha^{2/3}\varepsilon^{-1/3}, \qquad \frac{\varepsilon(1+\mu)^3}{\alpha^2} \le 8\zeta^3.$$

Together with (426), these imply the first condition in (431). Choosing the theorem's absolute constant c still smaller guarantees the second condition, $\varepsilon \le \zeta^3\alpha^2$, and the square-case condition $\varepsilon \le c_s\alpha^2$, since $\alpha \ge \rho/2$. The information argument then yields

$$\ell B = \Omega(mr) = \Omega\left(d\min\left\{d, \frac{\rho^{2/3}}{\lambda\varepsilon^{1/3}}\right\}\right).$$

The number ℓ is an absolute constant. The square construction proves the required bound in the remaining regime, completing the bit lower bound.

R.7 Exact grid and model qualifications

No rounding argument is needed for these inputs. In the square construction, p is a power of four, so $1/\sqrt{p}$ is dyadic; the hidden factor $b(I+\eta Z)$ has dyadic entries with $O(\log d)$ bits. In the rectangular construction, m is a power of four and the entries of the hidden vectors are b, zero, or $\pm b\tau/\sqrt{m}$, also on such a grid. The constants a, b, v, η, τ are dyadic. The coordinate a requires $O(1 + \log(1/\lambda))$ fractional bits, and all other coordinates need $O(\log d)$ bits. Taking a common refinement gives the stated $O(\log d + \log(1/\lambda))$ bound, independent of ε and of the hidden instance. Mirror pairs preserve exact noiselessness and all centering identities on this grid. Dataset duplication preserves these properties as well as all normalized losses and query guarantees.

The finite grid describes the input alphabet. The communication argument above separately charges every persistent data-dependent bit, including any retained model or randomness needed by the decoder. Fresh decoder coins and independent shared public randomness cannot evade its counting bound. All query records are fixed anchors common to the family, and the proof concerns a single batch of legal deletions; evaluating several possible queries in the mathematical information decoder is not a sequential-update assumption.

R.8 Rescaling to a unit-norm teacher

Corollary 112 (Unit-teacher lower bound)**.** *There is an absolute $c' > 0$ such that, for $d \ge d_0$, $0 < \lambda \le 1$, $0 < \rho \le 1/4$, and $0 < \varepsilon \le c'\lambda\rho^2$, the same sample-size, centering, common-query, and finite-grid conclusions hold for datasets satisfying $y = x^\top e_1$, hence having a public teacher of norm one. Their common original minimizer is*

$$w_S = \frac{a^2}{a^2 + \lambda}e_1,$$

and any summary attaining retained ridge excess at most ε with probability at least $2/3$ for each fixed dataset and query requires

$$\Omega\left(d \min\left\{d, \frac{\rho^{2/3}}{\lambda^{2/3}\varepsilon^{1/3}}\right\}\right)$$

bits. For fixed ρ, the sample size is $\Theta(d)$ and input precision is $O(\log d + \log(1/\lambda))$ bits per coordinate.

Proof. Multiply every label in Theorem 111, including query labels, by the public dyadic a. The features, sample counts, deletion fractions, and common-grid precision are unchanged. All transformed records satisfy $y = x_1$ and $|y| = a \le 1/2$. If F_T denotes the original objective and $F_T^{(a)}$ the transformed objective with the same λ, then

$$F_T^{(a)}(aw) = a^2 F_T(w), \qquad w_T^{(a)} = a w_T.$$

Consequently an excess-ε decoder for the transformed problem, followed by division of its output by a, gives an excess-ε/a^2 decoder for the original problem without adding data-dependent memory. Since $a^2 \ge \lambda/4$, choosing $c' \le c/4$ makes the original theorem applicable. Its lower bound becomes

$$\Omega\left(d \min\left\{d, \frac{\rho^{2/3}a^{2/3}}{\lambda\varepsilon^{1/3}}\right\}\right),$$

which has the claimed order because $a^2 = \Theta(\lambda)$. The reduction preserves the success probability for each fixed query. □

In this direct rescaling, label magnitudes are $\Theta(\sqrt{\lambda})$; the corollary does not assert a response variance bounded below independently of λ.

R.9 A BOUNDED TEACHER WITH NONVANISHING LABEL VARIANCE

Corollary 113 (Bounded teacher and label variance)**.** *There are absolute constants $c'' > 0$ and d_1 such that, for $d \ge d_1$, $0 < \lambda \le 1/4$, $0 < \rho \le 1/8$, and $0 < \varepsilon \le c''\lambda\rho^2$, there is a finite family of datasets of public size $n = \Theta(d/\rho)$ with the following properties. Their features have norm at most one, their labels have magnitude at most $1/2$, and every record satisfies $y = x^\top v$ for the same public vector $v = (e_0 + e_1)/2$, of norm $1/\sqrt{2}$. Every original dataset and every queried retained dataset has zero feature and label means and label variance at least $1/8$. Every original minimizer is the same public nonzero vector. The common legal deletion queries remove a fraction in $[\rho/2, \rho]$.*

Any summary with pointwise success probability at least $2/3$ for retained ridge excess at most ε requires

$$\Omega\left(d \min\left\{d, \frac{\rho^{2/3}}{\lambda^{2/3}\varepsilon^{1/3}}\right\}\right)$$

bits. The coordinates and labels lie on a common dyadic grid with $O(\log d + \log(1/\lambda))$ bits per entry. Thus, for fixed ρ, $n = \Theta(d)$ and the lower bound has the same dimension, regularization, and accuracy exponents as the unit-teacher lower bound.

Proof. Write the final feature space as $\mathbb{R}e_0 \oplus \mathbb{R}^{d-1}$, with e_1 denoting the first coordinate of the second summand. Apply Corollary 112 in dimension $d-1$ with deletion budget $\rho_h = 2\rho$. Its packets delete an α fraction, where $\rho \le \alpha \le 2\rho \le 1/4$. This fraction depends only on the public budget, through $L = \lceil 1/(4\rho) \rceil$ and $\alpha = 1/(2L)$. Set

$$\beta = 1 - \alpha, \qquad \gamma = \frac{\beta}{1+\beta}, \qquad \lambda_0 = \frac{\lambda}{\gamma}.$$

Then $3/4 \le \beta < 1$, $3/7 \le \gamma < 1/2$, and $\lambda_0 \le (7/3)\lambda \le 7/12 < 1$, so the hard construction at regularization λ_0 is valid. Its dyadic signal coordinate satisfies $a^2 = \Theta(\lambda_0) = \Theta(\lambda)$.

Let the hard unit-teacher dataset have n_h records (x, y) and the common original statistics

$$G_h = \mathrm{diag}(a^2, H), \qquad b_h = a^2 e_1.$$

Embed its features in the second summand and replace each label by $y/2$. Append n_h public records: $n_h/2$ copies each of $(e_0, 1/2)$ and $(-e_0, -1/2)$. The hard construction's sample size is even. The resulting size is $n = 2n_h = \Theta(d/\rho)$. Delete only the transformed hard query records. Their final deletion fraction is $\alpha/2 \in [\rho/2, \rho]$, and they are common to all instances. The final labels are exactly $x^\top(e_0+e_1)/2$; feature norm bounds and centering follow from those of the two component samples.

The original hard labels have magnitude a, whereas the appended labels have magnitude $1/2$. Thus the final original label variance is

$$\frac{1+a^2}{8} \geq \frac{1}{8}.$$

After deletion, the hard sample has βn_h records and the public sample still has n_h. Their means remain zero, and the retained label variance is

$$\frac{1+\beta a^2}{4(1+\beta)} \geq \frac{1}{8}.$$

For $w = (w_0, w_{\mathrm{weak}})$, the final retained objective decomposes exactly as

$$F_{\mathrm{final}}(w_0, w_{\mathrm{weak}}) = \frac{(1/2 - w_0)^2}{1+\beta} + \lambda w_0^2 + \frac{\gamma}{4} F_{h,\lambda_0}(2w_{\mathrm{weak}}), \tag{432}$$

where F_{h,λ_0} is the normalized retained hard objective before halving its labels. The hard loss has weight γ, label scaling contributes the factor $1/4$, and its regularizer becomes $\gamma\lambda_0\|w_{\mathrm{weak}}\|_2^2 = \lambda\|w_{\mathrm{weak}}\|_2^2$. This verifies all terms in (432).

The two blocks minimize separately. Therefore a final excess-ε output yields hard excess at most $4\varepsilon/\gamma$ by returning $2w_{\mathrm{weak}}$. A summary for the final instances would thus give a summary of the same bit length for the hard instances: training adds only public records and rescales labels, and the deletion decoder embeds and rescales the public query packet. Pointwise success probabilities are preserved. The hard theorem applies because

$$\frac{4\varepsilon}{\gamma} \leq c'\lambda_0(2\rho)^2 \quad \text{whenever} \quad \varepsilon \leq c'\lambda\rho^2.$$

Its inverse-rank factor is

$$\frac{(2\rho)^{2/3}}{\lambda_0^{2/3}(4\varepsilon/\gamma)^{1/3}} = \frac{\gamma\rho^{2/3}}{\lambda^{2/3}\varepsilon^{1/3}},$$

which differs from the claimed factor only by an absolute constant. Dimension $d-1$ can likewise be replaced by d in the asymptotic lower bound.

The original optimizer must be recomputed under the full mixture, rather than obtained by scaling the hard optimizer at λ_0. The full Gram blocks are $1/2$ on the public direction and $G_h/2$ on the hard directions; the corresponding response moments are $1/4$ and $b_h/4$. Consequently every original minimizer is exactly

$$w_S = \frac{1}{2(1+2\lambda)}e_0 + \frac{a^2}{2(a^2+2\lambda)}e_1,$$

independently of the hidden matrix. Finally, halving dyadic labels costs one extra fractional bit, the appended coordinates and labels are dyadic, and $\lambda_0 = \Theta(\lambda)$. The claimed common-grid precision is therefore unchanged in order. This proves all assertions. □

The public direction supplies the nonvanishing label variance. This corollary does not require a variance lower bound in every hidden direction; the information-bearing directions retain the spectral scale used in the hard construction.

R.10 Rare anchors with stronger signal

The unit-teacher reduction above gives every record the same absolute label. A bounded teacher also permits a small public population to carry stronger signal than the hidden records. Deleting that population changes the response moment more substantially, even while the original signal covariance remains at the ridge scale. The next construction improves the deletion-fraction dependence of the preceding lower bound; it does not assert a matching upper bound for vanishing deletion fractions.

Theorem 114 (Bounded teacher with rare signal-bearing anchors)**.** *There are absolute constants $c_{\mathrm{ra}} > 0$ and d_{ra} such that, for every $d \geq d_{\mathrm{ra}}$, $0 < \lambda \leq 1$, $0 < \rho \leq 1/4$, and*

$$0 < \varepsilon \leq c_{\mathrm{ra}} A_{\rho,\lambda}, \qquad A_{\rho,\lambda} = \min\{\rho^2, \rho\lambda\},$$

there is a finite family of bounded, centered datasets of one public size $n = \Theta(d/\rho)$, and common legal deletion requests, with these properties. Every feature satisfies $\|x\|_2 \leq 1$, and every label satisfies $y = x^\top e_1$ and $|y| \leq 1/2$. Every original minimizer is the same public nonzero vector, and every request deletes a fraction in $[\rho/2, \rho]$. The original and queried retained samples have zero feature and label means. All coordinates and labels belong to a common dyadic grid with $O(\log d + \log(1/\lambda))$ bits per entry.

Any summary with pointwise randomized success probability at least $2/3$ for retained ridge excess at most ε requires

$$\Omega\left(d \min\left\{d, \left(\frac{A_{\rho,\lambda}}{\lambda^3 \varepsilon}\right)^{1/3}\right\}\right) \tag{433}$$

persistent bits, including all data-dependent retained state. Independent shared public randomness and fresh decoder randomness are allowed.

Proof. We specify the modified records and their normalized statistics first, then prove the information readout with its changed scalar error.

Signal coordinates and common packets. Take L, α, β, q as in Appendix R.2. Thus $\rho/2 \leq \alpha \leq \rho \leq 1/4$, $\beta = 1 - \alpha \geq 3/4$, and $q = 1 - 2\alpha \geq 1/2$. Let a_h and a_a be the largest powers of two not exceeding, respectively,

$$\min\{\sqrt{\lambda}, 1/2\}, \qquad \min\{1/2, \sqrt{\lambda/(8\alpha)}\}. \tag{434}$$

The subscripts distinguish hidden records from public anchors. They obey

$$\lambda/4 \leq a_h^2 \leq \min\{\lambda, 1/4\}, \qquad \frac{1}{4}\min\{1/4, \lambda/(8\alpha)\} \leq a_a^2 \leq \min\{1/4, \lambda/(8\alpha)\}, \tag{435}$$

$$D := \alpha^2 a_a^2, \qquad \frac{A_{\rho,\lambda}}{128} \leq D \leq \frac{A_{\rho,\lambda}}{4}, \qquad D \leq \frac{\lambda}{32}. \tag{436}$$

Indeed, D lies between $\min\{\alpha^2, \alpha\lambda\}/32$ and $\min\{\alpha^2, \alpha\lambda\}/4$, and $\alpha \in [\rho/2, \rho]$. The last bound uses $D \leq \alpha\lambda/8$ and $\alpha \leq 1/4$.

Keep the nuisance vectors and multiplicities of the earlier construction: for each of k probe coordinates include effective anchors $+ve_j$ and $-ve_j$ once each, and effective hidden vectors $+z_j$ and $-z_j$ each $L - 1$ times, where $v = b = 1/2$. An effective anchor z now denotes

$$((a_a, z), a_a), \qquad ((-a_a, -z), -a_a),$$

and an effective hidden vector z denotes

$$((a_h, z), a_h), \qquad ((-a_h, -z), -a_h).$$

Every actual record is noiseless for the same teacher e_1 and belongs to a centered mirror pair. There are $n_0 = 4Lk$ actual records. The anchor population has mass 2α and the hidden population has mass q. Both signs of each nuisance vector occur with the same signal coordinate, so their signal–nuisance cross moments cancel.

Write $\Sigma_S = n_0^{-1} \sum_S xx^\top$ and $g_S = n_0^{-1} \sum_S xy$. With $P = JJ^\top$ as before, the normalized statistics are exactly

$$\Sigma_S = \mathrm{diag}(s_0, H), \qquad g_S = s_0 e_1, \qquad s_0 = 2\alpha a_a^2 + q a_h^2, \quad H = 2C + H_{\mathrm{hid}}, \tag{437}$$

where $C = \alpha v^2 P/k$ and $H_{\mathrm{hid}} = q \sum_{j=1}^k z_j z_j^\top / k$. In particular,

$$\lambda/8 \leq s_0 \leq 5\lambda/4, \qquad w_S = \frac{s_0}{s_0 + \lambda} e_1. \tag{438}$$

This common model depends only on public parameters and is nonzero.

For $s \in \{-1,1\}^k$, delete the actual mirror pair for anchor $s_j v e_j$ for every j. These are common public records; the query deletes $2k = \alpha n_0$ records and preserves centering. Set $t_s = \alpha v J s/k$ and $K = H - C + \beta\lambda I$. The normalized retained statistics are

$$\Sigma_T = \frac{1}{\beta}\begin{pmatrix} s_0 - \alpha a_a^2 & -a_a t_s^\top \\ -a_a t_s & H - C \end{pmatrix}, \qquad g_T = \frac{1}{\beta}\begin{pmatrix} s_0 - \alpha a_a^2 \\ -a_a t_s \end{pmatrix}. \tag{439}$$

Thus, writing $w_T = (u, h)$, the normal equations are

$$(s_0 - \alpha a_a^2 + \beta\lambda)u - a_a t_s^\top h = s_0 - \alpha a_a^2, \tag{440}$$

$$Kh = a_a(u-1)t_s. \tag{441}$$

For $W = (kK)^{-1}J$ and $r_s = u - 1$, this gives

$$h = \alpha v a_a r_s W s, \qquad r_s = -\frac{\beta\lambda}{s_0 - \alpha a_a^2 + \beta\lambda - a_a^2 t_s^\top K^{-1} t_s}. \tag{442}$$

The nuisance covariance is unchanged, so $K \succeq C + \beta\lambda I$ and $0 \le t_s^\top K^{-1} t_s \le \alpha$. This denominator is at least $q a_h^2 + \beta\lambda \ge \beta\lambda$ and at most $s_0 + \beta\lambda \le 9\lambda/4$. Therefore

$$-1 \le r_s \le -1/3. \tag{443}$$

Square family. If $\lambda(d-1) \le 1$, choose the largest power of four $p \le d-1$, set $k = p$, and take z_j as the columns of $b(I + \eta Z)$ for $Z \in \mathcal{Z}_p$, with $\eta = 1/64$. Every hidden nuisance vector has norm at most $5/8$ and every signal coordinate has magnitude at most $1/2$. Thus hidden feature norms are at most $\sqrt{1/4 + 25/64} < 1$ and anchor feature norms are at most $\sqrt{1/4 + 1/4} < 1$.

Neither K nor W depends on the changed signal coordinates. Their exact formulas are

$$W_Z^{-1} = (\alpha v^2 + \beta\lambda p)I + q b^2 (I + \eta Z)^2, \qquad H_{\mathrm{hid}} = \frac{q b^2}{p}(I + \eta Z)^2. \tag{444}$$

The spectral and separation argument in (419) applies to these identical matrices: $\|W_Z\|_{\mathrm{op}} \le R_0$, $\|W_Z - W_{Z'}\|_F \ge \gamma\sqrt{p}$, and $\log|\mathcal{Z}_p| = \Omega(p^2)$, for absolute constants. All hidden records remain after deletion, and their signal–nuisance cross terms cancel. Their contribution, together with regularization, gives

$$\Sigma_T + \lambda I \succeq \mathrm{diag}(\lambda, H_{\mathrm{hid}}/\beta) \succeq c_D\,\mathrm{diag}(\lambda, I_p/p). \tag{445}$$

Use constant repetition and the public medoid metric $\mathrm{diag}(\lambda, I_p/p)$ as in Appendix R.6. For each fixed dataset and query, the amplified good event has probability at least $1 - \delta$, where $\delta > 0$ is any chosen absolute constant. On that event,

$$|\widehat{u} - u|^2 \le C\varepsilon/\lambda, \qquad \|\widehat{h} - h\|_2^2 \le Cp\varepsilon.$$

Clip the scalar and project the readout:

$$\bar{r} = \mathrm{clip}(\widehat{u} - 1, [-2, -1/6]), \qquad Y = \Pi_{\{\|z\|_2 \le R_0\sqrt{p}\}}\left(\frac{\widehat{h}}{\alpha v a_a \bar{r}}\right). \tag{446}$$

The true residual belongs to the clipping interval, so $|\bar{r} - r_s| \le |\widehat{u} - u|$ and $|\bar{r}| \ge 1/6$. The target $W_Z s$ lies in the projection ball. On the good event,

$$\|Y - W_Z s\|_2 \le \frac{6}{\alpha v a_a}\|\widehat{h} - h\|_2 + 6\|W_Z s\|_2 |\bar{r} - r_s|,$$

$$\|Y - W_Z s\|_2^2 \le Cp\left(\frac{\varepsilon}{D} + \frac{\varepsilon}{\lambda}\right).$$

On failure, projection bounds squared error by $4R_0^2 p$. Hence

$$\mathbb{E}\|Y - W_Z s\|_2^2 \le Cp\left(\frac{\varepsilon}{D} + \frac{\varepsilon}{\lambda}\right) + 4R_0^2\delta p. \tag{447}$$

Choose δ sufficiently small for the packing constant and then require $\varepsilon \le c_s D$. Since $D \le \lambda/32$, both good-event terms are sufficiently small. The information argument of Appendix R.6, with public randomness included in its mathematical decoder, yields $B = \Omega(p^2) = \Omega(d^2)$.

Rectangular family. Suppose instead that $\lambda(d-1) > 1$. Fix an absolute constant $\zeta > 0$ below, and take the largest power of four r not exceeding

$$\min\left\{\frac{d-1}{4}, \frac{\zeta}{\lambda}\left(\frac{D}{\varepsilon}\right)^{1/3}\right\}. \tag{448}$$

We will require $\varepsilon \le \zeta^3 D$, so the second quantity is at least $1/\lambda \ge 1$; for sufficiently large d the first is also at least one. Let m_o be the largest power of four at most $d-1-r$. Then $m_o \ge r$, $m_o = \Theta(d)$, and

$$r = \Omega\left(\min\left\{d, \frac{1}{\lambda}\left(\frac{D}{\varepsilon}\right)^{1/3}\right\}\right). \tag{449}$$

Set $k = r$, $\tau = 1/1024$, and $z_j = b(e_j, \tau R_{\cdot j})$ for $R \in \mathcal{R}_{m_o,r}$. Hidden features have norm at most $\sqrt{1/4 + b^2(1+\tau^2)} < 1$, and the anchor bound remains valid.

With

$$\mu = \beta\lambda r, \quad c_* = \alpha v^2, \quad h_* = qb^2, \quad A_* = \mu + c_* + h_*,$$

the largest-power choice gives

$$\frac{3}{64} \le \mu \le \zeta(D/\varepsilon)^{1/3}. \tag{450}$$

If the dimension bound is active, $r > (d-1)/16$ and $\lambda(d-1) > 1$ give $\mu > 3/64$. If the accuracy bound is active, $\lambda r > \zeta(D/\varepsilon)^{1/3}/4 \ge 1/4$. The upper bound follows from (448).

The nuisance normal matrix is exactly the one previously inverted:

$$rK = \mu I + c_* P + h_* \begin{pmatrix} I_r \\ \tau R \end{pmatrix} \begin{pmatrix} I_r \\ \tau R \end{pmatrix}^\top.$$

The bottom block of $W = (rK)^{-1}J$ is consequently

$$-c_\mu f_\mu(R), \qquad c_\mu = \frac{h_*\tau}{A_*\mu}, \qquad f_\mu(R) = R(I + \kappa_\mu R^\top R)^{-1}, \quad \kappa_\mu = \frac{(\mu + c_*)h_*\tau^2}{A_*\mu}. \tag{451}$$

Equations (425) and (426) give an $\exp(\Omega(m_o r))$ packing of $f_\mu(R)$ with operator norm at most C_0, Frobenius separation at least $\gamma\sqrt{r}$, and signal

$$E_\mu = \lambda r c_\mu^2 \ge \frac{\tau^2}{64(1+\mu)^3}. \tag{452}$$

Amplify with the Euclidean medoid metric. Since $\Sigma_T + \lambda I \succeq \lambda I$, the good event gives $\|\widehat{w} - w_T\|_2^2 \le 9\varepsilon/\lambda$. Use the same scalar clipping as in (446) and decode only the output nuisance coordinates:

$$Y = \Pi_{\{\|z\|_2 \le C_0\sqrt{r}\}}\left(-\frac{\widehat{h}_{\mathrm{out}}}{\alpha v a_a c_\mu \bar{r}}\right). \tag{453}$$

The target is $f_\mu(R)s$, because $h_{\mathrm{out}} = -\alpha v a_a r_s c_\mu f_\mu(R)s$. On the good event,

$$\|Y - f_\mu(R)s\|_2 \le \frac{6}{\alpha v a_a c_\mu}\|\widehat{h}_{\mathrm{out}} - h_{\mathrm{out}}\|_2 + 6\|f_\mu(R)s\|_2|\bar{r} - r_s|,$$

$$\|Y - f_\mu(R)s\|_2^2 \le C\left(\frac{\varepsilon}{D\lambda c_\mu^2} + \frac{r\varepsilon}{\lambda}\right).$$

Projection bounds the failure contribution by $4C_0^2\delta r$. Thus

$$\mathbb{E}\|Y - f_\mu(R)s\|_2^2 \le Cr\left(\frac{\varepsilon}{DE_\mu} + \frac{\varepsilon}{\lambda}\right) + 4C_0^2\delta r. \tag{454}$$

The scalar error of the new unit-teacher readout costs ε/λ, independently of μ.

Choose δ sufficiently small. For absolute $c_a, c_b > 0$, the information lemma applies whenever $\varepsilon \le c_a D E_\mu$ and $\varepsilon \le c_b\lambda$. Fix ζ sufficiently small that $8\zeta^3 \le c_a\tau^2/64$. If $\varepsilon \le \zeta^3 D$, then

$$1 + \mu \le 2\zeta(D/\varepsilon)^{1/3}, \qquad \frac{\varepsilon(1+\mu)^3}{D} \le 8\zeta^3.$$

Together with (452), these give $\varepsilon \le c_a D E_\mu$. A sufficiently small bound on ε/D, using $D \le \lambda/32$, also gives $\varepsilon \le c_b\lambda$. The information conclusion is $B = \Omega(m_o r)$, proving (433) by (449) and (436). Choose c_{ra} sufficiently small that $128c_{\mathrm{ra}} \le \min\{c_s, \zeta^3, 32c_b\}$; this enforces every accuracy restriction in both branches.

Sample size, precision, and randomness. In either branch, duplicate the entire base sample $\lceil d/k \rceil$ times and delete every copy of the selected anchors. Then

$$n = 4Lk\lceil d/k \rceil = \Theta(d/\rho),$$

with absolute constants. This public sample size is independent of the packed matrix. Duplication leaves all normalized statistics, objectives, original models, centering, and deletion fractions unchanged. The hidden nuisance coordinates are dyadic with $O(\log d)$ bits, since p, m_o, r are powers of four and b, η, τ are fixed dyadic numbers. Both signal coordinates in (434) are dyadic. Moreover, $a_h^2 \ge \lambda/4$ and $a_a^2 \ge \min\{1/16, \lambda/(32\alpha)\} \ge \lambda/16$, where the last inequality uses $\alpha \le 1/4$ and $\lambda \le 1$. Thus a common refinement uses $O(\log d + \log(1/\lambda))$ bits, without a dependence on ε or $\log(1/\rho)$. Labels are precisely these signed dyadic signal coordinates, preserving exact noiselessness.

The amplification uses a constant number of independent summaries and independent public tapes, and its projection bounds hold even on failed outputs. The mean squared error conclusions are pointwise in the fixed instance and query, averaged over all these coins. The information lemma conditions only its counting bound on the independent public tapes; it uses the averaged success bound. Neither a simultaneous-success assumption nor retained-data access enters the proof. □

Corollary 115 (Rare anchors with nonvanishing label variance)**.** *There are absolute constants $c_v > 0$ and d_v such that, for $d \ge d_v$, $0 < \lambda \le 1/4$, $0 < \rho \le 1/8$, and $0 < \varepsilon \le c_v \min\{\rho^2, \rho\lambda\}$, the lower bound* (433) *holds for centered datasets with a common public teacher $(e_0 + e_1)/2$, labels of magnitude at most $1/2$, and label variance at least $1/8$ before and after every queried deletion. The original optimizer is the same public nonzero vector throughout the family. The sample size, common-query, pointwise-randomness, and grid conclusions of Theorem 114 also hold.*

Proof. Apply Theorem 114 in dimension $d - 1$ with budget 2ρ. Its actual deletion fraction $\alpha \in [\rho, 2\rho]$ is fixed by this public budget. Define

$$\beta = 1 - \alpha, \qquad \gamma = \frac{\beta}{1+\beta}, \qquad \lambda_0 = \frac{\lambda}{\gamma}.$$

Here $3/7 \le \gamma < 1/2$ and $\lambda_0 \le 7\lambda/3 \le 7/12 < 1$, within that theorem's range. Let S_h be its hard dataset of even size n_h, at regularization λ_0. Its normalized original Gram and response moments are $\mathrm{diag}(s_0, H)$ and $s_0 e_1$, with $s_0 = 2\alpha a_a^2 + q a_h^2 = \Theta(\lambda_0)$.

Embed these records orthogonally to e_0, halve their labels, and append $n_h/2$ copies each of $(e_0, 1/2)$ and $(-e_0, -1/2)$. The new size is $n = 2n_h = \Theta(d/\rho)$. Delete only the transformed hard query records. The final deletion fraction is $\alpha/2 \in [\rho/2, \rho]$, and the teacher is $(e_0 + e_1)/2$, with norm $1/\sqrt{2}$. Features remain bounded, and mirror pairs give zero feature and label means.

The original label variance is $(1 + s_0)/8 \ge 1/8$. Deleting the hard anchors removes hard squared-label mass αa_a^2 before label halving. Therefore the retained label variance is exactly

$$\frac{1 + s_0 - \alpha a_a^2}{4(1+\beta)} \ge \frac{1}{8}.$$

For $w = (w_0, w_{\mathrm{weak}})$, the retained objective decomposes as

$$F_{\mathrm{final}}(w) = \frac{(1/2 - w_0)^2}{1+\beta} + \lambda w_0^2 + \frac{\gamma}{4} F_{h,\lambda_0}(2w_{\mathrm{weak}}).$$

Thus an excess-ε answer gives an excess-$4\varepsilon/\gamma$ answer for the hard sample. The hypothesis of Theorem 114 is satisfied if

$$\varepsilon \le c_{\mathrm{ra}} \min\{\gamma\rho^2, \rho\lambda/2\}.$$

This follows from the corollary's hypothesis for sufficiently small c_v, since $\gamma \geq 3/7$. The cubed inverse-rank factor becomes

$$\frac{\min\{(2\rho)^2, (2\rho)\lambda_0\}}{\lambda_0^3(4\varepsilon/\gamma)} = \frac{\min\{\gamma^4\rho^2, (\gamma^3/2)\rho\lambda\}}{\lambda^3\varepsilon} = \Theta\left(\frac{\min\{\rho^2, \rho\lambda\}}{\lambda^3\varepsilon}\right).$$

The reduction adds only public records and rescales public query labels, preserving summary length and pointwise success probability.

Finally, the full original covariance blocks are $1/2$ and $\mathrm{diag}(s_0, H)/2$, and their response moments are $1/4$ and $s_0 e_1/4$. Every original optimizer is consequently

$$w_S = \frac{1}{2(1+2\lambda)} e_0 + \frac{s_0}{2(s_0+2\lambda)} e_1.$$

The common grid gains only one bit from label halving, and $\lambda_0 = \Theta(\lambda)$ preserves its precision. □

The stronger lower bound reflects the signal carried by the deletable population. When $\rho \leq \lambda$, public anchors may have labels of constant magnitude while occupying only $O(\rho)$ of the data. When $\lambda \leq \rho$, their squared labels scale as λ/ρ. In both cases the original signal second moment stays $\Theta(\lambda)$, so the common ridge predictor has constant shrinkage. The separate public direction in Corollary 115 supplies total label variance bounded away from zero; it does not supply hidden information.

Corollary 116 (Relative accuracy with a bounded teacher)**.** *There are absolute constants $c > 0$ and d_0 such that, for $d \geq d_0$, $0 < \lambda \leq 1/4$, $0 < \rho \leq 1/8$, and*

$$0 < \eta \leq c\min\{\rho^2/\lambda, \rho\}, \qquad A = \min\{\rho^2, \rho\lambda\},$$

there is a finite family of bounded datasets for which any summary guaranteeing $F_{S-U}(\widehat{w}_U) \leq (1+\eta)F_{S-U}(w_{S-U})$ with probability at least $2/3$ for each fixed dataset and query requires

$$B = \Omega\left(d\min\left\{d, \left(\frac{A}{\lambda^4\eta}\right)^{1/3}\right\}\right). \tag{455}$$

The hard datasets are centered, have common public teacher $v = (e_0+e_1)/2$ of norm $1/\sqrt{2}$, and have label variance at least $1/8$ both originally and after each queried deletion. Their original optimizer is public, nonzero, and common to the family. They have $n = \Theta(d/\rho)$, common legal queries deleting a fraction in $[\rho/2, \rho]$, and input grid precision $O(\log d + \log(1/\lambda))$. Both the original and every queried retained optimal objective lie in $[\lambda/6, \lambda/2]$.

Proof. Use the family in Corollary 115. Exact linear labels give $F_S(v) = F_{S-U}(v) = \lambda\|v\|^2 = \lambda/2$, proving the upper bounds on the optima. In the notation of that corollary's proof, the minimum of the retained public-coordinate objective is

$$\min_z \left\{\frac{(1/2-z)^2}{1+\beta} + \lambda z^2\right\} = \frac{\lambda}{4(1+\lambda(1+\beta))} \geq \frac{\lambda}{6}.$$

The corresponding original minimum is $\lambda/[4(1+2\lambda)] \geq \lambda/6$; the remaining objective block is nonnegative. Thus relative tolerance η implies additive excess at most $\lambda\eta/2$. Choosing $c \leq \min\{1, 2c_v\}$ allows the preceding corollary at $\varepsilon = \lambda\eta/2$, which proves (455) with unchanged randomness and public-data conclusions. □

For fixed $\rho \leq 1/8$ and $\lambda \leq 1/4$, $4\rho^2 \leq \min\{\rho^2/\lambda, \rho\} \leq \rho$. Consequently (455) is

$$\Omega_\rho\left(d\min\{d, (\lambda\eta^{1/3})^{-1}\}\right).$$

The generic relative upper bound in Appendix A.4 and Corollary 11, with $n = \Theta_\rho(d)$ and the extended input grid above, gives

$$O_\rho\left(d\min\{d, (\lambda\eta^{1/3})^{-1}\}(\log d + \log(1/\lambda))\right).$$

Indeed, in the low-rank branch $(\lambda\eta^{1/3})^{-1} = O(d)$ bounds the required rounding precision by $O(\log d)$, and exact response storage adds the input-grid precision; the exact Gram branch uses

the same extended precision. For $\lambda \geq d^{-O(1)}$ these bounds match within $O_\rho(\log d)$. The relative lower bound follows by converting the additive tolerance. At fixed ρ, the unsaturated additive rank scales as $\lambda^{-2/3}\varepsilon^{-1/3}$, whereas relative tolerance gives $\lambda^{-1}\eta^{-1/3}$ because the optimal objective is $\Theta(\lambda)$.

Proposition 117 (A vanishing-fraction obstruction for spectral flooring)**.** *Fix $\lambda = 1/4$. For $\alpha = 4^{-j}$ with j sufficiently large, there is a centered noiseless dataset with public unit teacher, $d = 1/(4\alpha^2)+2$, $n = 2/\alpha^3 = \Theta(d/\alpha)$, and grid precision $O(\log d)$, and a legal deletion of fraction α, such that both original and retained label variances equal $1/4$. For every rank $1 \leq \ell \leq d-2$, ordinary spectral flooring* (7)–(8) *and centered flooring* (509) *have the same excess*

$$F_{S-U}(\widehat{w}_U) - F_{S-U}(w_{S-U}) = \left(\frac{1}{4} + o(1)\right)\alpha^7. \tag{456}$$

Rank zero has at least this excess; rank $d-1$ is exact. In particular, additive tolerance $\varepsilon = \alpha^7/8$ requires rank $d-1$ within this family of floor decoders, although

$$\left(\frac{\alpha^2}{\lambda^3\varepsilon}\right)^{1/3} = 8\alpha^{-5/3} = o(d).$$

Proof. Put $t = \alpha^2$, $q = 1/(4t)$, and use coordinates consisting of e_1, q dummy directions, and e_b. A population of mass p at $\pm x$ means mass $p/2$ at each of (x, x_1) and $(-x, -x_1)$; the teacher is e_1. Take one population of total mass α at $\pm x_+$ and another of total mass α at $\pm x_-$, where

$$x_+ = \tfrac{1}{2}e_1 + \tfrac{\sqrt{\alpha}}{2}e_b, \qquad x_- = \tfrac{1}{2}e_1 - \tfrac{\sqrt{\alpha}}{2}e_b.$$

Also take mass $1/4 - \alpha/2$ at $\pm e_1$, total mass t at the $\pm$ pair in each dummy direction, and the remaining mass $1/2 - 3\alpha/2$ at zero. These masses are nonnegative and sum to one. All features have norm at most one, all labels have magnitude at most one, and the means vanish. Directly,

$$\Sigma_S = \operatorname{diag}(1/4, tI_q, t/2), \qquad w_S = e_1/2.$$

Delete the entire x_+ population, of mass α. Mirror pairs preserve centering, and the retained label variance is $(1/4 - \alpha/4)/(1-\alpha) = 1/4$, equal to its original value.

These populations are exact finite multisets. At base size $n_0 = 2/t$, each sign of either anchor has $1/\alpha$ copies, each sign of e_1 has $q - 1/(2\alpha)$ copies, each sign of each dummy has one copy, and zero has $n_0(1/2 - 3\alpha/2)$ copies, all integers. Duplicate the base sample $1/\alpha$ times to obtain the stated n. Since $\sqrt{\alpha} = 2^{-j}$, the feature and label grid requires only $O(\log d)$ bits. All records and the request can be public.

For $1 \leq \ell \leq q = d-2$, the next spectral floor satisfies $\delta/n = t$. Ties among dummy eigenvalues do not matter: every such rank has normalized covariance error

$$E = (\widehat{G} - G_S)/n = (t/2)e_b e_b^\top.$$

In particular $Ew_S = 0$, so the centered and ordinary decoders coincide. Set $\beta = 1-\alpha$, $c = \alpha^{3/2}/4$, $a = \beta/2$, and $C = \alpha x_+ x_+^\top$. On coordinates (e_1, e_b) the retained normal matrix and response, normalized by the original sample size, are

$$Q' = \Sigma_S - C + \beta\lambda I = \begin{pmatrix} a & -c \\ -c & (\beta+t)/4 \end{pmatrix}, \qquad g' = \begin{pmatrix} \beta/4 \\ -c \end{pmatrix}.$$

Restrict all subsequent matrices and vectors to these two coordinates. Here $Q' \succ 0$, and $g' - Q'w_S = (0, -c/2)^\top$. Writing $\Delta_0 = \det Q'$ and $\Delta_1 = \det(Q'+E)$, the e_b coordinate of w_{S-U} is therefore $h = -ac/(2\Delta_0)$. The decoder error and true retained curvature are respectively

$$\widehat{w}_U - w_{S-U} = -(Q'+E)^{-1}Ew_{S-U}, \qquad Q'/\beta.$$

More generally, replace the covariance error $t/2$ along e_b by any $e \geq 0$. Since $(Q' + ee_be_b^\top)^{-1}e_b = (c,a)^\top/(\Delta_0 + ae)$ and $(c,a)Q'(c,a)^\top = a\Delta_0$, the excess is exactly

$$\frac{e^2a^3c^2}{4\beta\Delta_0(\Delta_0+ae)^2}.$$

For $e = t/2$ this becomes

$$\frac{t^2 a^3 c^2}{16\beta\Delta_0\Delta_1^2} = \left(\frac{1}{4} + o(1)\right)\alpha^7,$$

because $a \to 1/2$, $\beta \to 1$, $\Delta_0, \Delta_1 \to 1/8$, and $c^2 = \alpha^3/16$. The scalar expression is increasing in e. At rank zero the floor is $\delta/n = 1/4 = \lambda$, so its error still annihilates w_S and has $e = \lambda - t/2 \geq t/2$ in the e_b direction. The dummy block is decoupled and has zero response, so it contributes no error. Finally, at rank $q + 1 = d - 1$ the next floor is $t/2$, making $\widehat{G} = G_S$. For sufficiently large j, the displayed excess exceeds $\alpha^7/8$, proving the rank claim. □

Proposition 117 rules out attaining the rank scale in (433) uniformly in the deletion fraction merely by retuning the rank of the present spectral floor. For sufficiently large j, $\alpha \leq \lambda = 1/4$ and $\varepsilon = \alpha^7/8 \leq c_{\mathrm{ra}}\alpha^2$, so the comparison is within the rare-anchor theorem's accuracy regime. The minimax question remains open: this fully specified family supplies no memory lower bound, and its required floor rank does not imply an $\Omega(d^2)$ bit bound.

S MEMORY AS A FUNCTION OF THE ORIGINAL REGULARIZED FIT

Define the original optimal regularized training objective by

$$L(S) = F_S(w_S) = \min_{v\in\mathbb{R}^d}\left\{\frac{1}{n}\left\|X_S v - y_S\right\|^2 + \lambda\left\|v\right\|^2\right\}. \tag{457}$$

This quantity is determined before the deletion request; it is neither a retained-data loss nor a measure defined through the deletion decoder. In particular, $0 \leq L(S) \leq F_S(0) \leq 1$. We first give an instance-dependent upper bound, and then consider the class with a public promise $L(S) \leq L_0 \leq 1$.

S.1 A LOSS-SENSITIVE TAIL BOUND

Lemma 118 (Original-fit-loss bound)**.** *Use the spectral summary* (7) *and decoder* (8)*. For every nonempty $T = S - U$, write $m = |T|$, $\beta = m/n$, and $\tau = \delta/m$. Simultaneously for every legal deletion,*

$$F_T(\widehat{w}_U) - F_T(w_T) \leq \frac{L(S)}{\beta}\left(\frac{\tau}{\lambda}\right)^3. \tag{458}$$

Proof. Restricting the original residual sum at w_S gives $\beta F_T(w_S) \leq F_S(w_S) = L(S)$: the retained data contribution is a subsum, and its weighted regularizer is $\beta\lambda\left\|w_S\right\|^2$. Thus $F_T(w_T) \leq L(S)/\beta$. Substitution into Lemma 8 proves the stated excess bound. For use with quantization, equation (30) also gives

$$\left\|\Pi w_T\right\|^2 \leq \frac{\tau L(S)}{\beta\lambda^2}. \tag{459}$$

Neither w_S nor $L(S)$ is required by this decoder; they occur only in the bound. □

For exact linear labels with a teacher of norm at most R, $L(S) \leq \lambda R^2$. Thus the lemma recovers the order $R^2\tau^3/\lambda^2$ of the source-condition bound in Appendix A.6; that direct proof has the sharper constant $1/4$. No noiselessness assumption is needed in Lemma 118.

S.2 A PUBLIC LOSS PROMISE AND FINITE-BIT SUMMARIES

Proposition 119 (Finite precision under a loss promise)**.** *Fix $0 < \beta_0 \leq 1$, $0 < \lambda \leq 1$, $0 < L_0 \leq 1$, and $0 < \varepsilon \leq L_0$. Constants may depend on the fixed β_0. Suppose $L(S) \leq L_0$, all coordinates and labels belong to a common p_{in}-bit dyadic grid, and $m \geq \beta_0 n$. There is a deterministic summary satisfying excess at most ε simultaneously for all such deletions. Its numerical storage scale is*

$$d\min\left\{d, \frac{1}{\beta_0\lambda}\left(\frac{L_0}{\varepsilon}\right)^{1/3}\right\}. \tag{460}$$

More precisely, set $c_0 = 1/\beta_0$ and

$$q = \left\lceil \frac{1}{\beta_0 \lambda} \left(\frac{4c_0 L_0}{\varepsilon} \right)^{1/3} \right\rceil. \tag{461}$$

If $q \geq d$, exact statistics require $O(d^2(p_{\text{in}} + \log(2+n)))$ bits. Otherwise, with $\ell = q$, a sufficient bit count is

$$O\left(d(\ell+1) \log\left(2 + nd + \frac{nd}{\beta_0^{3/2}\lambda} \sqrt{\frac{L_0}{\varepsilon}} \right) + dp_{\text{in}} \right). \tag{462}$$

For fixed β_0, $n = \Theta(d)$, and $0 < \lambda \leq 1$, both branches are covered by

$$O\left(d \min\left\{ d, \frac{1}{\lambda} \left(\frac{L_0}{\varepsilon} \right)^{1/3} \right\} (\log d + p_{\text{in}}) \right). \tag{463}$$

If $L_0 = 0$, the zero vector is exact and no data-dependent state is needed.

Proof. In the low-rank branch, form K, δ as in (7). For the rounding error we also use

$$F_T(w_T) \leq F_T(w_S) \leq \frac{L(S)}{\beta}, \qquad \|w_T\| \leq \sqrt{\frac{L(S)}{\beta\lambda}}. \tag{464}$$

The first bound follows by restricting the residual sum at w_S; the regularization term also increases by at most the factor $1/\beta$.

Put

$$a = \frac{\beta_0 n \lambda}{2} \sqrt{\frac{\beta_0 \varepsilon}{L_0}}, \qquad a' = \min\{a, n\}. \tag{465}$$

Choose the largest dyadic power h not exceeding $a'/(16nd)$. Round each entry of K to the nearest multiple of h, obtaining K_q, and round δ upward to a multiple δ_q of h. Set

$$\xi = hnd + h^2 d^2/4, \qquad \widehat{G}_q = K_q^\top K_q + (\delta_q + \xi) I. \tag{466}$$

Because $\|K\|_F \leq \sqrt{n}$ and $\ell < d$,

$$\left\| K_q^\top K_q - K^\top K \right\|_{\text{op}} \leq h\sqrt{n\ell d} + h^2 \ell d/4 \leq \xi. \tag{467}$$

Thus $\widehat{G}_q - G_S = E_0 + J$ with $0 \preceq J \preceq rI$, where $r = 2\xi + h$. Our choice of h gives $\xi \leq a'/8$, $h \leq a'/16$, and hence $r \leq a' \leq a$. In particular the quantized retained system remains positive definite.

Apply Lemma 7 to $(E_0 + J)/m$ and use the triangle inequality, (459), and (464). The quantized decoder obeys

$$\sqrt{F_T(\widehat{w}_q) - F_T(w_T)} \leq \sqrt{\frac{L(S)}{\beta}} \left(\frac{\delta}{m\lambda} \right)^{3/2} + \frac{r}{m\lambda} \sqrt{\frac{L(S)}{\beta}}. \tag{468}$$

Since $\delta \leq n/(\ell+1)$, (461) bounds the first term by $\sqrt{\varepsilon}/2$. The inequality $r \leq a$ bounds the second by $\sqrt{\varepsilon}/2$.

The stored state consists of a branch flag, n, ℓ, the rounded factor, the rounded scalar, and exact b_S. The scale h and correction ξ are determined by n, ℓ and the public parameters. Entries of K_q have magnitude at most $\sqrt{n} + h/2$, and $\delta_q \leq n + h$. Since

$$\frac{1}{h} = O\left(\max\left\{ d, \frac{d}{\beta_0^{3/2}\lambda} \sqrt{\frac{L_0}{\varepsilon}} \right\} \right), \tag{469}$$

their encodings give the logarithm in (462). Every entry of b_S has magnitude at most n and denominator dividing $2^{2p_{\text{in}}}$, accounting for the remaining term. The exact branch uses the same denominator bound for G_S.

For the simplified bound, write $t = \varepsilon/L_0 \leq 1$. In the low-rank branch, $q < d$ forces $\lambda t^{1/3} > c/d$, for a constant depending only on β_0. Thus $\lambda \geq c_1/d$ and $t \geq c_2/d^3$ when $\lambda \leq 1$; the covariance-quantization logarithm is $O(\log d)$. The exact branch is covered because $q \geq d$ forces the minimum in (463) to be $\Omega(d)$. Finally, $L(S) = 0$ and $\lambda > 0$ imply $w_S = 0$ and all labels zero. □

The public promise avoids storing an infinite-precision value of $L(S)$. Corollary 11 gives relative optimization accuracy without such a promise. Here the promise selects a smaller rank for a specified absolute error tolerance. An instance-adaptive encoder could instead store a dyadic upper estimate $L(S) \le \overline{L} \le 2L(S)$ and use it in place of L_0; its exponent and the selected rank must be included in the state. We use the public promise for the instance-class statement below.

S.3 Label scaling and the corresponding instance class

For $0 < L_0 \le 1$, define the common dyadic precision

$$p_*(d, L_0) = \max\left\{p_{\mathrm{in}}(d), \left\lceil \tfrac{1}{2}\log_2(1/L_0)\right\rceil\right\}, \tag{470}$$

where $p_{\mathrm{in}}(d)$ is the base precision in (4). Let $B^*_{L_0,p_*}(d, \lambda, \varepsilon)$ denote the least worst-case summary length for $d \le n \le 3d$, bounded features and labels on this grid, and the public promise $L(S) \le L_0$. The deletion and pointwise randomized-success requirements are those in (3).

Corollary 120 (Memory on a class with bounded original fit loss)**.** *There are universal constants $c, C, c_0 > 0$ such that, for every sufficiently large even d, $0 < \lambda \le 1$, $0 < L_0 \le 1$, and $0 < \varepsilon \le c_0 L_0$, writing*

$$W = d \min\left\{d, \frac{1}{\lambda}\left(\frac{L_0}{\varepsilon}\right)^{1/3}\right\}, \tag{471}$$

one has

$$cW \le B^*_{L_0,p_*}(d, \lambda, \varepsilon) \le CW\big(\log d + \log(1/L_0)\big). \tag{472}$$

The lower bound holds on families with $w_S = 0$, $L_0/8 \le L(S) \le 2L_0/3$, and common legal deletion requests.

Proof. Set

$$j = \left\lceil \tfrac{1}{2}\log_2(1/L_0)\right\rceil, \qquad s = 2^{-j}, \qquad L_0/4 < s^2 \le L_0. \tag{473}$$

For a dataset S, let sS denote the dataset obtained by multiplying every label by s and leaving every feature unchanged. For every retained multiset T and every w, direct substitution gives

$$F_{sT}(sw) = s^2 F_T(w), \qquad w_{sT} = s w_T, \qquad L(sS) = s^2 L(S). \tag{474}$$

Apply the lower-bound construction for Theorem 1 at accuracy $\varepsilon' = \varepsilon/s^2$. Its original optimizer is zero. The square family of Appendix C has half of its labels equal to ± 1, and the rectangular family of Appendix D has two thirds of its labels equal to ± 1; all other labels are zero. Rounding changes only zero-label features. Therefore their original objectives are exactly $1/2$ and $2/3$, respectively. After scaling,

$$L_0/8 \le L(sS) \le 2L_0/3 \le L_0, \qquad w_{sS} = 0. \tag{475}$$

Scaled labels belong to $\{0, \pm s\}$, so all transformed coordinates and labels lie exactly on the grid (470). No additional approximation is needed.

An encoder for this loss-constrained class would give an encoder for the base hard family by applying it to sS. On a base query U, the decoder supplies the scaled query sU and divides the returned vector by s. Equation (474) shows that a successful answer has base excess at most ε/s^2. All queries remain legal and common across the family, and their scaling is public and independent of the hidden instance. The reduction leaves the state length and pointwise success probability unchanged.

Choose $c_0 \le \varepsilon_0/4$, where ε_0 is the constant in Theorem 1. Then $\varepsilon' \le 4\varepsilon/L_0 \le \varepsilon_0$, so that theorem implies

$$B \ge c_1 d \min\left\{d, \frac{s^{2/3}}{\lambda \varepsilon^{1/3}}\right\} \ge c_2 d \min\left\{d, \frac{(L_0/\varepsilon)^{1/3}}{\lambda}\right\}. \tag{476}$$

This proves the lower bound. The upper bound is Proposition 119 with $\beta_0 = 2/3$ and $p_{\mathrm{in}} = p_*(d, L_0) = O(\log d + \log(1/L_0))$. □

The grid qualification is part of the corollary. Arbitrarily small L_0 can require more than $O(\log d)$ input bits, so (472) does not assert only a $\log d$ gap uniformly in L_0. If $L_0 \ge d^{-C_0}$ for a fixed constant C_0, the required precision is $O_{C_0}(\log d)$; it need not be the particular fixed grid in (4). The result concerns bounded labels on the refined dyadic grid, rather than requiring labels to stay in $\{-1, 0, 1\}$. Finally, the matching statement is restricted to $\varepsilon \le c_0 L_0$: (464) already makes reuse of the original model sufficient at suitably larger error tolerances.

T ASSESSING A DELETION ANSWER FROM ITS SUMMARY

The uniform cubic bound can be sharpened after the deletion request is known. The summary's spectral error envelope controls the unknown quadratic residual, yielding a request-specific upper bound on optimization error. We call this bound an *error certificate*: its target is optimization utility, not distributional unlearning. We derive the exact-arithmetic bound first, then account for finite precision.

Fix $T = S - U$, $k = |U|$, $m = n - k > 0$, and write

$$Q = G_T/m + \lambda I, \qquad P = (\widehat{G} - G_S)/m, \qquad \widehat{Q} = Q + P, \qquad \tau = \delta/m.$$

For the floor in (7), $0 \preceq P \preceq \tau\Pi$ and $P = \Pi P \Pi$. The projector can be taken onto the complement of the row space of K: if a leading eigenvalue equals δ, its factor row is zero and the corresponding direction also has covariance at most δ.

T.1 CENTERING CONTRACTS THE ERROR OF THE STORED MODEL

Proposition 121 (Contraction in the retained objective)**.** *For the retained Hessian $Q \succeq \lambda I$, any error $0 \preceq P \preceq \tau I$, and reference vector θ, define*

$$\widehat{w} = (Q + P)^{-1}(Q w_T + P\theta). \tag{477}$$

Then

$$F_T(\widehat{w}) - F_T(w_T) \le \left(\frac{\tau}{\lambda + \tau}\right)^2 [F_T(\theta) - F_T(w_T)]. \tag{478}$$

The matrices Q and P need not commute. In particular, the exact centered floor (509) *never has larger retained objective than reusing $\theta = w_S$.*

Proof. Let $e = \theta - w_T$ and $C = Q^{-1/2} P Q^{-1/2}$. The error map satisfies

$$\widehat{w} - w_T = (Q + P)^{-1} P e,$$
$$Q^{1/2}(Q + P)^{-1} P Q^{-1/2} = (I + C)^{-1} C.$$

Here $0 \preceq C \preceq (\tau/\lambda) I$. The matrix $(I + C)^{-1} C$ is symmetric, with eigenvalues $c/(1 + c)$ for eigenvalues c of C; its operator norm is therefore at most $\tau/(\lambda + \tau)$. Apply this bound to $Q^{1/2} e$ and use $F_T(v) - F_T(w_T) = \left\| Q^{1/2}(v - w_T) \right\|^2$. □

This comparison is to the chosen reference model. It does not order centered and uncentered flooring, or guarantee monotonic error as the stored rank changes.

Remark 122 (Centering need not preserve relative accuracy). The relative guarantee for the uncentered decoder does not extend to centered flooring. For an exact dyadic example, take $d = 3$, $n = 9$, with three copies of $(e_1, 0)$, two of $(e_2, 0)$, one of $(e_3, 1)$, and three of $(0, 0)$. These records are noiseless for the public norm-one teacher $v = e_3$. At rank $\ell = 1$,

$$G_S = \mathrm{diag}(3, 2, 1), \qquad \delta = 2, \qquad E_0 = \mathrm{diag}(0, 0, 1), \qquad \theta = w_S = \frac{e_3}{1 + 9\lambda}.$$

Delete the single record $(e_3, 1)$, a legal request with $m = 8$. All retained labels are zero, so $w_T = 0$ and $F_T(w_T) = 0$, whereas

$$\widehat{w}_{\mathrm{cf}} = \frac{e_3}{(1 + 9\lambda)(1 + 8\lambda)}, \qquad F_T(\widehat{w}_{\mathrm{cf}}) - F_T(w_T) = \frac{\lambda}{(1 + 9\lambda)^2 (1 + 8\lambda)^2} > 0.$$

The uncentered answer is exact. Embedding these coordinates and adding zero records gives the same obstruction for every $d \ge 3$ with $n = 3d$; the factors 9 and 8 become $3d$ and $3d - 1$. Thus contraction toward the retained optimum alone cannot establish a uniform multiplicative relative guarantee, even for bounded noiseless data.

T.2 Absolute certificates from a deletion request

Proposition 123 (Tail and deletion-gradient certificates)**.** *Suppose $\theta = w_S$ is stored exactly and P is the exact spectral-floor error above. Set*

$$g = \frac{b_U - G_U\theta - k\lambda\theta}{m}, \qquad a = -\widehat{Q}^{-1}g, \qquad \widehat{w}_{\mathrm{cf}} = \theta + a. \tag{479}$$

Using only the summary and deleted records, compute

$$C_{\mathrm{tail}} = \frac{\tau^2}{\lambda}\left\|\Pi a\right\|^2, \qquad C_{\mathrm{grad}} = \left(\frac{\tau}{\lambda+\tau}\right)^2 \frac{\|g\|^2}{\lambda},$$
$$C_{\mathrm{cf}} = \min\{C_{\mathrm{tail}}, C_{\mathrm{grad}}\}. \tag{480}$$

Then $F_T(\widehat{w}_{\mathrm{cf}}) - F_T(w_T) \le C_{\mathrm{cf}}$. For the uncentered floor answer $\widehat{w}_{\mathrm{f}}$ in (8)*, the corresponding bound is*

$$F_T(\widehat{w}_{\mathrm{f}}) - F_T(w_T) \le C_{\mathrm{f}} := \frac{\tau^2}{\lambda}\left\|\Pi\widehat{w}_{\mathrm{f}}\right\|^2. \tag{481}$$

Proof. Original stationarity, $(G_S + n\lambda I)\theta = b_S$, gives $g = Q\theta - b_T/m$. Thus (479) is the centered decoder and $w_T - \theta = -Q^{-1}g$. Since $\widehat{Q}a = -g$,

$$Q(\widehat{w}_{\mathrm{cf}} - w_T) = -Pa.$$

The exact quadratic excess is consequently $a^\top PQ^{-1}Pa \le \|Pa\|^2/\lambda \le \tau^2\|\Pi a\|^2/\lambda$. Also $F_T(\theta) - F_T(w_T) = g^\top Q^{-1}g \le \|g\|^2/\lambda$; Proposition 121 proves the gradient bound. For the uncentered answer, $\widehat{Q}\widehat{w}_{\mathrm{f}} = b_T/m$ implies $Q(\widehat{w}_{\mathrm{f}} - w_T) = -P\widehat{w}_{\mathrm{f}}$, giving the last bound by the same argument. □

For example, a request satisfying $g = 0$ leaves the exact optimum unchanged, and the centered certificate is zero at every stored rank. More generally, little movement in the omitted subspace can make the tail certificate small even when the worst-case cubic bound is large. The certificates upper-bound error; failure to meet a tolerance is inconclusive. The centered gradient bound does not apply to the uncentered answer.

T.3 A certificate for the midpoint with PSD repair

The same residual argument also applies to the midpoint-plus-PSD comparator used in the experiments. In this subsection let $\Pi = I - V_\ell V_\ell^\top$ refer to its stored leading eigenspace, including any tied leading directions chosen by that comparator. Define

$$G_{\mathrm{mid}} = V_\ell\,\mathrm{diag}_{i\le\ell}(\sigma_i)V_\ell^\top + \frac{\delta}{2}\Pi, \qquad E_{\mathrm{mid}} = G_{\mathrm{mid}} - G_S. \tag{482}$$

The ordinary leading factor and δ determine G_{mid} and, when $\delta > 0$, Π. If $\delta = 0$, $G_{\mathrm{mid}} = G_S$ and terms involving Π vanish. The error is supported in the omitted subspace and satisfies $\|E_{\mathrm{mid}}a\| \le (\delta/2)\|\Pi a\|$ for every a. For a symmetric matrix, $(\cdot)_+$ denotes replacement of negative eigenvalues by zero. The request determines the repair and solve matrix

$$J_U = (G_U - G_{\mathrm{mid}})_+, \qquad H_{\mathrm{psd}} = \frac{G_{\mathrm{mid}} - G_U + J_U}{m} + \lambda I = \frac{(G_{\mathrm{mid}} - G_U)_+}{m} + \lambda I \succeq \lambda I. \tag{483}$$

Proposition 124 (Posterior certificate for a repaired midpoint)**.** *Let the true retained half-gradient $g = Q\theta - b_T/m$ be available exactly, and set $a = -H_{\mathrm{psd}}^{-1}g$. Then*

$$F_T(\theta + a) - F_T(w_T) \le C_{\mathrm{mid}} := \frac{\left[(\delta/2)\|\Pi a\| + \|J_U a\|\right]^2}{m^2\lambda}. \tag{484}$$

For $\theta = w_S$, equation (479) *supplies g from the stored model and deleted records. The same bound applies to the uncentered answer with $\theta = 0$ and $g = -(b_S - b_U)/m$.*

Proof. Since $H_{\mathrm{psd}} = Q + (E_{\mathrm{mid}} + J_U)/m$ and $H_{\mathrm{psd}}a = -g$, one has

$$Q(\theta + a - w_T) = -\frac{(E_{\mathrm{mid}} + J_U)a}{m}.$$

Apply the exact quadratic excess identity, $Q^{-1} \preceq \lambda^{-1}I$, and the triangle inequality with the bound on $E_{\mathrm{mid}}a$. □

Every term in (484) is available from the summary and request; J_U is temporary decoding workspace. The error $E_{\mathrm{mid}} + J_U$ need not be positive semidefinite, so the floor's contraction proposition does not apply to this comparator. Neither certificate establishes a uniform accuracy ordering between the two methods. With an inexact solve, add $m\,\|H_{\mathrm{psd}}a + g\|$ inside the numerator's brackets; reference and covariance rounding require their own error bounds as described below.

T.4 Relative certificates from one additional scalar

Store additionally $L = F_S(\theta)$, still assuming $\theta = w_S$ exactly. The value of the retained objective at that reference is then available:

$$F_\theta := F_T(\theta) = \frac{nL - \sum_{(x,y)\in U}(y - x^\top\theta)^2 - k\lambda\,\|\theta\|^2}{m}. \tag{485}$$

This adds one numerical scalar; it is not part of the base summary's storage count.

Proposition 125 (Relative accuracy from a lower bound on the optimum)**.** *With the exact quantities above, put*

$$J = F_\theta - g^\top\widehat{Q}^{-1}g, \qquad \underline{F} = \max\{0, J - \tau\,\|\Pi a\|^2 - C_{\mathrm{cf}}\}. \tag{486}$$

Then $\underline{F} \le F_T(w_T)$. Hence, for any $\zeta > 0$, the check $C_{\mathrm{cf}} \le \zeta\underline{F}$ implies $F_T(\widehat{w}_{\mathrm{cf}}) \le (1+\zeta)F_T(w_T)$.

Proof. Expand the objective at $\theta + a$ and use $\widehat{Q}a = -g$:

$$F_T(\theta + a) = F_\theta + 2g^\top a + a^\top Qa = J - a^\top Pa \ge J - \tau\,\|\Pi a\|^2.$$

Subtract the upper bound C_{cf} on its excess. Since $F_T(w_T)$ is nonnegative, taking the maximum with zero preserves the lower bound. The relative conclusion follows from $F_T(\widehat{w}_{\mathrm{cf}}) - F_T(w_T) \le C_{\mathrm{cf}} \le \zeta\underline{F} \le \zeta F_T(w_T)$. □

The lower bound can be zero even when the optimum is positive. In that case this test can certify only zero excess; it does not assert that relative accuracy is impossible.

An alternative scalar gives a lower bound that remains positive whenever the retained labels are nonzero. Retain $Y_S = \sum_S y^2$ in place of the additional scalar L, and compute

$$v_T = \frac{Y_S - \sum_U y^2}{m}. \tag{487}$$

This option does not require $F_S(\theta)$ or an exact original optimizer to obtain the lower bound; an absolute error certificate must still satisfy its own assumptions.

Proposition 126 (Relative accuracy from retained label energy)**.** *Let h be a known upper bound on $\|Q\|_{\mathrm{op}}$. Bounded features always permit $h = 1 + \lambda$. If an available matrix Q_{up} satisfies $Q \preceq Q_{\mathrm{up}}$, one may instead take*

$$h = \min\{1 + \lambda, \|Q_{\mathrm{up}}\|_{\mathrm{op}}\}. \tag{488}$$

Then

$$\underline{F}_{\mathrm{energy}} := \frac{\lambda}{h}v_T \le F_T(w_T). \tag{489}$$

For any answer $\widetilde{w}$ with a valid absolute certificate $F_T(\widetilde{w}) - F_T(w_T) \le C$, the check $C \le \zeta\underline{F}_{\mathrm{energy}}$ implies $F_T(\widetilde{w}) \le (1 + \zeta)F_T(w_T)$. On a p_{in}-bit dyadic input grid, Y_S can be stored exactly using $O(p_{\mathrm{in}} + \log(2 + n))$ additional bits.

Proof. Completing the ridge quadratic and using the dual normal equations give

$$F_T(w_T) = \frac{\lambda}{m} y_T^\top (X_T X_T^\top/m + \lambda I)^{-1} y_T.$$

The largest eigenvalue of $X_T X_T^\top/m + \lambda I$ is at most h, because $\left\|X_T X_T^\top\right\|_{\mathrm{op}} = \|G_T\|_{\mathrm{op}}$ and $\|Q\|_{\mathrm{op}} \le h$. Inverting proves the lower bound; combining it with C proves the relative implication. Finally, $Y_S \in [0, n]$ has denominator dividing $2^{2p_{\mathrm{in}}}$, which gives its bit count. □

For flooring, $Q_{\mathrm{up}} = \widehat{Q}$ is valid, including the upward rounding of Appendix A.4. The legal upper matrix Q_{hi} in (496) is another option. A midpoint or PSD-repaired solve matrix is not automatically an upper bound on Q; without a proved upper matrix, use $h = 1 + \lambda$. The energy lower bound is strictly positive when $v_T > 0$, but can be conservative and need not improve (486). If both optional scalars are stored, the maximum of the two lower bounds is valid. This check concerns the returned answer, and does not remove the centered relative-accuracy obstruction in Remark 122.

T.5 Finite precision and the implementation boundary

The preceding identities are exact-arithmetic statements. In particular, replacing w_S by a rounded vector without adjusting the right-hand side invalidates the stationarity identity used to obtain g from deleted records. If the stored reference is θ, define its original stationarity residual

$$r_S = (G_S + n\lambda I)\theta - b_S.$$

The true retained half-gradient is the expression in (479) plus r_S/m. A norm bound on this additional term must enter the residual certificate if the decoder still uses the uncorrected expression. For example, a known rounding bound $\|\theta - w_S\| \le s_\theta$ implies $\|r_S\|/m \le (n/m)(1+\lambda)s_\theta$, since $\|G_S\|_{\mathrm{op}} \le n$. Likewise, $F_S(w_S)$ cannot replace $F_S(\theta)$ in (485) after model rounding.

One finite-bit way to avoid this stationarity approximation is to choose a dyadic reference θ during encoding and additionally retain

$$z_S = b_S - G_S\theta, \qquad R_S = \sum_{(x,y)\in S} (y - x^\top\theta)^2. \tag{490}$$

For dyadic data these are finite rational quantities that can be stored exactly. They cost an additional d-vector and one scalar. If the input and reference fractional precisions are p_{in} and p_θ, their total size is $O(d[p_{\mathrm{in}} + p_\theta + \log(2+n) + \log(2+\|\theta\|)])$ bits. Indeed, z_S has denominator dividing $2^{2p_{\mathrm{in}}+p_\theta}$ and magnitude per coordinate at most $n(1+\|\theta\|)$; R_S has denominator dividing $2^{2p_{\mathrm{in}}+2p_\theta}$ and is at most $n(1+\|\theta\|)^2$. The decoder can now obtain the true quantities for this reference from

$$\begin{aligned} g &= \frac{m\lambda\theta - z_S + b_U - G_U\theta}{m}, \\ F_\theta &= \frac{R_S - \sum_U (y - x^\top\theta)^2}{m} + \lambda\|\theta\|^2. \end{aligned} \tag{491}$$

The reference need not be an exact original optimizer. With this true g, solving $\widehat{Q}a = -g$ again satisfies (477). As throughout the exact-arithmetic model, the public λ is treated exactly in these identities.

For covariance rounding of the floor, write the implemented error as $P = P_0 + P_1$, where $0 \preceq P_0 \preceq \tau\Pi$ and $0 \preceq P_1 \preceq \nu I$ for a known ν; the upward rounding in Appendix A.4 provides such a full-space correction. For a possibly inexact solve let $s = \widehat{Q}a + g$. Then

$$\begin{aligned} Q(\theta + a - w_T) &= s - Pa, \\ F_T(\theta + a) - F_T(w_T) &\le \frac{[\tau\|\Pi a\| + \nu\|a\| + \|s\|]^2}{\lambda}. \end{aligned} \tag{492}$$

This follows from the exact excess identity and $Q^{-1} \preceq \lambda^{-1} I$. If only the uncorrected model gradient $\widetilde{g}$ is available, replace s by the measured residual $\widehat{Q}a + \widetilde{g}$ and add $\|r_S\|/m$ inside the brackets. For an exact solve with the true gradient g, the contraction also holds with $\tau + \nu$ in place of τ.

A retained rounded factor does not by itself supply the exact original projector Π. To evaluate (492), the implementation needs a proved upper bound on $\|\Pi a\|$. The universal choice $\|a\|$ requires no subspace metadata but loses the tail refinement. A stored approximation $\widetilde{\Pi}$ with a certified operator error ω permits $\|\Pi a\| \le \left\|\widetilde{\Pi} a\right\| + \omega \|a\|$; that representation and its precision must also be charged to memory.

With exact g, F_θ and any computed a, define $D = F_\theta + 2g^\top a + a^\top \widehat{Q} a$ and let C be a valid bound in (492). The same expansion gives

$$F_T(w_T) \ge \max\{0, D - \tau \|\Pi a\|^2 - \nu \|a\|^2 - C\}. \tag{493}$$

Certified upper bounds on the subspace norm can again be substituted. The bound accounts explicitly for residual statistics, covariance error, subspace error, and solve error. Numerical certification additionally requires rigorous roundoff enclosures for the evaluated quantities. The experiments evaluate the ideal formulas in floating point, without these enclosures or the additional residual-statistics representation.

U What a legal request reveals about the omitted covariance

A deleted feature matrix supplies more information than the term to subtract. Its legality requires the unknown retained Gram matrix to be positive semidefinite. With an exact spectral summary, this requirement gives a computable lower bound on the omitted covariance. We use the classical generalized Schur-complement criterion (Carlson et al., 1974) to obtain this bound. The resulting interval supports a stable midpoint decoder and improves the floor's certificates as well.

U.1 An interval determined by the summary and request

We first use exact arithmetic and the exact summary in (7). Let V be an orthonormal basis for the nonzero row space of K, and let Z be any orthonormal basis for its orthogonal complement. Write $q = \operatorname{rank} K$, $O = [V\ Z]$, and $\Pi = ZZ^\top$. Then

$$O^\top G_S O = \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}, \qquad 0 \preceq C \preceq \delta I, \qquad A = V^\top K^\top K V + \delta I. \tag{494}$$

Thus A is available from the summary. Using the nonzero row space handles tied eigenvalues: any selected direction with eigenvalue δ has a zero factor weight and may be included in Z. The covariance on this enlarged complement still has eigenvalues at most δ. The empty leading block when $q = 0$ is allowed.

For a legal request U, set $m = n - |U| > 0$ and partition its Gram matrix:

$$O^\top G_U O = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}, \qquad B = A - D_{11}.$$

All matrices G_S, G_U, A, C, D_{ij} in this subsection are unnormalized.

Proposition 127 (Covariance interval from legality)**.** *Let $B^\dagger$ denote the Moore–Penrose pseudoinverse. Every legal request satisfies*

$$B \succeq 0, \qquad (I - BB^\dagger) D_{12} = 0, \qquad L_U := D_{22} + D_{21} B^\dagger D_{12} \preceq C \preceq \delta I. \tag{495}$$

In particular $0 \preceq L_U \preceq \delta I$. Conversely, with the displayed B and range condition, every C in this interval yields a positive-semidefinite retained Gram matrix with these blocks. Define

$$\begin{aligned} &W_U = \delta I - L_U, \\ &G_{\mathrm{lo}} = O \operatorname{diag}(A, L_U) O^\top, \qquad G_{\mathrm{hi}} = O \operatorname{diag}(A, \delta I) O^\top, \\ &Q_{\mathrm{lo}} = (G_{\mathrm{lo}} - G_U)/m + \lambda I, \qquad Q_{\mathrm{hi}} = (G_{\mathrm{hi}} - G_U)/m + \lambda I. \end{aligned} \tag{496}$$

Then $G_{\mathrm{hi}} = \widehat{G}$ is the spectral floor and

$$\lambda I \preceq Q_{\mathrm{lo}} \preceq Q_T \preceq Q_{\mathrm{hi}}, \qquad Q_{\mathrm{hi}} - Q_{\mathrm{lo}} = Z W_U Z^\top / m.$$

Proof. Legality gives

$$O^\top (G_S - G_U) O = \begin{pmatrix} B & -D_{12} \\ -D_{21} & C - D_{22} \end{pmatrix} \succeq 0.$$

Its leading principal block is PSD. If $u \in \ker B$, testing its quadratic form on (tu, z) for both signs and arbitrary magnitude of t gives $u^\top D_{12} z = 0$. Hence $\mathrm{range}(D_{12}) \subseteq \mathrm{range}(B)$, which is the range condition in (495). For every u, z, that condition gives the exact completion of squares

$$\begin{aligned} &u^\top B u - 2u^\top D_{12} z + z^\top (C - D_{22}) z \\ &\quad = (u - B^\dagger D_{12} z)^\top B (u - B^\dagger D_{12} z) + z^\top (C - L_U) z. \end{aligned}$$

This proves both directions of the generalized Schur criterion, including singular B. Also $D_{22} \succeq 0$ and $D_{21} B^\dagger D_{12} \succeq 0$, so $L_U \succeq 0$. Choosing $C = L_U$ proves $G_{\mathrm{lo}} - G_U \succeq 0$; the remaining orders follow by adding the tail PSD differences and λI. For $q = 0$, the formula reads $L_U = G_U$ in the tail basis. □

The converse is a statement about compatible Gram matrices; it does not assert that every completion is realizable by the prescribed bounded, quantized records or by the stored labels. The interval is an available outer description of the unknown covariance. It uses features only.

U.2 A STABLE MIDPOINT WITH TAIL-SUPPORTED ERROR

Choose the midpoint after the request is known:

$$G_{\mathrm{lm}} = \tfrac{1}{2}(G_{\mathrm{lo}} + G_{\mathrm{hi}}), \qquad H_U = \tfrac{1}{2}(Q_{\mathrm{lo}} + Q_{\mathrm{hi}}), \qquad \widehat{w}_{\mathrm{lm}} = H_U^{-1}(b_S - b_U)/m. \tag{497}$$

Unlike a fixed midpoint followed by a PSD projection, this decoder keeps its covariance error in the omitted subspace. It satisfies $H_U \succeq \lambda I$ for every legal request, without a condition on $\delta/(m\lambda)$.

Proposition 128 (Relative cubic error of the legal midpoint)**.** *For every bounded dataset and legal request with $m > 0$, the exact decoder* (497) *satisfies*

$$F_T(\widehat{w}_{\mathrm{lm}}) - F_T(w_T) \le \frac{1}{4} F_T(w_T) \left(\frac{\delta}{m\lambda} \right)^3, \qquad T = S - U. \tag{498}$$

Proof. Put $Q = Q_T$, $H = H_U$, and $P = H - Q$. The interval gives

$$P = Z P_{\mathrm{tail}} Z^\top, \qquad -W_U/(2m) \preceq P_{\mathrm{tail}} \preceq W_U/(2m), \qquad \|P\|_{\mathrm{op}} \le \delta/(2m). \tag{499}$$

Although P need not be PSD, $Q \preceq Q_{\mathrm{hi}} = 2H - Q_{\mathrm{lo}} \preceq 2H - \lambda I$. Consequently

$$H^{-1} Q H^{-1} \preceq 2H^{-1} - \lambda H^{-2} \preceq \lambda^{-1} I. \tag{500}$$

The last inequality follows from $2/s - \lambda/s^2 \le 1/\lambda$ for every eigenvalue s of H. Since $\widehat{w}_{\mathrm{lm}} - w_T = -H^{-1} P w_T$, the exact excess is at most $\|P w_T\|^2 / \lambda$. The retained tail-response estimate (11) applies to the complement in (494) and yields $\|\Pi w_T\|^2 \le \delta F_T(w_T)/(m\lambda^2)$. Combining these inequalities proves the claim. □

This result gives the same cubic rank dependence as flooring. Its constant $1/4$ is larger than the floor's universal constant $4/27$ obtained from Lemma 9; it is not an improvement to that worst-case bound. Nor does it order actual errors: an uncentered floor can be exact on a flat spectrum while a midpoint introduces bias.

U.3 CENTERED UPDATES AND CERTIFICATES FOR BOTH DECODERS

Let $\theta = w_S$ be stored exactly, let $k = |U|$, and compute the true retained half-gradient from the request:

$$g = \frac{b_U - (G_U + k\lambda I)\theta}{m} = Q_T(\theta - w_T).$$

The same conclusions below hold for any reference θ if its true g is available, for example through the residual statistics in Appendix T.5. Write $\mathcal{E}_T(w) = F_T(w) - F_T(w_T)$ and $v_U = \|W_U\|_{\mathrm{op}}/m$.

Proposition 129 (Centered legal midpoint)**.** *Set* $a_{\mathrm{lm}} = -H_U^{-1} g$*,* $a_{\mathrm{lm},t} = Z^\top a_{\mathrm{lm}}$*, and* $\widehat{w}_{\mathrm{lm},c} = \theta + a_{\mathrm{lm}}$*. Then*

$$\mathcal{E}_T(\widehat{w}_{\mathrm{lm},c}) \le \left(\frac{v_U}{2\lambda + v_U}\right)^2 \mathcal{E}_T(\theta), \tag{501}$$

$$\mathcal{E}_T(\widehat{w}_{\mathrm{lm},c}) \le C_{\mathrm{lm},U} := \min\left\{\frac{\|W_U\|_{\mathrm{op}}\, a_{\mathrm{lm},t}^\top W_U a_{\mathrm{lm},t}}{4m^2\lambda}, \left(\frac{v_U}{2\lambda + v_U}\right)^2 \frac{\|g\|^2}{\lambda}\right\}. \tag{502}$$

Thus the update never has larger retained objective than its reference.

Proof. Use Q, H, P from the preceding proof and set $J = ZW_UZ^\top/(2m)$. We have $-J \preceq P \preceq J$ and $H = Q_{\mathrm{lo}} + J \succeq \lambda I + J$. For $\gamma = v_U/(2\lambda + v_U)$, the scalar inequality on $0 \le s \le v_U/2$ gives $J \preceq \gamma(\lambda I + J) \preceq \gamma H$. Hence the eigenvalues of $H^{-1/2}PH^{-1/2}$ lie in $[-\gamma, \gamma]$.

The centered error is $\widehat{w}_{\mathrm{lm},c} - w_T = H^{-1}P(\theta - w_T)$. In the Q metric this map is the symmetric matrix $I - Q^{1/2}H^{-1}Q^{1/2}$. Its eigenvalues equal those of $I - H^{-1/2}QH^{-1/2} = H^{-1/2}PH^{-1/2}$, proving (501). Also $\mathcal{E}_T(\theta) = g^\top Q^{-1} g \le \|g\|^2/\lambda$.

The exact residual identity $Q(\theta + a_{\mathrm{lm}} - w_T) = -Pa_{\mathrm{lm}}$ bounds excess by $\|Pa_{\mathrm{lm}}\|^2/\lambda$. The signed order in (499) implies $mP_{\mathrm{tail}} = W_U^{1/2} R W_U^{1/2}/2$ for a symmetric contraction R, with the factorization restricted to $\mathrm{range}(W_U)$ if necessary. Therefore

$$\|Pa_{\mathrm{lm}}\|^2 \le \frac{\|W_U\|_{\mathrm{op}}}{4m^2} a_{\mathrm{lm},t}^\top W_U a_{\mathrm{lm},t}.$$

This proves the other posterior term, including singular W_U. □

Proposition 130 (The same request interval improves flooring)**.** *For the unchanged centered floor, put* $a_{\mathrm{f}} = -Q_{\mathrm{hi}}^{-1} g$*,* $a_{\mathrm{f},t} = Z^\top a_{\mathrm{f}}$*, and* $\widehat{w}_{\mathrm{f},c} = \theta + a_{\mathrm{f}}$*. Then*

$$\mathcal{E}_T(\widehat{w}_{\mathrm{f},c}) \le \left(\frac{v_U}{\lambda + v_U}\right)^2 \mathcal{E}_T(\theta), \tag{503}$$

$$\mathcal{E}_T(\widehat{w}_{\mathrm{f},c}) \le C_{\mathrm{f},U} := \min\left\{\frac{\|W_U\|_{\mathrm{op}}\, a_{\mathrm{f},t}^\top W_U a_{\mathrm{f},t}}{m^2\lambda}, \left(\frac{v_U}{\lambda + v_U}\right)^2 \frac{\|g\|^2}{\lambda}\right\}. \tag{504}$$

This posterior is no larger than the original centered-floor posterior in (480)*.*

Proof. The normalized floor error is $P_{\mathrm{f}} = Z(\delta I - C)Z^\top/m$, where $0 \preceq \delta I - C \preceq W_U$. Proposition 121 applies with v_U in place of τ. For $E = \delta I - C$, PSD order gives $E^2 \preceq \|W_U\|_{\mathrm{op}} E \preceq \|W_U\|_{\mathrm{op}} W_U$. The residual identity then proves the tail term as above. Finally $W_U \preceq \delta I$ and $v_U \le \delta/m$ make both new posterior terms no larger than their respective original terms. □

The two posterior formulas compare different update vectors. The factor $1/4$ in one formula therefore does not guarantee a smaller certificate or excess for every dataset. When $W_U = 0$, the interval collapses and both decoders are exact. At an empty request, both centered answers equal their reference and both posteriors are zero. The same tail residual formulas apply to uncentered answers with the answer itself replacing a_{lm} or a_{f}; the centered gradient terms do not apply to those answers.

U.4 A SHARP CONTRACTION OVER THE HESSIAN INTERVAL

The midpoint also optimally balances uncertainty over this Hessian interval. This is a deletion application of matrix-midrange principles (Mostajeran et al., 2020). The following statement gives its exact contraction factor and allows arbitrary deterministic updates from the available gradient in the matching lower bound.

Proposition 131 (Optimal contraction from the interval and gradient)**.** *Define the computable quantities*

$$\Delta_U = Q_{\mathrm{hi}} - Q_{\mathrm{lo}}, \qquad t_U = \left\| Q_{\mathrm{lo}}^{-1/2} \Delta_U Q_{\mathrm{lo}}^{-1/2} \right\|_{\mathrm{op}}, \qquad \gamma_U = \frac{t_U}{2 + t_U}. \tag{505}$$

For the centered legal midpoint and exact retained half-gradient g,

$$\mathcal{E}_T(\widehat{w}_{\mathrm{lm},c}) \le \gamma_U^2 \mathcal{E}_T(\theta) \le \gamma_U^2 g^\top Q_{\mathrm{lo}}^{-1} g \le \left(\frac{v_U}{2\lambda + v_U}\right)^2 \frac{\|g\|^2}{\lambda}. \tag{506}$$

Thus $\gamma_U^2 g^\top Q_{\mathrm{lo}}^{-1} g$ may replace the gradient term in (502) *at unchanged persistent storage.*

The factor γ_U^2 is sharp over the full interval of quadratic objectives with $Q_{\mathrm{lo}} \preceq Q \preceq Q_{\mathrm{hi}}$. Specifically, fix these endpoints. For any deterministic rule $a = \mathcal{A}(Q_{\mathrm{lo}}, Q_{\mathrm{hi}}, g)$, including nonlinear rules, there are a nonzero g and an endpoint $Q \in \{Q_{\mathrm{lo}}, Q_{\mathrm{hi}}\}$ for which

$$\frac{(a + Q^{-1}g)^\top Q (a + Q^{-1}g)}{g^\top Q^{-1} g} \ge \gamma_U^2. \tag{507}$$

The ratio is exactly the objective excess after the update divided by the excess at the reference, for a quadratic with half-gradient g there. The midpoint rule attains the factor uniformly over the entire interval and all nonzero gradients. Moreover, for any $0 \le c < \gamma_U^2$, a randomized rule using only these endpoints and g has some fixed endpoint and nonzero gradient at which its probability of attaining a contraction ratio at most c is at most $1/2$.

Proof. Write $L = Q_{\mathrm{lo}}$, $U = Q_{\mathrm{hi}}$, $\Delta = U - L$, $H = (L + U)/2$, and $J = \Delta/2$. Congruence by $L^{-1/2}$ sends the pair (J, H) to $(D/2, I + D/2)$, where $D = L^{-1/2}\Delta L^{-1/2}$. Consequently the eigenvalues of $H^{-1/2} J H^{-1/2}$ are $s/(2+s)$ for the eigenvalues $s \ge 0$ of D. Its operator norm is exactly γ_U.

For every $Q \in [L, U]$, put $P = H - Q$. The orders $-J \preceq P \preceq J$ imply $\left\|H^{-1/2} P H^{-1/2}\right\|_{\mathrm{op}} \le \gamma_U$. The centered update has error map $H^{-1}P$. In the Q metric, this is the symmetric matrix $I - Q^{1/2} H^{-1} Q^{1/2}$, whose eigenvalues equal those of $H^{-1/2} P H^{-1/2}$. This proves the first inequality in (506) without any commutativity assumption. For the remaining inequalities, use $Q^{-1} \preceq L^{-1}$, $\mathcal{E}_T(\theta) = g^\top Q^{-1} g$, $t_U \le \|\Delta\|_{\mathrm{op}} / \lambda = v_U/\lambda$, and $L^{-1} \preceq \lambda^{-1} I$.

To prove sharpness, set $\kappa = 1 + t_U$ and choose a generalized eigenvector x satisfying

$$Ux = \kappa L x, \qquad x^\top L x = 1.$$

Use the same gradient $g = Lx$ for both endpoint quadratics. Their optimum displacements from the reference are $-x$ and $-x/\kappa$, respectively. For the rule's common output a, set $z = -a$ and $\alpha = x^\top L z$. Cauchy–Schwarz in the L and U inner products, respectively, gives

$$\frac{(x - z)^\top L (x - z)}{x^\top L x} \ge (1 - \alpha)^2,$$
$$\frac{(x/\kappa - z)^\top U (x/\kappa - z)}{(x/\kappa)^\top U (x/\kappa)} \ge (1 - \kappa\alpha)^2.$$

For every real α,

$$\max\{|1 - \alpha|, |1 - \kappa\alpha|\} \ge \frac{\kappa - 1}{\kappa + 1} = \gamma_U.$$

This proves (507). The midpoint gives $z = H^{-1} g = 2x/(1+\kappa)$, attaining equality at both endpoints for this gradient. The argument includes $\Delta = 0$, when $\gamma_U = 0$ and the midpoint is exact.

For a randomized rule, the scalar α has the same distribution at the two endpoints because its inputs (L, U, g) are identical. Success with contraction ratio at most $c < \gamma_U^2$ requires, respectively, $|1-\alpha| \le \sqrt{c}$ and $|1 - \kappa\alpha| \le \sqrt{c}$. These two sets are disjoint by the preceding scalar inequality. Their probabilities sum to at most one, so at least one fixed endpoint has success probability at most $1/2$. □

Sharpness is over quadratics specified solely by the interval and gradient. Realizability by the prescribed records, stored labels, or additional task statistics can restrict that class further. The generalized eigenvalue and inverse quadratic form above are computed in temporary workspace. Their finite-precision evaluation has the same qualifications as the interval itself below.

U.5 Storage, computation, and finite precision

The interval uses only the exact original spectral factor, its floor, the record count, and the request. An uncentered decoder additionally uses the stored b_S; a centered decoder uses θ instead. Neither stores L_U, W_U, Z, or a completed Gram matrix persistently. They are recomputed as temporary decoding workspace for each request. Thus the numerical state has the same $\ell d + d + 2$ entries as the corresponding floor summary: an ℓ-by-d factor, a d-vector, one floor scalar, and one integer count. Zero-weight factor rows may be retained in this payload. The decoder can form a full tail basis and dense matrices; no improved decoding-time claim follows from the persistent-memory count.

These statements assume exact spectral information and exact arithmetic. They do not turn the midpoint into the finite-bit upper bound of Theorem 1, which remains supplied by the quantized one-sided floor in Appendix A.4. In particular, factor rounding perturbs the known leading subspace and the zero cross block in (494). A pseudoinverse cutoff can also discard a small eigenvalue of B with a nonzero corresponding cross block, invalidating the completion of squares for the implemented matrices. Entrywise or spectral clipping alone does not establish a valid interval for the true covariance. A finite-bit or machine-certified implementation must enclose these errors and the resulting interval, in addition to the reference and solve errors discussed in Appendix T.5.

Floating-point evaluation is consequently a diagnostic of the ideal formulas. Numerical range residuals and PSD checks can detect a material failure of the computed completion; the implementation then tries the original floor and reports numerical failure if that solve also fails. Passing such checks without rigorous roundoff enclosures is not a numerical proof of the displayed error bounds.

U.6 An additional task-dependent constraint

The feature interval ignores the labels and reference model. In ideal arithmetic, one extra residual-energy scalar can tighten it. This extension is not used in the reported experiments.

Proposition 132 (Residual information tightens the interval)**.** *Suppose the reference is exactly $\theta = w_S$ and retain $R_S = \sum_S (y - x^\top \theta)^2$ in addition to the centered summary. Let $r_U = y_U - X_U \theta$ and form the unnormalized retained residual moments*

$$h = n\lambda\theta - X_U^\top r_U, \qquad s = R_S - \|r_U\|^2, \qquad h_v = V^\top h, \quad h_t = Z^\top h.$$

With B, D_{21}, L_U as above, put

$$r = s - h_v^\top B^\dagger h_v, \qquad z = h_t + D_{21} B^\dagger h_v.$$

Then $r \geq 0$. If $r = 0$, one has $z = 0$ and sets $L_{\mathrm{task}} = L_U$; otherwise set

$$L_{\mathrm{task}} = L_U + zz^\top / r.$$

In both cases $L_U \preceq L_{\mathrm{task}} \preceq C \preceq \delta I$. Consequently replace L_U by L_{task} in G_{lo} and Q_{lo} and recompute the midpoint. The interval and midpoint certificates then hold with $W_{\mathrm{task}} = \delta I - L_{\mathrm{task}}$. The unchanged floor may directly use this smaller width in its certificate; the unchanged midpoint may not.

Proof. Original stationarity gives $X_S^\top (y_S - X_S \theta) = n\lambda\theta$, so $h = X_T^\top (y_T - X_T \theta)$ and $s = \|y_T - X_T \theta\|^2$ are the exact retained moments. The Gram matrix of the retained leading features, residual labels, and tail features is

$$\begin{pmatrix} B & h_v & -D_{12} \\ h_v^\top & s & h_t^\top \\ -D_{21} & h_t & C - D_{22} \end{pmatrix} \succeq 0.$$

Its range conditions allow elimination of the possibly singular B, giving $\left(\begin{smallmatrix} r & z^\top \\ z & C - L_U \end{smallmatrix}\right) \succeq 0$. The scalar Schur criterion yields $z = 0$ at $r = 0$ and $C - L_U \succeq zz^\top / r$ otherwise. The earlier interval proofs now apply unchanged. □

The improvement depends on residual-feature correlation, not only a rescaling of label amplitude. The gain comes from one additional rank-one constraint within the same matrix-completion framework; empirical error need not improve uniformly. Small conditional residual energy r adds numerical sensitivity. The extra scalar leaves the reference-precision requirements unchanged: a rounded reference needs exact residual cross statistics or the explicit error enclosure in Appendix T.5.

V REPRODUCIBILITY

The appendices provide the complete constructions, information arguments, and finite-precision proofs. The accompanying scripts `experiments_deletion.py`, `experiments_frozen.py`, `experiments_task_alignment.py`, `experiments_response.py`, `experiments_dpp.py`, `experiments_importance.py`, `experiments_completion.py`, and `experiments_utility.py` specify the data construction or acquisition, preprocessing, seeds, parameter grids, decoders, and numerical summary accounting. The ordinary, importance, signed, and reference-utility encoders are implemented in `dpp_summary.py` and `dpp_importance_summary.py`. The eight experiment grids contain $90{,}764$ method evaluations in total; reused comparator rows are counted once. A separate source-geometry analysis covers 38 configurations and is not counted as additional method evaluations. The experiment environment uses Python 3.11, NumPy 2.4.6, Matplotlib 3.11.0, PyTorch 2.6.0 (CPU), torchvision 0.21.0 (CPU), and Pillow 12.3.0. The experiments compare floating-point implementations; the bit-complexity statements follow from the separate finite-precision proofs.

The separate SymPy prototype `actual_response_span.py` implements the rational actual-response span and exact weighted correction of Appendix K.3.9. It first reduces to a rational basis of the source feature span, including its nonorthogonal regularization metric. Run `python actual_response_span.py -check-only` from the project root. The checks cover zero and oversized deletion orders, repeated records, zero labels, singular gaps, nonorthogonal feature embeddings, and the explicit deletion-order chain. They compare responses and decoding against independent retained-record ridge fits. Its offline `positive_loss_certificate` API implements Proposition 85 and supplies a lower bound, not an optimizer. The checks include 1200 independent retraining/decoder comparisons and 166 certificate checks on eight small sources. All APIs require exact rational inputs. The lift is exponential in source rank; it is a research prototype, and the checks establish neither the bit-complexity theorem nor a high-rank runtime advantage. These checks add no performance-grid rows.

The same script's `singleton_responses`, `witness_loss_certificate`, and `singleton_loss_certificate` APIs implement the polynomial certificate path in Appendix K.3.10. They use shared rational solves, exact PSD tests, and optional scalar probability weights. The default singleton path also checks a greedy $r+1$-witness bank. Their result is labelled `positive` or `inconclusive`; an inconclusive partial bank does not assert a zero optimum. The returned rank describes the selected weighted bank, not the entire actual-response space. Additional checks comprise 113 independent singleton fits and 233 bank/spectral checks across 13 small sources, including bounded full-source-rank examples in dimensions 12 and 16 and three bounded weak-axis sources with the original spectral floor. The spectral-tail weight-optimization SDP and the complete implicit encoder are not implemented by these certificate APIs.

The same check entry point reproduces six dyadic sources from Theorem 94, with $k = 1, 2, 3, 4, 6, 8$. Rational square-root enclosures certify the four rounded filler coordinates; the bounded-precision fixture fails explicitly if rounding cannot be resolved. These checks include 48 independent retained-data fits, six lower and 48 directional upper moment bounds, and 32 exact weighted-loss identities for four coupled embeddings in dimensions 4 and 7 at ranks 1 and 2. They check finite instances of the construction and embedding, without optimizing the real correction class or adding performance-grid evaluations.

W SUPPLEMENTARY SYNTHETIC DIAGNOSTICS

W.1 SYNTHETIC DELETION DIAGNOSTICS

We measure deletion accuracy as the original covariance spectrum varies at a fixed storage budget. The controlled grid isolates finite-dimensional decoder behavior; minimax rates require a separate worst-case analysis. The implementation uses NumPy on CPU and compares exact additive statistics, reuse of the original model, global spectral shift, unshifted truncation, spectral flooring, centered global shift, centered flooring, and two centered summaries with query-time positive-semidefinite (PSD) repair, and the request-conditioned legal midpoint. Centered variants store w_S in place of b_S; centered flooring combines the existing flooring and centering operations.

For the repaired baselines, the original proxy keeps the leading eigenvalues and fills the tail with $\mu = \delta/2$ (midpoint) or $\mu = 0$ (truncation). The fixed midpoint follows the low-rank-plus-isotropic approximation principle of Luo et al. (2019); its coefficient is fixed before the comparisons. Writing this proxy as $\widetilde{G}$, both baselines return $w_S + [(\widetilde{G} - G_U)_+ + m\lambda I]^{-1}[(G_U + |U|\lambda I)w_S - b_U]$, where $(\cdot)_+$ clips negative eigenvalues to zero. They use the same persistent payload as centered flooring. Comparing the two repaired methods isolates the tail coefficient. Comparing repaired truncation with unshifted truncation also changes centering when $w_S \neq 0$, so it is not a pure repair ablation in that case. Neither repaired baseline is assigned the floor's tail-supported cubic bound. Both receive a summary-only residual bound: with $N = (\widetilde{G} - G_U)_+ - (\widetilde{G} - G_U)$ and $a = \widehat{w} - w_S$, the bound is $(\|Na\| + r\|\Pi a\|)^2/(m^2\lambda)$, where $r = \delta/2$ for midpoint and $r = \delta$ for truncation, and Π is the complement of that factor's row space. PSD repair by itself is not a utility certificate; this inequality bounds the unknown retained gradient residual.

The legal midpoint uses exactly the centered floor's stored factor, predictor, scalar, and count. It computes the omitted-covariance interval implied by the request and uses its midpoint (Appendix U). For a fair certificate comparison, the unchanged flooring decoders use the same request interval to refine their posterior bounds. The uniform floor and centered-floor bounds use the stronger $4/27$ constant from the spectral-shape lemma. We retain the original posterior and the earlier coarse cubic estimate in separately named columns.

We use $d = 48$, $n = 384$, and delete 96 records. Features have the form $x = R\,\mathrm{diag}(\sqrt{\omega})s$, where R is orthogonal, $s \in \{-1, 1\}^d$, and $\sum_j \omega_j = 1$; hence $\|x\| = 1$. Population spectra are flat, proportional to $j^{-3/2}$, or proportional to $\exp(-(j-1)/3)$. All sketches use the actual original sample Gram matrix. The signal family has labels $y = 0.2\sum_{j=1}^{4} s_j + \xi$, with $\xi \sim \mathrm{Unif}[-0.03, 0.03]$. The zero-model family pairs every feature with labels $+1$ and -1, giving $w_S = 0$ exactly in real arithmetic. Thus both families satisfy the feature and label bounds.

We evaluate $\lambda \in \{0.002, 0.01, 0.05\}$ and $\ell \in \{4, 12, 24\}$ over seven independent data seeds. Queries delete either a uniformly random subset, records with the largest signed original residual $y - x^\top w_S$, or records with the largest feature energy outside the original top-ℓ eigenspace. Ties are randomized. This gives 1134 configurations and 11340 method evaluations. Feature-only deletion on paired zero-model data usually preserves complete label pairs and is a benign control.

Each decoder receives only its own numeric summary and the deleted records. Retained data are used separately to compute the exact optimum and evaluate $F_T(\widehat{w}) - F_T(w_T)$, with the loss normalized by $|T|$. Storage includes the factor, the full b_S or w_S vector, the spectral-tail scalar, and n. At rank 12, every sketch stores 626 numeric elements (5008 bytes); exact packed symmetric Gram statistics store 1225 (9800 bytes), and model reuse stores 48 (384 bytes). These are float64/int64 payload sizes, excluding transient workspace and Python object overhead; they are not a quantized bit-complexity claim. The evaluator records $L = F_S(w_S)$ and the retained optimum separately. Posterior certificates are computed by a separate function whose only inputs are the numeric summary, deleted records, public λ, and decoded model. This function uses the stored factor's row space, including rank-zero and repeated-eigenvalue cases, and never receives the evaluator's optimum. The experiments use the interval-width posterior in (502); the generalized-eigenvalue refinement in Proposition 131 is a further theoretical option and is not used in the reported certificate values. For the Schur interval, the implementation records the pivot eigenvalues, discarded-mode range residual, lower-endpoint and width PSD checks, and any roundoff correction. The pseudoinverse cutoff is $64uds$ and the check tolerance is $256uds$, where u is float64 machine precision and $s = \max\{1, \|A\|_{\mathrm{op}}, \|G_U\|_{\mathrm{F}}\}$ (with δ replacing $\|A\|_{\mathrm{op}}$ for an empty leading block). A material check failure makes the legal-midpoint implementation return centered flooring and its original posterior, recording the fallback. A near-singular legal toy exercises this path; a compatible exactly singular toy succeeds. These guards diagnose floating-point limitations rather than proving the rounded interval legal. Independently, the actual solve matrix must have a positive computed minimum eigenvalue and pass Cholesky factorization; the computed solution must be finite and pass a normwise backward-residual check. A failed solve, including a failed flooring fallback, returns no model and an explicit numerical-failure status. No ridge adjustment is used. A bounded two-dimensional stress case at $\lambda = 10^{-28}$ passes the envelope diagnostics but fails the solve-matrix check, illustrating why the two checks are separate. These tests do not provide rigorous error enclosures.

Table 2: Normalized retained ridge excess for the nonzero signal family, $\lambda = 0.01$, $\ell = 12$, and signed-residual deletion. Entries are medians over seven seeds. Sketch methods use equal numeric storage; empirical 10th–90th percentile bands appear in Figure 2.

Spectrum	Reuse	Floor	C. floor	Midpoint + PSD	Trunc. + PSD	Legal midpoint
Flat	2.00×10^{-3}	3.74×10^{-3}	2.15×10^{-4}	5.90×10^{-4}	2.28×10^{-3}	9.05×10^{-5}
Power law	1.81×10^{-4}	4.67×10^{-6}	3.77×10^{-6}	1.22×10^{-6}	6.49×10^{-6}	1.15×10^{-6}
Steep	5.45×10^{-5}	1.23×10^{-7}	1.11×10^{-7}	5.21×10^{-8}	2.02×10^{-7}	4.56×10^{-8}

Table 2 shows substantial improvements from the midpoint baseline over flooring under decaying spectra. There is no uniform dominance: on the flat signal family, both model reuse and centered global shift outperform uncentered flooring. Centered flooring improves on uncentered flooring in all three displayed settings; centering has no effect when $w_S = 0$. On the paired zero-model power-law family at the same setting, flooring reduces median excess from 1.73×10^{-2} for reuse to 1.15×10^{-3} despite identical original models. The legal midpoint has the smallest median excess in the three displayed signal settings. Across all 567 signal configurations it improves on centered flooring in 479 and on the fixed midpoint in 545; neither comparison is uniformly ordered.

This spectrum sweep also changes learnability: the signal teacher norms are $2.77, 1.26, 1.00$ for the flat, power-law, and steep spectra, and their median L values at $\lambda = 0.01$ are $0.0506, 0.0141, 0.00956$. It therefore does not isolate the effect of task alignment. The subsequent controlled experiment holds the covariance and label multiset fixed.

Both flooring variants, both PSD-repaired baselines, and the legal midpoint remain positive definite in all 1134 configurations. Unshifted truncation is non-positive-definite in 487 cases and positive definite but outside its $\tau < \lambda$ certificate in another 182. We record these statuses separately and return no model for unstable cases. Exact summary decoding agrees with direct retraining to numerical precision; all applicable excess bounds, including the posterior values, pass direct-objective checks. No Schur-envelope fallback or width PSD correction occurs in this grid. PSD repair is active in 225 fixed-midpoint configurations and all 1134 truncation configurations, using operator norm greater than 10^{-12} as the numerical activity threshold. The checks confirm numerical consistency with the ideal-arithmetic inequalities on this grid. Rigorous roundoff certification and the sharpness of the cubic dependence are separate from these evaluations.

For the additional comparators, let C be the ordinary leading-rank Gram factor and $\delta = \sigma_{\ell+1}(G_S)$. Unshifted truncation uses $[C^\top C - G_U + m\lambda I]^{-1}(b_S - b_U)$ when its numerical system is positive definite. Centered global shift stores w_S in place of b_S and returns

$$w_S + [C^\top C - G_U + (m\lambda + \delta)I]^{-1}[(G_U + |U|\lambda I)w_S - b_U]. \tag{508}$$

This expression uses only the stored predictor, factor, scalar, count, and deleted records. It agrees with the original predictor at an empty request and reduces to global shift when $w_S = 0$.

Centered spectral flooring. Centering can also be applied to the spectral floor. Store (K, δ, w_S, n), let $\widehat{G} = K^\top K + \delta I$, and return

$$\widehat{w}_{\mathrm{cf}} = [\widehat{G} - G_U + m\lambda I]^{-1}[(\widehat{G} + n\lambda I)w_S - b_U]. \tag{509}$$

This uses the same number of numerical entries as spectral flooring, with w_S replacing b_S, and exactly returns the stored predictor for an empty request. It completes the comparison between centering and the choice of spectral correction; this combination is not a separate algorithmic novelty claim.

Its unquantized error also has cubic rank dependence. Set $E_0 = \widehat{G} - G_S$, $P = E_0/m$, $Q = Q_T$, $\tau = \delta/m$, and $\beta = m/n$. Write $r_A = \sum_A (y - x^\top w_A)^2/|A| \le 1$ for $A = S, T$. With $\kappa_\ell = \max_{i>\ell} \sigma_i(\delta - \sigma_i)^2 \le 4\delta^3/27$, we have $E_0 G_A E_0 \preceq \kappa_\ell I$. Stationarity gives $\|E_0 w_A\| \le$

$\sqrt{\kappa_\ell r_A/|A|}/\lambda$. Therefore (12) yields

$$\widehat{w}_{\mathrm{cf}} - w_T = (Q+P)^{-1}P(w_S - w_T), \tag{510}$$

$$\|E_0(w_S - w_T)\| \le \frac{\sqrt{\kappa_\ell}}{\lambda\sqrt{m}}(\sqrt{\beta r_S} + \sqrt{r_T}), \tag{511}$$

$$F_T(\widehat{w}_{\mathrm{cf}}) - F_T(w_T) \le \frac{\|E_0(w_S - w_T)\|^2}{m^2\lambda} \le \frac{4}{27}(1+\sqrt{\beta})^2\left(\frac{\tau}{\lambda}\right)^3. \tag{512}$$

Thus the numerical storage rate remains $O(d\min\{d, (\lambda\varepsilon^{1/3})^{-1}\})$ at a fixed retained fraction. A finite-bit implementation must also specify the precision of the stored predictor. The bit upper bound of Theorem 1 is supplied by the quantized construction in Appendix A.4; the ablation compares the actual floating-point payloads.

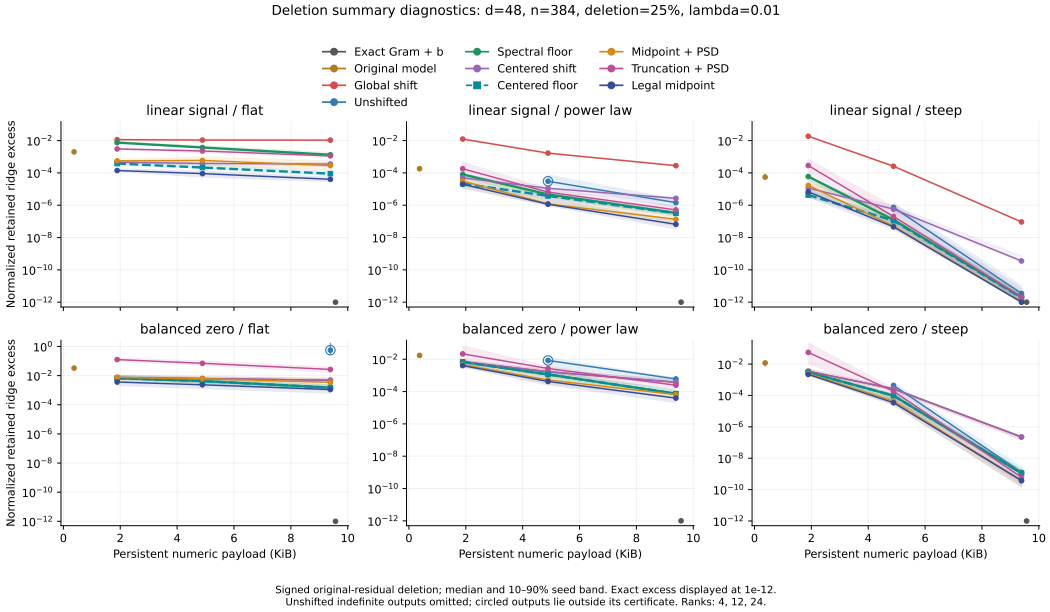

Figure 2: Deletion accuracy across the three synthetic spectra and two label families at $d = 48$, $n = 384$, $\lambda = 0.01$, and deletion fraction 25%. Queries select the largest signed original residuals. Lines connect ranks $4, 12, 24$; markers and shading show medians and empirical 10th–90th percentiles over seven seeds. The horizontal axis counts persistent numerical payload. Non-positive-definite unshifted outputs are omitted; circles mark positive-definite outputs outside the sufficient $\tau < \lambda$ condition. Exact-summary excess is displayed at 10^{-12} for visibility.

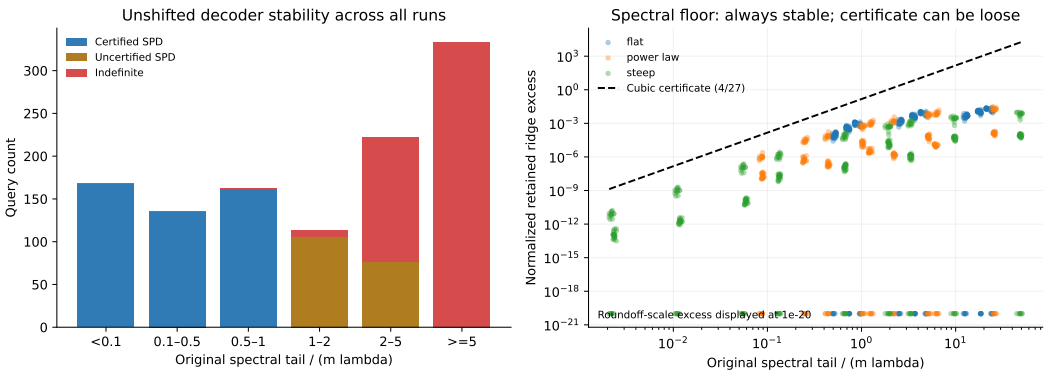

Figure 3: Stability and certificate diagnostics across 1134 synthetic configurations. Left: unshifted truncation has 487 non-positive-definite systems and 182 additional positive-definite systems outside its sufficient perturbation condition. Right: the spectral-floor systems remain positive definite, while their empirical excess can be far below the cubic upper bound. Roundoff-scale excess is displayed at 10^{-20}. "Certified" in this figure refers to a matrix perturbation condition, not a distributional guarantee of unlearning.

Table 3: Frozen features: median retained ridge excess over three seeds at $\lambda = 10^{-3}$, $\ell = 32$, and 25% deletion. Both random and signed-residual requests are shown. Exact-summary entries reflect numerical agreement with retraining, not a numerical certification procedure.

		Cat/dog		Airplane/truck	
Method	KiB	Random	Signed	Random	Signed
Exact Gram + b	1030.0	$< 10^{-20}$	$< 10^{-20}$	$< 10^{-20}$	$< 10^{-20}$
Original model	4.0	1.40×10^{-2}	1.25×10^{-1}	5.56×10^{-3}	4.37×10^{-2}
Global shift	132.0	3.67×10^{-2}	2.40×10^{-2}	1.16×10^{-2}	9.04×10^{-3}
Spectral floor	132.0	1.81×10^{-2}	9.44×10^{-3}	5.47×10^{-3}	3.23×10^{-3}
Centered shift	132.0	3.69×10^{-3}	9.87×10^{-3}	1.36×10^{-3}	3.87×10^{-3}
Centered floor	132.0	3.18×10^{-3}	7.96×10^{-3}	1.13×10^{-3}	2.68×10^{-3}
Midpoint + PSD	132.0	1.80×10^{-3}	4.38×10^{-3}	5.59×10^{-4}	1.39×10^{-3}
Truncation + PSD	132.0	5.83×10^{-3}	1.54×10^{-2}	1.42×10^{-3}	4.27×10^{-3}
Legal midpoint	132.0	1.57×10^{-3}	3.93×10^{-3}	5.34×10^{-4}	1.25×10^{-3}

X SUPPLEMENTARY FROZEN-FEATURE DIAGNOSTICS

X.1 FROZEN IMAGE FEATURES

We also evaluate CIFAR-10 cat/dog and airplane/truck tasks (Krizhevsky, 2009) using public ImageNet-pretrained ResNet-18 features (He et al., 2016). Torchvision's `IMAGENET1K_V1` weights, evaluation mode, and official resize/crop and normalization define a fixed 512-dimensional representation before the classifier. Each feature is then individually ℓ_2-normalized; no representation transformation is fitted on S. Labels are ± 1. Three seeds independently draw balanced subsets of 1,000 training and 500 test examples per task. We cross $\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}\}$, $\ell \in \{8, 32, 128\}$, deletion fractions $5\%, 25\%$, and random or largest signed-original-residual deletion, giving 216 configurations and 1,944 method evaluations. The diagnostic hyperparameter grid is fixed independently of test accuracy. Seed subsets can overlap within the official train or test pool, so the reported ranges describe the sampled outcomes rather than confidence intervals.

Decoders receive only their numeric summaries and deleted records. The principal metric remains normalized retained ridge excess; test accuracy is auxiliary. Persistent storage counts float64/int64 array payloads, excluding the shared public extractor and evaluation data. All seven spectral summaries have the same numeric storage at each rank. The midpoint and truncation baselines use the PSD repair described above. The legal midpoint and the refined flooring posterior use the same request-conditioned covariance interval. The certificate comparison therefore gives both methods the same request information.

Table 3 illustrates both the memory–error tradeoff and the importance of a strong stable comparator. Increasing flooring rank from 8 to 128 raises its payload from 36.0 to 516.0 KiB and lowers median excess from 4.09×10^{-2} to 3.70×10^{-4} on cat/dog, and from 1.32×10^{-2} to 1.08×10^{-4} on airplane/truck under the displayed signed-residual requests. Across the full grid, the fixed midpoint with PSD repair improves on centered flooring in 207 of 216 configurations and on repaired truncation in all 216. Centered flooring improves on uncentered flooring in 199 configurations. The legal midpoint improves on centered flooring in all 216 configurations but on the fixed midpoint in only 123: $31/108$ at 5% deletion and $92/108$ at 25% deletion. Conditioning on the legal request therefore need not improve actual accuracy over the fixed midpoint. All these comparisons are observations on the stated grid, not universal dominance claims. Both repaired baselines now receive their own summary-only residual certificates.

The posterior diagnostic is more informative than the uniform estimate. Across all 216 configurations, centered flooring's original posterior is a median 3.40 times its actual excess, with range 1.90–16.31; the tail-response term is smaller than the gradient term in every configuration. The same request interval reduces this median ratio to 2.85, with range 1.82–13.80, without changing its returned model. The fixed midpoint's ratio is 4.13 (2.38–26.16), and the legal midpoint's is 3.33 (2.17–14.54). The legal midpoint has a smaller posterior value than both other methods in all 216 configurations, despite losing some actual-error comparisons to the fixed midpoint. In Table 4, uniform centered-floor bounds are 10.6 and 6.77, whereas its refined posterior values range from

Table 4: Median actual excess and certificate values at the same rank, regularization, and deletion fraction as Table 3. Posterior values use only decoder inputs: flooring and legal midpoint use the same request interval, while fixed midpoint uses its repaired-residual bound. Uniform entries use the cubic label-bound estimates for flooring with the spectral-shape constant $4/27$; dashes mean that this bound is not assigned to that method. These are floating-point evaluations of ideal-arithmetic bounds.

Task	Deletion	Method	Excess	Posterior	Uniform
Cat/dog	Random	Centered floor	3.18×10^{-3}	9.49×10^{-3}	1.06×10^{1}
Cat/dog	Random	Midpoint + PSD	1.80×10^{-3}	8.58×10^{-3}	–
Cat/dog	Random	Legal midpoint	1.57×10^{-3}	5.59×10^{-3}	–
Cat/dog	Signed	Centered floor	7.96×10^{-3}	2.32×10^{-2}	1.06×10^{1}
Cat/dog	Signed	Midpoint + PSD	4.38×10^{-3}	2.00×10^{-2}	–
Cat/dog	Signed	Legal midpoint	3.93×10^{-3}	1.38×10^{-2}	–
Airplane/truck	Random	Centered floor	1.13×10^{-3}	2.76×10^{-3}	6.77×10^{0}
Airplane/truck	Random	Midpoint + PSD	5.59×10^{-4}	2.20×10^{-3}	–
Airplane/truck	Random	Legal midpoint	5.34×10^{-4}	1.56×10^{-3}	–
Airplane/truck	Signed	Centered floor	2.68×10^{-3}	7.60×10^{-3}	6.77×10^{0}
Airplane/truck	Signed	Midpoint + PSD	1.39×10^{-3}	6.53×10^{-3}	–
Airplane/truck	Signed	Legal midpoint	1.25×10^{-3}	4.23×10^{-3}	–

0.00276 to 0.0232, and legal-midpoint values from 0.00156 to 0.0138. The original optimal losses at this λ are a median 0.480 for cat/dog and 0.190 for airplane/truck. The evaluator also records $(4/27)F_T(w_T)(\tau/\lambda)^3$ and $(4/27)(L/\beta)(\tau/\lambda)^3$ separately; these theoretical comparison values are not inputs to the posterior function.

We fixed absolute tolerances $\varepsilon \in \{0.001, 0.01, 0.1\}$ before running the revised grid. At $\varepsilon = 0.01$, centered flooring's original posterior accepts 77 of 108 random and 60 of 108 signed-residual configurations. Its refined posterior accepts 80 and 61, compared with 21 of each for its uniform bound. Fixed midpoint accepts 78 and 61, and legal midpoint 82 and 64. Actual excess meets the tolerance in $100/84$ cases for centered flooring, $102/89$ for fixed midpoint, and $102/92$ for legal midpoint (random/signed). No accepted configuration violates its tolerance in this grid. Figure 4 shows all three tolerances; these counts describe the evaluated grid rather than a probability guarantee for new requests.

At the table's signed-residual setting, median test accuracies for centered flooring/flooring/exact retraining/model reuse are $79.2/78.8/78.8/85.2\%$ on cat/dog and $95.4/95.4/95.0/97.0\%$ on airplane/truck: matching the retained objective need not maximize balanced-test accuracy after label-oriented deletion. Here 99.6–100% of deleted cat/dog records and 82.4–84.0% of deleted airplane/truck records have positive labels. Exact decoding agrees with retraining to numerical precision, and all applicable excess bounds pass numerical consistency checks. These observations concern deletion with fixed representations; worst-case tightness and end-to-end neural-network unlearning require different evidence. No Schur-envelope fallback occurs in the frozen grid; the largest numerical width correction is 2.31×10^{-14}. At the 10^{-12} operator-norm activity threshold, PSD repair is inactive for the fixed midpoint in all 216 configurations and active for truncation in all 216. Thus the frozen grid tests the fixed midpoint only in the inactive-repair regime.

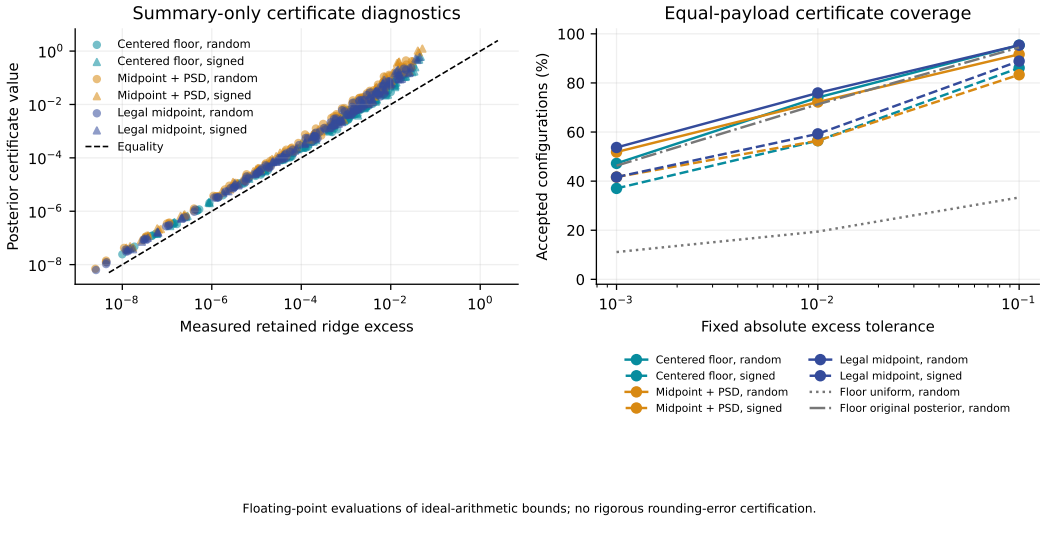

Figure 4: Posterior certificate diagnostics across the frozen-feature grid. Left: actual excess versus summary-only posterior for centered flooring, fixed midpoint, and legal midpoint, with requests distinguished by markers. Flooring receives the same request interval as legal midpoint. Right: their coverage at three predeclared absolute tolerances, using 108 configurations for each deletion strategy. Gray reference curves show the floor's uniform and original posterior coverage for random requests. The numerical evaluations do not include rigorous rounding-error enclosures.

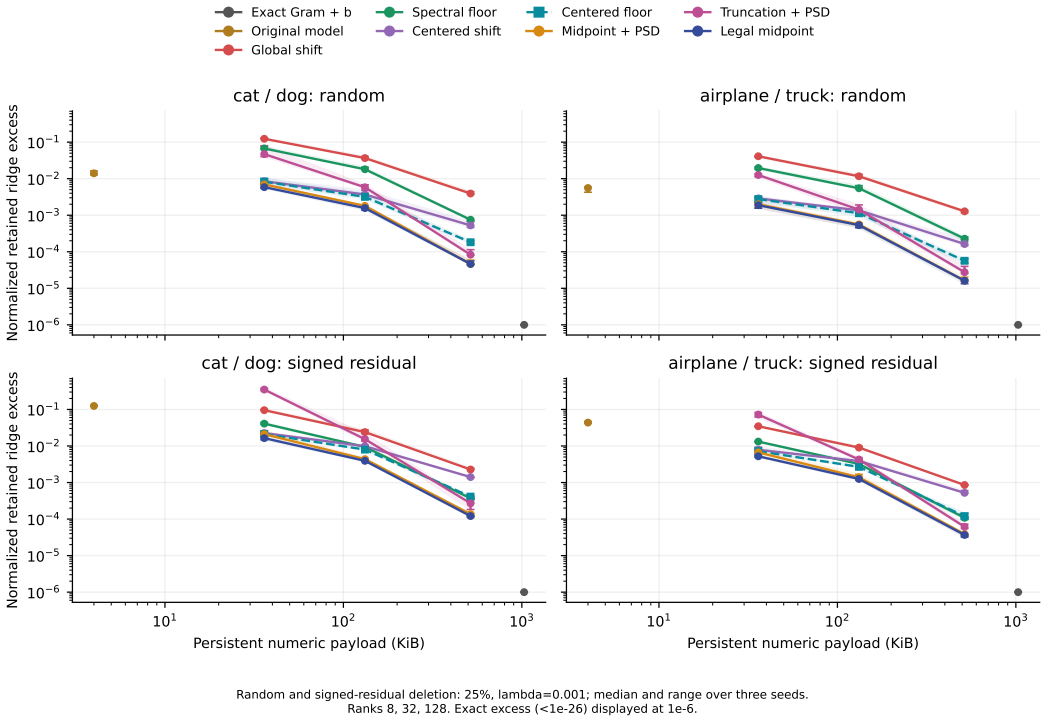

Figure 5: Memory–accuracy curves for the two frozen ResNet-18 tasks at $\lambda = 0.001$ and 25% random or signed-residual deletion. Lines connect ranks $8, 32, 128$; markers show medians and intervals show the full range across three seeds. Exact-summary excess is displayed at 10^{-6} solely for visibility; its measured value is below 10^{-26}. Storage counts numeric payload and excludes the common public feature extractor.

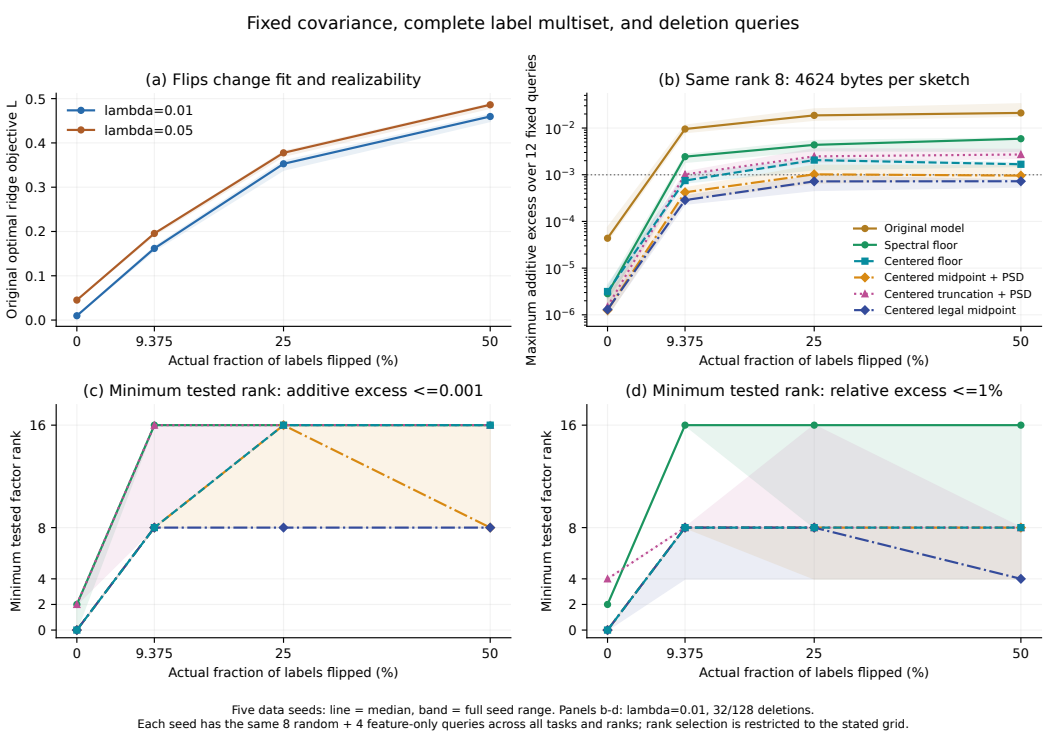

Figure 6: Task alignment with fixed covariance, full label multiset, and deletion masks. Curves show medians and full five-seed ranges after taking each seed's maximum over twelve common requests. Panels (b)–(d) use $\lambda = 0.01$ and 25% deletion. Panel (b) compares equal rank-eight sketch payloads; panels (c),(d) select the minimum rank from the stated grid. These storage comparisons are specific to the finite request bank.

Y FIXED-GEOMETRY TASK-ALIGNMENT CONTROL

We isolate task alignment from the covariance spectrum and label amplitude. For each of five seeds, $d = 64$ and $n = 128$ feature vectors are 64 rotated Rademacher vectors and their antipodes. The population weights satisfy $\omega_1 = 1/2$ and $\omega_j \propto (j-1)^{-3/2}$ for $j \geq 2$, with $\sum_{j\geq 2} \omega_j = 1/2$. Every feature has norm one. The clean task has $v = Re_1$, $\|v\| = 1$, and $y = x^\top v \in \{-1/\sqrt{2}, +1/\sqrt{2}\}$. We then flip equal numbers of positive and negative labels using fixed nested random permutations. Nominal flip fractions $0, 0.1, 0.25, 0.5$ correspond to $0, 6, 16, 32$ flips of each sign, so the actual fractions are $0, 9.375\%, 25\%, 50\%$. The entire label multiset, its second moment $1/2$, the feature matrix, and every sample eigenvalue remain unchanged. The norm-one noiseless teacher is a guarantee for the clean condition only. Balanced flips are selected independently within the two sign groups and need not preserve opposite labels on antipodal features. They can therefore introduce an even label component, which no linear predictor on these features can fit. This intervention changes realizability as well as alignment; it does not isolate L as the sole cause of deletion difficulty.

The fixed experimental grid is $\lambda \in \{0.01, 0.05\}$ and $\ell \in \{0, 2, 4, 8, 16, 32, 64\}$. Deletion counts 6 and 32 give actual fractions 4.6875% and 25%. For each seed and count, one common bank contains eight uniform random requests and four feature-only requests: largest energy outside a fixed rank-eight eigenspace, largest leading-component energy, largest signed leading coordinate, and largest ridge leverage at fixed pilot $\lambda = 0.01$. These masks are selected without labels and reused across every task, regularization parameter, and tested rank. All ten decoder methods are evaluated; rank-independent baselines are evaluated once. This gives $55{,}680$ method–query evaluations and $4{,}640$ seedwise method–rank bank summaries.

We report each seed's maximum excess over its twelve requests, then the median and full range across the five seeds. The primary rank criterion is relative excess at most 1% of the retained opti-

mum; the secondary criterion is additive excess at most 10^{-3}. The minimum is over the stated rank grid and common finite bank, not all legal deletions. Direct retained-objective differences agree with the quadratic excess, and all applicable theoretical and posterior inequalities pass numerical checks. Additive posterior bounds use only the numeric summary and deleted records. Dividing them by the true retained optimum produces an *oracle-normalized diagnostic*, not a relative certificate available to the decoder. Original and retained optima are evaluator quantities; the optional additional-scalar relative test is not implemented here. These checks use floating-point arithmetic and do not certify roundoff.

Task fit and deletion difficulty. At $\lambda = 0.01$ and 25% deletion, median original optimal objectives are $0.00976, 0.16178, 0.35284, 0.45966$ in increasing flip order. Centered flooring's median minimum additive ranks are $0, 8, 16, 16$; its relative ranks are $0, 8, 8, 8$ (Figure 6). Reusing the original model meets both tolerances on all five clean-task banks and fails both on all fifteen corrupted-task banks in this setting. The stronger midpoint-plus-PSD baseline has median minimum additive ranks $0, 8, 16, 8$ and smaller median rank-eight excess than centered flooring in each displayed condition. In particular, its selected rank decreases from 25% to 50% flips while L increases. The legal-midpoint method has median minimum additive ranks $0, 8, 8, 8$. At rank eight, it improves on midpoint-plus-PSD in the three corrupted conditions, while midpoint-plus-PSD is slightly better on the clean task. Thus the experiment supports an effect of task alignment at fixed geometry; it does not show that L alone monotonically determines empirical deletion difficulty, or that flooring dominates other stable summaries.

Certificates and relative normalization. The spectral ratio τ/λ is identical across tasks for each seed, rank, and deletion count. At rank eight in the displayed setting, its median is 1.364. The floor's original-loss bound $(4/27)(L/\beta)(\tau/\lambda)^3$ changes with task fit, but its relative theorem is simply $(4/27)(\tau/\lambda)^3$: division by $F_T(w_T)$ cancels the retained optimum factor (Figure 7). The floor and centered-floor uniform comparisons use the same improved $4/27$ spectral-shape constant as the other experiments. The geometry-only relative theorem is available to the decoder without evaluating $F_T(w_T)$. In contrast, the plotted posterior divided by $F_T(w_T)$ uses that evaluator-only denominator and is explicitly labeled oracle-normalized. All posterior-based *certified* ranks in the figure use the additive criterion. Separate CSV columns record the ranks selected by oracle-normalized relative diagnostics. For fair additive comparisons, we use the smallest applicable classical and legal-envelope posterior bound for each answer, including the midpoint-plus-PSD, truncation-plus-PSD, and legal-midpoint comparators. Centered flooring's median minimum ranks certified at additive tolerance 10^{-3} are $0, 16, 16, 16$, while the legal midpoint's are $0, 8, 16, 16$. The latter improves a tested storage threshold in the 9.375%-flip condition, but still requires rank 16 to certify the two most corrupted conditions where measured bank errors already pass at rank eight. None of the $20{,}160$ legal-envelope evaluations triggers the numerical fallback; this observation does not establish stability outside this grid.

Every rank-eight sketch uses 4624 bytes; rank zero uses 528 bytes, the original predictor 512 bytes, and packed exact statistics 17160 bytes. Rank sixteen uses 8720 bytes. At rank 64, the factor payload is 33296 bytes, so exact statistics are a cheaper exact-storage fallback. Counts include all persistent float64/int64 arrays and exclude transient workspace. These are array-payload counts rather than finite-bit optimality bounds.

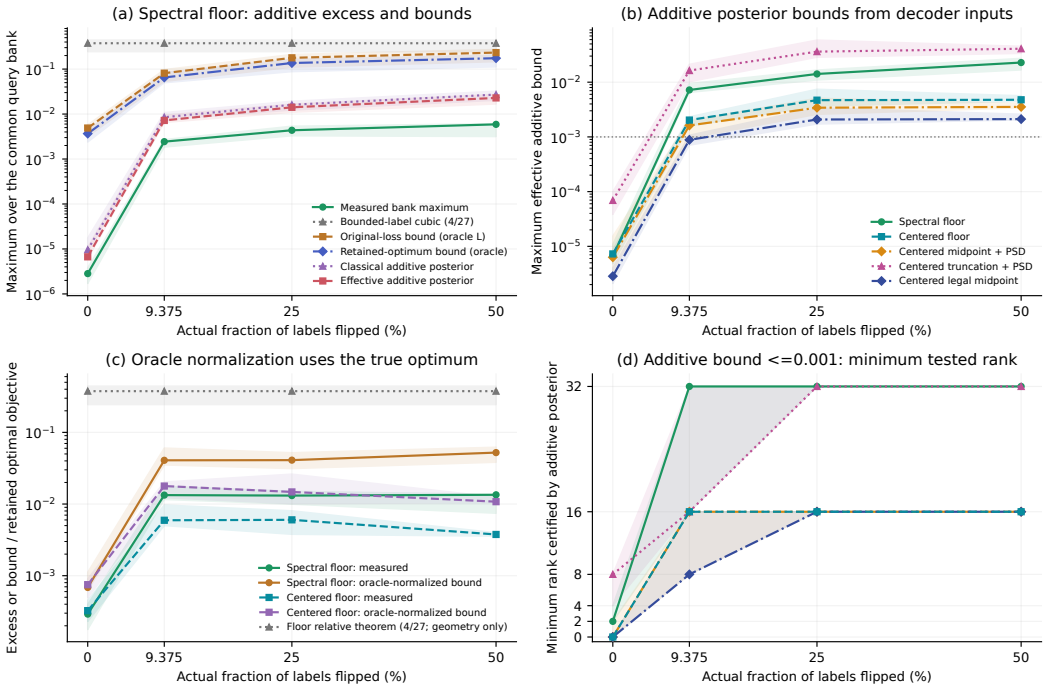

Figure 7: Measured errors, additive certificates, and oracle-normalized diagnostics at $\lambda = 0.01$ and 25% deletion. Panels (a)–(c) use rank eight; panel (d) selects a tested rank using the additive posterior. Each curve first takes the maximum over the same twelve queries per seed, then the median and full seed range. Original-loss and retained-optimum bounds use evaluator quantities; the additive posterior uses only summary and deleted records. The posterior normalization in (c) additionally uses the evaluator's true retained optimum and is not a decoder-available relative certificate. The geometry-only relative floor theorem is constant across tasks.

Z PRACTICAL RESPONSE ENVELOPES UNDER COMMON MEMORY CAPS

We test whether the computable response envelope provides an empirical advantage after charging its covariance-completion representation. The experiment uses only the existing normalized ResNet-18 feature cache: cat–dog and airplane–truck classification, three seeds, and $n = 1000, d = 512$ per training set. The fixed (λ, ρ) pairs are $(0.001, 0.005)$, $(0.01, 0.025)$, $(0.1, 0.05)$, $(0.1, 0.25)$, and $(0.01, 0.25)$. The last pair is a control expected to fail the sufficient contraction condition; its failure is recorded rather than assigned an accuracy. Each configuration uses one uniformly random and one signed-original-residual request, shared by every method and memory cap. Actual deletion counts are $5, 25, 50, 250, 250$, respectively.

For seed 11 on both tasks we additionally set $y = Xe_0$, using a fixed public unit teacher and the same geometry, parameter pairs, and caps. These labels retain their natural amplitude; we do not rescale them to the binary class labels' second moment of one. Thus the noiseless control also changes label scale. We do not assert that the real class labels satisfy the bounded-teacher assumption. The teacher-label second moments are 0.002335 for cat–dog and 0.0006147 for airplane–truck.

Memory and computation. For $m \in \{2, 4, 8, 16, 32\}$, the public cap is $8[2d(2m)+(2m)^2+d+6]$ bytes, namely $37040, 70192, 137264, 274480, 561200$ bytes. The six non-factor entries account for the reference vector's companion scalars and metadata, including the explicit weighted-mode flag; the d-vector itself is counted separately in the formula. The operator encoder selects the union of m covariance and m response directions. Comparator ranks are chosen conservatively before evaluating accuracy: first-response ranks $1, 3, 7, 15, 31$, and centered floor, fixed midpoint plus PSD repair, and legal-midpoint ranks $8, 16, 32, 66, 136$. The first-response comparator uses the inverse-Hessian residual moment in (154): it preserves the covariance action on the leading r directions of both the covariance and that moment, together with $A^{-1}\theta$, for a union of at most $2r + 1$ directions. It uses no higher operator iterates. We record actual union dimensions and payload bytes; a common cap does not imply identical realized storage. No full-rank summary is silently replaced by exact statistics. Packed exact statistics require 1054728 bytes, exceeding all caps, while the original predictor uses 4096 bytes. These references are evaluated once per query, giving 2160 unique method–query–cap/reference records rather than counting repeated references.

The operator preparation uses optimized series weights, relative stopping tolerance 10^{-8}, absolute tolerance 10^{-12}, and at most 256 iterations. A failed sufficient contraction or stopping check returns no summary. Preparation is reused across ranks to reduce experiment cost, but every reported standalone encoding time includes the full preparation plus that rank's selection time. Baselines receive the same timing treatment. Preparation, selection, decoding, and additive-bound computation are timed separately in one CPU run. The operator applies its map with $O(nd + d^2)$ temporary storage; it never builds a $d^2 \times d^2$ matrix. The CSV reports numeric bytes in the returned encoder-only context, which is neither peak allocation nor process RSS and is unavailable to decoders.

Evaluation and numerical boundary. Retained optima are computed independently by least squares on the augmented ridge design. We check the quadratic excess against direct objective differences and check every applicable additive bound. The operator bound includes its stopping-tail correction and uses only the stored summary and request. Dividing any additive bound by the true retained optimum is explicitly an oracle-normalized diagnostic, not a decoder-available relative certificate. Floating-point residuals, eigenvalues, and bounds are numerical diagnostics, not machine-certified enclosures. An accuracy win requires at least a 1% reduction in excess and an absolute difference greater than 10^{-16}; smaller differences are ties. Unavailable outputs are counted separately from accuracy wins and losses.

Results: an improved response baseline, but no class-task advantage over midpoint summaries. All 32 configurations from the first four parameter pairs pass the operator preparation check; all eight $(0.01, 0.25)$ configurations fail its sufficient condition. Successful preparations take 12–26 iterations. These failures do not establish that another response envelope or stopping rule could not work.

Table 5 includes all 300 real-task comparisons per comparator, with 60 unavailable operator outputs kept separate. Among the 240 available comparisons, the operator improves on the first-response summary in 191 cases but loses to both midpoint comparators in every case. It also loses to centered

flooring in 232 cases. On the noiseless controls it wins 57 of 80 available comparisons against first response, 13 against centered flooring, and 9 against each midpoint comparator. Thus this benchmark provides no general practical memory advantage for the operator summary over the stronger spectral comparators. Its useful theoretical response envelope does not by itself ensure competitive accuracy after the full representation is charged to the same byte cap.

Table 5: Real binary-class tasks at common byte caps, across both tasks, three seeds, five parameter pairs, five caps, and two query types. Wins and losses use the stated 1% and 10^{-16} thresholds. There are 300 comparisons per row; unavailable outputs are not classified as accuracy losses. Actual bytes may differ below the cap.

Comparator	Operator wins	Ties	Operator losses	Unavailable
First response	191	5	44	60
Centered floor	7	1	232	60
Fixed midpoint + PSD	0	0	240	60
Legal midpoint	0	0	240	60

For timing, we compare methods on the same available operator configurations and count each encoder once per configuration and cap, deduplicating its two queries. Median standalone encoding times are 3.397 seconds for the operator, 0.179 for first response, and about 0.0633 for each spectral comparator. The operator's median preparation and selection components are 3.394 seconds and 0.997 milliseconds; the corresponding first-response values are 0.176 seconds and 1.132 milliseconds. Component medians need not sum to the median total. The operator's returned encoder-only context contains 8400904 numeric bytes, excluding input records and other temporary allocations; this is not a peak-memory measurement.

All applicable additive bounds pass the numerical checks, but the operator bound remains loose: its median bound-to-excess ratio is 346 on the 240 available real-task comparisons, versus 2.63 for legal midpoint on those same requests and caps. Exact packed-statistics decoding has maximum excess 1.35×10^{-27} against independent augmented least squares. Figure 8 summarizes accuracy, computation, and additive-bound looseness. The script also exports all per-pair memory curves and individual records for inspecting the reported negative comparisons.

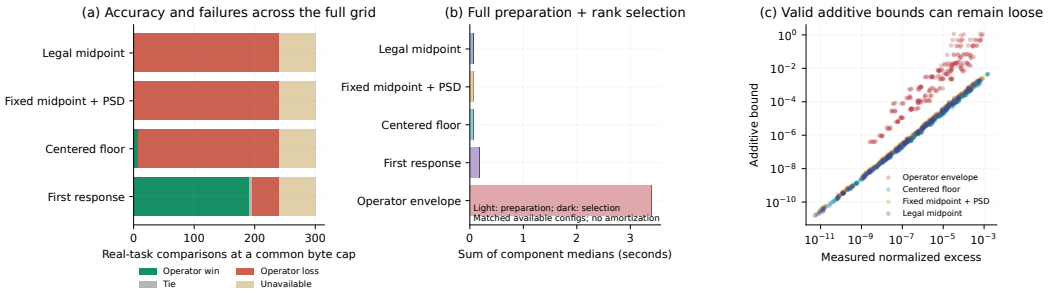

Figure 8: Practical response-envelope comparisons on the real binary-class tasks. Left: all comparisons, including failed sufficient conditions. Middle: sums of preparation and selection component medians on matched available operator configurations, with each encoder counted once per cap; full preparation is charged even when reused experimentally. Right: additive numerical bounds against measured excess. The operator improves its first-response baseline frequently, but has greater encoding cost and loses every available class-task accuracy comparison against the two midpoint summaries at these caps.

Table 6: DPP wins/ties/losses/unavailable outcomes against each comparator at a common numeric byte cap. Each sampler seed is a separate DPP outcome; the deterministic comparator is reused for pairing, not duplicated in the source evaluations. Frozen teacher, exploratory labels, and synthetic teacher have respectively 400, 200, and 96 comparisons per column and sampling budget.

Scope	s	First response	Floor	Fixed midpoint	Legal midpoint
Frozen teacher	128	140/5/255/0	27/0/373/0	7/1/392/0	6/1/393/0
Frozen teacher	512	162/5/233/0	55/1/344/0	17/2/381/0	18/2/380/0
Exploratory labels	128	98/5/97/0	0/0/200/0	0/0/200/0	0/0/200/0
Exploratory labels	512	99/5/96/0	0/0/200/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	67/3/26/0	11/0/85/0	21/1/74/0	2/1/93/0
Synthetic teacher	512	69/6/21/0	15/1/80/0	21/1/74/0	3/1/92/0

Z.1 Finite-sample DPP summaries

We fix the protocol before running the comparisons and evaluate ordinary DPP draws. Residual importance sampling from Appendix K.1 is evaluated separately. Sample covariances are computed in floating point for empirical comparison. The primary tasks have the public unit teacher $y_i = x_i^\top e_0$, without label rescaling. Two cached, pointwise-normalized ResNet-18 feature tasks (cat/dog and airplane/truck), with $n = 1000, d = 512$ and dataset seeds 11 and 23, give four primary cells. The original binary labels, evaluated only at seed 11 on each task, form an exploratory comparison outside the unit-teacher assumption. The five (λ, ρ) pairs are $(.001, .005), (.01, .025), (.1, .05), (.01, .125), (.001, .125)$. The latter two intentionally stress large deletion-to-regularization ratios. Each cell uses random and signed-original-residual deletion requests, shared by every method and budget, with $|U| = \lfloor \rho n \rfloor$.

The five numeric byte caps are exactly those of the response benchmark, $8(2d(2m)+(2m)^2+d+6)$ for $m = 2, 4, 8, 16, 32$. The DPP state stores $(V, Z, J, \delta, \theta_S, n)$ and costs $8(2dk + k^2 + d + 2)$ bytes, with union dimension $k \le \min(d, 2r)$. Each method uses its largest conservative rank that fits the same cap; actual payload size is recorded. Comparators are first response, centered floor, fixed midpoint with PSD repair, and legal midpoint. Exact statistics and the original model are evaluated once per query; deterministic comparators occur once per query and cap, not once per sampler seed.

Sampling uses $s \in \{128, 512\}$ and independent encoder seeds 101 and 211. These finite budgets are prescribed independently of outcomes. We record $\lceil 12(r+1)\log(2d/.05)\rceil$ and whether s reaches it, but do not infer a floating-point certificate from this flag. All preparation failures, decoder failures, and unfavorable outcomes remain in the results. Preparation may be reused across ranks and deletion fractions computationally; every standalone summary is charged its entire measured preparation cost. The measurements are single-run CPU timings, not latency benchmarks.

Synthetic controls use $n = 96, d = 32$, Gaussian features scaled either uniformly or geometrically from 1 to 10^{-3} and normalized by $\max(1, \|x_i\|)$, seeds 11 and 23, and the same unit teacher. Their pairs are $(.01, .05)$ and $(.0001, .125)$ and cap indices are $m = 1, 2, 3$. Together the frozen and synthetic grids contain 2936 evaluations. Retained optima are computed independently by augmented least squares. We report the quadratic retained excess and check its agreement with direct objective differences. Pairwise wins require both a one-percent improvement and an absolute difference exceeding 10^{-16}; failures are separate. The script `experiments_dpp.py` supports check-only and plot-only regeneration from the complete CSV results.

Observed accuracy and failures. All 2936 prescribed evaluations completed: 1392 DPP rows, 1392 deterministic comparator rows, and 152 exact/model references. There were no preparation, selection, or decoder failures. The 5568 paired comparisons are reported without filtering in Table 6. Increasing the sampler budget improves some outcomes against first response, but does not make DPP competitive with the strongest centered summaries at these byte caps. For example, on frozen unit-teacher tasks at $s = 512$, DPP wins 162 of 400 comparisons against first response, but only 18 of 400 against legal midpoint, with 380 losses. On exploratory class labels it loses all 200 comparisons against each centered comparator at each sampling budget. Synthetic controls favor DPP over first response more often, while legal midpoint still wins 92 of 96 comparisons at $s = 512$. These negative findings concern this fixed finite-sample construction and grid; they do not refute the theorem's larger-sample or asymptotic guarantees.

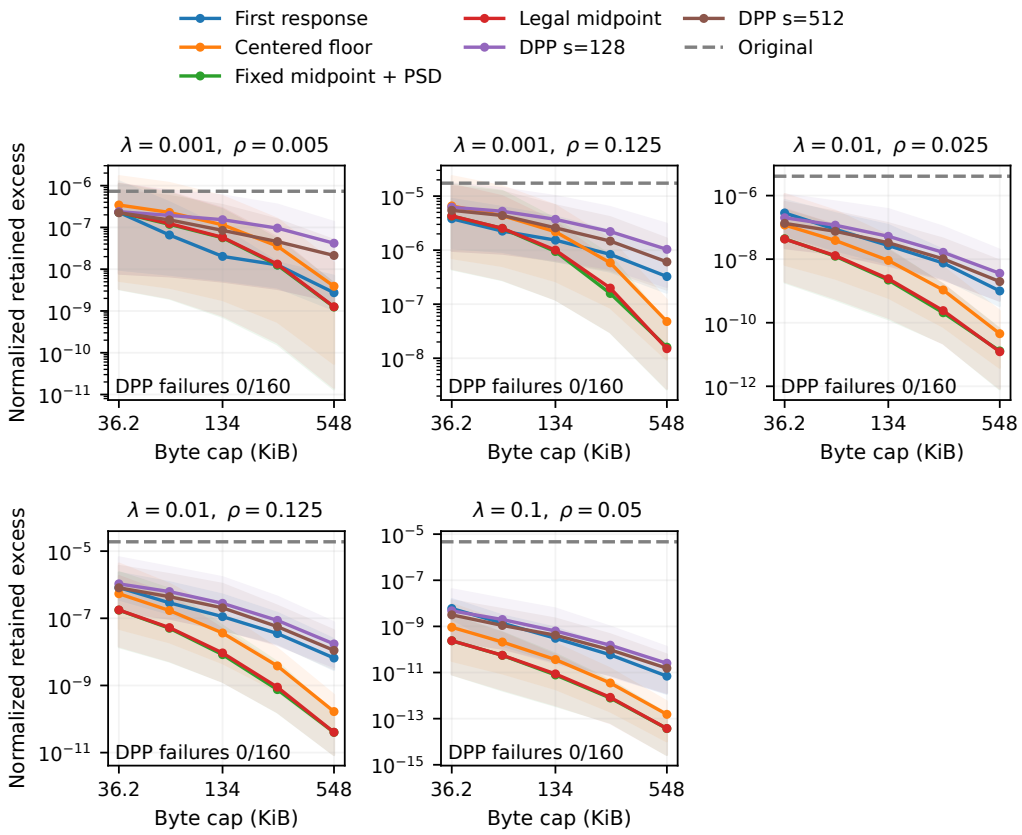

Figure 9: Frozen unit-teacher DPP tasks: all five predeclared parameter pairs, with medians and full ranges. Failure counts include every prescribed sampler outcome. The original-model line is a reference, not a capped summary; exact statistics lie above the sketch caps. Excesses below 10^{-16} are displayed at 10^{-16}.

Costs and numerical diagnostics. On frozen teacher tasks, median standalone encoding times are 1.079 and 2.710 seconds for $s = 128$ and 512, compared with 0.181 seconds for first response and approximately 0.060 seconds for each centered comparator. For exploratory class labels the corresponding DPP times are 0.973 and 2.694 seconds, versus 0.181 seconds and 0.061–0.062 seconds; synthetic teacher times are 0.186 and 0.818 seconds, versus 0.000808 seconds and 0.000358–0.000364 seconds. These medians use one measurement per task/seed/label/λ/cap/method/sampler configuration, excluding duplicate deletion fractions and queries. They include the full preparation cost and rank-selection cost. Decoder timers stop before retained-data oracle evaluation. Every frozen DPP payload is 32 bytes below its common cap; there is no uncharged exact-statistics branch. The maximum exact-summary excess against the independent retained least-squares oracle is 7.3×10^{-28}.

All 80 distinct unit-teacher preparations pass the reported numerical teacher diagnostic; none of the 24 exploratory preparations do. The largest sampled fit norm is 0.625 on frozen teacher data and $1 + 4.5 \times 10^{-15}$ on synthetic teacher data, compared with 29.76 for exploratory class labels. Maximum sampled centered norms are respectively 0.397, 0.705, and 27.42. Thus the observed class-label sample magnitudes do not satisfy the unit-teacher bounded-sample premise. The worst sampled row condition number is approximately 886 in the synthetic controls; maximum interpolation residuals remain below 4.2×10^{-13} across all scopes. No $s = 128$ configuration reaches the recorded theorem sample threshold. At $s = 512$, 80 of 400 frozen-teacher rows and all 96 synthetic teacher rows reach it; reaching this threshold is not itself a verified concentration event or a machine-certified enclosure. The smallest measured completion-error eigenvalue is -3.4×10^{-16} to rounding precision, and every additive-certificate field is intentionally unasserted.

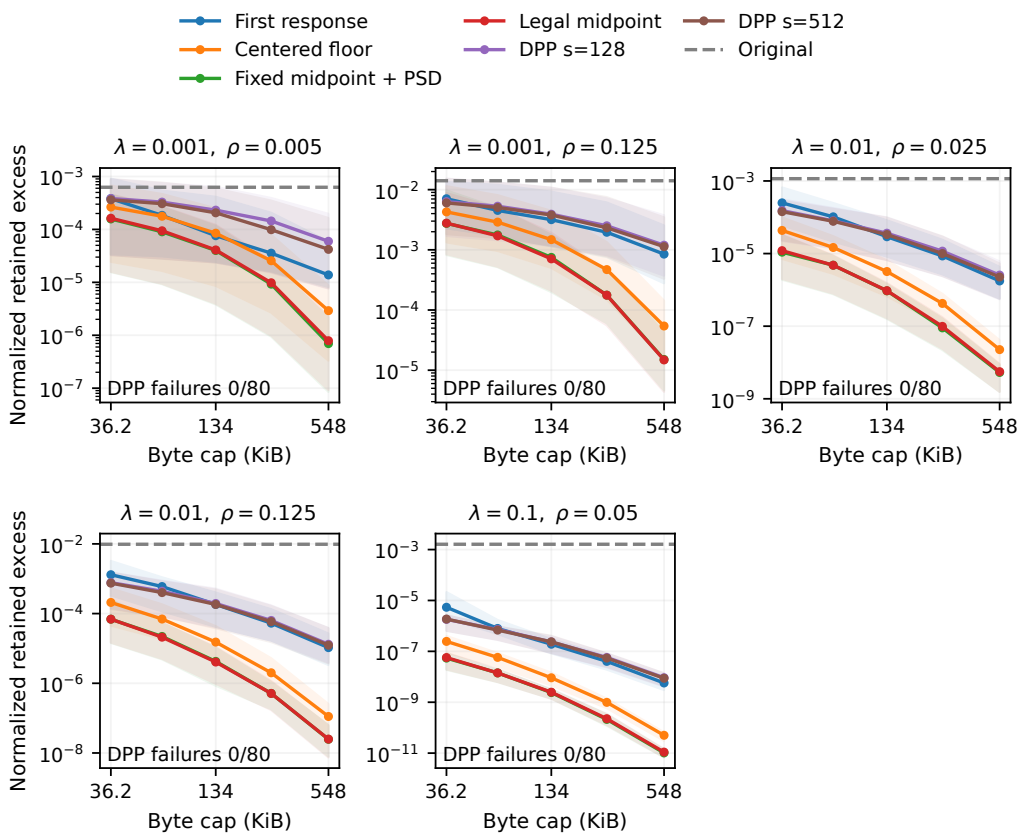

Figure 10: Exploratory class-label DPP tasks: all five predeclared parameter pairs, with medians and full ranges. These labels are outside the asserted unit-teacher scope. Failure counts include every prescribed sampler outcome; the original model is an uncapped reference.

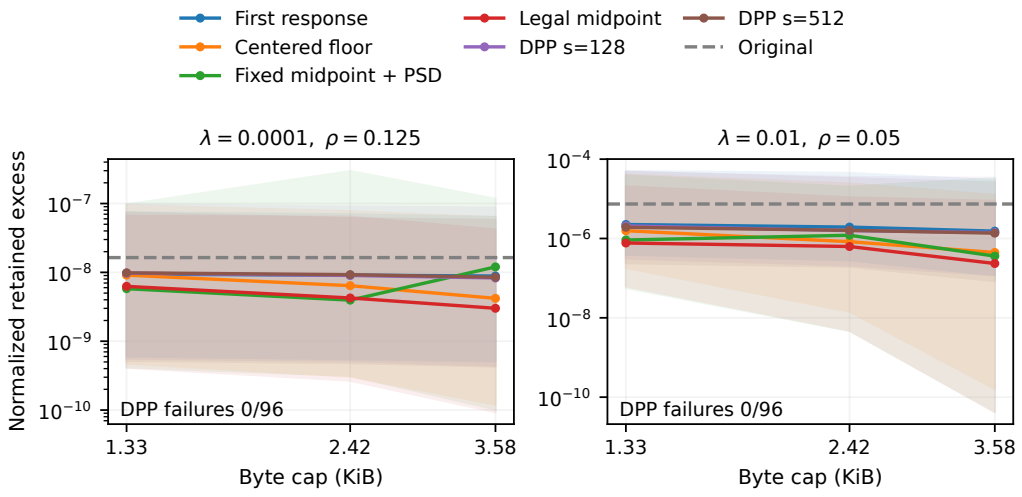

Figure 11: Synthetic unit-teacher DPP controls: both predeclared parameter pairs, with medians and full ranges across flat and ill-conditioned feature spectra, dataset seeds, and deletion queries. Failure counts include every prescribed sampler outcome; the original model is an uncapped reference.

Z.2 RESIDUAL-IMPORTANCE SAMPLING AT FIXED BYTE CAPS

Using a protocol fixed before the comparisons, we repeat the datasets, label families, parameter pairs, deletion queries, byte caps, sample counts $s \in \{128, 512\}$, and encoder seeds 101 and 211 from Section Z.1. These comprise frozen unit-teacher tasks, frozen exploratory class labels, and synthetic unit-teacher controls. Class labels satisfy the importance theorem's bounded-label assumption. The floating-point experiment measures performance; certifying the sampling law, arithmetic, and subspace bounds requires the enclosures in Appendix K.1.2.

Two methods use the spectral importance sampler of Appendix K.1.1. The importance-union method replaces the ordinary empirical covariance by $s^{-1}\sum_j H_{J_j}$ and retains the existing union of leading covariance and response eigenspaces, with the same floor rule. Its comparison with ordinary DPP is a sampling-law ablation. The importance-mixed method instead takes the leading k eigenvectors of $\Sigma + s^{-1}\sum_j H_{J_j}/R$ and measures the largest omitted covariance eigenvalue for its completion floor. At each cap, union uses its largest conservative rank r, and mixed uses its largest fitting factor dimension $k = \min\{d, 2r\}$. The second comparison changes both subspace construction and floor; it is not a pure sampling-law ablation. Both states store only $(V, Z, J, \delta, \theta_S, n)$, with charged numeric size $8(2dk + k^2 + d + 2)$ bytes. No exact-statistics branch is uncharged.

The new grid has 2784 evaluations, 1392 per method. Ordinary-DPP and the four strong comparator evaluations are reused from the existing complete `dpp_runs.csv`; they are not rerun or duplicated in the new source CSV. Pairing uses task, dataset seed, label family, numerical (λ, ρ), deletion strategy, and byte cap, with matching sample count and encoder seed for ordinary DPP. Each new method has five comparator pairs per row; the mixed method is also paired with union, giving 15312 comparisons. Wins and losses require a one-percent difference and an absolute difference exceeding 10^{-16}. Tables retain wins, ties, losses, and unavailable outcomes separately by scope and sample count.

Preparation is shared computationally between the two new methods and cached by dataset, labels, λ, sample count, and encoder seed; it does not depend on deletion fraction or query. Each standalone encoding is charged the complete measured preparation and its own selection cost. CPU thread limits are set before NumPy is imported, following the existing benchmark. Timing summaries retain the first measurement per encoding identity, excluding repeated deletion fractions and queries. They are single-run CPU timings. Decoder timing stops before independent retained-data evaluation. Retained optima are recomputed by augmented least squares, and quadratic excess is checked against the direct objective difference.

All preparation, selection, and decoder failures remain in the incremental source CSV. Diagnostics include the residual scale R, normalized covariance trace, dependent-draw and exact-rank fallback counts, mixture probability and mass error, sampled condition numbers, and completion residuals. A dependent draw is retained as its prescribed zero contribution; an unresolved independent fit aborts preparation without replacing draws. The script `experiments_importance.py` supports check-only validation and plot-only regeneration, which requires the complete 2784-row grid with unique identities and valid payload caps.

Accuracy across the fixed grid. The sampling-law ablation gives moderate, scope-dependent improvements. Against ordinary DPP, importance-union wins 230 and 234 of 400 frozen-teacher comparisons at $s = 128$ and 512, with 123 and 108 losses. On class labels it wins 81 and 93 of 200, with 76 and 58 losses; on synthetic teachers it wins 58 and 54 of 96, with 32 and 36 losses. Thus the improvement is not uniform across configurations, and increasing the budget does not improve every aggregate count.

The mixed construction is more sensitive to the finite sampling budget. At $s = 128$ it loses to union in 258 of 400 frozen-teacher and 127 of 200 class-label comparisons. At $s = 512$ it wins 236 and 149 of those comparisons, respectively. Synthetic controls give mixed 54 and 55 wins out of 96 at the two budgets. Since these methods share the prepared covariance, this comparison tests the subspace and floor choices jointly, rather than another change in sampling law.

The strongest centered summaries remain substantially better at these caps. Both importance methods lose all 200 class-label comparisons against each of centered floor, fixed midpoint, and legal midpoint at each sampling budget. Against legal midpoint on frozen teachers, each importance

Table 7: Importance-union wins/ties/losses/unavailable outcomes against each reused comparator at the same numeric byte cap. Each column has 400, 200, or 96 pairs per sampling budget for frozen teacher, class labels, or synthetic teacher, respectively.

Scope	s	Ordinary	First response	Floor	Fixed	Legal
Frozen teacher	128	230/47/123/0	145/6/249/0	26/1/373/0	4/1/395/0	6/0/394/0
Frozen teacher	512	234/58/108/0	189/4/207/0	62/0/338/0	23/0/377/0	23/0/377/0
Class labels	128	81/43/76/0	98/6/96/0	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	512	93/49/58/0	103/10/87/0	0/0/200/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	58/6/32/0	75/1/20/0	16/1/79/0	27/1/68/0	8/2/86/0
Synthetic teacher	512	54/6/36/0	75/3/18/0	18/0/78/0	29/0/67/0	8/1/87/0

Table 8: Importance-mixed wins/ties/losses/unavailable outcomes against each reused comparator at the same numeric byte cap. The protocol changes both the selected subspace and its floor.

Scope	s	Ordinary	First response	Floor	Fixed	Legal
Frozen teacher	128	152/15/233/0	102/7/291/0	29/2/369/0	7/0/393/0	6/0/394/0
Frozen teacher	512	244/24/132/0	201/5/194/0	64/0/336/0	23/0/377/0	23/0/377/0
Class labels	128	68/4/128/0	77/3/120/0	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	512	156/6/38/0	142/1/57/0	0/0/200/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	64/1/31/0	70/6/20/0	16/0/80/0	24/0/72/0	8/0/88/0
Synthetic teacher	512	59/7/30/0	74/3/19/0	15/0/81/0	24/0/72/0	7/1/88/0

method has only 6 and 23 wins out of 400 at $s = 128$ and 512, with 394 and 377 losses. On synthetic teachers, union has 8 wins at each budget, while mixed has 8 and 7; their corresponding losses are 86/87 and 88/88 out of 96. The importance law therefore improves the ordinary finite-sample construction in parts of the grid without making it competitive with the strongest centered baseline overall. These observations concern the prescribed finite budgets and empirical completion rules; they neither refute the theoretical rank guarantee nor establish its floating-point hypotheses. Only 280 of the 2784 new rows meet their recorded conservative sample threshold, so the majority are below it; meeting that threshold still does not verify the concentration event or a floating-point certificate.

Importance-union's median standalone encoding costs are 1.22–1.94 times the reused ordinary-DPP medians across the six scope/budget groups. The modest accuracy gains therefore carry additional preparation cost, while the centered baselines retain their much smaller measured encoding costs.

Observed outcomes and costs. All 2784 prescribed new evaluations are recorded, with 0 unavailable outcomes. Tables 7–9 report all 15312 pairs. The outcome definitions, scopes, budgets, and method choices were fixed before these results. Frozen teacher, importance union: median standalone encoding times are 1.397 and 3.318 seconds for $s = 128$ and 512, respectively. Frozen teacher, importance mixed: median standalone encoding times are 1.513 and 3.414 seconds for $s = 128$ and 512, respectively. Class labels, importance union: median standalone encoding times are 1.408 and 3.284 seconds for $s = 128$ and 512, respectively. Class labels, importance mixed: median standalone encoding times are 1.505 and 3.384 seconds for $s = 128$ and 512, respectively. Synthetic teacher, importance union: median standalone encoding times are 0.3615 and 1.439 seconds for $s = 128$ and 512, respectively. Synthetic teacher, importance mixed: median standalone encoding times are 0.3618 and 1.44 seconds for $s = 128$ and 512, respectively. The largest recorded normalized covariance trace is 1. Frozen teacher preparations have maximum dependent-draw count 0, exact-rank fallback count 0, and sampled condition number 80.75. Class labels preparations have maximum dependent-draw count 0, exact-rank fallback count 0, and sampled condition number 69.83. Synthetic teacher preparations have maximum dependent-draw count 0, exact-rank fallback count 0, and sampled condition number 503.2. These diagnostic maxima and timing medians retain the fixed grid; they do not verify a floating-point concentration or arithmetic event.

Table 9: Importance-mixed versus importance-union at the same cap and using the same prepared importance covariance. All prescribed outcomes appear as wins/ties/losses/unavailable.

Scope	s	Wins	Ties	Losses	Unavailable
Frozen teacher	128	117	25	258	0
Frozen teacher	512	236	23	141	0
Class labels	128	65	8	127	0
Class labels	512	149	8	43	0
Synthetic teacher	128	54	11	31	0
Synthetic teacher	512	55	13	28	0

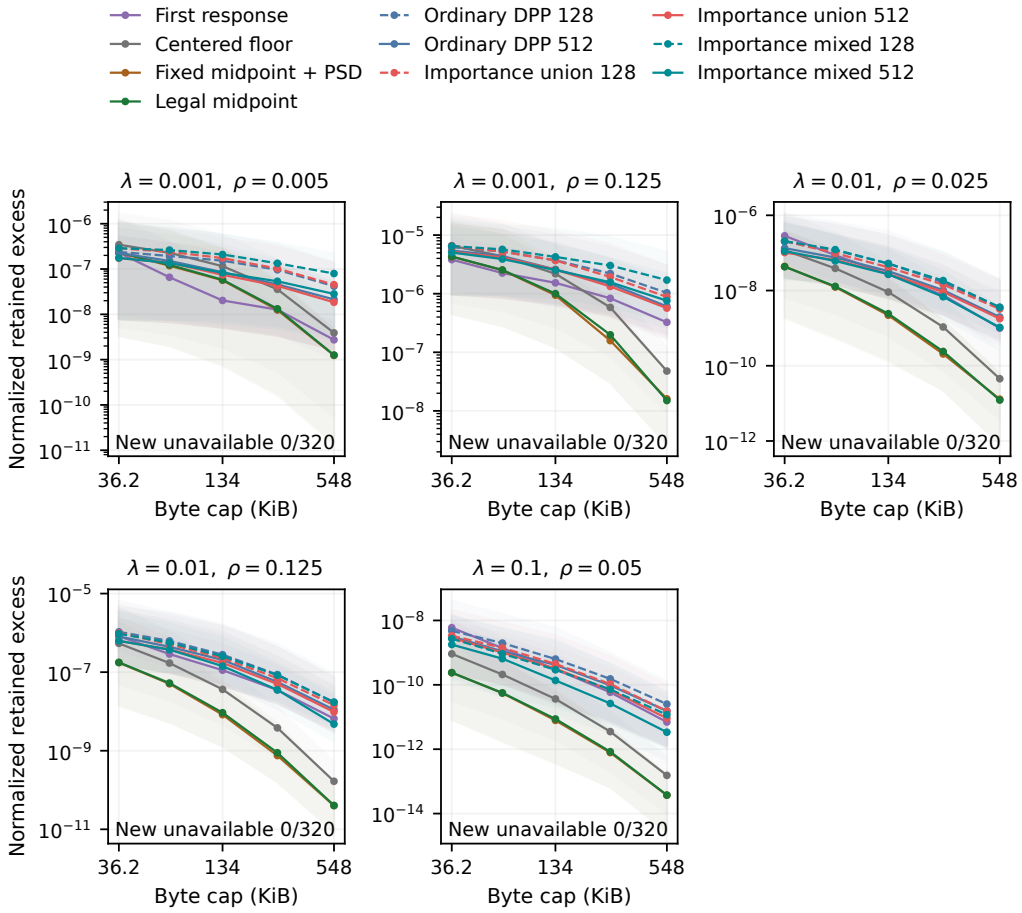

Figure 12: Frozen unit-teacher importance summaries, ordinary DPP, and the four strong comparators: all five parameter pairs, medians and full ranges, with unavailable outcomes counted. Values below 10^{-16} are displayed at 10^{-16}.

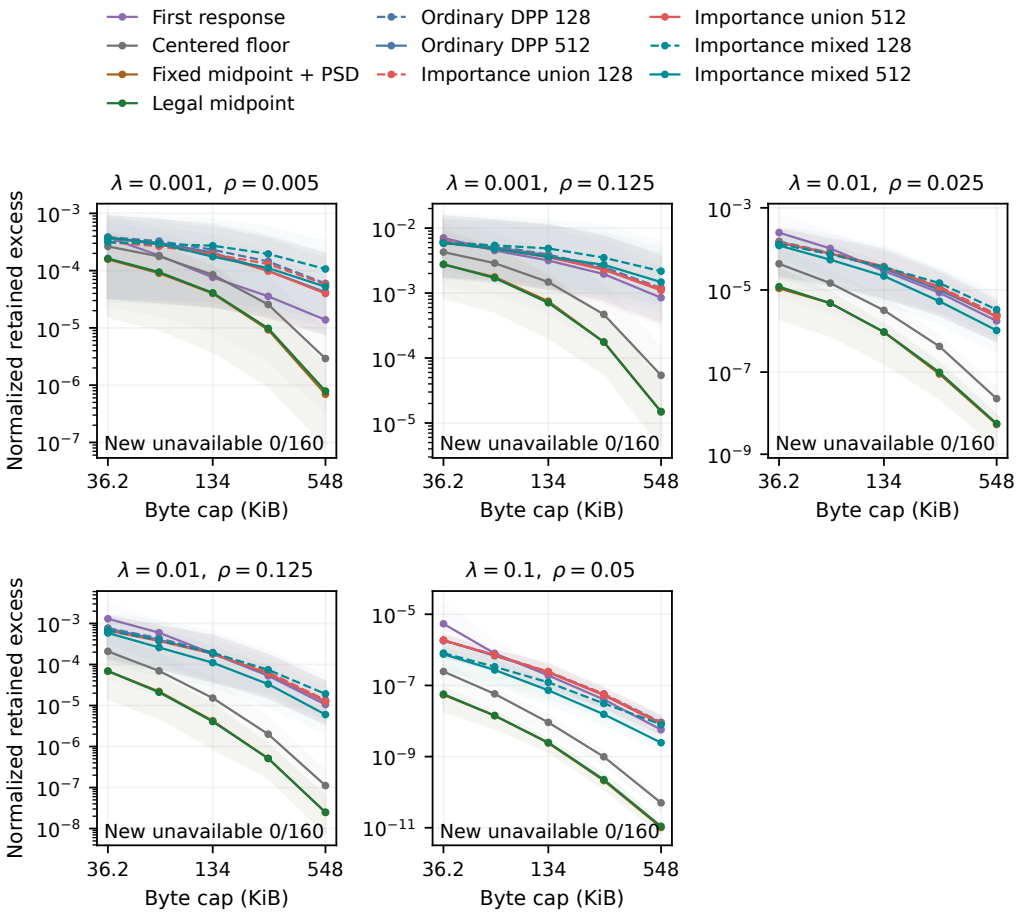

Figure 13: Exploratory class-label comparisons on the entire predeclared grid. Finite sampling and the recorded numerical diagnostics do not constitute a floating-point certificate.

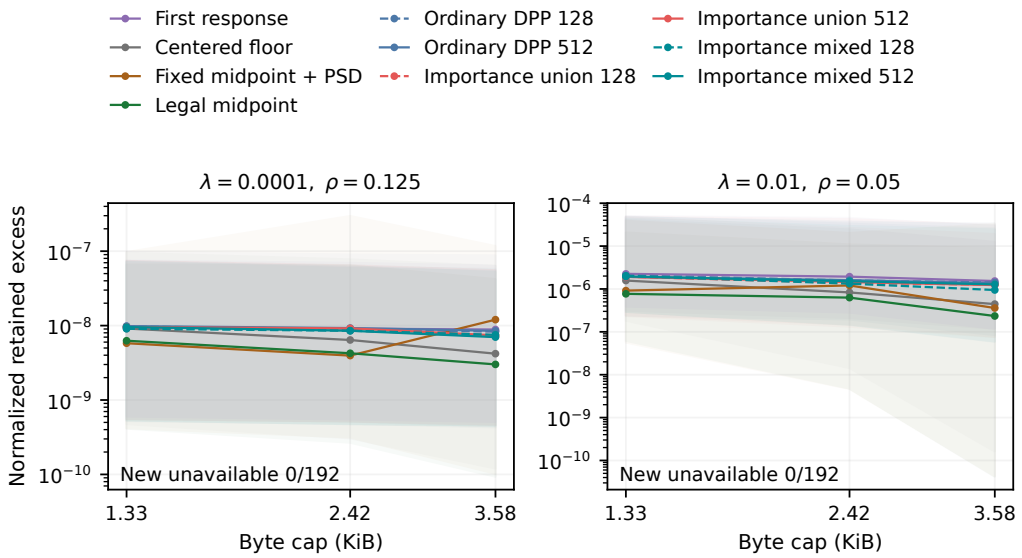

Figure 14: Synthetic unit-teacher controls: both parameter pairs, feature spectra, dataset seeds, and deletion queries. Medians and full ranges include both prescribed sampling budgets.

Z.3 SIGNED COMPLETIONS AT THE FIXED BYTE CAPS

We fix the comparison protocol in advance, reusing the datasets, label families, dataset seeds, (λ, ρ) pairs, deletion queries, five frozen-data byte caps, three synthetic byte caps, sample counts $s \in \{128, 512\}$, and encoder seeds 101 and 211 from Sections Z.1 and Z.2. Ranks, tolerances, and configurations are fixed independently of the outcomes.

Six methods combine three signed encoders with two decoders: residual Nyström floor, residual Nyström legal midpoint, residual metric floor, residual metric legal midpoint, response-only metric floor, and response-only metric legal midpoint. The first encoder uses the shorted floor residual of Appendix K.2; its matrix ingredients are classical generalized Schur-complement and Nyström constructions. The second uses the fixed response metric $2M/R + I/(m+1)$, where M is the empirical importance covariance and R its residual scale. The third fixes the covariance head rank at zero and allocates its entire factor budget to r response-metric correction columns, using metric $2M/R + I/(r+1)$. This is a fixed endpoint comparison, without split tuning. Its covariance floor may remain constant, so no uniform cubic spectral-trace rate is asserted for this endpoint. All three retain only $(F, L, \delta, \theta_S, n)$. Floor and midpoint share exactly the same encoded state. A full-covariance branch is used only when the encoder contract and charged factor shapes determine exact recovery; it is not triggered by an observed small error. The midpoint uses the interval implied by the signed completion and a legal deletion. It does not receive the original response subspace or retained records. These choices do not assert a universal ordering of actual retained excess.

The two balanced encoders select the largest integer $m \in \{0, \dots, d\}$ satisfying $8(d \min\{2m, d\} + d + 2)$ bytes. The resulting ranks are $4, 8, 16, 33, 68$ at $d = 512$, and $2, 4, 6$ at $d = 32$. The response-only endpoint instead selects the largest $r \le d$ satisfying $8(dr + d + 2)$ bytes, giving $8, 16, 32, 66, 136$ at $d = 512$ and $4, 8, 13$ at $d = 32$. Every actual factor entry is charged, including full-rank states. Actual dimensions and bytes are recorded and checked against the unchanged public cap. The sample threshold $\lceil 12(r+1)\log(2d/.05) \rceil$ uses the selected correction rank r (equal to m for the balanced encoders) and is recorded descriptively; meeting it does not certify the sampled or floating-point event.

The six methods have 1392 evaluations each, giving 8352 new rows. The complete 2936 ordinary-DPP/reference/baseline rows and 2784 importance-union/mixed rows are reused, without rerunning or duplicating them in the new source CSV. Every new row is paired with ordinary DPP, importance union, importance mixed, and the four established strong baselines at the same task, labels, numerical parameter pair, query, and byte cap. Sampled comparators additionally match sample count and encoder seed. These give 58464 comparisons. Pairing midpoint against floor within each encoder adds 4176; pairing balanced metric against Nyström within each decoder adds 2784; pairing response-only metric against balanced metric within each decoder adds another 2784, for 68208 pairs. Wins and losses require the existing one-percent and 10^{-16} absolute separation. Every tie and unavailable outcome remains in the scope- and sample-count-specific tables.

All six methods share the importance preparation, cached by dataset, labels, λ, sample count, and encoder seed, comprising 104 distinct preparations. The cache excludes deletion fraction and query. Each standalone state is charged the full measured preparation and its own encoder selection cost; sharing floor and midpoint does not discount this cost. Timing summaries retain the first measurement per encoding identity after excluding repeated deletion fractions and queries. CPU thread settings precede the NumPy import. Decoder timing ends before posterior diagnostics and the independent augmented least-squares retained oracle. Quadratic excess is checked against the direct objective difference.

Preparation, selection, decoder, oracle-check, and posterior failures remain explicitly recorded. Posterior calculations use the scalar-curvature specialization of the legal-interval bounds, without a generalized lower-curvature solve. They are numerical diagnostics with their own status, time, and bound-violation fields; they do not constitute floating-point certificates. A posterior failure does not erase a successfully evaluated decoder output. The source CSV is written incrementally. The script `experiments_completion.py` provides a check-only mode that validates grid identities, byte accounting, comparison counts, and small independent retained solves before the full run, and a plot-only mode that requires the complete unique 8352-row grid.

The main comparison is against the established legal-midpoint baseline. The floor-to-midpoint comparisons isolate decoder changes at a fixed encoded state; metric-to-Nyström comparisons isolate

the chosen response geometry at each decoder. Response-only versus balanced metric tests the fixed head/correction allocation. All three scopes and both sampling budgets are reported, including adverse results.

Observed accuracy at the same caps. The smaller Nyström floor improves on importance union in all 1392 prescribed pairs and on ordinary DPP in 1379, with two ties and 11 losses. These comparisons change the covariance completion and the rank affordable at the fixed cap while retaining the predeclared sampling budget. They are not a pure change of sampling law. The gain does not make these summaries competitive with the strongest centered baseline overall: Nyström floor has 105 wins, four ties, and 1283 losses against legal midpoint. The reused DPP states use their implemented $2dk + k^2 + d + 2$-entry representation. The exact invariant-block rewrite of the old completion in Appendix K.2, with $3dm + d + 2$ entries when both selected blocks have rank m, is not an evaluated comparator. These results therefore do not compare against every more economical encoding of that same old matrix. All six new methods lose every class-label comparison to legal midpoint: 200 losses per method at each sampling budget.

On frozen teachers, Nyström floor has 13/2/385 and 53/0/347 wins/ties/losses against legal midpoint at $s = 128$ and 512. Metric midpoint has 25/2/373 and 45/0/355, respectively. Synthetic teachers remain mixed: Nyström floor has 21/0/75 and 18/2/76, while metric midpoint has 13/1/82 at each budget. Table 10 gives every new method's comparison by scope and sample count. The additional tables retain all seven reused comparators.

Decoder, geometry, and allocation comparisons. At a fixed encoded state, midpoint versus floor gives 1098/20/274 wins/ties/losses for Nyström and 1264/8/120 for balanced metric. These gains are not uniform: on synthetic teachers, Nyström midpoint loses to its floor in 55 and 56 of 96 cases at the two sampling budgets. The response-only endpoint has 72/39/1281 across its midpoint-versus-floor comparisons, so its midpoint usually worsens actual error on this grid.

The preferred response geometry also depends on the decoder. Balanced metric floor versus Nyström floor gives 157/204/1031, whereas metric midpoint versus Nyström midpoint gives 1106/175/111. The envelope selection criterion therefore does not identify the more accurate floor decoder in this experiment. The response-only endpoint versus balanced metric gives 102/5/1285 for floor and 31/3/1358 for midpoint. This fixed allocation comparison includes the predeclared change in the rank-dependent metric floor. No alternative allocation was selected after observing these results. The reference-curvature utility rule of Appendix K.3.3 is not one of the six methods evaluated here.

Costs and precision qualifications. Across the three encoders, median standalone encoding times range approximately from 1.22 to 3.06 seconds on frozen teachers, 1.28 to 2.80 seconds on class labels, and 0.200 to 0.754 seconds on synthetic teachers, including the full importance preparation each time. The reused legal-midpoint medians are 0.0596, 0.0615, and 0.000364 seconds, respectively. These are the prescribed single-run CPU measurements, not optimized latency estimates. All payloads fit their fixed caps. Only 320 of 8352 rows meet the recorded sample threshold. The observed outcomes concern the fixed finite sample budgets and ordinary floating-point implementations; they do not establish the concentration event or the precision assumptions of the theoretical results.

Observed outcomes and numerical diagnostics. All 8352 prescribed new evaluations and 68208 pairs are recorded. There are 0 unavailable decoder/oracle outcomes, 0 unavailable numerical posterior values, and 0 observed numerical posterior bound violations. Exactly 320 rows meet their recorded sample threshold. This count is descriptive and does not establish a concentration event, a floating-point certificate, or the balanced cubic rank rule for the response-only endpoint. The scope tables include every prescribed comparison; no rank allocation or data cell was selected from these outcomes. Nyström floor has 0 unavailable outcomes among 1392 evaluations. Nyström midpoint has 0 unavailable outcomes among 1392 evaluations. Metric floor has 0 unavailable outcomes among 1392 evaluations. Metric midpoint has 0 unavailable outcomes among 1392 evaluations. Response only floor has 0 unavailable outcomes among 1392 evaluations. Response only midpoint has 0 unavailable outcomes among 1392 evaluations. Generalized Schur-complement order between completion errors does not imply order of realized retained excess; the empirical decoder and allocation ablations are reported separately.

Table 10: All six signed summaries against the established legal-midpoint baseline at the same byte cap. Cells are wins/ties/losses/unavailable. Each method has 400, 200, or 96 pairs per sampling budget for frozen teachers, class labels, or synthetic teachers, respectively.

Scope	Decoder	s	Nyström	Metric	Response only
Frozen teacher	Floor	128	13/2/385/0	10/1/389/0	3/1/396/0
Frozen teacher	Floor	512	53/0/347/0	28/1/371/0	18/0/382/0
Frozen teacher	Midpoint	128	20/4/376/0	25/2/373/0	2/1/397/0
Frozen teacher	Midpoint	512	34/4/362/0	45/0/355/0	10/0/390/0
Class labels	Floor	128	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	Floor	512	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	Midpoint	128	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	Midpoint	512	0/0/200/0	0/0/200/0	0/0/200/0
Synthetic teacher	Floor	128	21/0/75/0	16/1/79/0	19/0/77/0
Synthetic teacher	Floor	512	18/2/76/0	14/1/81/0	21/1/74/0
Synthetic teacher	Midpoint	128	8/1/87/0	13/1/82/0	4/1/91/0
Synthetic teacher	Midpoint	512	4/2/90/0	13/1/82/0	6/0/90/0

Table 11: Legal midpoint versus floor using the identical encoded state, for each of the three encoders. Every prescribed comparison is included as wins/ties/losses/unavailable.

Scope	s	Nyström	Metric	Response only
Frozen teacher	128	367/3/30/0	397/1/2/0	15/16/369/0
Frozen teacher	512	259/8/133/0	372/2/26/0	10/11/379/0
Class labels	128	200/0/0/0	200/0/0/0	0/7/193/0
Class labels	512	200/0/0/0	200/0/0/0	0/4/196/0
Synthetic teacher	128	35/6/55/0	45/4/47/0	25/1/70/0
Synthetic teacher	512	37/3/56/0	50/1/45/0	22/0/74/0

Table 12: The two fixed geometry/allocation comparisons at each decoder: balanced metric versus Nyström, and response-only versus balanced metric. Cells are wins/ties/losses/unavailable.

		Metric versus Nyström		Response versus metric	
Scope	s	Floor	Midpoint	Floor	Midpoint
Frozen teacher	128	34/47/319/0	325/43/32/0	1/1/398/0	0/0/400/0
Frozen teacher	512	28/45/327/0	328/37/35/0	12/1/387/0	0/0/400/0
Class labels	128	25/48/127/0	164/27/9/0	0/0/200/0	0/0/200/0
Class labels	512	23/26/151/0	161/24/15/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	25/17/54/0	70/21/5/0	42/2/52/0	16/0/80/0
Synthetic teacher	512	22/21/53/0	58/23/15/0	47/1/48/0	15/3/78/0

Table 13: Nyström floor versus all seven reused comparators at the same cap: wins/ties/losses/unavailable for every scope and sample count.

Scope	s	DPP	Union	Mixed	First	Floor	Fixed	Legal
Frozen teacher	128	397/0/3/0	400/0/0/0	400/0/0/0	314/2/84/0	78/0/322/0	16/1/383/0	13/2/385/0
Frozen teacher	512	399/0/1/0	400/0/0/0	400/0/0/0	357/0/43/0	139/1/260/0	52/4/344/0	53/0/347/0
Class labels	128	200/0/0/0	200/0/0/0	200/0/0/0	189/1/10/0	2/0/198/0	0/0/200/0	0/0/200/0
Class labels	512	200/0/0/0	200/0/0/0	197/3/0/0	188/0/12/0	0/1/199/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	91/2/3/0	96/0/0/0	95/0/1/0	88/0/8/0	40/1/55/0	36/1/59/0	21/0/75/0
Synthetic teacher	512	92/0/4/0	96/0/0/0	95/1/0/0	89/1/6/0	40/3/53/0	37/2/57/0	18/2/76/0

Table 14: Nyström legal midpoint versus all seven reused comparators, with the same complete pairing protocol.

Scope	s	DPP	Union	Mixed	First	Floor	Fixed	Legal
Frozen teacher	128	400/0/0/0	400/0/0/0	399/0/1/0	332/0/68/0	188/4/208/0	23/3/374/0	20/4/376/0
Frozen teacher	512	393/1/6/0	382/2/16/0	379/1/20/0	345/1/54/0	196/5/199/0	36/2/362/0	34/4/362/0
Class labels	128	200/0/0/0	200/0/0/0	200/0/0/0	194/0/6/0	70/2/128/0	0/0/200/0	0/0/200/0
Class labels	512	200/0/0/0	200/0/0/0	200/0/0/0	195/0/5/0	73/3/124/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	73/0/23/0	71/2/23/0	65/1/30/0	68/1/27/0	34/1/61/0	32/1/63/0	8/1/87/0
Synthetic teacher	512	66/2/28/0	66/1/29/0	67/3/26/0	68/3/25/0	32/2/62/0	31/1/64/0	4/2/90/0

Table 15: Balanced response-metric floor versus all seven reused comparators: wins/ties/losses/unavailable.

Scope	s	DPP	Union	Mixed	First	Floor	Fixed	Legal
Frozen teacher	128	382/4/14/0	386/3/11/0	381/1/18/0	270/4/126/0	50/2/348/0	10/0/390/0	10/1/389/0
Frozen teacher	512	371/1/28/0	365/1/34/0	355/0/45/0	315/2/83/0	80/4/316/0	29/2/369/0	28/1/371/0
Class labels	128	200/0/0/0	200/0/0/0	199/1/0/0	187/0/13/0	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	512	200/0/0/0	200/0/0/0	196/4/0/0	188/0/12/0	0/0/200/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	86/3/7/0	92/1/3/0	86/1/9/0	86/0/10/0	37/2/57/0	33/1/62/0	16/1/79/0
Synthetic teacher	512	88/2/6/0	88/0/8/0	87/4/5/0	86/2/8/0	36/4/56/0	32/1/63/0	14/1/81/0

Table 16: Balanced response-metric legal midpoint versus all seven reused comparators, retaining every prescribed outcome.

Scope	s	DPP	Union	Mixed	First	Floor	Fixed	Legal
Frozen teacher	128	400/0/0/0	400/0/0/0	399/0/1/0	336/2/62/0	204/4/192/0	26/3/371/0	25/2/373/0
Frozen teacher	512	398/0/2/0	397/1/2/0	392/1/7/0	353/0/47/0	225/4/171/0	50/0/350/0	45/0/355/0
Class labels	128	200/0/0/0	200/0/0/0	200/0/0/0	196/0/4/0	78/7/115/0	0/0/200/0	0/0/200/0
Class labels	512	200/0/0/0	200/0/0/0	200/0/0/0	196/0/4/0	84/4/112/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	76/0/20/0	76/3/17/0	73/2/21/0	74/2/20/0	40/0/56/0	39/1/56/0	13/1/82/0
Synthetic teacher	512	70/2/24/0	76/1/19/0	76/3/17/0	78/1/17/0	40/5/51/0	37/1/58/0	13/1/82/0

Table 17: The fixed response-only floor endpoint versus all seven reused comparators. The covariance head rank is zero at every cap.

Scope	s	DPP	Union	Mixed	First	Floor	Fixed	Legal
Frozen teacher	128	7/1/392/0	7/0/393/0	7/2/391/0	4/0/396/0	10/1/389/0	3/0/397/0	3/1/396/0
Frozen teacher	512	30/0/370/0	23/2/375/0	23/2/375/0	18/3/379/0	39/0/361/0	18/0/382/0	18/0/382/0
Class labels	128	0/0/200/0	0/0/200/0	0/0/200/0	3/0/197/0	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	512	3/0/197/0	1/1/198/0	2/0/198/0	4/0/196/0	0/0/200/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	86/2/8/0	85/0/11/0	80/4/12/0	87/1/8/0	34/2/60/0	35/1/60/0	19/0/77/0
Synthetic teacher	512	86/0/10/0	79/3/14/0	81/1/14/0	86/3/7/0	36/2/58/0	33/1/62/0	21/1/74/0

Table 18: The fixed response-only legal-midpoint endpoint versus all seven reused comparators. Adverse results are retained without changing the allocation.

Scope	s	DPP	Union	Mixed	First	Floor	Fixed	Legal
Frozen teacher	128	1/1/398/0	0/0/400/0	0/0/400/0	3/0/397/0	8/0/392/0	2/0/398/0	2/1/397/0
Frozen teacher	512	6/0/394/0	2/0/398/0	2/1/397/0	8/0/392/0	19/0/381/0	9/0/391/0	10/0/390/0
Class labels	128	0/0/200/0	0/0/200/0	0/0/200/0	2/0/198/0	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	512	1/0/199/0	0/1/199/0	0/0/200/0	4/0/196/0	0/0/200/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	51/1/44/0	50/1/45/0	42/2/52/0	57/1/38/0	23/2/71/0	26/0/70/0	4/1/91/0
Synthetic teacher	512	45/3/48/0	47/3/46/0	46/3/47/0	56/5/35/0	26/0/70/0	26/1/69/0	6/0/90/0

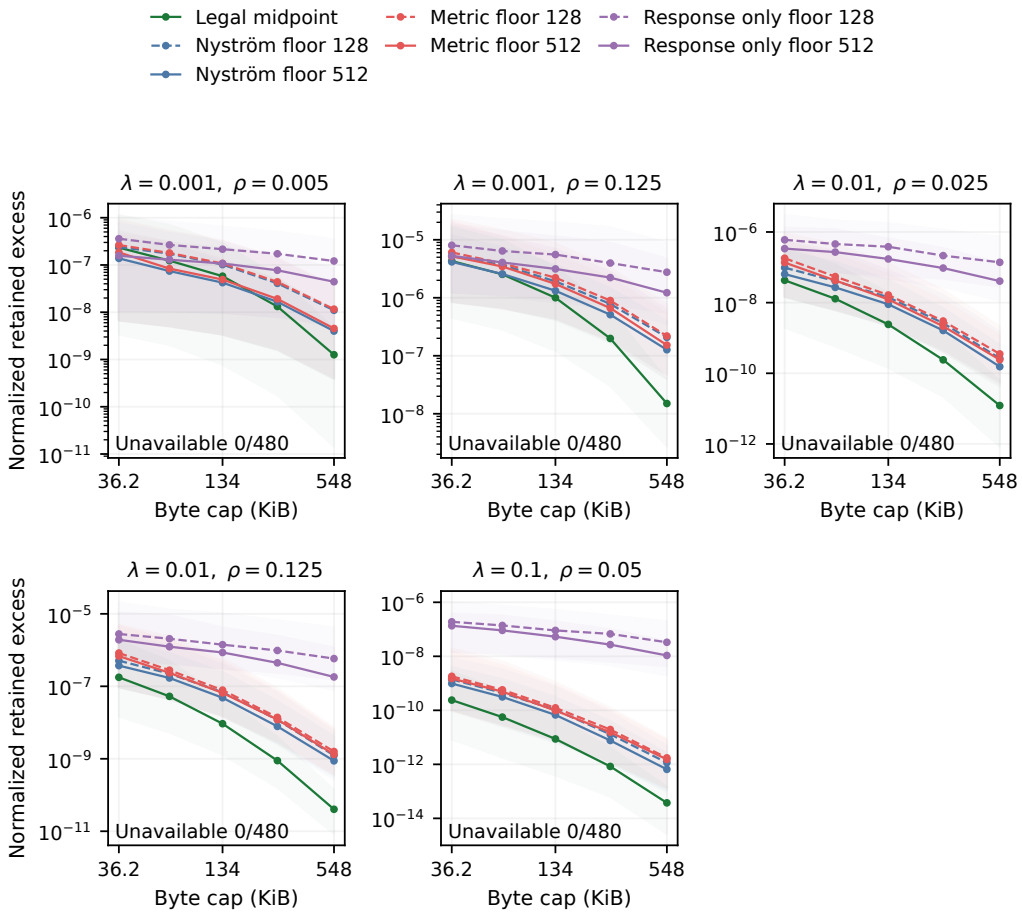

Figure 15: Frozen unit-teacher floor decoders and the established legal midpoint: all five parameter pairs, with medians and full ranges across the prescribed cells. Both sampling budgets are shown.

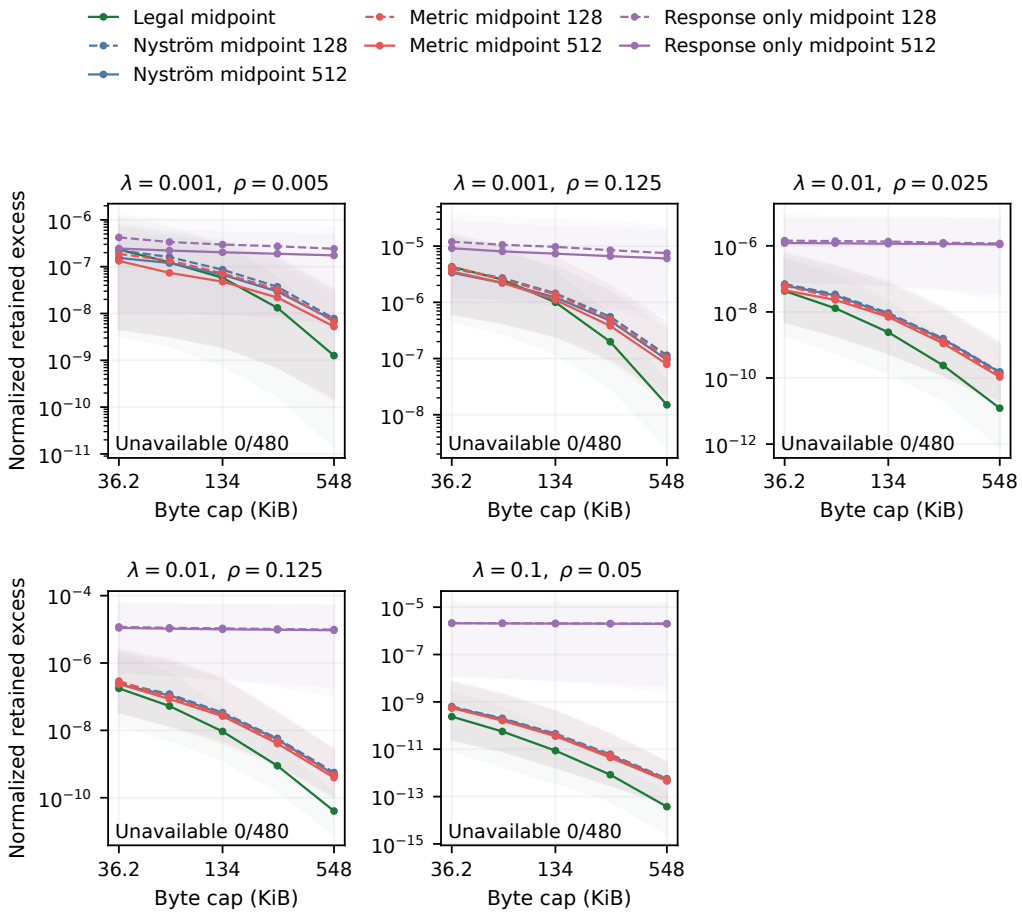

Figure 16: Frozen unit-teacher legal-midpoint decoders on the same complete grid. The response-only endpoint uses its fixed head-zero allocation; no allocation was tuned to these outcomes.

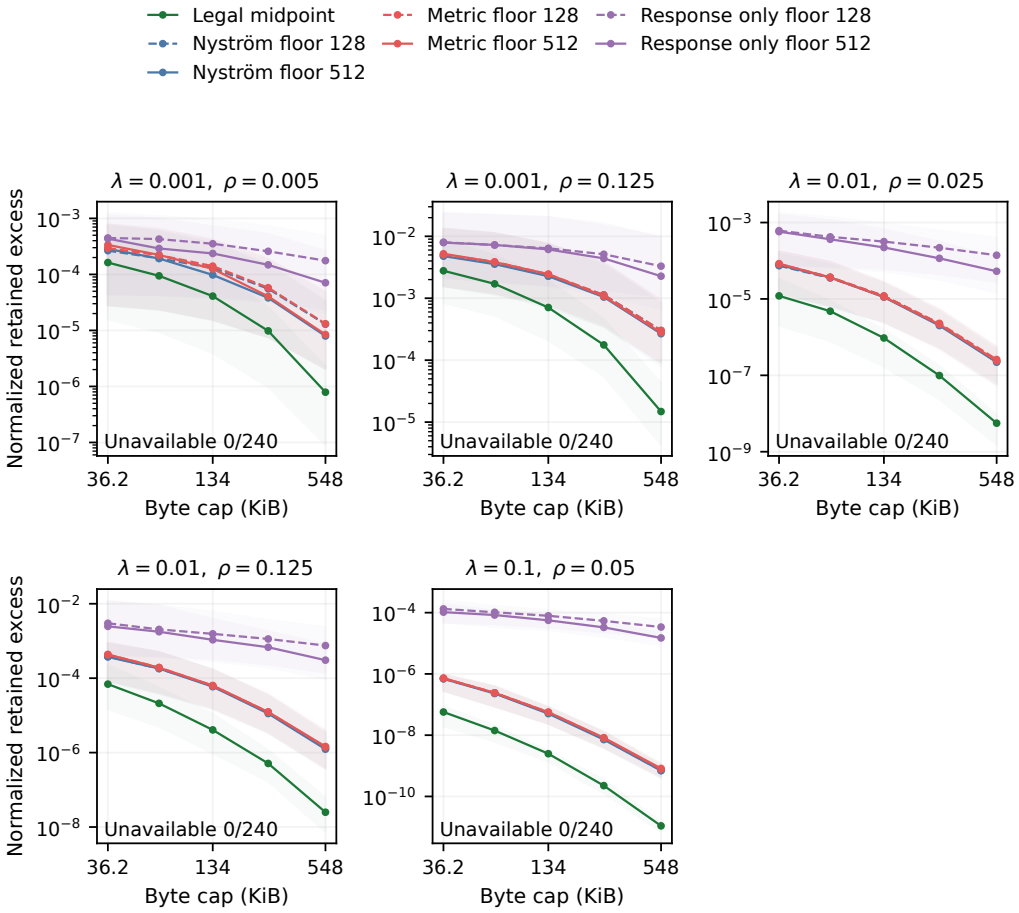

Figure 17: Class-label floor decoders against legal midpoint. The full prescribed ranges and both sample budgets are retained.

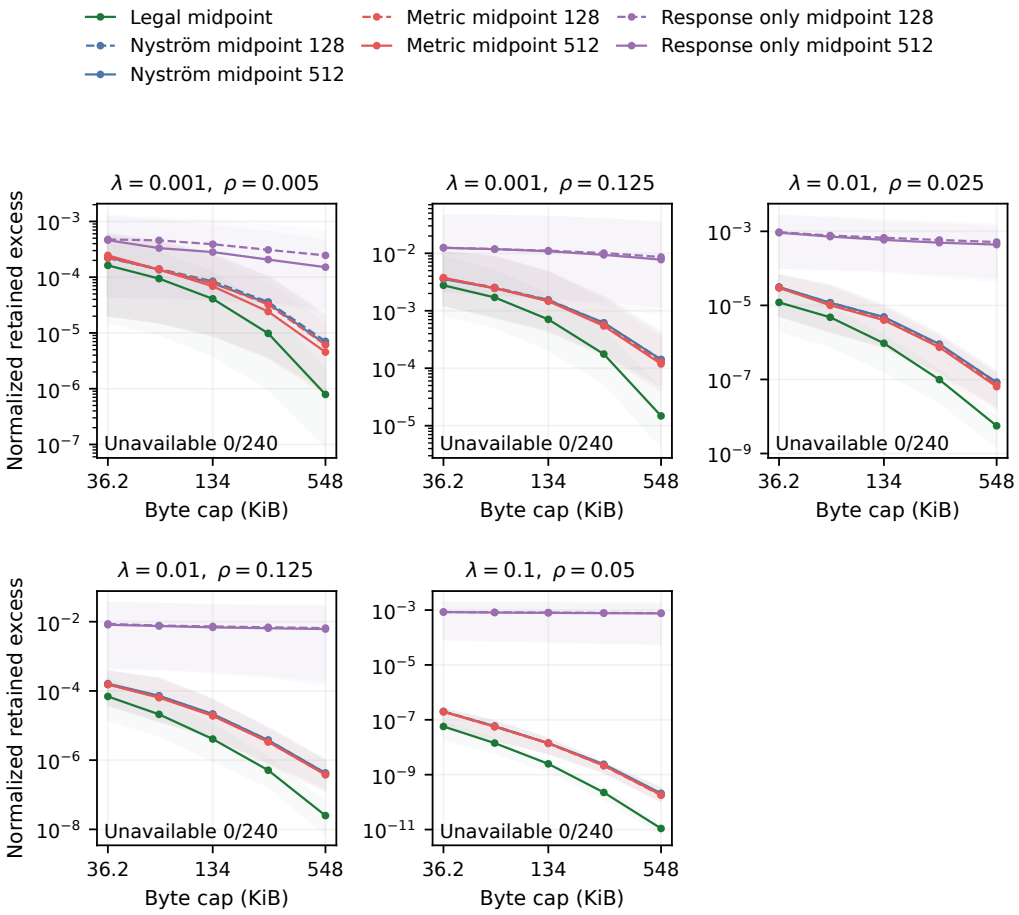

Figure 18: Class-label legal-midpoint decoders. All six new methods lose their class-label comparisons to the established legal-midpoint baseline at these caps.

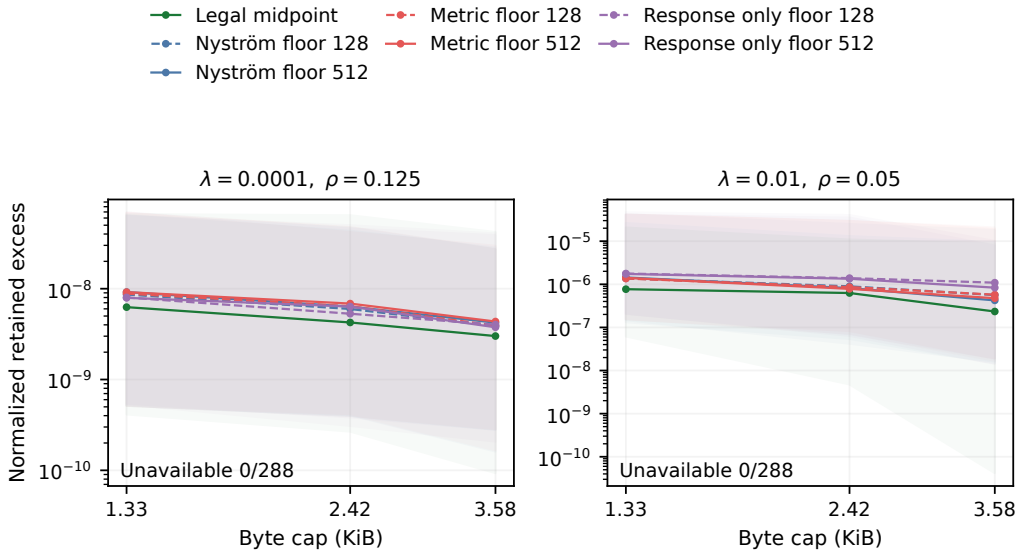

Figure 19: Synthetic unit-teacher floor decoders: both parameter pairs, spectra, dataset seeds, deletion queries, and sample budgets.

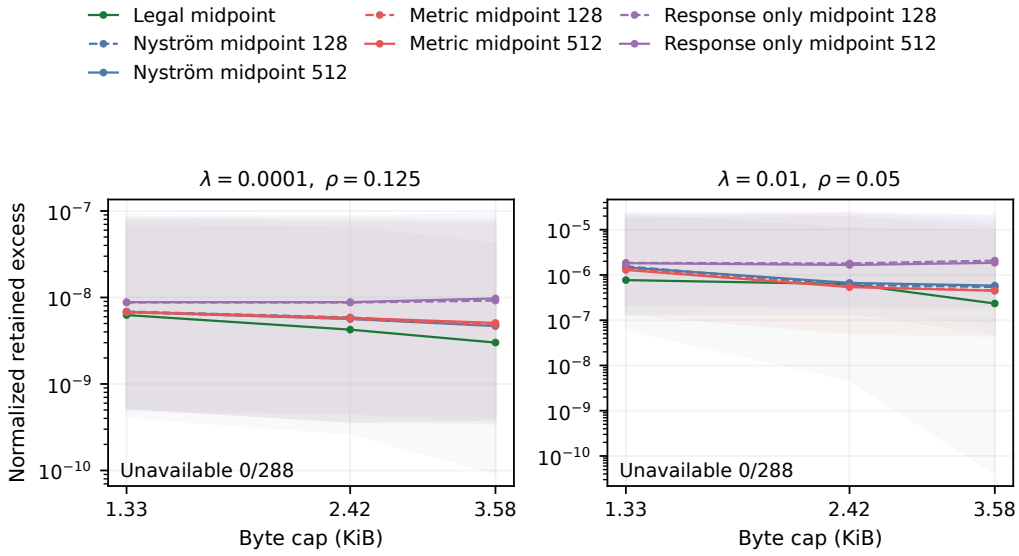

Figure 20: Synthetic unit-teacher legal-midpoint decoders with medians and full ranges. All six figures count unavailable outcomes explicitly; values below 10^{-16} are displayed at 10^{-16}. These numerical results do not constitute floating-point certificates.

Z.4 REFERENCE-UTILITY SELECTION AT THE FIXED BYTE CAPS

We fix the protocol before running the comparisons, reusing the datasets, label families, dataset seeds, parameter pairs, deletion queries, five frozen-data byte caps, three synthetic byte caps, sampling budgets $s \in \{128, 512\}$, and encoder seeds 101 and 211 from Sections Z.1, Z.2, and Z.3. Configurations, ranks, reference scales, and numerical tolerances are fixed independently of the new outcomes.

Four new methods use the same importance-prepared response covariance M, residual scale R, and public balanced ranks $r = m$. `utility_floor` uses $\delta = \sigma_{m+1}(\Sigma)$; `utility_inflated_floor` uses $\delta = 2\sigma_{m+1}(\Sigma)$. Their matching `utility_midpoint` and `utility_inflated_midpoint` methods change only the decoder, sharing identical encoded states and byte counts. Both encoders preserve the exact covariance head before making the feasible tail correction and use the reference $R_{\mathrm{ref}} = \Sigma + \lambda I$, that is, the predeclared reference scale $c = 1$. The covariance head may require both positive and negative factors under inflation. Every stored column is charged; neither method receives an uncharged full-covariance branch or encoder context at decoding. Only $(F_+, L_-, \delta, \theta_S, n)$ persists as numeric arrays. Each state is decoded with both the centered floor and the legal midpoint rules. The public method tag distinguishes ordinary and inflated floors; it carries no outcome-dependent metadata. Utility states do not inherit the metric encoder's full-covariance factor-count contract. The zero-floor utility endpoint does identify exact covariance.

Positive-factor directions are exact covariance head directions. For the inflated floor, positive eigenvalues of $L_-L_-^\top$ strictly below $\delta/2$ additionally identify exact negative-head directions, with source eigenvalues $\delta - s_i(L_-)^2$. Tail-correction eigenvalues are at least $\delta/2$. Equality directions are also exact in ideal arithmetic, but the numerical implementation omits a fixed tolerance band around that threshold. Any recovered subset remains a valid known head under the exact-arithmetic contract. The remaining block uses the same deletion-legality Schur-complement lower bound and completion upper bound as the earlier legal decoder. Recovery ranks, omitted threshold modes, tolerances, and head/off-block residuals are recorded. An invalid numerical interval is declined without replacing the decoder. Posterior computations use this interval and retain their own numerical status and timing; they are not floating-point certificates.

On the positive support of $G = \delta I - \Sigma_{\mathrm{tail}}$, set $q_i = g_i/(c(\sigma_i + \lambda) + g_i)$, $A = \mathrm{diag}(q_i)$, $B = 2M/R + I/(r+1)$, and $D = R_{\mathrm{ref,tail}}^{1/2} B_{\mathrm{tail}} R_{\mathrm{ref,tail}}^{1/2}$. The top r eigenvectors U of ADA give $H = \mathrm{orth}(\mathrm{diag}([q_i(1-q_i)]^{-1/2})U)$ and the feasible correction $G^{1/2}HH^\top G^{1/2}$. When $R = 0$, the exact zero-response endpoint requires $M = 0$ and $\theta_S = 0$, and defines $B = I/(r+1)$ without division by zero. Full heads, zero floors, empty positive support, and correction ranks larger than that support are handled explicitly. Numerical rank truncation is reported and does not silently inherit the nominal rank-r approximation guarantee.

The scalar reference loss is the exact-arithmetic expression

$$\ell_{\mathrm{ref}}(E) = \left\| B^{1/2} E (R_{\mathrm{ref}} + E)^{-1} R_{\mathrm{ref}} (R_{\mathrm{ref}} + E)^{-1} E B^{1/2} \right\|_2, \qquad E = \widehat{\Sigma} - \Sigma,$$

evaluated numerically. Selection diagnostics include this loss, the singular-value lower bound $\lambda_{r+1}(ADA)$, the positive-support condition number $\kappa(A)$, and the exact-arithmetic factor $\gamma_A^2 = (\kappa(A)+1)^2/(4\kappa(A))$. For the inflated floor, $\kappa(A) \le 2$ and hence $\gamma_A^2 \le 9/8$ in exact arithmetic. PSD-order residuals, eigenspace residuals, orthogonalization cutoffs, discarded singular values, and rank-truncation flags remain numerical diagnostics; no floating-point certificate is asserted.

The source covariance and the public deletion budget define

$$\eta = \frac{\rho(1+\lambda)}{\sigma_{\min}(\Sigma) + \lambda}.$$

This quantity does not use the observed deletion request or its outcome. The script records η and whether $\eta < 1$. Only under the theorem's remaining hypotheses and this strict stability condition does the $c = 1$ reference comparison transfer with the product factor $\gamma_A^2/(1-\eta)^2$, at most $9/[8(1-\eta)^2]$ for the inflated floor. Neither encoder uses the balanced reference scale $c = \sqrt{1-\eta}$; the sharper balanced-reference product factor is therefore not claimed for this grid. Keeping $c = 1$ for both encoders isolates floor inflation. Reference-loss approximation does not imply a universal

ordering of realized retained excess on individual deletions, nor does this floor-decoder reference theorem itself establish the same approximation factor for the legal-midpoint decoder.

All four methods use precisely the completed Nyström-floor rank policy: choose the largest $m \le d$ fitting $8(d\min\{2m, d\}+d+2)$ bytes and set $r = m$. This gives $m = 4, 8, 16, 33, 68$ at $d = 512$ and $m = 2, 4, 6$ at $d = 32$. The new representation uses at most $m+\min\{r, d-m\}$ columns, so the same conservative policy is valid despite signed heads. Actual factor dimensions, numeric elements, and bytes are checked. The sample threshold $\lceil 12(r+1)\log(2d/.05)\rceil$ is recorded descriptively. Meeting it does not establish a concentration event. The realizable unit-teacher scopes and exploratory class-label scope remain separate. General labels require the residual-scale envelope event and its general-label constants; they do not inherit the unit-teacher trace bound $1/4$.

Each new method has 1392 evaluations, giving exactly 5568 new rows. The source reuse comprises all 8352 completed signed-completion rows and all 5720 earlier rows (2936 ordinary-DPP/reference/baseline and 2784 importance rows), for 14072 reused rows. Each new evaluation has seven legacy comparators (ordinary DPP, importance union, importance mixed, and the four established baselines) and all six completed signed-completion comparators. These yield $5568 \times 7 = 38976$ legacy pairs and $5568 \times 6 = 33408$ completion pairs. Comparing midpoint against floor within each of the two floor variants adds 2784 decoder pairs. Comparing inflated against ordinary utility within each decoder adds 2784 inflation pairs, giving exactly 77952 pairs in total. The primary legal-midpoint comparison is the 5568-pair subset against `centered_legal_midpoint`; the within-family comparisons against residual Nyström floor and residual metric floor are an 11136-pair subset, without double counting.

Pair identities match task, labels, dataset seed, (λ, ρ), deletion strategy, and byte budget. Sampled methods additionally match sample count and encoder seed; deterministic baselines retain their original sample markers. Recomputed retained optima and query metadata are checked against the original grid. Wins and losses use the established one-percent and 10^{-16} absolute separation. All ties and unavailable outcomes remain in tables for every scope and both sampling budgets. Compact tables report all 13 comparator methods, both decoder comparisons, and both inflation comparisons, with separate primary and within-family CSV summaries. Decoder comparisons hold the stored state fixed; inflation comparisons hold the decoder fixed.

All four methods share exactly 104 importance preparations, cached by dataset, labels, λ, sample count, and encoder seed. Preparation excludes deletion fraction and query. Encoder selection is cached by that identity, floor multiplier, and selected ranks, so each encoder's floor and midpoint reuse a single measured selection. Every standalone state is charged the full preparation time plus its selection time, including selection diagnostics. Timing summaries deduplicate repeated fractions and queries; preparation summaries count each shared preparation once. Decoder timing excludes posterior diagnostics, the independent augmented least-squares oracle, and the direct-objective check. Posterior time and unavailability or violation counts are reported separately; a posterior failure does not erase a successfully evaluated decoder answer. Quadratic retained excess must agree with direct objective minus the independently fitted retained optimum. Preparation, selection, decoder, objective-check, and posterior failures remain explicit outcomes. No failed sampling draw is replaced.

The new source CSV is written incrementally without modifying any completed source grid. `experiments_utility.py -check-only` checks identities, pair multiplicities, fixed rank policy, storage, missing/duplicate rejection, and small independent retained solves with both decoders and their numerical posterior values. The full run is separate from this protocol check. `-plot-only` requires the complete unique new grid and regenerates the comparison tables and diagnostic summaries without rerunning encoders.

Utility experiment outcomes. All 5568 prescribed new evaluations and 77952 pairs are recorded; 0 evaluations are unavailable. Table cells report wins/ties/losses/unavailable, retaining every scope and sampling budget. Reference losses, conditioning factors, and posterior values are numerical diagnostics; no sampled event or floating-point certificate is asserted. There are 0 unavailable posterior values and 0 numerical posterior bound violations.

Interpretation of the completed grid. The four methods do not establish superiority to the strong legal-midpoint baseline. Ordinary utility floor has 69 wins, 5 ties, and 1318 losses against that base-

line; ordinary utility midpoint has 64/7/1321. Inflated utility floor has 1/0/1391, and inflated utility midpoint has 0/0/1392. Every method loses all 200 class-label comparisons at each sampling budget. At 512 samples in the frozen-teacher scope, ordinary utility floor and midpoint win 26/400 and 31/400 comparisons, respectively. These scope-specific results and all thirteen external comparators appear in the tables below.

The fixed ablations locate a substantial cost of inflation. Inflated floor loses 1384/1392 pairs to ordinary utility floor; inflated midpoint loses 1368/1392 to ordinary utility midpoint. Midpoint decoding improves its own floor in 1217/1392 ordinary and 1224/1392 inflated pairs, but this does not recover strong-baseline accuracy. At the original floor, utility selection wins 317 pairs and loses 848 to response-metric floor; with midpoint decoding it wins 221 and loses 822 to response-metric midpoint. Thus incorporating reference curvature is not a uniform empirical improvement over the earlier metric selection.

There are 912 distinct selected states, 456 per floor variant. For ordinary utility, the numerical ratio of reference loss to the singular-value lower bound has median 1.00021 and maximum 1.06499; for inflated utility, the corresponding values are 1.00042 and 1.08504. On this grid, the selected states are already close to the measured lower bound for the fixed reference objective. Further optimization has limited room to improve that objective; translating such gains into actual deletion accuracy remains uncertain. The experiment does not isolate the contributions of the common response envelope, its product pairing with curvature, the covariance-head allocation, and the admissible base. The $9/8$ theorem compares corrections within the inflated base; the inflation ablation illustrates why that comparison cannot imply superiority over the original base or other equal-budget summaries. All statements about these measured ratios remain floating-point diagnostics, not certified lower bounds on a numerical computation.

All 5568 evaluations respect their prescribed caps. There are no unavailable decoder or posterior values, observed posterior violations, or reported orthogonalization rank truncations. Exactly 256/5568 rows meet the recorded conservative sampling threshold, and 960/5568 meet the original crude curvature condition; neither count certifies a floating-point event. The largest disagreement between quadratic excess and independent direct objective difference is 6.89×10^{-16}. At 512 samples, median full encoding time on frozen teachers is 2.920 s for ordinary utility and 2.928 s for inflated utility, compared with 0.0596 s for the reused legal-midpoint baseline. These are single-environment CPU measurements, including selection diagnostics and all preparation costs. An independent reconstruction checked all 77952 pair identities and outcomes, all scope records, and every count in the sixteen generated tables.

Separate source-geometry diagnostics. After the fixed performance run, a separate analysis applies Theorem 56 to all 38 existing source/budget configurations, requiring 26 source/regularizer computations. It uses only original features, regularization, and the prescribed deletion budget. No encoder, sample draw, rank, reference scale, deletion request, or stored performance row is changed. The mode `-source-geometry` records every ridge leverage score summary, L_k, d_{eff}, a_{lev}, and both logarithmic response bounds in `utility_source_geometry.csv`. The effective budget in these new formulas is $q = \lfloor \rho n \rfloor / n$; the older crude condition retains its original prescribed ρ.

The leverage condition applies numerically to 20/38 configurations, compared with 6/38 for the crude minimum-eigenvalue condition. The other 18 configurations are explicitly unavailable for this new certificate. On the 20 applicable configurations, the new common multiplier is smaller than the old one in every case. This increases the reach of the theoretical condition; it does not change any predictor or turn the performance losses above into gains. In particular, a positive curvature comparison constant need not be close to one: here a_{lev} ranges from 0.0919 to 0.7298, and the unscaled reference comparison factor $9/(8a_{\mathrm{lev}}^2)$ has median 35.8 and maximum 133.3. The source computations agree with the independent trace identity to 2.14×10^{-13} and assert no floating certificate.

Table 19: Separate source-geometry analysis. N counts all configurations in the row; $L_k < 1$ and $\eta < 1$ count the new and old sufficient conditions. The maximum leverage mass includes every configuration. Both logarithmic multiplier ranges use the same subset satisfying $L_k < 1$; a dash means that subset is empty. Natural logarithms are used. Full source values, including unavailable new certificates, remain in the CSV.

Scope	Source	N	$L_k < 1$	$\eta < 1$	$\max L_k$	$\log(1 + \Gamma_{\mathrm{lev}})$	$\log(1 + \Gamma_{\mathrm{old}}) \mid_{L_k<1}$
Class labels	Air/truck	5	3	1	16.3	0.282–2.02	0.526–5.03
Class labels	Cat/dog	5	3	1	16.1	0.277–2.38	0.526–5.03
Frozen teacher	Air/truck	10	6	2	16.7	0.282–2.27	0.526–5.03
Frozen teacher	Cat/dog	10	6	2	16.2	0.274–2.38	0.526–5.03
Synthetic teacher	Flat	4	0	0	4.82	–	–
Synthetic teacher	Ill-cond.	4	2	0	4.17	1.03–1.12	4.35–4.35

Table 20: Utility variants against the established legal midpoint. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	9/1/390/0	15/2/383/0	0/0/400/0	0/0/400/0
Frozen teacher	512	26/2/372/0	31/2/367/0	1/0/399/0	0/0/400/0
Class labels	128	0/0/200/0	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	512	0/0/200/0	0/0/200/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	18/0/78/0	7/1/88/0	0/0/96/0	0/0/96/0
Synthetic teacher	512	16/2/78/0	11/2/83/0	0/0/96/0	0/0/96/0

Table 21: Utility variants against signed response-metric midpoint. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	4/0/396/0	78/106/216/0	0/0/400/0	0/0/400/0
Frozen teacher	512	20/2/378/0	68/78/254/0	2/0/398/0	0/0/400/0
Class labels	128	0/0/200/0	28/91/81/0	0/0/200/0	0/0/200/0
Class labels	512	0/0/200/0	20/58/122/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	52/2/42/0	15/7/74/0	24/1/71/0	5/0/91/0
Synthetic teacher	512	47/3/46/0	12/9/75/0	24/0/72/0	3/1/92/0

Table 22: Utility variants against signed Nyström midpoint. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	5/2/393/0	264/35/101/0	0/0/400/0	1/0/399/0
Frozen teacher	512	27/2/371/0	236/43/121/0	8/1/391/0	17/4/379/0
Class labels	128	0/0/200/0	149/31/20/0	0/0/200/0	0/0/200/0
Class labels	512	0/0/200/0	142/26/32/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	56/2/38/0	40/9/47/0	28/1/67/0	5/0/91/0
Synthetic teacher	512	58/2/36/0	45/4/47/0	31/1/64/0	5/1/90/0

Table 23: Utility variants against centered flooring. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	51/0/349/0	205/3/192/0	10/2/388/0	25/0/375/0
Frozen teacher	512	68/0/332/0	220/3/177/0	28/0/372/0	49/1/350/0
Class labels	128	0/0/200/0	76/7/117/0	0/0/200/0	0/0/200/0
Class labels	512	0/0/200/0	81/2/117/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	38/0/58/0	33/1/62/0	4/0/92/0	0/0/96/0
Synthetic teacher	512	40/1/55/0	33/2/61/0	10/0/86/0	2/0/94/0

Table 24: Utility variants against the first-response summary. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	248/4/148/0	328/2/70/0	142/6/252/0	251/3/146/0
Frozen teacher	512	272/3/125/0	343/1/56/0	160/8/232/0	307/1/92/0
Class labels	128	188/0/12/0	196/0/4/0	98/7/95/0	182/2/16/0
Class labels	512	188/0/12/0	196/0/4/0	105/6/89/0	184/0/16/0
Synthetic teacher	128	90/0/6/0	68/2/26/0	41/2/53/0	37/9/50/0
Synthetic teacher	512	90/1/5/0	73/0/23/0	41/3/52/0	40/8/48/0

Table 25: Utility variants against fixed midpoint with PSD repair. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	10/1/389/0	19/3/378/0	0/0/400/0	0/0/400/0
Frozen teacher	512	28/0/372/0	36/1/363/0	0/0/400/0	0/1/399/0
Class labels	128	0/0/200/0	0/0/200/0	0/0/200/0	0/0/200/0
Class labels	512	0/0/200/0	0/0/200/0	0/0/200/0	0/0/200/0
Synthetic teacher	128	36/2/58/0	33/1/62/0	16/0/80/0	19/0/77/0
Synthetic teacher	512	34/2/60/0	33/0/63/0	20/0/76/0	19/1/76/0

Table 26: Utility variants against ordinary DPP. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	381/1/18/0	393/1/6/0	196/12/192/0	360/3/37/0
Frozen teacher	512	350/3/47/0	381/1/18/0	173/13/214/0	349/2/49/0
Class labels	128	200/0/0/0	200/0/0/0	90/3/107/0	200/0/0/0
Class labels	512	200/0/0/0	200/0/0/0	86/12/102/0	199/0/1/0
Synthetic teacher	128	89/0/7/0	58/5/33/0	41/3/52/0	21/3/72/0
Synthetic teacher	512	88/0/8/0	50/3/43/0	41/5/50/0	22/10/64/0

Table 27: Utility variants against importance union. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	369/3/28/0	394/0/6/0	164/11/225/0	362/3/35/0
Frozen teacher	512	346/3/51/0	378/3/19/0	153/11/236/0	338/2/60/0
Class labels	128	200/0/0/0	200/0/0/0	89/8/103/0	197/0/3/0
Class labels	512	200/0/0/0	200/0/0/0	86/8/106/0	198/0/2/0
Synthetic teacher	128	91/1/4/0	52/5/39/0	26/4/66/0	21/5/70/0
Synthetic teacher	512	86/2/8/0	58/1/37/0	35/1/60/0	17/1/78/0

Table 28: Utility variants against importance mixed. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	361/5/34/0	394/1/5/0	204/7/189/0	360/2/38/0
Frozen teacher	512	323/2/75/0	370/1/29/0	106/6/288/0	330/2/68/0
Class labels	128	199/1/0/0	200/0/0/0	104/5/91/0	195/0/5/0
Class labels	512	197/3/0/0	200/0/0/0	45/3/152/0	182/2/16/0
Synthetic teacher	128	90/5/1/0	54/1/41/0	25/2/69/0	14/8/74/0
Synthetic teacher	512	90/4/2/0	54/4/38/0	30/4/62/0	15/7/74/0

Table 29: Utility variants against signed Nyström floor. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	40/41/319/0	365/2/33/0	0/0/400/0	7/4/389/0
Frozen teacher	512	26/36/338/0	259/3/138/0	0/0/400/0	12/4/384/0
Class labels	128	22/45/133/0	200/0/0/0	0/0/200/0	3/2/195/0
Class labels	512	24/21/155/0	200/0/0/0	0/0/200/0	10/2/188/0
Synthetic teacher	128	48/3/45/0	35/0/61/0	7/2/87/0	1/0/95/0
Synthetic teacher	512	46/7/43/0	29/1/66/0	9/0/87/0	0/1/95/0

Table 30: Utility variants against signed response-metric floor. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	71/64/265/0	385/0/15/0	0/0/400/0	38/13/349/0
Frozen teacher	512	64/50/286/0	356/1/43/0	1/0/399/0	79/20/301/0
Class labels	128	29/60/111/0	200/0/0/0	0/0/200/0	5/11/184/0
Class labels	512	40/41/119/0	200/0/0/0	0/0/200/0	19/9/172/0
Synthetic teacher	128	55/6/35/0	35/0/61/0	12/0/84/0	0/2/94/0
Synthetic teacher	512	58/6/32/0	32/1/63/0	16/0/80/0	3/0/93/0

Table 31: Utility variants against response-only metric floor. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	395/0/5/0	396/0/4/0	375/1/24/0	377/0/23/0
Frozen teacher	512	385/0/15/0	381/1/18/0	357/0/43/0	358/1/41/0
Class labels	128	200/0/0/0	200/0/0/0	200/0/0/0	200/0/0/0
Class labels	512	200/0/0/0	200/0/0/0	195/1/4/0	199/1/0/0
Synthetic teacher	128	57/4/35/0	34/2/60/0	12/0/84/0	7/1/88/0
Synthetic teacher	512	56/3/37/0	33/2/61/0	19/0/77/0	7/2/87/0

Table 32: Utility variants against response-only metric midpoint. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility floor	Utility midpoint	Inflated utility floor	Inflated utility midpoint
Frozen teacher	128	399/0/1/0	400/0/0/0	391/1/8/0	389/0/11/0
Frozen teacher	512	393/0/7/0	392/1/7/0	376/1/23/0	380/1/19/0
Class labels	128	200/0/0/0	200/0/0/0	200/0/0/0	200/0/0/0
Class labels	512	200/0/0/0	200/0/0/0	199/0/1/0	200/0/0/0
Synthetic teacher	128	70/1/25/0	64/2/30/0	44/1/51/0	30/1/65/0
Synthetic teacher	512	73/0/23/0	65/4/27/0	50/0/46/0	31/3/62/0

Table 33: Utility variants against ordinary utility floor. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Utility midpoint	Inflated utility floor
Frozen teacher	128	390/1/9/0	0/0/400/0
Frozen teacher	512	365/2/33/0	3/0/397/0
Class labels	128	200/0/0/0	0/0/200/0
Class labels	512	200/0/0/0	0/0/200/0
Synthetic teacher	128	31/2/63/0	1/0/95/0
Synthetic teacher	512	31/3/62/0	4/0/92/0

Table 34: Utility variants against ordinary utility midpoint. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Inflated utility midpoint
Frozen teacher	128	0/0/400/0
Frozen teacher	512	1/0/399/0
Class labels	128	0/0/200/0
Class labels	512	0/0/200/0
Synthetic teacher	128	7/5/84/0
Synthetic teacher	512	6/5/85/0

Table 35: Utility variants against inflated utility floor. Cells are wins/ties/losses/unavailable at the same prescribed byte cap. All scopes and both sample budgets are retained.

Scope	s	Inflated utility midpoint
Frozen teacher	128	382/2/16/0
Frozen teacher	512	360/1/39/0
Class labels	128	200/0/0/0
Class labels	512	200/0/0/0
Synthetic teacher	128	43/3/50/0
Synthetic teacher	512	39/5/52/0