WHEN DO WORST CASES ALIGN? EXACT COVARIANCE CONTACT IN VOLUME SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

We give the exact contact geometry of the universal covariance ceiling in volume-sampled least squares. Consider ordinary fixed-size volume sampling for unweighted least squares. Only the subset draw is random, while the design and response are fixed. We assume positive residual loss and no rank-critical row. The full contact space is the kernel of an explicit positive semidefinite operator. This space is unchanged across all strict-interior subset sizes. We show that varying the residual yields a finite orthogonal family of nonzero contact spaces. This family gives exact criteria for whole-query contact and the existence of a contact direction. Beyond exact contact, we quantify departure from the covariance ceiling. A leave-one-out quantity gives two-sided bounds on the normalized gap across all such subset sizes. These bounds also apply to designs with no contact direction. We also use this geometry to build certificates for expected query-MSE excess over the full-data fit. On fixed regression workloads, added information reduces sufficient sample sizes on fixed grids. Lower bounds can also certify that selected sizes exceed the risk allowance.

1 INTRODUCTION

Replacing the least-squares fit on all observations (Full) by a random subset fit can change predictions on fixed inputs, the *query*. Under ordinary fixed-size volume sampling with unweighted least squares, the fit is unbiased. Its expected squared-loss excess over Full is a query-weighted covariance trace. A universal covariance ceiling can be sharp over a class of designs yet remain unattainable on a fixed design. Attainment depends on the residual, which is Full's source-response error. Can one residual attain the ceiling in every query direction? Figure 1 summarizes the geometry. Directional contacts need not combine. A fixed residual determines the directions in which the covariance reaches the ceiling, and a whole query attains the ceiling only when its range is contained in that space.

For a fixed real design with no rank-critical row, the possible nonzero contact spaces form a finite, mutually orthogonal family. Each residual can realize at most one such space. With positive residual loss, its contact space is unchanged across all strict-interior sizes $d < s < m$. Exact contact is only part of the picture. A fixed design may have no nonzero contact direction. We therefore compare the covariance deficit at every strict-interior size with its leave-one-out value. The resulting two-sided bounds do not require an existing contact class.

Our contributions are:

(i) **Exact contact geometry.** An explicit kernel gives the full contact space of each residual. Whole-query contact is equivalent to containment in one complement-class space. Two-row deletion decides whether spectral contact is possible (Results 3–5).

(ii) **Quantitative geometry beyond contact.** We compare the covariance deficit at every strict-interior subset size with its leave-one-out deficit. The two-sided bounds also apply when the fixed design has no nonzero contact direction.

(iii) **Information-dependent query-risk certificates.** We derive certificates that become stronger as more information becomes available. On fixed regression workloads, tighter upper bounds can reduce sufficient sample sizes on fixed grids, while lower bounds can rule out selected sizes.

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

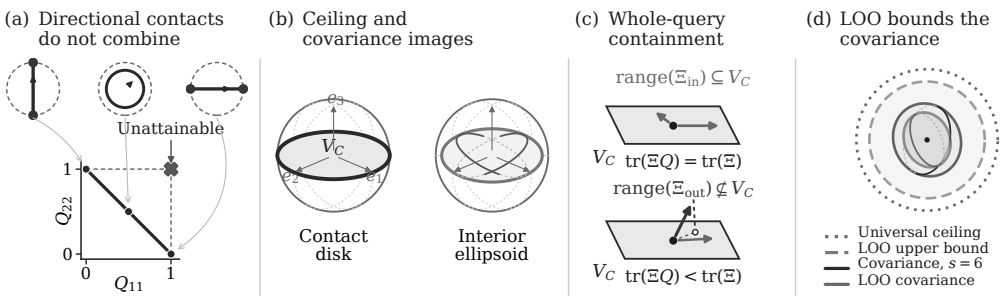


Figure 1: **Contact and covariance-body geometry.** Write $Q_s = \overline{M}_s/(\alpha_s L^*)$; a PSD matrix T is shown by $\{T^{1/2}u : \|u\|_2 \le 1\}$, not a confidence region. (a) On one design, different residuals contact either coordinate but not both. (b) On another design, a contact disk on V_C contrasts with an interior ellipsoid. (c) With that design and contact residual fixed, whole-query contact holds exactly when $\operatorname{range}\Xi \subseteq V_C$. (d) For a separate noncontact 8×3 design at $s = 6$, the projected three-dimensional bodies show $Q_7 \prec Q_6 \prec I_3 - \frac{3}{8}(I_3 - Q_7) \prec I_3$. Body widths are not deficit values; constructions are in Appendix P.

Relation to prior work. Ordinary-volume sampling gives unbiasedness and inverse-Gram moments (Dereziński & Warmuth, 2017; 2018b), while complement-frame theory supplies the underlying deletion geometry (Bodmann & Paulsen, 2005; Casazza et al., 2013). These results do not characterize fixed-response ceiling contact or relate covariance deficits across subset sizes. We show that the full contact kernel is invariant across all strict-interior sizes, classify the possible nonzero contact spaces across residuals, and compare every strict-interior deficit with its leave-one-out value. Appendix U gives detailed source relations.

2 Setting and the universal ceiling

Let $X \in \mathbb{R}^{m\times d}$ have full column rank and let $y \in \mathbb{R}^m$ be a deterministic response. The full-pool OLS fit, residual, and loss are

$$G = X^\top X, \qquad w^* = G^{-1}X^\top y, \qquad e = y - Xw^*, \qquad L^* = \|e\|_2^2. \tag{1}$$

For an indexed subset $S \subseteq [m]$ of size s, write $G_S = X_S^\top X_S$ and $D_S = \det(G_S)$. For $m > d$ and $d \le s \le m$, ordinary unrescaled indexed fixed-size volume sampling is

$$\mathbb{P}_X(S) = \frac{D_S}{Z_{X,s}}, \qquad Z_{X,s} = \sum_{|U|=s} D_U = \binom{m-d}{s-d}\det(G). \tag{2}$$

Only positive-volume subsets have positive probability. On such a subset we use selected unweighted OLS,

$$w_S = G_S^{-1}X_S^\top y_S, \qquad L_S = \|X_S w_S - y_S\|_2^2. \tag{3}$$

The subset draw is the only randomness. The centered coefficient covariance and its full-Gram-whitened form are

$$M_s = \mathbb{E}_X[(w_S - w^*)(w_S - w^*)^\top], \qquad \overline{M}_s = G^{1/2}M_s G^{1/2}. \tag{4}$$

For fixed query features $Z \in \mathbb{R}^{q\times d}$ with $q \ge 1$, define the whitened query metric $\Xi = G^{-1/2}Z^\top Z G^{-1/2}$. Selected-OLS unbiasedness then gives, for every fixed query target y_Q,

$$\mathbb{E}_X\|Zw_S - y_Q\|_2^2 - \|Zw^* - y_Q\|_2^2 = \operatorname{tr}(\Xi\overline{M}_s). \tag{5}$$

For one output, dividing (5) by q gives the expected query-MSE excess. The fixed query metric selects the covariance directions relevant to prediction.

Put $\alpha_s = (m-s)/(m-d)$. If $m = d$, $s = m$, or $L^* = 0$, covariance is zero; at $s = d < m$, supported square systems interpolate and the rank-$(d+1)$ augmentation is not invoked. Appendix A gives all support conventions.

For the residual-dependent analysis below, set $A = XG^{-1/2}$. Write $a_i^\top$ for row i of A and $\ell_i = \|a_i\|_2^2$ for its leverage. In the positive-loss branch, set $z = e/\sqrt{L^*}$, so $A^\top z = 0$ and $\|z\|_2 = 1$.

2.1 UNIVERSAL CLASS-SHARP CEILING

For symmetric matrices, $C \preceq D$ means that $D - C$ is positive semidefinite. The following common ceiling poses the fixed-design question.

Theorem 1 (Universal covariance ceiling)**.** *For every full-column-rank X, fixed y, $m > d$, and legal $d \le s \le m$, sampling by* (2) *followed by selected OLS* (3) *satisfies*

$$\mathbb{E}_X w_S = w^*, \qquad M_s \preceq \alpha_s L^* G^{-1}, \qquad \overline{M}_s \preceq \alpha_s L^* I_d. \tag{6}$$

The coefficient is sharp over the design class at every legal size; attainment on a fixed design is a separate question.

The proof, support conventions, attaining construction, general-position strictness and interior supremum are in Appendix A; its ordinary-law ingredients are standard (Dereziński & Warmuth, 2017; 2018b; Avron & Boutsidis, 2013; Li et al., 2017).

2.2 CAN ONE RESIDUAL MAXIMIZE A WHOLE QUERY?

In the duplicated-basis design, different residuals attain the ceiling in each coordinate direction, but no residual attains it for the whole isotropic query $\Xi = I_2$. Figure 1(a) shows the separation; Appendix P gives the exact calculation.

3 RESIDUAL AUGMENTATION

For $L^* > 0$, appending the compatible residual gives an orthonormal frame $B = [A\ z]$. Its row norms define the common geometry operator

$$h_i = \ell_i + z_i^2 \le 1, \qquad R_A(z) = \sum_{i=1}^{m} (1 - h_i) a_i a_i^\top \succeq 0. \tag{7}$$

For $d < s < m$, put $\beta_s = (s-d)/(m-d) = 1 - \alpha_s$ and $\gamma_s = (m-s)/(m-d-1)$. Let P_B be the auxiliary size-s volume law on $\det(B_S^\top B_S) > 0$; selected OLS is still drawn from P_X. Here *row general position* means that every indexed set of d rows of X is nonsingular.

Theorem 2 (Residual-augmented contraction)**.** *Suppose X has full column rank, $L^* > 0$, and $d < s < m$. Then*

$$\overline{M}_s \preceq L^* \left[I_d - \beta_s (I_d - \gamma_s R_A(z))^{-1} \right] \preceq \alpha_s L^* I_d - \beta_s \gamma_s L^* R_A(z). \tag{8}$$

With $K_S = A_S^\top A_S$, the supported-set measure change and first moment also hold for every such X:

$$P_X(S) L_S = \beta_s L^* P_B(S) \quad (P_X(S) > 0), \qquad \mathbb{E}_B K_S = I_d - \gamma_s R_A(z). \tag{9}$$

Under the additional condition of row general position, the covariance has the exact inverse-Gram representation

$$\frac{\overline{M}_s}{L^*} = I_d - \beta_s \mathbb{E}_B[K_S^{-1}]. \tag{10}$$

Appendix B.6 gives the raw-coordinate forms and support-aware boundary proof.

The linear bound gives one contact-kernel inclusion; supported omissions give the other, including at support-changing boundaries. The augmented moments also yield the upper certificates in Section 6.

4 EXACT CONTACT GEOMETRY

Let $\mathcal{Z}_A = \{z \in \ker(A^\top) : \|z\|_2 = 1\}$. For $L^* > 0$ and any $\theta \in \mathbb{R}^d$, write $\overline{M}_s^{A,z}$ for the covariance of $(A, A\theta + \sqrt{L^*} z)$ under the same law and OLS fit; it is independent of θ.

Theorem 3 (Fixed-residual contact kernel)**.** *Let $d \ge 1$, $m \ge d + 2$, and $d < s < m$. Suppose that $A \in \mathbb{R}^{m \times d}$ has orthonormal columns and no coloop: deleting any one row preserves column rank, or equivalently $\ell_i < 1$ for every row. For every fixed compatible residual $z \in \mathcal{Z}_A$,*

$$\boxed{\ker\left(\alpha_s I_d - \overline{M}_s^{A,z}/L^*\right) = \ker R_A(z).} \tag{11}$$

Here $R_A(z)$ is the residual operator in (7)*. The equality holds at every strict-interior size for this same fixed residual.*

Thus the covariance changes with subset size, but its contact space does not.

Why the kernel persists across sizes. For $v \in \ker R_A(z)$, every active row ($a_i^\top v \neq 0$) has $h_i = \ell_i + z_i^2 = 1$. A supported subset can omit at most one such row. Summing the omission contributions gives $v^\top \overline{M}_s^{A,z} v / L^* = \alpha_s \|v\|_2^2$ at every interior size; Theorem 2 gives the reverse inclusion. Appendices C and E give the support proof and query consequences.

4.1 Whole-query attainment

For a fixed query metric $\Xi \succeq 0$, whole-query equality asks whether one residual contacts every direction in $\operatorname{range}\Xi$. A residual selects one direction in the orthogonal complement of the feature columns. Let N have orthonormal columns spanning $\ker A^\top$, write its rows as $n_i^\top$, normalize $\widehat{n}_i = n_i/\sqrt{1-\ell_i}$, and partition the normalized rows into maximal parallel classes C (allowing opposite signs). Let $\widehat{n}_C$ be a unit representative of class C and set $V_C = \ker A_{-C}$.

Corollary 4 (Whole-query contact classification)**.** *Under the domain of Theorem 3, let $\Xi \succeq 0$ have $\operatorname{tr}\Xi > 0$. Whole-query contact is characterized by*

$$\boxed{\begin{array}{rl} & \exists z \in \mathcal{Z}_A : \quad \operatorname{tr}(\Xi \overline{M}_s^{A,z}) = \alpha_s L^* \operatorname{tr}\Xi \\ \iff & \operatorname{range}\Xi \subseteq V_C \quad \textit{for one class } C \textit{ with } |C| \geq 2. \end{array}} \tag{12}$$

Here $\dim V_C = |C| - 1$, and $z_* = \pm N\widehat{n}_C$ attains all queried directions simultaneously at every interior size.

Writing $z = N\xi$ with $\|\xi\| = 1$, saturation requires $\widehat{n}_i = \pm\xi$, so the saturated rows form at most one maximal class. Its witness gives

$$\ker R_A(N\widehat{n}_C) = V_C = \ker A_{-C}.$$

A singleton gives no nonzero contact. The nonzero spaces form a finite orthogonal family: their images under $v \mapsto Av$ have disjoint supports.

A PSD trace deficit vanishes exactly when it annihilates the query range. Consequently, a rank-r query needs one class with at least $r+1$ rows *and* $\operatorname{range}\Xi \subseteq V_C$; query rank alone does not decide contact. Full-rank query contact occurs precisely for a no-coloop $d+1$-row core plus zero padding, using its core dependence as residual. This recovers the class-sharp construction of Theorem 1. If every feature row is nonzero, a full-rank query is strictly below the universal trace ceiling.

4.2 Spectral attainment

Theorem 5 (Two-row deletion criterion)**.** *On the domain of Theorem 3,*

$$\exists z \in \mathcal{Z}_A : \ker R_A(z) \neq \{0\} \quad \iff \quad \exists i \neq j : \operatorname{rank}(A_{-\{i,j\}}) < d. \tag{13}$$

The kernel identity makes this spectral ceiling attainment at every interior size. The complement-frame rank argument and witnesses are in Appendix C.

Structural possibility versus observed attainment. On the same domain, for $\Xi \succeq 0$ with $\operatorname{tr}\Xi > 0$, define the single-output normalized ceiling deficit

$$g_s^{(1)}(z, \Xi) := 1 - \frac{\operatorname{tr}(\Xi \overline{M}_s^{A,z})}{\alpha_s L^* \operatorname{tr}\Xi}. \tag{14}$$

This is the unused fraction of the ceiling, not prediction error. Table 1 contrasts structural contact with observed attainment and shows why query range matters beyond rank and spectrum; construction details are in Appendix Q.1.

Table 1: **Contact depends on the fixed residual and query range.** "In fixed kernel?" tests $\operatorname{range}\Xi \subseteq \ker R_A(z)$. Witness responses are engineered; IEEE uses a topology-derived design. The exact deficits hold at $s = 21, 30, 39$ (Pontius), $16, 17, 18$ (IEEE), and $4, 5$ (matched). Matched queries have the same rank and spectrum.

Design / residual	Kernel dim.	In fixed kernel?	Deficit g
Pontius one-hot / observed	0	No	$17354/18443$
Same one-hot / class witness	1	Yes	0
IEEE-14-derived / class witness	2	Yes	0
Matched Ξ_{in} / class witness	2	Yes	0
Matched Ξ_{out} / same witness	2	No	$1/2$

5 Quantitative contact geometry

The classification in Section 4 tells us where the covariance ceiling is reached. We now bound the gap from that ceiling. The bound also applies when the fixed design has no nonzero contact direction.

5.1 All-size comparison with leave-one-out

Retain the no-coloop design and positive-loss assumptions of Theorem 3, and fix $z \in \mathcal{Z}_A$. With $c_i = 1 - \ell_i > 0$, define the leave-one-out (LOO) deficit operator

$$D_{\text{LOO}}(z) := \sum_{i=1}^{m} \left(1 - \frac{z_i^2}{c_i}\right) a_i a_i^\top. \tag{15}$$

The complement projection gives $0 \le z_i^2 \le c_i$, so this operator is positive semidefinite. For $\Xi \succeq 0$ with $\operatorname{tr}\Xi > 0$, put

$$\delta_{\text{LOO}}(z, \Xi) := \frac{\operatorname{tr}(\Xi D_{\text{LOO}}(z))}{\operatorname{tr}\Xi}. \tag{16}$$

The normalized ceiling deficit $g_s^{(1)}$ is defined in (14); it is distinct from Full-relative query MSE.

Theorem 6 (All-size covariance-deficit comparison)**.** *Under the assumptions above, for every integer $d < s < m$, let $b_s = \dfrac{s-d}{(m-s)(m-d-1)} > 0$. Then*

$$\boxed{b_s D_{\text{LOO}}(z) \preceq I_d - \frac{\overline{M}_s^{A,z}}{\alpha_s L^*} \preceq D_{\text{LOO}}(z).} \tag{17}$$

For every nonzero $\Xi \succeq 0$, this gives

$$b_s \delta_{\text{LOO}}(z, \Xi) \le g_s^{(1)}(z, \Xi) \le \delta_{\text{LOO}}(z, \Xi) = g_{m-1}^{(1)}(z, \Xi). \tag{18}$$

At $s = m - 1$, $b_s = 1$ and both matrix bounds are equalities.

Omission moments and a positive-semidefinite prediction-error remainder give the upper deficit bound. Sampling one omitted row before the remaining subset, then applying the universal ceiling, gives the lower bound. The predictor is exact at LOO. Appendix F proves both bounds with zero-volume subsets accounted for; neither uses the class classification.

Since $b_s, c_i > 0$, the covariance deficit, D_{LOO}, and $R_A(z)$ have the same kernel. Thus the comparison recovers the fixed-residual contact space; Section 4 classifies the spaces across residuals. Away from LOO, b_s can be small, and this comparison need not improve the resolvent bound. A small deficit means risk near the ceiling, not low risk.

To read the bound geometrically, let N have orthonormal columns spanning $\ker A^\top$, write $z = N\xi$ with $\|\xi\|_2 = 1$, and let $\widehat{n}_i$ be its unit complement row. With $u_i = a_i^\top \Xi a_i / \operatorname{tr}\Xi$, we have $u_i \ge 0$, $\sum_i u_i = 1$, and

$$\delta_{\text{LOO}}(z, \Xi) = \sum_i u_i \big(1 - |\widehat{n}_i^\top \xi|^2\big). \tag{19}$$

The query weights each row's angular error. Zero error gives whole-query contact; small error keeps the query-weighted covariance trace near its ceiling at every interior size. The converse loses a factor $1/b_s$, which can be large. Figure 1(d) shows the matrix nesting and Figure 2 gives finite noncontact and all-size examples. Angular-cap and deletion-subspace consequences are in Appendix F.

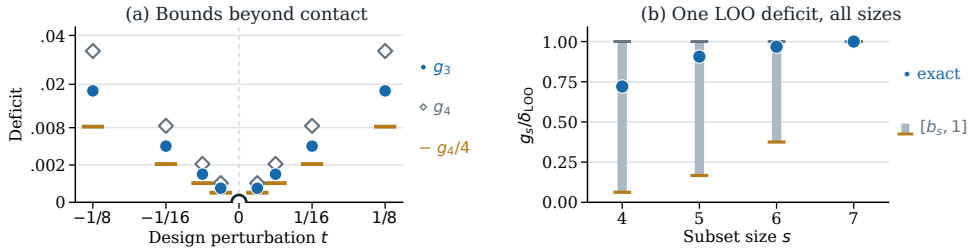


Figure 2: **Exact finite leave-one-out deficit comparisons.** Here g_s is the normalized ceiling deficit and $\delta_{\mathrm{LOO}} = g_{m-1}$. (a) With raw y, Z fixed, the 5×2 design X_t varies. Every tested $t \neq 0$ has no nonzero contact direction, yet $\delta_{\mathrm{LOO}}/4 \leq g_3 \leq \delta_{\mathrm{LOO}} = g_4$. Only tested points are shown; the vertical axis uses square-root coordinates with raw-deficit ticks. (b) A separate 8×3 design with $\delta_{\mathrm{LOO}} > 0$ shows $g_s/\delta_{\mathrm{LOO}}$ against the deterministic interval $[b_s, 1]$, not a sampling interval. Appendix P gives the exact constructions and two-row deletion checks.

5.2 Local joint alignment

The all-size bound needs no contact class. Near a proper class $0 < \dim V_C < d$, local bounds use the residual angle and query weight outside V_C; only the lower bound requires the local conditions in Appendix G.3. For any maximal class C, set $z_C = N\widehat{n}_C$ and $V_C = \ker A_{-C}$. LOO also has the exact law

$$\operatorname{range} \Xi \subseteq V_C \quad \Longrightarrow \quad g^{(1)}_{m-1}(z, \Xi) = 1 - (z^\top z_C)^2, \tag{20}$$

for every compatible unit residual and nonzero PSD query, without a local restriction or equal leverage (Appendix G.4).

5.3 Aggregate residual energy and orientation

For multi-output residuals $E = Y - XW_F$, rank at least two forces strict aggregate slack for every nonzero PSD query on the no-coloop, interior domain. Rank one still needs a class witness and query containment. Even equal residual spectra can give different query risks because orientation matters (Appendices G.5 and O.1).

6 Information-dependent query certificates

For fixed design and query, we first bound risk uniformly over the unknown residual, then use source responses and Full to bound absolute query-MSE excess.

6.1 Before source responses

On Theorem 3's structural domain, for nonzero $\Xi \succeq 0$ define

$$t_Q = \min_{z \in \mathcal{Z}_A} \frac{\operatorname{tr}(\Xi R_A(z))}{\operatorname{tr} \Xi}, \qquad B_{\mathrm{res}}(s; t) = 1 - \frac{\beta_s}{1 - \gamma_s t}. \tag{21}$$

The feature-only margin is $\nu_A = \min_{z \in \mathcal{Z}_A} \lambda_{\min}(R_A(z))$. Unlike ν_A, t_Q can be positive despite spectral contact (duplicated basis: $t_Q = 1/4$); it is a trace margin, not a Loewner lower bound.

Corollary 7 (Response-uniform query certificate)**.** *On this domain,* $\nu_A \leq t_Q < (m-d-1)/(m-d)$ *and*

$$\mathsf{T}^{\mathrm{RU}}_Q(A, s) := \max_{z \in \mathcal{Z}_A} \frac{\operatorname{tr}(\Xi \overline{M}^{A,z}_s)}{L^* \operatorname{tr} \Xi} \leq B_{\mathrm{res}}(s; t_Q), \qquad t_Q = 0 \Longleftrightarrow \mathsf{T}^{\mathrm{RU}}_Q(A, s) = \alpha_s. \tag{22}$$

The contact kernel gives the zero phase; weighted-trace Jensen gives the bound. Unlike $g_s^{(1)}$, $\mathsf{T}_Q^{\mathrm{RU}}$ is not divided by α_s. For $u_i = a_i^\top \Xi a_i / \operatorname{tr}\Xi$, the cheap endpoint

$$\underline{t}_Q^{\mathrm{cheap}} = \max\{0, 1 - \sum_i \ell_i u_i - \max_i u_i\} \le t_Q \tag{23}$$

avoids a complement eigensolve. The Feature comparator uses $t_{\mathrm{lev}} = \max\{0, 1 - 2\max_i \ell_i\} \le \nu_A$; Query uses $B_{\mathrm{res}}(s; \max\{t_{\mathrm{lev}}, \underline{t}_Q^{\mathrm{cheap}}\})$. Exact compression and further certificates are in Appendices E.4 and H.

The all-size comparison gives a second response-uniform statement. On the no-coloop domain of Theorem 6, let $\tau_Q = \min_{z\in\mathcal{Z}_A} \delta_{\mathrm{LOO}}(z, \Xi)$. Then

$$b_s \tau_Q \le \min_{z\in\mathcal{Z}_A} g_s^{(1)}(z, \Xi) = 1 - \mathsf{T}_Q^{\mathrm{RU}}(A, s)/\alpha_s \le \tau_Q. \tag{24}$$

This follows by minimizing the fixed-residual comparison. Unlike t_Q, τ_Q is the smallest LOO deficit and determines the worst-case normalized LOO risk. Its eigenvalue formula and a corresponding residual are in Appendix F; that residual need not maximize risk at each size.

6.2 Given source responses and Full

For fixed $Y \in \mathbb{R}^{m\times K}$ and query $Z \in \mathbb{R}^{q\times d}$, let W_F be Full, $E = Y - XW_F = [e_1, \dots, e_K]$, $L_j = \|e_j\|_2^2$, and $L_{\mathrm{src}} = \sum_j L_j = \|E\|_F^2$. Write $\overline{M}_{s,j}$ for output j's full-Gram-whitened coefficient covariance under the ordinary volume law. The absolute expected query-MSE excess and its source-query scale are

$$\mathcal{D}_s = \frac{\mathbb{E}_X\|Z(W_S - W_F)\|_F^2}{qK} = \frac{1}{qK}\sum_j \operatorname{tr}(\Xi\overline{M}_{s,j}), \qquad \Gamma_Q = \frac{L_{\mathrm{src}}\operatorname{tr}\Xi}{qK}. \tag{25}$$

By unbiasedness, $\mathcal{D}_s$ is expected query-MSE excess over Full. If $L_{\mathrm{src}} = 0$, every supported subset fit equals Full, so $\mathcal{D}_s = 0$ and no residual normalization is needed.

For $L_{\mathrm{src}} > 0$ and $d < s < m$, the upper bounds below hold for every full-column-rank X. Let $J_+ = \{j : L_j > 0\}$, $z_j = e_j/\sqrt{L_j}$, $R_j = R_A(z_j)$, and $\bar{R} = \sum_{j\in J_+}(L_j/L_{\mathrm{src}})R_j$. Under the analysis-only augmented law of $B_j = [A, z_j]$, set $K_S = A_S^\top A_S$ and

$$H_{s,j} = \mathbb{E}_{B_j} K_S = I - \gamma_s R_j, \qquad V_{s,j} = \mathbb{E}_{B_j}(K_S - H_{s,j})^2. \tag{26}$$

The square is a matrix square. These are augmented-law moments; $\mathcal{D}_s$ uses the ordinary law. In absolute MSE units, define

$$\begin{aligned} U_{\mathrm{agg}}(s) &= \frac{L_{\mathrm{src}}}{qK}\operatorname{tr}[\Xi\{I - \beta_s(I - \gamma_s\bar{R})^{-1}\}], \\ U_{\mathrm{per}}(s) &= \frac{1}{qK}\sum_{j\in J_+} L_j \operatorname{tr}[\Xi\{I - \beta_s H_{s,j}^{-1}\}], \\ C_s &= \frac{\beta_s}{qK}\sum_{j\in J_+} L_j \operatorname{tr}(\Xi H_{s,j}^{-1} V_{s,j} H_{s,j}^{-1}) \ge 0, \qquad U_2(s) = U_{\mathrm{per}}(s) - C_s. \end{aligned} \tag{27}$$

U_{agg} pools residuals; U_{per} keeps output-specific moments; U_2 also uses source-Gram fluctuations. All share the same risk units.

Corollary 8 (Response-aware query certificates)**.** *Under these conditions,*

$$\boxed{\mathcal{D}_s \le U_2(s) \le U_{\mathrm{per}}(s) \le U_{\mathrm{agg}}(s).} \tag{28}$$

Why the correction is valid. For $0 \prec K \preceq I$, $H = \mathbb{E}K$, and $V = \mathbb{E}(K - H)^2$,

$$\mathbb{E}K^{-1} - H^{-1} \succeq H^{-1}VH^{-1}. \tag{29}$$

The remainder is PSD, so this ordered bound is not a truncated Taylor approximation (Appendix B.5). On the common contact domain with $\operatorname{tr}\Xi > 0$, $U_{\mathrm{agg}}/\Gamma_Q \le B_{\mathrm{res}}(s; t_Q)$; observed residuals can improve the response-uniform guarantee.

Shared source moments avoid subset enumeration. Our dense construction has d^4 storage and d^5 work terms. Computation and conditional relative-loss extensions are in Appendices J.3, J.4, and S.

6.3 Omission lower certificates

For workload g with query $Z_g \in \mathbb{R}^{q_g \times d}$, let $\mathcal{D}_{s,g}$ be (25) with $Z = Z_g$ and $q = q_g$. For $d < s < m$ and no coloops, multiplying the upper-deficit bound by each output's absolute-risk ceiling and summing gives

$$\mathcal{D}_{s,g} \geq (m - s)\mathcal{D}_{m-1,g}. \tag{30}$$

Table 2 uses the no-weaker scalar refinement in Appendix K. The upper hierarchy holds for arbitrary full-rank designs; this lower bound requires no coloops.

For issued endpoints $L^-_{s,g} \leq \mathcal{D}_{s,g} \leq U^+_{s,g}$ and an absolute allowance ϵ_g, $L^-_{s,g} > \epsilon_g$ proves violation and $U^+_{s,g} \leq \epsilon_g$ proves sufficiency for this sampler's expected Full-query-MSE excess. Otherwise the size is unresolved; an upper endpoint above the allowance does not prove failure. No statement concerns individual subset losses.

7 Certified query-risk and sample-size decisions

We fix the ordinary volume sampler and unweighted OLS. Upper endpoints certify sufficient grid sizes; lower endpoints rule out selected sizes. Appendix Q fixes models, access, grids, and numerical procedures.

Decision objects. Here r indexes certificate policies and g query workloads. $U^+_{r,g}(s)$ and $F^+_{r,g}(s)$ are issued upper endpoints for the absolute risk (25) and the response-uniform target (22), respectively. Both are zero at Full. On a fixed grid $\mathcal{S}$ containing Full, choose the smallest $s \in \mathcal{S}$ with $U^+_{r,g}(s) \leq \epsilon_g$ for response-aware risk, or $\max_g F^+_{r,g}(s) \leq \tau$ for response-uniform risk. These are first certified grid sizes, not risk optima. The historical policy B uses the best valid Feature, Query, Energy, pooled, or per-output upper endpoint; B+2 also uses U^+_2.

Query information before responses. Bike uses 2011 source features ($m = 8{,}645, d = 20$) and eight 2012 query workloads, without response labels. At $\tau = 0.25$, the cheap Query margin cuts the common first-certified size from 433 to 135; at 172 rows it certifies all eight workloads and Feature none (Figure 3(a)). This uses the normalized response-uniform criterion, without solving for exact t_Q.

Second moments for an adequate Full model. Source-only validation selects a Tetouan model with quadratic weather and calendar features ($m = 39{,}312, q = 13{,}104, d = 19, K = 3$); query labels enter only final scoring. Full aggregate standardized MSE is 0.497537, versus 1.271863 for the source-mean predictor, with every output improved. At an allowance of 1% of mean source population variance, not Full query MSE, B+2 lowers the first sufficient size from $14{,}742$ to $11{,}757$ (Figure 3(b)). The LOO diagnostic uses a separate weather-only model.

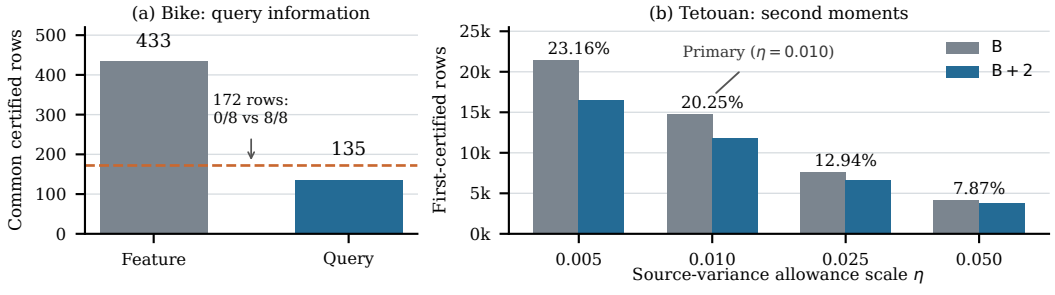


Figure 3: **Information changes first-certified sizes.** (a) Bike: eight workloads, features only, and a response-uniform threshold $\tau = 0.25$. (b) Tetouan: B$+2$ adds second moments to the B upper-bound policy; the primary allowance is $\eta = 0.01$ times mean source population variance. Bars are first sufficient fixed-grid sizes, not risk optima.

Selected matched lower and upper decisions. For each of 13 workloads, four sizes were chosen before rigorous evaluation: the largest float-screened scalar crossing, its next grid size, and the historical B+2 and B choices. Table 2 shows eight of 52 cells. The scalar lower certificate in Section 6.3 rules out 13 sizes; 26 policy choices are sufficient and 13 successors remain unresolved. The 12 Gas blocks share one source. Lower bounds do not reduce sizes already certified sufficient.

Decisions use directed endpoints for exact dyadic frozen binary64 arrays and rational allowances; parsing and preprocessing are outside this guarantee (Appendix R).

Table 2: **Selected expected Full-query-MSE excess bounds.** Lower: scalar omission; Upper: best valid candidate. The allowance is 1% of mean source population variance (Gas ≈ 0.635316; Tetouan ≈ 0.01). Values are rounded; decisions use exact rational endpoints. Selected sizes need not locate first passage.

Workload	s	Lower	Upper	Decision
Gas 2015 block 1	13,537	0.64228	4.4404	Violated
Gas 2015 block 1	13,648	0.6289	4.3007	Unresolved
Gas 2015 block 1	18,752	0.18442	0.62368	Sufficient
Gas 2015 block 4	5,104	0.64585	9.9313	Violated
Gas 2015 block 4	15,152	0.090221	0.62718	Sufficient
Tetouan	580	0.010076	0.33783	Violated
Tetouan	590	0.0099149	0.33266	Unresolved
Tetouan	11,757	0.00037991	0.0099221	Sufficient

Numerical calibration and scope. Gas and Tetouan have maximum leverage below $1/2$, so every two-row deletion retains rank and Theorem 5 rules out nonzero contact for any compatible residual and nonzero PSD query. This need not give tight risk estimates: b_s can be small far from LOO.

At Tetouan's selected size $s = 11{,}757$, a separate 256-path Monte Carlo mean displacement was about $0.0392\epsilon_g$ (Appendix R). This finite-path mean and Float64 LOO are diagnostics, not certificates; see also Appendix Q.3 for LOO precision.

Cost. Second moments add computation; no training speedup is claimed (Appendix Q.6). Certificates bound expected risk, not individual paths or the smallest size meeting the target.

Conclusion. For fixed designs without rank-critical rows, each residual has an exact covariance contact space that stays the same at every strict-interior subset size. Across residuals, the nonzero spaces form a finite orthogonal family that determines whole-query contact. The leave-one-out deficit gives two-sided Loewner bounds at every interior size, even when no nonzero contact direction exists. Response-uniform and response-aware upper certificates establish sufficient sample sizes; omission-based lower bounds rule out selected sizes whose expected query-MSE excess exceeds the allowance. These certificates concern expected query error, not the loss of every sampled fit.

AI USE STATEMENT

Generative AI was used to support proof development, verification, and manuscript preparation. The core proof was developed from the authors' study of prior work on volume sampling by Dereziński & Warmuth (2017; 2018b) and adapted to the fixed-design covariance contact setting. The main claims and proofs were formulated by the authors. AI helped explore intermediate proof steps and alternative arguments, and checked assumptions and boundary cases. The authors verified every theorem and proof in the final manuscript.

Authors designed experiments to evaluate how the theoretical upper bounds tighten when query information or second-moment information is added. AI was used to provide feedback on comparisons, controls, and experimental design. The authors made the final decisions on the experimental setup and the interpretation of the results.

AI was also used for manuscript editing to improve wording, clarity, and readability. The paper organization and emphasis of the contributions were determined by the authors. AI supported cross-checking related work, while the authors verified all relevant claims and citations against the original sources. AI assisted with experimental and analysis code. The authors checked the code for numerical consistency. The authors reviewed all AI-assisted material and take responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

Complete proofs and assumptions are given in the appendix. The experimental protocols specify preprocessing, information access, sampling and estimation; the numerical appendix describes the conditions for issuing a certificate. The reproducibility supplement's `README.md` and `geometry/reproducibility/RESULT_MAP.md` link the exact figures, selected-budget tables, and public-data risk experiment to their inputs and commands. They distinguish checks of stored results from fresh calculations using the stated external data and dependencies. It also includes inputs and scripts for numerical checks underlying the geometry figure and selected-budget tables. The supplement separates exact enumeration, rigorous bounds for frozen binary64 inputs, and nominal pointwise intervals from independent sampling paths.

REFERENCES

Jorge Antezana, Pedro Massey, Mariano Ruiz, and Demetrio Stojanoff. The Schur–Horn theorem for operators and frames with prescribed norms and frame operator. *Illinois Journal of Mathematics*, 51(2):537–560, 2007. URL https://arxiv.org/abs/math/0508646.

Haim Avron and Christos Boutsidis. Faster subset selection for matrices and applications. *SIAM Journal on Matrix Analysis and Applications*, 34(4):1464–1499, 2013. doi: 10.1137/120867287. URL https://epubs.siam.org/doi/10.1137/120867287.

Peter Balazs, Jean-Pierre Antoine, and Anna Gryboś. Weighted and controlled frames. *International Journal of Wavelets, Multiresolution and Information Processing*, 8(1):109–132, 2010. doi: 10.1142/S0219691310003377. URL https://doi.org/10.1142/S0219691310003377.

Bernhard G. Bodmann and Vern I. Paulsen. Frames, graphs and erasures. *Linear Algebra and its Applications*, 404:118–146, 2005. doi: 10.1016/j.laa.2005.02.016. URL https://doi.org/10.1016/j.laa.2005.02.016.

Peter G. Casazza, Matthew Fickus, Dustin G. Mixon, Jesse Peterson, and Ihar Smalyanau. Every Hilbert space frame has a Naimark complement. *Journal of Mathematical Analysis and Applications*, 406(1):111–119, 2013. doi: 10.1016/j.jmaa.2013.04.047. URL https://doi.org/10.1016/j.jmaa.2013.04.047.

Michał Dereziński and Manfred K. Warmuth. Unbiased estimates for linear regression via volume sampling. In *Advances in Neural Information Processing Systems 30*, pp. 3084–3093, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/54e36c5ff5f6a1802925ca009f3ebb68-Paper.pdf.

Michał Dereziński and Manfred K. Warmuth. Subsampling for ridge regression via regularized volume sampling. In Amos Storkey and Fernando Perez-Cruz (eds.), *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pp. 716–725. PMLR, 2018a. URL https://proceedings.mlr.press/v84/derezinski18a.html.

Michał Dereziński and Manfred K. Warmuth. Reverse iterative volume sampling for linear regression. *Journal of Machine Learning Research*, 19(23):1–39, 2018b. URL https://www.jmlr.org/papers/volume19/17-781/17-781.pdf.

Michał Dereziński, Manfred K. Warmuth, and Daniel J. Hsu. Leveraged volume sampling for linear regression. In *Advances in Neural Information Processing Systems 31*, pp. 2505–2514, 2018. URL https://papers.neurips.cc/paper_files/paper/2018/file/2ba8698b79439589fdd2b0f7218d8b07-Paper.pdf.

Michał Dereziński, Kenneth L. Clarkson, Michael W. Mahoney, and Manfred K. Warmuth. Minimax experimental design: Bridging the gap between statistical and worst-case approaches to least squares regression. In *Proceedings of the 32nd Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 1050–1069, 2019a. URL https://proceedings.mlr.press/v99/derezinski19b/derezinski19b.pdf.

Michał Dereziński, Manfred K. Warmuth, and Daniel Hsu. Correcting the bias in least squares regression with volume-rescaled sampling. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 944–953, 2019b. URL https://proceedings.mlr.press/v89/derezinski19a/derezinski19a.pdf.

Michał Dereziński, Manfred K. Warmuth, and Daniel Hsu. Unbiased estimators for random design regression. *Journal of Machine Learning Research*, 23:1–46, 2022. URL https://www.jmlr.org/papers/volume23/19-571/19-571.pdf.

Ethan N. Epperly. Adaptive randomized pivoting and volume sampling, 2026. URL https://arxiv.org/pdf/2510.02513v2. arXiv:2510.02513v2.

Matthew C. Fickus and Dustin G. Mixon. Numerically erasure-robust frames. *Linear Algebra and its Applications*, 437(6):1394–1407, 2012. doi: 10.1016/j.laa.2012.04.034. URL https://doi.org/10.1016/j.laa.2012.04.034.

Jack FitzGerald, Christopher Hench, Charith Peris, Scott Mackie, Kay Rottmann, Ana Sanchez, Aaron Nash, Liam Urbach, Vishesh Kakarala, Richa Singh, Swetha Ranganath, Laurie Crist, Misha Britan, Wouter Leeuwis, Gokhan Tur, and Prem Natarajan. MASSIVE: A 1M-example multilingual natural language understanding dataset with 51 typologically-diverse languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 4277–4302. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.acl-long.235. URL https://aclanthology.org/2023.acl-long.235/.

Chongming Gao. Kuairand supplementary files: Video captions and categories, 2026. URL https://zenodo.org/records/18159199. Version v1, published January 6, 2026.

Chongming Gao, Shijun Li, Yuan Zhang, Jiawei Chen, Biao Li, Wenqiang Lei, Peng Jiang, and Xiangnan He. KuaiRand: An unbiased sequential recommendation dataset with randomly exposed videos. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, pp. 3953–3957. ACM, 2022. doi: 10.1145/3511808.3557624. URL https://arxiv.org/abs/2208.08696.

GroupLens Research. MovieLens 100k dataset. https://files.grouplens.org/datasets/movielens/ml-100k/README, n.d.

Frank Hansen and Gert K. Pedersen. Jensen's operator inequality. *Bulletin of the London Mathematical Society*, 35(4):553–564, 2003. doi: 10.1112/S0024609303002200. URL https://doi.org/10.1112/S0024609303002200.

Richard V. Kadison. The Pythagorean theorem: I. the finite case. *Proceedings of the National Academy of Sciences*, 99(7):4178–4184, 2002. doi: 10.1073/pnas.032677199. URL https://doi.org/10.1073/pnas.032677199.

Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning*, 5(2–3):123–286, 2012. doi: 10.1561/2200000044. URL https://www.nowpublishers.com/article/Details/MAL-044.

Chengtao Li, Stefanie Jegelka, and Suvrit Sra. Polynomial time algorithms for dual volume sampling. In *Advances in Neural Information Processing Systems 30*, pp. 5038–5047, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/18bb68e2b38e4a8ce7cf4f6b2625768c-Paper.pdf.

MATPOWER. case14: IEEE 14-bus test case. https://matpower.org/docs/ref/matpower6.0/case14.html, n.d.

National Institute of Standards and Technology. Pontius: Statistical reference datasets, linear least squares. https://www.itl.nist.gov/div898/strd/lls/data/Pontius.shtml, n.d.

Patrick O. Perry and Art B. Owen. A rotation test to verify latent structure. *Journal of Machine Learning Research*, 11:603–624, 2010. URL https://www.jmlr.org/papers/v11/perry10a.html.

Guillaume Sagnol and Luc Pronzato. Fast screening rules for optimal design via quadratic lasso reformulation. *Journal of Machine Learning Research*, 24:1–32, 2023. URL https://www.jmlr.org/papers/volume24/22-1305/22-1305.pdf.

UCI Machine Learning Repository. Bike sharing dataset. https://archive.ics.uci.edu/dataset/275, n.d.a.

UCI Machine Learning Repository. Gas turbine CO and NOx emission data set. https://archive.ics.uci.edu/dataset/551, n.d.b.

UCI Machine Learning Repository. Power consumption of Tetouan city. https://archive.ics.uci.edu/dataset/849, n.d.c.

Kai Yu, Jinbo Bi, and Volker Tresp. Active learning via transductive experimental design. In *Proceedings of the 23rd International Conference on Machine Learning*, 2006. URL https://www.dbs.ifi.lmu.de/~tresp/papers/icml2006_actv.pdf.

PROOF SUPPORT AND SUPPLEMENTARY MATERIAL

Throughout, subsets are unordered sets of indexed rows. A zero-volume subset has probability zero, so no selected inverse or fit is defined for it.

Route	Entry points and destinations	Page
Core theory and proofs	Universal envelope (A); augmentation, contact, query geometry, all-size comparison, and stability (B–G)	13
Certificates and computation	Scalar certificates (H, p. 39); fixed-label risk (I); response-aware moments (J, p. 42); omission lower bounds (K, p. 45); numerical issuance (L)	39
Exact constructions and controls	Critical and matched examples (M–N); literal alignment controls (O, p. 53); finite constructions (P, p. 54)	47
Experimental evidence	Protocols (Q); selected risk cells (R, p. 61); relative-loss application (S)	56
Reproduction and scope	Reproduction details (T) and prior-work relations (U)	64

Structural contact results use no-coloop and strict-interior assumptions; upper certificates allow full-rank designs, while omission lower bounds require no coloops. Their local hypotheses are stated with each result.

Notation, units, and information regimes. In query-risk notation, q counts query rows, K counts outputs, and $n = m-d$; the shared-moment contraction uses $n_0 = m-d-1$. Local proofs may reuse q and K. The coefficient covariance is M_s, its full-Gram-whitened form is $\overline{M}_s = G^{1/2} M_s G^{1/2}$, and $\Xi = G^{-1/2} Z^\top Z G^{-1/2}$ is the fixed-query metric. The PSD operator $R_A(z)$ is dimensionless. Expected query SSE and MSE excess for one output are $\operatorname{tr}(\Xi \overline{M}_s)$ and $\operatorname{tr}(\Xi \overline{M}_s)/q$; for K outputs, entrywise MSE excess is $\mathcal{D}_s = \operatorname{tr}(\Xi \sum_j \overline{M}_{s,j})/(qK)$.

For each output, $M_{s,j}$ and $\overline{M}_{s,j}$ are raw and whitened covariances; $H_{s,j} = \mathbb{E}_{B_j}(A_S^\top A_S)$ and $V_{s,j} = \mathbb{E}_{B_j}(A_S^\top A_S - H_{s,j})^2$ are augmented-law source-Gram moments. All four are $d \times d$. The bounds $U_2, U_{\mathrm{per}}, U_{\mathrm{agg}}$ include $1/(qK)$. The absolute scale $\Gamma_Q = L_{\mathrm{src}} \operatorname{tr}(\Xi)/(qK)$ is not the Full query MSE; the response-aware C_s is a scalar, not an empirical coefficient moment. For an allowed query-label class $\mathcal{Y}$, $L_{\min}(W_F; \mathcal{Y})$ lower-bounds Full entrywise query MSE. A Full-relative ratio uses $L_{Q,F}$ only when it is finite and positive.

Feature/query certificates use only X, Z; response-aware ones also use Y and exact Full. Relative-MSE guarantees additionally need an allowed-label premise. Response-uniform, source-variance, and label-floor criteria differ. Use the endpoint or rank-size branch before normalizing (Appendix J); if $\Xi = 0$, unnormalized query excess is zero. Exact expectations, directed endpoints, exhaustive float64 LOO sums, and finite-path summaries are different evidence types.

The single-residual ceiling deficit $g_s^{(1)}$, the LOO matrix deficit $D_{\mathrm{LOO}}(z)$, and $\delta_{\mathrm{LOO}}(z, \Xi) = \operatorname{tr}(\Xi D_{\mathrm{LOO}}(z))/\operatorname{tr}\Xi$ are dimensionless, unlike absolute risk $\mathcal{D}_{m-1}$; none is Full-relative MSE or a response-optimized target. The valid all-size factor $b_s = (s-d)/[(m-s)(m-d-1)]$ is not claimed optimal.

GROUP I. CORE THEORY AND PROOFS

A THE UNIVERSAL FIXED-POOL COVARIANCE ENVELOPE

Fix a full-column-rank design $X \in \mathbb{R}^{m \times d}$ and a response $y \in \mathbb{R}^m$. Use $G = X^\top X$, the full fit w^*, residual e, and loss $L^* = \|e\|_2^2$ from (1). For an indexed set $S \subseteq [m]$ with $|S| = s$, put $G_S = X_S^\top X_S$ and $D_S = \det(G_S)$, as in Section 2. We use ordinary unrescaled size-s volume sampling without replacement over indexed rows, followed by unweighted least squares. Repeated rows remain distinct indices. This is the fixed-size law of (Dereziński & Warmuth, 2018b; Avron & Boutsidis, 2013; Li et al., 2017) in row or transposed-column notation. On its support $D_S > 0$, the selected fit w_S and loss L_S are given by (3). Zero-volume sets receive no fit. Every expectation below holds for the displayed fixed X and y.

The endpoint branches will be used without taking a quotient. If $m = d$, then $s = m$, X is invertible, $w_S = w^* = X^{-1} y$, and the covariance and both residual losses are zero; no expression containing

$(m-d)^{-1}$ is evaluated. If $m > d$ and $s = m$, the unique set is the full pool and its centered second moment is zero. If $L^* = 0$, then $y = Xw^*$ and every supported selected fit equals w^*, so again the centered second moment is zero. If $s = d < m$, every supported square selected system interpolates; no rank-$(d+1)$ representation is used here. These conventions also cover repeated rows and zero-volume sets.

A.1 Fixed-cardinality normalizer and theorem statement

Proposition 9 (Fixed-cardinality normalizer)**.** *For $m > d$ and $d \le s \le m$,*

$$Z_{X,s} := \sum_{|S|=s} D_S = \binom{m-d}{s-d} \det(G) > 0, \qquad \mathbb{P}_X(S) = \frac{D_S}{Z_{X,s}}. \tag{31}$$

Proof. For each indexed S, Cauchy–Binet gives

$$D_S = \sum_{\substack{T \subseteq S \\ |T|=d}} \det(X_T)^2. \tag{32}$$

Each indexed d-set T has $\binom{m-d}{s-d}$ indexed size-s supersets, so Cauchy–Binet again gives

$$\sum_{|S|=s} D_S = \binom{m-d}{s-d} \sum_{|T|=d} \det(X_T)^2 = \binom{m-d}{s-d} \det(X^\top X).$$

Full column rank makes the result positive; indexed repetitions and zero-determinant terms remain in the sum. □

Lemma 10 (Exclusion marginals)**.** *Let $F \in \mathbb{R}^{m\times p}$ satisfy $F^\top F = I_p$, and use its ordinary size-s volume law, $p \le s \le m$. For indexed rows $J \subseteq [m]$,*

$$\mathbb{P}_F(S \cap J = \varnothing) = \begin{cases} 0, & m - |J| < s, \\ \dfrac{\binom{m-|J|-p}{s-p}}{\binom{m-p}{s-p}} \det(I_p - F_J^\top F_J), & m - |J| \ge s. \end{cases} \tag{33}$$

Proof. If enough rows remain, apply the fixed-cardinality Cauchy–Binet calculation above to F_{-J}, whose Gram matrix is $I_p - F_J^\top F_J$, and divide by the normalizer $\binom{m-p}{s-p}$. The calculation also holds when F_{-J} is rank deficient: all determinants then vanish. If fewer than s rows remain, the event is empty. □

Until unbiasedness has been established, define only the centered second moment

$$C_s := \mathbb{E}_X\big[(w_S - w^*)(w_S - w^*)^\top\big]. \tag{34}$$

For $m > d$, let

$$\alpha = \frac{m-s}{m-d}, \qquad \beta = \frac{s-d}{m-d} = 1 - \alpha. \tag{35}$$

Theorem 11 (Universal covariance envelope for fixed-pool volume-sampled least squares)**.** *For every full-column-rank X, every fixed y, $m > d$, and $d \le s \le m$,*

$$\mathbb{E}_X w_S = w^*. \tag{36}$$

Consequently $M_s := C_s$ is the covariance of w_S, and

$$M_s \preceq \alpha L^* G^{-1}, \qquad \overline{M}_s := G^{1/2} M_s G^{1/2} \preceq \alpha L^* I_d. \tag{37}$$

Moreover,

$$\mathbb{E}_X \|Xw_S - y\|_2^2 \le (1 + d\alpha) L^* = \left(1 + \frac{d(m-s)}{m-d}\right) L^*. \tag{38}$$

The coefficient in (37) *is attained by the family in Proposition 18. For positive-loss row-general-position designs with $d \ge 2$ it is strictly unattained when $d < s < m$, but it remains their operator-norm supremum as described in Proposition 19.*

The directional ratio is $\lambda_{\max}(\overline{M}_s)/(\alpha L^*)$ only when $\alpha L^* > 0$; it differs from the raw Euclidean eigenvalue of M_s. No normalized ratio is formed at $s = m$ or $L^* = 0$.

A.2 Rank-size basis moments and uniform padding

The following rank-size calculation includes the off-diagonal second moments. Rank-size unbiasedness and volume-sampling moments are known ingredients (Dereziński & Warmuth, 2017; 2018b); the direct derivation fixes the signs and support needed below.

Lemma 12 (Rank-size basis moments)**.** *Let $A \in \mathbb{R}^{q\times d}$ have full column rank, suppose every indexed d-row submatrix is nonsingular, and fix $b \in \mathbb{R}^q$. Define*

$$K = A^\top A, \qquad v = K^{-1}A^\top b, \qquad r = b - Av.$$

Draw a d-set T with probability $\det(A_T)^2/\det(K)$ and put $v_T = A_T^{-1}b_T$. Then

$$\mathbb{E}v_T = v, \qquad \mathbb{E}[(v_T - v)(v_T - v)^\top] = \|r\|_2^2 K^{-1}. \tag{39}$$

Proof. The normal equations give $A^\top r = 0$. Let B_j replace column j of A by r. Cramer's rule in increasing row order gives

$$\det(A_T)(v_T - v)_j = \det((B_j)_T). \tag{40}$$

The polarized Cauchy–Binet identity

$$\sum_{|T|=d} \det(C_T)\det(D_T) = \det(C^\top D) \tag{41}$$

then gives

$$\det(K)\mathbb{E}[(v_T - v)_j] = \det(A^\top B_j) = 0,$$

because column j of $A^\top B_j$ is $A^\top r = 0$.

For every $j, k \in [d]$, the same identities give

$$\det(K)\mathbb{E}[(v_T - v)_j(v_T - v)_k] = \det(B_j^\top B_k). \tag{42}$$

Put $\rho = \|r\|_2^2$. For $j = k$, expand through the replaced row and column; for $j \neq k$, row j of $B_j^\top B_k$ has sole nonzero entry ρ in column k. Both cofactor expansions give

$$\det(B_j^\top B_k) = (-1)^{j+k}\rho\det(K_{-j,-k}) = \rho\det(K)(K^{-1})_{kj}.$$

The displayed cofactor sign and symmetry of K^{-1} complete all diagonal and off-diagonal entries of (39). □

Say that X is in *row general position* when every indexed set of d rows is nonsingular. On this domain, sample S from (31) and then $T \subseteq S$ by the rank-size volume law within X_S. The joint law is

$$\mathbb{P}_X(S,T) = \frac{D_S}{Z_{X,s}}\frac{\det(X_T)^2}{D_S} = \frac{\det(X_T)^2}{\binom{m-d}{s-d}\det(G)}. \tag{43}$$

Consequently

$$\mathbb{P}_X(T) = \frac{\det(X_T)^2}{\det(G)}, \qquad \mathbb{P}_X(S \mid T) = \binom{m-d}{s-d}^{-1} \quad (T \subseteq S). \tag{44}$$

Thus $S \mid T$ is uniform over the size-s supersets of the sampled basis, as in the basis-plus-uniform-padding representation of Dereziński et al. (2019b).

Proposition 13 (Exact row-general-position covariance decomposition)**.** *If X is in row general position, then*

$$\mathbb{E}_X w_S = w^*, \qquad M_s = L^*G^{-1} - \mathbb{E}_X[L_S G_S^{-1}]. \tag{45}$$

Proof. Apply Lemma 12 conditionally with $(A, b) = (X_S, y_S)$ and unconditionally with $(A, b) = (X, y)$. The coupling gives

$$\mathbb{E}[w_T \mid S] = w_S, \qquad \mathrm{Cov}(w_T \mid S) = L_S G_S^{-1}, \tag{46}$$

$$\mathbb{E}w_T = w^*, \qquad \mathrm{Cov}(w_T) = L^*G^{-1}. \tag{47}$$

Iterated expectation gives $\mathbb{E}w_S = w^*$, and matrix total covariance gives

$$L^*G^{-1} = \mathbb{E}_X[L_S G_S^{-1}] + \mathrm{Cov}(w_S),$$

proving (45) without taking a trace. □

A.3 THE SELECTED RESIDUAL FIRST MOMENT

Lemma 14 (Selected residual mean)**.** *For every full-column-rank X, $m > d$, and $d \le s \le m$,*

$$\mathbb{E}_X L_S = \beta L^*. \tag{48}$$

This identity does not require $L^* > 0$.

Proof. Let $H = [X \;\; y] \in \mathbb{R}^{m\times(d+1)}$. On a supported set, the Schur complement gives

$$\det(H_S^\top H_S) = D_S L_S, \qquad \det(H^\top H) = \det(G) L^*. \tag{49}$$

If $D_S = 0$, then $\operatorname{rank}(X_S) < d$, so $\operatorname{rank}(H_S) \le d$ and the determinant on the left is also zero. Thus, with unsupported terms understood as zero determinant weights, the first equality may be summed over all S.

For $s \ge d+1$, fixed-cardinality Cauchy–Binet applied to the $d+1$ columns of H yields

$$\sum_{|S|=s} \det(H_S^\top H_S) = \binom{m-d-1}{s-d-1} \det(H^\top H).$$

If $\operatorname{rank}(H) \le d$, both sides are zero, so this step does not divide by L^*. Dividing only by the positive ordinary normalizer from Proposition 9 gives

$$\mathbb{E}_X L_S = \frac{\binom{m-d-1}{s-d-1}}{\binom{m-d}{s-d}} L^* = \frac{s-d}{m-d} L^*.$$

When $s = d$, every supported square system interpolates and both sides of (48) are zero. □

A.4 THE LOEWNER ENVELOPE IN ROW GENERAL POSITION

For row-general-position X, Proposition 13 and Lemma 14, together with $G_S \preceq G$ and inverse order $G_S^{-1} \succeq G^{-1}$, give

$$\begin{aligned} \alpha L^* G^{-1} - M_s &= \mathbb{E}_X[L_S G_S^{-1}] - \beta L^* G^{-1} \\ &= \mathbb{E}_X\left[L_S (G_S^{-1} - G^{-1})\right] \succeq 0. \end{aligned} \tag{50}$$

Congruence by $G^{1/2}$ proves (37) on this class.

The full-pool loss consequence is also a matrix corollary. The normal equations $X^\top e = 0$ imply, for every w,

$$\|Xw - y\|_2^2 = L^* + \|w - w^*\|_G^2. \tag{51}$$

Therefore

$$\begin{aligned} \mathbb{E}_X\|Xw_S - y\|_2^2 &= L^* + \operatorname{tr}(G M_s) \\ &= L^* + \operatorname{tr}(\overline{M}_s) \le (1 + d\alpha) L^*, \end{aligned} \tag{52}$$

which proves (38) in row general position.

A.5 ARBITRARY FULL-RANK BOUNDARY DESIGNS

Unbiasedness and the covariance inequality require different boundary arguments. In particular, the exact decomposition (45) is not asserted when some selected Gram matrices are singular.

Proposition 15 (Determinant-weighted continuation of the mean)**.** *Equation* (36) *holds for every full-column-rank X, including designs outside row general position.*

Proof. For a supported set,

$$D_S w_S = \operatorname{adj}(G_S) X_S^\top y_S. \tag{53}$$

The right side is polynomial in the entries of X and y. It vanishes when G_S is singular: indeed, with $A_S = \operatorname{adj}(G_S) X_S^\top$,

$$A_S A_S^\top = \operatorname{adj}(G_S) G_S \operatorname{adj}(G_S) = 0,$$

so $A_S = 0$. Thus singular subsets contribute zero to the determinant-weighted numerator without receiving a selected fit.

On the dense set of row-general-position matrices, Proposition 13 and Proposition 9 imply

$$\sum_{|S|=s} \operatorname{adj}(G_S) X_S^\top y_S = \binom{m-d}{s-d} \operatorname{adj}(G) X^\top y. \tag{54}$$

Both sides are polynomial, so the identity holds for every X by polynomial continuation. For full-column-rank X, divide (54) by $Z_{X,s} = \binom{m-d}{s-d} \det(G)$ and use the preceding zero-numerator fact. The result is

$$\mathbb{E}_X w_S = G^{-1} X^\top y = w^*.$$

□

Lemma 16 (Lower semicontinuity of the centered second moment)**.** *Fix y, $d \le s \le m$, and full-column-rank X. If full-column-rank $X_a \to X$, let $M_{s,a}$ and M_s denote the ordinary-law second moments centered at their respective full fits. For every fixed $u \in \mathbb{R}^d$,*

$$u^\top M_s u \le \liminf_a u^\top M_{s,a} u. \tag{55}$$

Proof. Put $\mathcal{P} = \{S : |S| = s,\ D_S > 0\}$. For $S \in \mathcal{P}$, inverse continuity and $w_a^* \to w^*$ give $D_{S,a}\{u^\top(w_{S,a} - w_a^*)\}^2 \to D_S\{u^\top(w_S - w^*)\}^2$. Summing these finitely many terms and dropping only the nonnegative terms newly supported along the sequence yields the inequality for the unnormalized moments. Division by the converging positive normalizers $Z_{X_a,s} \to Z_{X,s} > 0$ proves (55). No fit or inverse is assigned to a zero-volume boundary set. □

Proposition 17 (One-sided boundary passage for the covariance)**.** *The Loewner inequalities* (37)*, and consequently the loss bound* (38)*, hold for every full-column-rank X.*

Proof. Let $V \in \mathbb{R}^{m \times d}$ be the Vandermonde matrix $V_{ij} = i^{j-1}$. For each d-set T, $\det(X_T + tV_T)$ is a nonzero polynomial in t, since its leading coefficient is $\det(V_T) \neq 0$. Choose a sequence $t_a \downarrow 0$ avoiding the finitely many roots of all these polynomials and set $X^{(a)} = X + t_a V$. Then every $X^{(a)}$ is in row general position and $X^{(a)} \to X$.

Keep y fixed. The full fit, residual loss, and Gram inverse converge, so Lemma 16 and the row-general-position envelope give

$$u^\top M_s u \le \liminf_a u^\top M_{s,a} u \le \lim_a \alpha L_a^* u^\top G_a^{-1} u = \alpha L^* u^\top G^{-1} u.$$

Because this holds for every fixed u, it is the first inequality in (37); congruence gives the second. Finally, Proposition 15 makes M_s a covariance, and (51)–(52) give the loss corollary. Lemma 16 handles the support change; the exact row-general-position identity (45) is not continued to the boundary. □

A.6 Exact matrix attainment, strict interior slack, and supremum

Proposition 18 (Core-plus-zero exact matrix attainment)**.** *Fix $m > d \ge 1$, let $q = m - d - 1$, and define*

$$X_0 = \begin{bmatrix} I_d \\ \mathbf{1}_d^\top \\ 0_{q \times d} \end{bmatrix}, \qquad y = \begin{bmatrix} \mathbf{1}_d \\ -1 \\ 0_q \end{bmatrix}. \tag{56}$$

For every $d \le s \le m$, this family satisfies

$$M_s = \alpha L^* G^{-1}. \tag{57}$$

Proof. The first $d+1$ rows are the core. Direct calculation gives

$$G = I_d + \mathbf{1}_d \mathbf{1}_d^\top, \qquad G^{-1} = I_d - \frac{\mathbf{1}_d \mathbf{1}_d^\top}{d+1}, \qquad w^* = 0, \qquad L^* = d+1. \tag{58}$$

A supported size-s set contains either exactly d core rows or all $d+1$. Every d-core minor has squared determinant one, whereas the all-core Gram determinant is $d+1$. With the convention that an out-of-range binomial coefficient is zero, the aggregate determinant weights are

$$W_d = (d+1)\binom{q}{s-d}, \qquad W_{d+1} = (d+1)\binom{q}{s-d-1}. \tag{59}$$

Pascal's identity therefore gives

$$\mathbb{P}_X(\text{exactly } d \text{ core rows}) = \frac{W_d}{W_d + W_{d+1}} = \frac{m-s}{m-d} = \alpha. \tag{60}$$

On the all-core event, the normal equations give $w_S = 0$. Conditional on exactly d core rows, the omitted core row is uniform among the $d+1$ possibilities and the corresponding interpolating fits are

$$a_0 = \mathbf{1}_d, \qquad a_i = \mathbf{1}_d - (d+1)e_i, \quad i = 1, \dots, d, \tag{61}$$

where e_i is the ith coordinate vector in $\mathbb{R}^d$. Their mean is zero, and their complete matrix second moment is

$$\begin{aligned} \frac{1}{d+1}\left(a_0 a_0^\top + \sum_{i=1}^{d} a_i a_i^\top\right) &= (d+1)I_d - \mathbf{1}_d\mathbf{1}_d^\top \\ &= L^* G^{-1}. \end{aligned} \tag{62}$$

Indeed, expanding the sum gives $(d+1)^2 I_d - (d+1)\mathbf{1}_d\mathbf{1}_d^\top$ before division by $d+1$. Multiplying (62) by the event probability (60) proves (57), including all off-diagonal entries. The same formulas cover $s = d$ and $s = m$ through $\alpha = 1$ and $\alpha = 0$, respectively. This proposition supplies an attaining family only. □

Proposition 19 (Strict interior slack and row-general-position supremum)**.** *If X is in row general position, $L^* > 0$, $d \geq 2$, and $d < s < m$, then*

$$M_s \prec \alpha L^* G^{-1}. \tag{63}$$

For every fixed (m, d, s) with $m > d$, $d \geq 2$, and $d \leq s < m$, row-general-position perturbations $X^{(a)} \to X_0$ of the family (56) *can be chosen so that, with the same fixed y,*

$$\left\| \frac{G_a^{1/2} M_{s,a} G_a^{1/2}}{L_a^*} - \alpha I_d \right\|_{\mathrm{op}} \longrightarrow 0. \tag{64}$$

Thus, for $d \geq 2$ and $d < s < m$, the coefficient is not attained in row general position but is its supremum. At the rank-size endpoint $s = d$, the basis identity already gives equality for every positive-loss row-general-position design; at $s = m$ no normalized ratio is formed.

Proof. For strictness, the slack is the positive-semidefinite expectation in (50). We show that it has no nonzero null direction. Let $H = [X \;\; e]$. Since $X^\top e = 0$ and $L^* > 0$, $\operatorname{rank}(H) = d+1$; since $d \geq 2$ and $d < s < m$, also $m \geq d+2$. The rank-$(d+1)$ row matroid of H has a circuit. No set of at most d rows of H can be dependent, because projection of such a dependence onto the first d coordinates would contradict row general position of X. Hence the circuit contains at least $d+1$ rows. Every circuit element is a non-coloop, so H has at least $d+1$ non-coloops.

Fix $u \neq 0$ and put $v = G^{-1}u \neq 0$. At most $d-1$ rows can satisfy $x_i^\top v = 0$: any d such rows would form a singular d-row submatrix. Thus some non-coloop i satisfies $x_i^\top v \neq 0$. Because i is not a coloop, H has a row basis R avoiding i. Extend it to a size-s set $R \subseteq S \subseteq [m] \setminus \{i\}$, possible because $d+1 \leq s \leq m-1$. Then $\operatorname{rank}(H_S) = d+1$, and the Schur identity (49) together with row general position gives $L_S > 0$ and $D_S > 0$.

Let $C_S = G - G_S = X_{S^c}^\top X_{S^c}$. Since $u = Gv$, direct expansion gives

$$\begin{aligned} u^\top(G_S^{-1} - G^{-1})u &= v^\top\{C_S + C_S G_S^{-1} C_S\}v \\ &\geq v^\top C_S v = \|X_{S^c} v\|_2^2 > 0, \end{aligned} \tag{65}$$

where the last inequality uses $i \notin S$ and $x_i^\top v \neq 0$. This S has positive probability and positive L_S. Every other summand in (50) is positive semidefinite, so the slack is positive in every nonzero direction u, proving (63).

For the supremum, use the Vandermonde perturbations from Proposition 17 with $X = X_0$. The directional liminf (55), applied to a fixed basis of coefficient directions and summed, together with exact attainment at the limit, gives

$$\liminf_a \operatorname{tr}(G_a M_{s,a}) \geq d\alpha L^*. \tag{66}$$

For completeness, one may take the fixed directions $u_j = G^{1/2} e_j$: their limiting sum is $\operatorname{tr}(GM_s) = d\alpha L^*$. Replacing G by G_a in the trace changes the sum by $o(1)$, because $G_a \to G$ and $M_{s,a} \preceq \alpha L_a^* G_a^{-1}$ keeps $M_{s,a}$ bounded. The envelope also gives the matching upper bound

$$\operatorname{tr}(G_a M_{s,a}) \leq d\alpha L_a^* \longrightarrow d\alpha L^*. \tag{67}$$

Consequently the matrices

$$R_a := \alpha L_a^* I_d - G_a^{1/2} M_{s,a} G_a^{1/2}$$

are positive semidefinite and satisfy $\operatorname{tr}(R_a) \to 0$. For a positive-semidefinite matrix, $\|R_a\|_{\mathrm{op}} \leq \operatorname{tr}(R_a)$. Since $L_a^* \to L^* = d + 1 > 0$, division by L_a^* proves (64). When $d \geq 2$ and $d < s < m$, combining this convergence with (63) distinguishes supremum from attainment. □

B Residual-augmented representation and response-aware resolvent envelope

Use the fixed X, y, full fit w^*, residual e, Grams G, G_S, and supported selected fit/loss from Appendix A; write $L^* = \|e\|^2$. The estimator uses the ordinary indexed volume law (31) and unweighted OLS, and its covariance is $M_s = C_s$ from (34). Zero-volume sets receive neither a fit nor an inverse. Throughout this section,

$$L^* > 0, \qquad d < s < m, \tag{68}$$

so $m \geq d + 2$ and

$$\alpha = \frac{m - s}{m - d}, \qquad \beta = \frac{s - d}{m - d} = 1 - \alpha, \qquad \gamma = \frac{m - s}{m - d - 1}. \tag{69}$$

The zero-loss and endpoint branches remain those of Appendix A; neither the normalized residual nor the rank-$(d + 1)$ augmented law below is invoked there.

B.1 Whitening and the exact change of measure

Assume first that X is in row general position, meaning that every indexed d-row submatrix is nonsingular, in addition to (68). Define

$$A = XG^{-1/2}, \qquad z = \frac{e}{\sqrt{L^*}}, \qquad B = [A \;\; z]. \tag{70}$$

The full normal equations give

$$A^\top A = I_d, \qquad A^\top z = 0, \qquad \|z\|_2 = 1, \qquad B^\top B = I_{d+1}. \tag{71}$$

For each S, let $K_S = A_S^\top A_S$. The augmented volume law is a different law, of rank $d + 1$, with its own support and normalizer:

$$Z_{B,s} = \sum_{|S|=s} \det(B_S^\top B_S) = \binom{m - d - 1}{s - d - 1}, \qquad \mathbb{P}_B(S) = \frac{\det(B_S^\top B_S)}{Z_{B,s}}. \tag{72}$$

Here P_B is a response-dependent analysis law induced by the full residual, not the sampler for w_S; the estimator continues to sample and fit under the ordinary feature-only law P_X. An augmented-supported set has $\operatorname{rank}(B_S) = d + 1$, hence $\operatorname{rank}(A_S) = d$ and $K_S \succ 0$. Thus every inverse under $\mathbb{E}_B$ below is taken only on augmented support. No inverse is assigned to a zero-augmented-volume set. Augmented support is contained in, and can be strictly smaller than, ordinary support.

Write $B = [X \;\; y]C$, where C is the invertible column transformation in (70), with $\det(C)^2 = 1/(\det(G)L^*)$. The supported-set Schur identity (49) therefore gives

$$\det(B_S^\top B_S) = \frac{D_S L_S}{\det(G)L^*} = \det(K_S)\frac{L_S}{L^*} \qquad (D_S > 0). \tag{73}$$

If S is not ordinary-supported, then $\operatorname{rank}(A_S) < d$ and $\operatorname{rank}(B_S) \le d$, so both determinants in the corresponding zero-equals-zero rank statement vanish; neither w_S nor L_S is defined there. Also $\det(K_S) = D_S/\det(G)$. Consequently, if the finite measure μ_X is defined by

$$\mu_X(\{S\}) = \begin{cases} \mathbb{P}_X(S)L_S, & D_S > 0, \\ 0, & D_S = 0, \end{cases} \tag{74}$$

then the two distinct normalizers in (31) and (72) yield the exact setwise change of measure

$$\boxed{\mathbb{P}_X(S)L_S = \beta L^* \mathbb{P}_B(S) \quad (D_S > 0), \qquad \mu_X(\{S\}) = \beta L^* \mathbb{P}_B(S) \quad (\text{all } S)}, \qquad \frac{Z_{B,s}}{\binom{m-d}{s-d}} = \beta. \tag{75}$$

The zero branch in (74) extends only a measure, not an inverse fit.

B.2 Exact row-general-position covariance transform

Proposition 13 gives unbiasedness and $M_s = L^* G^{-1} - \mathbb{E}_X[L_S G_S^{-1}]$ on the present row-general-position domain. Congruence by $G^{1/2}$, using $K_S^{-1} = G^{1/2} G_S^{-1} G^{1/2}$, and the setwise change of measure (75) therefore give

$$\boxed{\frac{\overline{M}_s}{L^*} = I_d - \beta\, \mathbb{E}_B[K_S^{-1}]}, \qquad \overline{M}_s = G^{1/2} M_s G^{1/2}. \tag{76}$$

This exact identity is established for row general position, $L^* > 0$, and $d < s < m$. Proposition 21 shows that it can fail on a full-column-rank boundary; row general position is a sufficient domain, not a necessary claim for each fixed design.

B.3 Augmented exclusion marginal and the augmented Gram first moment

Let $a_i^\top$ and $b_i^\top$ denote row i of A and B, respectively, and define

$$h_i = \|b_i\|_2^2, \qquad R = \sum_{i=1}^{m} (1 - h_i) a_i a_i^\top. \tag{77}$$

Lemma 10, with $F = B$, $p = d + 1$ and $J = \{i\}$, and the rank-one determinant lemma give

$$\mathbb{P}_B(i \notin S) = \gamma \det(I_{d+1} - b_i b_i^\top) = \gamma(1 - h_i). \tag{78}$$

Since $\sum_i a_i a_i^\top = I_d$, linearity now gives

$$\boxed{\mathbb{E}_B K_S = \sum_i \mathbb{P}_B(i \in S) a_i a_i^\top = I_d - \gamma R \succ 0.} \tag{79}$$

Strict positivity follows because every augmented-supported S has $K_S \succ 0$ and the augmented law has nonempty support.

B.4 Classical operator Jensen and two resolvent bounds

The classical operator Jensen inequality for the inversion map on the positive-definite cone (Hansen & Pedersen, 2003) gives

$$\mathbb{E}_B[K_S^{-1}] \succeq (\mathbb{E}_B K_S)^{-1} = (I_d - \gamma R)^{-1}. \tag{80}$$

The change of measure and augmented exclusion moment above supply its volume-specific input. The minus sign in (76) reverses the order, giving

$$\boxed{\overline{M}_s \preceq L^* \left[I_d - \beta (I_d - \gamma R)^{-1} \right].} \tag{81}$$

Moreover, $R \succeq 0$ and $I_d - \gamma R \succ 0$, so scalar functional calculus gives $(I_d - \gamma R)^{-1} \succeq I_d + \gamma R$. Hence the linear corollary used to derive the response-uniform slack in Appendix C is

$$\boxed{\overline{M}_s \preceq \alpha L^* I_d - \beta\gamma L^* R.} \tag{82}$$

No high-dimensional sharpness is claimed for (81).

B.5 Second-centered-Gram refinement of the response-aware resolvent

Proposition 20 (Contraction-specific second-moment refinement)**.** *Assume that X has full column rank and that $L^* > 0$ and $d < s < m$. Let A, z, B and $\mathbb{P}_B$ be as in* (70)–(72)*, and write*

$$\Omega_B = \{S \subseteq [m] : |S| = s,\ \det(B_S^\top B_S) > 0\}. \tag{83}$$

The law $\mathbb{P}_B$ is analysis-only, and every selected inverse below is restricted to Ω_B. Define

$$H_s = \mathbb{E}_B K_S, \qquad \Delta_S = K_S - H_s, \qquad V_s = \mathbb{E}_B[\Delta_S^2], \qquad \mathcal{Q}_s = H_s^{-1} V_s H_s^{-1}. \tag{84}$$

Then $0 \prec K_S \preceq I_d$ on Ω_B, and

$$\mathbb{E}_B[K_S^{-1}] - H_s^{-1} = \mathcal{Q}_s + H_s^{-1}\mathbb{E}_B\left[\Delta_S(K_S^{-1} - I_d)\Delta_S\right] H_s^{-1} \succeq \mathcal{Q}_s \succeq 0. \tag{85}$$

The coefficient of V_s is one: this is an exact ordered inverse remainder, not a Taylor approximation. If X is in row general position, the exact covariance transform gives

$$\overline{M}_s \preceq L^* \left[I_d - \beta H_s^{-1} - \beta \mathcal{Q}_s \right] \preceq L^* \left[I_d - \beta H_s^{-1} \right]. \tag{86}$$

The correction is computable from singleton and pair exclusions. Put $T = S^c$, $k = m - s$, $q = m - d - 1$, $D_i = a_i a_i^\top$, and retain $h_i = \|b_i\|_2^2$. The singleton exclusion probability is

$$\pi_i = \mathbb{P}_B(i \in T) = \frac{k}{q}(1 - h_i), \tag{87}$$

and, for $q \geq 2$, the pair exclusion probability is

$$\pi_{ij} = \mathbb{P}_B(i, j \in T) = \frac{k(k-1)}{q(q-1)} \left[(1 - h_i)(1 - h_j) - (b_i^\top b_j)^2 \right]. \tag{88}$$

When $m = d + 2$, the strict domain forces $q = k = 1$; in that branch set $\pi_{ij} = 0$ rather than evaluating (88)*. In either branch,*

$$V_s = \sum_i \pi_i (1 - \pi_i) D_i^2 + \sum_{i<j} (\pi_{ij} - \pi_i \pi_j)(D_i D_j + D_j D_i). \tag{89}$$

For every full-column-rank design in the same strict domain, without a row general position assumption, define

$$F_s = \mathbb{E}_B G_S, \qquad W_s = \mathbb{E}_B[(G_S - F_s) G^{-1} (G_S - F_s)] \succeq 0. \tag{90}$$

For the ordinary-law supported, full-fit-centered second moment M_s in (34)*, only the one-sided hierarchy*

$$M_s \preceq L^* \left[G^{-1} - \beta F_s^{-1} - \beta F_s^{-1} W_s F_s^{-1} \right] \preceq L^* \left[G^{-1} - \beta F_s^{-1} \right] \tag{91}$$

is asserted. On a boundary it is not an exact inverse-moment identity or covariance-gap decomposition (Remark 22). The proposition is not invoked at the existing zero branches $L^* = 0$, $s = m$, or $m = d$, nor at $s = d < m$; when $m = d + 1$ the strict domain is empty.

Proof. For $S \in \Omega_B$, augmented support implies $K_S \succ 0$, while $I_d - K_S = A_{S^c}^\top A_{S^c} \succeq 0$. Thus K_S is a positive contraction and $H_s \succ 0$. With $K = K_S$, $H = H_s$, and $\Delta = K - H$, direct multiplication, without commuting any factors, gives

$$K^{-1} = H^{-1} - H^{-1}\Delta H^{-1} + H^{-1}\Delta K^{-1}\Delta H^{-1}. \tag{92}$$

Taking expectations removes the centered linear term. Splitting $K_S^{-1} = I_d + (K_S^{-1} - I_d)$ in the last term yields (85). Its remainder is positive semidefinite because $K_S^{-1} - I_d \succeq 0$. Thus the first-order comparison is the Jensen bound (80); (85) retains the ordered, support-restricted remainder for these contractions. Combining it with (76) proves (86).

For the marginal formulas, apply Lemma 10 with $F = B$, $p = d + 1$ and $|J| = 1, 2$. The rank-one and rank-two determinant lemmas yield (87) and (88). When $q = k = 1$, only one row is excluded and the pair probability is zero; no factor $q(q-1)$ is divided by. Finally, $K_S = I_d - \sum_{i \in T} D_i$; expanding the square of its centered version retains both product orders and gives the anticommutator in (89).

On a row-general-position design, $F_s = G^{1/2} H_s G^{1/2}$ and $W_s = G^{1/2} V_s G^{1/2}$, so congruence of (86) gives the raw-coordinate hierarchy. The next subsection extends it by directional liminf. □

B.6 One-sided extension to arbitrary full-column-rank boundary designs

Now let X be an arbitrary full-column-rank design, still with $L^* > 0$ and $d < s < m$; row general position is no longer assumed. The quantities A, z, B, h_i remain defined by (70) and (77). In original coordinates put

$$Q = \sum_{i=1}^{m} (1 - h_i) x_i x_i^\top. \tag{93}$$

The augmented exclusion calculation itself did not require row general position. Because augmented support implies $G_S \succ 0$, it gives the exact first moment

$$F_s = \mathbb{E}_B G_S = G - \gamma Q \succ 0. \tag{94}$$

The first-order member of Proposition 20 is therefore the response-aware upper inequality

$$\boxed{M_s \preceq L^* \left[G^{-1} - \beta (G - \gamma Q)^{-1} \right].} \tag{95}$$

Choose row-general-position $X_a \to X$ as in Appendix A.5, keeping y fixed. The full quantities converge and $L_a^* > 0$ eventually. Lemma 16 gives, for every fixed coefficient direction u,

$$u^\top M_s u \le \liminf_a u^\top M_{s,a} u. \tag{96}$$

The finite-sum definitions of F_s and W_s in (90) are continuous under this perturbation. Nonempty augmented support gives $F_s \succ 0$, so its inverse is continuous as well; hence both right-hand sides of (91) are continuous. Applying the row-general-position hierarchy to X_a and passing to the directional liminf proves (91), and in particular (95). This passage does not continue selected inverses through zero support. The same passage applied to (82), or the elementary inequality $(G - \gamma Q)^{-1} \succeq G^{-1} + \gamma G^{-1} Q G^{-1}$, gives its original-coordinate linear form

$$M_s \preceq \alpha L^* G^{-1} - \beta \gamma L^* G^{-1} Q G^{-1}. \tag{97}$$

Proposition 21 (Boundary failure of the exact inverse-moment identity)**.** *Let*

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad y = (1, -1, 2, 0, 3)^\top, \qquad s = 3. \tag{98}$$

Then $G = 2I_2$, $L^* = 10$, and $M_s = I_2/6$, whereas

$$\mathbb{E}_B[G_S^{-1}] = \frac{39}{40} I_2, \qquad M_s - L^* \left\{ G^{-1} - \beta \mathbb{E}_B[G_S^{-1}] \right\} = -\frac{19}{12} I_2 \neq 0. \tag{99}$$

The boundary resolvent inequality nevertheless remains strict, with slack $(209/126) I_2 \succ 0$.

Proof. Here $w^* = (3/2, -1/2)^\top$ and $e = (-1/2, -1/2, 1/2, 1/2, 3)^\top$, which gives the stated G and L^*. Put $C_1 = \{1, 3\}$ and $C_2 = \{2, 4\}$. The eight ordinary-supported triples, whose normalizer is $Z_{X,3} = 12$, split as follows:

type of S	count	D_S	$w_S - w^*$	L_S
$C_1 \cup \{j\}, j \in C_2$	2	2	$(0, \pm 1/2)^\top$	$1/2$
$\{i\} \cup C_2, i \in C_1$	2	2	$(\pm 1/2, 0)^\top$	$1/2$
$\{i, j, 5\}, (i, j) \in C_1 \times C_2$	4	1	$(\pm 1/2, \pm 1/2)^\top$	9

All four sign combinations occur in the last row. Weighting by $\mathbb{P}_X(S) = D_S/12$, the three rows contribute respectively $\mathrm{diag}(0, 1/12)$, $\mathrm{diag}(1/12, 0)$, and $I_2/12$ to the centered second moment. Hence $M_s = I_2/6$.

For the augmented law, $Z_{B,3} = 1$. The setwise Schur determinant gives

$$\mathbb{P}_B(S) = \det(B_S^\top B_S) = \frac{D_S L_S}{\det(G) L^*} = \frac{D_S L_S}{40}. \tag{100}$$

The three table rows have per-triple augmented probabilities $1/40$, $1/40$, $9/40$ and inverse Gramians $\mathrm{diag}(1/2, 1)$, $\mathrm{diag}(1, 1/2)$, I_2, respectively. Thus

$$\mathbb{E}_B[G_S^{-1}] = \frac{2}{40}\,\mathrm{diag}(1/2, 1) + \frac{2}{40}\,\mathrm{diag}(1, 1/2) + \frac{36}{40} I_2 = \frac{39}{40} I_2. \tag{101}$$

With $\beta = 1/3$, the claimed defect is $I_2/6 - 10\{I_2/2 - (1/3)(39/40) I_2\} = -19 I_2/12$.

Finally, $h_i = x_i^\top G^{-1} x_i + e_i^2/L^* = 21/40$ for $i \le 4$, $h_5 = 9/10$, $\gamma = 1$, and $Q = (19/20) I_2$. Thus $G - \gamma Q = (21/20) I_2$, and the exact boundary slack is

$$L^* \left[G^{-1} - \beta (G - \gamma Q)^{-1} \right] - M_s = 10 \left(\frac{1}{2} - \frac{1}{3} \frac{20}{21} \right) I_2 - \frac{1}{6} I_2 = \frac{209}{126} I_2 \succ 0. \tag{102}$$

□

Remark 22 (Boundary scope of identities and inequalities). The Schur determinant, measure change, exclusion marginal, and Gram first moment remain exact on their stated support. The covariance identity (76) is proved under row general position and can fail on a full-column-rank boundary, as Proposition 21 shows. There, the one-sided upper inequalities (91), (95) and (97) still hold. No fit is assigned to an ordinary zero-volume set, and no exact covariance-gap decomposition is claimed there.

C Fixed-residual contact kernel and spectral phase

This section proves the fixed-residual kernel identity in Theorem 3 and the zero-margin spectral phase of Theorem 5. The one-sided bound and saturated-row omissions give the kernel; compactness then gives the spectral equivalence. Complement classes give pair deletion in Appendix D and whole-query contact in Appendix E. Sampling uses ordinary unrescaled indexed volume and selected unweighted OLS, with fits only on positive-volume sets.

C.1 Domain, support, and extrema

For the spectral formulation, define

$$\mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A, s) = \max_{z \in \mathcal{Z}_A} \frac{\lambda_{\max}(\overline{M}_s^{A,z})}{L^*}. \tag{103}$$

When the two-row deletion condition holds, there are one $z_* \in \mathcal{Z}_A$ and one unit coefficient direction v_* such that, simultaneously for every $s' \in \{d+1, \dots, m-1\}$,

$$\frac{v_*^\top \overline{M}_{s'}^{A,z_*} v_*}{L^*} = \frac{\lambda_{\max}(\overline{M}_{s'}^{A,z_*})}{L^*} = \frac{m - s'}{m - d}. \tag{104}$$

Throughout this proof, $A \in \mathbb{R}^{m \times d}$ has orthonormal columns and no coloops, so $\ell_i < 1$ for every row; also $d \ge 1$, $m \ge d + 2$, and $d < s < m$. For any $\theta \in \mathbb{R}^d$, let

$$\mathcal{Z}_A = \{z \in \ker(A^\top) : \|z\|_2 = 1\}, \qquad y_z = A\theta + \sqrt{L^*} z, \qquad L^* > 0. \tag{105}$$

The residual sphere is nonempty and compact because $m > d$. For a positive-volume indexed set S, define

$$p_A(S) = \frac{\det(A_S^\top A_S)}{\binom{m-d}{s-d}}, \qquad w_S(z) = (A_S^\top A_S)^{-1} A_S^\top (y_z)_S. \tag{106}$$

The fixed-cardinality Cauchy–Binet identity normalizes this supported law. For fixed A and S, $w_S(z)$ is affine in z, so the finite sum

$$\overline{M}_s^{A,z} = \sum_{\substack{|S|=s \\ \det(A_S^\top A_S)>0}} p_A(S)(w_S(z)-\theta)(w_S(z)-\theta)^\top \tag{107}$$

is continuous in z. Hence both the minimum defining ν_A and the maximum defining $\mathsf{T}_{\mathrm{spec}}^{\mathrm{RU}}(A,s)$ are attained.

For $z \in \mathcal{Z}_A$, the matrix $B = [A\ z]$ has orthonormal columns. Consequently $BB^\top$ is an orthogonal projection and

$$h_i := (BB^\top)_{ii} = \ell_i + z_i^2 \le 1. \tag{108}$$

It follows that

$$R_A(z) = \sum_{i=1}^m (1-h_i) a_i a_i^\top \succeq 0, \qquad \nu_A \ge 0. \tag{109}$$

C.2 Uniform slack and the strict direction

The response-aware covariance relation, in its boundary-valid one-sided form, gives for every $z \in \mathcal{Z}_A$

$$\frac{\overline{M}_s^{A,z}}{L^*} \preceq \alpha_s I_d - \beta_s \gamma_s R_A(z), \qquad \alpha_s = \frac{m-s}{m-d}, \quad \beta_s = \frac{s-d}{m-d}, \quad \gamma_s = \frac{m-s}{m-d-1}. \tag{110}$$

The one-sided bound holds after whitening a full-rank design, including support-changing boundaries; it does not extend the exact inverse-moment identity. Since $R_A(z) \succeq \nu_A I_d$ for every admissible z, (110) implies

$$\frac{\overline{M}_s^{A,z}}{L^*} \preceq (\alpha_s - \beta_s \gamma_s \nu_A) I_d. \tag{111}$$

Taking the largest eigenvalue and then the maximum over z proves (184). Because $\beta_s \gamma_s > 0$, it also proves $\nu_A > 0 \Rightarrow \mathsf{T}_{\mathrm{spec}}^{\mathrm{RU}}(A,s) < \alpha_s$.

C.3 Saturated-row contact kernel

Lemma 23 (Saturated-row contact kernel)**.** *Under the domain of this appendix, fix $z \in \mathcal{Z}_A$ and an interior size s. With*

$$C_s(z) = \overline{M}_s^{A,z}/L^*, \qquad \Delta_s(z) = \alpha_s I_d - C_s(z),$$

one has

$$\ker \Delta_s(z) = \ker R_A(z). \tag{112}$$

The equality holds for the same fixed residual at every interior size.

Proof. The one-sided inequality (110) gives $\Delta_s(z) \succeq \beta_s \gamma_s R_A(z)$, and hence $\ker \Delta_s(z) \subseteq \ker R_A(z)$. For the reverse inclusion, take a unit $v \in \ker R_A(z)$, set

$$r_i = a_i^\top v, \qquad J = \{i : r_i \neq 0\}, \tag{113}$$

and use (108). Every term in

$$0 = v^\top R_A(z) v = \sum_i (1-h_i) r_i^2 \tag{114}$$

is nonnegative. Therefore

$$h_i = 1 \quad (i \in J), \qquad \sum_{i \in J} r_i^2 = \|Av\|_2^2 = 1. \tag{115}$$

The no-coloop condition gives

$$z_i^2 = 1 - \ell_i > 0 \quad (i \in J). \tag{116}$$

For a saturated row, $BB^\top e_i = e_i$, hence

$$e_i = Aa_i + zz_i. \tag{117}$$

Let S be a positive-volume set and suppose it omits $i \in J$. Restricting (117) to S gives $z_S = -(1/z_i)A_S a_i$. The selected least-squares equations then yield

$$\frac{w_S(z) - \theta}{\sqrt{L^*}} = (A_S^\top A_S)^{-1} A_S^\top z_S = -\frac{a_i}{z_i}. \tag{118}$$

No positive-volume set can omit two distinct members $i, j \in J$. Otherwise the two error expressions would give $a_i/z_i = a_j/z_j = q$, while orthogonality of the two saturated augmented rows gives

$$a_i^\top a_j + z_i z_j = 0. \tag{119}$$

Its left side would then be $z_i z_j(\|q\|_2^2 + 1) \neq 0$, a contradiction. If S contains all of J, all omitted rows are orthogonal to v, so

$$(A_S^\top A_S)v = v, \qquad v^\top \frac{w_S(z) - \theta}{\sqrt{L^*}} = v^\top A_S^\top z_S = v^\top A^\top z = 0. \tag{120}$$

By symmetry, the first equality also gives $v^\top (A_S^\top A_S)^{-1} = v^\top$, which justifies the middle equality.

For $i \in J$, Lemma 10 with $F = A$ and $p = d$ gives

$$\mathbb{P}_A(i \notin S) = \alpha_s(1 - \ell_i) = \alpha_s z_i^2. \tag{121}$$

Here $A_{-i}^\top A_{-i} = I_d - a_i a_i^\top \succ 0$ by no coloops. The omissions in J are mutually exclusive, and the remaining support has zero directional error. Hence

$$\begin{aligned} v^\top C_s(z) v &= \sum_{i \in J} \mathbb{P}_A(i \notin S) \frac{r_i^2}{z_i^2} \\ &= \alpha_s \sum_{i \in J} r_i^2 = \alpha_s. \end{aligned} \tag{122}$$

Since $\Delta_s(z) \succeq 0$, this equality gives $v \in \ker \Delta_s(z)$. The calculation used only the positive-volume support and applies unchanged at every strict-interior size. □

C.4 Complement classes and their contact spaces

Let $N \in \mathbb{R}^{m \times n}$, $n = m - d$, have orthonormal columns spanning $\ker A^\top$. Write its rows as $n_i^\top$ and normalize them to $\widehat{n}_i = n_i/\sqrt{c_i}$, where $c_i = 1 - \ell_i = \|n_i\|^2 > 0$. For $z = N\xi$, $\|\xi\| = 1$,

$$z_i^2 = c_i \quad \Longleftrightarrow \quad |\widehat{n}_i^\top \xi| = 1 \quad \Longleftrightarrow \quad \widehat{n}_i = \pm\xi.$$

If the saturated set is empty, $R_A(z) \succ 0$. Otherwise it is one entire maximal parallel class C. For a class direction $\widehat{n}_C$, the residual $z = \pm N\widehat{n}_C$ saturates exactly C. Since every remaining coefficient in $R_A(z)$ is positive,

$$\ker R_A(z) = \{v : \operatorname{supp}(Av) \subseteq C\} = \ker A_{-C} = V_C.$$

The isometry $v \mapsto Av$ identifies V_C with vectors supported on C and orthogonal to N_C. Because N_C has rank one, $\dim V_C = |C| - 1$. Different classes give orthogonal spaces V_C, since their images under A have disjoint supports. In particular, a singleton class saturates one augmented row but has $V_C = \{0\}$, so it produces no nonzero covariance contact.

C.5 SPECTRAL CONTACT AND A COMMON WITNESS

Compactness, positive semidefiniteness, and Lemma 23 give, at every interior size,

$$\begin{aligned}\nu_A = 0 &\iff \exists z \in \mathcal{Z}_A : \ker R_A(z) \neq \{0\} \\ &\iff \exists z \in \mathcal{Z}_A : \ker \Delta_s(z) \neq \{0\} \iff \mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A,s) = \alpha_s.\end{aligned}$$

The last equivalence also uses the universal ceiling and attainment of the spectral maximum. A single unit pair (z, v) with $v \in \ker R_A(z)$ works at every $s' \in \{d+1, \dots, m-1\}$ by the fixed-residual kernel identity, proving (104). This is directional contact, not the full matrix equality $\overline{M}_s^{A,z} = \alpha_s L^* I_d$.

The converse holds even with coloops: at an attained spectral contact, (110) gives $0 \geq \beta_s\gamma_s u^\top R_A(z) u \geq 0$ for a unit top eigenvector u. Since $\beta_s\gamma_s > 0$, this forces $\nu_A = 0$. The no-coloop condition is needed for the forward support argument. The pair-deletion characterization follows in Appendix D.

C.6 SCOPE BOUNDARIES

The no-coloop condition is sufficient, not asserted to be necessary. It cannot be removed wholesale: for

$$d = 1, \qquad m = 3, \qquad s = 2, \qquad A = (1, 0, 0)^\top, \tag{123}$$

the margin is zero, but every positive-volume sample contains the leverage-one row and the selected coefficient is deterministic, so $\mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A, 2) = 0 < 1/2 = \alpha_2$. The theorem requires $L^* > 0$ and $d < s < m$; its normalized residual and centered full-Gram-whitened covariance ratio are undefined at $L^* = 0$ and not evaluated at $s = d$ or $s = m$.

D GLOBAL COMPLEMENT PHASE TEST AND CERTIFICATE

This appendix proves Proposition D.1. The argument uses the shared contact-class characterization on the real Parseval no-coloop domain and expresses the margin ν_A from Section 6.1 through complement geometry.

Proposition D.1 (Feature-only phase test)**.** *Under the assumptions of Theorem 5, put $P_\perp = I_m - AA^\top$, $c_i = (P_\perp)_{ii} = 1 - \ell_i$, and*

$$\begin{aligned}&\rho_\perp(A) = \max_{i \neq j} \frac{|(P_\perp)_{ij}|}{\sqrt{c_i c_j}}, \qquad \kappa_\perp(A) = \min_{k \in [m]} \lambda_{\min}\left(\sum_{i \neq k} c_i a_i a_i^\top\right), \\ &t_\perp(A) = (1 - \rho_\perp(A))\kappa_\perp(A).\end{aligned} \tag{124}$$

Then

$$\nu_A = 0 \iff \rho_\perp(A) = 1 \iff \det((P_\perp)_{\{i,j\},\{i,j\}}) = 0 \quad \textit{for some } i \neq j. \tag{125}$$

If $c_{\min} = \min_i c_i$, then

$$0 < c_{\min}, \qquad c_{\min}^2 \leq \kappa_\perp(A), \qquad 0 \leq (1 - \rho_\perp) c_{\min}^2 \leq t_\perp(A) \leq \nu_A. \tag{126}$$

Consequently,

$$\alpha_s - \mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A, s) \geq \beta_s\gamma_s t_\perp(A), \qquad d < s < m. \tag{127}$$

The complement Gram interpretation is standard (Casazza et al., 2013); $\kappa_\perp$ is an erasure-type lower-frame quantity (Fickus & Mixon, 2012). Appendix U compares these frame results with the pair-minor/deletion certificate below.

Proof of Proposition D.1. Use N, its rows $n_i^\top$, and $\widehat{n}_i = n_i/\sqrt{c_i}$ from Appendix C.4; put $n = m - d$ and $u_i := \widehat{n}_i \in \mathbb{R}^n$. Since $P_\perp = NN^\top$ and $c_i > 0$, the u_i are unit vectors and

$$(P_\perp)_{ij} = \sqrt{c_i c_j}\, u_i^\top u_j, \qquad \rho_\perp(A) = \max_{i \neq j} |u_i^\top u_j| \in [0, 1]. \tag{128}$$

For $z = N\xi \in \mathcal{Z}_A$ with $\|\xi\|_2 = 1$,

$$z_i = \sqrt{c_i}\, u_i^\top \xi, \qquad c_i - z_i^2 = c_i\{1 - (u_i^\top \xi)^2\} \geq 0. \tag{129}$$

Thus $R_A(z) \succeq 0$ for every $z \in \mathcal{Z}_A$, so $\nu_A \geq 0$.

By compactness and Appendix C.4, $\nu_A = 0$ holds exactly when a maximal parallel class C has $\dim V_C = |C| - 1 > 0$. This is equivalent to a parallel or antiparallel pair of normalized complement rows, hence to $\rho_\perp(A) = 1$.

For any distinct i, j, the relevant principal complement minor is

$$\det\big((P_\perp)_{\{i,j\},\{i,j\}}\big) = c_i c_j - (P_\perp)_{ij}^2 = c_i c_j\{1 - (u_i^\top u_j)^2\}. \tag{130}$$

Because $c_i c_j > 0$, this minor vanishes exactly when $|u_i^\top u_j| = 1$. This proves all equivalences in (125), including both parallel and antiparallel complement directions.

The same pair-minor has a direct row-deletion interpretation. For a pair $J = \{i, j\}$, let A_{-J} retain the indexed rows outside J. Since $P_\perp = I_m - AA^\top$ and $A^\top A = I_d$, Sylvester's determinant identity gives

$$\det\big((P_\perp)_{J,J}\big) = \det\big(I_2 - A_J A_J^\top\big) = \det\big(I_d - A_J^\top A_J\big) = \det\big(A_{-J}^\top A_{-J}\big). \tag{131}$$

Consequently, the vanishing-pair condition is equivalent to the existence of two indexed whitened rows whose deletion lowers the column rank of A. This determinant calculation is independent of the saturated-row contact classification; combined with the zero-margin/spectral equivalence proved in Appendix C, it completes the two-row statement (13).

It remains to prove the lower certificate. Let $c_{\min} := \min_i c_i > 0$ and

$$S_k := \sum_{i \neq k} c_i a_i a_i^\top.$$

Parsevalness and no-coloop give

$$S_k \succeq c_{\min} \sum_{i \neq k} a_i a_i^\top = c_{\min}(I_d - a_k a_k^\top) \succeq c_{\min} c_k I_d \succeq c_{\min}^2 I_d. \tag{132}$$

The middle spectral inequality follows from $a_k a_k^\top \preceq \|a_k\|_2^2 I_d$. Therefore $\kappa_\perp(A) \geq c_{\min}^2 > 0$.

We use the following weighted Gram estimate. If $u_1, \ldots, u_m$ are unit vectors with $\max_{i \neq j} |u_i^\top u_j| \leq \rho$ and $w_i \geq 0$, then

$$\lambda_{\max}\left(\sum_i w_i u_i u_i^\top\right) \leq \rho \sum_i w_i + (1 - \rho) \max_i w_i. \tag{133}$$

The right side is nondecreasing in ρ, so it suffices to prove the estimate for $\rho = \max_{i \neq j} |u_i^\top u_j| \in [0, 1]$. Let U have rows $u_i^\top$ and $D = \operatorname{diag}(w_1, \ldots, w_m)$. For unit $y \in \mathbb{R}^m$, taking absolute values of off-diagonal inner products and using Cauchy–Schwarz, without a common sign assumption, gives

$$\begin{aligned} y^\top D^{1/2} U U^\top D^{1/2} y &\leq \sum_i w_i y_i^2 + \rho \sum_{i \neq j} \sqrt{w_i w_j}|y_i y_j| \\ &= (1 - \rho) \sum_i w_i y_i^2 + \rho \left(\sum_i \sqrt{w_i}|y_i|\right)^2 \\ &\leq (1 - \rho) \max_i w_i + \rho \sum_i w_i. \end{aligned}$$

The nonzero spectra of $U^\top D U$ and $D^{1/2} U U^\top D^{1/2}$ agree, proving (133).

Fix unit $v \in \mathbb{R}^d$ and $z = N\xi \in \mathcal{Z}_A$, and set

$$w_i := c_i (a_i^\top v)^2, \qquad W := \sum_i w_i, \qquad w_\star := \max_i w_i.$$

Using (129) and then (133) with $\rho = \rho_\perp(A)$ gives

$$v^\top R_A(z)v = W - \xi^\top \left(\sum_i w_i u_i u_i^\top\right)\xi \geq (1-\rho_\perp(A))(W - w_\star). \tag{134}$$

Choose $k_\star$ with $w_{k_\star} = w_\star$. Then

$$W - w_\star = v^\top S_{k_\star} v \geq \lambda_{\min}(S_{k_\star}) \geq \kappa_\perp(A). \tag{135}$$

These bounds hold for every unit v and every $z \in \mathcal{Z}_A$, so

$$R_A(z) \succeq t_\perp(A) I_d \qquad (z \in \mathcal{Z}_A).$$

Minimizing over z proves $t_\perp(A) \leq \nu_A$, while $\rho_\perp(A) \in [0,1]$ and $\kappa_\perp(A) > 0$ give $t_\perp(A) \geq 0$. Together with (132), this proves the eigenvalue-free floor in (126); it discards the design-specific leave-one-out factor $\kappa_\perp(A)$. □

Covariance consequence. For $d < s < m$, the factors β_s and γ_s are positive. The linear bound in (8), together with $R_A(z) \succeq \nu_A I_d \succeq t_\perp I_d$ for every compatible z, gives (127) after maximizing over z and unit directions. The target is the centered, full-Gram-whitened spectral covariance, uniformly over compatible responses.

E Fixed-query contact and response-uniform query geometry

This section derives query and multi-output consequences from the saturated-row contact kernel proved once in Appendix C. It proves Corollaries 4, 24, and 7.

E.1 Query loss and the full contact kernel

Let $X \in \mathbb{R}^{m\times d}$ have full column rank, with $m \geq d$, and fix $Y \in \mathbb{R}^{m\times K}$, $Z \in \mathbb{R}^{q\times d}$, and $Y_Q \in \mathbb{R}^{q\times K}$, where $q, K \geq 1$. For each fixed $d \leq s \leq m$, draw one ordinary size-s volume subset and fit every output by unweighted OLS on that same positive-volume support. When $m = d$, this is just Full; $s = d < m$ uses supported square systems, and $s = m$ is exact Full. Write W_S and W_F for the selected and Full fits, $G = X^\top X$, $M_{s,j}$ for the coefficient covariance of output j, $\overline{M}_{s,j} = G^{1/2} M_{s,j} G^{1/2}$, and $\Xi = G^{-1/2} Z^\top Z G^{-1/2} \succeq 0$. With $L_{Q,S} = \|ZW_S - Y_Q\|_F^2/(qK)$ and $L_{Q,F} = \|ZW_F - Y_Q\|_F^2/(qK)$, expansion gives

$$\begin{aligned}\mathbb{E}_X L_{Q,S} - L_{Q,F} &= \frac{2\langle ZW_F - Y_Q, Z\mathbb{E}_X(W_S - W_F)\rangle_F + \mathbb{E}_X\|Z(W_S - W_F)\|_F^2}{qK} \\ &= \frac{1}{qK}\sum_{j=1}^K \operatorname{tr}(ZM_{s,j}Z^\top) = \frac{1}{qK}\sum_{j=1}^K \operatorname{tr}(\Xi\overline{M}_{s,j}) = \mathcal{D}_s.\end{aligned} \tag{136}$$

Selected-OLS unbiasedness cancels the cross term; the Frobenius column sum needs no independence between outputs or query rows. This identity needs neither no coloops, positive residual loss, nor a nonzero query metric. For $K = 1$, multiplying by q recovers the single-output SSE identity (5).

For the structural argument below, $A \in \mathbb{R}^{m\times d}$ has orthonormal columns, $d \geq 1$, $m \geq d+2$, $\ell_i < 1$, $d < s < m$, and $L^* > 0$. Only normalized query targets additionally require $W_Q := \operatorname{tr}\Xi > 0$.

Fix $z \in \mathcal{Z}_A$ and write $C_s(z) = \overline{M}_s^{A,z}/L^*$ and $\Delta_s(z) = \alpha_s I_d - C_s(z) \succeq 0$. By Lemma 23, $\ker\Delta_s(z) = \ker R_A(z)$ for the same fixed residual at every strict-interior size.

Finally,

$$\alpha_s L^* \operatorname{tr}\Xi - \operatorname{tr}(\Xi\overline{M}_s^{A,z}) = L^*\|\Delta_s(z)^{1/2}\Xi^{1/2}\|_F^2.$$

It vanishes exactly when $\operatorname{range}\Xi \subseteq \ker R_A(z)$.

E.2 Complement classes, dimensions, and query rank

Use the normalized complement rows and class spaces from Appendix C.4. For a residual saturating a class C, its identity $\ker R_A(z) = V_C$, together with the fixed-query kernel criterion above, proves the necessary-and-sufficient range condition in Corollary 4.

For a rank-r query, one class needs at least $r+1$ rows and must contain the query range in its deletion nullspace. For $\Xi \succ 0$, contact requires $V_C = \mathbb{R}^d$, hence $|C| = d+1$ and every row outside C is zero. A zero feature row has a complement direction orthogonal to all other complement rows, so cannot belong to such a nontrivial class. Conversely, a rank-d no-coloop core with $d+1$ nonzero rows has a unique dependence with no zero coordinate; its unit dependence extended by zeros saturates the core and gives $R_A(z) = 0$. This proves the full-rank classification. When all feature rows are nonzero, compactness of the residual sphere gives a positive response-uniform trace gap on that fixed design.

E.3 Multi-output strictness

Corollary 24 (Multi-output residual-rank strictness)**.** *Let $E = [e_1, \dots, e_K]$ be the exact Full residual matrix, let $L_j = \|e_j\|_2^2$, and let $L_{\mathrm{src}} = \sum_j L_j$. On the real Parseval, no-coloop structural domain with $m \geq d+2$ and $d < s < m$,*

$$\operatorname{rank}(E) \geq 2 \quad \Longrightarrow \quad \sum_{j=1}^{K} \overline{M}_{s,j} \prec \alpha_s L_{\mathrm{src}} I_d. \tag{137}$$

Proof. For $E \neq 0$, Proposition 32 gives (179). If $\operatorname{rank}(E) \geq 2$, then $\zeta_E > 0$; also $\vartheta_s, c_{\min} > 0$ on the stated domain. The right-hand side of that bound is positive definite, proving (137). Its trace against every nonzero $\Xi \succeq 0$ is positive. The converse fails: a rank-one residual may saturate a nontrivial complement class, while one with no saturated class has strict covariance deficit by the contact-kernel identity. □

E.4 The query residual margin and a sufficient envelope

The trace-minimum definition (21) has the equivalent complement-compression expression

$$u_i = \frac{a_i^\top \Xi a_i}{W_Q}, \qquad t_Q = \sum_i c_i u_i - \lambda_{\max}\big(N^\top \operatorname{diag}(u_i) N\big)\,. \tag{138}$$

Here $u_i \geq 0$ and $\sum_i u_i = 1$; the normalized query metric assigns weight u_i to whitened source row i. Indeed, the residual-dependent trace is

$$\frac{\operatorname{tr}(\Xi R_A(N\xi))}{W_Q} = \sum_i c_i u_i - \xi^\top N^\top \operatorname{diag}(u_i) N \xi.$$

Minimizing over the unit sphere evaluates the main definition and proves (138), with $t_Q = \min_z \operatorname{tr}(\Xi R_A(z))/W_Q \geq \nu_A \geq 0$. The minimum is attained. Averaging over a uniform residual direction gives

$$t_Q \leq \frac{n-1}{n} \sum_i c_i u_i < \frac{n-1}{n},$$

because $\sum_i \ell_i u_i > 0$. Thus the scalar denominator and the first-pass parameter $\delta = 1 - n t_Q/(n-1)$ are positive.

Diagonalize $R = \sum_k r_k v_k v_k^\top$ and use weights $\omega_k = v_k^\top \Xi v_k / W_Q \geq 0$, $\sum_k \omega_k = 1$. Scalar Jensen yields

$$\frac{\operatorname{tr}[\Xi(I - \gamma_s R)^{-1}]}{W_Q} = \sum_k \frac{\omega_k}{1 - \gamma_s r_k} \geq \frac{1}{1 - \gamma_s \operatorname{tr}(\Xi R)/W_Q}. \tag{139}$$

Apply the one-sided matrix resolvent and minimize its trace margin to obtain (22). No commutation assumption on Ξ and R is needed. In particular,

$$\alpha_s - \mathsf{T}_Q^{\mathrm{RU}}(A, s) \geq \frac{\beta_s \gamma_s t_Q}{1 - \gamma_s t_Q}.$$

The attained minimum has $t_Q = 0$ precisely when some residual has zero query trace slack in $R_A(z)$; Theorem 3 then gives exact contact with α_s. This proves both directions of the phase equivalence.

For a validated $0 \le t \le t_Q$, the algebra in Appendix H.2 applies with $\delta = 1 - nt/(n-1) > 0$. At normalized query tolerance $0 < \tau < 1$, the first sufficient size in $\{d+1, \dots, m\}$ is (194). Here normalization is by $L^* W_Q$ for single-output query SSE, with the corresponding q or qK factor for MSE. This is an upper-envelope pass. The trace margin need not lower-bound $R_A(z)$ in Loewner order.

For a realized multi-output response, the weighted average $\operatorname{tr}(\Xi \bar{R})/W_Q$ is at least t_Q. Hence

$$U_{\mathrm{agg}}(s) \le \frac{L_{\mathrm{src}} W_Q}{qK} B_{\mathrm{res}}\left(s; \frac{\operatorname{tr}(\Xi \bar{R})}{W_Q}\right) \le \frac{L_{\mathrm{src}} W_Q}{qK} B_{\mathrm{res}}(s; t_Q).$$

Thus the response-aware aggregate retains more information, while t_Q supplies a guarantee before source responses are known.

A lower endpoint without a complement eigensolve. Since $N^\top \operatorname{diag}(u_i) N \preceq (\max_i u_i) I$,

$$\max\left\{0, 1 - \sum_i \ell_i u_i - \max_i u_i\right\} \le t_Q. \tag{140}$$

The unit-trace PSD matrix Ξ/W_Q has operator norm at most one, so $u_i \le \ell_i$. Equation (140) is consequently at least $\max\{0, 1 - 2L_{\max}\}$. It can be combined by a maximum with other validated lower endpoints of t_Q, including any lower bound on ν_A. On the four frozen designs, the directed version of this cheap endpoint is selected above the feature-only endpoint, without constructing N; this observed numerical order is not assumed for arbitrary independently validated endpoints. If $\Xi = I_d$ and $\ell_i = d/m$ for every row, then $u_i = 1/m$ and the compression is exactly I_n/m, giving $t_Q = (m - d - 1)/m$.

E.5 A COMMON LEAVE-ONE-OUT IDENTITY

Let $A^\top A = I_d$, $c_i = 1 - \|a_i\|^2 > 0$, $n = m - d \ge 1$, and $s = m - 1$. Fix $z \in \ker A^\top$ with $\|z\| = 1$ and $y = A\theta + \sqrt{L^*} z$, $L^* > 0$. Lemma 10 gives $\mathbb{P}_A(S = -i) = c_i/n$, while the OLS deletion identity (Dereziński & Warmuth, 2017, Lemma 9), or Sherman–Morrison, gives

$$\frac{w_{-i} - \theta}{\sqrt{L^*}} = -z_i (I_d - a_i a_i^\top)^{-1} a_i = -\frac{z_i}{c_i} a_i.$$

Summing on this support and using $\sum_i a_i a_i^\top = I_d$ proves

$$\frac{\overline{M}_{m-1}^{A,z}}{L^*} = \frac{1}{n} \sum_i \frac{z_i^2}{c_i} a_i a_i^\top, \qquad n\left(\alpha_{m-1} I_d - \frac{\overline{M}_{m-1}^{A,z}}{L^*}\right) = \sum_i \frac{c_i - z_i^2}{c_i} a_i a_i^\top. \tag{141}$$

These are pointwise identities in z, with $\alpha_{m-1} = 1/n$. The case $n = 1$ is rank-size LOO; the strict-interior consequences below retain $n \ge 2$. No inverse or quotient by c_i is used when a coloop is present.

E.6 EXACT LEAVE-ONE-OUT QUERY RISK AND BOUNDARY EXAMPLES

Proposition 25 (Exact leave-one-out query risk)**.** *Let $A \in \mathbb{R}^{m \times d}$ have orthonormal columns and no coloops, let $m \ge d + 2$, $s = m - 1$, and let $\Xi \succeq 0$ have $W_Q = \operatorname{tr} \Xi > 0$. Then the response-uniform fixed-query target is*

$$\mathsf{T}_Q^{\mathrm{RU}}(A, m-1) = \frac{1}{n} \lambda_{\max}\left[N^\top \operatorname{diag}\left(\frac{u_i}{c_i}\right) N\right]. \tag{142}$$

If all $c_i = c$, its exact gap is $\alpha_{m-1} - \mathsf{T}_Q^{\mathrm{RU}} = t_Q/(nc)$. In particular, for equal leverage and every compatible positive-loss residual, $\operatorname{tr}(\overline{M}_{m-1})/L^* = d/(m-d)^2$.

Proof. The pointwise identity (141) gives $\mathsf{T}_Q^{\mathrm{RU}}(A, m-1) = \alpha_{m-1}(1 - \tau_Q(A, \Xi))$; the independent eigenvalue evaluation (160) yields (142). If $c_i = c$, then $t_Q = c - \lambda_{\max}(N^\top \operatorname{diag}(u_i) N)$, so direct substitution gives the stated gap. Under equal leverage, $\ell_i = d/m$, $c_i = n/m$, and $\sum_i z_i^2 = 1$; hence (141) gives $\operatorname{tr}(\overline{M}_{m-1})/L^* = (nc_i)^{-1} \sum_i z_i^2 \ell_i = d/n^2$. □

Duplicated-basis enumeration. For the control used in Section 2.2,

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} I_2 \\ I_2 \end{bmatrix}, \qquad (m, d, s) = (4, 2, 3), \qquad \Xi = I_2. \tag{143}$$

Every compatible unit residual is $z = 2^{-1/2}(u, -u)^\top$ for a unit $u = (u_1, u_2)^\top$. Enumeration of the supported size-three subsets gives

$$\frac{\overline{M}_3^{A,z}}{L^*} = \frac{1}{2}\operatorname{diag}(u_1^2, u_2^2), \tag{144}$$

so the unnormalized trace ratio is $1/2$ although simultaneous isotropic contact is impossible.

The core-plus-zero family (56) has full matrix contact at $(d, m, s) = (2, 4, 3)$; a repeated-coordinate extension has rank-two query contact at $(3, 5, 4)$. By contrast, the duplicated-basis design above has spectral target $1/2$ but isotropic normalized trace target $1/4$.

The assumptions matter: the coloop obstruction is (123). At rank size, the row-general-position design $X = [(1, 0); (0, 1); (1, 1); (1, 2)]$ and $y = (-1, -1, 1, 0)^\top$ have $\overline{M}_d/L^* = I_2$ although $R_A(z) \succ 0$. Full, zero loss, and the zero query metric have zero unnormalized excess; $m = d + 1$ has no strict-interior size.

F ALL-SIZE COMPARISON WITH THE LEAVE-ONE-OUT DEFICIT

This appendix proves Theorem 6 under the ordinary feature-only law, then treats approximate alignment and unknown responses. The comparison is independent of contact-class classification and of support-changing limits.

F.1 SUPPORTED-LAW PROOF OF THE MATRIX COMPARISON

Let $A \in \mathbb{R}^{m \times d}$ have orthonormal columns, $d \geq 1$, $n := m - d \geq 2$, and no coloop. Thus $c_i = 1 - \|a_i\|_2^2 > 0$. Fix $z \in \ker A^\top$ with $\|z\|_2 = 1$. For an integer $d < s < m$, put

$$k := m - s, \qquad \alpha_s := \frac{k}{n}, \qquad t_s := \frac{k-1}{n-1}, \qquad \eta_s := \frac{k(k-1)}{n(n-1)}.$$

Here $1 \leq k \leq n - 1$, so every displayed denominator is positive. The letter k counts omitted rows here, while K elsewhere counts output columns. On a supported set S, define

$$p_A(S) = \frac{\det(A_S^\top A_S)}{\binom{n}{s-d}}, \qquad v_S = (A_S^\top A_S)^{-1} A_S^\top z_S.$$

The corresponding selected fit for the source response $A\theta + \sqrt{L^*}z$ obeys $v_S = G^{1/2}(w_S - w^*)/\sqrt{L^*}$. The ordinary-law mean is zero, so

$$\mathbb{E}_A[v_S v_S^\top] = \frac{\overline{M}_s^{A,z}}{L^*}. \tag{145}$$

No fit is assigned to a zero-volume set.

Omission moments. Put $P := I_m - AA^\top$, so $P_{ii} = c_i$, and define

$$f_i := -\frac{z_i}{c_i} a_i, \qquad L(z) := \sum_i \frac{z_i^2}{c_i} a_i a_i^\top, \qquad F_S := \sum_{i \notin S} f_i. \tag{146}$$

Deleting row i leaves the positive definite Gram matrix $I_d - a_i a_i^\top$. Its full-pool least-squares fit for z_{-i} is f_i: the normal equations have right side $-z_i a_i$ and $(I_d - a_i a_i^\top)^{-1} a_i = a_i/c_i$. The indexed-volume normalizer on that deleted pool is $\binom{n-1}{s-d} c_i$. Thus the original law, conditioned on $i \notin S$, is exactly ordinary size-s volume sampling from the deleted pool. The boundary-valid ordinary-law unbiasedness and singleton exclusion identity give

$$\mathbb{E}_A[v_S \mid i \notin S] = f_i, \qquad \mathbb{P}_A(i \notin S) = \alpha_s c_i. \tag{147}$$

Consequently, for supported fits, $\mathbb{E}_A[v_S F_S^\top] = \alpha_s L(z)$.

For distinct i, j and $k \geq 2$, fixed-size Cauchy–Binet applied after deleting both rows gives

$$\mathbb{P}_A(i, j \notin S) = \eta_s \det(P_{\{i,j\},\{i,j\}}) = \eta_s(c_i c_j - P_{ij}^2). \tag{148}$$

It also holds at $k = 1$, when both sides are zero because two rows cannot be omitted. If the doubly deleted pool loses rank, its determinant and pair-exclusion probability are both zero. Define

$$H(z) := A^\top \operatorname{diag}(z_i/c_i)(P \circ P) \operatorname{diag}(z_i/c_i) A \succeq 0, \tag{149}$$

where $\circ$ is the entrywise product. The Schur product theorem makes $P \circ P$ positive semidefinite. Since $\sum_i c_i f_i = -A^\top z = 0$, expansion of the singleton and pair terms gives

$$\mathbb{E}_A[v_S F_S^\top] = \alpha_s L(z), \qquad \mathbb{E}_A[F_S F_S^\top] = \alpha_s L(z) - \eta_s H(z). \tag{150}$$

For clarity, the $c_i c_j$ part of the off-diagonal pair sum equals $-\sum_i c_i^2 f_i f_i^\top$; adding the P_{ij}^2 part and its diagonal gives $-H(z)$. At $k = 1$, the pair sum and η_s are zero.

Expanding the positive semidefinite square of $v_S - F_S$ now yields the exact supported-law decomposition

$$\frac{\overline{M}_s^{A,z}}{L^*} = \alpha_s L(z) + \eta_s H(z) + \mathbb{E}_A[(v_S - F_S)(v_S - F_S)^\top] \succeq \alpha_s L(z). \tag{151}$$

Since $A^\top A = I_d$, $D_{\mathrm{LOO}}(z) = I_d - L(z)$. Dividing (151) by α_s proves the right-hand inequality in (17).

Reverse deletion and total covariance. Choose the first omitted row $I = i$ with probability c_i/n. Conditional on $I = i$, draw S by ordinary size-s volume sampling from the deleted pool. Its mixture law is p_A: every supported S has k possible first omitted rows, and

$$\frac{k}{n\binom{n-1}{s-d}} = \frac{1}{\binom{n}{s-d}}.$$

Conditional unbiasedness gives $\mathbb{E}[v_S \mid I = i] = f_i$. The deleted pool's full-fit residual energy and inverse Gram are

$$e_i := 1 - \frac{z_i^2}{c_i} \geq 0, \qquad (I_d - a_i a_i^\top)^{-1} = I_d + \frac{a_i a_i^\top}{c_i}. \tag{152}$$

For $k \geq 2$, the universal covariance envelope in this full-rank pool applies to the conditional sample and gives

$$\operatorname{Cov}(v_S \mid I = i) \preceq t_s e_i \left(I_d + \frac{a_i a_i^\top}{c_i} \right). \tag{153}$$

If $e_i = 0$, every supported selected fit equals f_i, so its conditional covariance is zero. If $k = 1$, the conditional sample is the entire deleted pool and its covariance is again zero; thus (153) holds with $t_s = 0$ without applying an interior formula at the endpoint.

The mixture mean is $\sum_i (c_i/n) f_i = 0$. Also $\sum_i c_i e_i = n - 1$ and $\sum_i e_i a_i a_i^\top = I_d - L(z)$. Total covariance therefore gives

$$\begin{aligned} \frac{\overline{M}_s^{A,z}}{L^*} &\preceq \frac{L(z)}{n} + \frac{t_s}{n} \sum_i c_i e_i \left(I_d + \frac{a_i a_i^\top}{c_i} \right) \\ &= t_s I_d + \frac{n-k}{n(n-1)} L(z). \end{aligned} \tag{154}$$

Subtracting this expression from $\alpha_s I_d = (k/n) I_d$ gives

$$I_d - \frac{\overline{M}_s^{A,z}}{\alpha_s L^*} \succeq \frac{n-k}{k(n-1)} (I_d - L(z)) = b_s D_{\mathrm{LOO}}(z).$$

Together with the omission remainder, this proves the matrix comparison. Taking the trace against $\Xi \succeq 0$ proves the scalar comparison. When $k = 1$, $F_S = f_I = v_S$ and $\eta_s = t_s = 0$, so $\overline{M}_{m-1}^{A,z}/L^* = L(z)/n$ exactly. Hence

$$I_d - \frac{\overline{M}_{m-1}^{A,z}}{\alpha_{m-1} L^*} = D_{\mathrm{LOO}}(z), \qquad \alpha_{m-1} = 1/n. \tag{155}$$

The proof uses neither augmented-law inverse moments nor row general position; all moments stay on ordinary support.

The comparison also gives the fixed-residual contact kernel without using the class classification. Indeed, $D_{\mathrm{LOO}}(z) = \sum_i e_i a_i a_i^\top$ and $R_A(z) = \sum_i c_i e_i a_i a_i^\top$. All coefficients c_i are positive. Their kernels therefore coincide, and the two positive comparison factors imply

$$\ker\left(I_d - \frac{\overline{M}_s^{A,z}}{\alpha_s L^*}\right) = \ker D_{\mathrm{LOO}}(z) = \ker R_A(z). \tag{156}$$

The complement-class classification still requires the separate argument in Appendix C.4.

F.2 Angular concentration and approximate deletion spaces

Choose an orthonormal complement $N \in \mathbb{R}^{m\times n}$ with rows $n_i^\top$, and set $\widehat{n}_i = n_i/\sqrt{c_i}$. Write $z = N\xi$, $\|\xi\|_2 = 1$. For a nonzero query metric define

$$u_i := \frac{a_i^\top \Xi a_i}{\operatorname{tr}\Xi}, \qquad e_i = 1 - |\widehat{n}_i^\top \xi|^2.$$

Then $\sum_i u_i = 1$ and $\delta_{\mathrm{LOO}} = \sum_i u_i e_i$, as in (19). For $0 < \eta < 1$, let $C_\eta = \{i : e_i \le \eta\}$ and $E_\eta = \sum_{i\notin C_\eta} u_i$. Since $\eta E_\eta \le \delta_{\mathrm{LOO}} \le \eta + (1-\eta)E_\eta$, Theorem 6 gives

$$b_s \eta E_\eta \le g_s^{(1)}(z,\Xi) \le \eta + (1-\eta)E_\eta. \tag{157}$$

Thus a small deficit forces query row weight into a narrow angular cap, with loss $1/b_s$ in the reverse direction. Conversely, concentrated weight in a narrow cap gives a small deficit at every interior size.

The cap also has a low-dimensional approximate deletion space. The next result records the precise range where this follows from spectral separation. We use r_C for the number of cap rows, not query rank.

Proposition 26 (Approximate deletion space of a narrow cap)**.** *Fix a nonzero $\Xi \succeq 0$ and $0 \le \eta < 1/2$. Suppose $C = C_\eta$ is nonempty. Set*

$$r_C = |C|, \qquad \rho = \|z_C\|_2^2, \qquad \tau = \sum_{i\in C}(c_i - z_i^2), \qquad T_C = A_{-C}^\top A_{-C}.$$

Then $\rho > 0$, $r_C \le d+1$, and $0 \le \tau \le \eta\rho/(1-\eta) < \rho$. The bottom $r_C - 1$ eigenvalues of T_C are at most τ. If $r_C \le d$, every remaining eigenvalue is at least ρ. Let V_C^{app} be the bottom $r_C - 1$ eigenspace, with $V_C^{\mathrm{app}} = \{0\}$ when $r_C = 1$ and $V_C^{\mathrm{app}} = \mathbb{R}^d$ when $r_C = d+1$. Then

$$\frac{\operatorname{tr}[\Xi(I_d - \Pi_{V_C^{\mathrm{app}}})]}{\operatorname{tr}\Xi} \le \frac{E_\eta}{\rho}. \tag{158}$$

For a rank-r query, let λ_r be the least positive eigenvalue of $\Xi/\operatorname{tr}\Xi$. If $r_C \le r$, then

$$E_\eta \ge \rho(r - r_C + 1)\lambda_r. \tag{159}$$

Proof. The matrix $J := I_m - AA^\top - zz^\top$ is an orthogonal projector. Its $C\times C$ principal block has trace τ, since $J_{ii} = c_i - z_i^2$. On C, $z_i^2 = (1-e_i)c_i \ge (1-\eta)c_i > 0$, so $\rho > 0$ and $\tau = \sum_{i\in C} e_i c_i \le \eta\rho/(1-\eta) < \rho \le 1$. If $r_C > d+1$, a unit left-null vector of $[A\ z]_C$ is an eigenvector of $J_{C,C}$ with eigenvalue one. This contradicts $\operatorname{tr} J_{C,C} = \tau < 1$.

Because $A_C A_C^\top = I_C - z_C z_C^\top - J_{C,C}$ and $0 \preceq J_{C,C} \preceq \tau I_C$, its Rayleigh quotient on $z_C^\perp$ is at least $1-\tau$. Min–max gives $r_C - 1$ eigenvalues of $A_C A_C^\top$ at least $1-\tau$. On the other hand, $A_C A_C^\top \preceq I_C - z_C z_C^\top$, whose last eigenvalue is $1-\rho$. Transfer the nonzero eigenvalues to $A_C^\top A_C$ and use $T_C = I_d - A_C^\top A_C$. The stated two blocks follow. The conventions for $r_C = 1$ and $r_C = d+1$ avoid nonexistent eigenvalue indices.

In particular, $T_C \succeq \rho(I_d - \Pi_{V_C^{\mathrm{app}}})$ when the transverse space is nonzero. Since $\operatorname{tr}[(\Xi/\operatorname{tr}\Xi)T_C] = \sum_{i\notin C} u_i = E_\eta$, this proves (158); it is trivial when $V_C^{\mathrm{app}} = \mathbb{R}^d$. If $\Xi/\operatorname{tr}\Xi$ has rank r, it dominates λ_r times the projector onto its range. The trace of that projector outside an (r_C-1)-dimensional space is at least $r - r_C + 1$. Combining this with the previous inequality proves (159). □

At $\eta = 0$, the deletion space is exactly $\ker A_{-C}$, as in the contact classification. The approximate result requires $\eta < 1/2$ to separate the two spectral blocks. It does not assert that the design is close to a design with exact contact.

F.3 Response-uniform LOO margin

For fixed A and nonzero Ξ, define $\tau_Q(A, \Xi) := \min_{z \in \mathcal{Z}_A} \delta_{\mathrm{LOO}}(z, \Xi)$. By the pointwise identity (141), this also equals $\min_z g^{(1)}_{m-1}(z, \Xi)$; compactness gives attainment. Since $\delta_{\mathrm{LOO}}(N\xi, \Xi) = 1 - \xi^\top N^\top \operatorname{diag}(u_i/c_i) N\xi$,

$$\tau_Q(A, \Xi) = 1 - \lambda_{\max}\left[N^\top \operatorname{diag}(u_i/c_i) N\right]. \tag{160}$$

Taking minima over the same unit residual sphere in (18) gives

$$b_s \tau_Q(A, \Xi) \le 1 - \frac{\mathsf{T}^{\mathrm{RU}}_Q(A, s)}{\alpha_s} \le \tau_Q(A, \Xi). \tag{161}$$

A top eigenvector in (160) supplies one residual whose deficit is at most τ_Q at every interior size. It need not optimize the risk at every size. The minimized margin is an existential statement; it does not upper-bound the deficit of an arbitrary observed residual.

G Quantitative stability around complement contact classes

G.1 Domain and baseline premises

Fix A, s on the structural domain of Theorem 3, with its ordinary unrescaled indexed volume law and selected unweighted OLS, and fix $L^* > 0$. Retain $c_i > 0$, the complement basis N, its normalized rows $\widehat{n}_i$, and the residual sphere $\mathcal{Z}_A = \{N\xi : \|\xi\|_2 = 1\}$ from Appendix C.4. Retain $R_A(z)$ and the interior-size factors $\alpha_s, \beta_s, \gamma_s$ from (109) and (110). Put $c_{\min} = \min_i c_i$, $c_{\max} = \max_i c_i$, and $\vartheta_s = \beta_s\gamma_s/\alpha_s = (s-d)/(m-d-1) > 0$.

For $z = N\xi \in \mathcal{Z}_A$, define the covariance deficit

$$D^{\mathrm{cov}}_s(z) := \alpha_s I_d - \frac{\overline{M}^{A,z}_s}{L^*}. \tag{162}$$

Thus $D^{\mathrm{cov}}_s/\alpha_s$ is the normalized matrix ceiling deficit. For a nonzero PSD query metric Ξ, write $W_Q = \operatorname{tr}\Xi > 0$ and

$$g^{(1)}_s(z, \Xi) := \frac{\operatorname{tr}[\Xi D^{\mathrm{cov}}_s(z)]}{\alpha_s W_Q}. \tag{163}$$

This is exactly the main-text normalized deficit (14); the superscript distinguishes it from the aggregate deficit below.

The envelope, one-sided residual contraction, and interior contact kernel give

$$0 \preceq \frac{\overline{M}^{A,z}_s}{L^*} \preceq \alpha_s I_d, \qquad D^{\mathrm{cov}}_s(z) \succeq \beta_s\gamma_s R_A(z), \qquad \ker D^{\mathrm{cov}}_s(z) = \ker R_A(z). \tag{164}$$

These are Theorems 1 and 2 and Theorem 3. The complement coordinates give

$$R_A(N\xi) = \sum_{i=1}^m c_i\big(1 - |\widehat{n}_i^\top \xi|^2\big) a_i a_i^\top. \tag{165}$$

G.2 A fixed ordinary-law quadratic form

Lemma 27 (Ordinary-law residual quadratic form)**.** *Fix A, s and a nonzero $\Xi \succeq 0$. There is a matrix $\mathcal{K}_{s,\Xi} \in \mathbb{R}^{(m-d)\times(m-d)}$ such that*

$$0 \preceq \mathcal{K}_{s,\Xi} \preceq I_{m-d}, \qquad g^{(1)}_s(N\xi, \Xi) = \xi^\top \mathcal{K}_{s,\Xi}\xi \quad (\|\xi\|_2 = 1). \tag{166}$$

The sampling support in this representation is the fixed feature-only ordinary-volume support.

Proof. For an ordinary-supported s-set S, put

$$K_S = A_S^\top A_S, \qquad T_S = K_S^{-1} A_S^\top N_S, \qquad p_A(S) = \frac{\det K_S}{\binom{m-d}{s-d}}.$$

For the response $A\theta + \sqrt{L^*}N\xi$, selected least squares gives $(w_S - \theta)/\sqrt{L^*} = T_S\xi$. Hence the finite supported sum under the ordinary feature-only law gives

$$\frac{\operatorname{tr}(\Xi \overline{M}_s^{A,N\xi})}{L^* W_Q} = \xi^\top H_{s,\Xi}\xi, \qquad H_{s,\Xi} := \frac{1}{W_Q} \sum_{S:\det K_S>0} p_A(S) T_S^\top \Xi T_S \succeq 0.$$

The universal covariance ceiling, applied to every unit ξ, implies $H_{s,\Xi} \preceq \alpha_s I_{m-d}$. Therefore

$$\mathcal{K}_{s,\Xi} := I_{m-d} - H_{s,\Xi}/\alpha_s$$

is a positive contraction, and substitution in (163) proves (166). This argument never invokes the response-dependent augmented law or an inverse outside ordinary support. □

G.3 Joint residual–query stability

Fix a maximal projective class C of the normalized complement rows. Choose a unit representative $\widehat{n}_C$ and define

$$z_C := N\widehat{n}_C, \qquad V_C := \ker A_{-C}, \qquad \Pi_C := \operatorname{proj}_{V_C}.$$

Assume throughout this subsection that $\{0\} \neq V_C \neq \mathbb{R}^d$. In particular, $-C$ is nonempty. Define

$$\begin{aligned} \delta_C &:= \min_{i \notin C} \big(1 - |\widehat{n}_i^\top \widehat{n}_C|^2\big), \\ \sigma_C^2 &:= \min\big\{v^\top A_{-C}^\top A_{-C} v : v \in V_C^\perp,\ \|v\|_2 = 1\big\}, \\ \kappa_C &:= \frac{c_{\min}\delta_C \sigma_C^2}{4}. \end{aligned} \tag{167}$$

For $z = N\xi \in \mathcal{Z}_A$, set

$$\varepsilon_z := 1 - (z^\top z_C)^2 = 1 - (\xi^\top \widehat{n}_C)^2, \qquad \varepsilon_\Xi := \frac{\operatorname{tr}[\Xi(I_d - \Pi_C)]}{W_Q}. \tag{168}$$

Changing the signs of $\widehat{n}_C, z_C$ leaves ε_z, $R_A(z_C)$, and the covariance unchanged. In (167), σ_C^2 is the transverse minimum over $V_C^\perp$. The following fixed-design, proper-class result holds with constants independent of subset size.

Theorem 28 (Joint local residual–query alignment)**.** *For one fixed A, C, z_C, z, and Ξ, simultaneously for every $s \in \{d+1, \dots, m-1\}$, if*

$$\varepsilon_z \le \frac{\delta_C}{16}, \qquad \varepsilon_\Xi \le \frac{1}{2},$$

then

$$\vartheta_s \kappa_C (\varepsilon_z + \varepsilon_\Xi) \le g_s^{(1)}(z, \Xi) \le \min\left\{1, (\sqrt{\varepsilon_z} + \sqrt{\varepsilon_\Xi})^2\right\}. \tag{169}$$

The upper bound holds for every $z \in \mathcal{Z}_A$ and every nonzero $\Xi \succeq 0$, without either neighborhood restriction, on the same fixed design and proper-class domain.

Proof. The constants are positive. No coloops give $c_{\min} > 0$; maximality and finiteness give $\delta_C > 0$; and $\ker(A_{-C}^\top A_{-C}) = V_C$ gives $\sigma_C^2 > 0$ on $V_C^\perp$. Moreover, $c_{\min}, \delta_C, \sigma_C^2 \le 1$.

Orient $\widehat{n}_C$ so that $\xi^\top \widehat{n}_C \ge 0$ and write

$$\xi = \sqrt{1 - \varepsilon_z}\, \widehat{n}_C + \sqrt{\varepsilon_z}\, \eta, \qquad \eta \perp \widehat{n}_C, \quad \|\eta\|_2 = 1,$$

with the evident interpretation at $\varepsilon_z = 0$. If $i \in C$, then $\widehat{n}_i = \pm \widehat{n}_C$ and

$$1 - |\widehat{n}_i^\top \xi|^2 = \varepsilon_z. \tag{170}$$

If $i \notin C$, let $P_i = I - \widehat{n}_i \widehat{n}_i^\top$. The reverse triangle inequality gives

$$\begin{aligned}\sqrt{1 - |\widehat{n}_i^\top \xi|^2} &= \|P_i \xi\|_2 \\ &\geq \sqrt{1-\varepsilon_z}\, \|P_i \widehat{n}_C\|_2 - \sqrt{\varepsilon_z}\, \|P_i \eta\|_2 \\ &\geq \sqrt{1-\varepsilon_z}\sqrt{\delta_C} - \sqrt{\varepsilon_z} \geq \frac{\sqrt{\delta_C}}{2}.\end{aligned}$$

Indeed, $\varepsilon_z \leq \delta_C/16 \leq 1/16$ implies $\sqrt{1-\varepsilon_z} \geq 3/4$ and $\sqrt{\varepsilon_z} \leq \sqrt{\delta_C}/4$. Thus (165) yields

$$R_A(z) \succeq c_{\min} \left[\varepsilon_z A_C^\top A_C + \frac{\delta_C}{4} A_{-C}^\top A_{-C} \right]. \tag{171}$$

Let $H_C = A_{-C}^\top A_{-C}$. Since $0 \preceq H_C \preceq I$, $\ker H_C = V_C$, and $A_C^\top A_C = I - H_C$, the invariant orthogonal decomposition $V_C \oplus V_C^\perp$ gives

$$A_C^\top A_C \succeq \Pi_C, \qquad A_{-C}^\top A_{-C} \succeq \sigma_C^2 (I_d - \Pi_C).$$

Taking the Ξ-trace of (171), applying (164), and dividing by $\alpha_s W_Q$ gives the stronger proof-level estimate

$$g_s^{(1)}(z, \Xi) \geq \vartheta_s c_{\min} \left[\varepsilon_z (1 - \varepsilon_\Xi) + \frac{\delta_C \sigma_C^2}{4} \varepsilon_\Xi \right]. \tag{172}$$

Set $\lambda_C = \delta_C \sigma_C^2/4 \leq 1/4$. Since $\varepsilon_\Xi \leq 1/2$,

$$\varepsilon_z(1-\varepsilon_\Xi) + \lambda_C \varepsilon_\Xi - \lambda_C(\varepsilon_z + \varepsilon_\Xi) = \varepsilon_z(1 - \varepsilon_\Xi - \lambda_C) \geq 0.$$

This proves the frozen lower bound with $\kappa_C = c_{\min} \lambda_C$.

For the upper bound, exact contact at z_C gives $\ker D_s^{\mathrm{cov}}(z_C) = V_C$. Since $0 \preceq D_s^{\mathrm{cov}}(z_C) \preceq \alpha_s I_d$, its V_C block and cross blocks vanish, so

$$0 \preceq D_s^{\mathrm{cov}}(z_C) \preceq \alpha_s (I_d - \Pi_C), \qquad g_s^{(1)}(z_C, \Xi) \leq \varepsilon_\Xi. \tag{173}$$

Let $\mathcal{K} = \mathcal{K}_{s,\Xi}$ be the positive contraction from Lemma 27. Put $\rho = \xi^\top \widehat{n}_C$ and, when $\varepsilon_z > 0$, $\eta = (\xi - \rho \widehat{n}_C)/\sqrt{\varepsilon_z}$. Then

$$\begin{aligned}\sqrt{g_s^{(1)}(z, \Xi)} &= \|\mathcal{K}^{1/2} \xi\|_2 \\ &\leq |\rho|\, \|\mathcal{K}^{1/2} \widehat{n}_C\|_2 + \sqrt{\varepsilon_z}\, \|\mathcal{K}^{1/2} \eta\|_2 \\ &\leq \sqrt{g_s^{(1)}(z_C, \Xi)} + \sqrt{\varepsilon_z} \leq \sqrt{\varepsilon_\Xi} + \sqrt{\varepsilon_z}.\end{aligned}$$

The case $\varepsilon_z = 0$ is immediate. Squaring proves the second upper endpoint, while $0 \preceq \mathcal{K} \preceq I$ gives $g_s^{(1)} \leq 1$. Neither step used a neighborhood restriction. The geometric constants are independent of s, proving the simultaneous size quantifier. □

Corollary 29 (Query-only slice)**.** *At $z = z_C$, let*

$$r_C := \min\{v^\top R_A(z_C) v : v \in V_C^\perp,\ \|v\|_2 = 1\} > 0.$$

Then, for every nonzero $\Xi \succeq 0$ and every strict-interior size,

$$\vartheta_s r_C \varepsilon_\Xi \leq g_s^{(1)}(z_C, \Xi) \leq \varepsilon_\Xi. \tag{174}$$

Proof. The contact-kernel identity gives $\ker R_A(z_C) = \ker D_s^{\mathrm{cov}}(z_C) = V_C$. Hence $R_A(z_C) \succeq r_C(I_d - \Pi_C)$, while the PSD-kernel argument in (173) gives $D_s^{\mathrm{cov}}(z_C) \preceq \alpha_s(I_d - \Pi_C)$. Combine these relations with $D_s^{\mathrm{cov}}(z_C) \succeq \beta_s \gamma_s R_A(z_C)$, take the Ξ-trace, and divide by $\alpha_s W_Q$. The same comparison also gives $r_C \geq c_{\min} \delta_C \sigma_C^2 = 4\kappa_C$. □

Corollary 30 (Residual-only slice)**.** *Suppose* $\operatorname{range} \Xi \subseteq V_C$*, and define*

$$u_i := \frac{a_i^\top \Xi a_i}{W_Q}, \qquad \overline{c}_{\Xi, C} := \sum_{i \in C} c_i u_i.$$

Then $\sum_{i \in C} u_i = 1$, $\overline{c}_{\Xi,C} > 0$, and for every $z \in \mathcal{Z}_A$ and every strict-interior size,

$$\vartheta_s \overline{c}_{\Xi, C} \varepsilon_z \leq g_s^{(1)}(z, \Xi) \leq \varepsilon_z. \tag{175}$$

No neighborhood restriction is required.

Proof. For $i \notin C$, every vector in V_C is orthogonal to a_i, so PSD range containment gives $u_i = 0$. Since $A^\top A = I_d$, $\sum_i u_i = 1$ and thus $\sum_{i\in C} u_i = 1$. For $i \in C$, $\widehat{n}_i = \pm\widehat{n}_C$, and (165) gives

$$\frac{\operatorname{tr}[\Xi R_A(z)]}{W_Q} = \overline{c}_{\Xi,C}\varepsilon_z.$$

The one-sided contraction proves the lower bound. For the upper bound, range containment and exact contact give $\widehat{n}_C^\top \mathcal{K}_{s,\Xi}\widehat{n}_C = g_s^{(1)}(z_C, \Xi) = 0$, whence positivity gives $\mathcal{K}_{s,\Xi}\widehat{n}_C = 0$. Writing $\xi = \rho\widehat{n}_C + \sqrt{\varepsilon_z}\eta$ therefore yields

$$g_s^{(1)}(z, \Xi) = \varepsilon_z \eta^\top \mathcal{K}_{s,\Xi}\eta \le \varepsilon_z.$$

□

G.4 Exact unequal-leverage leave-one-out consequences

Set $n = m - d$ and $s = m - 1$, so $\alpha_{m-1} = 1/n$.

Proposition 31 (Pointwise and minimized LOO inequalities and angle law)**.** *For every $z \in \mathcal{Z}_A$,*

$$\frac{R_A(z)}{c_{\max}} \preceq nD_{m-1}^{\mathrm{cov}}(z) \preceq \frac{R_A(z)}{c_{\min}}. \tag{176}$$

The matrix sandwich also yields trace and smallest-eigenvalue inequalities pointwise in z. For a fixed nonzero $\Xi \succeq 0$, recall the margins t_Q from (21) *and $\nu_A = \min_{z\in\mathcal{Z}_A} \lambda_{\min}[R_A(z)]$. Use the response-uniform $\tau_Q(A, \Xi)$ from* (160)*, equivalently $\min_{z\in\mathcal{Z}_A} g_{m-1}^{(1)}(z, \Xi)$, and define $\tau_{\mathrm{spec}}(A) := \min_{z\in\mathcal{Z}_A} n\lambda_{\min}[D_{m-1}^{\mathrm{cov}}(z)]$. The minima are attained and satisfy*

$$\frac{t_Q}{c_{\max}} \le \tau_Q(A, \Xi) \le \frac{t_Q}{c_{\min}}, \qquad \frac{\nu_A}{c_{\max}} \le \tau_{\mathrm{spec}}(A) \le \frac{\nu_A}{c_{\min}}. \tag{177}$$

Equivalently,

$$\tau_Q(A, \Xi) = 1 - \frac{\mathsf{T}_Q^{\mathrm{RU}}(A, m-1)}{\alpha_{m-1}}, \qquad \tau_{\mathrm{spec}}(A) = 1 - \frac{\mathsf{T}_{\mathrm{spec}}^{\mathrm{RU}}(A, m-1)}{\alpha_{m-1}}.$$

Finally, let C be a maximal projective class of normalized complement rows, with unit representative $\widehat{n}_C$, anchor $z_C = N\widehat{n}_C$, and $V_C = \ker A_{-C}$. For any nonzero $\Xi \succeq 0$ with $\operatorname{range}\Xi \subseteq V_C$, put $\varepsilon_z = 1 - (z^\top z_C)^2$. Then

$$g_{m-1}^{(1)}(z, \Xi) = 1 - (z^\top z_C)^2 = \varepsilon_z \qquad (z \in \mathcal{Z}_A). \tag{178}$$

No equal-leverage assumption is used.

Proof. By the pointwise identity (141), $nD_{m-1}^{\mathrm{cov}}(z) = \sum_i (c_i - z_i^2)a_i a_i^\top/c_i$, whereas $R_A(z) = \sum_i (c_i - z_i^2)a_i a_i^\top$. Since $z_i^2 \le \|n_i\|_2^2 = c_i$, comparison of the positive summands proves (176). Taking PSD traces and smallest eigenvalues for each z, then minima over the compact residual sphere, proves (177); continuity gives attainment.

For the angle law, set $u_i = a_i^\top \Xi a_i / W_Q$. Range containment gives $u_i = 0$ off C and $\sum_{i\in C} u_i = 1$. On every row of the maximal class, $z_i^2/c_i = |\widehat{n}_C^\top \xi|^2 = (z^\top z_C)^2$. The Ξ-trace of the first identity in (141), with $\alpha_{m-1} = 1/n$, proves (178). □

G.5 Multi-output residual energy fraction

The following proposition gives the matrix form of the aggregate gap described in Section 5.3; its assumptions and endpoint exclusions are explicit.

Proposition 32 (Residual energy-fraction gap)**.** *Let $d \ge 1$, $m \ge d + 2$, and let $A \in \mathbb{R}^{m\times d}$ be real with $A^\top A = I_d$ and $c_i = 1 - \|a_i\|_2^2 > 0$ for every row. Use the same ordinary unrescaled indexed size-s volume law and selected unweighted OLS at an integer $d < s < m$. Let*

$$E = [e_1, \dots, e_K] \ne 0, \qquad A^\top E = 0,$$
$$L_j := \|e_j\|_2^2, \qquad L_{\mathrm{src}} := \sum_j L_j = \|E\|_F^2 > 0.$$

Zero residual columns have zero covariance and are omitted. For each retained column, let $\overline{M}_{s,j}$ be its full-Gram-whitened selected-coefficient covariance. Define

$$\zeta_E := 1 - \frac{\|E\|_2^2}{\|E\|_F^2}, \qquad B_A := A^\top \operatorname{diag}(c_1, \dots, c_m) A.$$

Then

$$I_d - \frac{\sum_j \overline{M}_{s,j}}{\alpha_s L_{\mathrm{src}}} \succeq \vartheta_s \zeta_E B_A \succeq \vartheta_s \zeta_E c_{\min} I_d. \tag{179}$$

Here $\alpha_s = (m-s)/(m-d) > 0$ and $\vartheta_s = (s-d)/(m-d-1) > 0$ are repeated locally because the displayed normalization is invalid at an endpoint. Moreover, $\zeta_E > 0$ exactly when $\operatorname{rank}(E) \geq 2$.

Proof. For each $L_j > 0$, put $z_j = e_j/\sqrt{L_j}$. Applying the boundary-valid contraction output by output and summing gives

$$\sum_j \overline{M}_{s,j} \preceq \alpha_s L_{\mathrm{src}} I_d - \beta_s \gamma_s \sum_{j: L_j > 0} L_j R_A(z_j).$$

Define the pooled residual operator

$$\overline{R}_E := \frac{1}{L_{\mathrm{src}}} \sum_{j: L_j > 0} L_j R_A(z_j).$$

Expanding row by row gives the exact identity

$$\overline{R}_E = A^\top \operatorname{diag}\left(c_i - \frac{(EE^\top)_{ii}}{L_{\mathrm{src}}} \right) A. \tag{180}$$

Let $P_\perp = I_m - AA^\top$. Since $A^\top E = 0$, $P_\perp E = E$. For every $x \in \mathbb{R}^m$,

$$x^\top E E^\top x = \|E^\top P_\perp x\|_2^2 \leq \|E\|_2^2 \|P_\perp x\|_2^2,$$

so

$$\frac{EE^\top}{L_{\mathrm{src}}} \preceq (1 - \zeta_E) P_\perp.$$

Taking diagonal entries and using $(P_\perp)_{ii} = c_i$ gives $(EE^\top)_{ii}/L_{\mathrm{src}} \leq (1-\zeta_E)c_i$. Substitution into (180) yields $\overline{R}_E \succeq \zeta_E B_A$. Also

$$B_A = \sum_i c_i a_i a_i^\top \succeq c_{\min} \sum_i a_i a_i^\top = c_{\min} I_d.$$

Divide the summed contraction by $\alpha_s L_{\mathrm{src}} > 0$ and use $\beta_s \gamma_s / \alpha_s = \vartheta_s$ to prove (179). Finally, for nonzero E, equality $\|E\|_2^2 = \|E\|_F^2$ holds exactly at rank one. □

For every nonzero query metric $\Xi \succeq 0$, define the aggregate deficit separately from (163) by

$$g_s^{\mathrm{agg}}(E, \Xi) := 1 - \frac{\operatorname{tr}(\Xi \sum_j \overline{M}_{s,j})}{\alpha_s L_{\mathrm{src}} W_Q}. \tag{181}$$

PSD trace monotonicity gives

$$g_s^{\mathrm{agg}}(E, \Xi) \geq \vartheta_s \zeta_E \sum_i c_i \frac{a_i^\top \Xi a_i}{W_Q}. \tag{182}$$

G.6 Excluded and degenerate branches

Proper-class constants require $\{0\} \neq V_C \neq \mathbb{R}^d$, hence $d > 1$. Appendix A handles endpoints: $s = d$ gives no asserted positive gap, while $s = m$ has $\alpha_m = 0$, zero covariance and unnormalized deficit, and no normalized deficit. When $m = d + 1$ there is no interior size. Coloops can invalidate complement normalization and LOO denominators.

For $\Xi = 0$, unnormalized query excess is zero but $g_s^{(1)}$, ε_Ξ, and query weights are undefined. For $L^* = 0$, every supported fit equals Full and covariance is zero, but the unit residual and normalized single-output covariance are undefined. For $E = 0$, all output covariances are zero, but ζ_E, g_s^{agg}, and (179) are undefined. Zero columns of nonzero E are omitted; nonzero rank-one E has $\zeta_E = 0$ and a trivial energy floor. The joint result is for one fixed design and proper class, not design perturbations, design-independent conditioning, or a global near-contact classification.

GROUP II. CERTIFICATES AND COMPUTATION

H SCALARIZATION OF A DIRECTED FEATURE-ONLY CERTIFICATE

This section proves the scalar upper-envelope family used in Section 6.1. It starts with a proved lower certificate $0 \le t \le \nu_A$.

H.1 FEATURE-ONLY MARGINS AND THE COMPLETE SCALAR HIERARCHY

The kernel margin and its spectral consequence are

$$\nu_A = \min_{z \in \mathcal{Z}_A} \lambda_{\min}(R_A(z)), \tag{183}$$

$$0 \le \beta_s \gamma_s \nu_A \le \alpha_s - \mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A, s). \tag{184}$$

Zero is exactly the contact phase. The complement-pair certificate $t_\perp = (1-\rho_\perp)\kappa_\perp \le \nu_A$, built from the largest complement correlation and worst post-deletion conditioning, has the same zero phase (Appendix D).

Maximum-leverage lower bound. Let $L_{\max} = \max_i \ell_i$. Then

$$\nu_A \ge t_{\mathrm{lev}}(A) := \max\{0, 1 - 2L_{\max}\}. \tag{185}$$

For every $z \in \mathcal{Z}_A$, Parsevalness gives $\sum_i \ell_i a_i a_i^\top \preceq L_{\max} I_d$. Also $a_i a_i^\top \preceq L_{\max} I_d$ and $\sum_i z_i^2 = 1$, so $\sum_i z_i^2 a_i a_i^\top \preceq L_{\max} I_d$. Thus $R_A(z) \succeq (1 - 2L_{\max}) I_d$; the orthonormal augmented frame $[A\ z]$ also gives $R_A(z) \succeq 0$. Minimizing over z proves (185).

For any validated $0 \le t \le \nu_A$, define

$$B_{\mathrm{univ}}(s) = \alpha_s, \quad B_{\mathrm{lin}}(s;t) = \alpha_s - \beta_s \gamma_s t, \quad B_{\mathrm{res}}(s;t) = 1 - \frac{\beta_s}{1 - \gamma_s t}. \tag{186}$$

Lemma 33 (Certificate-envelope monotonicity)**.** *Under the assumptions of Theorem 5, put $n = m - d$. If $0 \le t \le \nu_A$, then*

$$t \le \nu_A < \frac{n-1}{n}.$$

Consequently, for $r = s - d \in \{1, \dots, n-1\}$, all denominators in $B_{\mathrm{res}}(s;t)$ are positive and $B_{\mathrm{univ}}(s)$, $B_{\mathrm{lin}}(s;t)$, and $B_{\mathrm{res}}(s;t)$ are strictly decreasing in r.

Proof. Since $d \ge 1$, the identity I_d is a nonzero PSD query. The margin-range bound established before the first-pass calculation in Appendix E.4 gives

$$0 \le t \le \nu_A \le t_{I_d} < \frac{n-1}{n}.$$

Write

$$D_r = 1 - \frac{n-r}{n-1} t > \frac{r}{n} > 0.$$

For $1 \le r \le n-2$, direct subtraction gives

$$B_{\mathrm{univ}}(d+r) - B_{\mathrm{univ}}(d+r+1) = \frac{1}{n},$$

$$B_{\mathrm{lin}}(d+r;t) - B_{\mathrm{lin}}(d+r+1;t) = \frac{(n-1) + t(n-2r-1)}{n(n-1)},$$

$$B_{\mathrm{res}}(d+r;t) - B_{\mathrm{res}}(d+r+1;t) = \frac{1 - nt/(n-1)}{n D_r D_{r+1}}.$$

The first and third quantities are positive. The middle numerator is at least $(n-1) - t(n-3) \ge 3(n-1)/n > 0$ when $n \ge 3$; for $n = 2$ there is no adjacent pair. This proves the claim. □

For a fixed positive-loss response, let $z = e/\sqrt{L^*}$ and $q = m - s$. For $1 \le r = s - d \le n - 1$, the boundary-valid resolvent inequality gives

$$\frac{\overline{M}_s}{L^*} \preceq I_d - \frac{n-q}{n}\left(I_d - \frac{q}{n-1}R_A(z)\right)^{-1}. \tag{187}$$

Since $R_A(z) \succeq tI_d$ and $I_d - \gamma_s R_A(z) \succ 0$ by (79), inversion reverses Loewner order. The lemma gives $1 - qt/(n-1) > 0$, so

$$\begin{aligned}\frac{\overline{M}_s}{L^*} &\preceq \left(1 - \frac{(n-q)/n}{1 - qt/(n-1)}\right) I_d \\ &= \frac{q(n-1-nt)}{n(n-1-qt)} I_d = B_{\mathrm{res}}(s;t) I_d.\end{aligned} \tag{188}$$

This uses the one-sided resolvent at arbitrary full-rank support boundaries, not an exact inverse-moment covariance identity there.

Set $x = \gamma_s t \in [0, 1)$. Direct subtraction gives

$$B_{\mathrm{lin}}(s;t) - B_{\mathrm{res}}(s;t) = \frac{\beta_s x^2}{1-x} \ge 0, \qquad B_{\mathrm{univ}}(s) - B_{\mathrm{lin}}(s;t) = \beta_s x \ge 0. \tag{189}$$

Thus the scalarization gives the hierarchy:

$$\frac{\overline{M}_s}{L^*} \preceq B_{\mathrm{res}}(s;t) I_d \preceq B_{\mathrm{lin}}(s;t) I_d \preceq B_{\mathrm{univ}}(s) I_d. \tag{190}$$

At Full, the covariance and all three envelopes are zero. At rank size $s = d$, use the separate rank-size branch; the strict-interior resolvent argument does not invoke rank-$(d+1)$ augmentation there.

Corollary 34 (First upper-envelope passes)**.** *Fix one admissible integer domain and tolerance τ. If $s_{\mathrm{Res}}, s_{\mathrm{Lin}}, s_{\mathrm{Univ}}$ are the first integers at which the corresponding envelopes in* (186) *are at most τ, then*

$$s_{\mathrm{Res}} \le s_{\mathrm{Lin}} \le s_{\mathrm{Univ}}. \tag{191}$$

On a common ordered integer domain, the pointwise order in Equation (190) nests the three pass sets. Their first integers therefore obey Equation (191); only these upper-envelope predicates are ordered.

H.2 Closed-form resolvent pass and leverage-scale envelope

Here we derive the exact resolvent first pass and the maximum-leverage consequence

$$s_{\mathrm{Res}}(\tau; t_{\mathrm{lev}}) \le \min\left\{m,\ d + \left\lceil \frac{2(1-\tau)\mu d}{\tau} \right\rceil \right\}, \qquad \mu = mL_{\max}/d, \quad L_{\max} < 1/2. \tag{192}$$

Take a full-column-rank real design under Theorem 5, a certificate $0 \le t \le \nu_A$, and $0 < \tau < 1$. Search $s \in \{d+1, \dots, m\}$, assigning Full the exact bound zero. The inherited theorem domain has $n = m - d \ge 2$, and the monotonicity lemma gives $t < (n-1)/n$. Put $\delta = 1 - nt/(n-1) \in (0, 1]$ and $r = s - d$. Simplifying Equation (188) gives

$$\begin{aligned} B_{\mathrm{res}}(d+r;t) &= \frac{\delta(n-r)}{n\delta + (1-\delta)r}, \\ B_{\mathrm{res}}(d+r;t) \le \tau \quad &\Longleftrightarrow \quad r \ge \frac{n(1-\tau)\delta}{\tau + (1-\tau)\delta}. \end{aligned} \tag{193}$$

The rational denominator is positive for $1 \le r \le n$; at $r = n$ the expression is zero, agreeing with the exact Full bound. Since $0 < (1-\tau)\delta < \tau + (1-\tau)\delta$, the crossing lies in $(0, n)$. Its ceiling is the first legal $r \in \{1, \dots, n\}$; if that ceiling is n, the pass is Full with its exact zero bound:

$$s_{\mathrm{Res}}(\tau; t) = d + \left\lceil \frac{n(1-\tau)\delta}{\tau + (1-\tau)\delta} \right\rceil. \tag{194}$$

For a reported certificate, the stored floating-point endpoint is interpreted as its exact rational value before evaluating the ceiling predicate, so evaluation rounding cannot alter the cardinality for that stored certificate.

For the maximum-leverage certificate assume $L_{\max} < 1/2$ and put $t = t_{\mathrm{lev}} = 1 - 2L_{\max}$ and $\mu = mL_{\max}/d$. Thus μ measures maximum leverage relative to the average d/m, and $\mu d = mL_{\max}$ remains explicit in the sufficient-size scale. Since the average row leverage is d/m, $\mu \geq 1$. The definition of δ and $L_{\max} < 1/2$ give

$$n\delta = \frac{2n^2 L_{\max} - n}{n-1} < 2nL_{\max} < 2mL_{\max} = 2\mu d, \tag{195}$$
$$\frac{n(1-\tau)\delta}{\tau + (1-\tau)\delta} \leq \frac{n(1-\tau)\delta}{\tau} < \frac{2(1-\tau)\mu d}{\tau}.$$

The second line bounds the crossing in Equation (193). Taking ceilings and assigning the Full endpoint if needed proves Equation (192), which gives a sufficient cardinality for the scalar-envelope predicate. It is not an m-independent $O(d)$ statement unless μ is separately controlled, and it makes no optimal sample-complexity claim.

I From covariance bounds to expected query loss

For fixed source responses Y, query features Z, and query targets Y_Q, use the common-support identity (136). Write $M_{s,j}$ for the coefficient covariance of output j, L_j^* for its Full residual energy, and $L_{Q,s}$ for query MSE normalized by qK. On the real no-coloop domain with $d < s < m$ and $0 \leq t \leq \nu_A$, scalarization yields

$$M_{s,j} \preceq B_{\mathrm{res}}(s;t) L_j^* G^{-1} \preceq B_{\mathrm{lin}}(s;t) L_j^* G^{-1}. \tag{196}$$

Zero-loss columns contribute zero. Summing and contracting gives $\mathcal{D}_s \leq B(s;t)\Gamma_Q$ for either scalar envelope, where $\Gamma_Q = (\sum_j L_j^*)\,\mathrm{tr}(ZG^{-1}Z^\top)/(qK)$. If $L_{Q,F} > 0$,

$$C_Q = \Gamma_Q / L_{Q,F}, \qquad \frac{\mathbb{E}L_{Q,s} - L_{Q,F}}{L_{Q,F}} \leq B(s;t) C_Q. \tag{197}$$

The additive identity remains valid when a relative ratio is undefined.

I.1 Nested sizes and paired loss

The paired experiments draw one rank-d volume basis and one uniformly permuted complement, then take the required prefixes (Dereziński & Warmuth, 2017; 2018b). For fixed $d \leq a < b \leq m$ and positive-volume U of size b, this gives

$$\mathbb{P}(S_a = T \mid S_b = U) = \frac{\det(X_T^\top X_T)}{\binom{b-d}{a-d} \det(X_U^\top X_U)}. \tag{198}$$

Indeed, summing over bases inside $T \subseteq U$ by Cauchy–Binet and counting complement orders gives the joint mass

$$\mathbb{P}(S_a = T, S_b = U) = \frac{\det(X_T^\top X_T)}{\det(G)\binom{m-d}{a-d}\binom{m-a}{b-a}}. \tag{199}$$

Division by the ordinary size-b mass and the binomial identity $\binom{m-d}{a-d}\binom{m-a}{b-a} = \binom{m-d}{b-d}\binom{b-d}{a-d}$ prove the conditional law. Zero-volume subsets have zero mass. Conditional selected-OLS unbiasedness therefore gives

$$\mathbb{E}[W_a - W_b \mid S_b] = 0. \tag{200}$$

Expanding the squared query loss and cancelling the conditional cross term,

$$H_{a,b} := \frac{\|Z(W_a - W_b)\|_F^2}{qK}, \qquad \mathbb{E}[L_{Q,a} - L_{Q,b}] = \mathbb{E}H_{a,b}. \tag{201}$$

On the scalar-certificate domain with $d < a < b \leq m$, the same argument gives

$$\mathbb{E}H_{a,b} \leq \mathbb{E}L_{Q,a} - L_{Q,F} \leq B_{\mathrm{res}}(a;t)\Gamma_Q, \tag{202}$$

since $\mathbb{E}L_{Q,b} \geq L_{Q,F}$. This does not subtract separately constructed upper bounds or equate realized loss differences with squared displacement.

Lemma 35 (Conditional Pythagoras)**.** *For the preceding path, let $\mathcal{G}_s = \sigma(S_s, \dots, S_m)$ and $P_s = ZW_s$. For fixed $a < b$, $\mathbb{E}[W_a \mid \mathcal{G}_b] = W_b$. For any square-integrable $\mathcal{G}_b$-measurable matrix A shaped as P_s,*

$$\mathbb{E}[\|P_a - A\|_F^2 \mid \mathcal{G}_b] = \|P_b - A\|_F^2 + \mathbb{E}[\|P_a - P_b\|_F^2 \mid \mathcal{G}_b].$$

Consequently, for $a < b < c$, $\mathbb{E}[H_{a,c} \mid \mathcal{G}_b] = \mathbb{E}[H_{a,b} \mid \mathcal{G}_b] + H_{b,c}$; with $I_s = P_s - P_{s+1}$, $\mathbb{E}H_{a,b} = \sum_{s=a}^{b-1} \mathbb{E}\|I_s\|_F^2/(qK)$.

Proof. Given S_b, the inner path is conditionally independent of the later complement order. Thus conditional unbiasedness extends to $\mathcal{G}_b$. Expanding $P_a - A = (P_a - P_b) + (P_b - A)$ cancels the conditional cross term. Taking $A = P_c$ gives additivity, and iterating it gives the increment sum. □

Corollary 36 (Coefficient-covariance increments)**.** *On this fixed full-rank, common-support path, for each output and fixed $d \le a < b \le m$,*

$$M_{a,j} - M_{b,j} = \mathbb{E}[(w_{a,j} - w_{b,j})(w_{a,j} - w_{b,j})^\top] \succeq 0. \tag{203}$$

Proof. Expand $w_{a,j} - w_j^* = (w_{a,j} - w_{b,j}) + (w_{b,j} - w_j^*)$. The cross terms vanish conditional on $\mathcal{G}_b$; summing over outputs also preserves the identity and positive-semidefinite order. □

These nested-law statements need no margin or no-coloop assumption, but sizes must be fixed independently of path outcomes. For example, with $X = \mathbf{1}_3$, $y = (0, 0, 3)^\top$, choose size one when its fitted value is 3 and size two otherwise. The selected coefficient then has expectation $\frac{1}{3}\,3 + \frac{2}{3}\,\frac{3}{4} = \frac{3}{2}$, whereas the Full coefficient is 1. Fixed-size unbiasedness therefore does not justify outcome-adaptive stopping.

J Response-aware query certificates and relative-loss guarantees

We prove the hierarchy in Corollary 8 and its relative-loss consequence. Throughout, the target is the ordinary fixed-size volume law with common-support unweighted OLS.

Endpoint conventions. Zero-loss output columns are omitted before residual normalization. If $L_{\mathrm{src}} = 0$, every supported selected output equals Full and $\mathcal{D}_s = 0$; Full ($s = m$) also has exact zero excess. In either case $U_2, U_{\mathrm{per}}, U_{\mathrm{agg}}$ are set to zero. At $m = d$ only Full is available, while $m = d + 1$ has no strict-interior size. At $s = d < m$, use the separate rank-size bound without invoking the augmented rank-$(d+1)$ formulas. These branches are evaluated before any inverse with an undefined denominator.

J.1 Output aggregation

Let e_j be column j of the exact Full residual matrix E, put $L_j = \|e_j\|_2^2$, and let $J_+ = \{j : L_j > 0\}$. For $j \in J_+$, the output-specific response geometry from Appendix B is

$$R_j = \sum_{i=1}^{m} \left(1 - \ell_i - \frac{e_{ij}^2}{L_j}\right) a_i a_i^\top. \tag{204}$$

$$\bar{R} = \frac{1}{L_{\mathrm{src}}} \sum_{j \in J_+} L_j R_j = A^\top \operatorname{diag}\left(c_i - \frac{(EE^\top)_{ii}}{L_{\mathrm{src}}}\right) A. \tag{205}$$

The boundary-valid output-wise resolvent inequality and trace monotonicity give

$$\mathcal{D}_s \le U_{\mathrm{per}}(s) := \frac{1}{qK} \sum_{j \in J_+} L_j \operatorname{tr}\left[\Xi\{I - \beta_s(I - \gamma_s R_j)^{-1}\}\right]. \tag{206}$$

For $L_{\mathrm{src}} > 0$, set $p_j = L_j / L_{\mathrm{src}}$. Operator convexity of inversion gives

$$\sum_{j \in J_+} p_j (I - \gamma_s R_j)^{-1} \succeq \left(I - \gamma_s \sum_{j \in J_+} p_j R_j\right)^{-1} = (I - \gamma_s \bar{R})^{-1}. \tag{207}$$

Multiplication by $-\beta_s$, congruence with $\Xi^{1/2}$, and the trace prove $U_{\mathrm{per}} \le U_{\mathrm{agg}}$. Thus aggregation replaces the $|J_+|$ output-specific inverses by one inverse, at the possible cost of a looser bound. A directly checked aggregate condition $0 \le t < 1/\gamma_s$ and $\bar{R} \succeq tI$ gives the scalar comparison

$$U_{\mathrm{agg}}(s) \le U_{\mathrm{scalar}}(s;t) := \frac{L_{\mathrm{src}}}{qK} B_{\mathrm{res}}(s;t)\,\mathrm{tr}(\Xi). \tag{208}$$

Indeed, $(I-\gamma_s\bar{R})^{-1} \succeq (1-\gamma_s t)^{-1} I$ proves (208). A common output-wise lower bound $R_j \succeq tI$ is sufficient but is not required or asserted by the reported calculation.

J.2 Output-wise second-moment refinement

For each $j \in J_+$, put $z_j = e_j/\sqrt{L_j}$ and $B_j = [A, z_j]$. Under the analysis-only augmented law $\mathbb{P}_{B_j}$, recall $H_{s,j}$ and $V_{s,j}$ from Equation (26), and put $\mathcal{Q}_{s,j} = H_{s,j}^{-1} V_{s,j} H_{s,j}^{-1}$. Here $K_S = A_S^\top A_S$; selected inverses under this law are restricted to positive augmented volume. For every full-column-rank design and $d < s < m$, $H_{s,j} \succ 0$ and $\mathcal{Q}_{s,j} \succeq 0$. Equation (27) defines C_s and $U_2 = U_{\mathrm{per}} - C_s$. The boundary-safe part of Proposition 20 gives

$$\mathcal{D}_s \le U_2(s) \le U_{\mathrm{per}}(s) \le U_{\mathrm{agg}}(s). \tag{209}$$

Indeed, its raw-coordinate matrices obey $F_{s,j} = G^{1/2} H_{s,j} G^{1/2}$ and $W_{s,j}^{\mathrm{raw}} = G^{1/2} V_{s,j} G^{1/2}$, directly from their finite-sum definitions. Congruence by $G^{1/2}$, contraction with $\Xi \succeq 0$, and summation over outputs prove the first inequality. The second uses $C_s \ge 0$; the last is Equation (207). These bounds hold for every full-column-rank design on the stated interior domain; the section's endpoint conventions apply separately.

J.3 Shared computation of the second moment

Write $b_{ij}^\top$ for row i of B_j, and set $h_{ij} = \|b_{ij}\|_2^2$, $r_{ij} = 1 - h_{ij}$, and $D_i = a_i a_i^\top$. The following matrices do not depend on the card size:

$$\begin{aligned} R_j &= \sum_i r_{ij} D_i, & U &= \sum_i D_i^2, & W_j &= \sum_i h_{ij} D_i^2, \\ T_{pt}^{(j)} &= \sum_i b_{ij,p} b_{ij,t} D_i, & S_j &= \sum_{p,t=1}^{d+1} (T_{pt}^{(j)})^2. \end{aligned} \tag{210}$$

Every square here is an ordered matrix product, not an entrywise square. Put $n_0 = m - d - 1$, $k = m - s$, $c = k/n_0 = \gamma_s$, and $c_2 = k(k-1)/[n_0(n_0-1)]$ when $n_0 \ge 2$. When $n_0 = k = 1$, set $c_2 = 0$ before arithmetic. The symbol n_0 is unrelated to the query-row count q. Then

$$V_{s,j} = (c_2 - c^2) R_j^2 - c_2 S_j + (c - c_2) U + (2c_2 - c) W_j. \tag{211}$$

To verify this identity, let $J_T = \sum_{i \in S^c} D_i$, so that $K_S = I - J_T$ and $\mathbb{E}_{B_j} J_T = c R_j$. The proposition's singleton and pair exclusion probabilities give

$$\mathbb{E}_{B_j} J_T^2 = c(U - W_j) + c_2 \sum_{i \ne l} \{ r_{ij} r_{lj} - (b_{ij}^\top b_{lj})^2 \} D_i D_l.$$

Expanding the block squares yields $S_j = \sum_{i,l} (b_{ij}^\top b_{lj})^2 D_i D_l$. The diagonal contribution to $R_j^2 - S_j$ is $\sum_i (r_{ij}^2 - h_{ij}^2) D_i^2 = U - 2W_j$. Subtracting $(cR_j)^2$ proves Equation (211), with both product orders retained. In the $n_0 = k = 1$ branch, the excluded set has one row and $V_{s,j} = U - W_j - R_j^2$ directly. More generally, when $k = 1$, $c_2 = 0$ and forming S_j is unnecessary.

The A columns are common across outputs. Define

$$\begin{aligned} T_{pt}^{AA} &= \sum_i a_{i,p} a_{i,t} D_i, & S_{AA} &= \sum_{p,t=1}^{d} (T_{pt}^{AA})^2, \\ M_{p,j} &= \sum_i a_{i,p} z_{ij} D_i, & N_j &= \sum_i z_{ij}^2 D_i. \end{aligned} \tag{212}$$

Thus

$$S_j = S_{AA} + 2\sum_{p=1}^{d} M_{p,j}^2 + N_j^2. \tag{213}$$

The factor two counts the (p, z) and (z, p) blocks. A reproducible quadratic-feature convention is the *unscaled* upper triangle $v_i = (a_{i,p}a_{i,t})_{p\le t}$, of length $d(d+1)/2$. For $F = \sum_i v_i v_i^\top$, symmetrically unpacking column (p, t) gives T_{pt}^{AA}. Consequently $S_{AA} = \sum_p (T_{pp}^{AA})^2 + 2\sum_{p<t}(T_{pt}^{AA})^2$. This convention has no $\sqrt{2}$ scaling. The mixed feature products $\sum_i z_{ij} a_i v_i^\top$ similarly unpack to the $M_{p,j}$ blocks. Since $D_i^2 = \ell_i D_i$, the remaining moments satisfy

$$U = \sum_i \ell_i D_i, \qquad R_j = I - U - N_j, \qquad W_j = \sum_i \ell_i^2 D_i + \sum_i z_{ij}^2 \ell_i D_i.$$

The first term of W_j is shared as well.

One source-row stream accumulates the quadratic Gram for S_{AA} and shared U; per positive-loss output, it accumulates $M_{p,j}$, N_j, and the remaining W_j. At $k = 1$, S_j is skipped. For each interior size, (211) yields $V_{s,j}$; then the ordered product $H_{s,j}^{-1} V_{s,j} H_{s,j}^{-1}$ is contracted with the fixed query metric. No source-pair loop or $m \times m$ source Gram is required.

For $K_+ = |J_+|$ and C interior cards, this direct dense construction has arithmetic cost

$$O(md^4 + K_+ md^3 + d^5 + K_+ d^4 + CK_+ d^3). \tag{214}$$

These terms respectively cover shared quadratic-feature products, mixed products, shared block squares, output-specific block squares, and per-card matrix inverses and contractions, after the common Full fit and query geometry have been prepared. Row blocks allow accumulation without an $m \times m$ row Gram or an $m \times d(d+1)/2$ stored feature matrix. Apart from inputs and streamed blocks, the dense state is $O(d^4 + K_+ d^3 + Cd^2)$. The dependence on source-row count is linear for fixed d, K_+; avoiding m^2 storage does not remove the explicit d^4 and d^5 limitations in (214).

A common information hierarchy. On the common real Parseval/no-coloop domain with $L_{\mathrm{src}} > 0$, $\operatorname{tr}\Xi > 0$, $d < s < m$, and $0 \le t \le \nu_A$,

$$\mathcal{D}_s/\Gamma_Q \le U_2/\Gamma_Q \le U_{\mathrm{per}}/\Gamma_Q \le U_{\mathrm{agg}}/\Gamma_Q \le B_{\mathrm{res}}(s; t_Q) \le B_{\mathrm{res}}(s; t) \le B_{\mathrm{lin}}(s; t) \le \alpha_s. \tag{215}$$

Rightward bounds discard second moments, output-specific inverses, realized geometry, resolvents, and slack (weighted-trace link: Appendix E.4).

J.4 Known-label floors and the expected guarantee

Put $F = ZW_F$, let $\mathcal{Y}$ be a specified class of allowed fixed query label matrices, and retain the original row/output normalization:

$$L_{\min}(W_F; \mathcal{Y}) = \min_{Y_Q \in \mathcal{Y}} \frac{\|F - Y_Q\|_F^2}{qK}. \tag{216}$$

For rowwise allowed alphabets $\mathcal{A}_i$,

$$L_{\min}^{\mathrm{row}} = \frac{1}{qK} \sum_{i=1}^{q} \min_{y \in \mathcal{A}_i} \|F_{i,:} - y\|_2^2. \tag{217}$$

This covers the 60 one-hot MASSIVE labels and binary KuaiRand labels. If a fixed partition $\{G_g\}$ is known to share one one-hot class within each group, the sharper floor is

$$L_{\min}^{\mathrm{group}} = \frac{1}{qK} \sum_g \min_{c \in [K]} \sum_{i \in G_g} \|F_{i,:} - e_c\|_2^2 \ge L_{\min}^{\mathrm{row}}. \tag{218}$$

Every row retains its original weight. The group condition is an assumption on the allowed label set; its dataset-specific premise is stated with the relative-loss experiment.

For every allowed labeling, $L_{Q,F} \ge L_{\min}$. The fixed-label identity (136) and $\mathcal{D}_s \le U_2$ give, for $\eta \ge 0$,

$$U_2(s) \le \eta L_{\min} \quad \Longrightarrow \quad \mathbb{E}_X L_{Q,S} \le (1+\eta) L_{Q,F} \quad \text{for every fixed } Y_Q \in \mathcal{Y}. \tag{219}$$

The same conclusion follows if U_{per} or U_{agg} is at most $\eta L_{\min}$. If the floor is zero, the multiplied inequality is retained; a finite ratio is not manufactured from $0/0$.

One may therefore fix a finite card family $\mathcal{S}$ containing Full and, before sampling, choose

$$s_U^*(\eta) = \min\{s \in \mathcal{S} : U(s) \le \eta L_{\min}\}. \tag{220}$$

The guarantee then uses the ordinary size-$s_U^*(\eta)$ law. The label-floor criterion and source-variance allowances used for continuous responses have different denominators and are evaluated separately.

K OMISSION LOWER CERTIFICATES

The upper certificates in Section 6.2 apply to every full-column-rank design. The lower results here additionally require $m \ge d+2$ and no coloops: every one-row deletion retains column rank. Write $G = X^\top X$, let $x_i^\top$ be row i, and set

$$\begin{aligned} &P = I - XG^{-1}X^\top, \quad c_i = P_{ii} > 0, \quad n = m - d, \\ &k = m - s, \qquad\qquad \alpha = k/n, \qquad \lambda = (k-1)/(n-1). \end{aligned} \tag{221}$$

Here $d < s < m$, so $1 \le k < n$. Ordinary indexed volume sampling assigns probability only to positive-volume subsets. We define no inverse or fit on the zero-volume subsets.

Supported-law omission moments. For one response column, write $\beta_F = w^*$ and $\beta_S = w_S$ for the full and supported subset fits of Section 2. Let $e = y - X\beta_F$, $v_S = \beta_S - \beta_F$, and

$$f_i = -\frac{e_i}{c_i} G^{-1}x_i, \qquad F_S = \sum_{i \notin S} f_i, \qquad L_0 = \sum_i \frac{e_i^2}{c_i} G^{-1}x_i x_i^\top G^{-1}. \tag{222}$$

If $e = 0$, every supported fit equals β_F, so $v_S = f_i = F_S = 0$ and $L_0 = L_1 = 0$ in the formulas below. We can therefore assume $L^* := \|e\|_2^2 > 0$ and set

$$A = XG^{-1/2}, \qquad z = e/\sqrt{L^*}, \qquad T = \sqrt{L^*}\, G^{-1/2}.$$

Then $A^\top A = I$, $A^\top z = 0$, and its residual projector is the same P. For every supported S, $\det(X_S^\top X_S) = \det(G)\det(A_S^\top A_S)$. The factor $\det(G) > 0$ cancels from the volume law, so the supported sets and their probabilities agree in the two coordinates. Writing $v_S^{A,z}$, $f_i^{A,z}$, and $F_S^{A,z}$ for the predictors in Section F, direct substitution gives

$$v_S = T v_S^{A,z}, \qquad f_i = T f_i^{A,z}, \qquad F_S = T F_S^{A,z}, \qquad L_0 = T L(z) T^\top, \qquad L_1 = T H(z) T^\top.$$

Here $L(z)$ and $H(z)$ are defined in (146) and (149), respectively. The singleton and pair exclusion probabilities from that supported law read

$$\Pr(i \notin S) = \alpha c_i, \qquad \Pr(i, j \notin S) = \alpha\lambda(c_i c_j - P_{ij}^2) \quad (i \ne j). \tag{223}$$

In particular, $\eta_s = \alpha\lambda$. Applying the exact normalized moments (150) through T gives

$$\begin{aligned} &\mathbb{E}[v_S F_S^\top] = \alpha L_0, \\ &\mathbb{E}[F_S F_S^\top] = \alpha(L_0 - \lambda L_1), \quad L_1 = \sum_{i,j} \frac{e_i e_j P_{ij}^2}{c_i c_j} G^{-1}x_i x_j^\top G^{-1}. \end{aligned} \tag{224}$$

At $k = 1$, only row i is omitted, $\lambda = 0$, and the selected fit is the full fit of X_{-i}; hence $v_S = f_i = F_S$ on each supported set. All identities above include this endpoint and assign no fit to an unsupported set.

Let $C = P \circ P$, $D = \mathrm{diag}(c_i)$, and let row i of B be $(e_i/c_i)x_i^\top G^{-1}$. Then $L_0 = B^\top D B$ and $L_1 = B^\top C B$. The Schur product theorem gives $C \succeq 0$. Since P is a projector, $\sum_j P_{ij}^2 = c_i$ and

$$h^\top(D - C)h = \frac{1}{2}\sum_{i,j} P_{ij}^2 (h_i - h_j)^2 \ge 0. \tag{225}$$

Thus $0 \preceq L_1 \preceq L_0$. For any fixed query matrix Z, define $U_0 = \mathrm{tr}(ZL_0Z^\top)$ and $U_1 = \mathrm{tr}(ZL_1Z^\top)$. In particular, $0 \le U_1 \le U_0$.

Scalar projection and the basic lower bound. For every real multiplier a, the two moments imply the exact identity

$$\mathbb{E}\|Z(v_S - aF_S)\|_2^2 = \mathbb{E}\|Zv_S\|_2^2 - \alpha\{(2a - a^2)U_0 + a^2\lambda U_1\}. \tag{226}$$

If $U_0 > 0$, its rightmost brace is maximized at $a_* = U_0/(U_0 - \lambda U_1)$; the denominator is at least $(1 - \lambda)U_0 > 0$. Hence

$$\mathbb{E}\|Zv_S\|_2^2 \geq \alpha\frac{U_0^2}{U_0 - \lambda U_1} \geq \alpha(U_0 + \lambda U_1) \geq \alpha U_0. \tag{227}$$

If $U_0 = 0$, then $U_1 = 0$ and the displayed lower bound is defined to be zero. Actual interior risk may still be positive. At leave-one-out, $k = 1$, $\lambda = 0$, and direct one-row deletion gives $\mathbb{E}\|Zv_{m-1}\|_2^2 = U_0/n$. Thus the last inequality gives $\mathbb{E}\|Zv_S\|_2^2 \geq k\mathbb{E}\|Zv_{m-1}\|_2^2$. The scalar optimization concerns only predictors aF_S. Equality in its first inequality requires $Z(v_S - a_*F_S) = 0$ on every supported subset. The volume-law moments in (224) make this scalar projection a certificate.

For a fixed workload g, take $Z = Z_g$ with q_g rows. For K response columns, apply the definitions of U_0, U_1 to each residual column to obtain $U_{0,j}, U_{1,j}$. A zero-loss output contributes zero. With $\alpha_s = \alpha = (m - s)/(m - d)$ and $\lambda_s = \lambda = (m - s - 1)/(m - d - 1)$, summing (227) and dividing by the same q_gK used for the upper certificate and allowance gives

$$\mathcal{D}_{s,g} \geq \mathcal{L}_{s,g}^{\mathrm{sc}} := \frac{\alpha_s}{q_gK}\sum_{j=1}^{K}\begin{cases} U_{0,j}^2/(U_{0,j} - \lambda_s U_{1,j}), & U_{0,j} > 0, \\ 0, & U_{0,j} = 0. \end{cases} \tag{228}$$

The zero branch is a zero lower bound, not a claim of zero actual risk. The last inequality in (227) also proves (30); the exact LOO identity covers a zero query. Optimizing each output separately can only improve on a shared multiplier.

Raw moments without a row-pair loop. For $q_i = ZG^{-1}x_i$, the scalar moments are

$$U_0 = \sum_i \frac{e_i^2}{c_i}\|q_i\|_2^2, \qquad U_1 = \sum_{i,j}\frac{e_ie_jP_{ij}^2}{c_ic_j}q_i^\top q_j. \tag{229}$$

Put $T_a = \sum_i (e_i/c_i)(q_i)_a x_i x_i^\top$ for query row a. Expanding $P_{ij} = \mathbf{1}_{i=j} - x_i^\top G^{-1}x_j$ gives the exact identity

$$U_1 = \sum_i \frac{(2c_i - 1)e_i^2}{c_i^2}\|q_i\|_2^2 + \sum_{a=1}^{q_g}\mathrm{tr}(T_aG^{-1}T_aG^{-1}). \tag{230}$$

The signed diagonal term matters. This formula supports streaming source moments and exact rational checks; a floating-point evaluation alone is not an outward-rounded certificate.

Safe numerical enclosure. Suppose exact rational endpoints enclose $U_0 \in [l_0, u_0]$ and $U_1 \in [l_1, u_1]$. They must be obtained with rigorous arithmetic. The function $x^2/(x - \lambda y)$ need not increase in x: its derivative is $x(x - 2\lambda y)/(x - \lambda y)^2$. Consequently, inserting l_0, l_1 without further work is unsafe. Intersect the box with $0 \leq y \leq x$. Set

$$\begin{aligned} y_{\min} &= \max(0, l_1), \quad & x_{\min} &= \max(0, l_0, y_{\min}), \\ x_{\max} &= u_0, & x_* &= \min(x_{\max}, \max(x_{\min}, 2\lambda y_{\min})). \end{aligned} \tag{231}$$

Reject an inconsistent box if $y_{\min} > u_1$ or $x_{\min} > x_{\max}$. Otherwise the exact minimum of the scalar expression over the structural box is zero when $x_* = 0$, and $x_*^2/(x_* - \lambda y_{\min})$ when $x_* > 0$. For fixed $x > 0$ the expression increases in y, and for $y = y_{\min}$ its derivative changes sign at $x = 2\lambda y_{\min}$. The structural inequality $y \leq x$ makes the zero endpoint continuous. Every final multiplication by $\alpha/(q_gK)$ and every comparison with the allowance is exact rational arithmetic.

An alternative is to fix an exact rational a before evaluation. With $b = 2a - a^2$ and $x_a = l_0$ if $b \geq 0$ (otherwise $x_a = u_0$), (226) gives the safe lower $\alpha\max\{0, bx_a + a^2\lambda l_1\}$. The selected-cell implementation uses the constrained-box minimum; it also records this independent fixed-multiplier path. For the issued data, exact dyadic frozen binary64 arrays produce rational Gram, fit, and clearance checks. Outward ball arithmetic encloses the raw moments, followed by rational box minimization. This certifies those frozen numeric arrays, not real-number error in the earlier public-CSV parsing and preprocessing.

L NUMERICAL EVALUATION AND SUFFICIENT-SIZE SELECTION

Numerical coordinates. Symmetric whitening and numerical QR are congruent coordinates for the same displacement. If $X = QR$ is a positive-diagonal reduced QR factorization, $A = XG^{-1/2}$, and $O = G^{1/2}R^{-1}$, then O is orthogonal and $Q = AO$. With $\Delta_A = G^{1/2}(W_s - W_F)$ and $\Delta_{\mathrm{QR}} = R(W_s - W_F)$,

$$\Delta_{\mathrm{QR}} = O^\top \Delta_A, \qquad \operatorname{tr}(\Xi\, \mathbb{E}\Delta_A \Delta_A^\top) = \operatorname{tr}\big[R^{-\top} Z^\top Z R^{-1}\, \mathbb{E}\big(\Delta_{\mathrm{QR}} \Delta_{\mathrm{QR}}^\top\big)\big]\,.$$

This bridge does not equate an empirical QR moment with an outward-certified endpoint.

Feature-only margins. For a fixed source transform, let $\underline{c}_{\min} > 0$ and $\overline{\rho}_\perp < 1$ enclose the complement norm and coherence in Proposition D.1. Its analytic bound gives

$$\underline{t}_{\mathrm{pair}} = \operatorname{down}\big[(1 - \overline{\rho}_\perp)\operatorname{down}(\underline{c}_{\min}^2)\big] \le t_\perp \le \nu_A, \tag{232}$$

where each subtraction and product is rounded downward. Similarly,

$$\underline{t}_{\mathrm{lev}} = \max\{0, \operatorname{down}(2\underline{c}_{\min} - 1)\} \le \nu_A. \tag{233}$$

One may take the maximum of valid lower endpoints. After the common Gram factorization, leverage evaluation costs $O(md^2)$; the pair route costs $O(m^2 d)$ and can stream its dot products without storing an $m \times m$ matrix.

Directed moment bounds. Floating-point approximations are accompanied by outward error enclosures. The common Gram enclosure must certify positive definiteness and the inverse perturbation domain; the exact Full solution, residuals, query Gram, and moment contractions are enclosed using the same transform. Normalization requires a strictly positive residual-loss lower bound, unless that output is verified to have exactly zero loss. Positive inverse denominators and finite error radii are required throughout. These conditions validate

$$\mathcal{D}_s \le U_2(s) \le U_2^+(s) := U_{\mathrm{per}}^+(s) - C_s^-, \qquad 0 \le C_s^- \le C_s.$$

A negative lower correction can be replaced by zero because $C_s \ge 0$. An unresolved correction does not turn a first-order endpoint into a second-moment certificate. A negative issued upper bound indicates failure. For relative-loss decisions, the label-floor lower endpoint includes the Full-prediction error allowance; a nonpositive lower floor is set to zero. Full and verified zero-loss cases use their exact values before any strict-interior normalization. Failed positivity or enclosure checks return no certificate; nominal point estimates are not substituted.

First passing size. Candidates, tolerance, and information access are fixed before the subsequent subset draw. For a scalar envelope searched over every admissible integer, exact rational interpretation of a stored endpoint verifies

$$B(s; \underline{t}) \le \tau < B(s-1; \underline{t}). \tag{234}$$

The predecessor is omitted at $s = d+1$; monotonicity then proves minimality. For a restricted grid or a moment policy, evaluate candidates in increasing order and return the first directed upper endpoint below the allowance. This certifies a sufficient grid size without assuming monotonicity of U_2 or claiming an actual-risk optimum. The moment contraction and its cost are derived in Appendix J.3.

GROUP III. EXACT CONSTRUCTIONS AND CONTROLS

M CRITICAL EQUAL-LEVERAGE SPECIALIZATION

The following specialization is supplementary geometric intuition for the general phase theorem; it is not used to form the four source-bound cards.

Theorem 37 (Critical residual geometry)**.** *Let $d \ge 2$ and let $A \in \mathbb{R}^{2d \times d}$ have orthonormal columns and $\|a_i\|_2^2 = 1/2$ for every row. With $\rho_A = \max_{i \ne j} 2|a_i^\top a_j|$,*

$$\frac{1 - \rho_A}{4} \le \nu_A \le \frac{(1 - \rho_A)(3 + \rho_A)}{8}. \tag{235}$$

Moreover, $\nu_A = 0$ if and only if there are distinct i, j, a unit v, and signs $\sigma_i, \sigma_j \in \{\pm 1\}$ such that

$$a_i = \frac{\sigma_i}{\sqrt{2}} v, \qquad a_j = \frac{\sigma_j}{\sqrt{2}} v, \qquad a_k^\top v = 0 \quad (k \notin \{i, j\}). \tag{236}$$

Corollary 38 (Equal-leverage leave-one-out near-contact)**.** *Let $d \geq 2$, $m = 2d$, $s = m - 1$, and let $A \in \mathbb{R}^{2d\times d}$ have orthonormal columns with $\|a_i\|_2^2 = 1/2$ for every row. Define the coherence indicator $\rho_A = \max_{i\neq j} 2|a_i^\top a_j|$. Then*

$$\frac{1-\rho_A}{2} \leq 1 - \frac{\mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A, m-1)}{\alpha_{m-1}} \leq \frac{(1-\rho_A)(3+\rho_A)}{4}. \tag{237}$$

Throughout the proof of Theorem 37, $A \in \mathbb{R}^{2d\times d}$ has orthonormal columns, each row has squared norm $1/2$, $P = AA^\top$, and $P_\perp = I_{2d} - P$.

M.1 Fixed-complement residual weighting

For $z \in \mathcal{Z}_A$, $P_\perp z = z$, so subtracting $zz^\top$ gives an orthogonal projection Q_z with $Q_z^2 = Q_z$, $Q_z \succeq 0$, and rank $d-1$. Its diagonal gives the fixed-complement representation:

$$Q_z = P_\perp - zz^\top, \tag{238}$$

$$(Q_z)_{ii} = 1 - \ell_i - z_i^2 = \tfrac{1}{2} - z_i^2, \tag{239}$$

$$R_A(z) = \sum_i (\tfrac{1}{2} - z_i^2) a_i a_i^\top = A^\top \operatorname{Diag}(\operatorname{diag} Q_z) A. \tag{240}$$

In particular, $z_i^2 \leq 1/2$ and $R_A(z) \succeq 0$.

Complement and projection-diagonal theory (Casazza et al., 2013; Kadison, 2002; Antezana et al., 2007) provide related background. Appendix U distinguishes the residual-generated weights in (240) from freely specified frame weights (Balazs et al., 2010).

M.2 Coherence sandwich

Put $u_i = \sqrt{2}a_i$. Then $\|u_i\|_2 = 1$, $\sum_i u_i u_i^\top = 2I_d$, and $\rho_A = \max_{i\neq j} |u_i^\top u_j|$. For unit $z \in \mathcal{Z}_A$ and unit $v \in \mathbb{R}^d$, set

$$p_i = z_i^2, \qquad x_i = (u_i^\top v)^2. \tag{241}$$

Then $0 \leq p_i \leq 1/2$, $\sum_i p_i = 1$, $\sum_i x_i = 2$, and

$$v^\top R_A(z) v = \frac{1}{2} - \frac{1}{2}\sum_i p_i x_i. \tag{242}$$

The maximum of $\sum_i p_i x_i$ over the capped simplex $0 \leq p_i \leq 1/2$, $\sum_i p_i = 1$, places mass $1/2$ on the two largest x_i. For any pair,

$$x_i + x_j \leq \lambda_{\max}(u_i u_i^\top + u_j u_j^\top) = 1 + |u_i^\top u_j| \leq 1 + \rho_A. \tag{243}$$

Thus $\sum_i p_i x_i \leq (1+\rho_A)/2$, and (242) gives

$$\nu_A \geq \frac{1-\rho_A}{4}. \tag{244}$$

For the other direction, choose i, j attaining ρ_A and choose $\epsilon \in \{\pm 1\}$ so that $u_i^\top(\epsilon u_j) = \rho_A$. Define

$$b = \frac{e_i - \epsilon e_j}{\sqrt{2}}, \qquad h = b^\top P_\perp b = \frac{1+\rho_A}{2}, \qquad z = \frac{P_\perp b}{\sqrt{h}}. \tag{245}$$

The vector z is a unit residual, and a direct coordinate calculation gives

$$z_i^2 = z_j^2 = \frac{1+\rho_A}{4}. \tag{246}$$

For

$$v = \frac{u_i + \epsilon u_j}{\sqrt{2(1+\rho_A)}}, \tag{247}$$

one has $x_i = x_j = (1+\rho_A)/2$. Keeping only these two nonnegative terms in (242) yields

$$\nu_A \leq \frac{1}{2} - \frac{(1+\rho_A)^2}{8} = \frac{(1-\rho_A)(3+\rho_A)}{8}. \tag{248}$$

Together, (244) and (248) prove (235).

M.3 Zero locus

Since $0 \le \rho_A \le 1$, the coherence sandwich gives $\nu_A = 0$ if and only if $\rho_A = 1$. In that case, two unit rows $u_i = \sqrt{2}a_i$ and $u_j = \sqrt{2}a_j$ are parallel up to sign. Thus there are a unit v and signs σ_i, σ_j with $a_i = \sigma_i v/\sqrt{2}$ and $a_j = \sigma_j v/\sqrt{2}$. Parsevalness and nonnegativity of the remaining energy give

$$\sum_k (a_k^\top v)^2 = \|Av\|_2^2 = 1 = (a_i^\top v)^2 + (a_j^\top v)^2, \tag{249}$$

$$a_i = \frac{\sigma_i}{\sqrt{2}} v, \quad a_j = \frac{\sigma_j}{\sqrt{2}} v, \quad a_k^\top v = 0 \quad (k \notin \{i,j\}). \tag{250}$$

Conversely, assume (250) and define z_* by

$$(z_*)_i = \frac{\sigma_i}{\sqrt{2}}, \qquad (z_*)_j = -\frac{\sigma_j}{\sqrt{2}}, \qquad (z_*)_k = 0 \quad (k \notin \{i,j\}). \tag{251}$$

Then $A^\top z_* = 0$, $\|z_*\|_2 = 1$, and $v^\top R_A(z_*) v = 0$: the only rows nonorthogonal to v have zero weight in (240). Since $R_A(z_*) \succeq 0$, $\nu_A = 0$. This proves the equivalence in Theorem 37.

For a uniform Parseval frame, the exact binary two-erasure formula of Bodmann & Paulsen (2005) gives, in the present normalization, binary error $(1+\rho_A)/2$ and retained binary lower bound $(1-\rho_A)/2$. It therefore has the same projective-duplicate singular boundary; the functional comparison is in Appendix U.

M.4 Leave-one-out near-contact corollary

At $m = 2d$ and $s = m-1$, the exact leave-one-out identity (141) with equal leverage gives

$$\begin{aligned} \frac{\overline{M}_{m-1}^{A,z}}{L^*} &= \frac{1}{d}\left(I_d - 2R_A(z)\right) = \alpha_{m-1}\left(I_d - 2R_A(z)\right), \\ 1 - \frac{\mathsf{T}_{\mathrm{spec}}^{\mathrm{RU}}(A, m-1)}{\alpha_{m-1}} &= 2\nu_A. \end{aligned} \tag{252}$$

Multiplying the critical-coherence sandwich (235) by two proves Corollary 38; $1-\rho_A$ is its coherence indicator, with no metric-distance interpretation.

M.5 Ordinary-volume sharp witness

Assume the zero geometry (250), fix $\theta \in \mathbb{R}^d$ and $L^* > 0$, and take the single residual z_* of (251), fixed before sampling for every budget $s = d + r$, $1 \le r \le d-1$. Every positive-volume set S contains at least one of i, j, since the other rows lie in $v^\perp$. Write $q_S = \mathbf{1}_{\{i\in S\}} + \mathbf{1}_{\{j\in S\}} \ge 1$ and $G_S = A_S^\top A_S$. Parsevalness and the pair geometry, followed by the normal equations, give

$$\sum_{k\notin\{i,j\}} a_k a_k^\top = I_d - vv^\top, \tag{253}$$

$$G_S v = \tfrac{q_S}{2} v, \qquad A_S^\top (z_*)_S = \tfrac{\mathbf{1}_{\{i\in S\}} - \mathbf{1}_{\{j\in S\}}}{2} v.$$

Since G_S is invertible on this support, $(w_S(z_*) - \theta)/\sqrt{L^*} = G_S^{-1} A_S^\top (z_*)_S = (\mathbf{1}_{\{i\in S\}} - \mathbf{1}_{\{j\in S\}})v/q_S \in \{0, \pm v\}$. Thus every supported error lies in $\operatorname{span}(v)$, and $\overline{M}_s^{A,z_*}/L^* = \lambda_s vv^\top$. As $v \in \ker R_A(z_*)$, Lemma 23 now gives $\lambda_s = \alpha_s$. Hence

$$\overline{M}_s^{A,z_*} = \frac{d-r}{d} L^* vv^\top = \frac{2d-s}{d} L^* vv^\top, \qquad \frac{\lambda_{\max}(\overline{M}_s^{A,z_*})}{L^*} = \frac{2d-s}{d}. \tag{254}$$

The same z_* works for every $r \in \{1, \dots, d-1\}$, giving the critical equal-leverage instance of the size-invariant contact kernel in Theorem 3. This support calculation is valid even when the repeated-pair design is outside row general position; no exact interior inverse-moment identity is extended to that boundary.

Equation (254) is a rank-one centered, full-Gram-whitened covariance under the fixed-response ordinary-volume law. Its directional contact leaves the trace-derived scalar-loss envelope strict. The ordinary-volume identities of Dereziński & Warmuth (2018b) and the scalar-loss obstruction of Dereziński et al. (2018) are compared with this witness in Appendix U.

N MATCHED EXACT PHASE ILLUSTRATION

N.1 MATCHED EQUAL-LEVERAGE FAMILY AND EXACT ENUMERATION

For every real t, define once and for all

$$\begin{aligned} G(t) &= \frac{1}{1+t^2}\begin{bmatrix} 1-t^2 & -2t \\ 2t & 1-t^2 \end{bmatrix}, \quad Q_t = \operatorname{diag}(G(t), G(t)), \\ A_t &= 2^{-1/2}\begin{bmatrix} I_4 \\ Q_t \end{bmatrix}, \qquad\qquad J_t = 2^{-1/2}\begin{bmatrix} I_4 \\ -Q_t \end{bmatrix}. \end{aligned} \tag{255}$$

Thus $A_t^\top A_t = J_t^\top J_t = I_4$, $A_t^\top J_t = 0$, and J_t is an orthonormal basis for $\ker(A_t^\top)$. Every row of A_t has leverage $1/2$. The row and column order in this display is used throughout the matched-family calculations below; Figure 4 uses the subdomain $t \in [0, 1]$.

Proposition 39 (Exact margin of the matched rotation family)**.** *For every real t,*

$$\nu_{A_t} = \frac{1 - |\cos(4\arctan t)|}{8} = \frac{1}{8}\left(1 - \frac{|1 - 6t^2 + t^4|}{(1+t^2)^2}\right). \tag{256}$$

In particular, $\nu_{A_0} = 0$ and $\nu_{A_{1/2}} = 9/100$.

Proof. Every unit compatible residual is uniquely $z = J_t v$ for a unit $v \in \mathbb{R}^4$. With $D_x = \operatorname{diag}(x_i^2)$, substitution into (7) gives

$$R_{A_t}(J_t v) = \frac{1}{2} I_4 - \frac{1}{4}\left(D_v + Q_t^\top D_{Q_t v} Q_t\right). \tag{257}$$

For unit u, v, set

$$\begin{aligned} H(u, v) &= \sum_i u_i^2 v_i^2 + \sum_i (Q_t u)_i^2 (Q_t v)_i^2, \\ F(u) &= H(u, u). \end{aligned}$$

Cauchy–Schwarz applied to the concatenated squared-coordinate vectors gives $H(u, v) \le \sqrt{F(u)F(v)} \le \max_{\|w\|=1} F(w)$, with equality at $u = v = w$ for a maximizer. Hence

$$\nu_{A_t} = \frac{1}{2} - \frac{1}{4}\max_{\|w\|=1} F(w). \tag{258}$$

Split w across the two identical rotation blocks, with squared norms $r_1 + r_2 = 1$, and let $F^{(2)}_{\max}$ be the maximum on one unit block. Degree-four homogeneity gives $F(w) \le F^{(2)}_{\max}(r_1^2 + r_2^2) \le F^{(2)}_{\max}$. Equality follows by placing all mass on a one-block maximizer. Writing $w = (\cos\phi, \sin\phi)$ there and $\theta = 2\arctan t$,

$$F(w) = \frac{3}{2} + \frac{1}{2}\cos(2\theta)\cos(4\phi + 2\theta). \tag{259}$$

The maximum is $3/2 + |\cos(2\theta)|/2$. Substitution in (258), followed by the tangent half-angle identity, proves (256). At $t = 1/2$, the equality direction $w = (3, 1, 0, 0)^\top/\sqrt{10}$ gives $F(w) = 41/25$ and hence $\nu = 9/100$. □

For the family in Figure 4, $\theta = 2\arctan t$ and the critical coherence is

$$\rho_{A_t} = \max\{|\cos\theta|, |\sin\theta|\}. \tag{260}$$

The critical-frame theorem therefore gives the pair floor $(1 - \rho_{A_t})/4$, while the maximum-leverage bound (185) gives zero for every t. Panel (a) in Figure 4 computes the exact margin ν_{A_t}, not $\mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A_t, s)$; panel (b) separately shows exact covariance targets and upper envelopes.

We now preserve the complete support calculation for the two manuscript members $t \in \{0, 1/2\}$ under ordinary volume sampling and selected unweighted OLS at $s \in \{5, 6, 7\}$. There are 92 candidate indexed subsets across these budgets for each design. The positive/zero-volume support counts are

design	$s=5$	$s=6$	$s=7$
$t=0$	$32/24$	$24/4$	$8/0$
$t=1/2$	$48/8$	$28/0$	$8/0$

The ordinary determinant normalizers are $4, 6, 4$, respectively. Exact support-only calculation gives probability mass one and centered selected-OLS unbiasedness on each row. Zero-volume sets receive probability zero and no inverse or selected fit.

For $t = 0$, take $z_* = J_0 e_1$. With $L^* = 1$, exact enumeration gives

$$\overline{M}_s^{A_0, z_*} = \alpha_s e_1 e_1^\top, \qquad \alpha_s = (8-s)/4, \qquad s = 5, 6, 7. \tag{261}$$

This is directional contact, not full-matrix equality with $\alpha_s I_4$. For $t = 1/2$, Proposition 39 gives $\nu_{A_{1/2}} = 9/100$. The linear bound in (8) gives $\mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}/\alpha_s \le 1 - [(s-d)/(m-d-1)]\nu_A$. Componentwise for $s = (5, 6, 7)$,

$$\left(\frac{\mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A_{1/2}, s)}{\alpha_s} \right)_{s=5,6,7} \le \left(\frac{97}{100}, \frac{47}{50}, \frac{91}{100} \right). \tag{262}$$

These are response-uniform covariance upper bounds, not optimized covariance values. Using only the pair floor $1/20$ gives the looser ratio vector $(59/60, 29/30, 19/20)$. The support and contact calculations use exact rational arithmetic; no residual-grid maximization is used.

N.2 Exact covariance targets in the matched rotation family

After simultaneous row and column permutation, the family in Appendix N.1 is a direct sum of two copies of $B_t = 2^{-1/2}[I_2; G(t)]$. A local complement coordinate $v \in \mathbb{R}^2$ represents the actual residual $r(v) = 2^{-1/2}[v; -G(t)v] \in \mathbb{R}^4$, with $\|r(v)\|^2 = \|v\|^2$. Writing $b_i^\top$ for the local design rows, define

$$\mathcal{W}_t(v) := \sum_{i=1}^{4} r_i(v)^2 b_i b_i^\top = \frac{1}{4}\{\mathrm{diag}(v_i^2) + G(t)^\top \mathrm{diag}((G(t)v)_i^2) G(t)\}, \qquad \mathrm{tr}\, \mathcal{W}_t(v) = \frac{1}{2}\|v\|^2.$$

For $0 < t < 1$, every two-row local basis has positive volume. The local covariances at sizes $2, 3, 4$ are respectively $\|v\|^2 I_2$, $\mathcal{W}_t(v)$ (by (141)), and zero. The local volume normalizers $1, 2, 1$ give allocation weights $(1/2, 1/2)$ at global size 5, $(1/6, 4/6, 1/6)$ at size 6, and $(1/2, 1/2)$ at size 7. Conditional on the two sizes, local sampling factorizes and both fits are unbiased, so cross-block covariances vanish. For residual coordinates $v^{(j)}$ with energies $E_j = \|v^{(j)}\|^2$, this proves

$$\begin{gathered} \overline{M}_s = \mathrm{diag}_{j=1,2}\{a_s E_j I_2 + c_s \mathcal{W}_t(v^{(j)})\}, \\ (a_5, c_5) = (1/2, 1/2), \quad (a_6, c_6) = (1/6, 2/3), \quad (a_7, c_7) = (0, 1/2). \end{gathered} \tag{263}$$

Homogeneity places a response-uniform spectral optimizer in one block. Proposition 39 gives $\max_{\|v\|=1} \lambda_{\max}(\mathcal{W}_t(v)) = 1/2 - \nu_{A_t}$; substitution into (263) yields

$$\mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A_t, s) = \alpha_s - 2\beta_s \gamma_s \nu_{A_t}, \qquad s \in \{5, 6, 7\}, \quad 0 < t < 1. \tag{264}$$

At $t = 0, 1$, $\nu_{A_t} = 0$ and the separate common-witness result supplies the equality in (264).

At $t = 1/2$, $\nu_{A_t} = 9/100$, so the exact targets are

$$\mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A_{1/2}, 5) = 141/200, \quad \mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A_{1/2}, 6) = 11/25, \quad \mathsf{T}^{\mathrm{RU}}_{\mathrm{spec}}(A_{1/2}, 7) = 41/200. \tag{265}$$

Their ratios to α_s are $47/50$, $22/25$, and $41/50$. The common normalized residual direction $z = J_{1/2}(3, 1, 0, 0)^\top/\sqrt{10}$ and its matching covariance-eigenvector direction attain these three values.

N.3 Equal-leverage leave-one-out separation

For no-coloop $A^\top A = I_d$, the common LOO identity (141) specializes under equal leverage ℓ to $(m-d)^{-1}[I_d - R_A(z)/(1-\ell)]$. Optimizing its top eigenvalue over compatible unit residuals gives the spectral counterpart of Proposition 25.

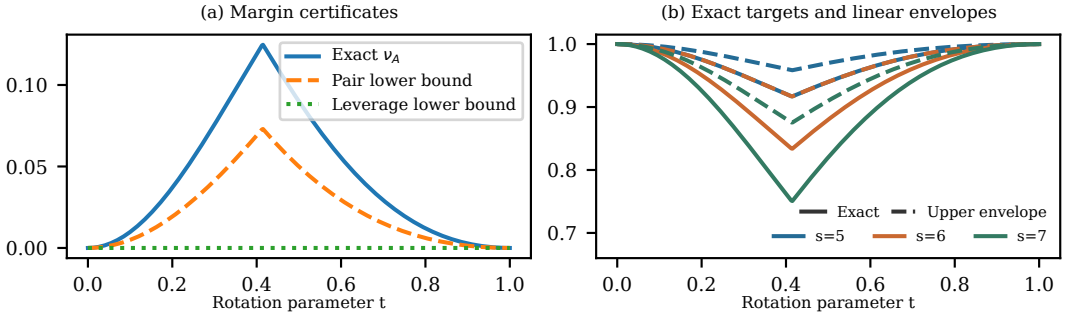


Figure 4: **Contact, exact covariance targets, and linear envelopes at fixed leverage.** For the 8×4 family in Appendix N.1, t is a rotation parameter and all row leverages are $1/2$. (a) The exact margin and Pair certificate vary, while the leverage bound is zero. (b) Solid curves are exact response-uniform target/ceiling ratios; dashed curves are their linear upper envelopes.

For a union of $k \geq 2$ orthobases scaled by $k^{-1/2}$, every leverage is $1/k$. If μ is the largest cross-basis absolute inner product, direct frame algebra gives

$$\rho_\perp = \frac{\mu}{k-1}, \qquad \kappa_\perp = \left(\frac{k-1}{k}\right)^2, \qquad t_\perp - t_{\mathrm{lev}} = \frac{1-(k-1)\mu}{k^2}. \tag{266}$$

Thus pair information is quantitatively stronger whenever $\mu < 1/(k-1)$, including regimes with positive maximum-leverage certification.

For $k = 3, d = 16$, compare three identical repeated row orthobases with three real mutually unbiased row orthobases. Write $A = 3^{-1/2}[I; H; HD]$ for the latter. Both have $(m, d, s) = (48, 16, 47)$ and the complete leverage vector $1/3$. In the repeated construction the rows are $e_j/\sqrt{3}$. Compatibility makes each coordinate triple sum to zero; if E_j is its residual energy, the residual operator is diagonal with entry $2/3 - E_j/3$. Thus its smallest eigenvalue is at least $1/3$, and a unit sum-zero residual supported on one triple attains equality. Hence $\nu_A = 1/3$ and the exact target is $1/64$, or ratio $1/2$ to α_{47}.

For completeness, take $U_1 = I, U_2 = H$, and $U_3 = HD$, where $H_{xy} = (-1)^{x\cdot y}/4$ for $x, y \in \{0,1\}^4$ and $D_{yy} = (-1)^{y_1y_2+y_3y_4}$. The Walsh transform of this bent sign vector has absolute value 4, so direct cross-inner-product calculation gives $\mu = 1/4$. This mutually unbiased construction has $t_\perp = 7/18 > t_{\mathrm{lev}} = 1/3$, and therefore target at most $5/384$, or ratio at most $5/12$, versus the exact repeated-design target $1/64$. This is at least one sixth smaller relative to the repeated-basis target. At tolerance $1/2$, on this incoherent construction the pair-route resolvent first pass is 28 and the maximum-leverage-route first pass is 29. Repeated bases and mutually unbiased frames are classical erasure geometry (Bodmann & Paulsen, 2005; Fickus & Mixon, 2012); the displayed consequence uses the response-uniform ordinary-volume covariance target of this paper.

N.4 Exact trace and fixed-query loss separation

For $0 < t < 1$, write $c = (1-t^2)/(1+t^2)$, $h = 2t/(1+t^2)$ and $J_t = 2^{-1/2}[I_4; -Q_t]$. Every compatible residual is $J_t v$, with $L^* = \|v\|^2$. Taking the trace in the common block identity (263) and using $\operatorname{tr} \mathcal{W}_t(v^{(j)}) = E_j/2$ proves, for every nonzero fixed residual,

$$\frac{\operatorname{tr} \overline{M}_s}{L^*} = \begin{cases} 5/4, & s = 5, \\ 2/3, & s = 6, \\ 1/4, & s = 7, \end{cases} \qquad 0 < t < 1. \tag{267}$$

At $t = 0, 1$, the rotation is a signed permutation and the local rank-size covariance becomes $\operatorname{diag}(v_i^2)$; vanished zero-volume bases are omitted. Hence $\overline{M}_s = \alpha_s \operatorname{diag}(v_i^2)$ and the trace ratio is $3/4, 1/2, 1/4$ at $s = 5, 6, 7$. The first two second moments therefore do not continue across the changing support.

For the fully matched comparison, fix $X = A_t$ and the same numeric response $y = \sqrt{2}[e_1; 0_4]$. Direct multiplication gives $w^* = e_1$ and residual $J_t e_1$, of energy one, for every t. At $s = 7$, each

omitted row has probability $(1-1/2)/(8-4) = 1/8$. Put $p = c^2h^2$; the nonzero covariance block is

$$\overline{M}_7\big|_{1:2,1:2} = \frac{1}{8}\begin{bmatrix} 2-2p & ch(h^2-c^2) \\ ch(h^2-c^2) & 2p \end{bmatrix}, \qquad \lambda_{\max}(\overline{M}_7) = \frac{1+\sqrt{1-3p}}{8}. \tag{268}$$

At $t_1 = 1/4$ and $t_2 = 1/2$, $p_1 = 14400/83521 < p_2 = 144/625$, so the covariance maximum strictly decreases. The feature margins are $16/289$ and $9/100$; the separate response-uniform spectral targets are $257/1156$ and $41/200$.

Tracing the fixed-response covariance against the three fixed query metrics gives the exact SSE increments and rounded values in Table 3. The raw response, residual energy, Full coefficient, each query, dimensions, size, leverage vector, and entire indexed sampling distribution are fixed across designs. The design and its residual projection vary. Every fixed numerical query target is allowed in (5); the Full query predictions are unchanged. At sizes $5, 6$, the trace identities still hold but the indexed probability vectors vary with t.

Exhaustive rational enumeration of the positive-volume indexed support verifies probability mass, unbiasedness, covariance quadratic coefficients, and the query traces. This directly computes both the fixed-response and response-uniform spectral values.

Table 3: **Matched designs, different query behavior.** Expected query excess SSE at $s = 7$. Each design cell shows the exact fraction above its six-decimal rounded value; change is relative to $t = 1/4$.

Z	$t = 1/4$	$t = 1/2$	Change
	$1/4$	$1/4$	
I_4	(0.250000) 72721/83521	(0.250000) 517/625	+0.00%
$\mathrm{diag}(2,1,1,1)$	(0.870691) 126721/334084	(0.827200) 1057/2500	−5.00%
$\mathrm{diag}(1,2,1,1)$	(0.379309)	(0.422800)	+11.47%

O Exact controls for alignment and moment bounds

O.1 Same spectrum, different query risk

This six-row design with a three-row contact class generates Figure 5 and the matched rank-two control in Appendix Q.1:

$$A = \begin{pmatrix} 2/3 & 2/3 & 0 \\ 1/3 & -2/3 & 0 \\ -2/3 & 1/3 & 0 \\ 0 & 0 & 2/3 \\ 0 & 0 & -2/3 \\ 0 & 0 & 1/3 \end{pmatrix}, \qquad N = \begin{pmatrix} 1/3 & 0 & 0 \\ 2/3 & 0 & 0 \\ 2/3 & 0 & 0 \\ 0 & 1/3 & 2/3 \\ 0 & 2/3 & 1/3 \\ 0 & 2/3 & -2/3 \end{pmatrix}.$$

Here $A^\top A = N^\top N = I_3$, $A^\top N = 0$, $C = \{1,2,3\}$, and $V_C = \mathrm{span}(e_1, e_2)$. Let $X = A$, $z_C = N_{:,1}$, $w = N_{:,2}$, and $\Xi = (e_1+e_2)(e_1+e_2)^\top/2$. Compatible residuals make the Full coefficient zero. For $z_t = (z_C + tw)/\sqrt{1+t^2}$, the LOO states $t = 0, 1/4, 1$ have $(\varepsilon_z, g_5^{(1)}) = (0,0), (1/17, 1/17), (1/2, 1/2)$. This realizes (20), which requires neither equal leverage nor the joint-local neighborhood. Appendix Q.2 gives broader perturbation checks.

For the multi-output comparison, fix $s = 5$, $K = 2$, and

$$E_A(a) = [z_C, aw], \qquad E_B(a) = [az_C, w], \qquad a \in \{0, 1/3, 1/2, 1\}.$$

Each pair has the same design, query, law, size, output count, total energy $1 + a^2$, singular spectrum $\{1, a\}$, and $\zeta_E = a^2/(1+a^2)$. At $a = 1/3$, $(1 - g_A^{\mathrm{agg}}, 1 - g_B^{\mathrm{agg}}) = (9/10, 1/10)$, so spectrum and energy do not determine fixed-query risk. For $a < 1$, the residual matrices and raw responses differ; only their spectra match. The pooled residual operator distinguishes the arms, so this does not show that second moments are necessary or that a learning algorithm improves.

Exact values sum over positive-volume subsets with $K_S = A_S^\top A_S$,

$$p_A(S) = \frac{\det K_S}{\binom{m-d}{s-d}}, \qquad W_S = K_S^{-1} A_S^\top E_S.$$

Normalized risk is $\operatorname{tr}[\Xi \sum_S p_A(S) W_S W_S^\top]/(\alpha_s L_{\mathrm{src}} \operatorname{tr} \Xi)$. Zero-volume subsets receive no inverse or fit. Figure inputs retain the exact rationals. For the exact-construction figures, $Q_s = \overline{M}_s/(\alpha_s L^*)$ denotes normalized covariance, not the LOO deficit.

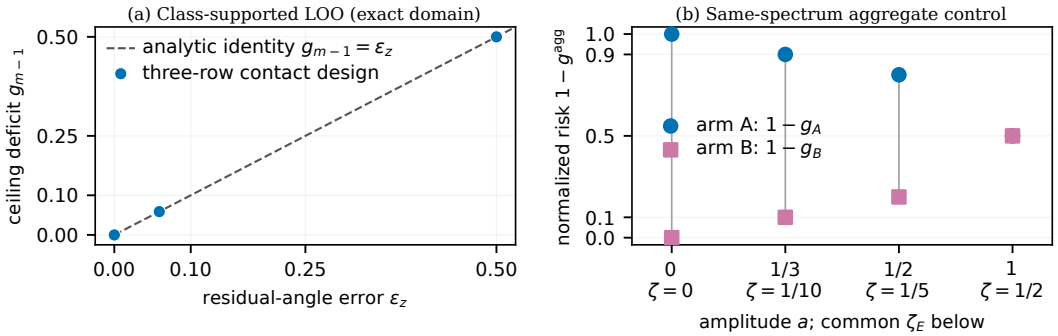


Figure 5: **Alignment and retained information.** (a) At $s = m - 1 = 5$, class-supported LOO deficits equal the residual-angle error ε_z (dashed). (b) At each a, both arms share the design, query, sampling law, size, energy, and spectrum; energy and spectrum change across a.

O.2 A SINGLETON-EXCLUSION BOUNDARY CHECK

For $m = d + 2 = 4$, $s = 3$, and one output, take

$$A = \frac{1}{3}\begin{pmatrix} 1 & 2 \\ 2 & -1 \\ 2 & 0 \\ 0 & 2 \end{pmatrix}, \qquad Y = \frac{1}{3\sqrt{5}}\begin{pmatrix} -2 \\ -4 \\ 5 \\ 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Here $A^\top A = I$, $A^\top Y = 0$, and Full is zero. The ordinary law gives expected query MSE; the analysis-only augmented law gives bound moments:

$U_{\mathrm{agg}} = U_{\mathrm{per}}$	U_2	Expected query MSE	$U_2 - \mathrm{MSE}$
0.468190	0.418307	0.388889	0.029418

Values are rounded query MSE (query SSE divided by $qK = 2$). This tests the singleton-exclusion branch with noncommuting H, V. The positive gap is analytic slack, not floating-point error; the finite enumeration is not a general proof.

P EXACT CONSTRUCTIONS FOR FIGURES 1 AND 2

Figure 1(a–c) uses two contact constructions. Its panel (d) and Figure 2(a–b) use three other designs. The two 8×3 designs are different. Only subsets with positive Gram determinant enter the supported-fit sum; zero-volume subsets have probability zero and no selected inverse.

The two contact constructions. For Figure 1(a), use the design and size in (143) with $z = 2^{-1/2}(u, -u)^\top$, $\|u\|_2 = 1$. The exact covariance in (144) and $\alpha_3 = 1/2$ give $Q_3 = \overline{M}_3/(\alpha_3 L^*) = \operatorname{diag}(u_1^2, u_2^2)$. Thus $Q_{11} + Q_{22} = 1$. Different residuals attain the endpoints $(1, 0)$ and $(0, 1)$, but none attains the ceiling corner $(1, 1)$.

Figure 1(b–c) uses the six-row, three-column Parseval frame A, N displayed in Appendix O.1. Set $X = A$ and take the compatible unit residual $z_C = N_{:,1}$. Its leave-one-out normalized covariance is exactly

$$Q_5 = \operatorname{diag}(1, 1, 0), \qquad V_C = \operatorname{span}(e_1, e_2).$$

In panel (b), the covariance disk is $\{Q_5^{1/2}u : u \in \mathbb{R}^3,\ \|u\|_2 \le 1\}$ for this fixture, with contact space V_C. For the strictly interior ellipsoid in the same panel, take $w = N_{:,2}$ and $z_{\mathrm{int}} = (z_C + w/2)/\sqrt{5/4}$. Its normalized covariance is $Q_5^{\mathrm{int}} = \mathrm{diag}(4/5, 4/5, 1/10) \prec I_3$. Panel (c) uses the same A, z_C: a query range contained in V_C reaches the whole-query ceiling, while one with a component outside V_C does not. These shapes are specific to this fixture.

The signed 5×2 path. Figure 2(a) fixes the raw response $y = (1, -1, 0, 0, 0)^\top$ and query $Z = (1, 0)$, and changes only the design:

$$X_t = \begin{pmatrix} 1 & 0 \\ 1 & t \\ t & 1 \\ 0 & 1 \\ -t & 1 \end{pmatrix}, \qquad t \in \{0, \pm 1/64, \pm 1/32, \pm 1/16, \pm 1/8\}.$$

At each signed t, we recomputed full OLS, its residual, the query scale, and the ordinary indexed size-s volume law in exact rational arithmetic. At $t = 0$, one pair deletion loses rank and $g_3 = g_4 = \delta_{\mathrm{LOO}} = 0$. Each tested $t \neq 0$ has full rank after every pair deletion and has no nonzero contact direction. The two signs are distinct exact states; their deficits agree by symmetry. The single zero-volume size-three subset at $t = 0$ is excluded from the supported-fit sum. A separate 5×2 noncontact control in Table 4 uses $X = [(1, 0); (0, 1); (1, 1); (1, 2); (2, 1)]$, $y = (1, -1, 2, 0, 1)^\top$, and $Z = (1, 2)$. Here $m = 5$, $d = 2$, the displayed interior size is $s = 3$, the LOO size is 4, and $b_3 = 1/4$. All ten size-three and five LOO subsets are supported. Exact enumeration gives the strict order $Q_4 \prec Q_3 \prec I_2 - \frac{1}{4} D_{\mathrm{LOO}} \prec I_2$.

The separate all-size state. Figure 2(b) uses

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 3 \\ 2 & 1 & 1 \end{pmatrix}, \quad y = (1, -1, 2, 0, 1, -2, 3, 1)^\top, \quad Z = (1, 2, -1).$$

Here $d = 3$, $m = 8$, and $\delta_{\mathrm{LOO}} = 645665506289/703030801760 > 0$. For the four interior sizes $s = 4, 5, 6, 7$, the lower ratios b_s are $1/16, 1/6, 3/8, 1$; at $s = 7$ the actual ratio and both endpoints equal one. This constructed example was not preregistered and is not a representative sample of designs.

Enumeration and comparison. For each raw full-rank design, write $G = X^\top X$, let $\widehat{\beta}_F$ be the full fit, and set $e = y - X\widehat{\beta}_F$. At size s, let $p_X(S) = \mathbb{P}_X(S)$ be the ordinary volume probability in (2). With $\alpha_s = (m - s)/(m - d)$, the exact one-row query risk and ceiling are

$$R_s = \sum_{S : p_X(S) > 0} p_X(S) \big[Z(\widehat{\beta}_S - \widehat{\beta}_F) \big]^2, \qquad C_s = \alpha_s \|e\|_2^2 \, ZG^{-1}Z^\top, \qquad g_s = 1 - R_s/C_s.$$

Figure 2 and the table use $b_s = (s - d)/[(m - s)(m - d - 1)]$ and $\delta_{\mathrm{LOO}} = g_{m-1}$. The old resolvent lower column comes from $\theta_s h/(1 - \gamma_s h)$, where $\theta_s = (s-d)/(m-d-1)$, $\gamma_s = (m-s)/(m-d-1)$, and $h = \sum_i (c_i - z_i^2) u_i$ in whitened coordinates: $c_i = 1 - \|a_i\|_2^2$, $z = e/\|e\|_2$, and $u_i = a_i^\top \Xi a_i / \operatorname{tr} \Xi$. Here a_i is row i of $XG^{-1/2}$ and Ξ is the corresponding whitened query metric. In the non-LOO rows below, the old lower is stronger than the basic $b_s \delta_{\mathrm{LOO}}$ lower. These comparisons do not show that the new bound always improves the old one.

Table 4: **Exact finite deficits and two lower bounds.** Values are rounded; comparisons use exact rational values. Each $\pm t$ row represents two separately enumerated states. "Old lower" is the resolvent deficit bound, not a risk bound.

Case	s	$b_s\delta_{\mathrm{LOO}}$	g_s	δ_{LOO}	Old lower
$5 \times 2, t = 0$	3	0	0	0	0
$5 \times 2, t = \pm 1/64$	3	.000132	.000285	.000529	.000156
$5 \times 2, t = \pm 1/32$	3	.000528	.001138	.002112	.000623
$5 \times 2, t = \pm 1/16$	3	.002100	.004532	.008400	.002488
$5 \times 2, t = \pm 1/8$	3	.008215	.017838	.032861	.009865
Separate 5×2	3	.214353	.626314	.857412	.331202
Separate 8×3	4	.057400	.661743	.918403	.360663
	5	.153067	.832317	.918403	.530129
	6	.344401	.888738	.918403	.628579
	7	.918403	.918403	.918403	.692920

Figure 1(d) uses this separate illustrative 8×3 design and response:

$$X = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 2 & 0 \\ 2 & -2 & 2 \\ -1 & -1 & -2 \\ -2 & 2 & 2 \\ -1 & 1 & -2 \\ 1 & 2 & 1 \\ 2 & -1 & -2 \end{pmatrix}, \qquad y = (0, 3, 4, 4, -2, -3, -3, 3)^\top.$$

Every two-row deletion leaves full column rank, so there is no nonzero contact class. All 28 size-six and eight LOO subsets are supported. Exact enumeration gives $Q_7 \prec Q_6 \prec I_3 - \frac{3}{8}(I_3 - Q_7) \prec I_3$. The figure projects these covariance bodies onto one plane; its silhouettes and surface guides come from the matrices, not fitted drawing primitives.

The accompanying source gives the 24-row rational table, raw designs and responses, subset probabilities, supported fits, and pair-deletion checks. The numerical verifier recomputes the deficit fractions. The exact pilot checked 11 design states, 24 size states, 311 supported fits, and one zero-volume subset excluded from the selected-fit sum. These finite checks illustrate the theorem; they do not prove the all-design statement or a perturbation law between the marked states.

GROUP IV. EXPERIMENTAL PROTOCOLS AND EVIDENCE

Q CONTACT AND INFORMATION-DEPENDENT CERTIFICATION EXPERIMENTS

Every subset fit uses ordinary indexed fixed-size volume sampling and unweighted OLS. The augmented law supplies certificate moments only. Fixed grids and allowances select sizes through directed endpoints; evaluation does not change them (Table 5).

Gas and Tetouan study registry. Gas Turbine (UCI Machine Learning Repository, n.d.b) uses 2011–2013 rows as one source ($m = 22{,}191$), an intercept and nine source-standardized predictors ($d = 10$), and raw CO/NOx outputs ($K = 2$). *Gas whole query* uses all 2014–2015 rows ($q = 14{,}542$) for LOO calibration. *Gas positional blocks* split each query year in order into six blocks and interleave them by year into twelve requests with different q_g. They serve the policy, cost, and matched-risk studies; they are neither months nor independent datasets. Tetouan (UCI Machine Learning Repository, n.d.c) uses the earliest 39,312 rows as source and the latest 13,104 as query. Its three outputs are scaled by their source population SD. *Tetouan weather* uses an intercept and five weather predictors ($d = 6$) for LOO calibration. The *Tetouan selected model* uses source-selected quadratic weather plus calendar features ($d = 19$, $K = 3$) for size selection and matched risk. All transforms use source data only. The LOO studies use thin-QR Parseval coordinates, and both designs pass full-rank, no-coloop, and domain checks.

Table 5: **Information, decision, and evidence.** Source responses and query labels enter only where stated. An allowance based on source variance is not a lower bound on query loss.

Study	Information used	Quantity / decision	Evidence
Contact and stability	Public-derived X; observed or engineered residual; fixed query	Contact kernel / normalized deficit	Exact small panels; numerical MovieLens alignment
Multi-output LOO	Source X, Y; Full fit; query Z	Energy slack and LOO displacement	Exhaustive float64 at LOO; bounds only elsewhere
Bike	Source X; query features and workload labels; no source or query responses	Response-uniform normalized threshold τ	Directed first pass
Tetouan	Source-only model selection; source Y, Full fit, and query Z	Aggregate source-variance allowance	Directed first pass; 256 paths; query labels only for final score
Gas	Source Y, Full fit, and 12 query-feature blocks; query Y never read	Same allowance; marginal row/cost comparison	Directed first pass; five timing repeats; 256 shared paths

For response-aware selection, the allowance uses source population variance: $\epsilon = \eta K^{-1} \sum_j \mathrm{Var}_{\mathrm{src}}(Y_j)$, not Full query MSE. At $\eta = .01$, it is about 0.6353156262 for Gas and 0.01 for source-standardized Tetouan; matched-risk issuance uses exact rational allowances. The three separate evaluations use *earlier selected-Tetouan paths* (256 paths at 11,757 rows), *earlier Gas block paths* (256 shared paths over a 108-size union), and *matched-risk paths* (256 independent paths per source for 48 Gas and five Tetouan cells). Their estimates are not pooled.

Q.1 Public contact construction and matched query control

Pontius is a 40-row NIST load-cell record (National Institute of Standards and Technology, n.d.). With $X = [1, x/(1.5 \times 10^6), (x/(1.5 \times 10^6))^2]$, all 780 row-pair deletions retain rank; four fixed queries at $s = 39$ have exact deficits 0.968–0.973. A derived one-hot model uses 20 first-seen load levels, each observed twice. Its 20 size-two classes admit contact, but the observed residual has zero contact-kernel dimension at $s = 21, 30, 39$. The first-class query $2e_0$ has exact deficit $17354/18443$ at each size. A compatible unit-witness response instead gives a one-dimensional kernel and zero deficit at all three sizes. The one-hot model does not claim better load-cell prediction.

For IEEE-14 (MATPOWER, n.d.), removing the leaf edge and grounding the connected incidence design gives $m = 19, d = 12$. A size-three complement class has a two-dimensional contact space. An engineered class-witness response has zero deficit for two rank-one directions and their equal-weight rank-two query at $s = 16, 17, 18$. These topology controls are not measured responses. With $H = X^\top X$, a raw attaining query lies in $H \ker X_{-C}$, generally not $\ker X_{-C}$.

For the MovieLens 100K `u1.base` graph (GroupLens Research, n.d.), bridge filtering and a grounded additive incidence map give $m = 79{,}859, d = 2{,}451$. The 76 degree-two pair witnesses are not a full census of maximal classes. Their residual-angle errors are $0.999862/0.999994/0.9999999997$ (minimum/median/maximum), so squared alignment is at most 0.0138%. Fitted-value relative disagreement is 2.02×10^{-12}, and the normal-equation infinity norm is 1.25×10^{-9}. These numerical checks do not prove noncontact. Queries summarize fitted values on retained training edges, not held-out recommendation performance.

The matched rank-two control uses the three-row contact class in Appendix O. At $s = 4, 5$, it fixes the design and witness residual $z_C = N_{:,1}$ and compares $\Xi_{\mathrm{in}} = \mathrm{diag}(1/2, 1/2, 0)$ with $\Xi_{\mathrm{out}} = \mathrm{diag}(1/2, 0, 1/2)$. Both have rank two, trace one, and spectrum $(1/2, 1/2, 0)$. Their normalized risks are 1 inside the class space and $1/2$ outside. Thus only containment changes.

Q.2 Fixed-design contact stability

The stability experiment fixes the Pontius first one-hot class, IEEE size-three class, feature matrices, source order, and contact sizes. It uses the residual grid $\{0, \delta_C/256, \delta_C/64, \delta_C/16, \min(1/4, \delta_C/4), 1\}$ and query grid $\{0, 1/64, 1/16, 1/4, 1/2, 3/4, 1\}$.

The local lower bound applies only when $\varepsilon_z \leq \delta_C/16$ and actual $\varepsilon_\Xi \leq 1/2$. In the IEEE equal-weight rank-two slice, actual ε_Ξ is half the query-grid value.

Of 504 states, 264 lie in the local domain and 252 have positive lower endpoints. All exact inequalities hold. The 18 class-supported rank-one LOO states also satisfy the exact angle law. Positive g/L_{wr} spans 7.11–352.45. IEEE uses $\underline{\sigma}_C^2 = 927/9646 \approx 0.09610$ instead of the restricted value ≈ 0.71144, a 7.40 factor. At the origin, $g = L_{\mathrm{wr}} = U = 0$. These finite-grid checks do not test feature-shift robustness or fitted convergence. Upper-utilization ranges and medians are in `geometry/reproducibility/supplementary_experimental_details.md`.

The local bound in Figure 6 is conservative, especially with IEEE's certified constant enclosure. The full 504-state record, including undisplayed sizes and directions, accompanies a numerical verifier.

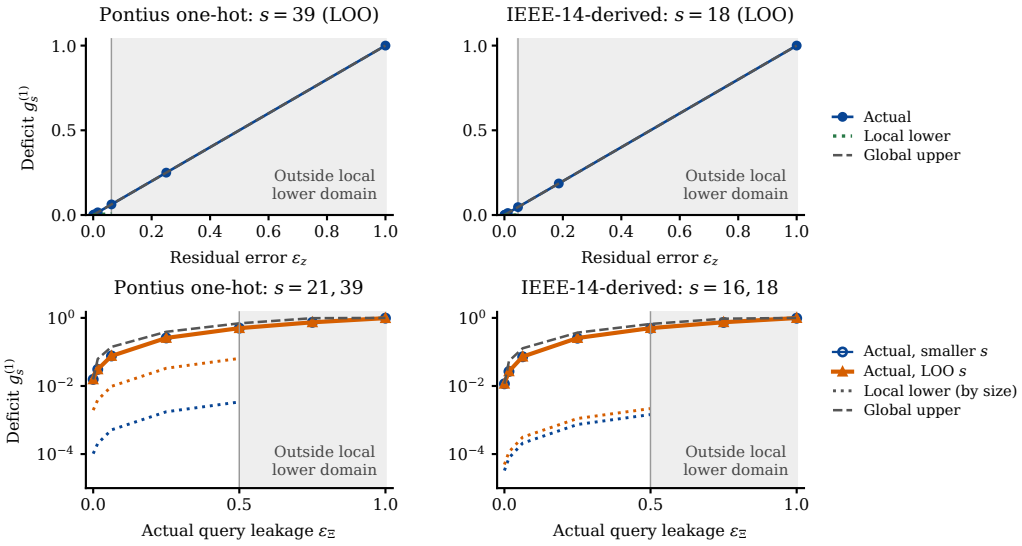


Figure 6: **Fixed-design contact stability.** Top: residual-only LOO slices at zero query leakage (linear scale, including the zero origin). Bottom: query leakage at fixed nonzero $\varepsilon_z = \delta_C/64$ for the indicated actual sizes (log scale). Curves show exact deficits, applicable local lower bounds, and global upper bounds. Shading excludes the local lower bound when $\varepsilon_z > \delta_C/16$ (top) or actual $\varepsilon_\Xi > 1/2$ (bottom).

Q.3 Natural multi-output slack and LOO calibration

For Gas whole query and Tetouan weather, the residual energy fractions are 0.0278083 and 0.1194896, and the query factors are 0.9983001 and 0.9998103. At LOO, $\vartheta_{m-1} = 1$, giving the floors in Table 6. Each 20-size analytic grid includes $d+1, d+2, 2d$, fixed fractions, $m-1$, and m. Actual expected displacement is computed only at LOO, from all 22,191 and 39,312 supported deletions in float64. Independent weighted-row calculations and three direct refits per task agree.

At LOO, $D = 1.6613854 \times 10^{-5}$ for Gas and 2.1245730×10^{-9} for Tetouan. Gas has a resolved $U_2 - D = 1.08657 \times 10^{-6}$. Tetouan's $\approx 3.867 \times 10^{-12}$ gap is below the frozen $\approx 10^{-10}$ tolerance, so its 1.00182 ratio is descriptive. At non-Full sizes, pooled-to-per-output gains are 4.85×10^{-10}–3.66×10^{-5} for Gas and at most 0.0115% for Tetouan. LOO separation is unresolved for both; these gains are much smaller than the second-moment corrections.

Q.4 Bike: response-uniform workload decisions

Bike Sharing (UCI Machine Learning Repository, n.d.a) uses 8,645 source rows from 2011 and 8,734 query rows from 2012. Its fixed $d = 20$ map has an intercept; source-standardized temperature, apparent temperature, humidity, and windspeed; first and second hourly sine/cosine harmonics; three season and two weather indicators; workingday and holiday flags; and four workingday–hour-

Table 6: **Multi-output slack and LOO bound calibration.** The floor is a lower bound on the unused ceiling fraction. Ratio denominators are actual expected Full-query-MSE excess at LOO, computed over all supported deletions in float64. †: Tetouan's gap is below numerical tolerance, so its ratio is descriptive.

Fixed model (d, K)	rank E	LOO deficit floor	$U_{\mathrm{per}}/\mathcal{D}_{m-1}$	$U_2/\mathcal{D}_{m-1}$
Gas $(10, 2)$	2	0.027761	17.8833	1.06540
Tetouan-weather $(6, 3)$	3	0.119467	7.65795	1.00182†

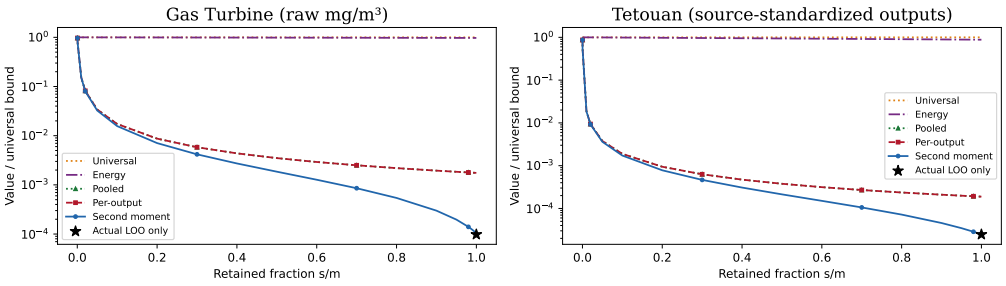


Figure 7: **Information-dependent bounds.** Both panels use log scales and omit zero-valued Full. The black LOO stars are the only actual expected risks; all curves are bounds. Energy uses its own ceiling and lies outside the ordered certificate family.

harmonic interactions. The eight workloads cross season with workingday. Certification reads only predictors, never casual, registered, or total counts. Corollary 7 removes the unknown positive residual scale without estimating it from labels. The cheap Query margin is 0.995414–0.997288, versus Feature's 0.983271.

At $\tau = 0.25$, Query first passes span 87–135 rows (Figure 3(a)). At an off-grid 432 rows, both rules cover all eight workloads; this does not change first passage. A trace-matched isotropic control gives 99 rows but is not a real workload. Later fixed-task evaluation has maximum pointwise normalized CI upper 0.01111 and does not change response-blind certification.

A separate response-aware whole-query comparison at source-variance allowance $\eta = 0.01$ changes $5{,}793 \to 4{,}409$ rows. Its 256-path mean displacement is 0.07515ϵ, with pointwise interval $[0.07209, 0.07820]\epsilon$. This uses different information and criteria from response-blind selection. Full-task scores are in `geometry/reproducibility/supplementary_experimental_details.md`. Observed weather defines the task; it does not forecast weather.

Q.5 Tetouan: source-only selection and moment gain

On the Tetouan split, source-only selection compares four fixed maps, all with an intercept: linear weather ($d = 6$), linear weather plus calendar ($d = 14$), quadratic weather ($d = 11$), and quadratic weather plus calendar ($d = 19$). Calendar features are daily, weekly, and annual harmonics. Quadratic maps add squares of the five source-standardized weather columns. The two folds are $0{:}19{,}656 \to 19{,}656{:}29{,}484$ and $0{:}29{,}484 \to 29{,}484{:}39{,}312$. Centering and scaling use each fold's training data, then the full source. There are no response lags or outcome-driven features.

The score gives equal weight to standardized MSE across both folds and all three outputs. Eligibility requires at least 5% aggregate improvement over the training mean, with no output worse. All candidates qualify. Scores are 1.111192 for quadratic weather plus calendar, 1.114790 for linear weather plus calendar, 1.303560 for linear weather, and 1.467331 for quadratic weather. The fixed rule selects quadratic weather plus calendar; its small lead does not show a general benefit from squared features. Query labels enter only the final scorer. Full MSE is below every source-mean comparator; exact scores are in `geometry/reproducibility/supplementary_experimental_details.md`.

On the same size grid, the selected model gives:

η	s_{B}	$s_{\mathsf{B}+2}$	Rows removed	Reduction (%)
.005	21,426	16,463	4,963	23.1634
.010	14,742	11,757	2,985	20.2483
.025	7,605	6,621	984	12.9389
.050	4,128	3,803	325	7.8731

At the primary allowance, a separate all-output policy changes $16{,}315 \to 12{,}973$ rows; the aggregate certificate gives 11,757 rows. On the earlier selected-Tetouan paths, primary aggregate displacement is 0.0003662 with pointwise 95% interval $[0.0003417, 0.0003907]$. Every evaluated cell meets the frozen half-width limit 0.1ϵ, and mean/allowance ratios are 0.0316–0.0441. The fixed-label loss identity holds in expectation; finite-path loss and displacement can differ.

Q.6 Gas: marginal cost of second-moment certification

The Gas cost experiment uses the positional blocks and a 548-size grid at $\eta = .01$. Query responses are never loaded.

For $J_+ = \{j : L_j > 0\}$, write the certificate using query-independent source/card operators:

$$\begin{aligned} \mathcal{P}_s &= \sum_{j \in J_+} L_j (I - \beta_s H_{s,j}^{-1}), \\ \mathcal{C}_s &= \beta_s \sum_{j \in J_+} L_j H_{s,j}^{-1} V_{s,j} H_{s,j}^{-1} \succeq 0, \\ U_{\mathrm{per},g}(s) &= \frac{\mathrm{tr}(\Xi_g \mathcal{P}_s)}{q_g K}, \qquad U_{2,g}(s) = \frac{\mathrm{tr}[\Xi_g (\mathcal{P}_s - \mathcal{C}_s)]}{q_g K}. \end{aligned} \tag{269}$$

This is a rearrangement of (27). The ordered inverse remainder validates the correction; the source/card factorization makes it reusable across requests. Zero-loss outputs are omitted. Issuance subtracts a validated lower correction from an upper per-output endpoint, then takes the minimum with existing endpoints.

We compare a source-statistics cache with a cache that also stores the query-independent operators by candidate size. Neither stores decisions; B stores no second-moment state. Both give identical directed endpoints and decisions in all 12×548 cases under 256-bit interval arithmetic. All B winners are per-output, and all B+2 winners are second-moment. Median row reduction is 19.0210%, or 3,753 rows.

Timing boundary and interpretation. Five paired repeats use prefixes $N = 1, 2, 6, 12$. The primary arm-median ratio is 1.167520. Its 16.75% premium combines moment work, visited sizes, and earlier stopping; totals exclude untimed pipeline work. This fixed-host $d = 10, K = 2$ study is not a kernel microbenchmark and does not test asymptotic scaling, label costs, or net training acceleration. Exact timings are in `geometry/reproducibility/supplementary_experimental_details.md`.

Shared-path risk diagnostics. The earlier Gas block cohort uses 256 independent nested ordinary-volume paths per cell, shared across blocks, with pointwise two-sided $t(255)$ intervals. At B+2-selected sizes, mean displacement is 0.1598–0.2905 of the allowance, with CI half-width 0.0287–0.0527; all 24 selected analytic groups pass the precision check. Path maxima are 1.614–3.340 times the allowance: the expectation certificate allows individual paths to exceed it. Across the full grid, 857 of 3,888 cells lack precision. Selected-cell precision does not cover the full grid, and output diagnostics do not make the aggregate policy all-output.

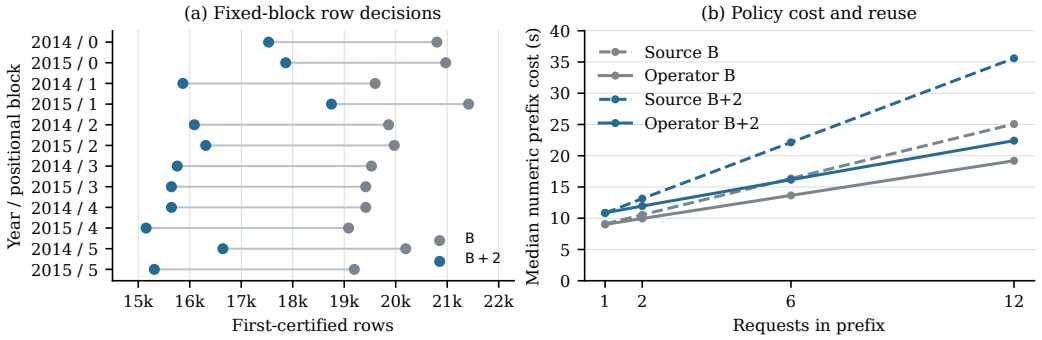


Figure 8: **Second moments separate information gain from operator reuse.** (a) Paired decisions on 12 blocks from one source. (b) Five-repeat median numeric prefix costs, including source build and one Full fit per arm. Adding moments raises cost under the same cache; operator reuse lowers repeated work.

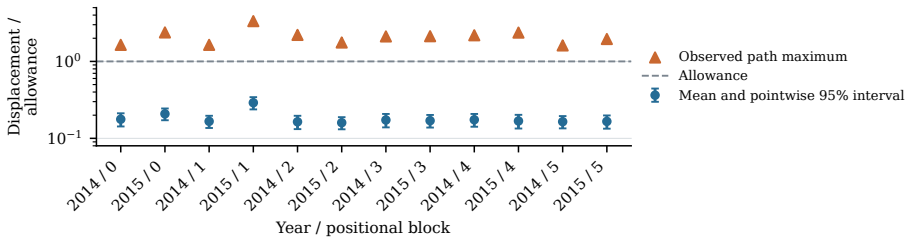


Figure 9: **Expected displacement and individual-path behavior.** At each B+2-selected size, points and bars are 256-path mean displacement and pointwise nominal 95% intervals; triangles are observed path maxima. The same paths serve all 12 blocks. The allowance constrains the sampling expectation, not every realized fit.

R MATCHED EXPECTED-RISK EVIDENCE

Workloads and decision units. Gas positional blocks and the selected Tetouan model (Appendix Q) define 13 workloads, not independent datasets. Each fixes X, Y, Z_g. In (30), $\mathcal{D}_{s,g}$ is expected Full-query-MSE excess from ordinary indexed fixed-size volume sampling, averaged over query rows and outputs. Query labels affect neither certificates nor allowances.

The registry computes its $\eta = .01$ allowance with exact rational arithmetic. Original B minimizes five valid upper candidates; B+2 adds Second. The displayed upper minimizes all seven, including Universal; Second wins all 52 selected cells. Original policy selections and sufficiency checks remain. At Tetouan $s = 11{,}757$, B $\approx 0.0139511 > \epsilon$ but Second $\approx 0.00992206 \leq \epsilon$; the sufficient-size reduction comes from the corrected upper, not the new lower bound.

Selection and numerical boundary. Before inspecting rigorous endpoints, each workload supplied four cells: its largest scalar *float diagnostic* crossing, next original-grid size, and original B+2 and B sufficient sizes. Each of 12 Gas grids has 548 sizes and Tetouan has 614; all 7,190 screening values were float diagnostics, not risk certificates. Table 2 shows the 2015 Gas blocks with largest and smallest pre-loss float crossings; their display rule was not separately recorded as preregistered.

Exact rational arithmetic treats frozen binary64 X, Y, Z_g as dyadics and verifies each selected cell's Gram inverse, Full normal equations, positive clearances, and allowance. Directed ball arithmetic encloses raw $U_{0,j}, U_{1,j}$ moments in (230); rational interval endpoints and (231) yield the scalar lower endpoint. The directed upper engine reissues all valid candidates. Both endpoints use the same $q_g K$ divisor; all 52 pairs satisfy $L^-_{s,g} \leq U^+_{s,g}$. Certification starts at the frozen arrays; it excludes public CSV parsing and preprocessing, individual subset losses, other samplers, and feature changes.

Table 7 reports all 52 decisions: 13 violated, 13 unresolved, and 26 sufficient. Its source JSON is the issued record; the companion CSV holds exact endpoints, allowances, roles, and decisions; the generator checks verdicts by exact fraction comparison. Each selected crossing certifies failure only at that size. Its next original-grid size remains unresolved, with lower endpoint at or below the allowance and upper endpoint above it. None of the 26 original B or B+2 choices is ruled out; the new lower bound does not reduce a sufficient size.

Table 7: All 52 rigorous certificate cells, read down (a), then (b). W: G14.k or G15.k means Gas 2014 or 2015 block k; T means Tetouan. R: C crossing, N successor, 2 B+2, B historical B. D: V violated, U unresolved, S sufficient. Bounds are rounded; the companion CSV has exact endpoints and allowances.

(a) First 28 cells						(b) Remaining 24 cells					
W	s	Lo	Up	R	D	W	s	Lo	Up	R	D
G14.0	9,543	0.6407396	7.354029	C	V	G15.1	13,537	0.6422758	4.440448	C	V
G14.0	9,609	0.6330421	7.231221	N	U	G15.1	13,648	0.6288981	4.300659	N	U
G14.0	17,531	0.1288306	0.6316429	2	S	G15.1	18,752	0.1844239	0.623682	2	S
G14.0	20,800	0.0324272	0.0718628	B	S	G15.1	21,415	0.03645081	0.05756336	B	S
G14.1	6,340	0.6380248	8.702048	C	V	G15.2	7,140	0.644414	8.170139	C	V
G14.1	6,436	0.6247761	8.47345	N	U	G15.2	7,213	0.6348424	8.012853	N	U
G14.1	15,867	0.1022022	0.6209222	2	S	G15.2	16,311	0.1106304	0.6191644	2	S
G14.1	19,601	0.03390401	0.1057477	B	S	G15.2	19,972	0.03411566	0.09469333	B	S
G14.2	6,769	0.6435544	8.556038	C	V	G15.3	5,992	0.6373832	8.780448	C	V
G14.2	6,862	0.6310687	8.343241	N	U	G15.3	6,094	0.6228549	8.530101	N	U
G14.2	16,089	0.1076014	0.6347968	2	S	G15.3	15,645	0.09915375	0.6142284	2	S
G14.2	19,861	0.03330374	0.09703717	B	S	G15.3	19,418	0.03386283	0.109709	B	S
G14.3	6,103	0.644758	8.927072	C	V	G15.4	5,104	0.6458549	9.931279	C	V
G14.3	6,214	0.6289648	8.652545	N	U	G15.4	5,201	0.630329	9.640742	N	U
G14.3	15,756	0.100404	0.6195587	2	S	G15.4	15,152	0.09022146	0.6271791	2	S
G14.3	19,529	0.03353106	0.1066058	B	S	G15.4	19,085	0.03162873	0.1159738	B	S
G14.4	5,992	0.6490169	8.753599	C	V	G15.5	5,305	0.6376063	9.618238	C	V
G14.4	6,094	0.6342227	8.504093	N	U	G15.5	5,326	0.6343261	9.557582	N	U
G14.4	15,645	0.1009592	0.6135538	2	S	G15.5	15,312	0.09055145	0.6128817	2	S
G14.4	19,418	0.03447924	0.1099492	B	S	G15.5	19,196	0.03146862	0.1117785	B	S
G14.5	7,767	0.6400144	7.866062	C	V	T	580	0.01007601	0.3378327	C	V
G14.5	7,878	0.6261971	7.641933	N	U	T	590	0.009914859	0.3326554	N	U
G14.5	16,644	0.1152497	0.6198003	2	S	T	11,757	0.0003799073	0.009922057	2	S
G14.5	20,194	0.03421554	0.08959933	B	S	T	14,742	0.0002703728	0.006339062	B	S
G15.0	10,821	0.6363874	5.896907	C	V						
G15.0	10,874	0.630347	5.816783	N	U						
G15.0	17,864	0.1469543	0.6143608	2	S						
G15.0	20,971	0.03530895	0.06847105	B	S						

Fixed-size monotonicity and first grid passage. For fixed design, responses, query, and ordinary law, Corollary 36 gives $M_{a,j} - M_{b,j} \succeq 0$ for $d \le a < b \le m$. Its fixed-query trace summed over outputs makes $\mathcal{D}_{s,g}$ nonincreasing in size. On the original grid, the first truly passing size lies in $(s_F, s_P]$: s_F is the largest *issued selected* failure and s_P the smallest issued selected sufficient size. This locates neither the last failing size nor an integer optimum, and gives no outcome-adaptive stopping rule or monotonicity of the normalized deficit.

Separate initial actual-risk measurement. The frozen manifest selects 256 independent ordinary-volume paths per source (512 total): each Gas path serves all 12 blocks, each Tetouan path five cells. Positive-volume OLS fits use certificate risk units. Table 8 reports 53 means with nominal *pointwise* Student-t 95% intervals. The `geometry/reproducibility/monte_carlo_cells.csv`, sourced from `geometry/reproducibility/monte_carlo_summary.json`, gives standard errors, quantiles, maxima, and extreme-path contributions. Forty cells overlap certificates (36 Gas, four Tetouan); 13 MC middle sizes lack certificates, and 12 certified Gas B sizes lack new MC summaries.

The prespecified nominal halfwidth target 0.1ϵ passed 23 of 48 Gas cells and all five Tetouan cells; 25 Gas cells missed it. Among certificate overlaps, 13 correlated Gas sample means fall below rigorous lower endpoints, but no nominal pointwise interval lies wholly below its endpoint. These finite-path

results neither refute a bound nor establish general scalar tightness. The exploratory intervals lack rigorous, simultaneous, or sequential coverage; two unrun larger-path stages are not pooled. MC means are not exact expectations and replace no deterministic bracket.

At Tetouan's B+2 size 11,757, Table 8 uses matched-risk paths, while the earlier estimate in Appendix Q.5 uses separate selected-Tetouan paths. The registry identifies both LOO variants and path sets.

Table 8: All 53 initial 256-path MC summaries, read down (a), then (b). W uses Table 7's workload codes. Lo and Hi are nominal pointwise 95% Student-t limits. Pr is P if the halfwidth is at most 0.1ϵ, F otherwise. These estimates do not certify a budget.

(a) First 28 cells						(b) Remaining 25 cells					
W	s	Mean	Lo	Hi	Pr	W	s	Mean	Lo	Hi	Pr
G14.0	9,543	0.609962	0.506353	0.713571	F	G15.1	13,537	0.723205	0.615545	0.830865	F
G14.0	9,609	0.591154	0.491808	0.690499	F	G15.1	13,648	0.735949	0.628045	0.843854	F
G14.0	13,315	0.331423	0.280709	0.382137	P	G15.1	16,079	0.429248	0.362684	0.495813	F
G14.0	17,531	0.150382	0.125694	0.17507	P	G15.1	18,752	0.193433	0.160693	0.226172	P
G14.1	6,340	0.634302	0.542902	0.725701	F	G15.2	7,140	0.644394	0.552306	0.736483	F
G14.1	6,436	0.605884	0.517596	0.694171	F	G15.2	7,213	0.640619	0.546131	0.735108	F
G14.1	10,763	0.284309	0.239317	0.329301	P	G15.2	11,429	0.302684	0.2557	0.349667	P
G14.1	15,867	0.110522	0.0936001	0.127445	P	G15.2	16,311	0.122461	0.10331	0.141611	P
G14.2	6,769	0.62238	0.527014	0.717745	F	G15.3	5,992	0.638569	0.543461	0.733677	F
G14.2	6,862	0.615533	0.522651	0.708415	F	G15.3	6,094	0.620466	0.529466	0.711465	F
G14.2	11,096	0.294955	0.245793	0.344116	P	G15.3	10,430	0.265856	0.223796	0.307916	P
G14.2	16,089	0.117331	0.0982878	0.136374	P	G15.3	15,645	0.106951	0.0899254	0.123976	P
G14.3	6,103	0.630738	0.539004	0.722472	F	G15.4	5,104	0.683216	0.580692	0.78574	F
G14.3	6,214	0.625884	0.53192	0.719848	F	G15.4	5,201	0.667041	0.564738	0.769345	F
G14.3	10,541	0.272735	0.227104	0.318367	P	G15.4	9,654	0.240442	0.200668	0.280216	P
G14.3	15,756	0.107388	0.0897439	0.125032	P	G15.4	15,152	0.100267	0.0834781	0.117055	P
G14.4	5,992	0.649566	0.559624	0.739509	F	G15.5	5,305	0.668011	0.570679	0.765342	F
G14.4	6,094	0.627844	0.5413	0.714387	F	G15.5	5,326	0.662923	0.56544	0.760407	F
G14.4	10,430	0.27121	0.226831	0.315588	P	G15.5	9,801	0.241967	0.202694	0.281241	P
G14.4	15,645	0.111295	0.0930816	0.129509	P	G15.5	15,312	0.100742	0.0843793	0.117105	P
G14.5	7,767	0.612965	0.522994	0.702936	F	T	580	0.0110456	0.0103269	0.0117642	P
G14.5	7,878	0.599922	0.507771	0.692074	F	T	590	0.0108441	0.0101343	0.0115539	P
G14.5	11,947	0.322682	0.270222	0.375142	P	T	3,377	0.00176283	0.0016469	0.00187875	P
G14.5	16,644	0.13642	0.114122	0.158718	P	T	11,757	0.00039184	0.000366674	0.000417007	P
G15.0	10,821	0.70012	0.586157	0.814083	F	T	14,742	0.000285119	0.000268044	0.000302195	P
G15.0	10,874	0.691479	0.579055	0.803903	F						
G15.0	14,203	0.398156	0.337913	0.458398	P						
G15.0	17,864	0.172841	0.145092	0.20059	P						

S RELATIVE-LOSS APPLICATION: MASSIVE AND KUAIRAND

These $d = 65$ experiments use the label-floor implication (219), ordinary fixed-size volume sampling, and unweighted, unregularized OLS. The three MASSIVE representations belong to one intent task, not independent datasets.

Data / outputs	Source and query	Fixed features
MASSIVE; 60-class one-hot	11,514 English training utterances; 148,700 test utterances in 50 other locales	LaBSE (A), paraphrase MPNet (B), or multilingual E5 (C); source-centered PCA64 plus intercept
KuaiRand; binary long-view	295,497 standard-exposure events; 1,186,059 random-exposure events, April 22–May 8	Intercept plus elementwise product of 64-dimensional item and historical-user-profile vectors

MASSIVE uses official splits (FitzGerald et al., 2023); query transforms use source-fitted means and components. KuaiRand (Gao et al., 2022; Gao, 2026) uses LaBSE item-text embeddings with PCA fitted to 7,583 static items. User profiles average positive long-view item vectors from April 8–21, before the source/query period; current outcomes do not enter the representation. Encoder

revisions, pooling, normalization, item mapping, and exact construction appear in `geometry/reproducibility/relative_loss_methods.md`.

Information and decision criterion. Bounds use source responses and the exact Full fit; query targets enter only evaluation. Under the recorded shared-label premise, MASSIVE fixes one intent within each of 2,974 original-ID groups of 50 query rows. KuaiRand uses a binary rowwise alphabet without group constraints. These premises set $L_{\min}$ without changing U_2, predictions, or row weights. All bounds use the 206-size grid $\{d + \lceil j(m-d)/200 \rceil : j = 0, \ldots, 200\}$ augmented by six prespecified protocol sizes. The first directed endpoint below $0.25L_{\min}$ is accepted only after checking every earlier candidate; U_2 need not be monotone. Full has bound zero.

Paired fidelity comparison. For displayed finite-path risks, a hat marks an empirical mean. For MASSIVE paraphrase-MPNet, sufficient sizes 6,534 and 4,989 were fixed before fitting or scoring. Each of 128 independent paths shares a volume basis and uniformly permuted complement across both sizes. Full and selected fits share response normalization. Error bars are pointwise 95% $t(127)$ intervals over paired paths and fixed queries, not population-task intervals.

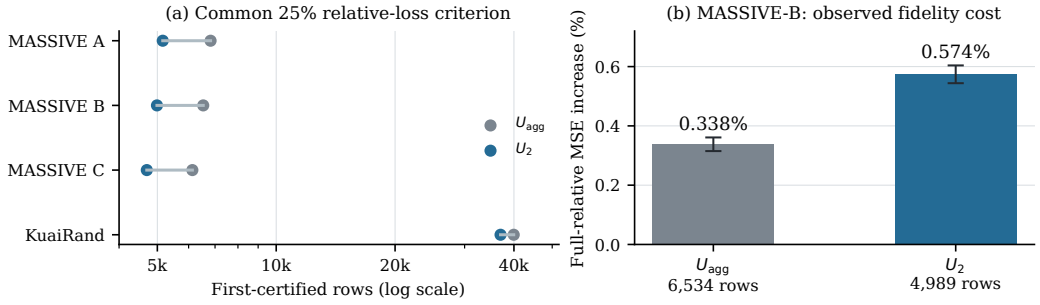


Figure 10: **Relative-loss budgets and observed fidelity.** (a) First-certified sizes under the common 25% criterion; aggregate and per-output bounds agree. (b) Paraphrase-MPNet mean Full-relative MSE increase with pointwise 95% intervals over 128 paired paths. A smaller sufficient budget can have worse observed fidelity.

Scope and adverse evidence. The label-floor guarantee holds in expectation; signed cross terms can persist on finite paths, and a zero floor can make the criterion uninformative. Query-normalized and Full-relative risks have different denominators: earlier small normalized-query budgets produced 59.53–129.63% mean Full-relative MSE increases. In a 90-pool study, source-only pair margins failed their practical criterion for all three representations; independent spectral-direction signals were weak. These adverse results limit broad claims of practical geometry gain; no net training speedup is established.

GROUP V. REPRODUCIBILITY AND SCOPE

T REPRODUCIBILITY

The supplement's `README.md` and `geometry/reproducibility/RESULT_MAP.md` give the numerical commands and their scope. Exact finite-law calculations, directed bounds, exhaustive float64 LOO sums, and MC estimates support different claims. Stored-endpoint checks do not rerun the directed engine; released path losses reproduce MC summaries, not draws or path independence. MC intervals condition on fixed data and use whole paths as units. Separate public-data commands reissue Figure 3 decisions and the selected Gas/Tetouan risk experiment. Other external-data studies may need further inputs, including encoders.

U RELATION TO PRIOR WORK

Ordinary volume sampling and deficits. Classical results under Section 2's law give unbiased selected least squares, inverse-Gram moments, and rank-size fixed-response control (Dereziński & Warmuth, 2018b, Thms. 5–8); see also Dereziński & Warmuth (2017). Uniformly padded rank-size bases give the nested law (Dereziński et al., 2019b; Dereziński & Warmuth, 2018b); Kulesza & Taskar (2012, §2.4) sample projection-DPP bases. Those identities do not determine fixed-response contact or all-size covariance deficits; our comparison uses ordinary-law omission moments, conditional unbiasedness, and total covariance. The inverse-Gram equality in Dereziński & Warmuth (2018b, Thm. 6) requires full support; no-coloop designs permit zero-volume subsets. The scalar lower certificate uses the omitted-row predictor's ordinary-law moments and optimizes its coefficient by classical least-squares projection. This scalar optimization does not enter the all-size matrix comparison.

Related regression guarantees. Rescaled estimators (Dereziński et al., 2018), regularized-volume ridge (Dereziński & Warmuth, 2018a), and volume-rescaled random designs (Dereziński et al., 2019b; 2022) change estimator, law, or randomness. The minimax construction of Dereziński et al. (2019a) combines volume and rescaled draws; Epperly (2026, Thm. 3.2) discusses selected OLS at rank size. Our strict-interior OLS is unweighted; only the indexed subset is random. The ordinary-sampling selected-OLS scalar-loss obstruction of Dereziński et al. (2018) has our sharp witness's numerical excess factor at specialized dimensions, but uses a highly nonuniform-leverage two-block limit. At equal block scale, whitening yields duplicated-pair geometry; the cited obstruction uses the unequal-scale limit. Our witness yields exact centered rank-one covariance for one fixed residual, specializing the fixed-design phase theorem.

Frames and queries. Complement-frame Gram geometry (Casazza et al., 2013), exact two-erasure formulas (Bodmann & Paulsen, 2005), and erasure conditioning (Fickus & Mixon, 2012) provide context; projection-diagonal admissibility (Kadison, 2002; Antezana et al., 2007) and prescribed frame weights (Balazs et al., 2010) allow broader choices. Our $Q_z = P_\perp - zz^\top$ has corank one in fixed $P_\perp$ and residual-compatible fractional weights, distinct from arbitrary frame weights and binary erasures. For uniform Parseval frames, binary two-erasure error $(1 + \rho_A)/2$ and retained lower bound $(1 - \rho_A)/2$ share ν_A's projective-duplicate singular boundary, not its residual-weighted margin. Our complement certificate links pair minors, deletion geometry, and erasure stability to an ordinary-volume covariance bound. Query-weighted traces are design objectives (Yu et al., 2006; Sagnol & Pronzato, 2023); we study fixed-response ceiling contact.

Inverse moments. We combine classical inverse-map Jensen (Hansen & Pedersen, 2003, Thm. 2.1) with volume-specific residual augmentation, supported omissions, and an ordered remainder for the second-moment correction. With full size-s support, centered moments (Dereziński & Warmuth, 2018b, Thm. 6) and residual rotations (Perry & Owen, 2010) yield residual-averaged calibration, not a fixed-response bound.