Submitted to *Econometrica*

FEASIBLE POWER FRONTIERS FOR LEARNED SCALAR ANDERSON–RUBIN SCORES UNDER HIGH-DIMENSIONAL WEAK IDENTIFICATION

TAMER ÇETIN

Weak-identification theory usually fixes the score before studying robust inference. In high-dimensional IV designs, scalar Anderson–Rubin scores are often selected from growing score dictionaries. This paper characterizes one finite-information frontier: sample splitting preserves weak-ID size conditional on the learned score, but oracle scalar-AR power is limited by what the training sample can learn about the weak-drift direction. In a canonical Gaussian weak-drift experiment with sparse drifts, minimax normalized noncentrality regret is of order $\min\{1, \sigma_n^2 s_n \log(ep_n/s_n)/r_n^2\}$, with corresponding power regret on compact oracle-noncentrality ranges. Hence nonvanishing training noise and growing sparse complexity rule out uniform oracle-power recovery under nondegenerate weak-ID oracle power. Bayes and finite-grid criteria are decision-theoretic benchmarks. Growing-dictionary PLIV gives the central econometric realization through $B_{n,\Delta}^{-1/2} g_n$.

KEYWORDS: weak identification, weak instruments, high-dimensional inference, feasible power, score learning, minimax regret, Anderson–Rubin.
JEL CLASSIFICATION. C12, C13, C14, C26, C36.

1. INTRODUCTION

Weak-identification robust inference is usually formulated after a score or moment process has been chosen. Anderson–Rubin, score, LM, conditional likelihood-ratio, conditional linear-combination, and weak-GMM procedures deliver tests and confidence sets with correct size when identification is weak, partial, or singular, taking the moment process as the object to which the robust statistic is applied (Anderson and Rubin, 1949, Staiger and Stock, 1997, Stock and Wright, 2000, Kleibergen, 2002, Moreira, 2003, Andrews, Moreira, and Stock, 2006, Andrews, 2016, Andrews and Mikusheva, 2016, Andrews and Cheng, 2012, Andrews and Mikusheva, 2022, Kaji, 2021). In modern IV applications, however, the score is often constructed before the robust statistic is reported. Instruments are expanded into transformations and interactions, high-dimensional controls are residualized, orthogonal scores are formed, and regularization selects among many candidate score representations. Under weak identification, these choices are first-order because they determine the local drift and covariance of the statistic used for robust inference.

This paper studies the finite-information power frontier for learned scalar Anderson–Rubin scores. Training data map noisy information about the weak drift into a linear score direction, and an independent inference sample evaluates the corresponding one-dimensional AR statistic. Sample splitting preserves the fixed-score AR null law conditional on training, while feasible local power depends on how accurately the training sample learns the weak identifying direction.

The frontier is a scalarization frontier. A full-vector Anderson–Rubin statistic avoids learning a score direction, but its critical value grows with the dictionary dimension. An oracle scalar Anderson–Rubin statistic concentrates power in the best direction, but that direction depends on the weak drift. A learned scalar Anderson–Rubin statistic is the intermediate design: sample splitting preserves weak-ID size, while training information selects the one-dimensional direction used for inference. The quantity η_n below is the price of this intermediate design.

The analysis is organized around a Gaussian weak-drift experiment. A growing score dictionary has dimension p_n, and the training signal is $Y_n = G_n + \rho_{\mathrm{tr}}^{-1/2} \Sigma_n^{1/2} \xi_n,\ \xi_n \sim N(0, I_{p_n})$.

Tamer Çetin: tamercetin@berkeley.edu

A feasible rule δ_n chooses a coefficient vector $\widehat{\beta}_n = \delta_n(Y_n)$. For denominator matrix B_n and inference-sample fraction $\rho_{\inf}$, the scalar Anderson–Rubin noncentrality generated by a score coefficient β at drift g is

$$Q_n(\beta; g, B_n) = \rho_{\inf} \frac{(\beta' g)^2}{\beta' B_n \beta}. \tag{1.1}$$

The associated level-α local power is $h_\alpha\{Q_n(\beta; g, B_n)\}$, $h_\alpha(q) = 1 - F_{\chi^2_1(q)}(c_{1-\alpha})$. The oracle score direction is proportional to $B_n^{-1} g$. The feasible problem is to choose a direction from the noisy training signal Y_n, and to quantify the resulting loss in local Anderson–Rubin power.

The paper is organized around one unavoidable frontier. Weak-ID robust size is protected by sample splitting, but oracle scalar-AR power is feasible only to the extent that the training sample learns the projective direction of the weak drift. In sparse canonical coordinates the finite-information difficulty is

$$\eta_n = \frac{\sigma_n^2 s_n \log(e p_n / s_n)}{r_n^2}, \tag{1.2}$$

and the frontier is $\min\{1, \eta_n\}$. The Bayes, PLIV, finite-grid, and numerical sections are organized around this single object: Bayes clarifies the power decision, PLIV supplies the central econometric realization, finite grids study common-score use over several local directions, and the numerical experiments illustrate the same difficulty index.

The organizing theorem is a sparse high-dimensional minimax frontier for learned scalar AR directions. In the canonical Gaussian sequence experiment $Y = g + \sigma_n \xi$, $\xi \sim N(0, I_{p_n})$, let $\mathcal{G}_n(s_n, r_n) = \{g \in \mathbb{R}^{p_n} : \|g\|_0 \le s_n,\ \|g\|_2 = r_n\}$. For normalized noncentrality regret,

$$L_n(\delta, g) = 1 - \frac{Q_n(\delta(Y); g, I_{p_n})}{Q_n^\star(g)} = 1 - \frac{\{\delta(Y)' g\}^2}{\|\delta(Y)\|^2 \|g\|^2},$$

the minimax envelope is

$$\inf_{\delta_n} \sup_{g \in \mathcal{G}_n(s_n, r_n)} \mathbb{E}_g L_n(\delta_n, g) \asymp \min\left\{1, \frac{\sigma_n^2 s_n \log(e p_n / s_n)}{r_n^2}\right\}. \tag{1.3}$$

The lower bound is projective: signs of score directions are irrelevant for scalar Anderson–Rubin power. The upper bound is attained, up to constants, by hard thresholding. On compact oracle-noncentrality ranges, the same rate transfers to power regret because the scalar AR power map has derivative bounded above and away from zero. Consequently, under nondegenerate weak-ID oracle power and nonvanishing training noise, growing sparse complexity rules out uniform recovery of oracle scalar-score power.

The scalar restriction creates a sharp econometric comparison. Full-vector or many-moment AR avoids learning a score direction but pays a growing critical-value cost when the dictionary is large. Oracle scalar AR avoids that cost but requires the weak drift. Learned scalar AR is the operational middle ground: sample splitting preserves weak-ID size conditional on the selected score, and the sparse frontier quantifies the finite-information cost of making the scalar direction feasible. The Bayes analysis is a supporting decision-theoretic benchmark for the same frontier. Given a posterior law for the drift G_n after observing Y_n, the Bayes rule maximizes $\mathbb{E}[h_\alpha\{Q_n(\beta; G_n, B_n)\} \mid Y_n]$ over score directions. This posterior expected-power criterion need not coincide with the posterior Rayleigh rule that maximizes posterior expected noncentrality. A two-direction posterior example shows that the power-optimal score can depend on the nominal level α, because the scalar AR power map is nonlinear in the noncentrality.

The main econometric realization is growing-dictionary partially linear IV. The structural environment is $Y = (\theta_0 + \Delta)D + g_0(X) + U,\ \mathbb{E}[U \mid Z, X] = 0$, with local first stage $D - m_0(X) = N^{-1/2}\pi_n(Z, X) + V$. For a centered dictionary $b_{p_n}(Z, X)$, the training sample learns $g_n = \mathbb{E}[b_{p_n}(Z, X)\pi_n(Z, X)]$, while the inference-side local drift is $G_{n,\Delta} = \Delta g_n$ and the scalar-AR denominator is $B_{n,\Delta} = \mathbb{E}[\omega_\Delta(Z, X)b_{p_n}(Z, X)b_{p_n}(Z, X)'],\ \omega_\Delta(Z, X) = \mathbb{E}[(U + \Delta V)^2 \mid Z, X]$. Thus scalar-AR power depends on the denominator-normalized first-stage direction $\widetilde{\gamma}_{n,\Delta} = B_{n,\Delta}^{-1/2} g_n$, not on raw instrument coordinates. Under aligned PLIV conditions, the canonical sparse frontier applies with difficulty index

$$\eta_{n,\Delta} = \frac{\rho_{\mathrm{tr}}^{-1}\sigma_\Delta^2 s_n \log(ep_n/s_n)}{\|\widetilde{\gamma}_{n,\Delta}\|^2}.$$

This normalization is substantive: sparsity is a property of the AR-denominator geometry, not of an arbitrary raw dictionary. Section 5.7 gives a rotated-factor example in which the raw drift is dense while the denominator-normalized drift is sparse. A final finite-grid extension studies direction-free score design. For a finite set of local alternatives or tested values, the paper defines integrated and maximin feasible-power envelopes over common training-measurable score rules. The result gives existence and a collapse condition under which all local directions rank score coefficients in the same order. This extension connects the pointwise score-learning frontier to finite-grid test design while preserving the scalar sample-split Anderson–Rubin structure.

The paper connects four literatures. Strong-identification optimal-instrument and optimal-moment theory studies efficiency when the identifying direction is regular (Chamberlain, 1987, Newey, 1990, Donald and Newey, 2001, Belloni, Chen, Chernozhukov, and Hansen, 2012, Chen, Chen, and Lewis, 2020). DML, locally robust moments, and automatic Riesz methods construct orthogonal scores for regular or locally regular targets (Chernozhukov et al., 2018, Chernozhukov, Newey, and Robins, 2022, Chernozhukov, Newey, and Singh, 2022). Weak-identification robust inference and weak-GMM decision theory study size control, power, and optimal decisions once the score or moment process is specified (Anderson and Rubin, 1949, Staiger and Stock, 1997, Stock and Wright, 2000, Kleibergen, 2002, Moreira, 2003, Andrews, Moreira, and Stock, 2006, Andrews, 2016, Andrews and Mikusheva, 2016, Andrews and Cheng, 2012, Andrews and Mikusheva, 2022, Kaji, 2021). Many-instrument and many-weak-moment work studies validity, estimation, or test construction with many instruments, moments, or covariates (Bekker, 1994, Chao and Swanson, 2005, Hansen, Hausman, and Newey, 2008, Newey and Windmeijer, 2009, Mikusheva and Sun, 2024, Carrasco and Tchuente, 2016, Lim, Wang, and Zhang, 2024, Ayyar, Matsushita, and Otsu, 2025, Ma, 2023, Boot and Nibbering, 2024, Zhang, Li, and Sun, 2026, Wang, Chan, and Ye, 2025, van der Laan, Kallus, and Bibaut, 2025, Baiardi, Clarke, Naghi, and Polselli, 2026). The present contribution adds a finite-information score-learning layer to weak-identification theory: before the final robust statistic is evaluated, high-dimensional score construction imposes a quantifiable constraint on attainable local power.

The numerical section reports limiting-experiment designs that correspond to the main theorems. One design compares posterior-power and posterior-Rayleigh rules. A sparse design varies the difficulty index in (1.2). Additional designs illustrate dense shrinkage, finite-grid direction-free criteria, and finite-sample PLIV transfer in aligned dictionaries.

The paper proceeds as follows. Section 2 defines the canonical finite-information experiment for learned scalar AR power and proves the sample-split scalar AR link. Section 3 gives the Bayes decision benchmark for posterior power-optimal score choice. Section 4 proves the sparse minimax frontier, its power-regret transfer, and the split-design implications. Section 5 transports the frontier to growing-dictionary PLIV through denominator-normalized score coordinates. Section 6 studies finite-grid direction-free score design. Section 7 gives numerical local experiments. Appendix A contains the main proofs, and the Online Appendix gives the full sparse minimax derivation and auxiliary PLIV and direction-free proofs.

2. A FINITE-INFORMATION EXPERIMENT FOR FEASIBLE POWER

2.1. *Training signal, score rules, and power regret*

Let $p_n \to \infty$. For each n, let B_n and Σ_n be symmetric positive-definite $p_n \times p_n$ matrices. The weak drift is a vector $G_n \in \mathbb{R}^{p_n}$, either fixed in minimax analysis or random under a Bayes prior. The training experiment is

$$Y_n = G_n + \tau_n \Sigma_n^{1/2} \xi_n, \qquad \xi_n \sim N(0, I_{p_n}), \qquad \tau_n = \rho_{\mathrm{tr}}^{-1/2}. \tag{2.1}$$

The constants ρ_{tr} and ρ_{inf} are fixed nondegenerate training and inference split-fraction limits satisfying $0 < \rho_{\mathrm{tr}}, \rho_{\mathrm{inf}} \leq 1$.

REMARK 2.1—Training and inference fractions: *The main statements are written for fixed nondegenerate split fractions because the sample-split AR statistic is evaluated conditionally on the learned score. The split is nevertheless part of the same frontier. Corollary 4.5 records the single-split tradeoff*

$$\eta_n(\rho) = \frac{\rho^{-1}\sigma_n^2 s_n \log(ep_n/s_n)}{r_n^2}, \qquad q_n^\star(\rho) = (1-\rho) r_n^2.$$

A larger training fraction reduces score-learning noise but lowers inference-sample noncentrality. Thus the split fraction is governed by the same finite-information frontier, rather than by a separate identification argument.

A feasible score rule is a measurable map

$$\delta_n : \mathbb{R}^{p_n} \to \mathbb{R}^{p_n} \setminus \{0\}, \qquad \widehat{\beta}_n = \delta_n(Y_n). \tag{2.2}$$

The action $\widehat{\beta}_n$ is a linear score direction used in a scalar Anderson–Rubin statistic. Sample splitting makes the selected score fixed conditional on training, so the procedure class isolates the feasible-power effect of high-dimensional score learning while preserving transparent weak-ID size control.

Because Anderson–Rubin noncentrality is homogeneous in the score coefficient, the action space can be normalized to

$$\mathcal{B}_n := \{\beta \in \mathbb{R}^{p_n} : \beta' B_n \beta = 1\}. \tag{2.3}$$

For a nonzero coefficient β and drift g, define

$$Q_n(\beta; g, B_n) = \rho_{\mathrm{inf}} \frac{(\beta' g)^2}{\beta' B_n \beta}. \tag{2.4}$$

The oracle noncentrality at drift g is

$$Q_n^\star(g) = \rho_{\mathrm{inf}} g' B_n^{-1} g, \tag{2.5}$$

attained by any nonzero scalar multiple of $B_n^{-1} g$ when $g \neq 0$.

Let

$$h_\alpha(q) = 1 - F_{\chi^2_1(q)}(c_{1-\alpha}), \qquad q \geq 0. \tag{2.6}$$

The local power induced by β is

$$\pi_{\alpha,n}(\beta, g) = h_\alpha\{Q_n(\beta; g, B_n)\}. \tag{2.7}$$

The oracle local power is

$$\pi^\star_{\alpha,n}(g) = h_\alpha\{Q^\star_n(g)\}. \tag{2.8}$$

DEFINITION 2.2—Power regret: *For a feasible rule δ_n and fixed drift g, the power regret is*

$$\mathcal{R}_{\alpha,n}(\delta_n, g) = \pi^\star_{\alpha,n}(g) - \mathbb{E}_g\big[\pi_{\alpha,n}\{\delta_n(Y_n), g\}\big], \tag{2.9}$$

where $\mathbb{E}_g$ denotes expectation under (2.1) *with $G_n = g$.*

THEOREM 2.3—Sample-split scalar AR experiment for learned scores: *Let $\widehat{\beta}_n = \delta_n(Y_n)$ be any training-measurable score rule, and let $\mathcal{F}_n = \sigma(Y_n)$. Suppose that, under the null,*

$$d_{\mathrm{BL}}\big(\mathcal{L}\{T_{n,\widehat{\beta}_n} \mid \mathcal{F}_n\}, N(0,1)\big) \xrightarrow{p} 0. \tag{2.10}$$

Under the local alternative indexed by fixed drift g, suppose

$$d_{\mathrm{BL}}\big(\mathcal{L}\{T_{n,\widehat{\beta}_n} \mid \mathcal{F}_n\}, N\{\mu_n(g), 1\}\big) \xrightarrow{p} 0, \tag{2.11}$$

where the training-measurable signed shift satisfies

$$\mu_n(g)^2 = Q_n(\widehat{\beta}_n; g, B_n). \tag{2.12}$$

Let $AR_{n,\widehat{\beta}_n} = T^2_{n,\widehat{\beta}_n}$. Then $\mathrm{Pr}_0\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha}\} = \alpha + o(1)$, and $\mathrm{Pr}_g\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha}\} = \mathbb{E}_g[h_\alpha\{Q_n(\delta_n(Y_n); g, B_n)\}] + o(1)$. Consequently the power criterion in (2.9) *is the local rejection probability generated by the learned-score scalar AR experiment. The proof is in Appendix A.1.*

REMARK 2.4—Uniform reading: *Theorem 2.3 is used below in its uniform form over the relevant triangular-array classes. If the conditional bounded-Lipschitz approximations in* (2.10) *and* (2.11) *hold uniformly over classes $\mathcal{P}_{n,0}$ and $\mathcal{P}_{n,\Delta}$, then the null-size and local-power conclusions also hold uniformly:*

$$\sup_{P \in \mathcal{P}_{n,0}} \big|P\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha}\} - \alpha\big| \to 0,$$

and

$$\sup_{P \in \mathcal{P}_{n,\Delta}} \Big|P\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha}\} - E_P h_\alpha\{Q_{n,P}(\widehat{\beta}_n)\}\Big| \to 0.$$

The proof is unchanged: once the conditional bounded-Lipschitz distance is controlled uniformly, the rejection-region bracketing argument and the normal boundary bound are deterministic and uniform over the conditional mean.

REMARK 2.5—Why study learned scalar AR scores?: *The rule class is intentionally narrower than the class of all weak-ID robust tests. Scalar AR keeps size control transparent after sample splitting: conditional on the training sample, the inference statistic is one-dimensional. The restriction also creates a sharp comparison. A full-vector AR statistic avoids learning a direction but pays a growing critical-value cost when the dictionary is large. An oracle scalar AR statistic concentrates power in the best direction but requires the weak drift. A learned scalar AR statistic is the intermediate procedure: it preserves sample-split weak-ID size while using training information to select a low-dimensional direction for inference. Proposition 4.8 formalizes the many-moment side of this scalarization tradeoff. The results below are frontiers over training-measurable scalar score rules, not dominance claims relative to CLR, conditional linear-combination, or modern many-instrument procedures.*

2.2. *Canonical normalization and sparse classes*

Many statements are simplest after the transformation

$$\theta = B_n^{1/2}\beta, \qquad \gamma = B_n^{-1/2}g, \qquad Z = B_n^{-1/2}Y_n. \tag{2.13}$$

Then $\theta'\theta = 1$ on the normalized action space and

$$Q_n(\beta; g, B_n) = \rho_{\inf}(\theta'\gamma)^2. \tag{2.14}$$

The transformed training experiment is

$$Z = \gamma + \tau_n \Omega_n^{1/2}\xi_n, \qquad \Omega_n = B_n^{-1/2}\Sigma_n B_n^{-1/2}. \tag{2.15}$$

REMARK 2.6—Coordinate meaning of sparsity: *The sparse minimax theory is stated in the score coordinates in which the denominator geometry is normalized. Equivalently, sparsity is a property of the canonical drift $\gamma = B_n^{-1/2}g$, not of an arbitrary raw dictionary. An arbitrary correlated denominator matrix can rotate the coordinate system and destroy raw-coordinate sparsity. In econometric embeddings, the sparse theory therefore applies to dictionaries that are orthogonalized or approximately diagonalized in the relevant score-denominator geometry, or to raw dictionaries for which B_n is sufficiently aligned with the chosen coordinates. The sparse power-regret frontier is therefore a canonical-coordinate result, with the econometric content supplied by dictionaries aligned with the score-denominator geometry.*

DEFINITION 2.7—Aligned canonical sparsity: *A raw dictionary is said to be aligned for the sparse frontier at denominator matrix B_n if the canonical drift $B_n^{-1/2}g$, rather than merely the raw drift g, is sparse or approximately sparse. In PLIV this means sparsity of $B_{n,\Delta}^{-1/2}g_n$. Alignment is automatic for an orthonormalized score dictionary with B_n approximately diagonal in the chosen coordinates, but it can fail when denominator normalization rotates a sparse raw first stage into a dense canonical direction. All sparse minimax statements below are conditional on this aligned canonical sparsity.*

The main high-dimensional frontier is stated for sparse shells in the canonical geometry, denoted $\mathcal{G}_n(s_n, r_n) = \{g \in \mathbb{R}^{p_n} : \|g\|_0 \le s_n,\ \|g\|_2 = r_n\}$. The sparse shell captures score dictionaries in which a small subset of canonical transformations carries identifying drift. Dense ellipsoid alternatives are retained as numerical benchmarks and are naturally paired with shrinkage rules; the main high-dimensional frontier theorem package below focuses on sparse shells.

Throughout the paper, G_n denotes a random weak drift under a prior, g denotes a fixed drift in minimax analysis, and $G_{n,\Delta}$ denotes the PLIV local drift associated with local direction Δ. The matrix B_n is always a denominator matrix; the normalized action set is $\mathcal{B}_n$.

3. BAYES DECISION BENCHMARK FOR THE FRONTIER

This section gives a decision-theoretic benchmark for the same learned scalar-score frontier. The minimax result in Section 4 characterizes worst-case finite-information loss over sparse weak-drift classes. The Bayes criterion studied here fixes a posterior law for the weak drift after training and asks which scalar AR direction maximizes posterior expected power. The comparison clarifies why feasible weak-ID score choice is a power problem rather than only a Rayleigh-quotient or expected-noncentrality problem.

3.1. *Exact posterior envelope*

Let G_n be a random weak drift with prior Π_n, independent of ξ_n. Given the training realization $Y_n = y$, a score coefficient $\beta \in \mathcal{B}_n$ delivers posterior expected power

$$\mathcal{P}_{\alpha,n}(\beta; y) = \mathbb{E}\left[h_\alpha\{Q_n(\beta; G_n, B_n)\} \mid Y_n = y\right]. \tag{3.1}$$

The Bayes feasible power frontier is the random value

$$\mathcal{V}_{\alpha,n}(y) = \sup_{\beta \in \mathcal{B}_n} \mathcal{P}_{\alpha,n}(\beta; y). \tag{3.2}$$

ASSUMPTION 3.1—Bayes regularity for posterior power: *For each n, the posterior kernel $y \mapsto \Pi_n(\cdot \mid y)$ is regular. The matrices B_n and Σ_n are positive definite.*

THEOREM 3.2—Bayes feasible power frontier: *Under Assumption 3.1, for each n and almost every y, the supremum in* (3.2) *is attained on $\mathcal{B}_n$. There exists a measurable Bayes power rule δ_n^P such that*

$$\delta_n^P(y) \in \operatorname*{arg\,max}_{\beta \in \mathcal{B}_n} \mathcal{P}_{\alpha,n}(\beta; y) \tag{3.3}$$

for almost every y. The ex ante Bayes feasible power value is

$$\mathbb{E}\big[\mathcal{V}_{\alpha,n}(Y_n)\big], \tag{3.4}$$

and no measurable score rule has larger prior expected local power. The proof is in Appendix A.2.

For comparison with posterior Rayleigh rules, assume additionally that

$$\mathbb{E}[\|G_n\|_2^2 \mid Y_n = y] < \infty \tag{3.5}$$

for the training realizations under consideration.

3.2. *Posterior power and posterior Rayleigh rules*

The posterior Rayleigh rule maximizes posterior expected noncentrality,

$$\mathcal{Q}_n(\beta; y) = \mathbb{E}[Q_n(\beta; G_n, B_n) \mid Y_n = y] = \rho_{\inf} \frac{\beta' M_n(y)\beta}{\beta' B_n \beta}, \qquad M_n(y) = \mathbb{E}[G_n G_n' \mid Y_n = y]. \tag{3.6}$$

It is a generalized eigenvector of $(M_n(y), B_n)$. The next result separates posterior expected power from posterior expected noncentrality.

PROPOSITION 3.3—Two-direction posterior and size-dependent power rules: *Let $B_n = I_2$, suppress n, and suppose that conditional on $Y = y$,*

$$G = \begin{cases} ae_1, & \textit{with probability } p, \\ be_2, & \textit{with probability } 1-p, \end{cases}$$

where $a, b > 0$, $p \in (0,1)$, and e_1, e_2 are coordinate vectors. Write a normalized score direction as $\beta(t) = (\sqrt{t}, \sqrt{1-t})'$, $\quad t \in [0,1]$. The posterior Rayleigh rule maximizes $pa^2 t + (1-p)b^2(1-t)$, and therefore selects $t = 1$ if $pa^2 > (1-p)b^2$, selects $t = 0$ if $pa^2 < (1-p)b^2$, and is indifferent if equality holds. The posterior expected-power rule maximizes

$$\Phi_\alpha(t) = ph_\alpha(\rho_{\inf} a^2 t) + (1-p)h_\alpha\{\rho_{\inf} b^2(1-t)\}. \tag{3.7}$$

Any interior maximizer satisfies

$$pa^2 h'_\alpha(\rho_{\inf} a^2 t) = (1-p)b^2 h'_\alpha\{\rho_{\inf} b^2(1-t)\}. \tag{3.8}$$

Consequently the posterior expected-power rule can depend on the nominal size α. If

$$pa^2 > (1-p)b^2 \quad \textit{but} \quad pa^2 h'_\alpha(\rho_{\inf} a^2) < (1-p)b^2 h'_\alpha(0), \tag{3.9}$$

then posterior Rayleigh selects $t = 1$, whereas posterior expected power does not. The proof is in Appendix A.3.

For example, with $\rho_{\inf} = 1$, $p = 0.95$, $a = 0.8$, and $b = 3$, direct maximization of $\Phi_\alpha(t)$ over $t \in [0,1]$ selects $t = 0$ at $\alpha = 0.01$ and $t = 1$ at $\alpha = 0.05$. The Rayleigh rule is unchanged and selects $t = 1$, because $pa^2 > (1-p)b^2$.

REMARK 3.4—When posterior power and Rayleigh approximately coincide: *If posterior uncertainty is concentrated on a single direction, or if all posterior noncentralities are uniformly small, the posterior-power objective is locally close to a Rayleigh objective. Indeed, for small q, $h_\alpha(q) = \alpha + h'_\alpha(0)q + o(q)$, so maximizing posterior expected power is first-order equivalent to maximizing posterior expected noncentrality. The noncoincidence in Proposition 3.3 therefore arises from posterior directional uncertainty together with nonlinearity of the power map away from the local linear region. This is the decision-theoretic role of the Bayes section in the paper.*

4. SPARSE MINIMAX FRONTIERS FOR LEARNED SCORE DIRECTIONS

This section proves the paper's organizing frontier in the Gaussian weak-drift experiment

$$Y = g + \sigma_n \xi, \qquad \xi \sim N(0, I_{p_n}).$$

The mathematical engine is the standard sparse normal-means geometry, but the payoff is a weak-ID scalar-AR payoff: squared projective direction loss is exactly normalized noncentrality loss, and on compact noncentrality ranges it is equivalent to local power regret. The single quantity

$$\eta_n = \sigma_n^2 s_n \log(ep_n/s_n)/r_n^2$$

therefore governs both the finite-information cost of learning the oracle scalar AR projection and the no-oracle-recovery implication under nondegenerate weak-ID oracle power.

The loss is normalized noncentrality regret,

$$L_n(\delta, g) = 1 - \frac{Q_n(\delta(Y); g, I_{p_n})}{Q_n^\star(g)} = 1 - \frac{\{\delta(Y)'g\}^2}{\|\delta(Y)\|^2 \|g\|^2}, \tag{4.1}$$

with the convention that $L_n = 1$ if $\delta(Y) = 0$. This is squared projective angular loss: score directions δ and $-\delta$ are equivalent for scalar AR power.

Let

$$\mathcal{G}_n(s_n, r_n) = \{g \in \mathbb{R}^{p_n} : \|g\|_0 \le s_n,\ \|g\|_2 = r_n\}, \tag{4.2}$$

and define

$$d_n = s_n \log(ep_n/s_n), \qquad \eta_n = \frac{\sigma_n^2 d_n}{r_n^2}. \tag{4.3}$$

The quantity η_n is the sparse finite-information difficulty index.

THEOREM 4.1—Sparse minimax frontier for normalized regret: *Suppose* $2 \le s_n \le p_n/4$. *There exist universal constants* $0 < c < C < \infty$ *such that*

$$c\min\{1, \eta_n\} \le \inf_{\delta_n} \sup_{g \in \mathcal{G}_n(s_n, r_n)} \mathbb{E}_g L_n(\delta_n, g) \le C\min\{1, \eta_n\}. \tag{4.4}$$

The infimum is over all measurable score rules $\delta_n : \mathbb{R}^{p_n} \to \mathbb{R}^{p_n}$, *with* $L_n = 1$ *when* $\delta_n(Y) = 0$. *The upper bound is attained up to constants by hard thresholding. Appendix A.4 gives the concise proof. The Online Appendix, Section 2, gives the full canonical-coordinate proof, including the projective-packing, Fano, perturbation-packing, and hard-thresholding derivations.*

The hard-thresholding upper bound is in the sparse normal-means thresholding tradition of Donoho and Johnstone (1994); the rate reflects the usual sparse high-dimensional complexity also familiar from minimax theory for sparse regression and related models (Raskutti, Wainwright, and Yu, 2011).

REMARK 4.2—Constructive role of hard thresholding: *The top-*s_n *hard-thresholding rule is used as an attainability device for the sparse minimax envelope: it shows that the canonical direction-learning rate in Theorem 4.1 is feasible in the Gaussian experiment. The lower-bound and no-oracle-recovery implications do not depend on this particular implementation. The theorem is a frontier result for score-direction learning; tuning-rule selection is separate from the econometric lower-bound object.*

LEMMA 4.3—Power-map comparability on compact noncentrality ranges: *Let* $h_\alpha(q) = 1 - F_{\chi_1^2(q)}(c_{1-\alpha})$. *For every* $\overline{q} < \infty$, *there exist constants* $0 < m_{\alpha,\overline{q}} \le M_{\alpha,\overline{q}} < \infty$ *such that, for all* $q \in [0, \overline{q}]$ *and* $\ell \in [0,1]$,

$$m_{\alpha,\overline{q}}\, q\ell \le h_\alpha(q) - h_\alpha\{q(1-\ell)\} \le M_{\alpha,\overline{q}}\, q\ell. \tag{4.5}$$

The proof is in Appendix A.5.

COROLLARY 4.4—Power-regret frontier on compact oracle-noncentrality ranges: *Suppose there exist constants* $0 < \underline{q} < \overline{q} < \infty$ *such that* $\rho_{\inf} r_n^2 \in [\underline{q}, \overline{q}]$. *Then there exist constants* $0 < c_\alpha < C_\alpha < \infty$, *depending only on* $\alpha, \underline{q}, \overline{q}$, *such that*

$$c_\alpha \min\{1, \eta_n\} \le \inf_{\delta_n} \sup_{g \in \mathcal{G}_n(s_n, r_n)} \mathcal{R}_{\alpha,n}(\delta_n, g) \le C_\alpha \min\{1, \eta_n\}. \tag{4.6}$$

The proof is in Appendix A.6; the Online Appendix, Corollary 2.7, records the corresponding canonical-coordinate transfer.

COROLLARY 4.5—Split-fraction frontier: *Let* $\rho \in (0,1)$ *denote the training fraction and* $1-\rho$ *the inference fraction in a single-split design. In the canonical sparse Gaussian experiment, define* $Y_\rho = g + \rho^{-1/2}\sigma_n \xi,\ \xi \sim N(0, I_{p_n})$, *and* $q_n^\star(\rho) = (1-\rho) r_n^2,\ \eta_n(\rho) = \frac{\rho^{-1}\sigma_n^2 s_n \log(ep_n/s_n)}{r_n^2}$. *Fix a nonempty split set* $\mathcal{R}_n = \left\{\rho \in [\underline{\rho}, 1-\underline{\rho}] : q_n^\star(\rho) \in [\underline{q}, \overline{q}]\right\}$, *where* $0 < \underline{\rho} < 1/2$ *and* $0 < \underline{q} < \overline{q} < \infty$. *Let* $\mathcal{R}^\rho_{\alpha,n}(\delta, g) = h_\alpha\{q_n^\star(\rho)\} - E_g h_\alpha\{Q_{n,\rho}(\delta(Y_\rho); g, I_{p_n})\}$, *where* $Q_{n,\rho}(\beta; g, I_{p_n}) = (1-\rho)\frac{(\beta' g)^2}{\|\beta\|_2^2}$. *Then, uniformly over* $\rho \in \mathcal{R}_n$, $\inf_{\delta_{n,\rho}} \sup_{g \in \mathcal{G}_n(s_n,r_n)} \mathcal{R}^\rho_{\alpha,n}(\delta_{n,\rho}, g) \asymp \min\{1, \eta_n(\rho)\}$. *Consequently, if* $\mathcal{V}^\rho_{\alpha,n} = \sup_{\delta_{n,\rho}} \inf_{g \in \mathcal{G}_n(s_n,r_n)} E_g h_\alpha\{Q_{n,\rho}(\delta_{n,\rho}(Y_\rho); g, I_{p_n})\}$, *then there exist constants* $0 < c_\alpha < C_\alpha < \infty$, *depending only on* $\alpha, \underline{q}, \overline{q}$, *such that, uniformly over* $\rho \in \mathcal{R}_n$, $h_\alpha\{q_n^\star(\rho)\} - C_\alpha \min\{1, \eta_n(\rho)\} \le \mathcal{V}^\rho_{\alpha,n} \le h_\alpha\{q_n^\star(\rho)\} - c_\alpha \min\{1, \eta_n(\rho)\}$. *Thus increasing the training fraction reduces score-learning noise but lowers inference-sample noncentrality. The exact optimizing split is constant-sensitive, but every split-design rule is governed by the displayed tradeoff between* $q_n^\star(\rho)$ *and* $\eta_n(\rho)$. *The proof is in Appendix A.7.*

COROLLARY 4.6—Sparse signal-to-noise frontier: *The minimax envelope implies the following signal-to-noise frontier.*

(i) If $\eta_n \ge 1$ *eventually, finite information prevents uniform recovery of the oracle score direction:* $\inf_{\delta_n} \sup_{g \in \mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(\delta_n, g) \ge c_0$ *for a universal constant* $c_0 > 0$.

(ii) If $\eta_n \to 0$, *the hard-thresholding upper-bound rule* H_{s_n} *satisfies* $\sup_{g \in \mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(H_{s_n}, g) = O(\eta_n)$.

(iii) If, in addition, $\rho_{\inf} r_n^2 \in [\underline{q}, \overline{q}] \subset (0, \infty)$, *the same low-signal and high-signal conclusions transfer to power regret through Corollary 4.4.*

The proof is in Appendix A.8.

COROLLARY 4.7—No oracle-score recovery under nondegenerate weak-ID power: *Suppose there exist constants* $0 < \underline{q} < \overline{q} < \infty$, $\underline{\rho} > 0$, *and* $\underline{\sigma} > 0$ *such that, eventually,* $\rho_{\inf} \ge \underline{\rho}$, $\rho_{\inf} r_n^2 \in [\underline{q}, \overline{q}]$, $\sigma_n^2 \ge \underline{\sigma}^2$. *If* $s_n \log(ep_n/s_n) \to \infty$, *then* $\eta_n \to \infty$, *and there exists* $c_\alpha > 0$ *such that* $\inf_{\delta_n} \sup_{g \in \mathcal{G}_n(s_n,r_n)} \mathcal{R}_{\alpha,n}(\delta_n, g) \ge c_\alpha$ *eventually. Thus, under nondegenerate weak-ID oracle power and nonvanishing training noise, growing sparse complexity rules out uniform oracle-power recovery by learned scalar scores. The proof is in Appendix A.9.*

PROPOSITION 4.8—Scalarization benefit against many-moment AR critical values: *Let* $p_n \to \infty$. *Suppose a full-vector AR statistic has local limit* $\chi^2_{p_n}(q_n)$, $q_n \to q \in (0, \infty)$, *and rejects at the level-*α *central critical value* $c_{p_n,1-\alpha}$. *Then* $\Pr\{\chi^2_{p_n}(q_n) > c_{p_n,1-\alpha}\} \to \alpha$. *By contrast, an oracle scalar AR score with the same limiting noncentrality* q *has limiting power* $h_\alpha(q) = \Pr\{\chi^2_1(q) > c_{1-\alpha}\} > \alpha$. *Thus scalar score learning trades a many-moment critical-value cost for the finite-information cost of learning the relevant direction in this benchmark. The proposition isolates the scalarization tradeoff: many-moment AR pays a growing critical-value cost, while learned scalar AR pays a finite-information score-learning cost. The proof is in Appendix A.10.*

REMARK 4.9—Weak-ID scaling and high-dimensional recoverability: *Corollary 4.7 isolates the conventional weak-ID implication of the sparse frontier. The quantity* $\eta_n = \sigma_n^2 s_n \log(ep_n/s_n)/r_n^2$ *measures training-side direction-learning difficulty, whereas* $\rho_{\inf} r_n^2$ *measures oracle inference-side noncentrality. In exact local-weak sequences with nondegenerate oracle power,* $r_n^2 = O(1)$. *Therefore growing sparse complexity requires either increasing effective training information, decreasing training noise, or a larger weak-drift norm to recover oracle scalar-score power.*

REMARK 4.10—Canonical-coordinate and procedure scope: *The sparse minimax envelope is stated in canonical Gaussian sequence coordinates. Correlated denominator and training-covariance geometries must first be transformed to the normalized coordinates described in Section 2 and in the PLIV embedding. The theorem is a frontier for learned linear score directions entering scalar Anderson–Rubin procedures. Broader weak-identification robust procedures have different critical-value and conditioning structures.*

The next section maps this canonical frontier into an econometric weak-IV model. In PLIV, the abstract radius r_n^2 becomes the squared norm of the denominator-normalized first-stage drift,

$$\|B_{n,\Delta}^{-1/2} g_n\|^2,$$

and the sparse difficulty becomes

$$\eta_{n,\Delta} = \frac{\rho_{\mathrm{tr}}^{-1} \sigma_\Delta^2 s_n \log(ep_n/s_n)}{\|B_{n,\Delta}^{-1/2} g_n\|^2}.$$

5. GROWING-DICTIONARY PLIV LEARNED-SCORE FRONTIER

This section is the central econometric realization of the canonical frontier. The primitive experiment and sparse minimax theorem identify the finite-information cost of learning a scalar AR direction. The PLIV embedding shows when that canonical experiment is generated by a growing instrument dictionary. The key object is the denominator-normalized first-stage drift $B_{n,\Delta}^{-1/2} g_n$; it is this object, not the raw first-stage coefficient vector, whose complexity determines feasible scalar-AR power. The section is organized in three layers. The first subsections define the PLIV benchmark, the training statistic, and the feasible sample-split AR statistic. The operational reduction then shows that, conditional on the learned score, the feasible PLIV statistic has the scalar-AR local-power representation of Section 2. The remaining subsections give sufficient primitive alignment and coupling conditions, a rotated canonical-factor example, and the PLIV learned-score frontier.

5.1. *PLIV benchmark: model, centering, and score dictionary*

Let $Y = (\theta_0 + \Delta)D + g_0(X) + U$, $\mathbb{E}[U \mid Z, X] = 0$. Start from the local first-stage representation

$$D - m_0(X) = N^{-1/2} \pi_n^{\mathrm{raw}}(Z, X) + V, \qquad \mathbb{E}[V \mid Z, X] = 0. \tag{5.1}$$

The raw drift in (5.1) need not be conditionally centered. Define $\bar{\pi}_n(X) = \mathbb{E}[\pi_n^{\mathrm{raw}}(Z, X) \mid X]$, $\pi_n^{\mathrm{c}}(Z, X) = \pi_n^{\mathrm{raw}}(Z, X) - \bar{\pi}_n(X)$, and $m_N(X) = m_0(X) + N^{-1/2}\bar{\pi}_n(X)$. Then

$$D - m_N(X) = N^{-1/2} \pi_n^{\mathrm{c}}(Z, X) + V, \qquad \mathbb{E}[\pi_n^{\mathrm{c}}(Z, X) \mid X] = 0. \tag{5.2}$$

Let $b_{p_n}(Z,X) \in \mathbb{R}^{p_n}$ be a growing dictionary satisfying

$$\mathbb{E}[b_{p_n}(Z,X) \mid X] = 0. \tag{5.3}$$

Because the dictionary is conditionally centered, $\mathbb{E}[b_{p_n}(Z,X)\pi_n^{\mathrm{c}}(Z,X)] = \mathbb{E}[b_{p_n}(Z,X)\pi_n^{\mathrm{raw}}(Z,X)]$. Below, write π_n for the centered drift π_n^{c} and take m_N as the residualization target. Define

$$\begin{aligned} g_n &= \mathbb{E}[b_{p_n}(Z,X)\pi_n(Z,X)], \\ G_{n,\Delta} &= \Delta g_n, \\ B_{n,\Delta} &= \mathbb{E}[\omega_\Delta(Z,X) b_{p_n}(Z,X) b_{p_n}(Z,X)'], \\ C_n &= \mathbb{E}[\sigma_V^2(Z,X) b_{p_n}(Z,X) b_{p_n}(Z,X)'], \end{aligned} \tag{5.4}$$

where

$$\omega_\Delta(Z,X) = \mathbb{E}[(U+\Delta V)^2 \mid Z,X], \qquad \sigma_V^2(Z,X) = \mathbb{E}[V^2 \mid Z,X].$$

For a score coefficient β, the PLIV local noncentrality is

$$Q_{n,\Delta}^{\mathrm{pliv}}(\beta) = \rho_{\inf} \frac{(\beta' G_{n,\Delta})^2}{\beta' B_{n,\Delta}\beta} = \rho_{\inf}\Delta^2 \frac{(\beta' g_n)^2}{\beta' B_{n,\Delta}\beta}. \tag{5.5}$$

The training statistic is $\widehat{g}_n = \frac{\sqrt{N}}{n_{\mathrm{tr}}} \sum_{i \in I_{\mathrm{tr}}} b_{p_n}(Z_i,X_i)\{D_i - \widehat{m}(X_i)\}$. Throughout Section 5, $\frac{n_{\mathrm{tr}}}{N} \to \rho_{\mathrm{tr}} \in (0,1)$, $\frac{n_{\inf}}{N} \to \rho_{\inf} \in (0,1]$. The training and inference samples are independent conditional on the triangular-array law.

5.2. *Operational learned-score PLIV statistic*

Let $\widehat{\beta}_n = \delta_n(\widehat{g}_n)$ be a training-measurable score coefficient and $\widehat{s}_n(Z,X) = b_{p_n}(Z,X)'\widehat{\beta}_n$. On the inference sample define

$$\widehat{T}_{n,\widehat{\beta}_n}(\theta_0) = \frac{n_{\inf}^{-1/2} \sum_{i \in I_{\inf}} \widehat{\beta}_n' b_{p_n}(Z_i,X_i)\{Y_i - \widehat{\mu}_\Delta(X_i) - \theta_0[D_i - \widehat{m}_\Delta(X_i)]\}}{\{\widehat{\beta}_n' \widehat{B}_{n,\Delta}^{\inf} \widehat{\beta}_n\}^{1/2}}, \tag{5.6}$$

where $\widehat{\mu}_\Delta$, $\widehat{m}_\Delta$, and $\widehat{B}_{n,\Delta}^{\inf}$ are cross-fit relative to the inference observations. Let $\widehat{AR}_{n,\widehat{s}_n}(\theta_0) = \widehat{T}_{n,\widehat{\beta}_n}(\theta_0)^2$. The oracle residualization targets under $P_{N,\Delta}$ are $\mu_{N,\Delta}(X) = \mathbb{E}_{P_{N,\Delta}}[Y \mid X]$, $\quad m_N(X) = \mathbb{E}_{P_{N,\Delta}}[D \mid X]$. The Δ-indexed notation in (5.6) is used for a fixed-local-alternative power expansion. The local direction is not observed by the practitioner; it indexes the sequence along which local power is evaluated. In confidence-set inversion the same notation is read pointwise: for each tested value one forms the residuals and variance estimator appropriate to that tested value, while the local-power calculation evaluates the resulting statistic along a fixed sequence $P_{N,\Delta}$.

5.3. *Operational conditions for the PLIV reduction*

For fixed Δ, define the denominator-bounded score set $\mathcal{H}_{n,\Delta}(c,C) = \{\beta \in \mathbb{R}^{p_n} : c \leq \beta' B_{n,\Delta}\beta \leq C\}$. Let

$$\mathcal{F}_{\rm tr} = \sigma\{(Y_i, D_i, Z_i, X_i) : i \in I_{\rm tr}\},$$
$$\mathcal{D}_{\rm tr} = \sigma\{(Z_i, X_i) : i \in I_{\rm tr}\},$$
$$\mathcal{D}_{\rm inf} = \sigma\{(Z_i, X_i) : i \in I_{\rm inf}\}.$$

ASSUMPTION 5.1—Operational PLIV reduction conditions: *Fix Δ. The following conditions hold for constants $0 < c < C < \infty$ and for the realized score rules considered below.*

(i) Splitting, centering, and eigenvalues. *The split samples are independent conditional on the triangular-array law, the centered first-stage representation* (5.2) *holds, and*

$$cI_{p_n} \preceq C_n \preceq CI_{p_n}, \qquad cI_{p_n} \preceq B_{n,\Delta} \preceq CI_{p_n}.$$

(ii) Training representation. *The drift sampling and training residualization errors satisfy*

$$\left\| C_n^{-1/2} \left\{ \frac{1}{n_{\rm tr}} \sum_{i \in I_{\rm tr}} b_{p_n}(Z_i, X_i)\pi_n(Z_i, X_i) - g_n \right\} \right\|_2 = o_p(1) \tag{5.7}$$

and

$$\left\| C_n^{-1/2} \frac{\sqrt{N}}{n_{\rm tr}} \sum_{i \in I_{\rm tr}} b_{p_n}(Z_i, X_i)\{\widehat{m}(X_i) - m_N(X_i)\} \right\|_2 = o_p(1). \tag{5.8}$$

Moreover,

$$d_{\rm BL}\left(\mathcal{L}\left\{ C_n^{-1/2} \frac{\sqrt{N}}{n_{\rm tr}} \sum_{i \in I_{\rm tr}} b_{p_n}(Z_i, X_i)V_i \,\middle|\, \mathcal{D}_{\rm tr} \right\}, N(0, \rho_{\rm tr}^{-1} I_{p_n}) \right) \xrightarrow{p} 0, \tag{5.9}$$

where $\mathcal{D}_{\rm tr} = \sigma\{(Z_i, X_i) : i \in I_{\rm tr}\}$. Condition (5.9) *is an operational high-dimensional Gaussian approximation; Propositions 5.4 and 5.5 below give sufficient conditions and separate operator-norm covariance control from the stronger trace/Frobenius control needed for exact canonical equivalence.*

(iii) Inference drift, variance, and nuisance remainders. *Let $\widehat{g}_n^{\rm inf} = \frac{1}{n_{\rm inf}} \sum_{i \in I_{\rm inf}} b_{p_n}(Z_i, X_i)\pi_n(Z_i, X_i)$. Then*

$$\left\| B_{n,\Delta}^{-1/2}(\widehat{g}_n^{\rm inf} - g_n) \right\|_2 = o_p(1), \tag{5.10}$$

and, uniformly over $\beta \in \mathcal{H}_{n,\Delta}(c,C)$,

$$\left| \frac{\beta' \widehat{B}_{n,\Delta}^{\rm inf} \beta}{\beta' B_{n,\Delta}\beta} - 1 \right| = o_p(1), \tag{5.11}$$

$$\frac{\left| n_{\rm inf}^{-1/2} \sum_{i \in I_{\rm inf}} \beta' b_{p_n}(Z_i, X_i)\{[\widehat{\mu}_\Delta - \mu_{N,\Delta}](X_i) - \theta_0[\widehat{m}_\Delta - m_N](X_i)\} \right|}{(\beta' B_{n,\Delta}\beta)^{1/2}} = o_p(1). \tag{5.12}$$

(iv) Conditional scalar CLT. *For $R_{i,\Delta} = U_i + \Delta V_i$, uniformly over $\beta \in \mathcal{H}_{n,\Delta}(c,C)$,*

$$d_{\mathrm{BL}}\left(\mathcal{L}\left\{\frac{n_{\inf}^{-1/2}\sum_{i\in I_{\inf}}\beta' b_{p_n}(Z_i,X_i)R_{i,\Delta}}{(\beta' B_{n,\Delta}\beta)^{1/2}}\,\middle|\,\mathcal{F}_{\mathrm{tr}}\vee\mathcal{D}_{\inf}\right\}, N(0,1)\right)\xrightarrow{p}0. \tag{5.13}$$

(v) Bounded local frontier. *The oracle PLIV local noncentrality is bounded:*

$$G'_{n,\Delta}B^{-1}_{n,\Delta}G_{n,\Delta} = O(1). \tag{5.14}$$

All remainders above are uniform over the PLIV classes used in the minimax statements below.

THEOREM 5.2—Operational PLIV learned-score AR reduction: *Suppose Assumption 5.1 holds. Let $\widehat{\beta}_n = \delta_n(\widehat{g}_n)$ be training-measurable and suppose $\Pr\{\widehat{\beta}_n \in \mathcal{H}_{n,\Delta}(c,C)\} \to 1$. Then:*

(i) Training representation. *Let $Z_{n,\mathrm{tr}} = C_n^{-1/2}\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_{p_n}(Z_i,X_i)V_i$. Then $C_n^{-1/2}(\widehat{g}_n - g_n) = Z_{n,\mathrm{tr}} + a_{n,\mathrm{tr}}$, $\|a_{n,\mathrm{tr}}\|_2 = o_p(1)$, and $d_{\mathrm{BL}}\big(\mathcal{L}\{Z_{n,\mathrm{tr}} \mid \mathcal{D}_{\mathrm{tr}}\}, N(0,\rho_{\mathrm{tr}}^{-1}I_{p_n})\big) \xrightarrow{p} 0$. No Euclidean coupling is asserted in this operational representation. Such a coupling is imposed only under the stronger sparse-frontier conditions in Theorem 5.12 and Assumption 5.8.*

(ii) Feasible signed-shift approximation. *Conditional on the training sigma-field,*

$$d_{\mathrm{BL}}\left(\mathcal{L}\{\widehat{T}_{n,\widehat{\beta}_n}(\theta_0)\mid\mathcal{F}_{\mathrm{tr}}\}, N\left(\sqrt{\rho_{\inf}}\frac{\widehat{\beta}'_n G_{n,\Delta}}{(\widehat{\beta}'_n B_{n,\Delta}\widehat{\beta}_n)^{1/2}}, 1\right)\right)\xrightarrow{p}0. \tag{5.15}$$

(iii) Size and local power. *For the null size statement, Assumption 5.1 is understood with $\Delta = 0$. For the local-power statement, it is understood at the fixed local direction Δ. Under the null, $\Pr_{N,0}\{\widehat{AR}_{n,\widehat{s}_n}(\theta_0) > c_{1-\alpha}\} = \alpha + o(1)$. Under $P_{N,\Delta}$,*

$$\Pr_{N,\Delta}\{\widehat{AR}_{n,\widehat{s}_n}(\theta_0) > c_{1-\alpha}\} = \mathbb{E}_{N,\Delta}\left[h_\alpha\left(\rho_{\inf}\frac{\{\delta_n(\widehat{g}_n)'G_{n,\Delta}\}^2}{\delta_n(\widehat{g}_n)'B_{n,\Delta}\delta_n(\widehat{g}_n)}\right)\right] + o(1). \tag{5.16}$$

Thus the feasible PLIV statistic satisfies the primitive learned-score scalar AR experiment with C_n, inference drift $G_{n,\Delta}$, and denominator $B_{n,\Delta}$. The proof is in Appendix A.11.

REMARK 5.3—Uniform PLIV reduction: *The operational PLIV reduction is pointwise in notation but is used in its class-uniform form. When Assumption 5.1 holds uniformly over a PLIV class $\mathfrak{P}_{n,\Delta}$, the $o_p(1)$ remainder in the rejection-probability representation is uniform over that class. In particular,*

$$\sup_{P\in\mathfrak{P}_{n,\Delta}}\left|P\{\widehat{AR}_{n,\widehat{s}_n}(\theta_0) > c_{1-\alpha}\} - E_P h_\alpha\left(\rho_{\inf}\frac{(\widehat{\beta}'_n G_{n,\Delta})^2}{\widehat{\beta}'_n B_{n,\Delta}\widehat{\beta}_n}\right)\right| \to 0.$$

Thus weak-ID uniformity is imposed or verified through the uniform operational and primitive PLIV conditions, not inferred from pointwise convergence.

5.4. *Primitive verification and canonical coupling*

The operational conditions above are intentionally stated in the exact score-weighted quantities that enter the statistic. This subsection records primitive sufficient conditions. Two distinctions are important. First, operator-norm covariance consistency gives a covariance-perturbed Gaussian experiment, but it does not by itself give bounded-Lipschitz convergence to the exact canonical $N(0, I_{p_n})$ law when $p_n \to \infty$. Exact canonical equivalence requires a trace or Frobenius condition. Second, the sparse upper bound for hard thresholding requires an L_2-coupling remainder, not only weak convergence of the training law.

PROPOSITION 5.4—Gaussian first-stage covariance and canonical equivalence: *Suppose, conditional on* $(Z_i, X_i)_{i\in I_{\rm tr}}$*, the first-stage residuals* V_i *are independent Gaussian random variables with* $\mathbb{E}[V_i \mid Z_i, X_i] = 0$, $\mathbb{E}[V_i^2 \mid Z_i, X_i] = \sigma_V^2(Z_i, X_i)$. *Let* $\widehat{C}_n^{\rm tr} = \frac{1}{n_{\rm tr}} \sum_{i\in I_{\rm tr}} b_{p_n}(Z_i, X_i) b_{p_n}(Z_i, X_i)' \sigma_V^2(Z_i, X_i)$ *and* $R_{C,n} = C_n^{-1/2} \widehat{C}_n^{\rm tr} C_n^{-1/2} - I_{p_n}$, $\widetilde{R}_{C,n} = \rho_{\rm tr} \frac{N}{n_{\rm tr}} (I_{p_n} + R_{C,n}) - I_{p_n}$. *Assume* (5.7), (5.8), *and* $\|R_{C,n}\|_{\rm op} = o_p(1)$. *Then*

$$
\begin{gathered}
C_n^{-1/2}(\widehat{g}_n - g_n) = \zeta_n + r_{n,\rm tr}, \\
\zeta_n \mid \mathcal{D}_{\rm tr} \sim N\{0, \rho_{\rm tr}^{-1}(I_{p_n} + \widetilde{R}_{C,n})\}, \\
\|r_{n,\rm tr}\|_2 = o_p(1).
\end{gathered} \tag{5.17}
$$

If, in addition, $\mathrm{tr}(\widetilde{R}_{C,n}^2) = o_p(1)$, *then* (5.9) *holds. The proof is in Online Appendix 4.1.*

PROPOSITION 5.5—Non-Gaussian training approximation and coupling: *Let* $X_{ni} = C_n^{-1/2} b_{p_n}(Z_i, X_i) V_i$, $i \in I_{\rm tr}$, *and suppose* $\mathbb{E}[X_{ni} \mid \mathcal{D}_{\rm tr}] = 0$. *Assume a conditional multivariate normal approximation for the normalized score-weighted array:*

$$
d_{\rm BL}\left(\mathcal{L}\left\{ \frac{\sqrt{N}}{n_{\rm tr}} \sum_{i\in I_{\rm tr}} X_{ni} \,\middle|\, \mathcal{D}_{\rm tr} \right\}, N\left(0, \frac{N}{n_{\rm tr}} C_n^{-1/2} \widehat{C}_n^{\rm tr} C_n^{-1/2}\right)\right) \xrightarrow{p} 0. \tag{5.18}
$$

A primitive input for high-dimensional Gaussian approximation, Berry–Esseen, or Yurinskii-type bounds is the vector third-moment quantity $\mathfrak{b}_{n,\rm tr} := \frac{N^{3/2}}{n_{\rm tr}^3} \sum_{i\in I_{\rm tr}} \mathbb{E}\big[\|X_{ni}\|_2^3 \mid \mathcal{D}_{\rm tr}\big] = o_p(1)$. *Set* $\widetilde{R}_{C,n} = \rho_{\rm tr} \frac{N}{n_{\rm tr}} C_n^{-1/2} \widehat{C}_n^{\rm tr} C_n^{-1/2} - I_{p_n}$. *If* (5.18) *holds and* $\mathrm{tr}(\widetilde{R}_{C,n}^2) = o_p(1)$, *then* (5.9) *follows. In particular, if* $K_{\sigma C,n} := \sup_{z,x} \sigma_V(z,x) \|C_n^{-1/2} b_{p_n}(z,x)\|_2$ *and* $\sup_{z,x} \frac{\mathbb{E}[|V|^3 | Z=z, X=x]}{\sigma_V^3(z,x)} < \infty$, *then* $\mathfrak{b}_{n,\rm tr} = O_p(K_{\sigma C,n}^3 / \sqrt{n_{\rm tr}})$. *For the usual weighted leverage scale* $K_{\sigma C,n} = O(\sqrt{p_n})$, *this is the familiar requirement* $p_n^{3/2} / \sqrt{n_{\rm tr}} \to 0$.

For sparse upper bounds, the bounded-Lipschitz approximation in (5.18) *is not enough. The needed additional input is a quadratic Wasserstein, or equivalently Euclidean coupling, rate in the denominator-normalized coordinates used by the sparse theorem. Here and below,* W_2 *denotes quadratic Wasserstein distance on* $\mathbb{R}^{p_n}$ *induced by the Euclidean norm:*

$$
W_2^2(P, Q) = \inf_{\pi\in\Pi(P,Q)} \int \|x - y\|_2^2 \, d\pi(x, y),
$$

where $\Pi(P, Q)$ *is the set of couplings of* P *and* Q. *Conditional* W_2 *distances are interpreted after conditioning on the relevant design sigma-field; an outer expectation averages over that sigma-field. Let*

$$
S_{n,\Delta}^B = B_{n,\Delta}^{-1/2} \frac{\sqrt{N}}{n_{\rm tr}} \sum_{i\in I_{\rm tr}} b_{p_n}(Z_i, X_i) V_i, \qquad \Gamma_{n,\Delta}^B = \rho_{\rm tr}^{-1} \sigma_\Delta^2 (I_{p_n} + \Theta_{n,\Delta}^{\rm tr}),
$$

with $\Theta^{\mathrm{tr}}_{n,\Delta}$ as in (5.24) *below, and with r_n and $\eta_{n,\Delta}$ defined by the sparse PLIV problem. If, in addition,*

$$\frac{\mathbb{E}\, W_2^2\big(\mathcal{L}\{S^B_{n,\Delta} \mid \mathcal{D}_{\mathrm{tr}}\}, N(0,\Gamma^B_{n,\Delta})\big)}{r_n^2} = o\{\min(1,\eta_{n,\Delta})\}, \tag{5.19}$$

then the non-Gaussian branch supplies the non-Gaussian part of the Euclidean coupling. Combined with the denominator-normalized covariance perturbation rate (5.27)*, this gives the canonical Euclidean coupling required in the sparse PLIV upper bound.*

Condition (5.19) *may be verified in applications using quantitative strong-approximation tools, such as Yurinskii- or Zaitsev-type couplings, under suitable moment, leverage, and dimension-growth restrictions. It is not implied by the bounded-Lipschitz approximation alone (Yurinskii, 1977, Zaitsev, 1987, Chernozhukov, Chetverikov, and Kato, 2017). The proof is in Online Appendix 4.2.*

PROPOSITION 5.6—Scalar inference CLT from leverage and third moments: *Fix Δ. Suppose, conditional on the inference-side design, the variables $R_{i,\Delta} = U_i + \Delta V_i$ are independent, mean zero, with $\mathbb{E}[R^2_{i,\Delta} \mid Z_i, X_i] = \omega_\Delta(Z_i, X_i)$.*

Let $K_{\omega B,n} = \sup_{z,x} \omega_\Delta(z,x)^{1/2} \|B^{-1/2}_{n,\Delta} b_{p_n}(z,x)\|_2$. Assume

$$K_{\omega B,n}/\sqrt{n_{\mathrm{inf}}} \to 0, \qquad \sup_{z,x} \frac{\mathbb{E}[|R_{i,\Delta}|^3 \mid Z_i = z, X_i = x]}{\omega_\Delta(z,x)^{3/2}} < \infty,$$

and the oracle covariance concentration

$$\left\| B^{-1/2}_{n,\Delta} \left\{ \frac{1}{n_{\mathrm{inf}}} \sum_{i \in I_{\mathrm{inf}}} b_i b_i' \omega_\Delta(Z_i, X_i) - B_{n,\Delta} \right\} B^{-1/2}_{n,\Delta} \right\|_{\mathrm{op}} = o_p(1).$$

Then the uniform conditional scalar CLT (5.13) *holds. If the feasible variance estimator is operator-norm consistent in the same geometry, then* (5.11) *also holds. The proof is in Online Appendix 4.3.*

PROPOSITION 5.7—Denominator-noise alignment: operator and trace forms: *Suppose, for fixed Δ, there exists $\kappa_\Delta > 0$ such that*

$$\epsilon_{n,\Delta} := \sup_{z,x} \left| \frac{\sigma_V^2(z,x)}{\omega_\Delta(z,x)} - \kappa_\Delta \right| \to 0,$$

and that $B_{n,\Delta} = \mathbb{E}[\omega_\Delta b_{p_n} b'_{p_n}]$ has eigenvalues bounded away from zero and infinity. Then

$$\left\| B^{-1/2}_{n,\Delta} C_n B^{-1/2}_{n,\Delta} - \kappa_\Delta I_{p_n} \right\|_{\mathrm{op}} \le \epsilon_{n,\Delta}. \tag{5.20}$$

Moreover,

$$\mathrm{tr}\left[\left\{ B^{-1/2}_{n,\Delta} C_n B^{-1/2}_{n,\Delta} - \kappa_\Delta I_{p_n} \right\}^2 \right] \le p_n \epsilon^2_{n,\Delta}. \tag{5.21}$$

Thus operator alignment follows from $\epsilon_{n,\Delta} \to 0$, while trace alignment for exact canonical coupling requires the stronger rate implied by the right side of (5.21)*. The proof is in Online Appendix 4.4.*

5.5. *A primitive aligned PLIV verification*

For the sparse frontier, normalize the training statistic by the inference denominator:

$$\widetilde{Z}_{n,\Delta} = B_{n,\Delta}^{-1/2}\widehat{g}_n, \qquad \widetilde{\gamma}_{n,\Delta} = B_{n,\Delta}^{-1/2} g_n.$$

The next assumption gives primitive rates that verify both the operational AR reduction and the stronger L_2-coupling needed by the sparse upper bound.

ASSUMPTION 5.8—Primitive aligned PLIV design with strong training coupling: *Fix* Δ. *The following conditions hold.*

(i) Centering, eigenvalues, and canonical alignment. *The centered representation* (5.2) *and dictionary centering* (5.3) *hold. The eigenvalues of* $B_{n,\Delta}$ *and* C_n *are bounded away from zero and infinity. For some* $0 < \underline{\sigma} \leq \sigma_\Delta \leq \overline{\sigma} < \infty$,

$$\left\| B_{n,\Delta}^{-1/2} C_n B_{n,\Delta}^{-1/2} - \sigma_\Delta^2 I_{p_n} \right\|_{\text{op}} \to 0. \tag{5.22}$$

(ii) Drift and nuisance rates. *The operational drift and nuisance remainders in* (5.7)–(5.12) *hold. For sparse-frontier upper bounds, with* $\eta_{n,\Delta} = \rho_{\text{tr}}^{-1}\sigma_\Delta^2 s_n \log(ep_n/s_n)/r_n^2$, $r_n = \|\widetilde{\gamma}_{n,\Delta}\|_2$, *the stronger* $B_{n,\Delta}$*-normalized training rate holds:*

$$\frac{\mathbb{E}\left\| B_{n,\Delta}^{-1/2}\left[\left\{\frac{1}{n_{\text{tr}}}\sum_{i\in I_{\text{tr}}} b_i\pi_i - g_n\right\} - \frac{\sqrt{N}}{n_{\text{tr}}}\sum_{i\in I_{\text{tr}}} b_i\{\widehat{m}(X_i) - m_N(X_i)\}\right]\right\|_2^2}{r_n^2} = o\{\min(1, \eta_{n,\Delta})\}. \tag{5.23}$$

(iii) Training covariance trace coupling. *Let* $\widehat{C}_n^{\text{tr}} = \frac{1}{n_{\text{tr}}}\sum_{i\in I_{\text{tr}}} b_i b_i' \sigma_V^2(Z_i, X_i)$ *and*

$$\Theta_{n,\Delta}^{\text{tr}} = \rho_{\text{tr}}\frac{N}{n_{\text{tr}}}\sigma_\Delta^{-2} B_{n,\Delta}^{-1/2}\widehat{C}_n^{\text{tr}} B_{n,\Delta}^{-1/2} - I_{p_n}. \tag{5.24}$$

Also define the C_n*-normalized covariance perturbation* $\widetilde{R}_{C,n}^{\text{tr}} = \rho_{\text{tr}}\frac{N}{n_{\text{tr}}} C_n^{-1/2}\widehat{C}_n^{\text{tr}} C_n^{-1/2} - I_{p_n}$. *For the operational Gaussian approximation in* (5.9), *require*

$$\|\widetilde{R}_{C,n}^{\text{tr}}\|_{\text{op}} = o_p(1), \qquad \text{tr}\{(\widetilde{R}_{C,n}^{\text{tr}})^2\} = o_p(1). \tag{5.25}$$

This is the C_n*-normalized trace condition used by Propositions 5.4 and 5.5. The denominator-normalized perturbation* $\Theta_{n,\Delta}^{\text{tr}}$ *is the one used by the sparse upper-bound coupling. For it, require*

$$\|\Theta_{n,\Delta}^{\text{tr}}\|_{\text{op}} = o_p(1), \qquad \text{tr}\{(\Theta_{n,\Delta}^{\text{tr}})^2\} = o_p(1). \tag{5.26}$$

For sparse-frontier upper bounds, the stronger mean trace rate holds:

$$\frac{\rho_{\text{tr}}^{-1}\sigma_\Delta^2 \mathbb{E}\,\text{tr}\{(\Theta_{n,\Delta}^{\text{tr}})^2\}}{r_n^2} = o\{\min(1, \eta_{n,\Delta})\}. \tag{5.27}$$

(iv) Training law, strong-coupling branch, and inference CLT. *For the operational AR reduction, either the first-stage residuals are conditionally Gaussian as in Proposition 5.4, or the direct bounded-Lipschitz vector approximation in Proposition 5.5 holds. For the sparse upper-bound conclusion, the denominator-normalized mean trace rate* (5.27) *in part (iii) is* *required in all cases. In addition, the training side satisfies one of two strong-coupling branches: either the first-stage residuals are conditionally Gaussian, or the non-Gaussian Wasserstein coupling rate* (5.19) *holds. The non-Gaussian branch replaces Gaussianity of the training law; it does not replace the covariance-perturbation trace rate in* (5.27)*. On the inference side, either* $R_{i,\Delta}$ *is conditionally Gaussian, or the leverage and moment conditions of Proposition 5.6 hold. Variance consistency* (5.11) *holds.*

(v) Bounded frontier and canonical sparsity. $G'_{n,\Delta}B^{-1}_{n,\Delta}G_{n,\Delta} = O(1)$. *For sparse-frontier statements,* $\widetilde{\gamma}_{n,\Delta} \in \mathcal{G}_n(s_n, r_n)$*, with* $2 \le s_n \le p_n/4$.

THEOREM 5.9—Primitive aligned PLIV verification: *Under Assumption 5.8, the operational conditions in Assumption 5.1 hold, and Theorem 5.2 applies.*

For the sparse-frontier components, suppose that (5.23) *holds, that the denominator-normalized mean trace rate* (5.27) *in Assumption 5.8(iii) holds, and that the training law satisfies one of the two strong-coupling branches in Assumption 5.8(iv). Then there exists a coupling with* $\xi_n \sim N(0, I_{p_n})$ *such that, uniformly over the primitive aligned sparse PLIV class,*

$$\widetilde{Z}_{n,\Delta} = \widetilde{\gamma}_{n,\Delta} + \rho_{\mathrm{tr}}^{-1/2}\sigma_\Delta \xi_n + e_{n,\Delta}, \qquad \frac{\mathbb{E}\|e_{n,\Delta}\|_2^2}{r_n^2} = o\{\min(1, \eta_{n,\Delta})\}. \tag{5.28}$$

Consequently the upper-bound part of Theorem 5.12 applies. The direct bounded-Lipschitz non-Gaussian approximation alone verifies the operational AR reduction, but not this Euclidean coupling. The proof is in Online Appendix 4.5.

REMARK 5.10—Interpretation of the aligned design: *Assumption 5.8 characterizes aligned dictionaries for which the denominator-normalized sparse frontier can be transported from the Gaussian sequence experiment to PLIV. The conditions cover score dictionaries that are close to the AR denominator geometry and nuisance estimates that are small in the score-weighted directions entering the scalar statistic. Raw-coordinate sparsity is not enough when* $B^{-1/2}_{n,\Delta}$ *is highly rotational; the sparse theorem applies to the canonical PLIV drift* $B^{-1/2}_{n,\Delta}g_n$.

5.6. *Aligned orthonormalized dictionary example*

This aligned design arises in a stylized but useful case after residualizing controls and orthonormalizing the instrument dictionary in the AR denominator geometry. Suppose $b_{p_n}(Z, X)$ is conditionally centered and normalized so that $\mathbb{E}[b_{p_n} b'_{p_n}] = I_{p_n}$. Suppose also that, for fixed Δ, the AR residual variance is approximately constant over the dictionary support, $\omega_\Delta(z, x) = \omega_{\Delta,n}\{1 + \varepsilon_{\omega,n}(z, x)\}$, $\|\varepsilon_{\omega,n}\|_\infty \le \delta_{\omega,n}$. and the first-stage residual variance is approximately proportional to it, $\sigma_V^2(z, x) = \sigma_\Delta^2\, \omega_\Delta(z, x)\{1 + \varepsilon_{V,n}(z, x)\}, \qquad \|\varepsilon_{V,n}\|_\infty \le \delta_{V,n}$. Then $\left\|\omega_{\Delta,n}^{-1} B_{n,\Delta} - I_{p_n}\right\|_{\mathrm{op}} \le \delta_{\omega,n}, \qquad \left\|B^{-1/2}_{n,\Delta} C_n B^{-1/2}_{n,\Delta} - \sigma_\Delta^2 I_{p_n}\right\|_{\mathrm{op}} \lesssim \delta_{V,n} + o(1)$. The corresponding trace conditions follow when $p_n\delta^2_{\omega,n} \to 0$ and $p_n\delta^2_{V,n} \to 0$. If the centered first-stage drift has the dictionary expansion $\pi_n(Z, X) = b_{p_n}(Z, X)'\theta_n + a_n^\pi(Z, X), \qquad \|\,\mathbb{E}[b_{p_n} a_n^\pi]\,\|_2 = o(\|\theta_n\|_2)$, with θ_n s_n-sparse, then $g_n = \theta_n + o(\|\theta_n\|_2)$. Since $B^{-1/2}_{n,\Delta} = \omega^{-1/2}_{\Delta,n}\{I_{p_n} + o_{\mathrm{op}}(1)\}$, the denominator normalization is asymptotically scalar in this example. The exact sparse theorem should still be read in the orthonormalized denominator coordinates; the display explains why raw dictionary sparsity and canonical sparsity coincide exactly in the ideal proportional case and remain close when the perturbations are negligible at the stated risk scale. This example covers orthonormalized instrument transformations whose first-stage noise and AR residual noise are approximately proportional; general correlated raw dictionaries require the denominator-normalized canonical coordinates described above.

5.7. *Rotated canonical-factor example*

The sparse frontier is formulated in denominator-normalized coordinates. This normalization can rotate the reported raw dictionary. Let $a_{p_n}(Z,X)$ be a conditionally centered factor dictionary with $E[a_{p_n}(Z,X)a_{p_n}(Z,X)'] = I_{p_n}$, and let the raw dictionary be $b_{p_n}(Z,X) = L_n a_{p_n}(Z,X)$, where L_n is nonsingular and $L_n L_n'$ is not diagonal in the reported raw coordinates. Suppose, for clarity, that $\omega_\Delta(z,x) = \omega_{\Delta,n}$, $\sigma_V^2(z,x) = \sigma_\Delta^2\omega_{\Delta,n}$. Then $B_{n,\Delta} = \omega_{\Delta,n}L_nL_n'$, $C_n = \sigma_\Delta^2\omega_{\Delta,n}L_nL_n'$, and therefore $B_{n,\Delta}^{-1/2}C_nB_{n,\Delta}^{-1/2} = \sigma_\Delta^2 I_{p_n}$. The denominator-normalized factors are $\psi_{p_n}(Z,X) = B_{n,\Delta}^{-1/2}b_{p_n}(Z,X) = \omega_{\Delta,n}^{-1/2}(L_nL_n')^{-1/2}L_na_{p_n}(Z,X)$, an orthonormal rotation and scaling of the raw factors. Suppose the centered first-stage drift loads sparsely on these normalized factors: $\pi_n(Z,X) = \omega_{\Delta,n}^{1/2}\psi_{p_n}(Z,X)'\theta_n + a_n^\pi(Z,X)$, $\|E[b_{p_n}a_n^\pi]\|_2 = o(\|\theta_n\|_2)$, where θ_n is s_n-sparse. Then $g_n = E[b_{p_n}\pi_n] = (L_nL_n')^{1/2}\theta_n + o(\|\theta_n\|_2)$, and $B_{n,\Delta}^{-1/2}g_n = \omega_{\Delta,n}^{-1/2}\theta_n + o(\|\theta_n\|_2)$. Thus a sparse canonical drift can correspond to a dense raw coefficient vector whenever $(L_nL_n')^{1/2}\theta_n$ is dense. The sparse frontier is a statement about the denominator-standardized score factors that carry identification, rather than sparsity in arbitrary raw instrument coordinates.

5.8. *Operational sparse PLIV frontier*

The sparse class below is a class for the normalized PLIV drift $\widetilde{\gamma}_{n,\Delta} = B_{n,\Delta}^{-1/2}g_n$, not for an arbitrary raw first-stage coefficient vector. Let $\mathfrak{P}_{n,\Delta}^{\mathrm{pliv}}(s_n,r_n)$ denote a class of PLIV triangular arrays satisfying Assumption 5.1, the alignment condition $\left\|B_{n,\Delta}^{-1/2}C_nB_{n,\Delta}^{-1/2} - \sigma_\Delta^2 I_{p_n}\right\|_{\mathrm{op}} = o(1)$, $\quad 0 < \underline{\sigma} \le \sigma_\Delta \le \overline{\sigma} < \infty$, and the sparse normalized drift restriction $\widetilde{\gamma}_{n,\Delta} = B_{n,\Delta}^{-1/2}g_n \in \mathcal{G}_n(s_n,r_n)$, $\quad r_n = \|\widetilde{\gamma}_{n,\Delta}\|_2$. For a normalized-coordinate rule $d_n : \mathbb{R}^{p_n} \to \mathbb{R}^{p_n}$, define $\widetilde{Z}_{n,\Delta} = B_{n,\Delta}^{-1/2}\widehat{g}_n$. Fix a deterministic unit vector $e_{1,n} \in \mathbb{R}^{p_n}$. Implement the original PLIV score coefficient by the normalized representative

$$\widehat{\beta}_n(d_n) = \begin{cases} B_{n,\Delta}^{-1/2}d_n(\widetilde{Z}_{n,\Delta})/\|d_n(\widetilde{Z}_{n,\Delta})\|_2, & d_n(\widetilde{Z}_{n,\Delta}) \neq 0, \\ B_{n,\Delta}^{-1/2}e_{1,n}, & d_n(\widetilde{Z}_{n,\Delta}) = 0. \end{cases} \tag{5.29}$$

Then $\widehat{\beta}_n(d_n)'B_{n,\Delta}\widehat{\beta}_n(d_n) = 1$. Define normalized PLIV noncentrality regret by

$$L_{n,\Delta}^{\mathrm{pliv}}(d_n) = 1 - \frac{Q_{n,\Delta}^{\mathrm{pliv}}\{\widehat{\beta}_n(d_n)\}}{Q_{n,\Delta}^{\star,\mathrm{pliv}}}, \qquad Q_{n,\Delta}^{\star,\mathrm{pliv}} = \rho_{\inf}\Delta^2 r_n^2.$$

DEFINITION 5.11—Balanced canonical PLIV lower-bound subclass: *For fixed $\Delta \neq 0$, $\sigma_\Delta > 0$, and $\widetilde{\gamma} \in \mathcal{G}_n(s_n,r_n)$, the balanced canonical PLIV subclass is the triangular array with known nuisances, $X_i \equiv 0$, $p_n + 1 \le \min(n_{\mathrm{tr}}, n_{\mathrm{inf}})$ eventually, and deterministic split designs satisfying*

$$\frac{1}{n_a}\sum_{i\in I_a} b_i = 0, \qquad \frac{1}{n_a}\sum_{i\in I_a} b_ib_i' = I_{p_n}, \qquad a \in \{\mathrm{tr}, \mathrm{inf}\},$$

first-stage drift $\pi_i = b_i'\widetilde{\gamma}$, first-stage residuals $V_i \overset{ind}{\sim} N(0,\sigma_\Delta^2)$, and inference residuals $R_{i,\Delta} = U_i + \Delta V_i \overset{ind}{\sim} N(0,1)$, independent of V_i conditional on the design. Then $B_{n,\Delta} = I_{p_n}$, $C_n = \sigma_\Delta^2 I_{p_n}$, $g_n = \widetilde{\gamma}$, and $\widetilde{Z}_{n,\Delta} = \widetilde{\gamma} + \sqrt{N/n_{\mathrm{tr}}}\,\sigma_\Delta\xi_n$, $\xi_n \sim N(0,I_{p_n})$. exactly. If exact balance *is inconvenient, an asymptotically balanced version with total-variation equivalence to this Gaussian experiment gives the same lower bound. This subclass is used as a least-favorable local experiment for the lower bound, not as a claim that empirical PLIV designs literally have deterministic balanced instruments and Gaussian residuals.*

THEOREM 5.12—PLIV learned scalar-AR power frontier: *Fix* $\Delta \neq 0$ *and suppose* $2 \leq s_n \leq p_n/4$. *Let*

$$\eta_{n,\Delta} = \frac{\rho_{\mathrm{tr}}^{-1}\sigma_\Delta^2 s_n \log(ep_n/s_n)}{r_n^2}. \tag{5.30}$$

Upper bound. *Suppose, uniformly over* $\mathfrak{P}^{\mathrm{pliv}}_{n,\Delta}(s_n, r_n)$, *the normalized training signal admits the coupled decomposition*

$$\widetilde{Z}_{n,\Delta} = \widetilde{\gamma}_{n,\Delta} + \rho_{\mathrm{tr}}^{-1/2}\sigma_\Delta \xi_n + e_{n,\Delta}, \qquad \xi_n \sim N(0, I_{p_n}), \tag{5.31}$$

with

$$\sup_{P \in \mathfrak{P}^{\mathrm{pliv}}_{n,\Delta}(s_n,r_n)} \frac{\mathbb{E}_P\|e_{n,\Delta}\|_2^2}{r_n^2} = o\{\min(1, \eta_{n,\Delta})\}. \tag{5.32}$$

Then the top-s_n *hard-thresholding rule in normalized coordinates satisfies*

$$\sup_{P \in \mathfrak{P}^{\mathrm{pliv}}_{n,\Delta}(s_n,r_n)} \mathbb{E}_P L^{\mathrm{pliv}}_{n,\Delta}(H_{s_n}) \leq C \min\{1, \eta_{n,\Delta}\} + o(1).$$

The primitive aligned verification theorem verifies (5.31)–(5.32) *for the primitive aligned sparse class only under one of its stated strong-coupling branches.*

Lower bound. *Suppose* $\mathfrak{P}^{\mathrm{pliv}}_{n,\Delta}(s_n, r_n)$ *contains the balanced canonical PLIV subclass in Definition 5.11, with* $\sqrt{N/n_{\mathrm{tr}}}$ *replaced by its nondegenerate limit* $\rho_{\mathrm{tr}}^{-1/2}$ *at an* $o(1)$ *cost. Then there exists a universal* $c > 0$ *such that*

$$\inf_{d_n} \sup_{P \in \mathfrak{P}^{\mathrm{pliv}}_{n,\Delta}(s_n,r_n)} \mathbb{E}_P L^{\mathrm{pliv}}_{n,\Delta}(d_n) \geq c \min\{1, \eta_{n,\Delta}\} - o(1).$$

The same lower bound holds if the balanced subclass is replaced by a PLIV subclass uniformly asymptotically equivalent in total variation to the canonical Gaussian sequence experiment over $\mathcal{G}_n(s_n, r_n)$. *If* $Q^{\star,\mathrm{pliv}}_{n,\Delta} = \rho_{\inf}\Delta^2 r_n^2 \in [\underline{q}, \overline{q}] \subset (0, \infty)$, *then the corresponding upper and lower bounds transfer to PLIV power regret through Corollary 4.4. The proof is in Appendix A.12.*

COROLLARY 5.13—Nondegenerate weak-ID impossibility in operational PLIV: *Fix* $\Delta \neq 0$. *Suppose* $Q^{\star,\mathrm{pliv}}_{n,\Delta} = \rho_{\inf}\Delta^2 r_n^2 \in [\underline{q}, \overline{q}] \subset (0, \infty)$, $\qquad \sigma_\Delta^2 \geq \underline{\sigma}^2 > 0$, *and* $s_n \log(ep_n/s_n) \to \infty$. *Then* $\eta_{n,\Delta} \to \infty$. *Under the lower-bound conditions of Theorem 5.12, learned scalar AR scores cannot uniformly recover oracle PLIV scalar-score power: there exists* $c_\alpha > 0$ *such that* $\inf_{d_n} \sup_{P \in \mathfrak{P}^{\mathrm{pliv}}_{n,\Delta}(s_n,r_n)} \mathcal{R}^{\mathrm{pliv}}_{\alpha,n,\Delta}(d_n) \geq c_\alpha - o(1)$, *where* $\mathcal{R}^{\mathrm{pliv}}_{\alpha,n,\Delta}(d_n) = h_\alpha(Q^{\star,\mathrm{pliv}}_{n,\Delta}) - \mathbb{E}_P h_\alpha\Big(Q^{\mathrm{pliv}}_{n,\Delta}\{\widehat{\beta}_n(d_n)\}\Big)$.

REMARK 5.14—Split fractions in PLIV: *For a single-split PLIV design with training fraction* ρ *and inference fraction* $1-\rho$, *the same tradeoff becomes*

$$q^{\star,\mathrm{pliv}}_{n,\Delta}(\rho) = (1-\rho)\Delta^2\|B^{-1/2}_{n,\Delta} g_n\|^2, \qquad \eta_{n,\Delta}(\rho) = \frac{\rho^{-1}\sigma_\Delta^2 s_n \log(ep_n/s_n)}{\|B^{-1/2}_{n,\Delta} g_n\|^2}.$$

Increasing the training fraction improves learnability of the denominator-normalized PLIV drift but reduces inference-sample noncentrality. The split choice is therefore an operational use of the same learned-score frontier.

REMARK 5.15—PLIV interpretation of the lower bound: *The lower bound is an econometric lower bound over PLIV classes that contain, or are uniformly asymptotically equivalent to, the denominator-normalized canonical sparse training experiment. Every training-measurable PLIV score rule then induces a measurable rule in the canonical Gaussian experiment, and the PLIV normalized noncentrality regret is the canonical projective regret after the transformation* $\widetilde{\gamma}_{n,\Delta} = B_{n,\Delta}^{-1/2} g_n$*. Thus the sparse lower bound identifies the least-favorable power cost of learned scalar AR score construction in aligned growing-dictionary PLIV. General PLIV designs with different denominator-normalized geometries correspond to different canonical drift classes.*

REMARK 5.16—Training drift and local-alternative scaling: *The training sample learns the first-stage drift* g_n*. The local-alternative index* Δ *scales the inference-sample drift and the resulting AR noncentrality, but it does not make the first-stage direction easier to learn. This is why* $\eta_{n,\Delta}$ *is defined using* $\widetilde{\gamma}_{n,\Delta} = B_{n,\Delta}^{-1/2} g_n$*, whereas oracle power contains the factor* Δ^2.

REMARK 5.17—Cross-fitting, orthogonality, and discontinuous score rules: *The primitive conditions above are stated in score-weighted form because these are the remainders that enter the feasible statistic. Standard orthogonalization and cross-fitting are one way to verify* (5.8) *and* (5.12)*. The hard-thresholding upper bound uses the stronger coupling condition* (5.32)*; this is why the primitive aligned theorem imposes trace-covariance and* L_2*-remainder rates rather than only a bounded-Lipsch training approximation.*

REMARK 5.18—Moment-selection mapping: *The same decision problem can arise in growing-moment GMM. If* $m_{p_n}(W,\theta) \in \mathbb{R}^{p_n}$ *is a growing moment vector and, along a weak-GMM local sequence,*

$$\sqrt{n}\,\mathbb{E}[m_{p_n}(W,\theta_0)] = g_n + o(1),$$

$$n^{-1/2} \sum_{i=1}^{n} \{m_{p_n}(W_i,\theta_0) - \mathbb{E} m_{p_n}(W_i,\theta_0)\} \approx \Sigma_n^{1/2} \xi_n.$$

then selecting a linear combination of moments maps into the primitive experiment. This remark records transportability of the learned-score decision problem. The main econometric embedding in this paper is the growing-dictionary PLIV model.

6. SUPPORTING FINITE-GRID DIRECTION-FREE FEASIBLE POWER DESIGN

This section records the finite-grid extension of the scalar-score decision problem. A score chosen for one local direction need not be appropriate when a common score is used over a finite set of tested values or local alternatives; integrated and maximin criteria formalize this common-score design problem.

Pointwise score rules can depend on the local direction through the drift and denominator pair $(G_{n\ell}, B_{n\ell})$. For multi-direction alternatives, a common rule can be chosen by an integrated or maximin criterion. Let $\ell \in \{1,\dots,L\}$ index a finite grid of local directions. Under direction ℓ, the training signal has law $P_{n\ell}$, and the payoff from score β is $h_\alpha\{Q_{n\ell}(\beta; G_{n\ell}, B_{n\ell})\}$. For weights

$w_\ell > 0$ summing to one, define $\mathcal{P}_{w,n}(\delta_n) = \sum_{\ell=1}^{L} w_\ell \mathbb{E}_{n\ell}[h_\alpha\{Q_{n\ell}(\delta_n(Y_n); G_{n\ell}, B_{n\ell})\}]$, where $\mathbb{E}_{n\ell}$ denotes expectation under $P_{n\ell}$. The maximin criterion is

$$\mathcal{P}_{\min,n}(\delta_n) = \min_{1\le\ell\le L} \mathbb{E}_{n\ell}[h_\alpha\{Q_{n\ell}(\delta_n(Y_n); G_{n\ell}, B_{n\ell})\}]\,. \tag{6.1}$$

THEOREM 6.1—Integrated finite-grid score envelope and collapse: *Suppose $L < \infty$, each $B_{n\ell}$ is positive definite, and the laws $P_{n\ell}$ are dominated by a common measure ν_n with densities $f_{n\ell}$. Let $K_n \subset \mathbb{R}^{p_n} \setminus \{0\}$ be a compact normalized action set, and restrict feasible rules to take values in K_n. Then an integrated-power optimal rule exists. For ν_n-almost every y, it can be chosen from* $\arg\max_{\beta\in K_n} \sum_{\ell=1}^{L} w_\ell f_{n\ell}(y) h_\alpha\{Q_{n\ell}(\beta; G_{n\ell}, B_{n\ell})\}$. *If there exist $c_\ell > 0$, $a_\ell \neq 0$, a positive-definite matrix B_{n0}, and a vector G_{n0} such that $B_{n\ell} = c_\ell B_{n0}$, $G_{n\ell} = a_\ell G_{n0}$for all ℓ, then every pointwise power-optimal score direction for (G_{n0}, B_{n0}) within the action set, that is, every element of* $\arg\max_{\beta\in K_n} Q_n(\beta; G_{n0}, B_{n0})$, *is integrated-power optimal. Any such maximizer also solves the maximin criterion* (6.1). *The proof is in Online Appendix 3.1.*

COROLLARY 6.2—Restricted inversion validity: *Consider a sample-split AR confidence set formed by using a common training-measurable direction-free score rule over a finite grid of tested values. If the conditional CLT, variance consistency, and oracle-equivalence conditions hold uniformly over that grid, then the inverted AR confidence set has asymptotic coverage at least $1 - \alpha$ on the grid. The proof is in Online Appendix 3.2.*

7. NUMERICAL LOCAL EXPERIMENTS

The numerical exercises illustrate the same frontier in several forms. The sparse designs vary the difficulty index η_n and display the corresponding loss in frontier attainment and scalar AR power. The PLIV design checks that the denominator-normalized canonical prediction tracks finite-sample learned-score power in an aligned IV design. The Bayes and finite-grid designs report decision-theoretic benchmarks for the same scalar-score choice problem.

7.1. *Posterior power versus posterior Rayleigh*

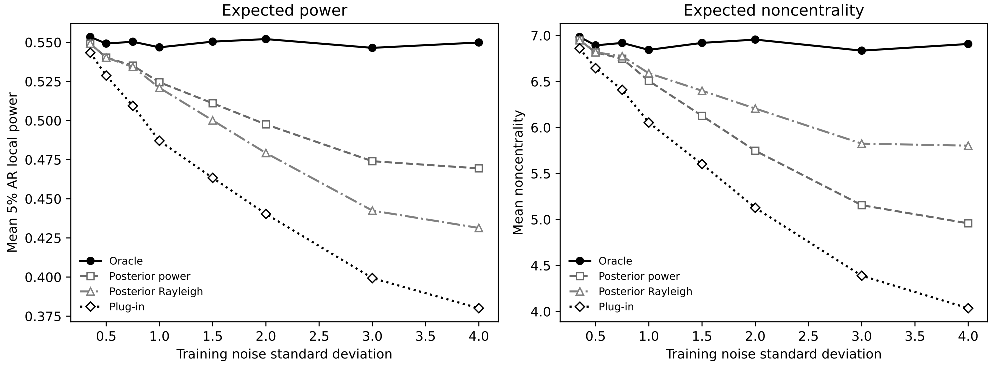

FIGURE 1.—Posterior-power and posterior-Rayleigh rules. The left panel reports mean 5 percent AR local power. The right panel reports mean noncentrality. The posterior-power rule optimizes the left-panel criterion, whereas the posterior-Rayleigh rule optimizes the right-panel criterion. The two rules become visibly distinct when posterior uncertainty over non-collinear drift directions remains.

TABLE I

POSTERIOR POWER AND POSTERIOR RAYLEIGH RULES IN THE LIMITING EXPERIMENT

Training noise	Rule	Mean power	Mean noncentrality	Mean frontier attainment
$\sigma = 1.0$	Oracle	0.547 (0.002)	6.844 (0.045)	1.000 (0.000)
	Posterior power	0.524 (0.002)	6.507 (0.045)	0.920 (0.001)
	Posterior Rayleigh	0.521 (0.002)	6.590 (0.046)	0.910 (0.001)
	Plug-in	0.487 (0.002)	6.053 (0.045)	0.782 (0.002)
$\sigma = 2.0$	Oracle	0.552 (0.002)	6.955 (0.046)	1.000 (0.000)
	Posterior power	0.498 (0.002)	5.747 (0.040)	0.808 (0.001)
	Posterior Rayleigh	0.479 (0.003)	6.206 (0.048)	0.749 (0.002)
	Plug-in	0.440 (0.002)	5.127 (0.041)	0.642 (0.002)
$\sigma = 3.0$	Oracle	0.546 (0.002)	6.835 (0.045)	1.000 (0.000)
	Posterior power	0.474 (0.002)	5.155 (0.037)	0.740 (0.002)
	Posterior Rayleigh	0.442 (0.003)	5.824 (0.048)	0.633 (0.003)
	Plug-in	0.399 (0.002)	4.388 (0.038)	0.579 (0.002)

Notes: Entries are computed from the limiting Gaussian training experiment with three non-collinear posterior support points. The posterior-power rule maximizes posterior expected rejection probability, while the posterior-Rayleigh rule maximizes posterior expected noncentrality. Monte Carlo standard errors are in parentheses; the simulation uses seed 20260508 and 20,000 draws per training-noise value.

The first experiment uses a two-dimensional posterior with three non-collinear drift support points. The denominator is normalized to the identity, and the training signal is

$$Y = G + \sigma\xi, \qquad \xi \sim N(0, I_2).$$

For each realization of Y, three feasible rules are compared: the posterior-power rule that maximizes posterior expected rejection probability, the posterior-Rayleigh rule that maximizes posterior expected noncentrality, and the plug-in rule Y. The oracle rule, which uses the realized drift direction, is infeasible and is shown only as a benchmark. Figure 1 reports mean local power and mean noncentrality as the training-noise standard deviation varies. Figure 1 and Table I illustrate Proposition 3.3. The posterior-Rayleigh rule can deliver larger expected noncentrality while the posterior-power rule delivers larger expected rejection probability. The plug-in rule is natural but treats the noisy training signal as if it were the drift, and it deteriorates as training noise grows.

7.2. *Sparse high-dimensional drift*

Studying a sparse high-dimensional weak drift, this experiment corresponds to the sparse theory in Section 4. Plug-in learning is not the right high-dimensional benchmark: it estimates all coordinates of the weak drift, including many noise coordinates. Thresholding uses the sparse structure of the drift class and therefore preserves more of the feasible power frontier. The canonical denominator and noise-shape matrices are $B_n = I_{p_n}$ and $\Sigma_n = I_{p_n}$. The true drift has $s = 8$ nonzero coordinates, each equal to 1.5, while the remaining coordinates are zero. The training signal is

$$Y = g + \sigma\xi, \qquad \xi \sim N(0, I_{p_n}),$$

and the inference fraction is set to $\rho_{\text{inf}} = 0.25$, so oracle power is away from both size and saturation.

The rules are the infeasible oracle, the plug-in rule Y, a top-s hard-thresholding rule that keeps the eight largest coordinates of Y, and a coordinatewise soft-thresholding rule

$$S_\lambda(Y)_j = \operatorname{sign}(Y_j)(|Y_j| - \lambda)_+, \qquad \lambda = \sigma\sqrt{2\log p_n}.$$

TABLE II

SPARSE HIGH-DIMENSIONAL SCORE LEARNING

Dimension	η_n	Rule	Mean power	Mean frontier attainment	Median frontier attainment
$p = 250$	0.32	Oracle	0.564 (0.000)	1.000 (0.000)	1.000
		Top-s threshold	0.443 (0.001)	0.738 (0.002)	0.776
		Soft threshold	0.313 (0.001)	0.484 (0.002)	0.481
		Plug-in	0.220 (0.000)	0.312 (0.000)	0.312
$p = 1000$	0.41	Oracle	0.564 (0.000)	1.000 (0.000)	1.000
		Top-s threshold	0.379 (0.001)	0.611 (0.002)	0.607
		Soft threshold	0.251 (0.001)	0.370 (0.001)	0.363
		Plug-in	0.104 (0.000)	0.102 (0.000)	0.101
$p = 2000$	0.46	Oracle	0.564 (0.000)	1.000 (0.000)	1.000
		Top-s threshold	0.341 (0.001)	0.539 (0.002)	0.579
		Soft threshold	0.224 (0.001)	0.319 (0.001)	0.311
		Plug-in	0.078 (0.000)	0.054 (0.000)	0.053

Notes: The sparse design has eight nonzero drift coordinates of amplitude 1.5, $\rho_{\mathrm{inf}} = 0.25$, and training-noise standard deviation $\sigma = 0.40$ in the dictionary-dimension design. The table reports simulations of the limiting Gaussian weak-drift experiment. The soft-thresholding rule uses $S_\lambda(Y)_j = \mathrm{sign}(Y_j)(|Y_j| - \lambda)_+$ with $\lambda = \sigma\sqrt{2\log p_n}$. If all coordinates are thresholded to zero, the numerical implementation uses the fixed representative $e_{1,n}$, which has frontier attainment $1/8$ in the displayed support design. Monte Carlo standard errors are in parentheses; the simulations use 8,000 draws per dimension.

For the dictionary-dimension design in Figure 2 and Table II, the training-noise standard deviation is $\sigma = 0.40$. If $S_\lambda(Y) = 0$, the numerical implementation uses the fixed representative $e_{1,n}$. Since the displayed sparse support is $\{1, \ldots, 8\}$, this fallback has normalized frontier attainment $1/8$. The fallback is a reporting convention for zero selected scores in the simulation. The dictionary dimension p_n varies from 50 to 2000, and Table II reports the sparse difficulty index η_n. Figure 3 fixes $p_n = 1000$ and varies the training-noise standard deviation, thereby varying η_n directly while keeping oracle noncentrality nondegenerate. The figure displays the signal-to-noise implication of Theorem 4.1 and Corollary 4.6: as η_n increases, feasible rules lose a larger fraction of the oracle frontier.

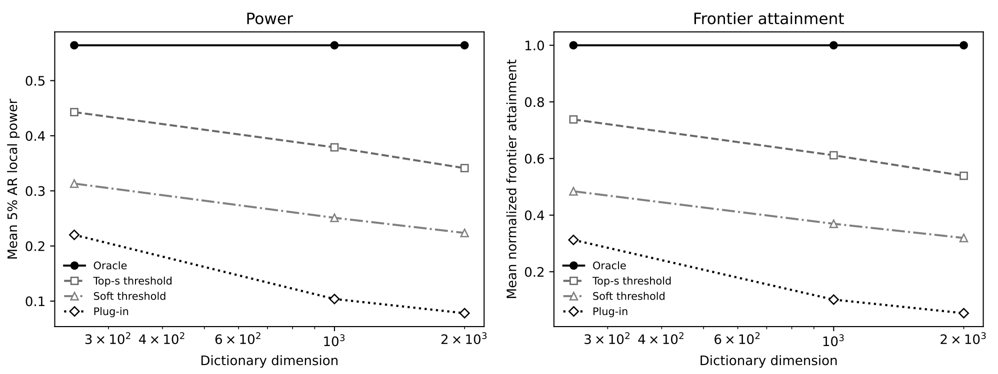

FIGURE 2.—Sparse high-dimensional score learning. The left panel reports mean 5 percent AR local power. The right panel reports normalized frontier attainment $Q_n(\widehat{\beta}; g, I)/Q_n^\star(g)$. As the dictionary dimension grows, the plug-in rule accumulates high-dimensional noise, while structured thresholding rules avoid part of the noise accumulation.

7.3. *Dense drift and shrinkage*

The third experiment uses a dense ellipsoid drift. The drift coordinates are proportional to j^{-1}, scaled to have squared norm 25, and $p_n = 600$. The shrinkage rule uses the ellipsoid prior

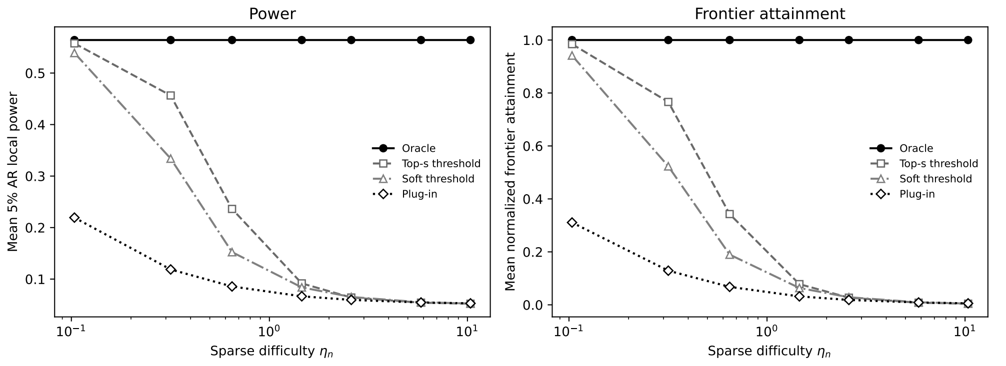

FIGURE 3.—Sparse signal-to-noise frontier. The horizontal axis is the sparse difficulty index $\eta_n = \sigma_n^2 s_n \log(ep_n/s_n)/r_n^2$. The design fixes $p_n = 1000$, $s_n = 8$, and $\rho_{\text{inf}} = 0.25$, and varies the training-noise standard deviation. Small η_n corresponds to recoverable oracle power; large η_n corresponds to finite-information power loss.

variances as coordinate-specific shrinkage weights. Figure 4 compares oracle, plug-in, ellipsoid shrinkage, and thresholding rules as training noise varies. The dense experiment is included to contrast the sparse minimax frontier in Section 4 with a smooth dense regime. It illustrates how the geometry of the weak-drift class changes the appropriate feasible score rule: thresholding is natural for sparse drift, whereas shrinkage is natural for dense ellipsoid drift.

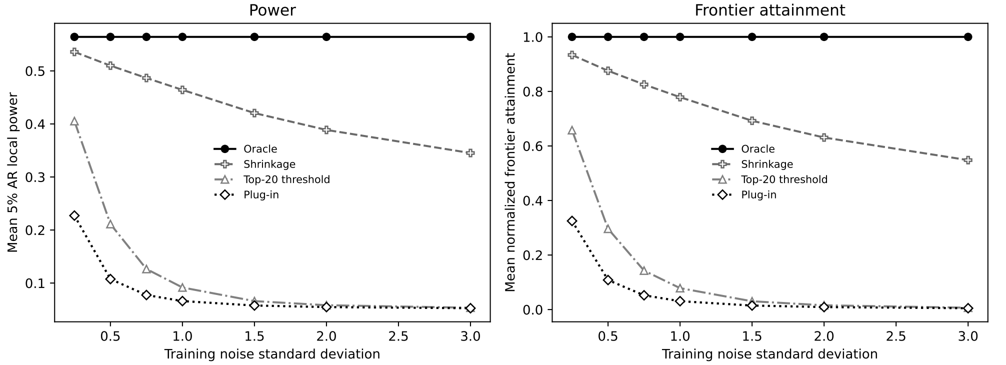

FIGURE 4.—Dense ellipsoid drift. Shrinkage dominates sparse thresholding in the dense ellipsoid design, while the plug-in rule deteriorates rapidly as training noise grows. The experiment illustrates that the optimal feasible score rule depends on the geometry of the weak-drift class.

7.4. *Direction-free score design*

The limiting-experiment check in this subsection illustrates direction-free score design with two local directions. Each direction has its own drift and denominator matrix. For every score direction in a fine grid, the experiment computes the power against direction 1 and direction 2. Figure 5 plots the same grid frontier used to compute Table III; the four highlighted markers are the pointwise, integrated, and maximin choices reported in the table.

TABLE III

DIRECTION-FREE SCORE DESIGN

Rule	Power direction 1	Power direction 2	Integrated power	Minimum power
Pointwise for direction 1	0.583	0.053	0.318	0.053
Pointwise for direction 2	0.197	0.582	0.390	0.197
Integrated	0.249	0.556	0.403	0.249
Maximin	0.357	0.357	0.357	0.357

Notes: Table entries and Figure 5 are computed from the same grid of score directions and the same four selected rules.

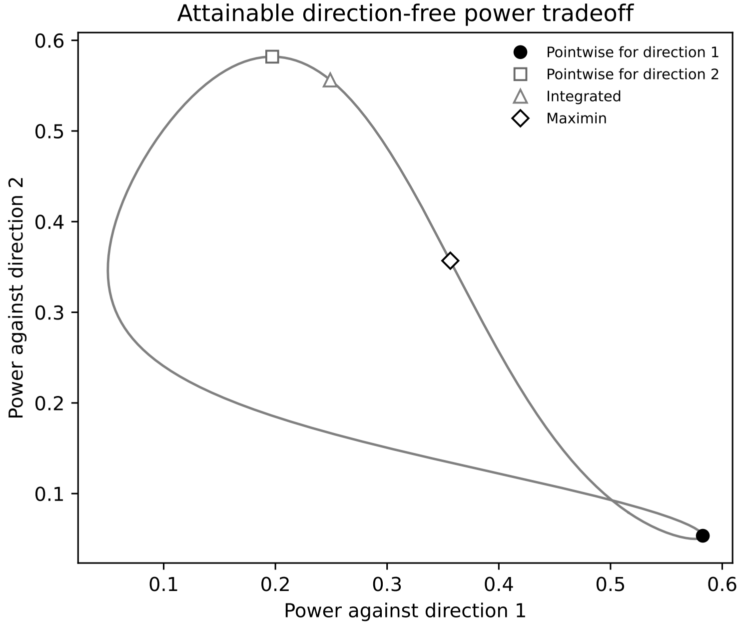

FIGURE 5.—Direction-free power tradeoff. Pointwise optimal scores maximize power against one direction but can perform poorly against the other. Integrated and maximin criteria select interior tradeoffs on the attainable power frontier.

7.5. *Finite-sample PLIV transfer check*

The preceding simulations are limiting Gaussian experiments. This final check simulates the sample-split PLIV design underlying Section 5. The design sets $B_{n,\Delta} = I_{p_n}$, uses a sparse first-stage drift, and varies the first-stage residual standard deviation. The horizontal axis is the operational sparse difficulty index

$$\eta_{n,\Delta} = \frac{\rho_{\mathrm{tr}}^{-1}\sigma_\Delta^2 s_n \log(ep_n/s_n)}{\|B_{n,\Delta}^{-1/2} g_n\|_2^2}.$$

Figure 6 reports a finite-sample PLIV transfer check. Panel A plots normalized frontier attainment, $Q_{n,\Delta}^{\mathrm{pliv}}(\widehat{\beta})/Q_{n,\Delta}^{\star,\mathrm{pliv}}$, for the reported score rules. Panel B plots the maximum and mean absolute differences between the finite-sample PLIV quantities and their canonical predictions over the same rules. The figure illustrates the denominator-normalized reduction used in Theorem 5.2: as the operational sparse difficulty $\eta_{n,\Delta}$ grows, feasible learned scores lose frontier attainment, while the canonical approximation error remains small.

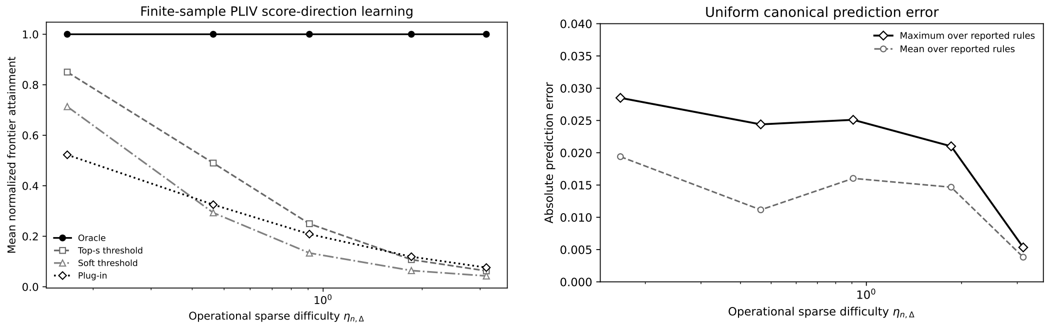

FIGURE 6.—Finite-sample PLIV frontier attainment and canonical approximation. Panel A reports mean normalized frontier attainment, $Q^{\text{pliv}}_{n,\Delta}(\widehat{\beta})/Q^{\star,\text{pliv}}_{n,\Delta}$, for the reported score rules. Panel B reports the maximum and mean absolute canonical prediction error over the same rules. The design has $N = 1000$, $p = 120$, $s = 8$, $\rho_{\text{tr}} = \rho_{\text{inf}} = 1/2$, and oracle scalar-AR noncentrality normalized to four.

8. CONCLUSION

This paper characterizes one finite-information power frontier for learned scalar Anderson–Rubin scores under weak identification. Sample splitting preserves transparent weak-ID size control conditional on the selected score, but it does not make the selected score as powerful as the infeasible oracle direction. In the canonical Gaussian weak-drift experiment, the exact sparse minimax frontier is governed by

$$\min\left\{1, \frac{\sigma_n^2 s_n \log(ep_n/s_n)}{r_n^2}\right\}.$$

On compact oracle-noncentrality ranges, this projective direction-learning frontier is also a local power-regret frontier. Hence nonvanishing training noise and growing sparse complexity rule out uniform recovery of oracle scalar-AR power under nondegenerate weak-ID oracle power.

The Bayes and finite-grid results clarify the decision problem faced by a learned score rule: posterior expected-power maximization need not coincide with posterior Rayleigh maximization, and a common score across several local directions solves an integrated or maximin power problem. The PLIV embedding gives the central econometric realization. In growing-dictionary PLIV, feasible scalar-AR power is governed by the denominator-normalized first-stage drift

$$B_{n,\Delta}^{-1/2} g_n,$$

not by the raw first-stage coefficient vector. Thus, within learned scalar-AR designs, feasible weak-ID power depends not only on the robust statistic evaluated at the end, but also on the finite information available for constructing the score that enters it.

REFERENCES

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20, 46–63. [1, 3]

Andrews, I. (2016). Conditional linear combination tests for weakly identified models. *Econometrica* 84, 2155–2182. [1, 3]

Andrews, D. W. K. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* 80, 2153–2211. [1, 3]

Andrews, I. and A. Mikusheva (2016). Conditional inference with a functional nuisance parameter. *Econometrica* 84, 1571–1612. [1, 3]

Andrews, I. and A. Mikusheva (2022). Optimal decision rules for weak GMM. *Econometrica* 90, 715–748. [1, 3]

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74, 715–752. [1, 3]

Ayyar, S., Y. Matsushita, and T. Otsu (2025). Conditional likelihood ratio test with many weak instruments. *Econometric Theory*, published online. [3]

Baiardi, A., P. S. Clarke, A. A. Naghi, and A. Polselli (2026). Double machine learning for static panel data with instrumental variables: New method and applications. arXiv:2603.20464. [3]

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62, 657–681. [3]

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80, 2369–2429. [3]

Boot, T. and D. Nibbering (2024). Inference on LATEs with covariates. arXiv:2402.12607. [3]

Carrasco, M. and G. Tchuente (2016). Regularization based Anderson–Rubin tests for many instruments. University of Kent School of Economics Discussion Paper 1608. [3]

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34, 305–334. [3]

Chen, J., D. L. Chen, and G. Lewis (2020). Mostly harmless machine learning: Learning optimal instruments in linear IV models. arXiv:2011.06158. [3]

Chao, J. C. and N. R. Swanson (2005). Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1692. [3]

Chernozhukov, V., D. Chetverikov, and K. Kato (2017). Central limit theorems and bootstrap in high dimensions. *Annals of Probability* 45, 2309–2352. [16]

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal* 21, C1–C68. [3]

Chernozhukov, V., W. Newey, and J. Robins (2022). Debiased machine learning of global and local parameters using regularized Riesz representers. *The Econometrics Journal* 25, 576–601. [3]

Chernozhukov, V., W. Newey, and R. Singh (2022). Automatic debiased machine learning of causal and structural effects. *Econometrica* 90, 967–1027. [3]

Donoho, D. L. and I. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, 425–455. [9]

Donald, S. G. and W. K. Newey (2001). Choosing the number of instruments. *Econometrica* 69, 1161–1191. [3]

Hansen, C., J. Hausman, and W. K. Newey (2008). Estimation with many instrumental variables. *Journal of Business & Economic Statistics* 26, 398–422. [3]

Kaji, T. (2021). Theory of weak identification in semiparametric models. *Econometrica* 89, 733–763. [1, 3]

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70, 1781–1803. [1, 3]

Lim, D., W. Wang, and Y. Zhang (2024). A conditional linear combination test with many weak instruments. *Journal of Econometrics* 238(2), 105602. [3]

Ma, Y. (2023). Identification-robust inference for the LATE with high-dimensional covariates. arXiv:2302.09756. [3]

Mikusheva, A. and L. Sun (2024). Weak identification with many instruments. *The Econometrics Journal* 27, C1–C28. [3]

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048. [1, 3]

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58, 809–837. [3]

Newey, W. K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica* 77, 687–719. [3]

Raskutti, G., M. J. Wainwright, and B. Yu (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Transactions on Information Theory* 57, 6976–6994. [9]

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586. [1, 3]

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica* 68, 1055–1096. [1, 3]

van der Laan, L., N. Kallus, and A. Bibaut (2025). Nonparametric instrumental variable inference with many weak instruments. arXiv:2505.07729. [3]

Wang, R., K. C. G. Chan, and T. Ye (2025). GMM with many weak moment conditions and nuisance parameters: General theory and applications to causal inference. arXiv:2505.07295. [3]

Yurinskii, V. V. (1977). On the error of the Gaussian approximation for convolutions. *Theory of Probability and Its Applications* 22, 236–247. [16]

Zaitsev, A. Yu. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality conditions. *Probability Theory and Related Fields* 74, 535–566. [16]

Zhang, D., X. Li, and B. Sun (2026). Multi-task learning for high-dimensional regression with many weak instruments. arXiv:2504.18107. [3]

APPENDIX A: PROOFS

This appendix contains the main conceptual and econometric bridge proofs. The full projective-packing, Fano, perturbation-packing, thresholding, direction-free, and primitive PLIV verification derivations are collected in the Online Appendix.

A.1. *Proof of Theorem 2.3*

PROOF: Let

$$\mathcal{F}_n := \sigma(Y_n), \qquad \widehat{\beta}_n = \delta_n(Y_n).$$

Then $\widehat{\beta}_n$ is $\mathcal{F}_n$-measurable. All conditional laws and conditional probabilities below are regular conditional objects given $\mathcal{F}_n$. Write

$$c := c_{1-\alpha}, \qquad A_c := \{x \in \mathbb{R} : x^2 > c\}.$$

For probability measures P, Q on $\mathbb{R}$, use the bounded-Lipschitz metric

$$d_{\mathrm{BL}}(P, Q) := \sup_{\|f\|_{\mathrm{BL}} \le 1} \left| \int f\, dP - \int f\, dQ \right|, \qquad \|f\|_{\mathrm{BL}} := \|f\|_\infty + \sup_{x \ne y} \frac{|f(x) - f(y)|}{|x - y|}.$$

First prove a conditional tail-transfer step. Let X_n be a scalar statistic and let μ_n be $\mathcal{F}_n$-measurable. Suppose

$$d_{\mathrm{BL}}\{\mathcal{L}(X_n \mid \mathcal{F}_n), N(\mu_n, 1)\} \xrightarrow{p} 0.$$

Let $K_n(\cdot) := N(\mu_n, 1)(\cdot)$. For $\varepsilon > 0$, define

$$A_c^{-\varepsilon} := \{x \in A_c : \mathrm{dist}(x, A_c^c) > \varepsilon\}, \qquad A_c^{\varepsilon} := \{x : \mathrm{dist}(x, A_c) \le \varepsilon\}.$$

Choose Lipschitz functions $\varphi_\varepsilon^-, \varphi_\varepsilon^+ : \mathbb{R} \to [0, 1]$ satisfying

$$1\{x \in A_c^{-\varepsilon}\} \le \varphi_\varepsilon^-(x) \le 1\{x \in A_c\} \le \varphi_\varepsilon^+(x) \le 1\{x \in A_c^{\varepsilon}\},$$

and

$$\|\varphi_\varepsilon^-\|_{\mathrm{BL}} \le 1 + \varepsilon^{-1}, \qquad \|\varphi_\varepsilon^+\|_{\mathrm{BL}} \le 1 + \varepsilon^{-1}.$$

Then

$$\begin{aligned}
\Pr\{X_n^2 > c \mid \mathcal{F}_n\} &\le \int \varphi_\varepsilon^+ \, d\mathcal{L}(X_n \mid \mathcal{F}_n) \\
&\le \int \varphi_\varepsilon^+ \, dK_n + (1 + \varepsilon^{-1}) d_{\mathrm{BL}}\{\mathcal{L}(X_n \mid \mathcal{F}_n), K_n\} \\
&\le K_n(A_c^{\varepsilon}) + (1 + \varepsilon^{-1}) d_{\mathrm{BL}}\{\mathcal{L}(X_n \mid \mathcal{F}_n), K_n\},
\end{aligned}$$

and similarly

$$\begin{aligned}
\Pr\{X_n^2 > c \mid \mathcal{F}_n\} &\ge \int \varphi_\varepsilon^- \, d\mathcal{L}(X_n \mid \mathcal{F}_n) \\
&\ge \int \varphi_\varepsilon^- \, dK_n - (1 + \varepsilon^{-1}) d_{\mathrm{BL}}\{\mathcal{L}(X_n \mid \mathcal{F}_n), K_n\} \\
&\ge K_n(A_c^{-\varepsilon}) - (1 + \varepsilon^{-1}) d_{\mathrm{BL}}\{\mathcal{L}(X_n \mid \mathcal{F}_n), K_n\}.
\end{aligned}$$

Hence

$$\left|\Pr\{X_n^2 > c \mid \mathcal{F}_n\} - K_n(A_c)\right| \le (1+\varepsilon^{-1}) d_{\mathrm{BL}}\{\mathcal{L}(X_n \mid \mathcal{F}_n), K_n\} + K_n(A_c^{\varepsilon} \setminus A_c^{-\varepsilon}). \tag{A.1}$$

Now

$$\partial A_c = \{-\sqrt{c}, \sqrt{c}\}.$$

Moreover,

$$A_c^{\varepsilon} \setminus A_c^{-\varepsilon} \subset [-\sqrt{c}-\varepsilon, -\sqrt{c}+\varepsilon] \cup [\sqrt{c}-\varepsilon, \sqrt{c}+\varepsilon].$$

For every $\mu \in \mathbb{R}$, the $N(\mu,1)$ density is bounded by $(2\pi)^{-1/2}$. Therefore

$$\sup_{\mu\in\mathbb{R}} N(\mu,1)(A_c^{\varepsilon} \setminus A_c^{-\varepsilon}) \le \frac{4\varepsilon}{\sqrt{2\pi}}. \tag{A.2}$$

Combining (A.1) and (A.2), for every fixed $\varepsilon > 0$,

$$\left|\Pr\{X_n^2 > c \mid \mathcal{F}_n\} - \Pr\{(Z+\mu_n)^2 > c \mid \mathcal{F}_n\}\right| \le (1+\varepsilon^{-1}) o_p(1) + \frac{4\varepsilon}{\sqrt{2\pi}},$$

where $Z \sim N(0,1)$ is independent of $\mathcal{F}_n$. Thus, for every $\eta > 0$, choose $\varepsilon > 0$ with $4\varepsilon/\sqrt{2\pi} < \eta/2$. Then

$$\Pr\Big(\left|\Pr\{X_n^2 > c \mid \mathcal{F}_n\} - \Pr\{(Z+\mu_n)^2 > c \mid \mathcal{F}_n\}\right| > \eta\Big) \to 0.$$

Consequently,

$$\left|\Pr\{X_n^2 > c \mid \mathcal{F}_n\} - \Pr\{(Z+\mu_n)^2 > c \mid \mathcal{F}_n\}\right| \xrightarrow{p} 0. \tag{A.3}$$

No boundedness condition on μ_n is used; the bound (A.2) is uniform in μ.

Under the null, the assumed conditional approximation is

$$d_{\mathrm{BL}}\big\{\mathcal{L}_0(T_{n,\widehat{\beta}_n} \mid \mathcal{F}_n), N(0,1)\big\} \xrightarrow{p} 0.$$

Taking $X_n = T_{n,\widehat{\beta}_n}$ and $\mu_n = 0$ in (A.3) gives

$$\begin{aligned}\Pr_0\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha} \mid \mathcal{F}_n\} &= \Pr_0\{T^2_{n,\widehat{\beta}_n} > c_{1-\alpha} \mid \mathcal{F}_n\} \\ &= \Pr\{Z^2 > c_{1-\alpha}\} + o_p(1) \\ &= \alpha + o_p(1).\end{aligned}$$

Let

$$R_{n,0} := \Pr_0\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha} \mid \mathcal{F}_n\} - \alpha.$$

Then $|R_{n,0}| \le 1$ and $R_{n,0} \xrightarrow{p} 0$. Hence

$$\mathbb{E}_0|R_{n,0}| \le \eta + \Pr_0(|R_{n,0}| > \eta) \to \eta \qquad \forall \eta > 0,$$

and therefore $\mathbb{E}_0|R_{n,0}| \to 0$. Iterated expectations yield

$$\Pr_0\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha}\} = \mathbb{E}_0\Big[\Pr_0\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha} \mid \mathcal{F}_n\}\Big] = \alpha + o(1).$$

Now fix a local drift g. Define the $\mathcal{F}_n$-measurable random noncentrality

$$\Lambda_n(g) := Q_n\{\widehat{\beta}_n; g, B_n\} = Q_n\{\delta_n(Y_n); g, B_n\}.$$

Let $\mu_n(g)$ denote the signed conditional shift in the theorem statement, so that

$$\mu_n(g)^2 = \Lambda_n(g).$$

The assumed conditional approximation is

$$d_{\mathrm{BL}}\big\{\mathcal{L}_g(T_{n,\widehat{\beta}_n} \mid \mathcal{F}_n), N\{\mu_n(g), 1\}\big\} \to_{p,g} 0.$$

Taking $X_n = T_{n,\widehat{\beta}_n}$ and $\mu_n = \mu_n(g)$ in (A.3) gives

$$\Pr_g\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha} \mid \mathcal{F}_n\} = \Pr_g\{T^2_{n,\widehat{\beta}_n} > c_{1-\alpha} \mid \mathcal{F}_n\} = \Pr\{(Z + \mu_n(g))^2 > c_{1-\alpha} \mid \mathcal{F}_n\} + o_{p,g}(1).$$

Since

$$(Z + \mu_n(g))^2 \mid \mathcal{F}_n \sim \chi^2_1\{\mu_n(g)^2\} = \chi^2_1\{\Lambda_n(g)\},$$

$$\Pr\{(Z + \mu_n(g))^2 > c_{1-\alpha} \mid \mathcal{F}_n\} = 1 - F_{\chi^2_1\{\Lambda_n(g)\}}(c_{1-\alpha}) = h_\alpha\{\Lambda_n(g)\} = h_\alpha(Q_n\{\delta_n(Y_n); g, B_n\})\,.$$

Thus

$$\Pr_g\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha} \mid \mathcal{F}_n\} = h_\alpha(Q_n\{\delta_n(Y_n); g, B_n\}) + r_{n,g}, \tag{A.4}$$

where $r_{n,g} = o_{p,g}(1)$. Since both terms in (A.4) are probabilities, $|r_{n,g}| \le 1$ after modifying $r_{n,g}$ on a null set. Hence, for every $\eta > 0$,

$$\mathbb{E}_g|r_{n,g}| \le \eta + \Pr_g(|r_{n,g}| > \eta),$$

so $\mathbb{E}_g|r_{n,g}| \to 0$. Taking expectations in (A.4) yields

$$\Pr_g\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha}\} = \mathbb{E}_g\left[\Pr_g\{AR_{n,\widehat{\beta}_n} > c_{1-\alpha} \mid \mathcal{F}_n\}\right] = \mathbb{E}_g[h_\alpha(Q_n\{\delta_n(Y_n); g, B_n\})] + o(1).$$

This proves the asserted local-power representation.

Finally, by Definition 2.2,

$$h_\alpha\{Q_n^\star(g)\}$$

is the oracle local rejection probability and

$$\mathbb{E}_g[h_\alpha(Q_n\{\delta_n(Y_n); g, B_n\})]$$

is the local rejection probability generated by the learned-score scalar AR experiment. Their difference is the power regret in (2.9). The theorem follows. *Q.E.D.*

A.2. *Proof of Theorem 3.2*

PROOF: Fix n. Let $\mu_{Y,n}$ denote the marginal law of Y_n. By Assumption 3.1, there exists a regular conditional distribution

$$\Pi_n(\cdot \mid y) = \mathcal{L}(G_n \mid Y_n = y).$$

Fix one version of $\Pi_n(\cdot \mid y)$, modified arbitrarily on a $\mu_{Y,n}$-null set. All statements involving y below hold on this version.

The normalized action set is

$$\mathcal{B}_n = \{\beta \in \mathbb{R}^{p_n} : \beta' B_n \beta = 1\}. \tag{A.5}$$

Since B_n is symmetric positive definite,

$$0 < \lambda_{\min}(B_n) \leq \lambda_{\max}(B_n) < \infty.$$

For every $\beta \in \mathcal{B}_n$,

$$1 = \beta' B_n \beta \geq \lambda_{\min}(B_n)\|\beta\|^2, \qquad \|\beta\| \leq \lambda_{\min}(B_n)^{-1/2}.$$

Thus $\mathcal{B}_n$ is bounded. Since $\beta \mapsto \beta' B_n \beta$ is continuous, $\mathcal{B}_n$ is closed. Hence $\mathcal{B}_n$ is compact. It is nonempty because, for any $v \neq 0$,

$$\frac{v}{(v' B_n v)^{1/2}} \in \mathcal{B}_n.$$

Define

$$f_n(\beta, g) := h_\alpha\{Q_n(\beta; g, B_n)\}. \tag{A.6}$$

On $\mathcal{B}_n$,

$$Q_n(\beta; g, B_n) = \rho_{\inf}(\beta' g)^2.$$

Hence $(\beta, g) \mapsto Q_n(\beta; g, B_n)$ is continuous on $\mathcal{B}_n \times \mathbb{R}^{p_n}$.

The map h_α is continuous. Indeed, if $Z \sim N(0,1)$, then, for $q \geq 0$,

$$h_\alpha(q) = 1 - F_{\chi^2_1(q)}(c_{1-\alpha}) = \Pr\{(Z + \sqrt{q})^2 > c_{1-\alpha}\}.$$

If $q_m \to q$, then

$$(Z + \sqrt{q_m})^2 \to (Z + \sqrt{q})^2 \qquad \text{a.s.}$$

Moreover,

$$\Pr\{(Z + \sqrt{q})^2 = c_{1-\alpha}\} = 0$$

because Z has a continuous density. Therefore

$$1\{(Z+\sqrt{q_m})^2 > c_{1-\alpha}\} \to 1\{(Z+\sqrt{q})^2 > c_{1-\alpha}\} \qquad \text{a.s.}$$

Dominated convergence gives $h_\alpha(q_m) \to h_\alpha(q)$. Also

$$0 \le h_\alpha(q) \le 1 \qquad \forall q \ge 0.$$

Consequently f_n is bounded, jointly Borel measurable on $\mathcal{B}_n \times \mathbb{R}^{p_n}$, and continuous in β for every fixed g.

For each y, define

$$\mathcal{P}_{\alpha,n}(\beta; y) := \int_{\mathbb{R}^{p_n}} f_n(\beta, g)\, \Pi_n(dg \mid y). \tag{A.7}$$

For every fixed β, the kernel property of $y \mapsto \Pi_n(\cdot \mid y)$ gives measurability of

$$y \mapsto \mathcal{P}_{\alpha,n}(\beta; y).$$

For every fixed y, if $\beta_m \to \beta$ in $\mathcal{B}_n$, then

$$f_n(\beta_m, g) \to f_n(\beta, g) \quad \forall g, \qquad |f_n(\beta_m, g)| \le 1.$$

Thus

$$\mathcal{P}_{\alpha,n}(\beta_m; y) = \int f_n(\beta_m, g)\, \Pi_n(dg \mid y) \to \int f_n(\beta, g)\, \Pi_n(dg \mid y) = \mathcal{P}_{\alpha,n}(\beta; y).$$

Hence $\beta \mapsto \mathcal{P}_{\alpha,n}(\beta; y)$ is continuous on $\mathcal{B}_n$. Since $\mathcal{P}_{\alpha,n}$ is measurable in y for each β, continuous in β for each y, and $\mathcal{B}_n$ is a compact metric space, $\mathcal{P}_{\alpha,n}$ is jointly Borel measurable.

Define the posterior value

$$\mathcal{V}_{\alpha,n}(y) := \sup_{\beta \in \mathcal{B}_n} \mathcal{P}_{\alpha,n}(\beta; y). \tag{A.8}$$

For each y, compactness of $\mathcal{B}_n$ and continuity of $\mathcal{P}_{\alpha,n}(\cdot; y)$ imply

$$\mathcal{V}_{\alpha,n}(y) = \max_{\beta \in \mathcal{B}_n} \mathcal{P}_{\alpha,n}(\beta; y).$$

Let D_n be a countable dense subset of $\mathcal{B}_n$. Since $\mathcal{P}_{\alpha,n}(\cdot; y)$ is continuous,

$$\mathcal{V}_{\alpha,n}(y) = \sup_{\beta \in D_n} \mathcal{P}_{\alpha,n}(\beta; y). \tag{A.9}$$

The right-hand side in (A.9) is a countable supremum of measurable functions. Hence $y \mapsto \mathcal{V}_{\alpha,n}(y)$ is measurable.

Let

$$A_n(y) := \arg\max_{\beta \in \mathcal{B}_n} \mathcal{P}_{\alpha,n}(\beta; y).$$

For every y, $A_n(y)$ is nonempty and compact. Its graph is

$$\mathrm{Gr}(A_n) = \{(y, \beta) : \beta \in \mathcal{B}_n,\ \mathcal{P}_{\alpha,n}(\beta; y) = \mathcal{V}_{\alpha,n}(y)\}. \tag{A.10}$$

Since $\mathcal{B}_n$ is compact metric and $(y,\beta) \mapsto \mathcal{P}_{\alpha,n}(\beta;y)$ is a CarathÃ©odory objective, the measurable maximum theorem implies that $y \mapsto \mathcal{V}_{\alpha,n}(y)$ is measurable and that the argmax correspondence

$$A_n(y) = \arg\max_{\beta \in \mathcal{B}_n} \mathcal{P}_{\alpha,n}(\beta;y)$$

is nonempty compact-valued and weakly measurable. The Kuratowski–Ryll-Nardzewski measurable selection theorem therefore gives a measurable selector

$$\delta_n^P(y) \in A_n(y) \qquad \mu_{Y,n}\text{-a.s.}$$

After arbitrary modification on the remaining null set,

$$\delta_n^P(y) \in \arg\max_{\beta \in \mathcal{B}_n} \mathcal{P}_{\alpha,n}(\beta;y) \qquad \mu_{Y,n}\text{-a.s.} \tag{A.11}$$

Let δ_n be any measurable feasible rule with values in $\mathbb{R}^{p_n} \setminus \{0\}$. Define

$$\widetilde{\delta}_n(y) := \frac{\delta_n(y)}{\{\delta_n(y)' B_n \delta_n(y)\}^{1/2}}.$$

Since B_n is positive definite, $\widetilde{\delta}_n(y)$ is well-defined and $\widetilde{\delta}_n(y) \in \mathcal{B}_n$. For every y and g,

$$Q_n\{\delta_n(y); g, B_n\} = Q_n\{\widetilde{\delta}_n(y); g, B_n\}. \tag{A.12}$$

Thus it suffices to consider rules taking values in $\mathcal{B}_n$.

For any measurable $\delta_n : Y_n \mapsto \mathcal{B}_n$,

$$\begin{aligned}
\mathbb{E}[h_\alpha\{Q_n(\delta_n(Y_n); G_n, B_n)\}] &= \mathbb{E}[\mathbb{E}[h_\alpha\{Q_n(\delta_n(Y_n); G_n, B_n)\} \mid Y_n]] \\
&= \mathbb{E}\left[\int h_\alpha\{Q_n(\delta_n(Y_n); g, B_n)\}\Pi_n(dg \mid Y_n)\right] \\
&= \mathbb{E}[\mathcal{P}_{\alpha,n}\{\delta_n(Y_n); Y_n\}].
\end{aligned}$$

For every y,

$$\mathcal{P}_{\alpha,n}\{\delta_n(y); y\} \le \sup_{\beta \in \mathcal{B}_n} \mathcal{P}_{\alpha,n}(\beta; y) = \mathcal{V}_{\alpha,n}(y).$$

Therefore

$$\mathbb{E}[h_\alpha\{Q_n(\delta_n(Y_n); G_n, B_n)\}] \le \mathbb{E}[\mathcal{V}_{\alpha,n}(Y_n)]. \tag{A.13}$$

By (A.11),

$$\mathcal{P}_{\alpha,n}\{\delta_n^P(y); y\} = \mathcal{V}_{\alpha,n}(y) \qquad \mu_{Y,n}\text{-a.s.}$$

Hence

$$\begin{aligned}
\mathbb{E}\big[h_\alpha\{Q_n(\delta_n^P(Y_n); G_n, B_n)\}\big] &= \mathbb{E}\big[\mathcal{P}_{\alpha,n}\{\delta_n^P(Y_n); Y_n\}\big] \\
&= \mathbb{E}[\mathcal{V}_{\alpha,n}(Y_n)].
\end{aligned}$$

Combining this equality with (A.13) proves that δ_n^P attains the ex ante Bayes feasible power value and that no measurable feasible rule can exceed it. *Q.E.D.*

A.3. *Proof of Proposition 3.3*

PROOF: Fix n and fix a realization $Y_n = y$ for which the stated posterior distribution holds. All posterior expectations below are conditional on $Y_n = y$. Thus

$$\Pr(G = ae_1 \mid Y_n = y) = p, \qquad \Pr(G = be_2 \mid Y_n = y) = 1 - p, \tag{A.14}$$

where $a, b > 0$ and $p \in (0,1)$.

Since $B_n = I_2$, the normalized action set is

$$\mathbb{S}^1 = \{\beta = (\beta_1, \beta_2)' \in \mathbb{R}^2 : \|\beta\| = 1\}. \tag{A.15}$$

For any $\beta \in \mathbb{S}^1$, define

$$t_\beta := \beta_1^2 \in [0,1], \qquad \beta_2^2 = 1 - t_\beta. \tag{A.16}$$

For $g = ae_1$,

$$\begin{aligned} Q_n(\beta; ae_1, I_2) &= \rho_{\inf} \frac{(\beta' ae_1)^2}{\beta'\beta} \\ &= \rho_{\inf} \frac{a^2\beta_1^2}{\beta_1^2 + \beta_2^2} \\ &= \rho_{\inf} a^2 t_\beta. \end{aligned} \tag{A.17}$$

For $g = be_2$,

$$\begin{aligned} Q_n(\beta; be_2, I_2) &= \rho_{\inf} \frac{(\beta' be_2)^2}{\beta'\beta} \\ &= \rho_{\inf} \frac{b^2\beta_2^2}{\beta_1^2 + \beta_2^2} \\ &= \rho_{\inf} b^2 (1 - t_\beta). \end{aligned} \tag{A.18}$$

For each $t \in [0,1]$, define

$$\beta(t) = (\sqrt{t}, \sqrt{1-t})'. \tag{A.19}$$

Then $\beta(t) \in \mathbb{S}^1$, and for every $\beta \in \mathbb{S}^1$,

$$Q_n(\beta; ae_1, I_2) = Q_n\{\beta(t_\beta); ae_1, I_2\}, \tag{A.20}$$

and

$$Q_n(\beta; be_2, I_2) = Q_n\{\beta(t_\beta); be_2, I_2\}. \tag{A.21}$$

Therefore

$$\begin{aligned} &\mathbb{E}[h_\alpha\{Q_n(\beta; G, I_2)\} \mid Y_n = y] \\ &\quad = p\, h_\alpha\{Q_n(\beta; ae_1, I_2)\} + (1-p) h_\alpha\{Q_n(\beta; be_2, I_2)\} \\ &\quad = p\, h_\alpha(\rho_{\inf} a^2 t_\beta) + (1-p) h_\alpha\{\rho_{\inf} b^2 (1 - t_\beta)\}. \end{aligned} \tag{A.22}$$

It follows that

$$\sup_{\beta\in\mathbb{S}^1} \mathbb{E}[h_\alpha\{Q_n(\beta;G,I_2)\}\,|\,Y_n=y] = \sup_{t\in[0,1]} \left[p\,h_\alpha(\rho_{\inf}a^2t)+(1-p)h_\alpha\{\rho_{\inf}b^2(1-t)\}\right]. \tag{A.23}$$

Define

$$\Phi_\alpha(t) = p\,h_\alpha(\rho_{\inf}a^2t)+(1-p)h_\alpha\{\rho_{\inf}b^2(1-t)\}. \tag{A.24}$$

Equations (A.23) and (A.24) prove (3.7). Since h_α is continuous and $[0,1]$ is compact,

$$\arg\max_{t\in[0,1]} \Phi_\alpha(t) \neq \emptyset. \tag{A.25}$$

The posterior-Rayleigh objective is posterior expected noncentrality. By (A.17)–(A.18),

$$\begin{aligned}\mathbb{E}[Q_n(\beta(t);G,I_2)\,|\,Y_n=y] &= p\,Q_n\{\beta(t);ae_1,I_2\}+(1-p)Q_n\{\beta(t);be_2,I_2\}\\ &= p\,\rho_{\inf}a^2t+(1-p)\rho_{\inf}b^2(1-t)\\ &= \rho_{\inf}\left[pa^2t+(1-p)b^2(1-t)\right].\end{aligned} \tag{A.26}$$

Since $\rho_{\inf}>0$,

$$\arg\max_{t\in[0,1]} \mathbb{E}[Q_n(\beta(t);G,I_2)\,|\,Y_n=y] = \arg\max_{t\in[0,1]} \Psi(t), \tag{A.27}$$

where

$$\Psi(t) := pa^2t+(1-p)b^2(1-t). \tag{A.28}$$

For $t,u\in[0,1]$,

$$\begin{aligned}\Psi(t)-\Psi(u) &= \{pa^2t+(1-p)b^2(1-t)\}-\{pa^2u+(1-p)b^2(1-u)\}\\ &= \{pa^2-(1-p)b^2\}(t-u).\end{aligned} \tag{A.29}$$

Hence, if $pa^2>(1-p)b^2$, then for all $t\in[0,1)$,

$$\Psi(t)-\Psi(1) = \{pa^2-(1-p)b^2\}(t-1)<0, \tag{A.30}$$

so

$$\arg\max_{t\in[0,1]} \Psi(t) = \{1\}. \tag{A.31}$$

If $pa^2<(1-p)b^2$, then for all $t\in(0,1]$,

$$\Psi(t)-\Psi(0) = \{pa^2-(1-p)b^2\}t<0, \tag{A.32}$$

so

$$\arg\max_{t\in[0,1]} \Psi(t) = \{0\}. \tag{A.33}$$

Finally, if $pa^2=(1-p)b^2$, then

$$\Psi(t) = (1-p)b^2 \qquad \forall t\in[0,1], \tag{A.34}$$

so

$$\arg\max_{t\in[0,1]} \Psi(t) = [0,1]. \tag{A.35}$$

It remains to derive the posterior expected-power first-order condition. Let

$$c := c_{1-\alpha}, \qquad x := \sqrt{c}, \qquad Z \sim N(0,1), \tag{A.36}$$

and write F_N and ϕ for the standard normal cdf and density. For $q \geq 0$,

$$\begin{aligned} h_\alpha(q) &= \Pr\{(Z+\sqrt{q})^2 > c\} \\ &= \Pr\{Z+\sqrt{q} > x\} + \Pr\{Z+\sqrt{q} < -x\} \\ &= \Pr\{Z > x-\sqrt{q}\} + \Pr\{Z < -x-\sqrt{q}\} \\ &= 1 - F_N(x-\sqrt{q}) + F_N(-x-\sqrt{q}). \end{aligned} \tag{A.37}$$

For $q > 0$, set $s = \sqrt{q}$. Then

$$h_\alpha(s^2) = 1 - F_N(x-s) + F_N(-x-s). \tag{A.38}$$

Differentiating with respect to s,

$$\begin{aligned} \frac{d}{ds} h_\alpha(s^2) &= -\phi(x-s)(-1) + \phi(-x-s)(-1) \\ &= \phi(x-s) - \phi(-x-s) \\ &= \phi(x-s) - \phi(x+s). \end{aligned} \tag{A.39}$$

Since

$$\frac{dq}{ds} = 2s, \qquad \frac{ds}{dq} = \frac{1}{2s}, \tag{A.40}$$

we obtain

$$h'_\alpha(q) = \frac{\phi(x-\sqrt{q}) - \phi(x+\sqrt{q})}{2\sqrt{q}}, \qquad q > 0. \tag{A.41}$$

The right derivative at zero is

$$h'_\alpha(0) := \lim_{q\downarrow 0} h'_\alpha(q) = \lim_{s\downarrow 0} \frac{\phi(x-s) - \phi(x+s)}{2s}. \tag{A.42}$$

Using

$$\phi(x-s) = \phi(x) - s\phi'(x) + o(s), \qquad \phi(x+s) = \phi(x) + s\phi'(x) + o(s), \tag{A.43}$$

we get

$$\begin{aligned} h'_\alpha(0) &= \lim_{s\downarrow 0} \frac{\{\phi(x) - s\phi'(x) + o(s)\} - \{\phi(x) + s\phi'(x) + o(s)\}}{2s} \\ &= \lim_{s\downarrow 0} \frac{-2s\phi'(x) + o(s)}{2s} \\ &= -\phi'(x). \end{aligned} \tag{A.44}$$

Since

$$\phi'(x) = -x\phi(x), \tag{A.45}$$

we have

$$h'_\alpha(0) = x\phi(x) = \sqrt{c_{1-\alpha}}\,\phi(\sqrt{c_{1-\alpha}}). \tag{A.46}$$

For $q > 0$,

$$x + \sqrt{q} > |x - \sqrt{q}|. \tag{A.47}$$

Since

$$\phi(u) = (2\pi)^{-1/2} \exp(-u^2/2) \tag{A.48}$$

is strictly decreasing in $|u|$, (A.41) implies

$$h'_\alpha(q) > 0 \qquad \forall q \geq 0. \tag{A.49}$$

For $t \in (0, 1)$, the chain rule gives

$$\frac{d}{dt} h_\alpha(\rho_{\inf} a^2 t) = \rho_{\inf} a^2 h'_\alpha(\rho_{\inf} a^2 t), \tag{A.50}$$

and

$$\frac{d}{dt} h_\alpha\{\rho_{\inf} b^2(1-t)\} = -\rho_{\inf} b^2 h'_\alpha\{\rho_{\inf} b^2(1-t)\}. \tag{A.51}$$

Combining (A.24), (A.50), and (A.51),

$$\Phi'_\alpha(t) = p\,\rho_{\inf} a^2 h'_\alpha(\rho_{\inf} a^2 t) - (1-p)\rho_{\inf} b^2 h'_\alpha\{\rho_{\inf} b^2(1-t)\}. \tag{A.52}$$

If $t \in (0, 1)$ maximizes Φ_α, then

$$\Phi'_\alpha(t) = 0. \tag{A.53}$$

Equations (A.52) and (A.53), together with $\rho_{\inf} > 0$, imply

$$pa^2 h'_\alpha(\rho_{\inf} a^2 t) = (1-p) b^2 h'_\alpha\{\rho_{\inf} b^2(1-t)\}, \tag{A.54}$$

which is (3.8).

The dependence on the nominal size is explicit from (A.41) and (A.46), because $x = \sqrt{c_{1-\alpha}}$. Equivalently, for $q > 0$,

$$\frac{h'_\alpha(q)}{h'_\alpha(0)} = \exp(-q/2)\frac{\sinh\{x\sqrt{q}\}}{x\sqrt{q}}, \qquad x = \sqrt{c_{1-\alpha}}, \tag{A.55}$$

which is not constant in x. Therefore the first-order equation (A.54) is not generally invariant to α.

Finally impose (3.9). From

$$pa^2 > (1-p)b^2 \tag{A.56}$$

and (A.31),

$$\arg\max_{t \in [0,1]} \Psi(t) = \{1\}. \tag{A.57}$$

The left derivative of Φ_α at $t=1$ is

$$\Phi'_\alpha(1-) = \lim_{t\uparrow 1}\left[p\,\rho_{\inf}a^2h'_\alpha(\rho_{\inf}a^2t) - (1-p)\rho_{\inf}b^2h'_\alpha\{\rho_{\inf}b^2(1-t)\}\right] \\ = \rho_{\inf}\left[pa^2h'_\alpha(\rho_{\inf}a^2) - (1-p)b^2h'_\alpha(0)\right]. \tag{A.58}$$

By (3.9),

$$pa^2h'_\alpha(\rho_{\inf}a^2) - (1-p)b^2h'_\alpha(0) < 0. \tag{A.59}$$

Hence

$$\Phi'_\alpha(1-) < 0. \tag{A.60}$$

By continuity of h'_α at zero and on $(0,\infty)$, there exists $\varepsilon_0 \in (0,1)$ such that

$$\Phi'_\alpha(u) < 0 \qquad \forall u \in (1-\varepsilon_0, 1). \tag{A.61}$$

For every $\varepsilon \in (0,\varepsilon_0)$,

$$\Phi_\alpha(1) - \Phi_\alpha(1-\varepsilon) = \int_{1-\varepsilon}^{1}\Phi'_\alpha(u)\,du \\ < 0. \tag{A.62}$$

Therefore

$$\Phi_\alpha(1-\varepsilon) > \Phi_\alpha(1), \qquad 1 \notin \arg\max_{t\in[0,1]}\Phi_\alpha(t). \tag{A.63}$$

Combining (A.57) and (A.63),

$$\arg\max_{t\in[0,1]}\Psi(t) = \{1\}, \qquad 1 \notin \arg\max_{t\in[0,1]}\Phi_\alpha(t). \tag{A.64}$$

This proves the stated noncoincidence. *Q.E.D.*

A.4. *Proof of Theorem 4.1*

The projective lower bound follows from the sparse projective packing, KL/Fano, and local perturbation arguments developed in Online Appendix Section 2. The hard-thresholding upper bound follows from the hard-thresholding oracle inequality in Online Appendix Lemma 2.5. The Online Appendix gives the full constant-level derivations of these auxiliary ingredients.

LEMMA A.1—Projective sparse lower bound: *In the Gaussian experiment*

$$Y = g + \sigma_n\xi, \qquad \xi \sim N(0, I_{p_n}),$$

over

$$\mathcal{G}_n(s_n, r_n) = \{g \in \mathbb{R}^{p_n} : \|g\|_0 \le s_n,\ \|g\|_2 = r_n\}, \qquad 2 \le s_n \le p_n/4,$$

there exists a universal constant $c>0$ such that

$$\inf_{\delta_n}\sup_{g\in\mathcal{G}_n(s_n,r_n)}\mathbb{E}_gL_n(\delta_n, g) \ge c\min\left\{1,\ \frac{\sigma_n^2 s_n\log(ep_n/s_n)}{r_n^2}\right\}.$$

The full proof is given by the projective packing, KL/Fano, and local perturbation arguments in Online Appendix Section 2.

LEMMA A.2—Hard-thresholding upper bound: *Let $H_{s_n}(Y)$ retain the s_n largest coordinates of Y in absolute value and set all others to zero. There exists a universal constant $C<\infty$ such that*

$$\sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(H_{s_n},g) \le C\min\left\{1, \frac{\sigma_n^2 s_n \log(ep_n/s_n)}{r_n^2}\right\}.$$

The full proof is given in Online Appendix Lemma 2.5.

PROOF: The lower bound in (4.4) is Lemma A.1. For the upper bound, use the hard-thresholding rule H_{s_n} in Lemma A.2. Since the minimax infimum is over all measurable score rules,

$$\inf_{\delta_n}\sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(\delta_n,g) \le \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(H_{s_n},g).$$

Combining the lower and upper bounds gives (4.4). *Q.E.D.*

A.5. *Proof of Lemma 4.3*

PROOF: Let

$$c := c_{1-\alpha}, \qquad x := \sqrt{c}, \qquad Z\sim N(0,1).$$

For $q\ge 0$,

$$\begin{aligned} h_\alpha(q) &= \Pr\{(Z+\sqrt{q})^2 > c\} \\ &= \Pr\{Z+\sqrt{q} > x\} + \Pr\{Z+\sqrt{q} < -x\} \\ &= 1-\Phi(x-\sqrt{q}) + \Phi(-x-\sqrt{q}). \end{aligned} \tag{A.65}$$

For $q>0$, set $s=\sqrt{q}$. Then

$$h_\alpha(s^2) = 1-\Phi(x-s)+\Phi(-x-s).$$

Hence

$$\begin{aligned} \frac{d}{ds}h_\alpha(s^2) &= -\phi(x-s)(-1)+\phi(-x-s)(-1) \\ &= \phi(x-s)-\phi(-x-s) \\ &= \phi(x-s)-\phi(x+s). \end{aligned} \tag{A.66}$$

Since $q=s^2$,

$$\frac{ds}{dq} = \frac{1}{2s} = \frac{1}{2\sqrt{q}}.$$

Combining this with (A.66),

$$h'_\alpha(q) = \frac{\phi(x-\sqrt{q})-\phi(x+\sqrt{q})}{2\sqrt{q}}, \qquad q>0. \tag{A.67}$$

For $q > 0$,

$$\begin{aligned}\phi(x-\sqrt{q})-\phi(x+\sqrt{q}) &= (2\pi)^{-1/2}\left[e^{-(x-\sqrt{q})^2/2}-e^{-(x+\sqrt{q})^2/2}\right]\\ &= (2\pi)^{-1/2}e^{-(x+\sqrt{q})^2/2}\left[e^{\{(x+\sqrt{q})^2-(x-\sqrt{q})^2\}/2}-1\right] \\ &= (2\pi)^{-1/2}e^{-(x+\sqrt{q})^2/2}\left[e^{2x\sqrt{q}}-1\right].\end{aligned} \tag{A.68}$$

Since $x = \sqrt{c_{1-\alpha}} > 0$, (A.68) gives

$$\phi(x-\sqrt{q})-\phi(x+\sqrt{q}) > 0 \qquad (q > 0).$$

Therefore

$$h'_\alpha(q) > 0 \qquad (q > 0). \tag{A.69}$$

It remains to compute the right derivative at zero. Using (A.68) with $s = \sqrt{q}$,

$$\begin{aligned}\lim_{q\downarrow 0} h'_\alpha(q) &= \lim_{s\downarrow 0}\frac{\phi(x-s)-\phi(x+s)}{2s}\\ &= \lim_{s\downarrow 0}\frac{(2\pi)^{-1/2}e^{-(x+s)^2/2}(e^{2xs}-1)}{2s}.\end{aligned} \tag{A.70}$$

Since

$$\lim_{s\downarrow 0} e^{-(x+s)^2/2} = e^{-x^2/2}, \qquad \lim_{s\downarrow 0}\frac{e^{2xs}-1}{2s} = x,$$

(A.70) gives

$$h'_\alpha(0) := \lim_{q\downarrow 0} h'_\alpha(q) = x(2\pi)^{-1/2}e^{-x^2/2} = x\phi(x) = \sqrt{c_{1-\alpha}}\,\phi(\sqrt{c_{1-\alpha}}) > 0. \tag{A.71}$$

Equations (A.67)–(A.71) imply that h'_α extends continuously to $q = 0$ and is strictly positive on $[0,\infty)$.

Fix $\bar{q} < \infty$. Define

$$m_{\alpha,\bar{q}} := \inf_{u\in[0,\bar{q}]} h'_\alpha(u), \qquad M_{\alpha,\bar{q}} := \sup_{u\in[0,\bar{q}]} h'_\alpha(u). \tag{A.72}$$

Continuity of h'_α on the compact interval $[0,\bar{q}]$, together with strict positivity, gives

$$0 < m_{\alpha,\bar{q}} \le M_{\alpha,\bar{q}} < \infty. \tag{A.73}$$

For $q \in [0,\bar{q}]$ and $\ell \in [0,1]$,

$$q(1-\ell) \in [0,q] \subset [0,\bar{q}].$$

By the fundamental theorem of calculus,

$$h_\alpha(q) - h_\alpha\{q(1-\ell)\} = \int_{q(1-\ell)}^{q} h'_\alpha(u)\,du. \tag{A.74}$$

Using (A.73) in (A.74),

$$m_{\alpha,\bar{q}}\{q-q(1-\ell)\} \le h_\alpha(q)-h_\alpha\{q(1-\ell)\} \le M_{\alpha,\bar{q}}\{q-q(1-\ell)\}.$$

Since

$$q-q(1-\ell)=q\ell,$$

the preceding display becomes

$$m_{\alpha,\bar{q}}\,q\ell \le h_\alpha(q)-h_\alpha\{q(1-\ell)\} \le M_{\alpha,\bar{q}}\,q\ell,$$

which is (4.5). *Q.E.D.*

A.6. *Proof of Corollary 4.4*

PROOF: Write $p=p_n$, $s=s_n$, $r=r_n$, and

$$q_n^\star := \rho_{\inf} r^2.$$

For every $g\in\mathcal{G}_n(s,r)$,

$$\|g\|_2=r.$$

Therefore

$$\begin{aligned} Q_n^\star(g) &= \sup_{\beta\neq 0}\rho_{\inf}\frac{(\beta' g)^2}{\|\beta\|_2^2} \\ &= \rho_{\inf}\|g\|_2^2 = q_n^\star. \end{aligned} \tag{A.75}$$

The supremum is attained by any nonzero scalar multiple of g.

Let $\delta=\delta(Y)$ be any measurable score rule. If $\delta(Y)\neq 0$, then

$$Q_n\{\delta(Y);g,I_p\}=\rho_{\inf}\frac{\{\delta(Y)'g\}^2}{\|\delta(Y)\|_2^2},$$

and

$$L_n(\delta,g)=1-\frac{\{\delta(Y)'g\}^2}{\|\delta(Y)\|_2^2\|g\|_2^2}.$$

Hence, on $\{\delta(Y)\neq 0\}$,

$$\begin{aligned} Q_n\{\delta(Y);g,I_p\} &= \rho_{\inf}\|g\|_2^2\frac{\{\delta(Y)'g\}^2}{\|\delta(Y)\|_2^2\|g\|_2^2} \\ &= Q_n^\star(g)\{1-L_n(\delta,g)\}. \end{aligned} \tag{A.76}$$

On $\{\delta(Y)=0\}$, the convention is $L_n(\delta,g)=1$ and $Q_n\{\delta(Y);g,I_p\}=0$, so (A.76) remains valid. Thus, for all realizations,

$$Q_n\{\delta(Y);g,I_p\}=q_n^\star\{1-L_n(\delta,g)\}. \tag{A.77}$$

By assumption,

$$q_n^\star\in[\underline{q},\bar{q}]. \tag{A.78}$$

Applying Lemma 4.3 with $\overline{q}$, $q = q_n^\star$, and $\ell = L_n(\delta, g)$, and using $0 \le L_n(\delta, g) \le 1$, gives

$$\begin{aligned} m_{\alpha,\overline{q}}\, q_n^\star L_n(\delta, g) &\le h_\alpha(q_n^\star) - h_\alpha\{q_n^\star(1 - L_n(\delta, g))\} \\ &\le M_{\alpha,\overline{q}}\, q_n^\star L_n(\delta, g). \end{aligned} \tag{A.79}$$

Using (A.78), this implies

$$\begin{aligned} m_{\alpha,\overline{q}}\, \underline{q}\, L_n(\delta, g) &\le h_\alpha(q_n^\star) - h_\alpha\{q_n^\star(1 - L_n(\delta, g))\} \\ &\le M_{\alpha,\overline{q}}\, \overline{q}\, L_n(\delta, g). \end{aligned} \tag{A.80}$$

By (A.77),

$$h_\alpha\{q_n^\star(1 - L_n(\delta, g))\} = h_\alpha(Q_n\{\delta(Y); g, I_p\})\,.$$

Thus (A.80) is equivalent to

$$\begin{aligned} m_{\alpha,\overline{q}}\, \underline{q}\, L_n(\delta, g) &\le h_\alpha\{Q_n^\star(g)\} - h_\alpha(Q_n\{\delta(Y); g, I_p\}) \\ &\le M_{\alpha,\overline{q}}\, \overline{q}\, L_n(\delta, g). \end{aligned} \tag{A.81}$$

Taking P_g-expectations,

$$\begin{aligned} m_{\alpha,\overline{q}}\, \underline{q}\, E_g L_n(\delta, g) &\le h_\alpha\{Q_n^\star(g)\} - E_g h_\alpha(Q_n\{\delta(Y); g, I_p\}) \\ &\le M_{\alpha,\overline{q}}\, \overline{q}\, E_g L_n(\delta, g). \end{aligned} \tag{A.82}$$

The middle term is exactly $\mathcal{R}_{\alpha,n}(\delta, g)$. Therefore

$$a_\alpha E_g L_n(\delta, g) \le \mathcal{R}_{\alpha,n}(\delta, g) \le A_\alpha E_g L_n(\delta, g), \tag{A.83}$$

where

$$a_\alpha := m_{\alpha,\overline{q}}\, \underline{q} > 0, \qquad A_\alpha := M_{\alpha,\overline{q}}\, \overline{q} < \infty.$$

Taking suprema over $g \in \mathcal{G}_n(s, r)$ gives, for every rule δ,

$$a_\alpha \sup_{g \in \mathcal{G}_n(s,r)} E_g L_n(\delta, g) \le \sup_{g \in \mathcal{G}_n(s,r)} \mathcal{R}_{\alpha,n}(\delta, g) \le A_\alpha \sup_{g \in \mathcal{G}_n(s,r)} E_g L_n(\delta, g). \tag{A.84}$$

Taking infima over all measurable rules δ, the left inequality gives

$$\inf_\delta \sup_{g \in \mathcal{G}_n(s,r)} \mathcal{R}_{\alpha,n}(\delta, g) \ge a_\alpha \inf_\delta \sup_{g \in \mathcal{G}_n(s,r)} E_g L_n(\delta, g). \tag{A.85}$$

Similarly, the right inequality gives

$$\inf_\delta \sup_{g \in \mathcal{G}_n(s,r)} \mathcal{R}_{\alpha,n}(\delta, g) \le A_\alpha \inf_\delta \sup_{g \in \mathcal{G}_n(s,r)} E_g L_n(\delta, g). \tag{A.86}$$

Applying Theorem 4.1 to the two displays yields

$$c_\alpha \min\{1, \eta_n\} \le \inf_\delta \sup_{g \in \mathcal{G}_n(s_n, r_n)} \mathcal{R}_{\alpha,n}(\delta, g) \le C_\alpha \min\{1, \eta_n\},$$

for constants $0 < c_\alpha < C_\alpha < \infty$ depending only on $\alpha, \underline{q}, \overline{q}$. This proves (4.6). *Q.E.D.*

A.7. *Proof of Corollary 4.5*

PROOF: For fixed ρ, the training experiment is the same canonical sparse Gaussian experiment as in Theorem 4.1, with noise variance $\rho^{-1}\sigma_n^2$ replacing σ_n^2. Hence Theorem 4.1 gives

$$\inf_{\delta_{n,\rho}} \sup_{g\in\mathcal{G}_n(s_n,r_n)} E_g L_n(\delta_{n,\rho}, g) \asymp \min\left\{1, \frac{\rho^{-1}\sigma_n^2 s_n \log(ep_n/s_n)}{r_n^2}\right\} = \min\{1, \eta_n(\rho)\}.$$

The constants are uniform over $\rho \in [\underline{\rho}, 1-\underline{\rho}]$. On $\mathcal{R}_n$, the oracle noncentrality

$$q_n^\star(\rho) = (1-\rho)r_n^2$$

lies in $[\underline{q}, \overline{q}]$. Lemma 4.3 therefore transfers normalized noncentrality regret to power regret uniformly over $\rho \in \mathcal{R}_n$. This gives

$$\inf_{\delta_{n,\rho}} \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathcal{R}^\rho_{\alpha,n}(\delta_{n,\rho}, g) \asymp \min\{1, \eta_n(\rho)\}.$$

Finally, since $q_n^\star(\rho)$ is fixed over the shell $\mathcal{G}_n(s_n, r_n)$,

$$\begin{aligned}\mathcal{V}^\rho_{\alpha,n} &= \sup_{\delta_{n,\rho}} \inf_{g\in\mathcal{G}_n(s_n,r_n)} E_g h_\alpha\{Q_{n,\rho}(\delta_{n,\rho}(Y_\rho); g, I_{p_n})\} \\ &= h_\alpha\{q_n^\star(\rho)\} - \inf_{\delta_{n,\rho}} \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathcal{R}^\rho_{\alpha,n}(\delta_{n,\rho}, g).\end{aligned}$$

Combining this identity with the preceding envelope gives the displayed value bounds. *Q.E.D.*

A.8. *Proof of Corollary 4.6*

PROOF: Let

$$\eta_n = \frac{\sigma_n^2 s_n \log(ep_n/s_n)}{r_n^2}. \tag{A.87}$$

Let $c_L > 0$ and $C_L < \infty$ denote universal constants such that Theorem 4.1 gives

$$c_L \min\{1, \eta_n\} \le \inf_{\delta_n} \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(\delta_n, g) \le C_L \min\{1, \eta_n\}. \tag{A.88}$$

Moreover, by the hard-thresholding upper-bound ingredient used in the proof of Theorem 4.1, there exists a universal constant $C_H < \infty$ such that

$$\sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(H_{s_n}, g) \le C_H \min\{1, \eta_n\}. \tag{A.89}$$

First suppose

$$\eta_n \ge 1 \tag{A.90}$$

eventually. Then

$$\min\{1, \eta_n\} = 1 \tag{A.91}$$

eventually. Combining (A.88) and (A.91),

$$\inf_{\delta_n} \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(\delta_n, g) \ge c_L \tag{A.92}$$

eventually. Thus part (i) holds with

$$c_0 := c_L. \tag{A.93}$$

Next suppose

$$\eta_n \to 0. \tag{A.94}$$

Then, for all sufficiently large n,

$$\min\{1, \eta_n\} = \eta_n. \tag{A.95}$$

Combining (A.89) and (A.95),

$$\sup_{g \in \mathcal{G}_n(s_n, r_n)} \mathbb{E}_g L_n(H_{s_n}, g) \le C_H \eta_n \tag{A.96}$$

eventually. Hence

$$\sup_{g \in \mathcal{G}_n(s_n, r_n)} \mathbb{E}_g L_n(H_{s_n}, g) = O(\eta_n), \tag{A.97}$$

which proves part (ii).

It remains to prove the power-regret transfer in part (iii). Suppose

$$\rho_{\inf} r_n^2 \in [\underline{q}, \overline{q}] \subset (0, \infty). \tag{A.98}$$

For $g \in \mathcal{G}_n(s_n, r_n)$,

$$Q_n^\star(g) = \rho_{\inf} \|g\|^2 = \rho_{\inf} r_n^2. \tag{A.99}$$

Set

$$q_n^\star := \rho_{\inf} r_n^2. \tag{A.100}$$

By (A.98),

$$q_n^\star \in [\underline{q}, \overline{q}]. \tag{A.101}$$

For any measurable score rule δ_n, the definition of normalized noncentrality regret gives, with the convention $L_n = 1$ when $\delta_n(Y) = 0$,

$$Q_n\{\delta_n(Y); g, I_{p_n}\} = q_n^\star\{1 - L_n(\delta_n, g)\}. \tag{A.102}$$

Indeed, if $\delta_n(Y) \neq 0$, then

$$\begin{aligned} Q_n\{\delta_n(Y); g, I_{p_n}\} &= \rho_{\inf} \frac{\{\delta_n(Y)' g\}^2}{\|\delta_n(Y)\|^2} \\ &= \rho_{\inf} \|g\|^2 \frac{\{\delta_n(Y)' g\}^2}{\|\delta_n(Y)\|^2 \|g\|^2} \\ &= q_n^\star\{1 - L_n(\delta_n, g)\}. \end{aligned} \tag{A.103}$$

If $\delta_n(Y) = 0$, then $Q_n\{\delta_n(Y); g, I_{p_n}\} = 0$ and $L_n(\delta_n, g) = 1$, so (A.102) also holds.

By Lemma 4.3, there exist constants $0 < m_{\alpha,\overline{q}} \le M_{\alpha,\overline{q}} < \infty$ such that, for all $q \in [0, \overline{q}]$ and all $\ell \in [0, 1]$,

$$m_{\alpha,\overline{q}} q\ell \le h_\alpha(q) - h_\alpha\{q(1-\ell)\} \le M_{\alpha,\overline{q}} q\ell. \tag{A.104}$$

Apply (A.104) with

$$q = q_n^\star, \qquad \ell = L_n(\delta_n, g).$$

Using (A.101), we obtain

$$\begin{aligned} m_{\alpha,\overline{q}}\, \underline{q}\, L_n(\delta_n, g) &\le h_\alpha(q_n^\star) - h_\alpha\{q_n^\star[1 - L_n(\delta_n, g)]\} \\ &\le M_{\alpha,\overline{q}}\, \overline{q}\, L_n(\delta_n, g). \end{aligned} \tag{A.105}$$

By (A.102),

$$\begin{aligned} m_{\alpha,\overline{q}}\, \underline{q}\, L_n(\delta_n, g) &\le h_\alpha\{Q_n^\star(g)\} - h_\alpha(Q_n\{\delta_n(Y); g, I_{p_n}\}) \\ &\le M_{\alpha,\overline{q}}\, \overline{q}\, L_n(\delta_n, g). \end{aligned} \tag{A.106}$$

Taking expectation under P_g,

$$\begin{aligned} m_{\alpha,\overline{q}}\, \underline{q}\, \mathbb{E}_g L_n(\delta_n, g) &\le h_\alpha\{Q_n^\star(g)\} - \mathbb{E}_g h_\alpha(Q_n\{\delta_n(Y); g, I_{p_n}\}) \\ &\le M_{\alpha,\overline{q}}\, \overline{q}\, \mathbb{E}_g L_n(\delta_n, g). \end{aligned} \tag{A.107}$$

The middle term is exactly

$$\mathcal{R}_{\alpha,n}(\delta_n, g) = h_\alpha\{Q_n^\star(g)\} - \mathbb{E}_g h_\alpha(Q_n\{\delta_n(Y); g, I_{p_n}\})\,. \tag{A.108}$$

Therefore

$$a_\alpha\, \mathbb{E}_g L_n(\delta_n, g) \le \mathcal{R}_{\alpha,n}(\delta_n, g) \le A_\alpha\, \mathbb{E}_g L_n(\delta_n, g), \tag{A.109}$$

where

$$a_\alpha := m_{\alpha,\overline{q}}\, \underline{q} > 0, \qquad A_\alpha := M_{\alpha,\overline{q}}\, \overline{q} < \infty. \tag{A.110}$$

Taking suprema over $g \in \mathcal{G}_n(s_n, r_n)$, for any fixed rule δ_n,

$$a_\alpha \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(\delta_n, g) \le \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathcal{R}_{\alpha,n}(\delta_n, g) \le A_\alpha \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(\delta_n, g). \tag{A.111}$$

The lower-bound side of (A.111) and (A.88) imply

$$\begin{aligned} \inf_{\delta_n} \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathcal{R}_{\alpha,n}(\delta_n, g) &\ge a_\alpha \inf_{\delta_n} \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(\delta_n, g) \\ &\ge a_\alpha c_L \min\{1, \eta_n\}. \end{aligned} \tag{A.112}$$

If $\eta_n \ge 1$ eventually, (A.112) gives

$$\inf_{\delta_n} \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathcal{R}_{\alpha,n}(\delta_n, g) \ge a_\alpha c_L \tag{A.113}$$

eventually.

For the high-signal side, apply the right-hand side of (A.111) to the hard-thresholding rule H_{s_n}. Combining with (A.89),

$$\begin{aligned} \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathcal{R}_{\alpha,n}(H_{s_n}, g) &\le A_\alpha \sup_{g\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}_g L_n(H_{s_n}, g) \\ &\le A_\alpha C_H \min\{1, \eta_n\}. \end{aligned} \tag{A.114}$$

If $\eta_n \to 0$, then (A.114) gives

$$\sup_{g \in \mathcal{G}_n(s_n, r_n)} \mathcal{R}_{\alpha,n}(H_{s_n}, g) = O(\eta_n). \tag{A.115}$$

Equations (A.113) and (A.115) prove the power-regret transfer in part (iii). *Q.E.D.*

A.9. *Proof of Corollary 4.7*

PROOF: Let $q_n^\star := \rho_{\inf} r_n^2$. By assumption, $0 < \underline{q} \le q_n^\star \le \overline{q} < \infty$. By the maintained nondegenerate split condition, there exists $\underline{\rho} > 0$ such that $\rho_{\inf} \ge \underline{\rho}$ eventually. Hence

$$r_n^2 = \frac{q_n^\star}{\rho_{\inf}} \le \frac{\overline{q}}{\underline{\rho}}$$

eventually. Therefore, using $\sigma_n^2 \ge \underline{\sigma}^2 > 0$,

$$\begin{aligned}\eta_n &= \frac{\sigma_n^2 s_n \log(ep_n/s_n)}{r_n^2} \\ &\ge \frac{\underline{\sigma}^2 \underline{\rho}}{\overline{q}} s_n \log(ep_n/s_n) \longrightarrow \infty.\end{aligned}$$

Thus $\eta_n \ge 1$ eventually. Corollary 4.4 gives

$$\inf_{\delta_n} \sup_{g \in \mathcal{G}_n(s_n, r_n)} \mathcal{R}_{\alpha,n}(\delta_n, g) \ge c_\alpha \min\{1, \eta_n\} = c_\alpha$$

eventually. This proves the claim. *Q.E.D.*

A.10. *Proof of Proposition 4.8*

PROOF: Let $Z_{p_n}(q_n) \sim \chi^2_{p_n}(q_n)$. Write

$$Z_{p_n}(q_n) = \sum_{j=1}^{p_n} (X_{nj} + a_{nj})^2, \qquad X_{nj} \overset{iid}{\sim} N(0,1), \qquad \sum_{j=1}^{p_n} a_{nj}^2 = q_n,$$

with $q_n \to q \in (0, \infty)$. Define $W_{nj} = (X_{nj} + a_{nj})^2 - (1 + a_{nj}^2)$. Then $\mathbb{E}W_{nj} = 0$, $\mathrm{Var}(W_{nj}) = 2 + 4a_{nj}^2$, and $v_n^2 := \sum_{j=1}^{p_n} \mathrm{Var}(W_{nj}) = 2p_n + 4q_n$. Moreover, for a universal $C < \infty$, $\mathbb{E}|W_{nj}|^3 \le C(1 + |a_{nj}|^6)$, so

$$\sum_{j=1}^{p_n} \mathbb{E}|W_{nj}|^3 \le C\left(p_n + \sum_{j=1}^{p_n} |a_{nj}|^6\right) \le C(p_n + q_n^3) = O(p_n).$$

Since $v_n^3 \asymp p_n^{3/2}$,

$$\frac{\sum_{j=1}^{p_n} \mathbb{E}|W_{nj}|^3}{v_n^3} \to 0.$$

Lyapunov's theorem gives

$$\frac{Z_{p_n}(q_n) - p_n - q_n}{\{2(p_n + 2q_n)\}^{1/2}} = \frac{\sum_{j=1}^{p_n} W_{nj}}{v_n} \xrightarrow{d} N(0,1).$$

Let $c_{p_n,1-\alpha}$ be the $(1-\alpha)$-quantile of $\chi^2_{p_n}$. For the central chi-square,

$$\frac{\chi^2_{p_n} - p_n}{\sqrt{2p_n}} \xrightarrow{d} N(0,1),$$

and hence

$$\frac{c_{p_n,1-\alpha} - p_n}{\sqrt{2p_n}} \to z_{1-\alpha}.$$

Since $q_n = O(1)$,

$$\frac{c_{p_n,1-\alpha} - p_n - q_n}{\{2(p_n + 2q_n)\}^{1/2}} = \frac{c_{p_n,1-\alpha} - p_n}{\sqrt{2p_n}} \frac{\sqrt{2p_n}}{\{2(p_n + 2q_n)\}^{1/2}} - \frac{q_n}{\{2(p_n + 2q_n)\}^{1/2}} \to z_{1-\alpha}.$$

Therefore

$$\Pr\{Z_{p_n}(q_n) > c_{p_n,1-\alpha}\} = \Pr\left\{\frac{Z_{p_n}(q_n) - p_n - q_n}{\{2(p_n + 2q_n)\}^{1/2}} > \frac{c_{p_n,1-\alpha} - p_n - q_n}{\{2(p_n + 2q_n)\}^{1/2}}\right\}$$
$$\to \Pr\{N(0,1) > z_{1-\alpha}\} = \alpha.$$

For the scalar statistic, put $x_\alpha = \sqrt{c_{1-\alpha}}$. For $q > 0$,

$$h_\alpha(q) = \Pr\{(Z + \sqrt{q})^2 > c_{1-\alpha}\} = 1 - \Phi(x_\alpha - \sqrt{q}) + \Phi(-x_\alpha - \sqrt{q}).$$

For $q > 0$,

$$h'_\alpha(q) = \frac{\phi(x_\alpha - \sqrt{q}) - \phi(x_\alpha + \sqrt{q})}{2\sqrt{q}} > 0,$$

and $h'_\alpha(0+) = x_\alpha \phi(x_\alpha) > 0$. Thus h_α is strictly increasing on $[0,\infty)$. Since $q > 0$, $h_\alpha(q) > h_\alpha(0) = \alpha$. This proves the proposition. *Q.E.D.*

A.11. *Proof of Theorem 5.2*

PROOF: Let $\mathcal{E}_n = \{\widehat{\beta}_n \in \mathcal{H}_{n,\Delta}(c,C)\}$. By assumption, $\Pr(\mathcal{E}_n^c) \to 0$. All uniform approximations below are applied on $\mathcal{E}_n$. Since rejection probabilities and bounded-Lipschitz test functions are bounded by constants, the contribution of $\mathcal{E}_n^c$ is $o(1)$. Let

$$\mathcal{F}_{\rm tr} := \sigma\{(Y_i, D_i, Z_i, X_i) : i \in I_{\rm tr}\}, \qquad b_i := b_{p_n}(Z_i, X_i), \qquad \pi_i := \pi_n(Z_i, X_i).$$

From (5.2), $D_i - m_N(X_i) = N^{-1/2}\pi_i + V_i$. Hence

$$\begin{aligned}\widehat{g}_n - g_n = &\left\{\frac{1}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_i\pi_i - g_n\right\} \\ &+ \frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_i V_i - \frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_i\{\widehat{m}(X_i) - m_N(X_i)\}.\end{aligned} \tag{A.116}$$

Premultiplying by $C_n^{-1/2}$ gives

$$C_n^{-1/2}(\widehat{g}_n - g_n) = A_{n,\mathrm{tr}} + Z_{n,\mathrm{tr}} - M_{n,\mathrm{tr}},$$

where

$$\begin{aligned}A_{n,\mathrm{tr}} &:= C_n^{-1/2}\left\{\frac{1}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_i\pi_i - g_n\right\}, \\ Z_{n,\mathrm{tr}} &:= C_n^{-1/2}\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_i V_i, \\ M_{n,\mathrm{tr}} &:= C_n^{-1/2}\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_i\{\widehat{m}(X_i) - m_N(X_i)\}.\end{aligned}$$

By Assumption 5.1,

$$\|A_{n,\mathrm{tr}}\|_2 = o_p(1), \qquad \|M_{n,\mathrm{tr}}\|_2 = o_p(1),$$

and

$$d_{\mathrm{BL}}\big(\mathcal{L}\{Z_{n,\mathrm{tr}} \mid \mathcal{D}_{\mathrm{tr}}\}, N(0, \rho_{\mathrm{tr}}^{-1} I_{p_n})\big) \xrightarrow{p} 0.$$

Set

$$a_{n,\mathrm{tr}} := A_{n,\mathrm{tr}} - M_{n,\mathrm{tr}}.$$

Then

$$C_n^{-1/2}(\widehat{g}_n - g_n) = Z_{n,\mathrm{tr}} + a_{n,\mathrm{tr}}, \qquad \|a_{n,\mathrm{tr}}\|_2 = o_p(1),$$

and the bounded-Lipschitz approximation for $Z_{n,\mathrm{tr}}$ is the display above. This proves part (i).

Now condition on $\mathcal{F}_{\mathrm{tr}}$. Then

$$\widehat{\beta}_n := \delta_n(\widehat{g}_n)$$

is fixed. Let

$$R_{i,\Delta} := U_i + \Delta V_i, \qquad B := B_{n,\Delta}, \qquad G := G_{n,\Delta} = \Delta g_n.$$

The oracle inference residual satisfies

$$\begin{aligned}Y_i - \mu_{N,\Delta}(X_i) - \theta_0\{D_i - m_N(X_i)\} &= U_i + \Delta\{D_i - m_N(X_i)\} \\ &= R_{i,\Delta} + \Delta N^{-1/2}\pi_i.\end{aligned}$$

For any $\beta \in \mathcal{H}_{n,\Delta}(c, C)$, define

$$S_n^o(\beta) := \frac{1}{\sqrt{n_{\text{inf}}}} \sum_{i \in I_{\text{inf}}} \beta' b_i \{Y_i - \mu_{N,\Delta}(X_i) - \theta_0 [D_i - m_N(X_i)]\}.$$

Then

$$S_n^o(\beta) = A_{n,\Delta}(\beta) + D_{n,\Delta}(\beta), \tag{A.117}$$

where

$$A_{n,\Delta}(\beta) := \frac{1}{\sqrt{n_{\text{inf}}}} \sum_{i \in I_{\text{inf}}} \beta' b_i R_{i,\Delta},$$

and

$$D_{n,\Delta}(\beta) := \sqrt{\frac{n_{\text{inf}}}{N}}\, \Delta\, \beta' \widehat{g}_n^{\text{inf}}, \qquad \widehat{g}_n^{\text{inf}} := \frac{1}{n_{\text{inf}}} \sum_{i \in I_{\text{inf}}} b_i \pi_i.$$

By (5.10),

$$\sup_{\beta \in \mathcal{H}_{n,\Delta}(c,C)} \left| \frac{\Delta \beta' (\widehat{g}_n^{\text{inf}} - g_n)}{(\beta' B \beta)^{1/2}} \right| = o_p(1).$$

Therefore, uniformly on $\mathcal{H}_{n,\Delta}(c, C)$,

$$\frac{D_{n,\Delta}(\beta)}{(\beta' B \beta)^{1/2}} = \sqrt{\frac{n_{\text{inf}}}{N}}\, \frac{\beta' G}{(\beta' B \beta)^{1/2}} + o_p(1).$$

By the bounded-frontier condition,

$$\sup_{\beta \neq 0} \left| \frac{\beta' G}{(\beta' B \beta)^{1/2}} \right| = (G' B^{-1} G)^{1/2} = O(1).$$

Since $n_{\text{inf}}/N \to \rho_{\text{inf}}$,

$$\frac{D_{n,\Delta}(\widehat{\beta}_n)}{(\widehat{\beta}_n' B \widehat{\beta}_n)^{1/2}} = \sqrt{\rho_{\text{inf}}}\, \frac{\widehat{\beta}_n' G}{(\widehat{\beta}_n' B \widehat{\beta}_n)^{1/2}} + o_p(1). \tag{A.118}$$

Let

$$\widehat{B} := \widehat{B}_{n,\Delta}^{\text{inf}}.$$

Variance consistency gives

$$\frac{\widehat{\beta}_n' \widehat{B} \widehat{\beta}_n}{\widehat{\beta}_n' B \widehat{\beta}_n} = 1 + o_p(1).$$

The score-weighted nuisance condition gives

$$\frac{\widehat{S}_n(\widehat{\beta}_n) - S_n^o(\widehat{\beta}_n)}{(\widehat{\beta}_n' B \widehat{\beta}_n)^{1/2}} = o_p(1),$$

where $\widehat{S}_n(\widehat{\beta}_n)$ denotes the feasible inference numerator using $(\widehat{\mu}_\Delta, \widehat{m}_\Delta)$. Hence

$$\begin{aligned} \widehat{T}_{n,\widehat{\beta}_n}(\theta_0) &= \frac{\widehat{S}_n(\widehat{\beta}_n)}{(\widehat{\beta}_n' \widehat{B} \widehat{\beta}_n)^{1/2}} \\ &= \frac{A_{n,\Delta}(\widehat{\beta}_n)}{(\widehat{\beta}_n' B \widehat{\beta}_n)^{1/2}} + \sqrt{\rho_{\inf}} \frac{\widehat{\beta}_n' G}{(\widehat{\beta}_n' B \widehat{\beta}_n)^{1/2}} + o_p(1). \end{aligned} \tag{A.119}$$

By the conditional scalar CLT in (5.13), applied at the training-measurable realized score $\widehat{\beta}_n$,

$$d_{\mathrm{BL}}\left(\mathcal{L}\left\{ \frac{A_{n,\Delta}(\widehat{\beta}_n)}{(\widehat{\beta}_n' B \widehat{\beta}_n)^{1/2}} \,\middle|\, \mathcal{F}_{\mathrm{tr}} \vee \mathcal{D}_{\inf} \right\}, N(0,1)\right) \xrightarrow{p} 0.$$

Taking conditional expectations over $\mathcal{D}_{\inf}$ transfers this approximation to conditioning only on $\mathcal{F}_{\mathrm{tr}}$. Indeed, for any f with $\|f\|_{\mathrm{BL}} \le 1$, the absolute difference between the $\mathcal{F}_{\mathrm{tr}}$-conditional expectation of f and its standard-normal expectation is bounded by the $\mathcal{F}_{\mathrm{tr}}$-conditional expectation of the preceding d_{BL} distance, which is $o_p(1)$ because the distance is bounded. Hence

$$d_{\mathrm{BL}}\left(\mathcal{L}\left\{ \frac{A_{n,\Delta}(\widehat{\beta}_n)}{(\widehat{\beta}_n' B \widehat{\beta}_n)^{1/2}} \,\middle|\, \mathcal{F}_{\mathrm{tr}} \right\}, N(0,1)\right) \xrightarrow{p} 0.$$

Combining this display with (A.119) yields

$$d_{\mathrm{BL}}\left(\mathcal{L}\{\widehat{T}_{n,\widehat{\beta}_n}(\theta_0) \mid \mathcal{F}_{\mathrm{tr}}\}, N\left[\sqrt{\rho_{\inf}} \frac{\widehat{\beta}_n' G_{n,\Delta}}{(\widehat{\beta}_n' B_{n,\Delta} \widehat{\beta}_n)^{1/2}}, 1\right]\right) \xrightarrow{p} 0,$$

which is (5.15).

The signed shift in the preceding display has square

$$\rho_{\inf} \frac{(\widehat{\beta}_n' G_{n,\Delta})^2}{\widehat{\beta}_n' B_{n,\Delta} \widehat{\beta}_n}.$$

The conditional version of Theorem 2.3 gives

$$\Pr_{N,\Delta}\{\widehat{AR}_{n,\widehat{\beta}_n} > c_{1-\alpha} \mid \mathcal{F}_{\mathrm{tr}}\} = h_\alpha\left(\rho_{\inf} \frac{(\widehat{\beta}_n' G_{n,\Delta})^2}{\widehat{\beta}_n' B_{n,\Delta} \widehat{\beta}_n}\right) + o_p(1).$$

Let $\varepsilon_{n,\Delta}$ denote the conditional approximation remainder. Since it is the difference of two conditional probabilities, $|\varepsilon_{n,\Delta}| \le 1$, and the preceding display gives $\varepsilon_{n,\Delta} \xrightarrow{p} 0$. Hence $\mathbb{E}|\varepsilon_{n,\Delta}| \to 0$. Taking expectations and using $\widehat{\beta}_n = \delta_n(\widehat{g}_n)$ gives

$$\Pr_{N,\Delta}\{\widehat{AR}_{n,\widehat{\beta}_n} > c_{1-\alpha}\} = \mathbb{E}_{N,\Delta}\left[h_\alpha\left(\rho_{\inf} \frac{\{\delta_n(\widehat{g}_n)' G_{n,\Delta}\}^2}{\delta_n(\widehat{g}_n)' B_{n,\Delta} \delta_n(\widehat{g}_n)}\right)\right] + o(1).$$

Under the null, $G_{n,0} = 0$, so the same argument gives

$$\Pr_{N,0}\{\widehat{AR}_{n,\widehat{\beta}_n} > c_{1-\alpha}\} = h_\alpha(0) + o(1) = \alpha + o(1).$$

Q.E.D.

A.12. *Proof of Theorem 5.12*

PROOF: Write $B := B_{n,\Delta},\ g := g_n,\ G := G_{n,\Delta} = \Delta g_n,\ \widetilde{\gamma} := B^{-1/2}g,\ r_n := \|\widetilde{\gamma}\|_2$. For any nonzero original coefficient β, set $\theta = B^{1/2}\beta$. Then $\beta' B\beta = \|\theta\|_2^2, \qquad \beta' g = \theta'\widetilde{\gamma}$. Thus

$$\frac{(\beta' g)^2}{\beta' B\beta} = \frac{(\theta'\widetilde{\gamma})^2}{\|\theta\|_2^2}. \tag{A.120}$$

Since $G = \Delta g$,

$$Q_{n,\Delta}^{\star,\mathrm{pliv}} = \sup_{\beta\neq 0} \rho_{\inf}\frac{(\beta' G)^2}{\beta' B\beta} = \rho_{\inf}\Delta^2\|\widetilde{\gamma}\|_2^2 = \rho_{\inf}\Delta^2 r_n^2.$$

Moreover,

$$L_{n,\Delta}^{\mathrm{pliv}}(\beta) = 1 - \frac{(\theta'\widetilde{\gamma})^2}{\|\theta\|_2^2\|\widetilde{\gamma}\|_2^2}.$$

Thus PLIV normalized noncentrality regret is exactly canonical projective angular loss in coordinates $(\theta,\widetilde{\gamma})$.

For the upper bound, use the normalized signal representation $\widetilde{Z}_{n,\Delta} = \widetilde{\gamma} + \sigma_{n,\Delta}\xi_n + e_{n,\Delta},\ \sigma_{n,\Delta}^2 := \rho_{\mathrm{tr}}^{-1}\sigma_\Delta^2$, where $\xi_n \sim N(0, I_{p_n})$. Let $\widehat{\theta}_n := H_{s_n}(\widetilde{Z}_{n,\Delta}),\ S := \mathrm{supp}(\widetilde{\gamma}),\ \widehat{S} := \mathrm{supp}(\widehat{\theta}_n),\ A := S\cup\widehat{S}$. Then $|A| \le 2s_n$. Since H_{s_n} is a best s_n-sparse Euclidean approximation, $\|\widetilde{Z}_{n,\Delta} - \widehat{\theta}_n\|_2^2 \le \|\widetilde{Z}_{n,\Delta} - \widetilde{\gamma}\|_2^2$. Put $h_n := \widehat{\theta}_n - \widetilde{\gamma},\ w_n := \sigma_{n,\Delta}\xi_n + e_{n,\Delta}$. Then $\|w_n - h_n\|_2^2 \le \|w_n\|_2^2$, and therefore $\|h_n\|_2^2 \le 2w_n' h_n = 2w_{n,A}' h_{n,A} \le 2\|w_{n,A}\|_2\|h_n\|_2$. Thus $\|h_n\|_2^2 \le 4\|w_{n,A}\|_2^2$. Also

$$\|w_{n,A}\|_2^2 \le 2\sigma_{n,\Delta}^2\|\xi_{n,A}\|_2^2 + 2\|e_{n,\Delta,A}\|_2^2 \le 2\sigma_{n,\Delta}^2 \max_{|J|\le 2s_n}\sum_{j\in J}\xi_{n,j}^2 + 2\|e_{n,\Delta}\|_2^2.$$

Consequently,

$$\|\widehat{\theta}_n - \widetilde{\gamma}\|_2^2 \le 8\sigma_{n,\Delta}^2 \max_{|J|\le 2s_n}\sum_{j\in J}\xi_{n,j}^2 + 8\|e_{n,\Delta}\|_2^2.$$

By the maximal chi-square bound in Online Appendix Lemma 2.5,

$$\mathbb{E}\max_{|J|\le 2s_n}\sum_{j\in J}\xi_{n,j}^2 \le C s_n \log(ep_n/s_n).$$

Hence, uniformly over $\widetilde{\gamma}\in\mathcal{G}_n(s_n, r_n)$,

$$\mathbb{E}\|\widehat{\theta}_n - \widetilde{\gamma}\|_2^2 \le C\rho_{\mathrm{tr}}^{-1}\sigma_\Delta^2 s_n\log(ep_n/s_n) + 8\mathbb{E}\|e_{n,\Delta}\|_2^2.$$

For every nonzero u,

$$1 - \frac{(u'\widetilde{\gamma})^2}{\|u\|_2^2\|\widetilde{\gamma}\|_2^2} \le \min\left\{1, \frac{\|u-\widetilde{\gamma}\|_2^2}{r_n^2}\right\}.$$

Therefore

$$\begin{aligned}\sup_{\widetilde{\gamma}\in\mathcal{G}_n(s_n,r_n)} \mathbb{E}L_{n,\Delta}^{\mathrm{pliv}}(\widehat{\theta}_n) &\le C\min\left\{1, \frac{\rho_{\mathrm{tr}}^{-1}\sigma_\Delta^2 s_n\log(ep_n/s_n)}{r_n^2}\right\} + 8\sup_{\widetilde{\gamma}}\frac{\mathbb{E}\|e_{n,\Delta}\|_2^2}{r_n^2}\\ &= C\min\{1,\eta_{n,\Delta}\} + o\{\min(1,\eta_{n,\Delta})\},\end{aligned}$$

where the last equality uses (5.32). Implementing this canonical rule in original PLIV coordinates by (5.29) gives the stated upper bound. The zero case uses the fixed default direction and has loss bounded by one, so it is absorbed by the displayed bound.

Now consider the lower bound. In the balanced canonical PLIV subclass of Definition 5.11, for every $\widetilde{\gamma}\in\mathcal{G}_n(s_n,r_n)$, $g_n=\widetilde{\gamma}$, $B_{n,\Delta}=I_{p_n}$, $C_n=\sigma_\Delta^2 I_{p_n}$. The training statistic satisfies

$$\begin{aligned}\widehat{g}_n&=\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_i\{N^{-1/2}b_i'\widetilde{\gamma}+V_i\}\\&=\left(\frac{1}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_ib_i'\right)\widetilde{\gamma}+\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_iV_i\\&=\widetilde{\gamma}+\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_iV_i.\end{aligned}$$

Because $V_i\overset{ind}{\sim}N(0,\sigma_\Delta^2)$ and $n_{\mathrm{tr}}^{-1}\sum_{i\in I_{\mathrm{tr}}}b_ib_i'=I_{p_n}$,

$$\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_iV_i\sim N\left(0,\frac{N}{n_{\mathrm{tr}}}\sigma_\Delta^2 I_{p_n}\right).$$

Since $N/n_{\mathrm{tr}}\to\rho_{\mathrm{tr}}^{-1}$, this is the canonical Gaussian sequence experiment with noise variance $\sigma_{n,\Delta}^2=(N/n_{\mathrm{tr}})\sigma_\Delta^2=\rho_{\mathrm{tr}}^{-1}\sigma_\Delta^2\{1+o(1)\}$. The inference residual in this subclass is $R_{i,\Delta}=\varepsilon_i$, so the scalar AR reduction is exact and the denominator is the identity. Apply the projective sparse lower-bound construction from Theorem 4.1 to the Gaussian experiment

$$\widehat{g}_n=\widetilde{\gamma}+\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sigma_\Delta\xi_n,\qquad \widetilde{\gamma}\in\mathcal{G}_n(s_n,r_n).$$

Equivalently, the proof uses the same packing $\widetilde{\gamma}_m=r_nv_m$, $m=1,\ldots,M$, with

$$\log M\ge c_1s_n\log(ep_n/s_n),\qquad 1-(v_m'v_k)^2\ge c_2\min\{1,\eta_{n,\Delta}\},\quad m\ne k,$$

and the KL bound changes only by the factor $N/n_{\mathrm{tr}}=\rho_{\mathrm{tr}}^{-1}\{1+o(1)\}$. Fano's inequality therefore gives, for every normalized-coordinate rule d_n,

$$\sup_{\widetilde{\gamma}\in\mathcal{G}_n(s_n,r_n)} E_{\widetilde{\gamma}}L(d_n,\widetilde{\gamma})\ge c\min\{1,\eta_{n,\Delta}\}-o(1).$$

Since the balanced canonical subclass is contained in $\mathfrak{P}_{n,\Delta}^{\mathrm{pliv}}(s_n,r_n)$,

$$\inf_{d_n}\sup_{P\in\mathfrak{P}_{n,\Delta}^{\mathrm{pliv}}(s_n,r_n)} E_PL_{n,\Delta}^{\mathrm{pliv}}(d_n)\ge c\min\{1,\eta_{n,\Delta}\}-o(1).$$

The total-variation alternative follows directly. Let P_γ^0 denote the canonical Gaussian law and P_γ^{pliv} the normalized PLIV law. If

$$\varepsilon_n:=\sup_{\gamma\in\mathcal{G}_n(s_n,r_n)}\|P_\gamma^{\mathrm{pliv}}-P_\gamma^0\|_{\mathrm{TV}}\to 0,$$

then, for every rule d_n and every bounded loss $0 \le L \le 1$,

$$\left| E_{P_\gamma^{\mathrm{pliv}}} L(d_n, \gamma) - E_{P_\gamma^0} L(d_n, \gamma) \right| \le \varepsilon_n.$$

Taking suprema over γ, then infima over d_n, transfers the canonical lower bound with an $o(1)$ error. Finally suppose $Q_{n,\Delta}^{\star,\mathrm{pliv}} = \rho_{\mathrm{inf}} \Delta^2 r_n^2 \in [\underline{q}, \overline{q}] \subset (0, \infty)$. For any rule, $Q_{n,\Delta}^{\mathrm{pliv}} = Q_{n,\Delta}^{\star,\mathrm{pliv}} \{1 - L_{n,\Delta}^{\mathrm{pliv}}\}$. Lemma 4.3 gives constants $0 < a_\alpha < A_\alpha < \infty$ such that

$$a_\alpha L_{n,\Delta}^{\mathrm{pliv}} \le h_\alpha(Q_{n,\Delta}^{\star,\mathrm{pliv}}) - h_\alpha(Q_{n,\Delta}^{\mathrm{pliv}}) \le A_\alpha L_{n,\Delta}^{\mathrm{pliv}}.$$

Taking expectations, suprema, and infima converts the normalized noncentrality-regret frontier into the stated PLIV power-regret frontier. *Q.E.D.*

A.13. *Proof of Corollary 5.13*

PROOF: Let $q_{n,\Delta}^\star := Q_{n,\Delta}^{\star,\mathrm{pliv}} = \rho_{\mathrm{inf}} \Delta^2 r_n^2$. By assumption, $0 < \underline{q} \le q_{n,\Delta}^\star \le \overline{q} < \infty$. Since $\Delta \neq 0$ is fixed and the inference split is nondegenerate, there exists $\underline{\rho} > 0$ such that $\rho_{\mathrm{inf}} \ge \underline{\rho}$ eventually. Hence

$$r_n^2 = \frac{q_{n,\Delta}^\star}{\rho_{\mathrm{inf}} \Delta^2} \le \frac{\overline{q}}{\underline{\rho} \Delta^2}$$

eventually. Also the training split is nondegenerate, so $\rho_{\mathrm{tr}}^{-1} \ge c_\rho > 0$ eventually. If $\sigma_\Delta^2 \ge \underline{\sigma}^2 > 0$, then

$$\begin{aligned} \eta_{n,\Delta} &= \frac{\rho_{\mathrm{tr}}^{-1} \sigma_\Delta^2 s_n \log(ep_n/s_n)}{r_n^2} \\ &\ge c_\rho \underline{\sigma}^2 \frac{s_n \log(ep_n/s_n)}{r_n^2} \\ &\ge c s_n \log(ep_n/s_n) \longrightarrow \infty \end{aligned}$$

for some $c > 0$. Hence $\eta_{n,\Delta} \ge 1$ eventually. The lower bound in Theorem 5.12, together with the compact-range power-regret transfer, implies

$$\liminf_{n\to\infty} \inf_{d_n} \sup_{P \in \mathfrak{P}_{n,\Delta}^{\mathrm{pliv}}(s_n, r_n)} \mathcal{R}_{\alpha,n,\Delta}^{\mathrm{pliv}}(d_n) \ge c_\alpha > 0.$$

This is the stated first-order oracle-power nonrecoverability. *Q.E.D.*

SUPPLEMENT TO "FEASIBLE POWER FRONTIERS FOR LEARNED SCALAR ANDERSON–RUBIN SCORES UNDER HIGH-DIMENSIONAL WEAK IDENTIFICATION"

TAMER ÇETIN

1. PURPOSE OF THIS APPENDIX

This Online Appendix gives the full technical derivations and auxiliary results for "Feasible Power Frontiers for Learned Scalar Anderson–Rubin Scores under High-Dimensional Weak Identification." The main text is organized around one finite-information frontier for learned scalar AR power: the sparse difficulty index $\sigma_n^2 s_n \log(ep_n/s_n)/r_n^2$ governs normalized noncentrality regret and, on compact oracle-noncentrality ranges, power regret. The first technical section gives the canonical sparse projective minimax proof and the corresponding power-regret transfer. The next section gives the finite-grid direction-free and restricted-inversion proofs. The following section gives the primitive PLIV verification details used by the operational PLIV reduction and the PLIV learned-score frontier in the main text. The final section records supplementary numerical details.

For a nonzero score direction β and nonzero drift g, define normalized noncentrality attainment and regret by

$$A(\beta, g) = \frac{(\beta' g)^2}{\|\beta\|^2 \|g\|^2}, \qquad L(\beta, g) = 1 - A(\beta, g).$$

If $\beta = 0$, set $L(\beta, g) = 1$. Thus L is squared projective angular loss between the selected score direction and the drift direction, with β and $-\beta$ equivalent.

Let

$$\mathcal{G}_p(s, r) = \{g \in \mathbb{R}^p : \|g\|_0 \le s,\ \|g\|_2 = r\}, \qquad 2 \le s \le p/4.$$

The organization is as follows. Theorem 2.1 is the canonical proof companion to Theorem 4.1. Lemmas 2.2–2.5 provide the packing, information, and thresholding ingredients. Corollary 2.7 is the canonical proof companion to Corollary 4.4. The remaining sections prove the direction-free and PLIV primitive verification results stated in the main text.

2. SPARSE MINIMAX FRONTIER

This section gives the full canonical-coordinate derivation underlying Theorem 4.1 in the main text. Subsections 2.1–2.4 prove the sparse minimax envelope. Subsection 2.5 gives the corresponding canonical power-regret transfer underlying Corollary 4.4. The argument is written in the Gaussian sequence coordinates used in Section 4 of the main paper.

THEOREM 2.1—Canonical sparse minimax envelope: *Theorem 2.1 is the canonical-coordinate version of Theorem 4.1; the notation in this appendix suppresses the sequence index n.*

Consider the Gaussian sequence experiment

$$Y = g + \sigma\xi, \qquad \xi \sim N(0, I_p),$$

with $g \in \mathcal{G}_p(s, r)$, $2 \le s \le p/4$. Let

$$\eta(p, s, r, \sigma) = \frac{\sigma^2 s \log(ep/s)}{r^2}.$$

There exist universal constants $0 < c < C < \infty$ *such that*

$$c \min\{1, \eta(p,s,r,\sigma)\} \le \inf_{\delta} \sup_{g \in \mathcal{G}_p(s,r)} \mathbb{E}_g L\{\delta(Y), g\} \le C \min\{1, \eta(p,s,r,\sigma)\},$$

where the infimum is over all measurable rules $\delta : \mathbb{R}^p \to \mathbb{R}^p$.

2.1. *Auxiliary combinatorial and information inequalities*

LEMMA 2.2—Constant-weight Varshamov–Gilbert packing: *There exist universal constants* $c_0, c_1 > 0$ *such that, for integers* $1 \le m \le q/4$, *there are subsets* $S_1, \ldots, S_M \subset \{1, \ldots, q\}$, *each of cardinality* m, *satisfying*

$$|S_j \triangle S_k| \ge c_0 m \qquad \text{for all } j \ne k, \tag{2.1}$$

and

$$\log M \ge c_1 m \log(eq/m). \tag{2.2}$$

PROOF: Let

$$r_q := \lfloor q/m \rfloor. \tag{2.3}$$

Since $q/m \ge 4$,

$$r_q \ge q/m - 1 \ge \frac{3}{4}(q/m), \qquad r_q \ge 4. \tag{2.4}$$

Partition the first mr_q elements of $\{1, \ldots, q\}$ into m disjoint blocks

$$B_1, \ldots, B_m, \qquad |B_a| = r_q.$$

For each codeword $u = (u_1, \ldots, u_m) \in \{1, \ldots, r_q\}^m$, define

$$S(u) := \{\text{the } u_a\text{-th element of block } B_a : a = 1, \ldots, m\}. \tag{2.5}$$

Then

$$|S(u)| = m. \tag{2.6}$$

For $u, v \in \{1, \ldots, r_q\}^m$, let

$$d_H(u,v) := \#\{a : u_a \ne v_a\}.$$

Then

$$|S(u) \triangle S(v)| = 2 d_H(u,v). \tag{2.7}$$

Let

$$h := \lfloor m/16 \rfloor. \tag{2.8}$$

Choose a maximal subset $\mathcal{C} \subset \{1, \ldots, r_q\}^m$ satisfying

$$d_H(u,v) > h \qquad \text{for all distinct } u, v \in \mathcal{C}. \tag{2.9}$$

Maximality gives the covering inequality

$$|\mathcal{C}|\, V_{m,r_q}(h) \ge r_q^m, \tag{2.10}$$

where

$$V_{m,r_q}(h) := \sum_{a=0}^{h} \binom{m}{a} (r_q - 1)^a. \tag{2.11}$$

Since $h \le m/16$,

$$V_{m,r_q}(h) \le \left(\sum_{a=0}^{m} \binom{m}{a} \right) r_q^h \le 2^m r_q^{m/16}. \tag{2.12}$$

Equations (2.10)–(2.12) imply

$$|\mathcal{C}| \ge 2^{-m} r_q^{15m/16}. \tag{2.13}$$

Taking logarithms and using (2.4),

$$\begin{aligned} \log |\mathcal{C}| &\ge \frac{15m}{16} \log r_q - m \log 2 \\ &\ge m \left[\frac{15}{16} \{\log(q/m) - \log(4/3)\} - \log 2 \right]. \end{aligned} \tag{2.14}$$

Set $x = q/m$. Since $x \ge 4$, define

$$c_1 := \inf_{x \ge 4} \frac{\frac{15}{16}\{\log x - \log(4/3)\} - \log 2}{\log(ex)}. \tag{2.15}$$

The numerator in (2.15) is positive at $x = 4$, because

$$\frac{15}{16}\{\log 4 - \log(4/3)\} - \log 2 = \frac{15}{16} \log 3 - \log 2 > 0,$$

and the ratio has positive limit $15/16$ as $x \to \infty$. Hence $c_1 > 0$, and (2.14) gives

$$\log |\mathcal{C}| \ge c_1 m \log(eq/m). \tag{2.16}$$

Enumerate $\mathcal{C} = \{u^1, \dots, u^M\}$, and set $S_j = S(u^j)$. For distinct j, k, (2.9) gives

$$d_H(u^j, u^k) \ge h + 1.$$

If $m \ge 16$, then $h + 1 > m/16$, so (2.7) gives

$$|S_j \triangle S_k| > m/8.$$

If $1 \le m < 16$, then $h = 0$, so $d_H(u^j, u^k) \ge 1$ and

$$|S_j \triangle S_k| \ge 2 \ge m/8.$$

Thus (2.1) holds with $c_0 = 1/8$. Together with (2.16), this proves the lemma. *Q.E.D.*

LEMMA 2.3—Fano inequality in the form used below: *Let $\{P_1, \dots, P_M\}$ be probability measures and let J be uniform on $\{1, \dots, M\}$. Conditional on $J = m$, let $Y \sim P_m$. If*

$$\frac{1}{M^2} \sum_{j,k=1}^{M} \mathrm{KL}(P_j, P_k) \le a \log M \tag{2.17}$$

for some $a < 1$*, then, for every estimator* $\widehat{J}(Y)$*,*

$$\sup_{1 \le m \le M} P_m(\widehat{J} \ne m) \ge 1 - a - \frac{1}{\log M}. \tag{2.18}$$

In particular, if $a \le 1/4$ *and* $M \ge e^4$*, the right-hand side is at least* $1/2$*.*

PROOF: Let

$$\bar{P} := \frac{1}{M} \sum_{k=1}^{M} P_k.$$

The mutual information satisfies

$$I(J;Y) = \frac{1}{M} \sum_{j=1}^{M} \mathrm{KL}(P_j, \bar{P}). \tag{2.19}$$

By convexity of $Q \mapsto \mathrm{KL}(P_j, Q)$,

$$\mathrm{KL}(P_j, \bar{P}) = \mathrm{KL}\left(P_j, \frac{1}{M} \sum_{k=1}^{M} P_k\right) \le \frac{1}{M} \sum_{k=1}^{M} \mathrm{KL}(P_j, P_k). \tag{2.20}$$

Equations (2.19) and (2.20) give

$$I(J;Y) \le \frac{1}{M^2} \sum_{j,k=1}^{M} \mathrm{KL}(P_j, P_k) \le a \log M. \tag{2.21}$$

For any estimator $\widehat{J}$, let $P_e = \Pr(\widehat{J} \ne J)$. Then

$$\begin{aligned} \log M - I(J;Y) &= H(J \mid Y) \\ &\le H(J \mid \widehat{J}) \\ &\le \log 2 + P_e \log(M-1) \\ &\le 1 + P_e \log M. \end{aligned} \tag{2.22}$$

Combining (2.21) and (2.22),

$$\log M - a \log M \le 1 + P_e \log M,$$

so

$$P_e \ge 1 - a - \frac{1}{\log M}. \tag{2.23}$$

Finally,

$$\sup_m P_m(\widehat{J} \ne m) \ge \frac{1}{M} \sum_{m=1}^{M} P_m(\widehat{J} \ne m) = P_e. \tag{2.24}$$

Equations (2.23)–(2.24) prove (2.18). If $a \le 1/4$ and $M \ge e^4$, then

$$1 - a - 1/\log M \ge 1 - 1/4 - 1/4 = 1/2.$$

Q.E.D.

2.2. *Projective sparse packing*

LEMMA 2.4—Projective sparse perturbation packing: *There exist universal constants $c_0, c_1, c_2 > 0$ such that the following holds. Let $2 \le s \le p/4$, $r > 0$, $\sigma > 0$, and set*

$$\eta = \frac{\sigma^2 s \log(ep/s)}{r^2}, \qquad \theta^2 = c_0 \min\{1, \eta\},$$

where c_0 is chosen sufficiently small. Then there are vectors $v^1, \dots, v^M \in \mathbb{R}^p$ satisfying

$$\|v^m\|_0 \le s, \qquad \|v^m\|_2 = 1, \qquad \log M \ge c_1 s \log(ep/s),$$

$$1 - (v^{m\prime} v^k)^2 \ge c_2 \min\{1, \eta\} \qquad \text{for all } m \neq k, \tag{2.25}$$

and, with $P_m = N(rv^m, \sigma^2 I_p)$,

$$\frac{1}{M^2} \sum_{m,k=1}^{M} \mathrm{KL}(P_m, P_k) \le \frac{1}{4} \log M. \tag{2.26}$$

PROOF: Apply Lemma 2.2 with

$$q = p - 1, \qquad m = s - 1.$$

Because $2 \le s \le p/4$, $1 \le m \le q/4$ after changing only universal constants. There exist sets $S_1, \dots, S_M \subset \{2, \dots, p\}$, each of cardinality $m = s - 1$, such that

$$|S_j \triangle S_k| \ge c_S m \qquad (j \neq k), \tag{2.27}$$

and

$$\log M \ge c_M s \log(ep/s) \tag{2.28}$$

for universal constants $c_S, c_M > 0$. Take $c_0 \le 1/16$, so

$$0 < \theta^2 \le 1/16. \tag{2.29}$$

Define

$$v^j := (1 - \theta^2)^{1/2} e_1 + \frac{\theta}{\sqrt{m}} \sum_{\ell \in S_j} e_\ell. \tag{2.30}$$

Then

$$\|v^j\|_0 = s, \qquad \|v^j\|_2^2 = (1 - \theta^2) + \theta^2 = 1. \tag{2.31}$$

For $j \neq k$, let $d_{jk} := |S_j \triangle S_k|$. Since $|S_j \cap S_k| = m - d_{jk}/2$,

$$\begin{aligned} v^{j\prime} v^k &= (1 - \theta^2) + \frac{\theta^2}{m} |S_j \cap S_k| \\ &= 1 - \theta^2 + \theta^2 \left(1 - \frac{d_{jk}}{2m}\right) \\ &= 1 - \theta^2 \frac{d_{jk}}{2m}. \end{aligned} \tag{2.32}$$

Set

$$x_{jk} := \theta^2 \frac{d_{jk}}{2m}.$$

By (2.29), $0 \le x_{jk} \le \theta^2 \le 1/16$. Hence

$$\begin{aligned} 1 - (v^{j\prime} v^k)^2 &= 1 - (1 - x_{jk})^2 \\ &= 2x_{jk} - x_{jk}^2 \\ &\ge x_{jk} = \theta^2 \frac{d_{jk}}{2m} \ge \frac{c_S}{2}\theta^2. \end{aligned} \tag{2.33}$$

Since $\theta^2 = c_0 \min\{1, \eta\}$, (2.33) gives (2.25) with $c_2 = c_S c_0/2$.

For the KL bound,

$$\|v^j - v^k\|_2^2 = \frac{\theta^2}{m} |S_j \triangle S_k| \le 2\theta^2. \tag{2.34}$$

Therefore, for $P_j = N(rv^j, \sigma^2 I_p)$,

$$\begin{aligned} \mathrm{KL}(P_j, P_k) &= \frac{r^2 \|v^j - v^k\|_2^2}{2\sigma^2} \\ &\le \frac{r^2\theta^2}{\sigma^2} \\ &= c_0 \frac{r^2}{\sigma^2} \min\{1, \eta\}. \end{aligned} \tag{2.35}$$

If $\eta \le 1$, then

$$\frac{r^2}{\sigma^2}\eta = s \log(ep/s). \tag{2.36}$$

If $\eta > 1$, then

$$\frac{r^2}{\sigma^2} < s \log(ep/s). \tag{2.37}$$

Thus (2.35)–(2.37) imply

$$\mathrm{KL}(P_j, P_k) \le c_0 s \log(ep/s). \tag{2.38}$$

Choosing $c_0 \le c_M/4$ and using (2.28),

$$\mathrm{KL}(P_j, P_k) \le \frac{1}{4} \log M.$$

Averaging over (j, k) gives (2.26). *Q.E.D.*

2.3. *Hard-thresholding upper bound*

LEMMA 2.5—Hard-thresholding oracle inequality: *Let $H_s(Y)$ retain the s largest coordinates of Y in absolute value and set all others to zero. If $Y = g + \sigma\xi$, $\xi \sim N(0, I_p)$, and $\|g\|_0 \le s$, then*

$$\mathbb{E}_g \|H_s(Y) - g\|^2 \le C\sigma^2 s \log(ep/s) \tag{2.39}$$

for a universal constant $C < \infty$.

PROOF: Let $\widehat{g} := H_s(Y)$, $S := \text{supp}(g)$, and $\widehat{S} := \text{supp}(\widehat{g})$. Since $\widehat{g}$ is a best s-sparse Euclidean approximation to Y,

$$\|Y - \widehat{g}\|^2 \le \|Y - g\|^2. \tag{2.40}$$

Let $h := \widehat{g} - g$. Since $Y = g + \sigma\xi$, (2.40) gives

$$\|\sigma\xi - h\|^2 \le \|\sigma\xi\|^2. \tag{2.41}$$

Expanding (2.41),

$$\|h\|^2 \le 2\sigma\xi' h. \tag{2.42}$$

The support of h is contained in $S \cup \widehat{S}$, and

$$|S \cup \widehat{S}| \le 2s.$$

Hence

$$\xi' h = \xi'_{S\cup\widehat{S}} h_{S\cup\widehat{S}} \le \|\xi_{S\cup\widehat{S}}\| \|h\|. \tag{2.43}$$

Equations (2.42)–(2.43) imply

$$\|H_s(Y) - g\|^2 = \|h\|^2 \le 4\sigma^2 \|\xi_{S\cup\widehat{S}}\|^2. \tag{2.44}$$

Therefore

$$\|H_s(Y) - g\|^2 \le 4\sigma^2 W_{p,s}, \qquad W_{p,s} := \max_{|A|\le 2s} \sum_{j\in A} \xi_j^2. \tag{2.45}$$

For $0 < \lambda < 1/2$,

$$\begin{aligned} \mathbb{E} e^{\lambda W_{p,s}} &\le \sum_{m=0}^{2s} \sum_{|A|=m} \mathbb{E} \exp\left(\lambda \sum_{j\in A} \xi_j^2\right) \\ &= \sum_{m=0}^{2s} \binom{p}{m} (1 - 2\lambda)^{-m/2}. \end{aligned} \tag{2.46}$$

With $\lambda = 1/4$,

$$\begin{aligned} \mathbb{E} e^{W_{p,s}/4} &\le \sum_{m=0}^{2s} \binom{p}{m} 2^{m/2} \\ &\le (2s+1) \binom{p}{2s} 2^s \\ &\le (2s+1) \left(\frac{ep}{2s}\right)^{2s} 2^s. \end{aligned} \tag{2.47}$$

Jensen's inequality gives

$$e^{\mathbb{E} W_{p,s}/4} \le \mathbb{E} e^{W_{p,s}/4}, \qquad \mathbb{E} W_{p,s} \le 4 \log \mathbb{E} e^{W_{p,s}/4}.$$

Thus, using $s \le p/4$,

$$\begin{aligned}\mathbb{E}W_{p,s} &\le 4\log(2s+1) + 8s\log\left(\frac{ep}{2s}\right) + 4s\log 2 \\ &\le Cs\log(ep/s).\end{aligned} \tag{2.48}$$

Taking expectations in (2.45) and using (2.48) proves (2.39). *Q.E.D.*

2.4. *Proof of Theorem 2.1*

PROOF: Set

$$\eta = \frac{\sigma^2 s\log(ep/s)}{r^2}.$$

Upper bound. For any nonzero w and nonzero g,

$$\begin{aligned}L(w,g) &= 1 - \frac{(w'g)^2}{\|w\|^2\|g\|^2} \\ &= \min_{a\in\mathbb{R}}\left\|\frac{g}{\|g\|} - aw\right\|^2 \\ &\le \left\|\frac{g}{\|g\|} - \frac{w}{\|g\|}\right\|^2 = \frac{\|w-g\|^2}{\|g\|^2}.\end{aligned} \tag{2.49}$$

If $w = 0$, then $L(0,g) = 1 = \|0-g\|^2/\|g\|^2$. Thus (2.49) holds for all w. For $g \in \mathcal{G}_p(s,r)$, $\|g\| = r$, so

$$L\{H_s(Y), g\} \le \min\left\{1, \frac{\|H_s(Y)-g\|^2}{r^2}\right\}. \tag{2.50}$$

Using Lemma 2.5,

$$\begin{aligned}\sup_{g\in\mathcal{G}_p(s,r)} \mathbb{E}_g L\{H_s(Y), g\} &\le \min\left\{1, \frac{1}{r^2}\sup_{g\in\mathcal{G}_p(s,r)}\mathbb{E}_g\|H_s(Y)-g\|^2\right\} \\ &\le C\min\left\{1, \frac{\sigma^2 s\log(ep/s)}{r^2}\right\} \\ &= C\min\{1,\eta\}.\end{aligned} \tag{2.51}$$

This gives the upper bound.

Lower bound. Use Lemma 2.4. Let $v^1, \ldots, v^M$ be the packing, set

$$g^m := rv^m, \qquad P_m := N(g^m, \sigma^2 I_p), \tag{2.52}$$

and write

$$\Delta_p := c_2\min\{1,\eta\}. \tag{2.53}$$

Then

$$1-(v^{m\prime}v^k)^2 \geq \Delta_p \qquad (m \neq k), \tag{2.54}$$

and

$$\frac{1}{M^2}\sum_{m,k=1}^{M} \mathrm{KL}(P_m, P_k) \leq \frac{1}{4}\log M. \tag{2.55}$$

Let J be uniform on $\{1,\ldots,M\}$, and let $Y \mid J=m \sim P_m$. By Lemma 2.3, after reducing constants if necessary,

$$\inf_{\widehat{J}} \sup_{1\leq m\leq M} P_m(\widehat{J}\neq m) \geq c_F \tag{2.56}$$

for a universal constant $c_F>0$. If the constructed packing has $M<e^4$, then $s\log(ep/s)$ is bounded by a universal constant in the present construction. In that finite-entropy case, the same lower rate follows, after reducing the universal constant, from a two-point Le Cam argument applied to two separated sparse directions with KL divergence bounded by a sufficiently small constant. Thus the lower bound holds uniformly over the stated range $2\leq s\leq p/4$. For $u,v\neq 0$, define

$$a(u,v) := \arccos\left(\frac{|u'v|}{\|u\|\|v\|}\right) \in [0,\pi/2]. \tag{2.57}$$

Let

$$a_p := \arcsin(\sqrt{\Delta_p}). \tag{2.58}$$

By (2.54),

$$\sin^2 a(v^m, v^k) \geq \Delta_p, \qquad a(v^m,v^k)\geq a_p \qquad (m\neq k). \tag{2.59}$$

For $0\leq x\leq 1$,

$$\begin{aligned}\sin^2\left\{\frac{1}{2}\arcsin(\sqrt{x})\right\} &= \frac{1-\cos\{\arcsin(\sqrt{x})\}}{2}\\ &= \frac{1-\sqrt{1-x}}{2} = \frac{x}{2(1+\sqrt{1-x})} \geq \frac{x}{4}.\end{aligned} \tag{2.60}$$

Taking $x=\Delta_p$,

$$\sin^2(a_p/2) \geq \Delta_p/4. \tag{2.61}$$

For an arbitrary score rule $\delta(Y)$, define $\widehat{m}_\delta(Y)$ as follows. On the event $\{\delta(Y)\neq 0\}$, set

$$\widehat{m}_\delta(Y) \in \operatorname*{arg\,min}_{1\leq k\leq M} a\{\delta(Y), v^k\}, \tag{2.62}$$

with arbitrary tie-breaking. On the event $\{\delta(Y)=0\}$, assign $\widehat{m}_\delta(Y)$ an arbitrary value in $\{1,\ldots,M\}$.

Fix m. If

$$L\{\delta(Y), g^m\} < \sin^2(a_p/2), \tag{2.63}$$

then $\delta(Y)\neq 0$, because $L\{0,g^m\}=1$ and $\sin^2(a_p/2)<1$. Therefore, on the event in (2.63),

$$L\{\delta(Y), g^m\} = \sin^2 a\{\delta(Y), v^m\}. \tag{2.64}$$

Thus

$$a\{\delta(Y), v^m\} < a_p/2. \tag{2.65}$$

For any $k \neq m$, the projective triangle inequality gives

$$\begin{aligned} a\{\delta(Y), v^k\} &\geq a(v^m, v^k) - a\{\delta(Y), v^m\} \\ &> a_p - a_p/2 = a_p/2. \end{aligned} \tag{2.66}$$

Hence $\widehat{m}_\delta(Y) = m$. Consequently

$$\{\widehat{m}_\delta(Y) \neq m\} \subset \left\{L\{\delta(Y), g^m\} \geq \sin^2(a_p/2)\right\}. \tag{2.67}$$

Therefore

$$\begin{aligned} \mathbb{E}_m L\{\delta(Y), g^m\} &\geq \sin^2(a_p/2) P_m\{\widehat{m}_\delta(Y) \neq m\} \\ &\geq \frac{\Delta_p}{4} P_m\{\widehat{m}_\delta(Y) \neq m\}, \end{aligned} \tag{2.68}$$

where the second inequality uses (2.61). Taking suprema over m, then infima over δ, and using (2.56),

$$\begin{aligned} \inf_\delta \sup_{g \in \mathcal{G}_p(s,r)} \mathbb{E}_g L\{\delta(Y), g\} &\geq \inf_\delta \sup_{1 \leq m \leq M} \mathbb{E}_m L\{\delta(Y), g^m\} \\ &\geq \frac{\Delta_p}{4} \inf_{\widehat{J}} \sup_{1 \leq m \leq M} P_m(\widehat{J} \neq m) \\ &\geq \frac{c_F \Delta_p}{4} \geq c \min\{1, \eta\}. \end{aligned} \tag{2.69}$$

The lower and upper bounds (2.51) and (2.69) prove Theorem 2.1. *Q.E.D.*

2.5. *Power-regret transfer for Corollary 4.4*

LEMMA 2.6—Canonical power-map comparability: *Let*

$$h_\alpha(q) = 1 - F_{\chi_1^2(q)}(c_{1-\alpha}).$$

For every $0 < \underline{q} < \overline{q} < \infty$*, there exist constants* $0 < c < C < \infty$*, depending only on* $\alpha, \underline{q}, \overline{q}$*, such that, for every* $q \in [\underline{q}, \overline{q}]$ *and every* $\ell \in [0, 1]$*,*

$$c\ell \leq h_\alpha(q) - h_\alpha\{q(1-\ell)\} \leq C\ell. \tag{2.70}$$

PROOF: Let $c_\alpha = c_{1-\alpha}$, $x = \sqrt{c_\alpha}$, and $Z \sim N(0,1)$. For $q \geq 0$,

$$h_\alpha(q) = \Pr\{(Z + \sqrt{q})^2 > c_\alpha\} = 1 - \Phi(x - \sqrt{q}) + \Phi(-x - \sqrt{q}). \tag{2.71}$$

For $q > 0$,

$$h'_\alpha(q) = \frac{\phi(x - \sqrt{q}) - \phi(x + \sqrt{q})}{2\sqrt{q}}. \tag{2.72}$$

For the right derivative at zero, put $s=\sqrt{q}$. Then

$$\begin{aligned}\lim_{q\downarrow 0} h'_\alpha(q) &= \lim_{s\downarrow 0}\frac{\phi(x-s)-\phi(x+s)}{2s}\\ &= \lim_{s\downarrow 0}\frac{\{\phi(x)-s\phi'(x)+o(s)\}-\{\phi(x)+s\phi'(x)+o(s)\}}{2s}\\ &= -\phi'(x)=x\phi(x)>0.\end{aligned} \tag{2.73}$$

Set

$$h'_\alpha(0):=x\phi(x). \tag{2.74}$$

For $q>0$, $x+\sqrt{q}>|x-\sqrt{q}|$, so

$$\phi(x-\sqrt{q})>\phi(x+\sqrt{q}),$$

and hence $h'_\alpha(q)>0$. Thus h'_α extends continuously to zero and is strictly positive on $[0,\overline{q}]$. Therefore there exist $0<m<M<\infty$, depending only on α and $\overline{q}$, such that

$$m\le h'_\alpha(u)\le M \qquad \forall u\in[0,\overline{q}]. \tag{2.75}$$

For $q\in[\underline{q},\overline{q}]$ and $\ell\in[0,1]$,

$$h_\alpha(q)-h_\alpha\{q(1-\ell)\}=\int_{q(1-\ell)}^{q} h'_\alpha(u)\,du. \tag{2.76}$$

Using (2.75) in (2.76),

$$mq\ell\le h_\alpha(q)-h_\alpha\{q(1-\ell)\}\le Mq\ell. \tag{2.77}$$

Since $q\in[\underline{q},\overline{q}]$, (2.77) implies (2.70) with constants $c=m\underline{q}$ and $C=M\overline{q}$. *Q.E.D.*

COROLLARY 2.7—Canonical power-regret envelope; supplement to Corollary 4.4: *Let*

$$q^\star=\rho_{\inf}r^2.$$

Suppose

$$q^\star\in[\underline{q},\overline{q}]\subset(0,\infty).$$

Then there exist constants $0<c_\alpha<C_\alpha<\infty$, depending only on $\alpha,\underline{q},\overline{q}$, such that

$$\begin{aligned}c_\alpha\min\{1,\eta(p,s,r,\sigma)\} &\le \inf_\delta \sup_{g\in\mathcal{G}_p(s,r)}[h_\alpha(q^\star)-\mathbb{E}_g h_\alpha\{q^\star A(\delta(Y),g)\}]\\ &\le C_\alpha\min\{1,\eta(p,s,r,\sigma)\}.\end{aligned} \tag{2.78}$$

PROOF: For $g\in\mathcal{G}_p(s,r)$,

$$q^\star=\rho_{\inf}r^2. \tag{2.79}$$

For any rule δ,

$$A\{\delta(Y),g\}=1-L\{\delta(Y),g\}. \tag{2.80}$$

Thus

$$h_\alpha(q^\star) - h_\alpha\{q^\star A(\delta(Y), g)\} = h_\alpha(q^\star) - h_\alpha\left(q^\star[1 - L\{\delta(Y), g\}]\right). \tag{2.81}$$

By assumption, $q^\star \in [\underline{q}, \overline{q}]$. Applying Lemma 2.6 with $q = q^\star$ and $\ell = L\{\delta(Y), g\}$ gives

$$cL\{\delta(Y), g\} \le h_\alpha(q^\star) - h_\alpha\{q^\star A(\delta(Y), g)\} \le CL\{\delta(Y), g\}. \tag{2.82}$$

Taking expectations,

$$c\mathbb{E}_g L\{\delta(Y), g\} \le h_\alpha(q^\star) - \mathbb{E}_g h_\alpha\{q^\star A(\delta(Y), g)\} \le C\mathbb{E}_g L\{\delta(Y), g\}. \tag{2.83}$$

Taking suprema over $g \in \mathcal{G}_p(s, r)$, then infima over δ, gives

$$\begin{aligned} c \inf_\delta \sup_{g \in \mathcal{G}_p(s,r)} \mathbb{E}_g L\{\delta(Y), g\} &\le \inf_\delta \sup_{g \in \mathcal{G}_p(s,r)} \left[h_\alpha(q^\star) - \mathbb{E}_g h_\alpha\{q^\star A(\delta(Y), g)\}\right] \\ &\le C \inf_\delta \sup_{g \in \mathcal{G}_p(s,r)} \mathbb{E}_g L\{\delta(Y), g\}. \end{aligned} \tag{2.84}$$

Theorem 2.1 and (2.84) imply (2.78). *Q.E.D.*

3. DIRECTION-FREE AND INVERSION PROOFS

This section supplies the auxiliary proofs for the finite-grid direction-free results in Section 6 of the main text. Subsection 3.1 proves Theorem 6.1; Subsection 3.2 proves Corollary 6.2.

3.1. *Proof of Theorem 6.1*

PROOF: Let $K_n \subset \mathbb{R}^{p_n}$ be the common compact action set. For $\ell = 1, \dots, L$, write

$$dP_{n\ell}(y) = f_{n\ell}(y)\, d\nu_n(y), \qquad f_{n\ell} \ge 0, \qquad \int f_{n\ell}(y)\, d\nu_n(y) = 1.$$

The densities are taken as fixed measurable versions.

For $\beta \in K_n$, define

$$q_{n\ell}(\beta) := Q_{n\ell}(\beta; G_{n\ell}, B_{n\ell}) = \rho_{\inf} \frac{(\beta' G_{n\ell})^2}{\beta' B_{n\ell} \beta}.$$

Since $B_{n\ell}$ is positive definite and K_n is contained in the nonzero score directions,

$$\beta' B_{n\ell} \beta > 0 \qquad \forall \beta \in K_n.$$

Therefore

$$\beta_m \to \beta \in K_n \quad \Longrightarrow \quad q_{n\ell}(\beta_m) \to q_{n\ell}(\beta), \qquad \ell = 1, \dots, L. \tag{3.1}$$

Since h_α is continuous,

$$\beta_m \to \beta \in K_n \quad \Longrightarrow \quad h_\alpha\{q_{n\ell}(\beta_m)\} \to h_\alpha\{q_{n\ell}(\beta)\}. \tag{3.2}$$

Also

$$0 \le h_\alpha\{q_{n\ell}(\beta)\} \le 1, \qquad \forall \beta \in K_n, \quad \ell = 1, \dots, L. \tag{3.3}$$

For any measurable rule $\delta_n : \mathsf{Y}_n \to K_n$,

$$\begin{aligned}\mathcal{P}_{w,n}(\delta_n) &= \sum_{\ell=1}^{L} w_\ell \mathbb{E}_{n\ell}\left[h_\alpha\{q_{n\ell}(\delta_n(Y_n))\}\right] \\ &= \sum_{\ell=1}^{L} w_\ell \int h_\alpha\{q_{n\ell}(\delta_n(y))\} f_{n\ell}(y)\, d\nu_n(y) \\ &= \int \sum_{\ell=1}^{L} w_\ell f_{n\ell}(y) h_\alpha\{q_{n\ell}(\delta_n(y))\}\, d\nu_n(y).\end{aligned} \tag{3.4}$$

Define the pointwise integrated objective

$$\Psi_n(\beta, y) := \sum_{\ell=1}^{L} w_\ell f_{n\ell}(y) h_\alpha\{q_{n\ell}(\beta)\}. \tag{3.5}$$

Then (3.4) becomes

$$\mathcal{P}_{w,n}(\delta_n) = \int \Psi_n\{\delta_n(y), y\}\, d\nu_n(y). \tag{3.6}$$

For fixed y, $\beta \mapsto \Psi_n(\beta, y)$ is continuous on K_n by (3.2). For fixed β, $y \mapsto \Psi_n(\beta, y)$ is measurable because each $f_{n\ell}$ is measurable. Hence Ψ_n is a Carathéodory function, and therefore jointly Borel measurable.

Define

$$V_{w,n}(y) := \sup_{\beta \in K_n} \Psi_n(\beta, y). \tag{3.7}$$

Since K_n is compact and $\Psi_n(\cdot, y)$ is continuous,

$$V_{w,n}(y) = \max_{\beta \in K_n} \Psi_n(\beta, y). \tag{3.8}$$

Let D_n be a countable dense subset of K_n. By continuity,

$$V_{w,n}(y) = \sup_{\beta \in D_n} \Psi_n(\beta, y). \tag{3.9}$$

The right-hand side is a countable supremum of measurable functions. Thus

$$y \mapsto V_{w,n}(y) \quad \text{is measurable.} \tag{3.10}$$

Let

$$A_{w,n}(y) := \operatorname*{arg\,max}_{\beta \in K_n} \Psi_n(\beta, y). \tag{3.11}$$

For every y, $A_{w,n}(y)$ is nonempty and compact. Its graph is

$$\mathrm{Gr}(A_{w,n}) = \{(y, \beta) : \beta \in K_n,\ \Psi_n(\beta, y) = V_{w,n}(y)\}. \tag{3.12}$$

Since K_n is compact metric and Ψ_n is a Carathéodory objective, the measurable maximum theorem implies that $V_{w,n}$ is measurable and that the argmax correspondence

$$A_{w,n}(y) = \operatorname*{arg\,max}_{\beta \in K_n} \Psi_n(\beta, y)$$

is nonempty compact-valued and weakly measurable. The Kuratowski–Ryll-Nardzewski measurable selection theorem therefore gives a measurable selector

$$\delta_n^w(y) \in A_{w,n}(y) \qquad \nu_n\text{-a.e.} \tag{3.13}$$

Consequently, for every measurable rule $\delta_n : \mathsf{Y}_n \to K_n$,

$$\Psi_n\{\delta_n(y), y\} \leq \Psi_n\{\delta_n^w(y), y\} \qquad \nu_n\text{-a.e.} \tag{3.14}$$

Integrating (3.14) and using (3.6),

$$\mathcal{P}_{w,n}(\delta_n) \leq \mathcal{P}_{w,n}(\delta_n^w). \tag{3.15}$$

This proves existence of an integrated-power optimal rule and the pointwise characterization.

Now impose the collapse condition

$$B_{n\ell} = c_\ell B_{n0}, \qquad G_{n\ell} = a_\ell G_{n0}, \qquad c_\ell > 0, \quad a_\ell \neq 0. \tag{3.16}$$

For every $\beta \in K_n$,

$$\begin{aligned} q_{n\ell}(\beta) &= \rho_{\inf} \frac{(\beta' G_{n\ell})^2}{\beta' B_{n\ell} \beta} \\ &= \rho_{\inf} \frac{(\beta' a_\ell G_{n0})^2}{\beta' c_\ell B_{n0} \beta} \\ &= \frac{a_\ell^2}{c_\ell} \rho_{\inf} \frac{(\beta' G_{n0})^2}{\beta' B_{n0} \beta} \\ &= \kappa_\ell q_{n0}(\beta), \qquad \kappa_\ell := \frac{a_\ell^2}{c_\ell} > 0. \end{aligned} \tag{3.17}$$

Let

$$\beta^0 \in \arg\max_{\beta \in K_n} q_{n0}(\beta). \tag{3.18}$$

Compactness of K_n and continuity of q_{n0} imply that the argmax in (3.18) is nonempty. For every $\beta \in K_n$,

$$q_{n0}(\beta) \leq q_{n0}(\beta^0). \tag{3.19}$$

Multiplying (3.19) by $\kappa_\ell > 0$ and using (3.17),

$$q_{n\ell}(\beta) \leq q_{n\ell}(\beta^0), \qquad \forall \ell, \quad \forall \beta \in K_n. \tag{3.20}$$

The scalar AR power map is increasing. For $q > 0$,

$$h_\alpha'(q) = \frac{\phi(\sqrt{c_{1-\alpha}} - \sqrt{q}) - \phi(\sqrt{c_{1-\alpha}} + \sqrt{q})}{2\sqrt{q}} > 0,$$

and

$$h_\alpha'(0) = \sqrt{c_{1-\alpha}}\,\phi(\sqrt{c_{1-\alpha}}) > 0.$$

Thus (3.20) implies

$$h_\alpha\{q_{n\ell}(\beta)\} \leq h_\alpha\{q_{n\ell}(\beta^0)\}, \qquad \forall \ell, \quad \forall \beta \in K_n. \tag{3.21}$$

For integrated power, multiply (3.21) by $w_\ell f_{n\ell}(y) \geq 0$ and sum over ℓ:

$$\begin{aligned}\Psi_n(\beta, y) &= \sum_{\ell=1}^{L} w_\ell f_{n\ell}(y) h_\alpha\{q_{n\ell}(\beta)\} \\ &\leq \sum_{\ell=1}^{L} w_\ell f_{n\ell}(y) h_\alpha\{q_{n\ell}(\beta^0)\} \\ &= \Psi_n(\beta^0, y), \qquad \forall \beta \in K_n, \quad \nu_n\text{-a.e. } y.\end{aligned} \tag{3.22}$$

Therefore the constant rule

$$\delta_n^0(y) := \beta^0 \tag{3.23}$$

satisfies

$$\mathcal{P}_{w,n}(\delta_n) \leq \mathcal{P}_{w,n}(\delta_n^0) \qquad \text{for all measurable } \delta_n.$$

Hence δ_n^0 is integrated-power optimal under (3.16).

For the maximin criterion,

$$\mathcal{P}_{\min,n}(\delta_n) = \min_{1 \leq \ell \leq L} \mathbb{E}_{n\ell}\left[h_\alpha\{q_{n\ell}(\delta_n(Y_n))\}\right]. \tag{3.24}$$

For any measurable δ_n and every ℓ, (3.21) gives

$$h_\alpha\{q_{n\ell}(\delta_n(Y_n))\} \leq h_\alpha\{q_{n\ell}(\beta^0)\} \qquad P_{n\ell}\text{-a.s.} \tag{3.25}$$

Taking expectations under $P_{n\ell}$,

$$\mathbb{E}_{n\ell} h_\alpha\{q_{n\ell}(\delta_n(Y_n))\} \leq h_\alpha\{q_{n\ell}(\beta^0)\}, \qquad \ell = 1, \ldots, L. \tag{3.26}$$

Taking minima over ℓ,

$$\begin{aligned}\mathcal{P}_{\min,n}(\delta_n) &= \min_\ell \mathbb{E}_{n\ell} h_\alpha\{q_{n\ell}(\delta_n(Y_n))\} \\ &\leq \min_\ell h_\alpha\{q_{n\ell}(\beta^0)\} \\ &= \mathcal{P}_{\min,n}(\delta_n^0).\end{aligned} \tag{3.27}$$

Thus δ_n^0 is maximin optimal under the collapse condition. This completes the proof. *Q.E.D.*

3.2. *Proof of Corollary 6.2*

PROOF: Let

$$\Theta_L := \{\theta_1, \ldots, \theta_L\}$$

be the finite inversion grid, and define the grid confidence set

$$\mathcal{C}_N := \{\theta_j \in \Theta_L : AR_N(\theta_j) \leq c_{1-\alpha}\}. \tag{3.28}$$

Let $\mathcal{F}_{\mathrm{tr}}$ denote the training sigma-field. The direction-free score rule is $\mathcal{F}_{\mathrm{tr}}$-measurable. Hence, conditional on $\mathcal{F}_{\mathrm{tr}}$, the score used in each statistic $AR_N(\theta_j)$ is fixed.

The assumed uniform conditional null approximation gives, for every continuity point t of $F_{\chi^2_1}$,

$$\Delta_N(t) := \sup_{1\le j\le L} \left| \Pr_{\theta_j}\{AR_N(\theta_j) \le t \mid \mathcal{F}_{\rm tr}\} - F_{\chi^2_1}(t) \right| \xrightarrow{p} 0. \tag{3.29}$$

Since $F_{\chi^2_1}$ is continuous at $c_{1-\alpha}$,

$$F_{\chi^2_1}(c_{1-\alpha}) = 1-\alpha.$$

Set

$$\Delta_N := \Delta_N(c_{1-\alpha}).$$

Then (3.29) implies

$$\sup_{1\le j\le L} \left| \Pr_{\theta_j}\{AR_N(\theta_j) \le c_{1-\alpha} \mid \mathcal{F}_{\rm tr}\} - (1-\alpha) \right| = \Delta_N \xrightarrow{p} 0. \tag{3.30}$$

Moreover,

$$0 \le \Delta_N \le 1,$$

so

$$\mathbb{E}_{\theta_j}\Delta_N \to 0 \qquad \text{uniformly over } j=1,\dots,L. \tag{3.31}$$

Indeed, for every $\eta > 0$,

$$\mathbb{E}_{\theta_j}\Delta_N = \mathbb{E}_{\theta_j}\{\Delta_N 1(\Delta_N \le \eta)\} + \mathbb{E}_{\theta_j}\{\Delta_N 1(\Delta_N > \eta)\} \le \eta + \Pr_{\theta_j}(\Delta_N > \eta),$$

and the right side has $\limsup_N$ bounded by η; then let $\eta \downarrow 0$.

Fix $j \in \{1,\dots,L\}$ and suppose that the true value is θ_j. By (3.28),

$$1\{\theta_j \in \mathcal{C}_N\} = 1\{AR_N(\theta_j) \le c_{1-\alpha}\}. \tag{3.32}$$

Taking conditional expectations in (3.32),

$$\Pr_{\theta_j}\{\theta_j \in \mathcal{C}_N \mid \mathcal{F}_{\rm tr}\} = \Pr_{\theta_j}\{AR_N(\theta_j) \le c_{1-\alpha} \mid \mathcal{F}_{\rm tr}\}. \tag{3.33}$$

Combining (3.30) and (3.33),

$$\begin{aligned} \left| \Pr_{\theta_j}\{\theta_j \in \mathcal{C}_N \mid \mathcal{F}_{\rm tr}\} - (1-\alpha) \right| &= \left| \Pr_{\theta_j}\{AR_N(\theta_j) \le c_{1-\alpha} \mid \mathcal{F}_{\rm tr}\} - (1-\alpha) \right| \\ &\le \Delta_N. \end{aligned} \tag{3.34}$$

Taking expectations,

$$\begin{aligned} \left| \Pr_{\theta_j}\{\theta_j \in \mathcal{C}_N\} - (1-\alpha) \right| &= \left| \mathbb{E}_{\theta_j}\left[\Pr_{\theta_j}\{\theta_j \in \mathcal{C}_N \mid \mathcal{F}_{\rm tr}\} - (1-\alpha) \right] \right| \\ &\le \mathbb{E}_{\theta_j}\left| \Pr_{\theta_j}\{\theta_j \in \mathcal{C}_N \mid \mathcal{F}_{\rm tr}\} - (1-\alpha) \right| \\ &\le \mathbb{E}_{\theta_j}\Delta_N = o(1). \end{aligned} \tag{3.35}$$

Thus, for each fixed grid point,

$$\Pr_{\theta_j}\{\theta_j \in \mathcal{C}_N\} = 1 - \alpha + o(1). \tag{3.36}$$

The same calculation gives uniformity over the finite grid. Taking the supremum over $j = 1, \dots, L$ in (3.35),

$$\sup_{1 \le j \le L} \left| \Pr_{\theta_j}\{\theta_j \in \mathcal{C}_N\} - (1-\alpha) \right| \le \sup_{1 \le j \le L} \mathbb{E}_{\theta_j} \Delta_N = o(1). \tag{3.37}$$

Equations (3.36) and (3.37) prove the corollary. *Q.E.D.*

4. PRIMITIVE PLIV VERIFICATION PROOFS

This section proves the primitive verification results used in Section 5. The main distinction relative to an operator-norm-only Gaussian approximation is that exact canonical training in high dimension requires trace-scale covariance control, or an explicit Euclidean coupling. Operator-norm convergence controls worst-direction variance but does not by itself make the full p_n-dimensional Gaussian law close to the canonical law.

4.1. *Proof of Proposition 5.4*

PROOF: Write $n_t = n_{\mathrm{tr}}$, $I_t = I_{\mathrm{tr}}$, $b_i = b_{p_n}(Z_i, X_i)$, and $\pi_i = \pi_n(Z_i, X_i)$. From the centered first stage,

$$D_i - m_N(X_i) = N^{-1/2}\pi_i + V_i.$$

Therefore

$$\widehat{g}_n - g_n = \left\{ \frac{1}{n_t} \sum_{i \in I_t} b_i \pi_i - g_n \right\} + \frac{\sqrt{N}}{n_t} \sum_{i \in I_t} b_i V_i - \frac{\sqrt{N}}{n_t} \sum_{i \in I_t} b_i \{\widehat{m}(X_i) - m_N(X_i)\}.$$

Premultiplying by $C_n^{-1/2}$ gives

$$C_n^{-1/2}(\widehat{g}_n - g_n) = A_{n,t} + Z_{n,t} - M_{n,t},$$

where

$$A_{n,t} = C_n^{-1/2} \left\{ \frac{1}{n_t} \sum_{i \in I_t} b_i \pi_i - g_n \right\},$$

$$Z_{n,t} = C_n^{-1/2} \frac{\sqrt{N}}{n_t} \sum_{i \in I_t} b_i V_i,$$

$$M_{n,t} = C_n^{-1/2} \frac{\sqrt{N}}{n_t} \sum_{i \in I_t} b_i \{\widehat{m}(X_i) - m_N(X_i)\}.$$

The first two remainder assumptions give $\|A_{n,t}\|_2 = o_p(1)$ and $\|M_{n,t}\|_2 = o_p(1)$.

Conditional on $\mathcal{D}_{\mathrm{tr}}$, $Z_{n,t}$ is Gaussian with mean zero and covariance

$$\mathrm{Var}(Z_{n,t} \mid \mathcal{D}_{\mathrm{tr}}) = \frac{N}{n_t} C_n^{-1/2} \widehat{C}_n^{\mathrm{tr}} C_n^{-1/2}.$$

With the notation of Proposition 5.4,

$$C_n^{-1/2}\widehat{C}_n^{\mathrm{tr}}C_n^{-1/2} = I_{p_n} + R_{C,n},$$

and hence

$$\mathrm{Var}(Z_{n,t} \mid \mathcal{D}_{\mathrm{tr}}) = \rho_{\mathrm{tr}}^{-1}(I_{p_n} + \widetilde{R}_{C,n}), \qquad \widetilde{R}_{C,n} = \rho_{\mathrm{tr}}\frac{N}{n_t}(I_{p_n} + R_{C,n}) - I_{p_n}.$$

Since $n_t/N \to \rho_{\mathrm{tr}}$ and $\|R_{C,n}\|_{\mathrm{op}} = o_p(1)$, $\|\widetilde{R}_{C,n}\|_{\mathrm{op}} = o_p(1)$. This proves (5.17).

If $\mathrm{tr}(\widetilde{R}_{C,n}^2) = o_p(1)$, then the conditional Gaussian law is close to the canonical law. On the event $\|\widetilde{R}_{C,n}\|_{\mathrm{op}} \le 1/2$,

$$\begin{aligned}\mathrm{KL}\{N(0, I_{p_n} + \widetilde{R}_{C,n}), N(0, I_{p_n})\} &= \frac{1}{2}\left[\mathrm{tr}(\widetilde{R}_{C,n}) - \log\det(I_{p_n} + \widetilde{R}_{C,n})\right]\\ &\le C\,\mathrm{tr}(\widetilde{R}_{C,n}^2).\end{aligned}$$

Pinsker's inequality gives total-variation convergence, and bounded-Lipschitz convergence follows because $d_{\mathrm{BL}} \le 2\|\cdot\|_{\mathrm{TV}}$. Since $\|\widetilde{R}_{C,n}\|_{\mathrm{op}} = o_p(1)$, the complement of $\{\|\widetilde{R}_{C,n}\|_{\mathrm{op}} \le 1/2\}$ has probability $o(1)$. Hence the preceding total-variation bound holds in probability. Adding the two $o_p(1)$ Euclidean remainders preserves bounded-Lipschitz convergence by the Lipschitz property. Thus (5.9) holds. *Q.E.D.*

4.2. *Proof of Proposition 5.5*

PROOF: Let

$$S_{n,t} = C_n^{-1/2}\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_{p_n}(Z_i, X_i)V_i.$$

The direct vector approximation (5.18) gives conditional bounded-Lipschitz convergence from $S_{n,t}$ to a centered Gaussian vector with conditional covariance

$$\frac{N}{n_{\mathrm{tr}}}C_n^{-1/2}\widehat{C}_n^{\mathrm{tr}}C_n^{-1/2} = \rho_{\mathrm{tr}}^{-1}(I_{p_n} + \widetilde{R}_{C,n}).$$

If $\mathrm{tr}(\widetilde{R}_{C,n}^2) = o_p(1)$, the same Gaussian covariance comparison used in the proof of Proposition 5.4 gives

$$d_{\mathrm{BL}}\{N(0, \rho_{\mathrm{tr}}^{-1}(I_{p_n} + \widetilde{R}_{C,n})), N(0, \rho_{\mathrm{tr}}^{-1}I_{p_n})\} \xrightarrow{p} 0.$$

The triangle inequality proves (5.9).

Next consider the strong-coupling statement. In the denominator-normalized coordinates put

$$S_{n,\Delta}^B = B_{n,\Delta}^{-1/2}\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_{p_n}(Z_i, X_i)V_i, \qquad \Gamma_{n,\Delta}^B = \rho_{\mathrm{tr}}^{-1}\sigma_\Delta^2(I_{p_n} + \Theta_{n,\Delta}^{\mathrm{tr}}).$$

Here $W_2^2(P,Q) = \inf_{\pi\in\Pi(P,Q)}\int \|x-y\|_2^2\,d\pi(x,y)$, with $\Pi(P,Q)$ denoting the couplings of P and Q. In the conditional expression used in the main text, the design sigma-field is held fixed before taking the infimum. By the definition of quadratic Wasserstein distance, after enlarging the probability space if necessary, condition (5.19) gives a Gaussian vector $G^B_{n,\Delta} \mid \mathcal{D}_{\mathrm{tr}} \sim N(0, \Gamma^B_{n,\Delta})$ and a remainder $u_{n,\Delta}$ such that

$$S^B_{n,\Delta} = G^B_{n,\Delta} + u_{n,\Delta}, \qquad \frac{\mathbb{E}\|u_{n,\Delta}\|_2^2}{r_n^2} = o\{\min(1, \eta_{n,\Delta})\}.$$

Because $I_{p_n} + \Theta^{\mathrm{tr}}_{n,\Delta}$ is the normalized conditional covariance matrix of $G^B_{n,\Delta}$, it is positive semidefinite. Hence

$$G^B_{n,\Delta} = \rho_{\mathrm{tr}}^{-1/2} \sigma_\Delta (I_{p_n} + \Theta^{\mathrm{tr}}_{n,\Delta})^{1/2} \xi_n, \qquad \xi_n \sim N(0, I_{p_n}),$$

on an enlarged probability space. Every eigenvalue x of $\Theta^{\mathrm{tr}}_{n,\Delta}$ satisfies $x \geq -1$, and

$$|(1+x)^{1/2} - 1|^2 = \frac{x^2}{\{(1+x)^{1/2} + 1\}^2} \leq x^2, \qquad x \geq -1.$$

By spectral calculus,

$$\begin{aligned}\mathbb{E}\left[\left\|\{(I_{p_n} + \Theta^{\mathrm{tr}}_{n,\Delta})^{1/2} - I_{p_n}\}\xi_n\right\|_2^2 \middle| \mathcal{D}_{\mathrm{tr}}\right] &= \mathrm{tr}\left[\{(I_{p_n} + \Theta^{\mathrm{tr}}_{n,\Delta})^{1/2} - I_{p_n}\}^2\right] \\ &\leq \mathrm{tr}\{(\Theta^{\mathrm{tr}}_{n,\Delta})^2\}.\end{aligned}$$

Therefore, using (5.27),

$$\frac{\mathbb{E}\left\|\rho_{\mathrm{tr}}^{-1/2} \sigma_\Delta \{(I_{p_n} + \Theta^{\mathrm{tr}}_{n,\Delta})^{1/2} - I_{p_n}\}\xi_n\right\|_2^2}{r_n^2} = o\{\min(1, \eta_{n,\Delta})\}.$$

Combining the last two displays gives the Euclidean coupling required by the sparse upper-bound argument.

It remains only to justify the displayed order of the vector third-moment input. If

$$K_{\sigma C,n} := \sup_{z,x} \sigma_V(z,x) \|C_n^{-1/2} b_{p_n}(z,x)\|_2$$

and the conditional standardized third moments of V are uniformly bounded, then

$$\begin{aligned}\mathbb{E}[\|X_{ni}\|_2^3 \mid \mathcal{D}_{\mathrm{tr}}] &= \|C_n^{-1/2} b_{p_n}(Z_i, X_i)\|_2^3 \mathbb{E}[|V_i|^3 \mid Z_i, X_i] \\ &\leq C K^3_{\sigma C,n}.\end{aligned}$$

Hence

$$\mathfrak{b}_{n,\mathrm{tr}} \leq C \frac{N^{3/2}}{n_{\mathrm{tr}}^3} n_{\mathrm{tr}} K^3_{\sigma C,n} = O_p(K^3_{\sigma C,n}/\sqrt{n_{\mathrm{tr}}}),$$

because N/n_{tr} is bounded. If $K_{\sigma C,n} = O(\sqrt{p_n})$, this becomes $p_n^{3/2}/\sqrt{n_{\mathrm{tr}}} \to 0$. *Q.E.D.*

4.3. *Proof of Proposition 5.6*

PROOF: Fix Δ and condition on the inference-side design. Write $B = B_{n,\Delta}$, and

$$K_{\omega B,n} = \sup_{z,x} \omega_\Delta(z,x)^{1/2} \|B_{n,\Delta}^{-1/2} b_{p_n}(z,x)\|_2.$$

Write $\omega_i = \omega_\Delta(Z_i, X_i)$, $b_i = b_{p_n}(Z_i, X_i)$, and

$$\widehat{B}_n^o = \frac{1}{n_{\inf}} \sum_{i \in I_{\inf}} b_i b_i' \omega_i.$$

For $\beta \in \mathcal{H}_{n,\Delta}(c, C)$, define

$$X_i(\beta) = \frac{n_{\inf}^{-1/2} \beta' b_i R_{i,\Delta}}{(\beta' B \beta)^{1/2}}.$$

The conditional variance is

$$s_n^2(\beta) := \sum_{i \in I_{\inf}} \mathrm{Var}\{X_i(\beta) \mid \mathcal{D}_{\inf}\} = \frac{\beta' \widehat{B}_n^o \beta}{\beta' B \beta}.$$

The oracle covariance concentration assumption gives

$$\sup_{\beta \in \mathcal{H}_{n,\Delta}(c,C)} |s_n^2(\beta) - 1| = o_p(1).$$

Let

$$a_i(\beta) = \frac{n_{\inf}^{-1/2} \beta' b_i \omega_i^{1/2}}{(\beta' B \beta)^{1/2}}.$$

Then

$$\max_i |a_i(\beta)| \le \frac{K_{\omega B,n}}{\sqrt{n_{\inf}}},$$

and

$$\sum_{i \in I_{\inf}} a_i(\beta)^2 = s_n^2(\beta) = 1 + o_p(1)$$

uniformly over $\beta \in \mathcal{H}_{n,\Delta}(c, C)$. By the conditional Berry–Esseen inequality for scalar triangular arrays,

$$\sup_t \left| \Pr\left\{ \frac{\sum_i X_i(\beta)}{s_n(\beta)} \le t \middle| \mathcal{D}_{\inf} \right\} - \Phi(t) \right| \le C \sum_i \frac{\mathbb{E}[|X_i(\beta)|^3 \mid \mathcal{D}_{\inf}]}{s_n(\beta)^3}.$$

The standardized third-moment bound gives, uniformly over β,

$$\sum_i \mathbb{E}[|X_i(\beta)|^3 \mid \mathcal{D}_{\inf}] \le C \sum_i |a_i(\beta)|^3 \le C \max_i |a_i(\beta)| \sum_i a_i(\beta)^2 = O_p(K_{\omega B,n}/\sqrt{n_{\inf}}) = o_p(1).$$

Therefore the standardized sum is conditionally standard normal uniformly over $\beta \in \mathcal{H}_{n,\Delta}(c, C)$. Since $s_n(\beta) = 1 + o_p(1)$ uniformly, replacing the conditional standard deviation by one changes the bounded-Lipschitz distance by $o_p(1)$. This proves (5.13). If the feasible variance estimator is operator-norm consistent in the same geometry, then the Rayleigh quotient bound immediately gives (5.11). *Q.E.D.*

4.4. *Proof of Proposition 5.7*

PROOF: Write $b = b_{p_n}(Z,X)$, $B = B_{n,\Delta}$, $\omega = \omega_\Delta(Z,X)$, and $e_\Delta(Z,X) = \sigma_V^2(Z,X)/\omega_\Delta(Z,X) - \kappa_\Delta$. Then

$$C_n - \kappa_\Delta B = \mathbb{E}[e_\Delta(Z,X)\omega bb'].$$

Premultiplying and postmultiplying by $B^{-1/2}$ gives

$$M_{n,\Delta} := B^{-1/2}(C_n - \kappa_\Delta B)B^{-1/2} = \mathbb{E}[e_\Delta(Z,X)a(Z,X)a(Z,X)'],$$

where $a(Z,X) = B^{-1/2}\omega_\Delta(Z,X)^{1/2}b_{p_n}(Z,X)$. Since $\mathbb{E}[a(Z,X)a(Z,X)'] = I_{p_n}$, for every unit vector u,

$$|u'M_{n,\Delta}u| \le \|e_\Delta\|_\infty \mathbb{E}[(u'a)^2] = \|e_\Delta\|_\infty.$$

Thus

$$\|M_{n,\Delta}\|_{\mathrm{op}} \le \epsilon_{n,\Delta},$$

which proves (5.20). Because $M_{n,\Delta}$ is symmetric, its eigenvalues are bounded in absolute value by $\epsilon_{n,\Delta}$. Therefore

$$\mathrm{tr}(M_{n,\Delta}^2) \le p_n\epsilon_{n,\Delta}^2,$$

which is (5.21). *Q.E.D.*

4.5. *Proof of Theorem 5.9*

PROOF: The operational drift and nuisance conditions in Assumption 5.1 are included directly in Assumption 5.8. The inference scalar CLT follows from the conditionally Gaussian inference branch or from Proposition 5.6, and feasible variance consistency is imposed in the primitive aligned design. The bounded frontier and eigenvalue conditions are also part of the primitive aligned design.

It remains to verify the training Gaussian approximation and the stronger sparse-frontier coupling. Work in the denominator-normalized coordinates

$$\widetilde{Z}_{n,\Delta} = B_{n,\Delta}^{-1/2}\widehat{g}_n, \qquad \widetilde{\gamma}_{n,\Delta} = B_{n,\Delta}^{-1/2}g_n.$$

Using the first-stage decomposition,

$$\widetilde{Z}_{n,\Delta} - \widetilde{\gamma}_{n,\Delta} = r_{n,\Delta}^B + S_{n,\Delta}^B,$$

where

$$r_{n,\Delta}^B = B_{n,\Delta}^{-1/2}\left[\left\{\frac{1}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_i\pi_i - g_n\right\} - \frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_i\{\widehat{m}(X_i) - m_N(X_i)\}\right]$$

and

$$S_{n,\Delta}^B = B_{n,\Delta}^{-1/2}\frac{\sqrt{N}}{n_{\mathrm{tr}}}\sum_{i\in I_{\mathrm{tr}}} b_iV_i.$$

For the operational training approximation in Assumption 5.1, use the C_n-normalized covariance condition (5.25). This condition is exactly the trace/Frobenius canonical-equivalence condition in (5.25), with $\widetilde{R}_{C,n} = \widetilde{R}_{C,n}^{\mathrm{tr}}$. Since $n_{\mathrm{tr}}/N \to \rho_{\mathrm{tr}}$, its operator-norm part also implies the operator covariance condition required in Proposition 5.4. Hence, in the conditionally Gaussian branch, Proposition 5.4 gives (5.9). In the non-Gaussian branch, the direct bounded-Lipschitz vector approximation in Proposition 5.5, together with (5.25), gives the same operational approximation. Thus (5.9) holds, and hence Assumption 5.1 holds.

The denominator-normalized covariance condition (5.26) and the stronger rate (5.27) are used only in the sparse upper-bound coupling below.

It remains to prove the Euclidean coupling used by the sparse upper bound. In the Gaussian branch, conditional on the design,

$$S^B_{n,\Delta} = \rho_{\rm tr}^{-1/2}\sigma_\Delta(I_{p_n} + \Theta^{\rm tr}_{n,\Delta})^{1/2}\xi_n, \qquad \xi_n \sim N(0, I_{p_n}),$$

on an enlarged probability space. In the non-Gaussian branch, condition (5.19) gives a coupling

$$S^B_{n,\Delta} = \rho_{\rm tr}^{-1/2}\sigma_\Delta(I_{p_n} + \Theta^{\rm tr}_{n,\Delta})^{1/2}\xi_n + u_{n,\Delta}, \qquad \frac{\mathbb{E}\|u_{n,\Delta}\|_2^2}{r_n^2} = o\{\min(1, \eta_{n,\Delta})\}.$$

In the Gaussian branch set $u_{n,\Delta} = 0$. Therefore, in both strong-coupling branches,

$$\begin{aligned} \widetilde{Z}_{n,\Delta} &= \widetilde{\gamma}_{n,\Delta} + \rho_{\rm tr}^{-1/2}\sigma_\Delta\xi_n + e_{n,\Delta}, \\ e_{n,\Delta} &= r^B_{n,\Delta} + u_{n,\Delta} + \rho_{\rm tr}^{-1/2}\sigma_\Delta\{(I_{p_n} + \Theta^{\rm tr}_{n,\Delta})^{1/2} - I_{p_n}\}\xi_n. \end{aligned}$$

The preceding covariance-square-root term can be bounded without restricting to the event $\|\Theta^{\rm tr}_{n,\Delta}\|_{\rm op} \le 1/2$. Conditional on the design, $I_{p_n} + \Theta^{\rm tr}_{n,\Delta}$ is positive semidefinite, so every eigenvalue x of $\Theta^{\rm tr}_{n,\Delta}$ satisfies $x \ge -1$, and

$$|(1+x)^{1/2} - 1|^2 = \frac{x^2}{\{(1+x)^{1/2} + 1\}^2} \le x^2, \qquad x \ge -1.$$

By spectral calculus,

$$\begin{aligned} \mathbb{E}\left[\left\|\{(I_{p_n} + \Theta^{\rm tr}_{n,\Delta})^{1/2} - I_{p_n}\}\xi_n\right\|_2^2 \middle| \mathcal{D}_{\rm tr}\right] &= {\rm tr}\big[\{(I_{p_n} + \Theta^{\rm tr}_{n,\Delta})^{1/2} - I_{p_n}\}^2\big] \\ &\le {\rm tr}\{(\Theta^{\rm tr}_{n,\Delta})^2\}. \end{aligned}$$

Combining this bound with (5.23), (5.27), and, in the non-Gaussian case, (5.19), yields

$$\frac{\mathbb{E}\|e_{n,\Delta}\|_2^2}{r_n^2} = o\{\min(1, \eta_{n,\Delta})\}.$$

This is (5.28), and Theorem 5.12 therefore gives the stated sparse PLIV upper bound. The argument uses the non-Gaussian bounded-Lipschitz approximation only for the operational AR reduction; the sparse upper-bound step uses the displayed Euclidean coupling. *Q.E.D.*

5. SUPPLEMENTARY NUMERICAL BENCHMARK FOR THEOREM 2.1 AND COROLLARY 2.7

The following numerical experiment illustrates the canonical minimax envelope in Theorem 2.1 and the power-regret transfer in Corollary 2.7. The simulation is in the limiting Gaussian sequence experiment rather than a finite-sample PLIV design. It reports how normalized regret and power regret move with the difficulty index $\eta = \sigma^2 s \log(ep/s)/r^2$.

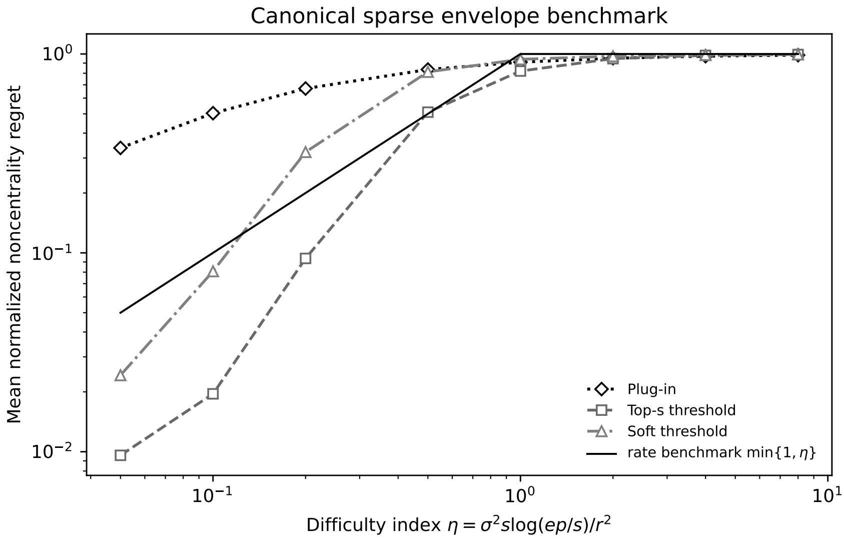

FIGURE 1.—Canonical sparse envelope benchmark: normalized regret. The dashed line is the rate benchmark $\min\{1, \eta\}$.

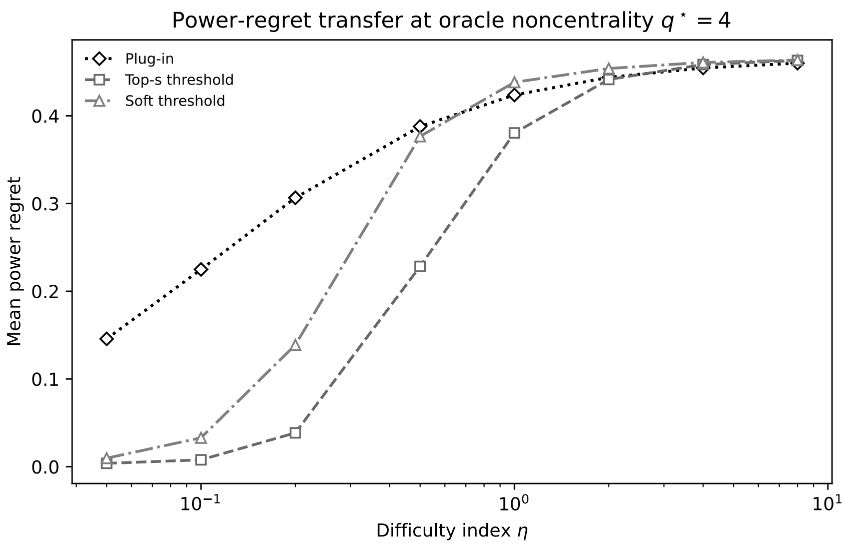

FIGURE 2.—Power-regret transfer at oracle noncentrality $q^\star = 4$.

5.1. *Finite-sample PLIV transfer check*

This subsection gives supplementary details for Figure 6 in the main text. The simulation uses

$$D_i = N^{-1/2} b_i' g + V_i, \qquad Y_i = \Delta D_i + U_i, \qquad b_i \sim N(0, I_p),$$

with $N = 1000$, $p = 120$, $s = 8$, $\Delta = 1$, and $\rho_{\mathrm{tr}} = \rho_{\inf} = 1/2$. The sparse drift has s equal nonzero coordinates and is scaled so that the oracle scalar-AR noncentrality is four. The variance of U_i is chosen so that $\mathrm{Var}(U_i + \Delta V_i) = 1$, hence $B_{n,\Delta} = I_p$ in the population design. The

TABLE I

CANONICAL SPARSE ENVELOPE BENCHMARK

η	Rule	Mean normalized regret	Mean power regret	Mean frontier attainment
0.10	Plug-in	0.504 (0.000)	0.225 (0.000)	0.496 (0.000)
	Top-s threshold	0.020 (0.000)	0.008 (0.000)	0.980 (0.000)
	Soft threshold	0.081 (0.000)	0.033 (0.000)	0.919 (0.000)
1.00	Plug-in	0.910 (0.000)	0.424 (0.000)	0.090 (0.000)
	Top-s threshold	0.821 (0.001)	0.380 (0.000)	0.179 (0.001)
	Soft threshold	0.941 (0.000)	0.438 (0.000)	0.059 (0.000)
4.00	Plug-in	0.975 (0.000)	0.455 (0.000)	0.025 (0.000)
	Top-s threshold	0.984 (0.000)	0.458 (0.000)	0.016 (0.000)
	Soft threshold	0.989 (0.000)	0.461 (0.000)	0.011 (0.000)

Notes: The benchmark is computed from the canonical Gaussian sequence experiment with $p = 1000$, $s = 20$, and oracle noncentrality normalized to $q^\star = 4$; the training-noise standard deviation is chosen to attain the displayed difficulty index. The soft-thresholding rule uses $\tau = \sigma\sqrt{2\log p}$ and the top-coordinate fallback. Monte Carlo standard errors are in parentheses; each row uses 8,000 draws.

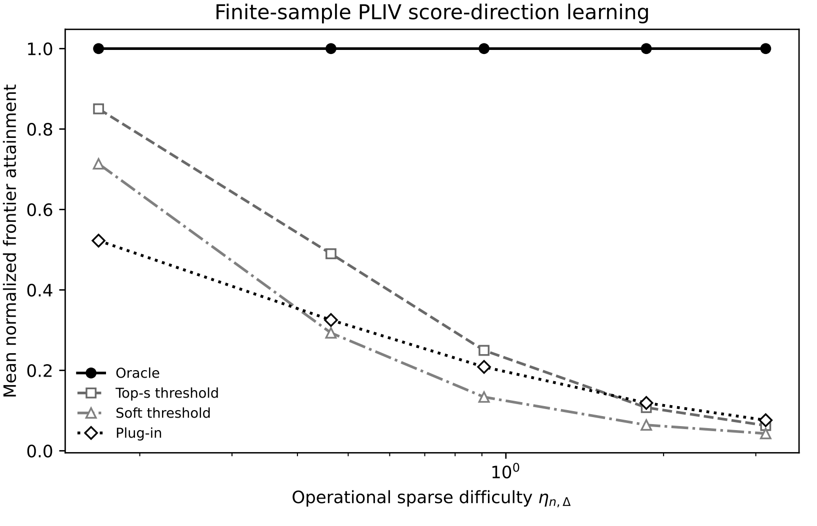

FIGURE 3.—Finite-sample PLIV score-direction learning. The plot reports mean normalized frontier attainment as the operational PLIV sparse difficulty index varies.

reported canonical power is $h_\alpha(q^\star A_n)$, where A_n is the realized normalized frontier attainment of the learned score.

TABLE II

FINITE-SAMPLE PLIV TRANSFER CHECK

$\eta_{n,\Delta}$	Rule	Null rejection	Empirical power	Canonical power	Mean frontier attainment
0.46	Oracle	0.056 (0.008)	0.506 (0.018)	0.516 (0.000)	1.000 (0.000)
	Top-s threshold	0.046 (0.007)	0.263 (0.016)	0.287 (0.003)	0.490 (0.006)
	Soft threshold	0.054 (0.008)	0.182 (0.014)	0.191 (0.002)	0.293 (0.005)
	Plug-in	0.045 (0.007)	0.209 (0.014)	0.207 (0.001)	0.325 (0.002)
1.85	Oracle	0.050 (0.008)	0.495 (0.018)	0.516 (0.000)	1.000 (0.000)
	Top-s threshold	0.041 (0.007)	0.084 (0.010)	0.101 (0.002)	0.108 (0.003)
	Soft threshold	0.041 (0.007)	0.072 (0.009)	0.080 (0.001)	0.064 (0.002)
	Plug-in	0.052 (0.008)	0.094 (0.010)	0.106 (0.001)	0.119 (0.002)

Notes: The table reports finite-sample rejection probabilities, canonical predictions, and mean normalized frontier attainment for the sample-split PLIV design described in Section 5.1.