RANK IS NOT ENOUGH: ADAPTIVE POLICY SELECTION AT SINGULAR TIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Rank is not enough to predict the difficulty of selecting among tied policies. We construct finite MDPs with identical state and action spaces, transition-rank budgets, symmetry multiplicities, Fisher information, exact policy tie, and comparison-gradient norm, yet whose fixed-radius local minimax regrets differ by a factor $\Theta_R(\sqrt{d/r})$. The governing object is the Fisher-relative geometry between the contingent cone $\mathcal{K}_0$ of feasible transition perturbations and the polytope $\mathfrak{G}_0$ of tied policy gradients. For equivariant sparse low-rank transition models, we first prove that the oracle-local root-n decision problem converges exactly to the Gaussian game determined by $(\mathcal{K}_0, \mathfrak{G}_0)$. We then construct one unknown-center rule that attains this oracle value uniformly on quantitative stable support–rank–policy strata. At singular rank intersections, however, we prove a positive adaptation tax: no branch-agnostic procedure can simultaneously attain the neighboring oracle risks. A commitment lift realizes the rank-diffusion gap in the complete deterministic policy class. Finally, we give a certified finite algorithm for the Gaussian game and finite-sample categorical experiments that recover adaptive attainment, adaptation failure, and the predicted $\sqrt{d/r}$ scaling.

1 INTRODUCTION

Rank is not enough. Two finite MDPs can have the same state and action spaces, transition-rank budgets, symmetry multiplicities, Fisher information, exact policy tie, and comparison-gradient norm, yet differ by a factor $\Theta_R(\sqrt{d/r})$ in fixed-radius local policy-selection difficulty. The missing object is the *Fisher-relative alignment* between feasible transition perturbations and policy-value gradients. We formalize this object, show when it can be learned at no first-order cost, and prove that learning it can become impossible at singular model intersections.

Let $\mathfrak{T} = (\mu, (r_t)_{t<T}, r_T, \Pi)$ denote a finite-horizon task with a finite policy class. At a strictly positive transition center P_0, define the Bellman–Le Cam pair

$$\mathfrak{B}_{\mathfrak{T}}(P_0, w) := (\mathcal{K}_0, \mathfrak{G}_{0,\mathfrak{T}}), \qquad \mathcal{K}_0 := \mathsf{W}_0 T_\Theta(P_0), \qquad \mathfrak{G}_{0,\mathfrak{T}} := \operatorname{conv}\{\gamma_\pi : \pi \in \Pi_0\}. \tag{1}$$

Here $\mathcal{K}_0$ is the Fisher-whitened contingent cone of the structured transition model and $\mathfrak{G}_{0,\mathfrak{T}}$ is the polytope of first-order gradients of the tied policies. The corresponding Gaussian loss is

$$L_{0,\mathfrak{T}}(\theta, q) := h_{\mathfrak{G}_{0,\mathfrak{T}}}(\theta) - \langle \bar{\gamma}_q, \theta \rangle, \qquad Z \sim N(\theta, I). \tag{2}$$

The oracle benchmark is the minimax value of this Gaussian game over $\theta \in \mathcal{K}_0 \cap R\mathbb{B}_2$.

Headline result: an adaptation phase transition. On quantitative support-saturated and rank-saturated strata, we construct one unknown-center rule that estimates the active stratum, tied face, Fisher-normal coordinate, and Gaussian game, and attains the oracle value uniformly. At a rank intersection, however, two noncollinear efficient directions can be statistically indistinguishable; every center-independent rule then pays a strictly positive first-order adaptation tax. Thus the oracle geometry is not merely descriptive:

$$\boxed{\begin{array}{c}\text{stable support–rank–policy strata} \Longrightarrow \text{adaptive oracle efficiency,} \\ \text{singular intersections} \Longrightarrow \text{an unavoidable adaptation penalty.}\end{array}} \tag{3}$$

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

$$\mathcal{K}_{d,r} = \{\Delta : \operatorname{rank}(\Delta) \le r\},$$
$$G_{\mathrm{conc}} = r^{-1/2}\operatorname{diag}(I_r, 0), \qquad G_{\mathrm{flat}} = d^{-1/2} I_d,$$
$$\|G_{\mathrm{conc}}\|_{\mathrm{F}} = \|G_{\mathrm{flat}}\|_{\mathrm{F}} = 1,$$
$$\omega_{\mathcal{K}_{d,r}}(G_{\mathrm{conc}}) = 1,$$
$$\omega_{\mathcal{K}_{d,r}}(G_{\mathrm{flat}}) = \sqrt{r/d}.$$

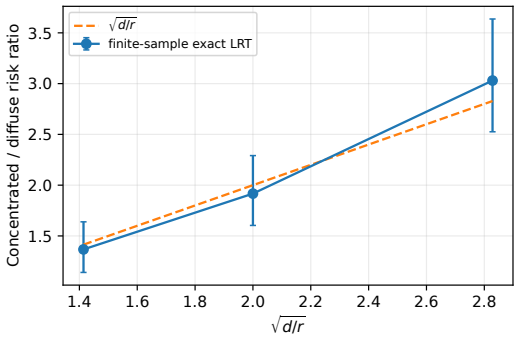

Figure 1: **Rank is not enough.** Equal Frobenius norm and a common rank-r feasible cone leave a $\sqrt{d/r}$ gap in restricted decision width. The full-policy finite-MDP experiment exhibits the same scaling at finite sample size.

Second result: a full-policy realization. A strictly positive commitment lift makes exactly two deterministic nonstationary policies globally optimal and gives every other deterministic policy a common positive base gap. Concentrated and diffuse tasks preserve the same complete base-policy value table, Fisher experiment, structural model, and comparison-gradient norm, but their limiting local risks satisfy a two-sided $\Theta_R(\sqrt{d/r})$ ratio. The phenomenon is therefore intrinsic to full MDP policy optimization, not to a hand-selected pair of experts.

Operational and empirical consequences. For arbitrary compact Gaussian mean sets we prove finite-net approximation, least-favorable-prior duality, a certified finite linear program, and a finite-termination exchange method. Simulations use the exact categorical transition experiments rather than direct Gaussian samples: the adaptive rule approaches the branch oracle on stable strata, the adaptation gap persists at the singular intersection, and the full-policy risk ratio tracks $\sqrt{d/r}$.

Relation to prior work. Local Gaussian decision theory underlies statistical treatment choice and best-arm regret (Hirano & Porter, 2009; Adusumilli, 2022); nonregular max-type inference and nonunique optimal decisions are studied by Song (2014); Luedtke & van der Laan (2016); Zhu et al. (2024); tangent cones govern singular statistical asymptotics (Drton, 2009; Olikier et al., 2026). Global low-rank and linear MDP theory instead studies exploration or representation learning (Jiang et al., 2017; Jin et al., 2020; Agarwal et al., 2020), while symmetry work studies equivariant policies and abstractions (van der Pol et al., 2020; Panangaden et al., 2024). Our contribution is the structured synthesis: stochastic equivariant sparse-determinantal geometry, Bellman regret, matching Gaussian decision transfers, adaptive attainment and its singularity barrier, and a full-policy rank-diffusion realization.

2 Structured policy selection and the oracle benchmark

Let $\mathcal{S}, \mathcal{A}$ be finite, $\mathcal{X} = \mathcal{S} \times \mathcal{A}$, and $T \ge 2$. The unknown transition kernel belongs to $\Theta_{G,s,\boldsymbol{r}}$: it is stochastic, equivariant under a known finite group, supported on at most s nontrivial real-isotypic sectors, and has multiplicity-block ranks at most r_ρ. Rewards, the initial law, and the finite candidate class Π are known. From every $x \in \mathcal{X}$ we observe independent categorical samples with $n_x/n \to w_x > 0$. Fix a strictly positive center P_0 and radius R.

For an external policy lottery $q \in \Delta(\Pi)$, let $J^q(P) = \sum_\pi q(\pi) J^\pi(P)$ and define

$$\boxed{\mathcal{R}^{\mathrm{all}}_{n,R}(P_0; \mathfrak{T}) := \inf_{\delta_n : \mathcal{Y}_n \rightsquigarrow \Delta(\Pi)} \sup_{P \in \Theta_{n,R}(P_0)} \mathbb{E}_P\left[\sqrt{n}\{J^\star_{\mathfrak{T}}(P) - J^{\delta_n(Y)}(P)\}\right].} \tag{4}$$

The expectation includes observation and internal randomization. In the oracle benchmark the rule may use P_0, its Fisher whitening W_0, and the tied face Π_0. The order of limits is

$$\boxed{\text{fix } (P_0, \mathfrak{T}, R, d, T, G, s, \boldsymbol{r}, w), \qquad n \to \infty.} \tag{5}$$

2.1 STRUCTURED AND BELLMAN GEOMETRY

In symmetry-adapted coordinates, write the active base support as Λ_0 and the blockwise determinantal tangent cones as $\mathcal{T}_\rho$. The normalized trivial sector contributes a linear nuisance space $\mathcal{N}_{\mathbf{1}}$.

Theorem 2.1 (Quantitative structured geometry)**.** *At a strictly positive center, under the real-Schur assumption,*

$$\mathcal{C}_0 = T_{\Theta_{G,s,\boldsymbol{r}}}(P_0) = \mathcal{B}_G\left(\mathcal{N}_{\mathbf{1}} \oplus \bigcup_{\substack{\Lambda \supseteq \Lambda_0 \\ |\Lambda| \le s}} \left[\prod_{\rho\in\Lambda} \mathcal{T}_\rho \times \prod_{\rho\notin\Lambda} \{0\}\right]\right). \tag{6}$$

For every bounded tangent direction there is a feasible path $P_u(t) = P_0 + tu + O(t^2)$ uniformly in Fisher norm, and the bounded rescaled local models converge to $\mathcal{C}_0$ at Hausdorff rate $O(t)$. The cone and its whitened image are closed and centrally symmetric.

Corollary 2.2 (Central symmetry)**.**

$$\mathcal{C}_0 = -\mathcal{C}_0, \qquad \mathcal{K}_0 := \mathsf{W}_0\mathcal{C}_0 = -\mathcal{K}_0. \tag{7}$$

For a policy π, its transition derivative is the occupancy–continuation pairing

$$\ell_\pi(u) = \sum_{t<T}\sum_x d_t^\pi(x) \sum_{s'} u(s' \mid x) V_{t+1}^\pi(s').$$

Let γ_π be its Fisher–Riesz representative.

Theorem 2.3 (Bellman loss and exact tie regularity)**.** *For the finite candidate class,*

$$DJ_{P_0}^\pi[u] = \ell_\pi(u), \qquad DJ_{P_0}^\star[u] = \max_{\pi\in\Pi_0} \ell_\pi(u) = h_{\mathfrak{G}_{0,\mathfrak{T}}}(\mathsf{W}_0 u). \tag{8}$$

The optimized value is Fréchet differentiable in the ambient normalized equivariant model iff all active derivatives agree. On the structured cone, all first-order pairwise regrets vanish iff $\langle \gamma_\pi - \gamma_{\pi'}, \theta\rangle = 0$ for every $\theta \in \mathcal{K}_0$. Along the corrected feasible paths, policy regret equals $tL_{0,\mathfrak{T}}(\mathsf{W}_0 u, q) + O(t^2)$ uniformly on bounded sets.

2.2 ORACLE GAUSSIAN DECISION LIMIT

Let Z_n be the observable Fisher central sequence and $\theta_n(P) = \mathsf{W}_0\sqrt{n}(P - P_0)$. Put $\mathcal{K}_R = \mathcal{K}_0 \cap R\mathbb{B}_2$ and

$$\mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_{0,\mathfrak{T}}) := \inf_\delta \sup_{\theta\in\mathcal{K}_R} \mathbb{E}_{Z\sim N(\theta,I)} L_{0,\mathfrak{T}}(\theta, \delta(Z)). \tag{9}$$

Theorem 2.4 (Oracle Bellman–Le Cam benchmark)**.** *For fixed R and fixed dimension, the actual local mean sets converge to $\mathcal{K}_R$, the categorical likelihoods satisfy uniform LAN, and*

$$\sup_{P\in\Theta_{n,R}(P_0)} d_{\mathrm{BL}}(\mathcal{L}_P(Z_n), N(\theta_n(P), I)) \to 0. \tag{10}$$

Moreover, uniformly in local alternatives and candidate lotteries,

$$\sqrt{n}\{J^\star(P) - J^q(P)\} = L_{0,\mathfrak{T}}(\theta_n(P), q) + O(n^{-1/2}),$$

and

$$\boxed{\lim_{n\to\infty} \mathcal{R}_{n,R}^{\mathrm{all}}(P_0; \mathfrak{T}) = \mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_{0,\mathfrak{T}}).} \tag{11}$$

The lower bound uses finite embedded Gaussian subexperiments and likelihood- ratio uniform integrability; the upper bound applies Lipschitz near-minimax Gaussian rules to Z_n. Full deficiency convergence of the complete nonconvex experiment is not asserted.

Why the oracle equality is nontrivial. Three interfaces are needed. First, the Schur-complement correction gives actual structured paths and the inverse correction proves that every bounded local alternative lies within $O(n^{-1/2})$ of the tangent cone. Second, the ambient categorical likelihood is regular even though the admissible mean set is a nonconvex union of cones; mean-set convergence and LAN must therefore be proved separately. Third, LAN alone yields only a testing lower bound. Exact minimax equality uses

$$\text{finite interior Gaussian games} \longrightarrow \text{lower transfer}$$

and

$$\text{Lipschitz Gaussian rules applied to } Z_n \longrightarrow \text{upper transfer.}$$

The finite-path likelihood ratios are uniformly L^2-bounded, while Gaussian risk functions have a common modulus of continuity. These facts permit finite net exhaustion of $\mathcal{K}_R$ and a diagonal observable rule sequence.

3 Adaptive efficiency and its singularity barrier

Stable strata. Let $\mathcal{P}_{\mathrm{stab}}$ be the finite union of compact support-saturated, rank-saturated strata in Appendix J. Uniformly on this class: probabilities and design masses are bounded away from zero; active block norms and smallest active singular values have positive margins; inactive policies have a positive value gap; and the active value-gap map has full row rank on the model tangent, with a uniform Fisher transversality margin. These assumptions make the discrete stratum, tied face, and Fisher-normal comparison space statistically identifiable. Quantitatively, every center obeys common margins

$$\begin{aligned} &p_{\min}(P_0) \ge p_\star, \qquad w_{\min} \ge w_\star, \qquad |\Lambda(P_0)| = s, \\ &\min_{\rho\in\Lambda} \|B_\rho(P_0)\|_{\mathrm{F}} \ge \beta_\star, \qquad \min_{\rho\in\Lambda} \sigma_{r_\rho}(B_\rho(P_0)) \ge \kappa_\star, \\ &\min_{\pi\notin\Pi_0} \{J^\star(P_0) - J^\pi(P_0)\} \ge \Delta_\star, \qquad \lambda_{\min}(\mathsf{A}_{P_0}\mathsf{A}_{P_0}^{*,P_0}) \ge \tau_\star^2. \end{aligned} \tag{12}$$

Here A_{P_0} is the differential of the active value-gap map restricted to the smooth model tangent. The final inequality makes the tie set a uniformly regular submanifold and gives a well-conditioned Fisher-normal comparison coordinate.

Theorem 3.1 (Adaptive attainment on stable strata)**.** *For every fixed $R < \infty$, there exists one Borel randomized rule $\widehat{\delta}_{n,R}$, independent of the unknown center, its whitening, its support/rank stratum, and its tied face, such that*

$$\boxed{\sup_{P_0\in\mathcal{P}_{\mathrm{stab}}} \left| \sup_{P\in\Theta_{n,R}(P_0)} \mathbb{E}_P\left[\sqrt{n}\{J^\star(P) - J^{\widehat{\delta}_{n,R}}(P)\}\right] - \mathfrak{R}_R(\mathcal{K}_0(P_0), \mathfrak{G}_{0,\mathfrak{T}}(P_0)) \right| \to 0.} \tag{13}$$

Estimator and proof. Use a rowwise pilot fraction $\rho_n = n^{-1/3}$ and threshold $t_n = n^{-1/6}$. The procedure is:

1. symmetrize the pilot empirical kernel and threshold block norms, singular values, and candidate gaps to recover the support, ranks, and tied face;
2. project the pilot estimate onto the corresponding safe tie patch;
3. on the independent main sample, compute unrestricted and tie-constrained likelihood maximizers $\widehat{P}_{M,n}$ and $\widehat{P}_{T,n}$;
4. form the estimated Fisher-normal statistic
$$\widehat{Z}_n = \widehat{U}_n^*\widehat{\mathcal{I}}_n\sqrt{N_n}(\widehat{P}_{M,n} - \widehat{P}_{T,n});$$
5. estimate the relative gradients, solve the estimated Gaussian game to a vanishing tolerance, and apply its rule to $\widehat{Z}_n$.

Pilot recovery fails with probability at most $Ce^{-cn^{1/3}}$. Uniform quadratic likelihood expansions in adapted charts give the conditional normal limit of $\widehat{Z}_n$; the game-stability bound below transfers the

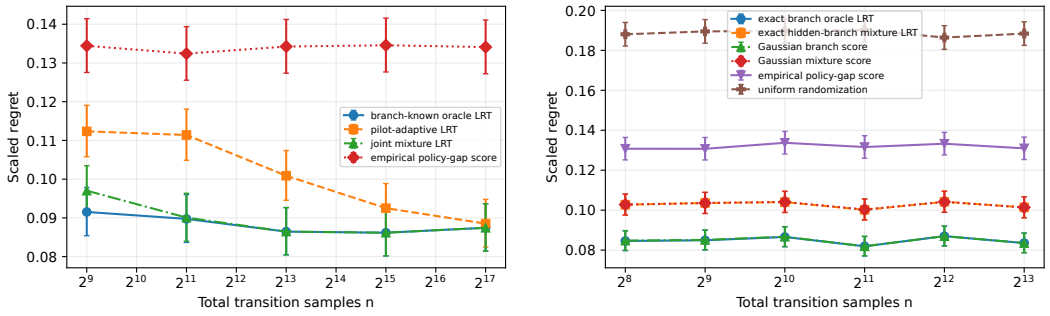

Figure 2: Finite transition-data manifestation of the phase transition. Left: on a stable rank-one stratum, the pilot-adaptive rule approaches the branch-known oracle. Right: at the singular intersection, the hidden-branch rule remains separated from the branch oracle. Bars are simultaneous exact 95% intervals over the pre-specified experiment suite.

estimated rule to the true game. A uniform finite-subexperiment lower bound rules out first-order superefficiency.

The proof separates four uniform statements. Stratum and face recovery is exponentially accurate; the unconstrained-minus-constrained MLE difference has a shifted Gaussian normal limit; estimated normal gradients converge after an orthogonal frame alignment; and a uniform finite-subexperiment lower bound precludes first-order superefficiency. The deterministic continuity bridge is quantitative. If compact mean sets are within Hausdorff distance ε after orthogonal alignment and matched gradients differ by at most η modulo a common translation, then

$$\boxed{|V(K,B) - V(\widetilde{K},\widetilde{B})| \le \left(2 + \sqrt{2/\pi}\, R\right) \Gamma\varepsilon + 2R\eta.} \tag{14}$$

Singular intersections. The stable margins cannot be removed uniformly. Appendix K constructs two smooth one-dimensional rank-one branches meeting at a rank-zero center. Their branch-specific efficient coordinates are noncollinear but statistically indistinguishable when the branch label is hidden.

Theorem 3.2 (A positive adaptation tax)**.** *For every $R > 0$ and $\vartheta \in (0, \pi/2)$, there is a strictly positive finite MDP with a C_2-equivariant rank-one transition model and two tied policies such that each branch oracle has limit*

$$\lim_n \mathcal{R}^{\mathrm{or,j}}_{n,R} = \frac{\cos\vartheta}{2} c_R, \qquad c_R = \max_{0 \le t \le R} 2t\Phi(-t), \tag{15}$$

but every branch-agnostic rule satisfies

$$\boxed{\liminf_n \left[\mathcal{R}^{\mathrm{ad}}_{n,R} - \max_j \mathcal{R}^{\mathrm{or,j}}_{n,R}\right] \ge \Delta_{\vartheta,R} > 0.} \tag{16}$$

Writing $t_R \in \arg\max_{t \le R} 2t\Phi(-t)$,

$$\Delta_{\vartheta,R} = \cos\vartheta\, t_R\{e_{\mathrm{mix}}(t_R) - \Phi(-t_R)\},$$

where hiding the noncollinear branch direction strictly increases the optimal sign-testing error. The same tax persists for smooth rank-one centers whose rank margin collapses faster than $n^{-1/2}$.

The strictness is explicit. If $d_j(z) = \varphi(z - t_R u_j) - \varphi(z + t_R u_j)$, then

$$e_{\mathrm{mix}}(t_R) - \Phi(-t_R) = \frac{1}{8}\int \left(|d_1| + |d_2| - |d_1 + d_2|\right) dz > 0. \tag{17}$$

Noncollinearity creates a positive-measure region on which d_1 and d_2 have opposite signs, so hiding the branch label strictly contracts total variation.

4 CERTIFIED COMPUTATION OF THE GAUSSIAN BENCHMARK

For a finite mean set $F = \{\theta_i\}_{i=1}^m$ and relative gradients $B = (b_a)_{a\in A}$, the Gaussian game has the exact least-favorable-prior dual

$$\boxed{V(F,B) = \max_{\lambda\in\Delta_m} \left\{ \sum_i \lambda_i h_i - \int \max_{a\in A} \left\langle b_a, \sum_i \lambda_i \varphi(z-\theta_i)\theta_i \right\rangle dz \right\},} \tag{18}$$

where $h_i = \max_a \langle b_a, \theta_i\rangle$. Projected supergradient ascent on the prior simplex, with averaged Bayes responses, returns lower and upper certificates with gap at most

$$2\sqrt{2}\,\Gamma R\sqrt{m/T} \tag{19}$$

after T iterations.

For an arbitrary compact $K \subseteq R\mathbb{B}_2$, Appendix M combines a parameter net, Gaussian smoothing, an observation grid, and one finite linear program. Its rule satisfies

$$\boxed{\mathcal{R}(K,B;q^{\mathrm{LP}}) \le V(K,B) + D_{\mathrm{obs}}(\sigma,h,M) + \Omega_{R,\Gamma}(\varepsilon_K),} \tag{20}$$

with all terms explicit; prescribed choices make the excess at most any $\varepsilon > 0$. An exchange variant alternates finite-game solution and worst-risk separation and terminates finitely because every nonterminal new mean must be separated from the current active set. The result is a certified finite algorithm, although no polynomial-time claim is made for arbitrary nonconvex sparse-rank unions.

Finite certificates. The observation grid partitions $[-M,M]^d$ into boxes. For a mean θ_i, each cell probability is a product of one-dimensional Gaussian cdf differences, so the discretized game is a literal finite LP. Its lower and upper errors decompose into Gaussian smoothing, within-cell Lipschitz error, tail truncation, and parameter-net error. The exchange implementation keeps finite lower and upper certificates at every iteration; before termination, every newly added worst-risk mean is separated from the active set by the inverse of the common Gaussian-risk modulus. Hence the method terminates after at most a packing-number number of additions. Sparse rank-cone nets use covering exponents governed by multiplicity-block ranks rather than only the ambient matrix dimension.

5 STRUCTURAL WIDTH AND FULL-POLICY RANK DIFFUSION

For an H-invariant task and an H-equivariant policy, subgroup-fixed spaces exclude raw derivative sectors and cap the rank of every potentially visible block. These statements transfer to whitened gradients only under rank-preserving separable Fisher geometry.

Theorem 5.1 (Task envelope and binary restricted width)**.** *For each representation sector,*

$$G_{\pi,\rho} = \sum_{t<T} D_{t,\pi,\rho}^{\top} W_{t+1,\pi,\rho}, \qquad \operatorname{rank}(G_{\pi,\rho}) \le (T-1)\dim V_\rho^H$$

(up to block dimensions), and $G_{\pi,\rho} = 0$ if $V_\rho^H = \{0\}$. For a binary tie with centered gradient g and a closed centrally symmetric cone $\mathcal{K}$,

$$c_R\omega_{\mathcal{K}}(g) \le \mathfrak{R}_R(\mathcal{K}, \operatorname{conv}\{-g,g\}) \le R\omega_{\mathcal{K}}(g). \tag{21}$$

At a maximally singular sparse block-rank center under separable Fisher geometry, the width is exactly

$$\boxed{\omega_{\mathcal{K}_0}(g)^2 = b_{\mathbf{1}}^2 + \sum_{\rho\in \mathrm{Top}_{\le s}(b)} \sum_{j\le r_\rho} \sigma_j(\widetilde{G}_\rho)^2.} \tag{22}$$

5.1 THE COMPLETE DETERMINISTIC POLICY CLASS

A strictly positive commitment lift turns the original comparison-gradient gadget into a full-policy MDP: an initial action selects a mode, all later state–time pairs have one mode-consistent action, and every inconsistency incurs a bounded penalty. Parameter-free commitment rows add no comparison gradient, while the uncertain rows preserve the original Fisher geometry up to one common scalar.

Table 1: Representative finite-sample results; intervals are simultaneous exact 95% intervals over the pre-specified suite.

Experiment/rule	Estimate	Simultaneous interval
Singular branch oracle, $n = 8192$	0.0835	$[0.0787, 0.0885]$
Hidden branch, $n = 8192$	0.1014	$[0.0962, 0.1067]$
Stable adaptive, $n = 131072$	0.0885	$[0.0825, 0.0948]$
Stable branch oracle	0.0874	$[0.0814, 0.0937]$
Diffuse cone-aware, $d = 8$	0.0227	$[0.0207, 0.0249]$
Diffuse raw gap score, $d = 8$	0.0412	$[0.0387, 0.0437]$

Theorem 5.2 (Full-policy rank-diffusion realization)**.** *For every fixed $d \geq r \geq 1$, there exist two finite-horizon MDP tasks with a common strictly positive C_2-equivariant transition model, common generative experiment, symmetry-breaking rewards, and the complete deterministic nonstationary Markov policy class such that: (i) exactly two deterministic policies are globally optimal; (ii) every other deterministic policy has a common positive base gap; (iii) the centered whitened comparison gradients are $\beta_d G_{d,r}^{\mathrm{conc}}$ and $\beta_d G_d^{\mathrm{flat}}$ for the same $\beta_d > 0$; and, for fixed $R > 0$,*

$$\boxed{\frac{c_R}{R}\sqrt{\frac{d}{r}} \leq \frac{\lim_n \widetilde{\mathcal{R}}_{n,R}^{\mathrm{conc,full}}}{\lim_n \widetilde{\mathcal{R}}_{n,R}^{\mathrm{flat,full}}} \leq \frac{R}{c_R}\sqrt{\frac{d}{r}}.} \tag{23}$$

The two tasks have the same complete base-policy value table and differ only in singular-value concentration relative to the common rank cone. The limit $n \to \infty$ is taken first for fixed (d, r).

Why the lift preserves the phenomenon. A commitment state chooses a global sign mode. Every later state–time pair has one mode-consistent action; all other actions incur a bounded penalty. Under the uniform base signal kernel, every mean-zero signal reward has zero conditional expectation, so a policy value is exactly minus one half of its expected number of consistency violations. Strict positivity gives every state–time pair positive occupancy, proving the global policy gap. The commitment rows are parameter free. On uncertain rows, both the Fisher form and the raw comparison derivative are multiplied by task-independent positive scalars, so concentrated and diffuse whitened gradients retain the same common factor β_d. The tasks therefore match the transition experiment, complete base-policy table, reward envelope, and gradient norm; only the singular-value profile relative to the rank cone changes.

6 Finite-sample evaluation

All observations are sampled from the exact categorical transition experiments; no Gaussian observations are simulated. The pre-specified suite contains 85 Bernoulli error estimands. Appendix N uses Bonferroni-adjusted exact binomial intervals, giving simultaneous coverage at least 0.95 over every reported risk and derived ratio.

At the singular rank-zero intersection with $R = 3$ and $\vartheta = \pi/3$, the branch-oracle and hidden-branch Gaussian limits are 0.08499 and 0.10380. At $n = 8192$, the exact categorical likelihood-ratio rules yield 0.0835 and 0.1014, while the empirical policy-gap score yields 0.1309. On stable rank-one branches, the pilot recovers the branch with probability 0.9845 at $n = 131072$; adaptive and branch-oracle risks are 0.0885 and 0.0874.

For the full-policy construction with $r = 1$, $d \in \{2, 4, 8\}$, and $n = 65536$, concentrated-to-diffuse ratios are $1.37, 1.92, 3.03$, versus $\sqrt{2}, 2, \sqrt{8}$. In the diffuse $d = 8$ task, the cone-aware score has risk 0.02274, compared with 0.04121 for the raw policy-gap score and 0.05271 for uniform randomization. These experiments illustrate the proved fixed-dimensional theory; they are not used in any proof.

7 Related work, scope, and conclusion

Closest antecedents. The table separates imported decision-theoretic ingredients from the new structured-control layer.

Prior line	Overlap	Additional result here
Treatment rules and best-arm regret (Hirano & Porter, 2009; Adusumilli, 2022)	Gaussian local games and minimax action choice	Nonconvex sparse-rank MDP cone, adaptive geometry estimation, and full-policy realization.
Nonregular maxima and nonunique decisions (Song, 2014; Luedtke & van der Laan, 2016; Zhu et al., 2024)	Ties and direction-dependent first-order behavior	Candidate-policy regret, stable-stratum attainment, and a positive adaptation tax.
Singular models and determinantal cones (Drton, 2009; Olikier et al., 2026)	Tangent-cone local asymptotics	Stochastic equivariant sparse union, quantitative actual-model approximation, and decision transfer.
Structured and symmetric RL (Jiang et al., 2017; Jin et al., 2020; Agarwal et al., 2020; van der Pol et al., 2020; Panangaden et al., 2024)	Global learnability, representations, and abstractions	Exact task-dependent local geometry and rank-diffusion difficulty.

What the rebuilt theorem package changes. The new results remove four weaknesses of the oracle-only formulation.

Previous limitation	New theorem	Remaining boundary
Center, Fisher geometry, stratum, and tied face known	One pilot/main-sample rule attains the oracle value uniformly on stable strata	Fixed margins and finite dimension.
No explanation of adaptation failure	A positive branch-agnostic excess-risk lower bound at a rank intersection	One explicit class of singular intersections.
Two preselected policies in the separation	Commitment lift with exactly two global deterministic optima and a positive gap to all others	Generative sampling and fixed (d, r).
Existential Gaussian minimax rule	Finite LP and exchange algorithms with certified error	Generic complexity may be exponential.

Interpretation. The stable theorem and the lower bound identify a statistical phase transition, not merely two unrelated examples. Fixed support, rank, and policy-face margins make the geometry a regular nuisance: it can be estimated on a vanishing pilot sample, while the main sample retains the full first-order comparison signal. When a rank margin collapses, the efficient direction itself becomes locally unidentified. The hidden direction then reduces total variation between the positive and negative policy-gap mixtures, producing a strict adaptation loss. The full-policy rank-diffusion family shows that even when the geometry is known, scalar summaries such as rank and gradient norm do not determine the oracle value.

Implications for structured reinforcement learning. The phase transition separates two kinds of structure that are often conflated. On a regular stratum, support, rank, and the active policy face are nuisance parameters with fixed margins. They can be recovered from a vanishing pilot sample, after which the remaining data solve the same first-order comparison problem as an oracle. At a singular intersection, by contrast, the efficient comparison direction is itself part of the unknown local parameter. Estimating that direction consumes information relevant to the decision and creates a strict loss. Thus the correct question is not simply whether an MDP has low rank, but whether its decision-relevant geometry is stably identifiable.

The full-policy construction gives a complementary message. Even after the geometry is known, equal model dimension, transition rank, Fisher spectrum, policy gap, and gradient norm need not imply equal difficulty. The controlling quantity is the energy of the comparison gradient that can be exposed by the feasible cone. For rank constraints this becomes top-r singular-value energy; for block-sparse models it additionally selects the most relevant representation sectors. Scalar complexity measures lose this alignment information.

The computational theorem makes these conclusions operational. The finite LP gives a globally certified benchmark, while the exchange method identifies least-favorable local alternatives. In the adaptive procedure the same solver is run on an estimated normal game. This provides a concrete blueprint for future trajectory-sampling work: the behavior policy should shape the Fisher metric so that decision-relevant directions are measured accurately, rather than merely maximizing undirected coverage.

Scope. The adaptive theorem is fixed-radius and fixed dimensional; it assumes finite spaces, a known finite group action and structural budgets, independent one-step generative samples, strictly positive kernels, real-Schur-type shared irreducibles, a finite candidate class, and quantitative stable-stratum margins. The lower bound proves that uniform oracle adaptation can fail when a rank margin collapses. The full-policy theorem removes the external-candidate restriction for the rank-diffusion phenomenon, but no trajectory-sampling, growing-dimensional, or nonasymptotic global minimax theorem is claimed. The certified Gaussian solver is finite but may be exponential in dimension.

Open directions. Under episodic trajectories, the design weights are replaced by feasible occupancies and the score becomes a martingale array. The natural extension is therefore a joint information-design problem: choose a behavior policy whose occupancy-weighted Fisher metric aligns with the policy-comparison gradients. The present results identify the objective but do not prove trajectory LAN or optimal adaptive exploration. Approximate equivariance raises a different question, because the symmetry bias and the root-n statistical radius compete on the same scale. Boundary kernels with zero probabilities may mix Gaussian support-preserving directions with Poisson activation directions, so strict positivity cannot simply be removed from the current proof.

A second direction is quantitative high-dimensional theory. The full-policy family takes $n \to \infty$ first for each fixed (d, r); its positivity margin, design mass, and horizon deteriorate with d. A joint theorem would require explicit experiment-distance bounds and a shorter realization. Finally, the certified Gaussian solver is generic but exponential. Exploiting the block-sparse rank geometry to obtain polynomial-time separation or exact least-favorable priors in invariant subclasses would connect the decision benchmark more directly to algorithms.

Conclusion. The local complexity of structured control is not determined by rank, dimension, symmetry counts, Fisher spectrum, policy gap, or gradient norm in isolation. It is governed by the relative Fisher geometry of feasible transition perturbations and policy-value gradients. On stable strata this geometry is learnable without first-order loss; at singular intersections the geometry itself becomes an unidentified nuisance and imposes a positive adaptation tax. In this precise sense,

$$\boxed{\text{rank is not enough: adaptive difficulty is geometric.}} \tag{24}$$

AI USE STATEMENT

Generative AI tools were used for research ideation, literature organization, exploratory conceptual development, candidate theorem formulation, proof-strategy brainstorming, reference formatting, source integration, code drafting, and language editing. All AI-assisted mathematical claims, proofs, constructions, citations, experiments, and computations in the final manuscript were subsequently re-derived, checked, and validated by the authors. Citations were checked against original or publisher sources, and the bundled regression tests and fixed-seed scripts were used to validate the displayed matrix identities, finite-MDP constructions, Gaussian-game algorithms, and finite-sample results. The authors take responsibility for the final content, including all text, claims, code, and artifacts produced with AI assistance.

REPRODUCIBILITY STATEMENT

Complete assumptions and proofs are given in the appendices. The supplementary source contains regression checks for the structured matrix paths, Bellman derivatives, local likelihood expansions, adaptive estimator, adaptation lower bound, full-policy realization, and Gaussian solver. A fixed-seed script reproduces every finite-sample figure and table from categorical transition data; Appendix N proves simultaneous exact coverage for all reported Monte Carlo intervals. No numerical result is used as a premise in a theorem proof.

REFERENCES

P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008. ISBN 978-0-691-13298-3. doi: 10.1515/9781400830244.

Karun Adusumilli. Neyman allocation is minimax optimal for best arm identification with two arms. *arXiv preprint arXiv:2204.05527*, 2022. URL https://arxiv.org/abs/2204.05527.

Alekh Agarwal, Sham Kakade, Akshay Krishnamurthy, and Wen Sun. FLAMBE: Structural complexity and representation learning of low rank MDPs. In *Advances in Neural Information Processing Systems 33*, pp. 20095–20107, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/e894d787e2fd6c133af47140aa156f00-Abstract.html.

C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934. doi: 10.1093/biomet/26.4.404.

Taco S. Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant CNNs on homogeneous spaces. In *Advances in Neural Information Processing Systems 32*, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/b9cfe8b6042cf759dc4c0cccb27a6737-Abstract.html.

Mathias Drton. Likelihood ratio tests and singularities. *The Annals of Statistics*, 37(2):979–1012, 2009. doi: 10.1214/07-AOS571.

Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. doi: 10.1007/BF02288367.

Charles J. Geyer. On the asymptotics of constrained m-estimation. *The Annals of Statistics*, 22(4):1993–2010, 1994. doi: 10.1214/aos/1176325768.

Rainer Hettich and Kenneth O. Kortanek. Semi-infinite programming: Theory, methods, and applications. *SIAM Review*, 35(3):380–429, 1993. doi: 10.1137/1035089.

Keisuke Hirano and Jack R. Porter. Asymptotics for statistical treatment rules. *Econometrica*, 77(5):1683–1701, 2009. doi: 10.3982/ECTA6630.

Nan Jiang, Akshay Krishnamurthy, Alekh Agarwal, John Langford, and Robert E. Schapire. Contextual decision processes with low Bellman rank are PAC-learnable. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1704–1713. PMLR, 2017. URL https://proceedings.mlr.press/v70/jiang17c.html.

Chi Jin, Zhuoran Yang, Zhaoran Wang, and Michael I. Jordan. Provably efficient reinforcement learning with linear function approximation. In *Proceedings of the Thirty-Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 2137–2143. PMLR, 2020. URL https://proceedings.mlr.press/v125/jin20a.html.

Patrick Kern, Axel Simroth, and Henryk Zähle. First-order sensitivity of the optimal value in a markov decision model with respect to deviations in the transition probability function. *Mathematical Methods of Operations Research*, 92(1):165–197, 2020. doi: 10.1007/s00186-020-00706-w. URL https://doi.org/10.1007/s00186-020-00706-w.

Lucien Le Cam. Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, 3(2):37–98, 1960.

Lucien Le Cam and Grace Lo Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer Series in Statistics. Springer, New York, 2 edition, 2000. doi: 10.1007/978-1-4612-1166-2.

Alexander R. Luedtke and Mark J. van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *The Annals of Statistics*, 44(2):713–742, 2016. doi: 10.1214/15-AOS1384.

Leon Mirsky. Symmetric gauge functions and unitarily invariant norms. *The Quarterly Journal of Mathematics*, 11(1):50–59, 1960. doi: 10.1093/qmath/11.1.50.

Leon Mirsky. A trace inequality of john von neumann. *Monatshefte für Mathematik*, 79:303–306, 1975. doi: 10.1007/BF01647331.

Jerzy Neyman and Egon S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694–706):289–337, 1933. doi: 10.1098/rsta.1933.0009.

Guillaume Olikier, Petar Mlinarić, P.-A. Absil, and André Uschmajew. The tangent cone to the real determinantal variety: Various expressions and a proof. *Set-Valued and Variational Analysis*, 34, 2026. doi: 10.1007/s11228-026-00796-4. Article 8.

Prakash Panangaden, Sahand Rezaei-Shoshtari, Rosie Zhao, David Meger, and Doina Precup. Policy gradient methods in the presence of symmetries and state abstractions. *Journal of Machine Learning Research*, 25(71):1–57, 2024. URL https://jmlr.org/papers/v25/23-1415.html.

R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 1998. doi: 10.1007/978-3-642-02431-3.

Reinhold Schneider and André Uschmajew. Convergence results for projected line-search methods on varieties of low-rank matrices via łojasiewicz inequality. *SIAM Journal on Optimization*, 25(1): 622–646, 2015. doi: 10.1137/140957822.

Jean-Pierre Serre. *Linear Representations of Finite Groups*, volume 42 of *Graduate Texts in Mathematics*. Springer, New York, 1977. doi: 10.1007/978-1-4684-9458-7.

Alexander Shapiro. On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66(3):477–487, 1990. doi: 10.1007/BF00940933.

Alexander Shapiro. On the asymptotics of constrained local m-estimators. *The Annals of Statistics*, 28(3):948–960, 2000. doi: 10.1214/aos/1015952006.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958. doi: 10.2140/pjm.1958.8.171. URL https://msp.org/pjm/1958/8-1/pjm-v8-n1-p14-s.pdf.

Kyungchul Song. Local asymptotic minimax estimation of nonregular parameters with translation-scale equivariant maps. *Journal of Multivariate Analysis*, 125:136–158, 2014. URL https://arxiv.org/abs/1403.2022.

Elise van der Pol, Daniel E. Worrall, Herke van Hoof, Frans A. Oliehoek, and Max Welling. MDP homomorphic networks: Group symmetries in reinforcement learning. In *Advances in Neural Information Processing Systems 33*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/2be5f9c2e3620eb73c2972d7552b6cb5-Abstract.html.

A. W. van der Vaart. *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. doi: 10.1017/CBO9780511802256.

John von Neumann. Some matrix-inequalities and metrization of matrix-space. *Tomsk University Review*, 1:286–300, 1937.

Yi Zhu, Jing Dong, and Henry Lam. Uncertainty quantification and exploration for reinforcement learning. *Operations Research*, 72(4):1689–1709, 2024. doi: 10.1287/opre.2023.2436.

A NOTATION, ASSUMPTIONS, AND THEOREM MAP

This appendix fixes the mathematical contract used by all subsequent proofs. All vector spaces are finite dimensional and real unless stated otherwise. We use the Bouligand–Peano contingent cone and finite-dimensional set convergence conventions of variational analysis (Rockafellar & Wets, 1998), the real isotypic decomposition of finite-group representations (Serre, 1977), and the local Gaussian-experiment framework of Le Cam–Yang and van der Vaart (Le Cam & Yang, 2000; van der Vaart, 1998). Randomized minimax arguments in Appendix F use the compact convex form of Sion's theorem (Sion, 1958).

A.1 GLOBAL CONVENTIONS

For a finite set E, $\mathbb{R}^E$ denotes the Euclidean space of real functions on E, identified with column vectors after an arbitrary but fixed ordering. For finite sets E, F, a matrix in $\mathbb{R}^{E\times F}$ has rows indexed by E and columns indexed by F. All matrix ranks are ordinary real ranks. Singular values are written in nonincreasing order,

$$\sigma_1(M) \geq \sigma_2(M) \geq \cdots \geq 0,$$

and are padded by zeros beyond $\min\{\#E, \#F\}$ when convenient. Empty sums are zero, $\min \varnothing := +\infty$, and the support budget $s = 0$ is allowed. If $b = (b_\rho)_\rho$ is a finite family of nonnegative numbers, $\mathrm{Top}_s(b)$ denotes any subset of at most s indices carrying the s largest values; all expressions used below depend only on the corresponding multiset of values and are therefore independent of tie breaking.

For a Euclidean space E, write

$$\mathbb{B}_E := \{x \in E : \|x\|_2 \leq 1\}, \qquad R\mathbb{B}_E := \{Rx : x \in \mathbb{B}_E\}.$$

When $E = \mathbb{R}^d$, we abbreviate $\mathbb{B}_E$ to $\mathbb{B}_2$. The Euclidean, Frobenius, Hilbert–Schmidt, operator, row-ℓ_1, and Fisher norms are written, respectively,

$$\|\cdot\|_2, \quad \|\cdot\|_{\mathrm{F}}, \quad \|\cdot\|_{\mathrm{HS}}, \quad \|\cdot\|_{\mathrm{op}}, \quad \|\cdot\|_{\infty,1}, \quad \|\cdot\|_{\mathcal{I}_{0,w}}.$$

All unqualified convergences of vectors or matrices are in any, hence every, finite-dimensional norm. For nonempty compact sets A, B in a normed space,

$$d_{\mathrm{H}}(A, B) := \max\left\{\sup_{a\in A}\inf_{b\in B}\|a-b\|, \sup_{b\in B}\inf_{a\in A}\|a-b\|\right\}. \tag{25}$$

A superscript on d_{H} specifies the norm. The bounded–Lipschitz metric between probability laws on $\mathbb{R}^d$ is

$$d_{\mathrm{BL}}(P, Q) := \sup_{\substack{\|f\|_\infty\leq 1\\ \mathrm{Lip}(f)\leq 1}} \left|\int f\, dP - \int f\, dQ\right|. \tag{26}$$

A.2 CONTROLLED PROCESS AND POLICY NOTATION

The controlled process is

$$M = \big(\mathcal{S}, \mathcal{A}, P, \{r_t\}_{t=1}^{T-1}, r_T, \mu\big), \qquad \mathcal{X} := \mathcal{S}\times\mathcal{A}, \tag{27}$$

where $\mathcal{S}$ and $\mathcal{A}$ are nonempty finite sets, $T \geq 2$, and $P \in [0,1]^{\mathcal{X}\times\mathcal{S}}$ is row stochastic. A deterministic nonstationary Markov policy is

$$\pi = (\pi_1, \ldots, \pi_{T-1}), \qquad \pi_t : \mathcal{S} \to \mathcal{A}.$$

The finite candidate class is denoted by Π, and the task object is

$$\mathfrak{T} := \big(\mu, (r_t)_{t<T}, r_T, \Pi\big). \tag{28}$$

The value and candidate-optimized value are

$$J^\pi(P) := \mathbb{E}_P^\pi\left[\sum_{t=1}^{T-1} r_t(S_t, A_t) + r_T(S_T)\right], \qquad J^\star_{\mathfrak{T}}(P) := \max_{\pi\in\Pi} J^\pi(P). \tag{29}$$

We suppress the task subscript when $\mathfrak{T}$ is fixed. At the base kernel P_0,

$$\Pi_0 := \arg\max_{\pi\in\Pi} J^\pi(P_0) \tag{30}$$

is the base-optimal face. For $q \in \Delta(\Pi)$, the value of the external policy lottery is

$$J^q(P) := \sum_{\pi\in\Pi} q(\pi) J^\pi(P). \tag{31}$$

No within-episode randomization is intended by this notation: a policy is sampled from q and then executed for the entire episode.

For a fixed policy π and the base kernel P_0, $d_t^\pi \in \Delta(\mathcal{X})$ denotes the stage-t state–action occupancy and V_t^π, Q_t^π denote the usual finite-horizon continuation values. The fixed-policy transition derivative is the linear functional

$$\ell_\pi(u) := \sum_{t=1}^{T-1} \sum_{x\in\mathcal{X}} d_t^\pi(x) \sum_{s'\in\mathcal{S}} u(s' \mid x) V_{t+1}^\pi(s'). \tag{32}$$

Appendix D proves that $\ell_\pi = DJ_{P_0}^\pi$.

A.3 Symmetry-adapted transition coordinates

A known finite group G acts on $\mathcal{S}$ and $\mathcal{A}$, diagonally on $\mathcal{X}$. The permutation representations are

$$(U_g^{\mathcal{S}} f)(s) := f(g^{-1}s), \qquad (U_g^{\mathcal{X}} q)(x) := q(g^{-1}x). \tag{33}$$

The conditional-expectation operator associated with P is

$$(K_P f)(x) := \sum_{s'\in\mathcal{S}} P(s' \mid x) f(s'). \tag{34}$$

Transition equivariance means

$$K_P U_g^{\mathcal{S}} = U_g^{\mathcal{X}} K_P \quad \forall g \in G, \tag{35}$$

equivalently $P(gs' \mid gx) = P(s' \mid x)$.

Throughout the core theorem package, every irreducible real representation appearing simultaneously in $\mathbb{R}^{\mathcal{S}}$ and $\mathbb{R}^{\mathcal{X}}$ is assumed to be of real Schur type. After fixing orthogonal isotypic coordinates,

$$K_P \cong \bigoplus_{\rho\in\widehat{G}_{\mathbb{R}}} I_{V_\rho} \otimes B_\rho(P), \qquad B_\rho(P) \in \mathbb{R}^{m_\rho^{\mathcal{X}} \times m_\rho^{\mathcal{S}}}. \tag{36}$$

The trivial representation is denoted by $\mathbf{1}$, and

$$\mathfrak{R} := \left\{ \rho \in \widehat{G}_{\mathbb{R}} \setminus \{\mathbf{1}\} : m_\rho^{\mathcal{X}} m_\rho^{\mathcal{S}} > 0 \right\} \tag{37}$$

indexes the nontrivial shared isotypic sectors. Stochastic normalization fixes only the image of the distinguished constant vector in the trivial multiplicity space. The remaining normalized trivial-isotypic directions form the nuisance space $\mathcal{N}_{\mathbf{1}}$ constructed in Appendix B.

Let $\bar{P}$ be the uniform kernel. Appendix B constructs a linear isomorphism

$$\mathcal{B}_G : \mathcal{Z}_G^0 := \mathcal{N}_{\mathbf{1}} \oplus \bigoplus_{\rho\in\mathfrak{R}} \mathbb{R}^{m_\rho^{\mathcal{X}} \times m_\rho^{\mathcal{S}}} \longrightarrow \mathcal{H}_G^0, \tag{38}$$

where

$$\mathcal{H}_G^0 := \left\{ u \in \mathbb{R}^{\mathcal{X}\times\mathcal{S}} : u\mathbf{1}_{\mathcal{S}} = 0,\ u(gs' \mid gx) = u(s' \mid x) \right\} \tag{39}$$

is the normalized equivariant perturbation space. Every normalized equivariant kernel has the unique form

$$P = \bar{P} + \mathcal{B}_G\big(Z_{\mathbf{1}}, (B_\rho)_\rho\big). \tag{40}$$

For budgets $s \in \{0, \dots, |\mathfrak{R}|\}$ and $0 \le r_\rho \le \min\{m_\rho^{\mathcal{X}}, m_\rho^{\mathcal{S}}\}$, define

$$\boxed{\Theta_{G,s,\boldsymbol{r}} := \left\{ P : \begin{array}{l} P \ge 0,\ P\mathbf{1}_{\mathcal{S}} = \mathbf{1}_{\mathcal{X}},\ P(gs' \mid gx) = P(s' \mid x), \\ |\{\rho \in \mathfrak{R} : B_\rho(P) \ne 0\}| \le s, \\ \operatorname{rank}(B_\rho(P)) \le r_\rho \quad \forall \rho \in \mathfrak{R} \end{array} \right\}.} \tag{41}$$

The support and rank constraints apply only to the nontrivial blocks; the normalized trivial nuisance block is unrestricted.

A.4 Local variational and Fisher geometry

Fix a base kernel $P_0 \in \Theta_{G,s,\boldsymbol{r}}$ satisfying

$$p_{\min} := \min_{x\in\mathcal{X},\, s'\in\mathcal{S}} P_0(s' \mid x) > 0. \tag{42}$$

The sampling design consists of independent samples $Y_{x,1:n_x} \overset{\text{iid}}{\sim} P(\cdot \mid x)$, independent across x, with

$$n := \sum_x n_x, \qquad w_{n,x} := n_x/n \to w_x, \qquad w_x \geq w_{\min} > 0, \qquad \sum_x w_x = 1. \tag{43}$$

The finite allocations need not be exactly G-invariant; only the limiting weights satisfy $w_{gx} = w_x$.

$$\max_x |w_{n,x} - w_x| \longrightarrow 0. \tag{44}$$

The Fisher inner product on $\mathcal{H}_G^0$ is

$$\langle u, v\rangle_{\mathcal{I}_{0,w}} := \sum_{x\in\mathcal{X}} w_x \sum_{s'\in\mathcal{S}} \frac{u(s' \mid x)v(s' \mid x)}{P_0(s' \mid x)}. \tag{45}$$

It is positive definite because $p_{\min}, w_{\min} > 0$. Set $d := \dim \mathcal{H}_G^0$ and fix any linear isometry

$$\mathsf{W}_0 : (\mathcal{H}_G^0, \langle\cdot,\cdot\rangle_{\mathcal{I}_{0,w}}) \longrightarrow (\mathbb{R}^d, \langle\cdot,\cdot\rangle_2). \tag{46}$$

For a subset S of a finite-dimensional Euclidean space and $x \in S$, the Bouligand tangent cone is

$$T_S(x) := \left\{ v : \exists\, t_k \downarrow 0,\ x_k \in S,\ x_k \to x,\ \frac{x_k - x}{t_k} \to v \right\}. \tag{47}$$

The structured and Fisher-whitened tangent cones are

$$\mathcal{C}_0 := T_{\Theta_{G,s,\boldsymbol{r}}}(P_0), \qquad \mathcal{K}_0 := \mathsf{W}_0\mathcal{C}_0. \tag{48}$$

For $R < \infty$, the finite-sample local model and its mean set are

$$\Theta_{n,R}(P_0) := \left\{ P \in \Theta_{G,s,\boldsymbol{r}} : \|\mathsf{W}_0\sqrt{n}(P - P_0)\|_2 \leq R \right\}, \tag{49}$$

$$\theta_n(P) := \mathsf{W}_0\sqrt{n}(P - P_0), \qquad \mathfrak{K}_{n,R} := \{\theta_n(P) : P \in \Theta_{n,R}(P_0)\}. \tag{50}$$

For each $\pi \in \Pi$, let $\gamma_\pi \in \mathbb{R}^d$ be the Fisher–Riesz representative of ℓ_π:

$$\ell_\pi(u) = \langle \gamma_\pi, \mathsf{W}_0 u\rangle_2 \quad \forall u \in \mathcal{H}_G^0. \tag{51}$$

The task-dependent Bellman gradient polytope and Bellman–Le Cam geometry are

$$\mathfrak{G}_{0,\mathfrak{T}} := \operatorname{conv}\{\gamma_\pi : \pi \in \Pi_0\}, \qquad \mathfrak{B}_{\mathfrak{T}}(P_0, w) := (\mathcal{K}_0, \mathfrak{G}_{0,\mathfrak{T}}). \tag{52}$$

Only $\mathcal{K}_0$ is determined by the transition experiment. We suppress $\mathfrak{T}$ from $\mathfrak{G}_{0,\mathfrak{T}}$ when the task and candidate class are fixed. For a compact convex set $C \subseteq \mathbb{R}^d$, its support function is

$$h_C(\theta) := \sup_{c\in C} \langle c, \theta\rangle_2. \tag{53}$$

A.5 Gaussian decision notation

A randomized Gaussian decision rule is a Borel map

$$\delta : \mathbb{R}^d \to \Delta(\Pi_0). \tag{54}$$

For $q \in \Delta(\Pi_0)$, define

$$\bar{\gamma}_q := \sum_{\pi\in\Pi_0} q(\pi)\gamma_\pi, \qquad L_{0,\mathfrak{T}}(\theta, q) := h_{\mathfrak{G}_{0,\mathfrak{T}}}(\theta) - \langle \bar{\gamma}_q, \theta\rangle_2. \tag{55}$$

If $Z \sim N(\theta, I_d)$, the radius-R Gaussian game is

$$\mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_{0,\mathfrak{T}}) := \inf_{\delta} \sup_{\theta\in\mathcal{K}_0\cap R\mathbb{B}_2} \mathbb{E}_\theta[L_{0,\mathfrak{T}}(\theta, \delta(Z))]. \tag{56}$$

The infimum ranges over all Borel randomized rules unless a theorem states a smaller class. A data-dependent rule δ_n is a Markov kernel from the full observation array to $\Delta(\Pi)$ or $\Delta(\Pi_0)$. Expectations in finite-sample local risks include both sample randomness and the internal randomization of δ_n. The corresponding fixed-radius finite-sample candidate-policy risk is

$$\mathcal{R}^{\text{all}}_{n,R}(P_0; \mathfrak{T}) := \inf_{\delta_n : \mathcal{Y}_n \rightsquigarrow \Delta(\Pi)} \sup_{P\in\Theta_{n,R}(P_0)} \mathbb{E}_P\left[\sqrt{n}\big(J^\star_{\mathfrak{T}}(P) - J^{\delta_n(Y)}(P)\big)\right], \tag{57}$$

where $\mathcal{Y}_n = \prod_x \mathcal{S}^{n_x}$ and the lottery is external.

A.6 STANDING ASSUMPTIONS AND THEOREM-SPECIFIC HYPOTHESES

Assumption A.1 (Finite controlled model and bounded rewards)**.** The sets $\mathcal{S}$, $\mathcal{A}$, the horizon T, and the candidate policy class Π are finite. The rewards and initial law are known, and

$$R_{\max} := \max_{t<T} \|r_t\|_\infty \vee \|r_T\|_\infty < \infty.$$

Assumption A.2 (Known exact symmetry and real Schur type)**.** The finite group G and its actions on $\mathcal{S}$ and $\mathcal{A}$ are known. The transition class is exactly G-equivariant. Every irreducible real representation appearing simultaneously in $\mathbb{R}^{\mathcal{S}}$ and $\mathbb{R}^{\mathcal{X}}$ is of real Schur type.

Assumption A.3 (Structured class and interior base point)**.** The unknown kernel belongs to $\Theta_{G,s,\boldsymbol{r}}$, and the fixed center $P_0 \in \Theta_{G,s,\boldsymbol{r}}$ satisfies $p_{\min} > 0$.

Assumption A.4 (Generative design)**.** For every $x \in \mathcal{X}$, the observations $Y_{x,1}, \dots, Y_{x,n_x}$ are i.i.d. from $P(\cdot \mid x)$ and independent across x. The deterministic allocation satisfies $w_{n,x} := n_x/n \to w_x \geq w_{\min} > 0$. Only the limiting design is assumed G-invariant: $w_{gx} = w_x$.

Assumption A.5 (Fixed-dimensional local regime)**.** The task $\mathfrak{T}$, sets, horizon, group, representation multiplicities, support and rank budgets, design limit, dimension, and local radius R are fixed while $n \to \infty$. The oracle benchmark at a center P_0 may use P_0, its Fisher isometry W_0, and its base-optimal face Π_0. The adaptive theorem instead uses one center-independent rule, uniformly over the quantitative stable class of Appendix J.

The assumptions above govern the central results in Sections 2–4. The following hypotheses are local to specific results and are not global:

1. the statewise max–plus derivative in Theorem D.11 takes Π to be the complete deterministic Markov policy class;
2. Theorem 5.1 assumes an H-invariant task and the existence of an H-equivariant deterministic policy;
3. the exact block score in Theorem 5.1 assumes a maximally singular center with vanishing nontrivial base blocks, a binary candidate-policy tie, the displayed support-budget cone, and a block-isotropic or left–right separable rank-preserving Fisher metric;
4. Theorem 5.2 uses its explicit C_2 construction and a two-policy candidate class;
5. Theorem 3.1 assumes the quantitative positivity, support, rank, inactive-policy-gap, and tie-transversality margins defining $\mathcal{P}_{\mathrm{stab}}$;
6. Theorem 3.2 uses two noncollinear smooth rank-one branch models meeting at a common rank-zero center;
7. Theorem 5.2 uses the explicit commitment lift and the complete deterministic nonstationary policy class;
8. the certified solver in Appendix M is finite and globally accurate but is not claimed polynomial time for arbitrary nonconvex mean sets.

Proposition A.6 (Coordinate invariance of the task-dependent Bellman–Le Cam geometry)**.** *Let $\widetilde{\mathsf{W}}_0$ be another Fisher isometry. Then there exists $O \in O(d)$ such that*

$$\widetilde{\mathsf{W}}_0 = O\mathsf{W}_0, \qquad \widetilde{\mathcal{K}}_0 = O\mathcal{K}_0, \qquad \widetilde{\gamma}_\pi = O\gamma_\pi, \qquad \widetilde{\mathfrak{G}}_{0,\mathfrak{T}} = O\mathfrak{G}_{0,\mathfrak{T}}. \tag{58}$$

For every R, the Gaussian Bellman game has the same value in the two coordinate systems. Hence $\mathfrak{B}_{\mathfrak{T}}(P_0, w)$ is intrinsic up to a common orthogonal transformation.

Proof. The map $O := \widetilde{\mathsf{W}}_0 \mathsf{W}_0^{-1}$ is a Euclidean isometry and hence orthogonal. The cone identity follows immediately. From

$$\ell_\pi(u) = \langle \gamma_\pi, \mathsf{W}_0 u \rangle = \langle O\gamma_\pi, O\mathsf{W}_0 u \rangle,$$

uniqueness of the Riesz representative gives $\widetilde{\gamma}_\pi = O\gamma_\pi$. Finally, if $Z \sim N(\theta, I_d)$, then $OZ \sim N(O\theta, I_d)$; transporting a decision rule by $z \mapsto O^\top z$ preserves every candidate-policy-regret risk. □

A.7 THEOREM MAP AND PROOF DEPENDENCIES

The main theorem package follows the chain

$$\boxed{\begin{array}{c}\text{equivariant stochastic coordinates} \implies \text{structured tangent cone} \\ \implies \text{Gaussian local experiment} \implies \text{minimax candidate-policy game}\end{array}} \tag{59}$$

The Bellman differential calculus supplies the loss attached to this statistical chain, while subgroup symmetry and matrix duality make the limit game interpretable and computable.

Main result	**Primary proof**	**Logical role**
Theorem 2.1	Appendices B–C	Exact sparse determinantal contingent cone.
Theorem 2.1	Appendix C	Common-domain rank-preserving stochastic paths with a uniform quadratic correction.
Theorem 2.1	Appendix C	Linear-rate convergence of the actual bounded local model to the tangent cone.
Theorem 2.3	Appendix D	Candidate-policy transition derivatives and support-function derivative of the candidate-optimized value.
Theorem 2.3	Appendix D	Uniform candidate-policy regret expansion along the structured feasible paths.
Theorem 2.4	Appendices C and E	Deterministic convergence of the admissible Gaussian mean sets.
Theorem 2.4	Appendix E	Uniform categorical LAN over the actual nonconvex local neighborhoods.
Theorem 2.4	Appendix E	Uniform shifted-central-sequence approximation and finite Gaussian subexperiment embedding.
Proposition 2.4	Appendices D–E	Uniform reduction of scaled MDP candidate-policy regret to the Gaussian loss.
Theorem 2.4	Appendix F	Matching lower and upper decision transfers and exact fixed-radius minimax equality.
Theorem 5.1	Appendix G	Exclusion and rank bounds for potentially visible raw derivative sectors.
Theorem 5.1	Appendix H	Explicit upper and lower minimax-risk bounds in terms of the restricted decision width.
Theorem 5.1	Appendix H	Exact top-s/top-r restricted-width formula under rank-preserving Fisher geometry.
Theorem 5.2 and Proposition 5.2	Appendix I	Exact finite-MDP comparison-gradient realization and matched-summary pair.
Theorem 5.2	Appendix I	Two-sided fixed-dimensional limiting-risk separation of order $\sqrt{d/r}$.
Theorem 3.1	Appendix J	Uniform unknown-center attainment of the oracle Gaussian value on stable support–rank–policy strata.
Theorem 3.2	Appendix K	A strictly positive adaptation gap at a singular rank intersection.
Theorem 5.2	Appendix L	Commitment-lifted realization with exactly two globally optimal policies.
Certified Gaussian-game computation	Appendix M	Finite dual, LP, and exchange certificates for arbitrary compact mean sets.
Finite-sample evaluation	Appendix N	Exact simultaneous intervals for adaptation and rank-diffusion experiments.

More explicitly, Appendix B supplies the affine coordinate map and Fisher geometry used by every later appendix. Appendix C and Appendix D are logically independent after this coordinate layer: the former characterizes feasible local directions, whereas the latter characterizes their first-order control effect. Appendix E combines both ingredients, and Appendix F converts the approximation into the oracle fixed-radius decision limit. Appendices G–I specialize and realize the structured Gaussian game. Appendix J proves that the oracle value is uniformly attainable on stable strata; Appendix K proves the singular adaptation barrier; Appendices L–N provide the full-policy realization, certified computation, and finite-sample evaluation.

A.8 SCOPE OF THE THEOREM PACKAGE

The oracle benchmark is fixed-radius, root-n, fixed dimensional, and pointwise at a strictly positive structured center. The adaptive theorem removes knowledge of the center, Fisher whitening, active support/ranks, and tied face uniformly on quantitative stable strata. The adaptation lower bound shows that such uniform attainment can fail at a singular rank intersection. The paper does not establish trajectory-constrained sampling, approximate equivariance, growing-dimensional LAN, a global nonasymptotic minimax theorem, or polynomial-time computation for arbitrary nonconvex Gaussian

games. Complex- and quaternionic-Schur-type blocks remain outside the coordinate contract. These are mathematical scope restrictions rather than informal simplifications.

B EQUIVARIANT STOCHASTIC TRANSITION OPERATORS

This appendix develops the linear and statistical coordinate system used throughout the paper. The representation-theoretic ingredients are classical (Serre, 1977); the same intertwining viewpoint underlies general equivariant linear architectures (Cohen et al., 2019) and group-symmetric Markov decision processes (van der Pol et al., 2020). Our contribution here is not a new form of Schur's lemma. Rather, we isolate the exact affine slice imposed by stochastic normalization, retain the full trivial-isotypic nuisance space for nontransitive actions, and identify precisely when the multinomial Fisher geometry respects the resulting multiplicity blocks.

Throughout this appendix, G is a finite group, $\mathcal{S}$ and $\mathcal{X}$ are nonempty finite G-sets, and

$$\mathcal{H}_{\mathcal{S}} := \mathbb{R}^{\mathcal{S}}, \qquad \mathcal{H}_{\mathcal{X}} := \mathbb{R}^{\mathcal{X}}$$

carry their standard Euclidean inner products. In the MDP application, $\mathcal{X} = \mathcal{S} \times \mathcal{A}$, but none of the results below requires a product structure. Matrices are indexed by rows $x \in \mathcal{X}$ and columns $s \in \mathcal{S}$.

B.1 FINITE-GROUP ACTIONS AND EQUIVARIANT KERNELS

For $g \in G$, define orthogonal permutation representations by

$$(U_g^{\mathcal{S}} f)(s) := f(g^{-1}s), \qquad (U_g^{\mathcal{X}} q)(x) := q(g^{-1}x). \tag{60}$$

For a matrix $P \in \mathbb{R}^{\mathcal{X}\times\mathcal{S}}$, let $K_P : \mathcal{H}_{\mathcal{S}} \to \mathcal{H}_{\mathcal{X}}$ denote the associated linear operator,

$$(K_P f)(x) := \sum_{s\in\mathcal{S}} P(s \mid x) f(s). \tag{61}$$

When $P(s \mid x) \geq 0$ and $\sum_s P(s \mid x) = 1$, this is the conditional-expectation operator of a Markov kernel.

Lemma B.1 (Equivalent formulations of transition equivariance)**.** *For $P \in \mathbb{R}^{\mathcal{X}\times\mathcal{S}}$, the following are equivalent:*

1. *$K_P U_g^{\mathcal{S}} = U_g^{\mathcal{X}} K_P$ for every $g \in G$;*
2. *$P(gs \mid gx) = P(s \mid x)$ for every $(g, x, s) \in G \times \mathcal{X} \times \mathcal{S}$;*
3. *P is constant on every orbit of the diagonal action $g \cdot (x, s) := (gx, gs)$ on $\mathcal{X} \times \mathcal{S}$.*

Proof. Let $e_s \in \mathcal{H}_{\mathcal{S}}$ be the indicator of s. Since $U_g^{\mathcal{S}} e_s = e_{gs}$,

$$(K_P U_g^{\mathcal{S}} e_s)(x) = P(gs \mid x), \qquad (U_g^{\mathcal{X}} K_P e_s)(x) = P(s \mid g^{-1}x).$$

Thus the intertwining identity is equivalent to $P(gs \mid x) = P(s \mid g^{-1}x)$ for every g, x, s, which becomes part 2 after replacing x by gx. Parts 2 and 3 are equivalent by the definition of an orbit. □

It is useful to regard equivariant matrices as fixed points of a representation on the full matrix space. Define

$$\mathcal{U}_g M := U_g^{\mathcal{X}} M U_{g^{-1}}^{\mathcal{S}}, \qquad M \in \mathbb{R}^{\mathcal{X}\times\mathcal{S}}. \tag{62}$$

Then $\mathcal{U} : G \to O(\mathbb{R}^{\mathcal{X}\times\mathcal{S}})$ under the Frobenius inner product, and

$$\mathrm{Hom}_G(\mathcal{H}_{\mathcal{S}}, \mathcal{H}_{\mathcal{X}}) = \mathrm{Fix}(\mathcal{U}) := \{M : \mathcal{U}_g M = M \ \forall g \in G\}. \tag{63}$$

B.2 REAL ISOTYPIC COORDINATES AND THE SCHUR NORMAL FORM

Assumption B.2 (Real-Schur-type core model)**.** Every irreducible real representation that occurs in both $\mathcal{H}_{\mathcal{S}}$ and $\mathcal{H}_{\mathcal{X}}$ is of real Schur type:

$$\mathrm{End}_G(V_\rho) = \mathbb{R} I_{V_\rho}.$$

Remark B.3 (Included and excluded examples). For $G = C_2$, both irreducible real representations are one dimensional and hence of real Schur type, so every permutation action of C_2 satisfies the assumption. By contrast, the two-dimensional real rotation representation of C_3 has commutant isomorphic to $\mathbb{C}$ (generated by the identity and the 90° complex structure after identifying the plane with $\mathbb{C}$), and is therefore of complex, not real, Schur type. A model in which that representation occurs on both $\mathcal{S}$ and $\mathcal{X}$ is excluded by the present theorem contract and requires complex multiplicity coordinates.

Because G is finite and the permutation representations are orthogonal, Maschke's theorem yields orthogonal isotypic decompositions. Fix orthogonal isomorphisms

$$\begin{aligned} \Psi_{\mathcal{S}} &: \bigoplus_{\rho \in \widehat{G}_{\mathbb{R}}} V_\rho \otimes \mathbb{R}^{m_\rho^{\mathcal{S}}} \longrightarrow \mathcal{H}_{\mathcal{S}}, \\ \Psi_{\mathcal{X}} &: \bigoplus_{\rho \in \widehat{G}_{\mathbb{R}}} V_\rho \otimes \mathbb{R}^{m_\rho^{\mathcal{X}}} \longrightarrow \mathcal{H}_{\mathcal{X}}. \end{aligned} \tag{64}$$

Terms with zero multiplicity are omitted. Write $d_\rho := \dim_{\mathbb{R}} V_\rho$.

Proposition B.4 (Real-type Schur normal form)**.** *Under Assumption B.2, a linear map $K : \mathcal{H}_{\mathcal{S}} \to \mathcal{H}_{\mathcal{X}}$ is G-equivariant if and only if there are unique matrices*

$$B_\rho(K) \in \mathbb{R}^{m_\rho^{\mathcal{X}} \times m_\rho^{\mathcal{S}}}$$

such that

$$\Psi_{\mathcal{X}}^\top K \Psi_{\mathcal{S}} = \bigoplus_{\rho \in \widehat{G}_{\mathbb{R}}} I_{V_\rho} \otimes B_\rho(K). \tag{65}$$

Consequently,

$$\dim \mathrm{Hom}_G(\mathcal{H}_{\mathcal{S}}, \mathcal{H}_{\mathcal{X}}) = \sum_{\rho \in \widehat{G}_{\mathbb{R}}} m_\rho^{\mathcal{X}} m_\rho^{\mathcal{S}}. \tag{66}$$

Moreover, if K, L have multiplicity blocks B_ρ, C_ρ, then

$$\langle K, L \rangle_{\mathrm{HS}} = \sum_{\rho \in \widehat{G}_{\mathbb{R}}} d_\rho \, \langle B_\rho, C_\rho \rangle_{\mathrm{F}}, \qquad \|K\|_{\mathrm{HS}}^2 = \sum_{\rho \in \widehat{G}_{\mathbb{R}}} d_\rho \, \|B_\rho\|_{\mathrm{F}}^2. \tag{67}$$

Proof. Write $K_{\sigma\rho}$ for the component of $\Psi_{\mathcal{X}}^\top K \Psi_{\mathcal{S}}$ from $V_\rho \otimes \mathbb{R}^{m_\rho^{\mathcal{S}}}$ to $V_\sigma \otimes \mathbb{R}^{m_\sigma^{\mathcal{X}}}$. Relative to multiplicity bases, every matrix entry of $K_{\sigma\rho}$ is a G-intertwiner $V_\rho \to V_\sigma$. Schur's lemma gives $K_{\sigma\rho} = 0$ when $\rho \not\cong \sigma$. When $\rho = \sigma$, Assumption B.2 implies that every such entry is a real scalar multiple of I_{V_ρ}. Collecting these scalars gives a unique matrix $B_\rho(K)$ and proves equation 65. Conversely, every operator of the displayed form plainly intertwines the two group actions. The dimension formula follows by counting the free entries of the B_ρ's.

Orthogonality of $\Psi_{\mathcal{S}}$ and $\Psi_{\mathcal{X}}$ preserves the Hilbert–Schmidt inner product. Different isotypic summands are orthogonal, and

$$\left\langle I_{V_\rho} \otimes B_\rho, I_{V_\rho} \otimes C_\rho \right\rangle_{\mathrm{HS}} = \mathrm{tr}(I_{V_\rho}) \, \mathrm{tr}(B_\rho^\top C_\rho) = d_\rho \, \langle B_\rho, C_\rho \rangle_{\mathrm{F}}.$$

Summing over ρ proves equation 67. □

Remark B.5 (Multiplicity-basis gauge). Replacing the multiplicity bases by orthogonal matrices $Q_\rho^{\mathcal{S}}$ and $Q_\rho^{\mathcal{X}}$ transforms

$$B_\rho(K) \longmapsto (Q_\rho^{\mathcal{X}})^\top B_\rho(K) Q_\rho^{\mathcal{S}}.$$

Hence the zero/nonzero status, rank, singular values, and Frobenius norm of a multiplicity block are intrinsic. These are exactly the block quantities used in the support and rank constraints of the structured model.

Remark B.6 (Why the real-type restriction is explicit). For a general irreducible real representation, $\mathrm{End}_G(V_\rho)$ is isomorphic to one of $\mathbb{R}, \mathbb{C}$, or the quaternions $\mathbb{H}$; the corresponding multiplicity maps are matrices over that division algebra (Serre, 1977). The ordinary real-matrix formula equation 65 is therefore not valid verbatim for complex or quaternionic Schur type. The core paper restricts to real type rather than silently conflating these cases.

B.3 Stochastic normalization and the full trivial-isotypic nuisance

Let $\mathbf{1}$ denote the one-dimensional trivial representation. The G-fixed subspace of $\mathbb{R}^Y$ for a finite G-set Y is denoted $(\mathbb{R}^Y)^G$.

Lemma B.7 (Invariant functions and orbit multiplicity)**.** *For every finite G-set Y,*

$$(\mathbb{R}^Y)^G = \{f : Y \to \mathbb{R} : f \text{ is constant on each } G\text{-orbit}\}.$$

Consequently, the multiplicity $m_{\mathbf{1}}^Y$ of the trivial representation equals the number of G-orbits in Y.

Proof. The identity $U_g^Y f = f$ for all g is equivalent to $f(gy) = f(y)$ for all g, y, which is precisely constancy on each orbit. The orbit-indicator functions form a basis of this fixed subspace. □

Let

$$c_{\mathcal{S}} := \Psi_{\mathcal{S},\mathbf{1}}^\top \mathbf{1}_{\mathcal{S}} \in \mathbb{R}^{m_{\mathbf{1}}^{\mathcal{S}}}, \qquad c_{\mathcal{X}} := \Psi_{\mathcal{X},\mathbf{1}}^\top \mathbf{1}_{\mathcal{X}} \in \mathbb{R}^{m_{\mathbf{1}}^{\mathcal{X}}} \tag{68}$$

be the multiplicity coordinates of the all-ones vectors. Since the coordinate maps are orthogonal,

$$\|c_{\mathcal{S}}\|_2 = \sqrt{|\mathcal{S}|}, \qquad \|c_{\mathcal{X}}\|_2 = \sqrt{|\mathcal{X}|}. \tag{69}$$

Proposition B.8 (Exact stochastic affine slice)**.** *Let K be G-equivariant, with multiplicity blocks $B_\rho(K)$. Then*

$$K\mathbf{1}_{\mathcal{S}} = \mathbf{1}_{\mathcal{X}} \quad \Longleftrightarrow \quad B_{\mathbf{1}}(K)c_{\mathcal{S}} = c_{\mathcal{X}}. \tag{70}$$

Choose normalization-adapted orthonormal multiplicity bases whose first vectors are

$$e_{\mathcal{S}} := c_{\mathcal{S}}/\sqrt{|\mathcal{S}|}, \qquad e_{\mathcal{X}} := c_{\mathcal{X}}/\sqrt{|\mathcal{X}|}.$$

Then equation 70 is equivalent to

$$B_{\mathbf{1}}(K) = \begin{bmatrix} \beta & a^\top \\ 0 & C \end{bmatrix}, \qquad \beta := \sqrt{\frac{|\mathcal{X}|}{|\mathcal{S}|}}, \tag{71}$$

where a and C are unrestricted and empty blocks are allowed. The linear space of normalized equivariant perturbations is

$$\mathcal{H}_G^0 := \{U \in \mathrm{Hom}_G(\mathcal{H}_{\mathcal{S}}, \mathcal{H}_{\mathcal{X}}) : U\mathbf{1}_{\mathcal{S}} = 0\}, \tag{72}$$

whose trivial multiplicity block belongs to

$$\mathcal{N}_{\mathbf{1}} := \{\Delta_{\mathbf{1}} : \Delta_{\mathbf{1}} c_{\mathcal{S}} = 0\}. \tag{73}$$

In normalization-adapted coordinates,

$$\Delta_{\mathbf{1}} = \begin{bmatrix} 0 & b^\top \\ 0 & D \end{bmatrix}. \tag{74}$$

Finally,

$$\begin{aligned} \dim \mathcal{N}_{\mathbf{1}} &= m_{\mathbf{1}}^{\mathcal{X}}(m_{\mathbf{1}}^{\mathcal{S}} - 1), \\ \dim \mathcal{H}_G^0 &= m_{\mathbf{1}}^{\mathcal{X}}(m_{\mathbf{1}}^{\mathcal{S}} - 1) + \sum_{\rho \neq \mathbf{1}} m_\rho^{\mathcal{X}} m_\rho^{\mathcal{S}}. \end{aligned} \tag{75}$$

Proof. The vector $\mathbf{1}_{\mathcal{S}}$ is G-fixed and hence has no component in a nontrivial isotypic summand. Applying the block form equation 65 to $\mathbf{1}_{\mathcal{S}}$ gives

$$\Psi_{\mathcal{X}}^\top K \mathbf{1}_{\mathcal{S}} = B_{\mathbf{1}}(K) c_{\mathcal{S}}$$

inside the trivial multiplicity space and zero in all nontrivial summands. This equals the coordinate vector of $\mathbf{1}_{\mathcal{X}}$ if and only if equation 70 holds.

In the adapted bases, the constraint reads $B_{\mathbf{1}}(K)e_{\mathcal{S}} = \beta e_{\mathcal{X}}$, so the first column is fixed to $(\beta, 0, \ldots, 0)^\top$ and all remaining columns are free. This is equation 71; subtracting two normalized operators gives equation 74.

The linear map $T : \mathbb{R}^{m_{\mathbf{1}}^{\mathcal{X}} \times m_{\mathbf{1}}^{\mathcal{S}}} \to \mathbb{R}^{m_{\mathbf{1}}^{\mathcal{X}}}$ defined by $T(\Delta) = \Delta c_{\mathcal{S}}$ is surjective because $T(y c_{\mathcal{S}}^\top / \|c_{\mathcal{S}}\|_2^2) = y$. Therefore its kernel has dimension $m_{\mathbf{1}}^{\mathcal{X}} m_{\mathbf{1}}^{\mathcal{S}} - m_{\mathbf{1}}^{\mathcal{X}}$, proving the first line of equation 75. Proposition B.4 gives the remaining summands. □

Remark B.9 (Why the trivial block cannot be discarded). If both G-actions are transitive, then $m_{\mathbf{1}}^{\mathcal{S}} = m_{\mathbf{1}}^{\mathcal{X}} = 1$ and $\mathcal{N}_{\mathbf{1}} = \{0\}$. For nontransitive actions, however, invariant functions include independent orbit indicators, and stochasticity fixes only the image of the single all-ones vector. The remaining trivial-isotypic directions are genuine nuisance parameters and must be retained.

B.4 AFFINE BLOCK COORDINATES AND LOCAL POSITIVITY

Define the uniform Markov kernel

$$\bar{P}(s \mid x) := \frac{1}{|\mathcal{S}|}, \qquad (K_{\bar{P}} f)(x) = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} f(s). \tag{76}$$

It is strictly positive, row stochastic, and equivariant under arbitrary G-actions. Its only nonzero multiplicity block is

$$B_{\mathbf{1}}(\bar{P}) = \frac{c_{\mathcal{X}} c_{\mathcal{S}}^{\top}}{|\mathcal{S}|}. \tag{77}$$

Let

$$\mathfrak{R} := \{\rho \in \widehat{G}_{\mathbb{R}} \setminus \{\mathbf{1}\} : m_\rho^{\mathcal{S}} m_\rho^{\mathcal{X}} > 0\} \tag{78}$$

and define the block-coordinate space

$$\mathcal{Z}_G^0 := \mathcal{N}_{\mathbf{1}} \oplus \bigoplus_{\rho \in \mathfrak{R}} \mathbb{R}^{m_\rho^{\mathcal{X}} \times m_\rho^{\mathcal{S}}}. \tag{79}$$

For $\mathbf{\Delta} = (\Delta_{\mathbf{1}}, (\Delta_\rho)_\rho) \in \mathcal{Z}_G^0$, define the reconstruction map

$$\mathcal{B}_G(\mathbf{\Delta}) := \Psi_{\mathcal{X}} \left[(I_{V_{\mathbf{1}}} \otimes \Delta_{\mathbf{1}}) \oplus \bigoplus_{\rho \in \mathfrak{R}} (I_{V_\rho} \otimes \Delta_\rho) \right] \Psi_{\mathcal{S}}^{\top}. \tag{80}$$

Proposition B.10 (Global affine parameterization)**.** *The map $\mathcal{B}_G : \mathcal{Z}_G^0 \to \mathcal{H}_G^0$ is a linear isomorphism. Every row-normalized equivariant matrix has a unique representation*

$$P = \bar{P} + \mathcal{B}_G(\mathbf{\Delta}), \qquad \mathbf{\Delta} \in \mathcal{Z}_G^0. \tag{81}$$

Moreover,

$$\|\mathcal{B}_G(\mathbf{\Delta})\|_{\mathrm{HS}}^2 = \|\Delta_{\mathbf{1}}\|_{\mathrm{F}}^2 + \sum_{\rho \in \mathfrak{R}} d_\rho \|\Delta_\rho\|_{\mathrm{F}}^2. \tag{82}$$

The set of G-equivariant Markov kernels is the compact convex polytope

$$\mathcal{P}_G = \left\{ \bar{P} + \mathcal{B}_G(\mathbf{\Delta}) : \mathbf{\Delta} \in \mathcal{Z}_G^0,\ \bar{P} + \mathcal{B}_G(\mathbf{\Delta}) \geq 0 \right\}. \tag{83}$$

Proof. Propositions B.4 and B.8 show that $\mathcal{B}_G$ is a bijection onto the normalized equivariant perturbation space. If P is normalized and equivariant, then $P - \bar{P} \in \mathcal{H}_G^0$, yielding the unique representation equation 81. Formula equation 82 is equation 67 restricted to $\mathcal{H}_G^0$. Finally, nonnegativity is a finite collection of linear inequalities in $\mathbf{\Delta}$. The resulting set is convex and closed; it is bounded because every normalized nonnegative row belongs to a probability simplex. □

Proposition B.11 (Strict positivity makes inequality constraints locally inactive)**.** *Let $P_0 \in \mathcal{P}_G$ satisfy*

$$p_{\min} := \min_{x \in \mathcal{X},\, s \in \mathcal{S}} P_0(s \mid x) > 0. \tag{84}$$

Then, for every $U \in \mathcal{H}_G^0$,

$$P_0 + tU \in \mathcal{P}_G \quad \textit{whenever} \quad |t| \, \|U\|_{\max} \leq p_{\min}. \tag{85}$$

Now suppose a design $w = (w_x)_{x \in \mathcal{X}}$ satisfies $w_x \geq w_{\min} > 0$, and define the Fisher norm in equation 89 below. Then

$$\|U\|_{\max} \leq \|U\|_{\mathrm{F}} \leq w_{\min}^{-1/2} \|U\|_{\mathcal{I}_{0,w}}. \tag{86}$$

Consequently, for every $R > 0$,

$$\|U\|_{\mathcal{I}_{0,w}} \leq R, \qquad |t| \leq \frac{p_{\min} \sqrt{w_{\min}}}{2R} \quad \Longrightarrow \quad P_0 + tU \geq \frac{p_{\min}}{2} \tag{87}$$

entrywise.

Proof. The matrix $P_0 + tU$ remains equivariant and normalized because $U \in \mathcal{H}_G^0$. If $|t| \, \|U\|_{\max} \leq p_{\min}$, then every entry is nonnegative, proving equation 85. Since $P_0(s \mid x) \leq 1$,

$$\|U\|_{\mathcal{I}_{0,w}}^2 = \sum_x w_x \sum_s \frac{U(s \mid x)^2}{P_0(s \mid x)} \geq w_{\min} \sum_{x,s} U(s \mid x)^2 = w_{\min} \|U\|_{\mathrm{F}}^2 .$$

Together with $\|U\|_{\max} \leq \|U\|_{\mathrm{F}}$, this proves equation 86. The final implication follows by substitution, with a factor $1/2$ retained as a positivity margin. □

Corollary B.12 (Uniform positivity for second-order model paths)**.** *Fix $R, C > 0$. Suppose that $U \in \mathcal{H}_G^0$, $\|U\|_{\mathcal{I}_{0,w}} \leq R$, and that $E(t) \in \mathcal{H}_G^0$ obeys $\|E(t)\|_{\max} \leq Ct^2$. Then there exists $t_0 = t_0(p_{\min}, w_{\min}, R, C) > 0$ such that*

$$P_0 + tU + E(t) \in \mathcal{P}_G \qquad \forall |t| \leq t_0.$$

The choice of t_0 is uniform over all such U and E.

Proof. Choose $t_0 > 0$ such that $t_0 R / \sqrt{w_{\min}} + Ct_0^2 \leq p_{\min}$. The normalization and equivariance constraints hold because $U, E(t) \in \mathcal{H}_G^0$, while the entrywise perturbation from P_0 is at most $p_{\min}$. □

B.5 Fisher geometry and its block-coordinate pullback

Let $P_0 \in \mathcal{P}_G$ be strictly positive, and let $w_x > 0$, $\sum_x w_x = 1$, be the asymptotic generative-design weights. On the full row-sum-zero space

$$\mathcal{H}^0 := \{U \in \mathbb{R}^{\mathcal{X} \times \mathcal{S}} : U\mathbf{1}_{\mathcal{S}} = 0\}, \tag{88}$$

define

$$\langle U, V \rangle_{\mathcal{I}_{0,w}} := \sum_{x \in \mathcal{X}} w_x \sum_{s \in \mathcal{S}} \frac{U(s \mid x) V(s \mid x)}{P_0(s \mid x)}. \tag{89}$$

This is the Fisher bilinear form of the product multinomial experiment on valid row-sum-zero directions.

Proposition B.13 (Fisher invariance and the equivariant fixed subspace)**.** *Assume that*

$$w_{gx} = w_x, \qquad P_0(gs \mid gx) = P_0(s \mid x) \quad \forall g, x, s. \tag{90}$$

Then:

1. *$\langle \cdot, \cdot \rangle_{\mathcal{I}_{0,w}}$ is an inner product on $\mathcal{H}^0$ and on $\mathcal{H}_G^0$;*
2. *the action $\mathcal{U}_g$ from equation 62 preserves $\mathcal{H}^0$ and is orthogonal for this Fisher inner product;*
3. *$\mathcal{H}_G^0 = (\mathcal{H}^0)^G$ is the fixed-point subspace of that action;*
4. *there exists a linear Fisher isometry*

$$\mathsf{W}_0 : (\mathcal{H}_G^0, \langle \cdot, \cdot \rangle_{\mathcal{I}_{0,w}}) \longrightarrow (\mathbb{R}^d, \langle \cdot, \cdot \rangle), \qquad d := \dim \mathcal{H}_G^0.$$

Proof. Strict positivity of P_0 and positivity of every w_x make equation 89 positive definite on the full matrix space, hence on either subspace. Since $U^{\mathcal{S}}_{g^{-1}} \mathbf{1}_{\mathcal{S}} = \mathbf{1}_{\mathcal{S}}$, the relation $U\mathbf{1}_{\mathcal{S}} = 0$ implies $(\mathcal{U}_g U)\mathbf{1}_{\mathcal{S}} = 0$. By a change of variables $(x, s) \mapsto (gx, gs)$ and equation 90,

$$\begin{aligned} \langle \mathcal{U}_g U, \mathcal{U}_g V \rangle_{\mathcal{I}_{0,w}} &= \sum_{x,s} w_x \frac{U(g^{-1}s \mid g^{-1}x) V(g^{-1}s \mid g^{-1}x)}{P_0(s \mid x)} \\ &= \sum_{x,s} w_x \frac{U(s \mid x) V(s \mid x)}{P_0(s \mid x)}. \end{aligned}$$

Part 3 is equation 63 intersected with the row-sum-zero space. Part 4 follows by choosing an orthonormal basis of the finite-dimensional Fisher inner-product space. □

Pull the Fisher form back through the block reconstruction map. Equip $\mathcal{Z}_G^0$ with the unweighted direct-sum Frobenius inner product

$$\langle \mathbf{\Delta}, \mathbf{\Gamma} \rangle_{\mathcal{Z}} := \langle \Delta_{\mathbf{1}}, \Gamma_{\mathbf{1}} \rangle_{\mathrm{F}} + \sum_{\rho \in \mathfrak{R}} \langle \Delta_\rho, \Gamma_\rho \rangle_{\mathrm{F}} \,. \tag{91}$$

There is a unique positive-definite self-adjoint operator $\mathbf{F}_{0,w} : \mathcal{Z}_G^0 \to \mathcal{Z}_G^0$ such that

$$\langle \mathcal{B}_G(\mathbf{\Delta}), \mathcal{B}_G(\mathbf{\Gamma}) \rangle_{\mathcal{I}_{0,w}} = \langle \mathbf{\Delta}, \mathbf{F}_{0,w} \mathbf{\Gamma} \rangle_{\mathcal{Z}} \,. \tag{92}$$

In general, $\mathbf{F}_{0,w}$ need not be block diagonal in the irrep label. The next example demonstrates that this failure can occur even when every irreducible representation is one dimensional.

Example B.14 (Symmetry does not force block-orthogonal Fisher information)**.** Let $G = C_2 \times C_2 = \{e, a, b, c\}$ act regularly on $\mathcal{S} = \mathcal{X} = G$. Every equivariant Markov kernel is a convolution kernel

$$P_0(y \mid x) = p(x^{-1}y).$$

Take the uniform design $w_x = 1/4$ and

$$(p(e), p(a), p(b), p(c)) = \left(\frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right).$$

An equivariant row-sum-zero perturbation has the form $U_q(y \mid x) = q(x^{-1}y)$ with $\sum_h q(h) = 0$. Let $\chi_0, \chi_1, \chi_2, \chi_3$ be the four real characters, ordered so that their character table is

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

The actual scalar intertwiner block in character χ_i is $b_i := \sum_h \chi_i(h) q(h)$; normalization gives $b_0 = 0$. Since $q = H^\top b/4$,

$$\|U_q\|^2_{\mathcal{I}_{0,w}} = \sum_{h \in G} \frac{q(h)^2}{p(h)} = \frac{1}{4} \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \tag{93}$$

Thus the Fisher form couples three inequivalent one-dimensional irreducible blocks. Group invariance of the sampling experiment is therefore insufficient to justify blockwise isotropy or even blockwise orthogonality.

Proof. For each x, the change of variables $h = x^{-1}y$ gives

$$\sum_y \frac{U_q(y \mid x)^2}{P_0(y \mid x)} = \sum_h \frac{q(h)^2}{p(h)},$$

and averaging over the four values of x leaves the same expression. With $D = \mathrm{diag}(2, 6, 6, 6)$ and $q = H^\top b/4$, this quadratic form is $b^\top H D H^\top b/16$. Direct multiplication gives diagonal entries $5/4$ and off-diagonal entries $-1/4$ on the nontrivial-character subspace, which is equation 93. □

The mechanism behind Example B.14 is structural: although the Fisher operator commutes with $\mathcal{U}_g$ on the full matrix space, the equivariant parameter space is the fixed-point space of $\mathcal{U}$; the induced group action on that fixed space is trivial and therefore imposes no further Schur-type restriction on its positive-definite Fisher operator.

B.6 A BLOCK-ISOTROPIC REGIME WITH RANK-PRESERVING WHITENING

The exact top-s/top-r formula in Section 5 requires a Fisher geometry whose whitening preserves block support and matrix rank. The following proposition gives a transparent sufficient condition and is the regime used by the realizable separation in Section 5.

Proposition B.15 (Uniform base kernel and balanced design)**.** *Suppose*

$$P_0(s \mid x) = \frac{1}{|\mathcal{S}|} \quad \forall x, s, \qquad w_x = \frac{1}{|\mathcal{X}|} \quad \forall x. \tag{94}$$

Let $U = \mathcal{B}_G(\mathbf{\Delta})$ and $V = \mathcal{B}_G(\mathbf{\Gamma})$ belong to $\mathcal{H}_G^0$. Then

$$\boxed{\langle U, V\rangle_{\mathcal{I}_{0,w}} = \frac{|\mathcal{S}|}{|\mathcal{X}|}\left[\langle \Delta_{\mathbf{1}}, \Gamma_{\mathbf{1}}\rangle_{\mathrm{F}} + \sum_{\rho\in\mathfrak{R}} d_\rho \langle \Delta_\rho, \Gamma_\rho\rangle_{\mathrm{F}}\right].} \tag{95}$$

Consequently, the block map

$$\mathsf{W}_{\mathrm{blk}}(U) := \left(\sqrt{\frac{|\mathcal{S}|}{|\mathcal{X}|}}\,\Delta_{\mathbf{1}}, \left(\sqrt{\frac{|\mathcal{S}|d_\rho}{|\mathcal{X}|}}\,\Delta_\rho\right)_{\rho\in\mathfrak{R}}\right) \tag{96}$$

is a Fisher isometry into the Euclidean direct sum of the block-coordinate spaces. It preserves the support of the nontrivial blocks and the rank of every block.

Proof. Under equation 94,

$$\langle U, V\rangle_{\mathcal{I}_{0,w}} = \frac{1}{|\mathcal{X}|}\sum_x\sum_s |\mathcal{S}| U(s \mid x) V(s \mid x) = \frac{|\mathcal{S}|}{|\mathcal{X}|}\langle U, V\rangle_{\mathrm{HS}}\,.$$

Substituting the Hilbert–Schmidt decomposition equation 67 gives equation 95. Equation equation 96 is therefore an isometry. Each block is multiplied by a nonzero scalar, which neither activates a zero block nor changes matrix rank. □

Corollary B.16 (Exact block weights for the structured-width theorem)**.** *In the setting of Proposition B.15, the block-isotropic Fisher coefficients in Theorem 5.1 may be chosen as*

$$\alpha_\rho = \frac{|\mathcal{S}|d_\rho}{|\mathcal{X}|} \quad (\rho \in \mathfrak{R}), \tag{97}$$

while the unrestricted trivial nuisance block has coefficient $|\mathcal{S}|/|\mathcal{X}|$ under its inherited Frobenius coordinates.

Remark B.17 (General Fisher whitening can destroy low rank). For arbitrary strictly positive P_0 and design w, a Fisher isometry W_0 always exists by Proposition B.13, but its pullback in multiplicity coordinates may mix irrep blocks or apply a general invertible linear transformation within a matrix block. Such a map need not preserve matrix rank. Therefore the abstract local Gaussian theorem uses an arbitrary Fisher-whitened cone $\mathsf{W}_0\mathcal{C}_0$, whereas the closed-form top-s/top-r calculation is invoked only under the additional block-isotropic or rank-preserving separable hypotheses stated explicitly in Theorem 5.1 and Appendix H.

B.7 Summary of the coordinate contract

The results above establish the exact coordinate contract used by the remaining appendices:

1. under the real-type assumption, equivariant transition operators have intrinsic multiplicity blocks B_ρ;
2. stochastic normalization imposes the single vector constraint $B_{\mathbf{1}} c_{\mathcal{S}} = c_{\mathcal{X}}$ and leaves the nuisance space $\mathcal{N}_{\mathbf{1}}$;
3. a strictly positive base kernel makes nonnegativity locally inactive, uniformly over Fisher-bounded directions;
4. the multinomial Fisher form is positive definite but need not respect the irrep blocks;
5. the uniform base kernel with balanced sampling yields the exact block-isotropic, rank-preserving whitening required by the structural-width and realizability arguments.

Appendix C will impose sparse determinantal constraints on the nontrivial blocks inside this affine coordinate system. This direct parameterization avoids any unsupported rule equating the tangent cone of an intersection with the intersection of separately computed tangent cones.

C TANGENT GEOMETRY OF EQUIVARIANT SPARSE LOW-RANK KERNELS

This appendix identifies the exact Bouligand tangent cone of the structured transition class and proves the stronger quantitative properties needed by the local statistical theory: every tangent direction admits a uniformly second-order accurate, rank-preserving stochastic path, and bounded local blow-ups of the model converge in Hausdorff distance to the tangent cone at a linear rate. The contingent-cone formalism is standard in variational analysis (Rockafellar & Wets, 1998). For a single bounded-rank matrix variety, the cone formula was stated by Schneider & Uschmajew (2015); a recent complete proof and equivalent descriptions are provided by Olikier et al. (2026). We give a self-contained block-matrix proof because our later arguments also require quantitative path and inverse-approximation bounds that are not supplied by the bare cone identity.

We work under the coordinate contract of Appendix B. In particular,

$$\mathcal{Z}_G^0 = \mathcal{N}_{\mathbf{1}} \oplus \bigoplus_{\rho \in \mathfrak{R}} \mathbb{R}^{m_\rho^{\mathcal{X}} \times m_\rho^{\mathcal{S}}}, \qquad \mathcal{B}_G : \mathcal{Z}_G^0 \longrightarrow \mathcal{H}_G^0 \tag{98}$$

is the block-reconstruction isomorphism from equation 79–equation 80. The uniform kernel is denoted by $\bar{P}$. For a normalized equivariant kernel P, write

$$\boldsymbol{B}(P) := \mathcal{B}_G^{-1}(P - \bar{P}) = \big(B_{\mathbf{1}}^{\circ}(P), (B_\rho(P))_{\rho \in \mathfrak{R}}\big). \tag{99}$$

For nontrivial ρ, this agrees with the multiplicity block of K_P, because $\bar{P}$ has no nontrivial isotypic component.

C.1 CONTINGENT CONES AND AFFINE REDUCTION

For a subset S of a finite-dimensional Euclidean space E and $x \in S$, we use the Bouligand–Peano contingent cone

$$T_S(x) := \left\{ v \in E : \exists\, t_k \downarrow 0,\ x_k \in S,\ x_k \to x,\ \frac{x_k - x}{t_k} \to v \right\}. \tag{100}$$

It is a closed cone, but need not be convex. If $L : E \to F$ is a linear isomorphism, then directly from the sequential definition,

$$T_{L(S)}(Lx) = L\, T_S(x). \tag{101}$$

Define the coordinate structured set

$$\begin{aligned} \mathfrak{V}_{s,\boldsymbol{r}} := \Big\{ \big(Z_{\mathbf{1}}, (Z_\rho)_\rho\big) \in \mathcal{Z}_G^0 :\ & |\{\rho \in \mathfrak{R} : Z_\rho \neq 0\}| \le s, \\ & \operatorname{rank}(Z_\rho) \le r_\rho \quad \forall \rho \in \mathfrak{R} \Big\}. \end{aligned} \tag{102}$$

The nuisance block $Z_{\mathbf{1}} \in \mathcal{N}_{\mathbf{1}}$ is unrestricted. By construction of the model in Section 2,

$$\Theta_{G,s,\boldsymbol{r}} = \left\{ \bar{P} + \mathcal{B}_G(\boldsymbol{Z}) : \boldsymbol{Z} \in \mathfrak{V}_{s,\boldsymbol{r}},\ \bar{P} + \mathcal{B}_G(\boldsymbol{Z}) \ge 0 \right\}. \tag{103}$$

Fix throughout a strictly positive base kernel $P_0 \in \Theta_{G,s,\boldsymbol{r}}$ and set

$$\boldsymbol{Z}^0 := \boldsymbol{B}(P_0) = \big(Z_{\mathbf{1}}^0, (B_\rho^0)_{\rho \in \mathfrak{R}}\big). \tag{104}$$

Proposition C.1 (Local removal of the positivity inequalities)**.** *There exists a neighborhood $\mathcal{U}$ of $\boldsymbol{Z}^0$ in $\mathcal{Z}_G^0$ such that*

$$\Theta_{G,s,\boldsymbol{r}} \cap \big(\bar{P} + \mathcal{B}_G(\mathcal{U})\big) = \bar{P} + \mathcal{B}_G\big(\mathfrak{V}_{s,\boldsymbol{r}} \cap \mathcal{U}\big). \tag{105}$$

Consequently,

$$\boxed{T_{\Theta_{G,s,\boldsymbol{r}}}(P_0) = \mathcal{B}_G\, T_{\mathfrak{V}_{s,\boldsymbol{r}}}(\boldsymbol{Z}^0).} \tag{106}$$

Proof. Let $p_{\min} := \min_{x,s} P_0(s \mid x) > 0$. Since $\mathcal{B}_G$ is linear and continuous, there exists a coordinate neighborhood $\mathcal{U}$ of $\boldsymbol{Z}^0$ such that

$$\big\|\mathcal{B}_G(\boldsymbol{Z} - \boldsymbol{Z}^0)\big\|_\infty < p_{\min} \qquad \forall \boldsymbol{Z} \in \mathcal{U}.$$

Every $\bar{P} + \mathcal{B}_G(\boldsymbol{Z})$ is already equivariant and row normalized; the last display makes it entrywise positive. Thus nonnegativity imposes no additional restriction in this neighborhood, proving equation 105. Tangent cones depend only on the germ of a set at the base point. Applying the linear-image rule equation 101 to the local equality yields equation 106. □

Pull back the Fisher metric through $\mathcal{B}_G$. Recall from equation 92 that the positive-definite self-adjoint operator $\mathbf{F}_{0,w}$ satisfies

$$\|\mathcal{B}_G(\boldsymbol{H})\|^2_{\mathcal{I}_{0,w}} = \langle \boldsymbol{H}, \mathbf{F}_{0,w}\boldsymbol{H}\rangle_{\mathcal{Z}}\,. \tag{107}$$

Let

$$\lambda_- := \lambda_{\min}(\mathbf{F}_{0,w}) > 0, \qquad \lambda_+ := \lambda_{\max}(\mathbf{F}_{0,w}) < \infty. \tag{108}$$

Then

$$\sqrt{\lambda_-}\,\|\boldsymbol{H}\|_{\mathcal{Z}} \le \|\mathcal{B}_G(\boldsymbol{H})\|_{\mathcal{I}_{0,w}} \le \sqrt{\lambda_+}\,\|\boldsymbol{H}\|_{\mathcal{Z}}\,. \tag{109}$$

These constants are fixed by (P_0, w) and are used only for quantitative uniformity; no block-diagonality of $\mathbf{F}_{0,w}$ is assumed.

C.2 A QUANTITATIVE TANGENT THEOREM FOR ONE DETERMINANTAL VARIETY

For integers $m, n \ge 1$ and $0 \le r \le \min\{m, n\}$, let

$$\mathcal{M}^{m,n}_{\le r} := \{X \in \mathbb{R}^{m\times n} : \operatorname{rank}(X) \le r\}. \tag{110}$$

Fix $X \in \mathcal{M}^{m,n}_{\le r}$ with $q := \operatorname{rank}(X)$. Choose a compact singular value decomposition and orthogonal complements,

$$X = U\Sigma V^\top, \qquad [U\ U_\perp] \in O(m), \qquad [V\ V_\perp] \in O(n), \tag{111}$$

where $\Sigma \in \mathbb{R}^{q\times q}$ is positive diagonal when $q > 0$. For $\Delta \in \mathbb{R}^{m\times n}$ write

$$[U\ U_\perp]^\top \Delta [V\ V_\perp] = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{112}$$

Empty blocks are interpreted in the natural way.

Theorem C.2 (Determinantal tangent cone with quantitative path lifting)**.** *The contingent cone of $\mathcal{M}^{m,n}_{\le r}$ at X is*

$$\boxed{T_{\mathcal{M}^{m,n}_{\le r}}(X) = \left\{\Delta \in \mathbb{R}^{m\times n} : \operatorname{rank}(U_\perp^\top \Delta V_\perp) \le r - q\right\}.} \tag{113}$$

If $q = 0$, the set in equation 113 is simply $\mathcal{M}^{m,n}_{\le r}$ and the exact path $X(t) := t\Delta$ suffices.

Suppose $q > 0$ and Δ satisfies equation 113. Whenever $\Sigma + tA$ is invertible, define

$$\mathsf{P}_{X,\Delta}(t) := [U\ U_\perp] \begin{bmatrix} \Sigma + tA & tB \\ tC & tD + t^2 C(\Sigma + tA)^{-1}B \end{bmatrix} [V\ V_\perp]^\top. \tag{114}$$

Then

$$\operatorname{rank}\big(\mathsf{P}_{X,\Delta}(t)\big) = q + \operatorname{rank}(tD) \le r. \tag{115}$$

In particular,

$$t \ne 0 \quad\Longrightarrow\quad \operatorname{rank}\big(\mathsf{P}_{X,\Delta}(t)\big) = q + \operatorname{rank}(D), \qquad \mathsf{P}_{X,\Delta}(0) = X. \tag{116}$$

The path is continuous at $t = 0$, and

$$\mathsf{P}_{X,\Delta}(t) = X + t\Delta + t^2 U_\perp C(\Sigma + tA)^{-1} B V_\perp^\top. \tag{117}$$

In particular, if $\|\Delta\|_{\mathrm{F}} \le M$ and

$$0 \le t \le \frac{\sigma_q(X)}{2M}, \tag{118}$$

with the usual convention when $M = 0$, then

$$\|\mathsf{P}_{X,\Delta}(t) - X - t\Delta\|_{\mathrm{F}} \le \frac{2M^2}{\sigma_q(X)} t^2. \tag{119}$$

Conversely, suppose $q > 0$, $\|\Delta\|_{\mathrm{F}} \le M$,

$$0 < t \le \frac{\sigma_q(X)}{2M}, \qquad X + t\Delta \in \mathcal{M}^{m,n}_{\le r}, \tag{120}$$

with the usual convention when $M = 0$. Define $\widehat{\Delta}$ by retaining A, B, C in equation 112 and replacing D by

$$\widehat{D} := D - tC(\Sigma + tA)^{-1}B. \tag{121}$$

Then

$$\widehat{\Delta} \in T_{\mathcal{M}^{m,n}_{\le r}}(X), \qquad \left\|\Delta - \widehat{\Delta}\right\|_{\mathrm{F}} \le \frac{2M^2}{\sigma_q(X)} t. \tag{122}$$

Proof. We first prove necessity in equation 113. Let $\Delta \in T_{\mathcal{M}_{\le r}^{m,n}}(X)$. By equation 100, there are $t_k \downarrow 0$ and $\Delta_k \to \Delta$ such that $X + t_k\Delta_k \in \mathcal{M}_{\le r}^{m,n}$. Write the blocks of Δ_k as A_k, B_k, C_k, D_k. For all large k, $\Sigma + t_k A_k$ is invertible. Block Gaussian elimination gives

$$\begin{aligned}\operatorname{rank}(X + t_k\Delta_k) &= q + \operatorname{rank}\big(t_k D_k - t_k^2 C_k(\Sigma + t_k A_k)^{-1} B_k\big) \\ &= q + \operatorname{rank}\big(D_k - t_k C_k(\Sigma + t_k A_k)^{-1} B_k\big)\,.\end{aligned} \tag{123}$$

Indeed, left and right multiplication by the invertible block-triangular matrices

$$\begin{bmatrix} I & 0 \\ -t_k C_k(\Sigma + t_k A_k)^{-1} & I \end{bmatrix}, \qquad \begin{bmatrix} I & -(\Sigma + t_k A_k)^{-1} t_k B_k \\ 0 & I \end{bmatrix}$$

reduces the matrix to block diagonal form. Since the left side of equation 123 is at most r, the corrected lower-right block has rank at most $r - q$. It converges to D; closedness of $\mathcal{M}_{\le r-q}^{m-q,n-q}$ yields $\operatorname{rank}(D) \le r - q$.

For sufficiency, assume $\operatorname{rank}(D) \le r - q$. The same block elimination applied to equation 114 reduces it to $\operatorname{diag}(\Sigma + tA, tD)$. Hence

$$\operatorname{rank}\big(\mathsf{P}_{X,\Delta}(t)\big) = q + \operatorname{rank}(tD),$$

which equals q at $t = 0$ and $q + \operatorname{rank}(D)$ when $t \neq 0$; this proves equation 115–equation 116. Expansion gives equation 117, hence Δ is tangent. When $\|\Delta\|_{\mathrm{F}} \le M$ and $t \le \sigma_q(X)/(2M)$,

$$\sigma_{\min}(\Sigma + tA) \ge \sigma_q(X) - t\,\|A\|_2 \ge \frac{\sigma_q(X)}{2},$$

so $\big\|(\Sigma + tA)^{-1}\big\|_2 \le 2/\sigma_q(X)$. Since $\|B\|_{\mathrm{F}}, \|C\|_{\mathrm{F}} \le M$, this proves equation 119.

Finally, assume the converse hypotheses, in particular $t > 0$. The Schur-complement identity gives

$$\operatorname{rank}(X + t\Delta) = q + \operatorname{rank}\Big(t\widehat{D}\Big) = q + \operatorname{rank}(\widehat{D}),$$

so $\operatorname{rank}(\widehat{D}) \le r - q$. Thus $\widehat{\Delta}$ belongs to the cone just proved, and

$$\Big\|\Delta - \widehat{\Delta}\Big\|_{\mathrm{F}} = t\,\big\|C(\Sigma + tA)^{-1}B\big\|_{\mathrm{F}} \le \frac{2M^2}{\sigma_q(X)}t.$$

This establishes equation 122. □

Remark C.3 (Imported and new components). The identity equation 113 is the classical bounded-rank tangent-cone formula; the modern reference Olikier et al. (2026) supplies a complete proof and explains a gap in an earlier proof of Schneider & Uschmajew (2015). The explicit path equation 114, the uniform remainder equation 119, and the inverse approximation equation 122 are recorded here in the precise form needed to prove a quantitative local blow-up theorem for the structured MDP model.

C.3 Block support sparsity

For each $\rho \in \mathfrak{R}$, set

$$E_\rho := \mathbb{R}^{m_\rho^{\mathcal{X}} \times m_\rho^{\mathcal{S}}}, \qquad \mathcal{M}_\rho := \{B \in E_\rho : \operatorname{rank}(B) \le r_\rho\}.$$

Define the nontrivial block variety

$$\mathcal{V}_{s,\boldsymbol{r}} := \left\{ (B_\rho)_{\rho\in\mathfrak{R}} \in \prod_{\rho\in\mathfrak{R}} E_\rho : |\{\rho : B_\rho \neq 0\}| \le s,\ B_\rho \in \mathcal{M}_\rho \right\}. \tag{124}$$

Let $\boldsymbol{B}^0 = (B_\rho^0)_\rho \in \mathcal{V}_{s,\boldsymbol{r}}$ and define

$$\Lambda_0 := \{\rho \in \mathfrak{R} : B_\rho^0 \neq 0\}, \qquad q_\rho := \operatorname{rank}(B_\rho^0), \tag{125}$$

with the convention $q_\rho = 0$ for $\rho \notin \Lambda_0$. For every ρ, let

$$\mathcal{T}_\rho := T_{\mathcal{M}_\rho}(B_\rho^0). \tag{126}$$

By Theorem C.2, if $B_\rho^0 = U_\rho \Sigma_\rho V_\rho^\top$, then

$$\mathcal{T}_\rho = \big\{H_\rho : \operatorname{rank}(U_{\rho,\perp}^\top H_\rho V_{\rho,\perp}) \le r_\rho - q_\rho\big\}\,. \tag{127}$$

At an inactive block, this reads $\mathcal{T}_\rho = \mathcal{M}_\rho$.

Theorem C.4 (Tangent cone of a sparse product of determinantal varieties)**.** *Let*

$$\mathfrak{L}_0 := \{\Lambda \subseteq \mathfrak{R} : \Lambda_0 \subseteq \Lambda,\ |\Lambda| \le s\}. \tag{128}$$

Then

$$\boxed{T_{\mathcal{V}_{s,\boldsymbol{r}}}(\boldsymbol{B}^0) = \bigcup_{\Lambda \in \mathfrak{L}_0} \left(\prod_{\rho \in \Lambda} \mathcal{T}_\rho \times \prod_{\rho \notin \Lambda} \{0\} \right).} \tag{129}$$

Proof. Let $\boldsymbol{H} \in T_{\mathcal{V}_{s,\boldsymbol{r}}}(\boldsymbol{B}^0)$. Choose $t_k \downarrow 0$ and $\boldsymbol{B}^k \in \mathcal{V}_{s,\boldsymbol{r}}$ such that $(\boldsymbol{B}^k - \boldsymbol{B}^0)/t_k \to \boldsymbol{H}$. Since every block in Λ_0 is nonzero, $B_\rho^k \neq 0$ for all $\rho \in \Lambda_0$ and all sufficiently large k. The family of supports is finite, so after passing to a subsequence there is a fixed $\Lambda \in \mathfrak{L}_0$ with $\operatorname{supp}(\boldsymbol{B}^k) = \Lambda$. For $\rho \notin \Lambda$, the scaled block is zero; for $\rho \in \Lambda$, the scaled block converges to a tangent direction of $\mathcal{M}_\rho$ at B_ρ^0. This proves inclusion in the right side of equation 129.

Conversely, fix $\Lambda \in \mathfrak{L}_0$ and a block direction in the corresponding product. For every $\rho \in \Lambda$, apply Theorem C.2 to obtain a common-small-t path $B_\rho(t) \in \mathcal{M}_\rho$ with derivative H_ρ; take $B_\rho(t) \equiv 0$ outside Λ. The combined path has at most $|\Lambda| \le s$ nonzero blocks and derivative $\boldsymbol{H}$, proving the reverse inclusion. □

C.4 Exact structured transition tangent cone

Combining Proposition C.1 and Theorem C.4 gives the exact local model geometry.

Theorem C.5 (Tangent cone of the equivariant sparse low-rank model)**.** *Let $P_0 \in \Theta_{G,s,\boldsymbol{r}}$ be strictly positive, with coordinates equation 104. Then*

$$\boxed{\mathcal{C}_0 := T_{\Theta_{G,s,\boldsymbol{r}}}(P_0) = \mathcal{B}_G\left(\mathcal{N}_{\mathbf{1}} \oplus \bigcup_{\Lambda \in \mathfrak{L}_0} \left[\prod_{\rho \in \Lambda} \mathcal{T}_\rho \times \prod_{\rho \notin \Lambda} \{0\} \right] \right).} \tag{130}$$

Equivalently, $u \in \mathcal{C}_0$ if and only if its coordinate direction $\boldsymbol{H} = \mathcal{B}_G^{-1} u$ has arbitrary nuisance component $H_{\mathbf{1}} \in \mathcal{N}_{\mathbf{1}}$ and there exists $\Lambda \in \mathfrak{L}_0$ such that

$$H_\rho = 0 \quad (\rho \notin \Lambda), \qquad \operatorname{rank}(U_{\rho,\perp}^\top H_\rho V_{\rho,\perp}) \le r_\rho - q_\rho \quad (\rho \in \Lambda). \tag{131}$$

Proof. The coordinate set equation 102 is the direct product of the unrestricted linear nuisance space $\mathcal{N}_{\mathbf{1}}$ and the nontrivial sparse variety equation 124. Its tangent cone is therefore the direct product of $\mathcal{N}_{\mathbf{1}}$ and the cone in Theorem C.4. The affine reduction equation 106 completes the proof. □

Corollary C.6 (Central symmetry of the structured tangent cone)**.** *At every strictly positive $P_0 \in \Theta_{G,s,\boldsymbol{r}}$,*

$$\boxed{\mathcal{C}_0 = -\mathcal{C}_0, \qquad \mathcal{K}_0 = -\mathcal{K}_0.} \tag{132}$$

Moreover, both $\mathcal{C}_0$ and $\mathcal{K}_0$ are closed cones.

Proof. The nuisance space $\mathcal{N}_{\mathbf{1}}$ is linear. For every ρ, the block condition

$$\operatorname{rank}(U_{\rho,\perp}^\top H_\rho V_{\rho,\perp}) \le r_\rho - q_\rho$$

is unchanged by $H_\rho \mapsto -H_\rho$, and the zero pattern associated with an admissible support Λ is unchanged by simultaneous sign reversal. Thus every component of the finite union in equation 130 is a closed centrally symmetric cone. A finite union of closed sets is closed and remains centrally symmetric. Finally, $\mathcal{B}_G$ is a linear homeomorphism and W_0 is a linear isometry, so both properties pass to $\mathcal{C}_0$ and $\mathcal{K}_0$. □

Remark C.7 (The two local singularity mechanisms). Formula equation 130 isolates two independent sources of singularity. If $|\Lambda_0| < s$, previously inactive irrep blocks may turn on, producing a finite union over support supersets. If an active block has $q_\rho < r_\rho$, its determinantal tangent cone contains a nonlinear rank constraint in the orthogonal–orthogonal corner. When the support is saturated and every nonvacuous active rank constraint is saturated, the local cone is a linear tangent space.

C.5 Uniform second-order path lifting

We next strengthen mere tangency to a uniform second-order path statement. Let

$$\kappa_- := \lambda_-^{-1/2}, \qquad \kappa_+ := \lambda_+^{1/2}, \qquad M_R := \kappa_- R. \tag{133}$$

If $\Lambda_0 \neq \varnothing$, define the smallest positive active singular value

$$\sigma_* := \min_{\rho \in \Lambda_0} \sigma_{q_\rho}(B_\rho^0) > 0, \tag{134}$$

and otherwise set $\sigma_* := +\infty$. Let

$$C_R^{\mathrm{geom}} := \begin{cases} \dfrac{2\kappa_+ \sqrt{|\Lambda_0|}\, M_R^2}{\sigma_*}, & \Lambda_0 \neq \varnothing, \\ 0, & \Lambda_0 = \varnothing. \end{cases} \tag{135}$$

We use the convention that a bound with a zero denominator is omitted when its numerator also vanishes. Define

$$\begin{aligned} \tau_R^{\mathrm{inv}} &:= \begin{cases} \sigma_*/(2M_R), & \Lambda_0 \neq \varnothing,\ R > 0, \\ +\infty, & \text{otherwise}, \end{cases} \\ \tau_R^{\mathrm{pos}} &:= \min \left\{ \begin{cases} \dfrac{p_{\min}\sqrt{w_{\min}}}{4R}, & R > 0, \\ +\infty, & R = 0, \end{cases} \quad \begin{cases} \sqrt{\dfrac{p_{\min}\sqrt{w_{\min}}}{4C_R^{\mathrm{geom}}}}, & C_R^{\mathrm{geom}} > 0, \\ +\infty, & C_R^{\mathrm{geom}} = 0, \end{cases} \right\}, \\ \tau_R^{\mathrm{geom}} &:= \min\{\tau_R^{\mathrm{inv}}, \tau_R^{\mathrm{pos}}\}. \end{aligned} \tag{136}$$

The superscript "geom" emphasizes that this radius enforces only block invertibility and stochastic positivity. Active-policy localization and Bellman Taylor control use the separate radius defined in Appendix D.

Theorem C.8 (Uniform rank-preserving stochastic path lifting)**.** *For every $R > 0$ and every $u \in \mathcal{C}_0$ with $\|u\|_{\mathcal{I}_{0,w}} \leq R$, there exists a path*

$$P_u(t) \in \Theta_{G,s,\boldsymbol{r}}, \qquad 0 \leq t \leq \tau_R^{\mathrm{geom}}, \tag{137}$$

such that

$$P_u(0) = P_0, \qquad P_u(t) = P_0 + tu + E_u(t), \qquad \|E_u(t)\|_{\mathcal{I}_{0,w}} \leq C_R^{\mathrm{geom}} t^2. \tag{138}$$

Moreover,

$$P_u(t)(s \mid x) \geq \frac{p_{\min}}{2} \qquad \forall (x,s), \quad 0 \leq t \leq \tau_R^{\mathrm{geom}}. \tag{139}$$

The path may be chosen blockwise and preserves a support pattern witnessing $u \in \mathcal{C}_0$.

Proof. Let $\boldsymbol{H} = \mathcal{B}_G^{-1} u$. By equation 109, $\|\boldsymbol{H}\|_{\mathcal{Z}} \leq M_R$, so every block has Frobenius norm at most M_R. Choose $\Lambda \in \mathfrak{L}_0$ witnessing membership in equation 130. Move the nuisance coordinate along the straight line $Z_{\mathbf{1}}^0 + tH_{\mathbf{1}}$. For $\rho \in \Lambda_0$, use the path equation 114; for $\rho \in \Lambda \setminus \Lambda_0$, use the exact conic path tH_ρ; and keep all blocks outside Λ equal to zero. For $t \leq \tau_R^{\mathrm{inv}}$, every active-base inverse exists and every block has rank at most its budget. The nontrivial support is contained in Λ, hence has cardinality at most s.

Only the blocks in Λ_0 contribute a quadratic correction. By equation 119, their direct-sum coordinate norm is at most

$$\frac{2\sqrt{|\Lambda_0|}\, M_R^2}{\sigma_*} t^2.$$

The upper inequality in equation 109 converts this into the Fisher bound $C_R^{\mathrm{geom}} t^2$ in equation 138.

Finally,

$$\|P_u(t) - P_0\|_\infty \leq \frac{tR + C_R^{\mathrm{geom}} t^2}{\sqrt{w_{\min}}} \leq \frac{p_{\min}}{2}$$

for $t \leq \tau_R^{\mathrm{pos}}$, using equation 86. Thus the reconstructed matrix is strictly positive; it is normalized and equivariant by construction. This proves equation 139 and completes the proof. □

C.6 Quantitative convergence of bounded local blow-ups

For $t > 0$ and $R \geq 0$, define the bounded blow-up of the actual model by

$$\mathcal{A}_t(R) := \left\{ \frac{P - P_0}{t} : P \in \Theta_{G,s,\boldsymbol{r}},\ \left\| \frac{P - P_0}{t} \right\|_{\mathcal{I}_{0,w}} \leq R \right\}. \tag{140}$$

For nonempty compact subsets A, B of $\mathcal{H}_G^0$, let

$$\mathrm{d}_{\mathrm{H}}^{\mathcal{I}}(A, B) := \max \left\{ \sup_{a \in A} \mathrm{dist}_{\mathcal{I}}(a, B), \sup_{b \in B} \mathrm{dist}_{\mathcal{I}}(b, A) \right\} \tag{141}$$

be the Hausdorff distance induced by the Fisher norm.

Theorem C.9 (Linear-rate local blow-up convergence)**.** *Fix $R > 0$. If $C_R^{\mathrm{geom}} > 0$, set*

$$\bar{\tau}_R^{\mathrm{geom}} := \min \left\{ \tau_R^{\mathrm{geom}}, \frac{R}{C_R^{\mathrm{geom}}} \right\}; \tag{142}$$

if $C_R^{\mathrm{geom}} = 0$, set $\bar{\tau}_R^{\mathrm{geom}} := \tau_R^{\mathrm{geom}}$. Then, for every $0 < t \leq \bar{\tau}_R^{\mathrm{geom}}$,

$$\boxed{\mathrm{d}_{\mathrm{H}}^{\mathcal{I}}(\mathcal{A}_t(R), \mathcal{C}_0 \cap R\mathbb{B}_{\mathcal{I}}) \leq 2C_R^{\mathrm{geom}} t.} \tag{143}$$

In particular,

$$\mathcal{A}_t(R) \xrightarrow[t \downarrow 0]{\text{Hausdorff}} \mathcal{C}_0 \cap R\mathbb{B}_{\mathcal{I}}. \tag{144}$$

For $R = 0$, both sets are $\{0\}$.

Proof. We prove the two directed-distance bounds.

Outer approximation. Take $u_t = (P_t - P_0)/t \in \mathcal{A}_t(R)$ and write $\boldsymbol{H}_t = \mathcal{B}_G^{-1} u_t$. By equation 109, $\|\boldsymbol{H}_t\|_{\mathcal{Z}} \leq M_R$. Let Λ_t be the support of the nontrivial blocks of P_t. For every $\rho \in \Lambda_0$ and $t \leq \tau_R^{\mathrm{inv}}$, Weyl's inequality gives

$$\sigma_{q_\rho}(B_\rho^0 + tH_{t,\rho}) \geq \sigma_* - tM_R \geq \frac{\sigma_*}{2} > 0.$$

Hence $\Lambda_0 \subseteq \Lambda_t$ and $\Lambda_t \in \mathfrak{L}_0$.

For each $\rho \in \Lambda_0$, apply the inverse correction equation 121 to $H_{t,\rho}$. For $\rho \in \Lambda_t \setminus \Lambda_0$, keep $H_{t,\rho}$ unchanged; it already has rank at most r_ρ because $B_\rho(P_t) = tH_{t,\rho}$. Set all blocks outside Λ_t to zero and leave the nuisance component unchanged. The resulting coordinate direction $\widehat{\boldsymbol{H}}_t$ belongs to the cone in equation 130. The quantitative estimate equation 122, summed over Λ_0 and converted by equation 109, yields

$$\|u_t - \widehat{u}_t\|_{\mathcal{I}_{0,w}} \leq C_R^{\mathrm{geom}} t, \qquad \widehat{u}_t := \mathcal{B}_G(\widehat{\boldsymbol{H}}_t) \in \mathcal{C}_0. \tag{145}$$

Since $\|\widehat{u}_t\|_{\mathcal{I}_{0,w}} \leq R + C_R^{\mathrm{geom}} t$, define

$$\beta_t := \begin{cases} 1, & \|\widehat{u}_t\|_{\mathcal{I}_{0,w}} \leq R, \\ \dfrac{R}{\|\widehat{u}_t\|_{\mathcal{I}_{0,w}}}, & \|\widehat{u}_t\|_{\mathcal{I}_{0,w}} > R. \end{cases}$$

The cone property gives $\beta_t \widehat{u}_t \in \mathcal{C}_0 \cap R\mathbb{B}_{\mathcal{I}}$, and

$$\begin{aligned} \|u_t - \beta_t \widehat{u}_t\|_{\mathcal{I}_{0,w}} &\leq \|u_t - \widehat{u}_t\|_{\mathcal{I}_{0,w}} + (1 - \beta_t) \|\widehat{u}_t\|_{\mathcal{I}_{0,w}} \\ &\leq C_R^{\mathrm{geom}} t + \left(\|\widehat{u}_t\|_{\mathcal{I}_{0,w}} - R \right)_+ \leq 2C_R^{\mathrm{geom}} t. \end{aligned}$$

This proves the first directed-distance bound.

Inner approximation. Take $u \in \mathcal{C}_0 \cap R\mathbb{B}_{\mathcal{I}}$. If $C_R^{\mathrm{geom}} = 0$, use Theorem C.8 directly. Otherwise, for $t \leq R/C_R^{\mathrm{geom}}$, set

$$\alpha_t := 1 - \frac{C_R^{\mathrm{geom}} t}{R} \in [0, 1].$$

Because $\mathcal{C}_0$ is a cone, $\alpha_t u \in \mathcal{C}_0$. Apply Theorem C.8 to this direction and define

$$v_t := \frac{P_{\alpha_t u}(t) - P_0}{t} = \alpha_t u + \frac{E_{\alpha_t u}(t)}{t}.$$

Then

$$\|v_t\|_{\mathcal{I}_{0,w}} \le \alpha_t R + C_R^{\mathrm{geom}} t = R,$$

so $v_t \in \mathcal{A}_t(R)$, and

$$\|v_t - u\|_{\mathcal{I}_{0,w}} \le (1-\alpha_t)R + C_R^{\mathrm{geom}} t = 2C_R^{\mathrm{geom}} t.$$

This proves the second directed-distance bound and hence equation 143. □

Corollary C.10 (First-order derivability of the structured MDP model)**.** *At every strictly positive $P_0 \in \Theta_{G,s,\boldsymbol{r}}$, the structured model is first-order derivable in the following strong sense:*

1. *every direction in $\mathcal{C}_0$ is realized by a feasible path with a uniform $O(t^2)$ error on Fisher-bounded sets;*
2. *every Fisher-bounded sequence of feasible t-scale displacements is $O(t)$-close to $\mathcal{C}_0$;*
3. *the bounded local parameter sets used by the statistical experiment converge to the tangent cone in Fisher-Hausdorff distance.*

Remark C.11 (Why Theorem C.9 is needed). The definition of a contingent cone supplies only subsequential first-order directions. The local asymptotic minimax theorem requires a stronger uniform statement over an entire bounded neighborhood. The explicit Schur-complement path and inverse correction are what upgrade pointwise tangency to the quantitative set convergence in equation 143. This is the geometric input used in Appendix E to replace the varying finite-sample parameter sets by the fixed Gaussian cone $\mathcal{K}_0$.

D Bellman differential calculus

This appendix establishes the deterministic analytic layer of the local experiment. We first differentiate a fixed finite-horizon policy exactly, then pass to the optimized value by a finite-maximum argument, and finally recover an equivalent statewise max–plus recursion when the policy class is the complete class of deterministic nonstationary Markov policies. First-order sensitivity of optimal Markov decision values with respect to transition perturbations has been studied in substantially more general measurable-state models (Kern et al., 2020). Our purpose here is narrower: we require finite-dimensional formulas with *uniform quadratic remainders* that can be composed with the structured paths of Appendix C and the Gaussian experiment of Appendix E. We use Hadamard directional differentiability in the sense formalized by Shapiro (1990); all finite-maximum calculations needed below are nevertheless proved directly.

Throughout this appendix, P_0 is the strictly positive base kernel from Section 2, Π is the fixed finite candidate class of deterministic nonstationary Markov policies, and

$$\Pi_0 := \arg\max_{\pi \in \Pi} J^\pi(P_0).$$

All rewards satisfy

$$\max_{1 \le t < T} \|r_t\|_\infty \vee \|r_T\|_\infty \le R_{\max}. \tag{146}$$

We write $\langle a, b\rangle := a^\top b$ for Euclidean duality on state vectors. Unless displayed explicitly, constants may depend on the fixed objects $(P_0, w, T, R_{\max}, \Pi, R, C)$ but not on the local direction, the active candidate policy, its selected support pattern, or n.

D.1 Policy-specific operators and perturbation norms

Fix $\pi = (\pi_1, \dots, \pi_{T-1}) \in \Pi$. Its stage-$t$ selection operator $E_t^\pi : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{S}}$ is

$$(E_t^\pi q)(s) := q(s, \pi_t(s)). \tag{147}$$

For a transition kernel P and a row-sum-zero perturbation $u \in \mathbb{R}^{\mathcal{X}\times\mathcal{S}}$, define

$$P_t^\pi(P) := E_t^\pi K_P, \qquad U_t^\pi(u) := E_t^\pi K_u, \tag{148}$$

so that

$$(P_t^\pi(P)v)(s) = \sum_{s'} P(s' \mid s, \pi_t(s))v(s'), \qquad (U_t^\pi(u)v)(s) = \sum_{s'} u(s' \mid s, \pi_t(s))v(s').$$

Let $r_t^\pi(s) := r_t(s, \pi_t(s))$. The policy value vectors satisfy

$$V_T^\pi(P) = r_T, \qquad V_t^\pi(P) = r_t^\pi + P_t^\pi(P)V_{t+1}^\pi(P), \quad t = T-1, \dots, 1, \tag{149}$$

and $J^\pi(P) = \langle \mu, V_1^\pi(P)\rangle$.

For a matrix $u \in \mathbb{R}^{\mathcal{X}\times\mathcal{S}}$, define the maximal row ℓ_1 norm

$$\|u\|_{\infty,1} := \max_{x\in\mathcal{X}} \sum_{s'\in\mathcal{S}} |u(s' \mid x)|\,. \tag{150}$$

This is the natural perturbation norm for sup-norm dynamic programming.

Lemma D.1 (Operator and Fisher domination)**.** *For every policy π, stage t, perturbation u, and state vector v,*

$$\|U_t^\pi(u)v\|_\infty \le \|u\|_{\infty,1}\, \|v\|_\infty\,. \tag{151}$$

If $u \in \mathcal{H}_G^0$ and $w_x \ge w_{\min} > 0$, then

$$\|u\|_{\infty,1} \le w_{\min}^{-1/2}\, \|u\|_{\mathcal{I}_{0,w}}\,. \tag{152}$$

Proof. The first assertion follows rowwise from

$$|(U_t^\pi(u)v)(s)| \le \sum_{s'} |u(s' \mid s, \pi_t(s))|\, \|v\|_\infty\,.$$

For the second, weighted Cauchy–Schwarz and $\sum_{s'} P_0(s' \mid x) = 1$ give, for every x,

$$\begin{aligned}\left(\sum_{s'} |u(s' \mid x)|\right)^2 &\le \left(\sum_{s'} \frac{u(s' \mid x)^2}{P_0(s' \mid x)}\right)\left(\sum_{s'} P_0(s' \mid x)\right)\\ &\le \frac{1}{w_x}\, \|u\|_{\mathcal{I}_{0,w}}^2\,.\end{aligned}$$

Taking the maximum over x proves equation 152. □

D.2 Smoothness and quantitative derivatives of a fixed policy

Proposition D.2 (Polynomiality and derivative recursions)**.** *For every fixed $\pi \in \Pi$ and $t \in \{1, \dots, T\}$, the map $P \mapsto V_t^\pi(P)$ is a vector-valued polynomial of degree at most $T - t$ in the entries of P. Consequently, $P \mapsto J^\pi(P)$ is a polynomial of degree at most $T - 1$ and is infinitely Fréchet differentiable on the ambient matrix space.*

At a row-stochastic kernel P, the first derivative in direction u is given by $\dot V_T^\pi[u] = 0$ and

$$\dot V_t^\pi[u] = U_t^\pi(u)V_{t+1}^\pi(P) + P_t^\pi(P)\dot V_{t+1}^\pi[u]. \tag{153}$$

For directions u, v, the second derivative is given by $\ddot V_T^\pi[u, v] = 0$ and

$$\begin{aligned}\ddot V_t^\pi[u, v] = U_t^\pi(u)\dot V_{t+1}^\pi[v] + U_t^\pi(v)\dot V_{t+1}^\pi[u]\\ + P_t^\pi(P)\ddot V_{t+1}^\pi[u, v].\end{aligned} \tag{154}$$

Finally,

$$DJ_P^\pi[u] = \langle \mu, \dot V_1^\pi[u]\rangle, \qquad D^2 J_P^\pi[u, v] = \langle \mu, \ddot V_1^\pi[u, v]\rangle. \tag{155}$$

Proof. Repeated substitution in equation 149 gives

$$V_t^\pi(P) = r_t^\pi + \sum_{k=t+1}^{T-1} P_t^\pi(P)\cdots P_{k-1}^\pi(P) r_k^\pi + P_t^\pi(P)\cdots P_{T-1}^\pi(P) r_T, \tag{156}$$

where an empty product is the identity. Every $P_j^\pi(P)$ depends linearly on P, proving polynomiality and the degree bound. Differentiating equation 149 once and twice yields equation 153 and equation 154; applying the fixed linear functional $v \mapsto \langle \mu, v\rangle$ gives equation 155. □

The next result provides the uniform constants used throughout the local analysis. Put

$$L_T := \frac{R_{\max} T(T-1)}{2}, \qquad H_T := \frac{R_{\max} T(T-1)(T-2)}{3}, \qquad Q_T := \frac{H_T}{2}. \tag{157}$$

In particular, $Q_2 = 0$, as expected because a horizon-two policy value is affine in P.

Lemma D.3 (Uniform derivative bounds)**.** *Let P be row stochastic. For every policy π, stage t, and directions u, v, writing $h := T - t$,*

$$\|V_t^\pi(P)\|_\infty \le (h+1) R_{\max}, \tag{158}$$

$$\|DV_t^\pi(P)[u]\|_\infty \le \frac{h(h+1)}{2} R_{\max} \|u\|_{\infty,1}, \tag{159}$$

$$\left\|D^2 V_t^\pi(P)[u,v]\right\|_\infty \le \frac{h(h-1)(h+1)}{3} R_{\max} \|u\|_{\infty,1} \|v\|_{\infty,1}. \tag{160}$$

Consequently,

$$|DJ_P^\pi[u]| \le L_T \|u\|_{\infty,1}, \qquad \left|D^2 J_P^\pi[u,v]\right| \le H_T \|u\|_{\infty,1} \|v\|_{\infty,1}. \tag{161}$$

All bounds are uniform over the finite policy class Π and, in fact, over all deterministic nonstationary Markov policies.

Proof. Since every $P_t^\pi(P)$ is a Markov operator, $\|P_t^\pi(P) z\|_\infty \le \|z\|_\infty$. Backward induction in equation 149 gives equation 158. Applying Lemma D.1 to equation 153 and inducting backward gives

$$\|DV_t^\pi(P)[u]\|_\infty \le R_{\max} \|u\|_{\infty,1} \sum_{j=1}^{h} j = \frac{h(h+1)}{2} R_{\max} \|u\|_{\infty,1}.$$

Similarly, using equation 154 and the preceding first-derivative bound,

$$\begin{aligned}\left\|D^2 V_t^\pi(P)[u,v]\right\|_\infty &\le 2R_{\max} \|u\|_{\infty,1} \|v\|_{\infty,1} \sum_{j=1}^{h} \frac{(j-1)j}{2} \\ &= \frac{h(h-1)(h+1)}{3} R_{\max} \|u\|_{\infty,1} \|v\|_{\infty,1}.\end{aligned}$$

Finally, μ is a probability vector, so $|\langle \mu, z\rangle| \le \|z\|_\infty$; substituting $t = 1$, hence $h = T - 1$, proves equation 161. □

Corollary D.4 (Uniform fixed-policy Taylor bounds)**.** *Let P and $P + h$ be row-stochastic kernels. Then, for every $\pi \in \Pi$,*

$$|J^\pi(P+h) - J^\pi(P)| \le L_T \|h\|_{\infty,1}, \tag{162}$$

$$|J^\pi(P+h) - J^\pi(P) - DJ_P^\pi[h]| \le Q_T \|h\|_{\infty,1}^2. \tag{163}$$

In particular, if $\|u\|_{\mathcal{I}_{0,w}} \le R$ and $P_0 + tu$ is stochastic, then

$$\left|J^\pi(P_0 + tu) - J^\pi(P_0) - tDJ_{P_0}^\pi[u]\right| \le \frac{Q_T R^2}{w_{\min}} t^2. \tag{164}$$

Proof. The segment $P_\tau = P + \tau h$, $\tau \in [0,1]$, is row stochastic by convexity. The fundamental theorem of calculus and the first bound in equation 161 imply equation 162. Taylor's theorem with integral remainder and the second bound in equation 161 give

$$J^\pi(P+h) - J^\pi(P) - DJ^\pi_P[h] = \int_0^1 (1-\tau) D^2 J^\pi_{P_\tau}[h,h] d\tau,$$

whose absolute value is at most $H_T \|h\|^2_{\infty,1}/2 = Q_T \|h\|^2_{\infty,1}$. Equation equation 164 follows from Lemma D.1. □

D.3 Occupancy–continuation gradients and Fisher–Riesz representatives

Let $\nu^\pi_t \in \Delta(\mathcal{S})$ be the stage-t state distribution under (P_0, π), and define the state-action occupancy

$$d^\pi_t(s,a) := \nu^\pi_t(s) \mathbf{1}\{a = \pi_t(s)\}. \tag{165}$$

Proposition D.5 (Occupancy–continuation derivative)**.** *For every $\pi \in \Pi$ and row-sum-zero direction u,*

$$\boxed{DJ^\pi_{P_0}[u] = \sum_{t=1}^{T-1} \sum_{x \in \mathcal{X}} d^\pi_t(x) \sum_{s' \in \mathcal{S}} u(s' \mid x) V^\pi_{t+1}(s').} \tag{166}$$

Equivalently, defining the raw gradient matrix

$$H_\pi(x,s') := \sum_{t=1}^{T-1} d^\pi_t(x) V^\pi_{t+1}(s'), \tag{167}$$

we have

$$DJ^\pi_{P_0}[u] = \langle H_\pi, u \rangle_{\mathrm{F}}. \tag{168}$$

Proof. Unrolling equation 153 at P_0 yields

$$\dot V^\pi_1[u] = \sum_{t=1}^{T-1} P^\pi_1(P_0) \cdots P^\pi_{t-1}(P_0) U^\pi_t(u) V^\pi_{t+1}(P_0),$$

where the product is the identity when $t = 1$. Left multiplication by $\mu^o p$ produces the state distribution $(\nu^\pi_t)^\top$, proving equation 166. Rearranging the finite sums gives equation 168. □

The matrix H_π is not itself the Fisher gradient because the statistical tangent space imposes row-sum-zero and equivariance constraints. The next result gives the exact representative.

Proposition D.6 (Explicit Fisher–Riesz representative)**.** *For each $x \in \mathcal{X}$, put*

$$\bar H_\pi(x) := \sum_{y \in \mathcal{S}} P_0(y \mid x) H_\pi(x,y), \tag{169}$$

and define $\Gamma^\circ_\pi \in \mathbb{R}^{\mathcal{X} \times \mathcal{S}}$ by

$$\Gamma^\circ_\pi(s' \mid x) := \frac{P_0(s' \mid x)}{w_x} \big(H_\pi(x,s') - \bar H_\pi(x) \big). \tag{170}$$

Then $\Gamma^\circ_\pi \mathbf{1}_{\mathcal{S}} = 0$ and, for every row-sum-zero u,

$$DJ^\pi_{P_0}[u] = \langle \Gamma^\circ_\pi, u \rangle_{\mathcal{I}_{0,w}}. \tag{171}$$

Assume P_0 and w are G-invariant and let

$$\mathrm{Av}_G Z := \frac{1}{|G|} \sum_{g \in G} \mathcal{U}_g Z, \qquad (\mathcal{U}_g Z)(s' \mid x) := Z(g^{-1}s' \mid g^{-1}x). \tag{172}$$

Then

$$\Gamma^G_\pi := \mathrm{Av}_G \Gamma^\circ_\pi \tag{173}$$

is the unique Fisher–Riesz representative of $DJ^\pi_{P_0}$ on $\mathcal{H}^0_G$. Consequently, under the Fisher isometry W_0 of Section 2,

$$\gamma_\pi = \mathsf{W}_0 \Gamma^G_\pi. \tag{174}$$

Proof. The row sum of Γ°_π is

$$\frac{1}{w_x}\left(\sum_{s'} P_0(s' \mid x) H_\pi(x, s') - \bar{H}_\pi(x)\right) = 0.$$

For any row-sum-zero u,

$$\begin{aligned}\langle \Gamma^\circ_\pi, u\rangle_{\mathcal{I}_{0,w}} &= \sum_x \sum_{s'} \big(H_\pi(x, s') - \bar{H}_\pi(x)\big) u(s' \mid x) \\ &= \sum_x \sum_{s'} H_\pi(x, s') u(s' \mid x),\end{aligned}$$

which equals equation 168. By Fisher invariance, proved in Appendix B, Av_G is the Fisher-orthogonal projection onto the fixed-point space $\mathcal{H}^0_G$. Hence, for $u \in \mathcal{H}^0_G$,

$$\langle \Gamma^\circ_\pi, u\rangle_{\mathcal{I}_{0,w}} = \langle \mathrm{Av}_G\, \Gamma^\circ_\pi, u\rangle_{\mathcal{I}_{0,w}}\,.$$

Uniqueness follows from positive definiteness of the Fisher form on $\mathcal{H}^0_G$. Finally, the isometry property of W_0 gives

$$DJ^\pi_{P_0}[u] = \left\langle \mathsf{W}_0 \Gamma^G_\pi, \mathsf{W}_0 u\right\rangle,$$

which identifies equation 174. □

D.4 Local stability of the base-optimal candidate face

If $\Pi \setminus \Pi_0 \neq \varnothing$, define the strict finite-class gap

$$\Delta_0 := \min_{\pi \in \Pi \setminus \Pi_0} \big(J^\star(P_0) - J^\pi(P_0)\big) > 0. \tag{175}$$

If every policy is base optimal, set $\Delta_0 := +\infty$.

Lemma D.7 (Active-set localization)**.** *Let P be row stochastic. If*

$$\|P - P_0\|_{\infty,1} \le \frac{\Delta_0}{4L_T}, \tag{176}$$

with the convention that the right side is $+\infty$ when $\Delta_0 = +\infty$, then

$$J^\star(P) = \max_{\pi \in \Pi_0} J^\pi(P). \tag{177}$$

Moreover, every $\pi \notin \Pi_0$ is separated from every maximizer over Π_0 by at least $\Delta_0/2$.

Proof. Fix $\pi_0 \in \Pi_0$ and $\pi \notin \Pi_0$. By Corollary D.4,

$$\begin{aligned}J^{\pi_0}(P) - J^\pi(P) &\ge J^{\pi_0}(P_0) - J^\pi(P_0) - 2L_T \|P - P_0\|_{\infty,1} \\ &\ge \Delta_0/2.\end{aligned}$$

Thus no policy outside Π_0 can maximize at P; this proves both claims. □

For $R, C \ge 0$, define the active-face radius

$$\tau^{\mathrm{face}}_{R,C} := \begin{cases} +\infty, & \Delta_0 = +\infty \text{ or } L_T = 0, \\ \dfrac{\Delta_0 \sqrt{w_{\min}}}{4L_T R}, & C = 0,\ R > 0,\ \Delta_0 < +\infty, \\ \dfrac{-R + \sqrt{R^2 + C\Delta_0 \sqrt{w_{\min}}/L_T}}{2C}, & C > 0,\ \Delta_0 < +\infty, \\ +\infty, & R = C = 0. \end{cases} \tag{178}$$

and the Bellman radius

$$\tau^{\mathrm{Bell}}_{R,C} := \min\{1, \tau^{\mathrm{face}}_{R,C}\}. \tag{179}$$

Indeed, if $0 \le t \le \tau^{\mathrm{Bell}}_{R,C}$ and $\|u\|_{\mathcal{I}_{0,w}} \le R$, $\|e(t)\|_{\mathcal{I}_{0,w}} \le Ct^2$, then

$$\|tu + e(t)\|_{\infty,1} \le \frac{Rt + Ct^2}{\sqrt{w_{\min}}} \le \frac{\Delta_0}{4L_T}, \tag{180}$$

with the usual interpretation when $\Delta_0 = +\infty$. Thus this radius controls active-policy localization and the common Taylor regime; it is distinct from the geometric feasibility radius of Appendix C.

D.5 HADAMARD DERIVATIVE AND UNIFORM LOCAL REGRET EXPANSION

We use the following one-sided tangential notion. A function F is Hadamard directionally differentiable at P_0 tangentially to a set $\mathcal{D}$ if there exists a continuous positively homogeneous map F'_{P_0} on $\mathcal{D}$ such that, for every $t_n \downarrow 0$ and every $u_n \to u \in \mathcal{D}$ with $P_0 + t_n u_n$ in the domain,

$$\frac{F(P_0 + t_n u_n) - F(P_0)}{t_n} \longrightarrow F'_{P_0}(u).$$

Theorem D.8 (Candidate-policy support-function calculus)**.** *Assume a finite horizon, bounded known rewards, and a fixed finite candidate class Π. The candidate-optimized finite-class value $J^\star(P) = \max_{\pi\in\Pi} J^\pi(P)$ is Hadamard directionally differentiable at P_0 tangentially to $\mathcal{H}_G^0$, and hence tangentially to every subset such as the structured cone $\mathcal{C}_0$. Its derivative is*

$$\boxed{DJ^\star_{P_0}[u] = \max_{\pi\in\Pi_0} DJ^\pi_{P_0}[u] = \max_{\pi\in\Pi_0} \ell_\pi(u) = h_{\mathfrak{G}_0}(\mathsf{W}_0 u).} \tag{181}$$

For every $R < \infty$, uniformly over $u \in \mathcal{H}_G^0$ with $\|u\|_{\mathcal{I}_{0,w}} \le R$ and $0 \le t \le \tau^{\mathrm{Bell}}_{R,0}$,

$$|J^\star(P_0 + tu) - J^\star(P_0) - t h_{\mathfrak{G}_0}(\mathsf{W}_0 u)| \le \frac{Q_T R^2}{w_{\min}} t^2, \tag{182}$$

whenever $P_0 + tu$ is stochastic and lies in the active-set neighborhood of Lemma D.7. This straight interpolation is used only for the ambient stochastic Taylor expansion; it need not remain in the structured support/rank class. For every $\pi \in \Pi_0$,

$$\begin{aligned} |J^\star(P_0 + tu) - J^\pi(P_0 + tu) - t\,[h_{\mathfrak{G}_0}(\mathsf{W}_0 u) - \langle \gamma_\pi, \mathsf{W}_0 u\rangle]| \\ \le \frac{2Q_T R^2}{w_{\min}} t^2. \end{aligned} \tag{183}$$

Proof. Let $t_n \downarrow 0$, $u_n \to u$ in Fisher norm, and suppose $P_n := P_0 + t_n u_n$ is stochastic. By Lemma D.1, (u_n) is bounded in $\|\cdot\|_{\infty,1}$ and P_n eventually belongs to the active-set neighborhood. Hence

$$J^\star(P_n) = \max_{\pi\in\Pi_0} J^\pi(P_n).$$

Corollary D.4 gives, uniformly over the finite set Π_0,

$$J^\pi(P_n) = J^\star(P_0) + t_n \ell_\pi(u_n) + O(t_n^2).$$

Therefore

$$\frac{J^\star(P_n) - J^\star(P_0)}{t_n} = \max_{\pi\in\Pi_0} \ell_\pi(u_n) + O(t_n) \longrightarrow \max_{\pi\in\Pi_0} \ell_\pi(u).$$

The limiting map is a maximum of finitely many continuous linear functionals, so it is continuous and positively homogeneous. This proves Hadamard directional differentiability. The support-function identity follows from the definitions of γ_π and $\mathfrak{G}_0$.

For the quantitative expansion, Lemma D.7 and equation 164 give

$$\begin{aligned} J^\star(P_0 + tu) = \max_{\pi\in\Pi_0} \{J^\star(P_0) + t\ell_\pi(u) + R_\pi(t,u)\}, \\ |R_\pi(t,u)| \le Q_T R^2 t^2 / w_{\min}. \end{aligned}$$

Taking the maximum preserves this uniform error bound and proves equation 182. Subtracting the corresponding fixed-policy expansion for $\pi \in \Pi_0$ proves equation 183. □

Proposition D.9 (Regularity criterion at a policy tie)**.** *Let Π_0 be the finite active candidate set at P_0.*

1. *The map $J^\star$ is Fréchet differentiable at P_0 relative to the open affine neighborhood of normalized equivariant kernels if and only if*

$$\ell_\pi = \ell_{\pi'} \quad \text{on } \mathcal{H}_G^0 \quad \forall \pi, \pi' \in \Pi_0. \tag{184}$$

In this case the common functional is the derivative.

2. *On the structured cone, all first-order pairwise candidate-policy regrets vanish if and only if*

$$\langle \gamma_\pi - \gamma_{\pi'}, \theta \rangle = 0 \quad \forall \theta \in \mathcal{K}_0, \quad \forall \pi, \pi' \in \Pi_0. \tag{185}$$

Equivalently, the support function $h_{\mathfrak{G}_0}$ restricts to the same linear functional $\langle \gamma_\pi, \cdot \rangle$ on $\mathcal{K}_0$ for every active policy. For a binary tie with $g = (\gamma_+ - \gamma_-)/2$, this condition is equivalent to

$$\omega_{\mathcal{K}_0}(g) = 0. \tag{186}$$

Thus a policy tie is only a potential source of first-order nonregularity; genuine first-order nonsmoothness requires disagreement of active derivatives on feasible local directions.

Proof. For part 1, sufficiency follows from the uniform finite-family expansion

$$J^\pi(P_0 + h) = J^\star(P_0) + \ell_\pi(h) + o(\|h\|)$$

and active-set localization: if all active linear terms coincide, taking the finite maximum preserves the common linear expansion. Conversely, if $J^\star$ is Fréchet differentiable with derivative L, then its Hadamard directional derivative satisfies

$$\max_{\pi \in \Pi_0} \ell_\pi(u) = L(u) \quad \forall u \in \mathcal{H}_G^0.$$

Applying the same identity to $-u$ yields $\min_{\pi \in \Pi_0} \ell_\pi(u) = L(u)$. Hence every active functional equals L on every direction.

For part 2, the first-order regret of active policy π in whitened direction $\theta \in \mathcal{K}_0$ is

$$h_{\mathfrak{G}_0}(\theta) - \langle \gamma_\pi, \theta \rangle .$$

These quantities vanish for every active π and every feasible θ if and only if all active linear functionals agree on $\mathcal{K}_0$, which is exactly equation 185. In the binary case the difference is $2\langle g, \theta \rangle$, and the condition is equivalent to the vanishing of its absolute support functional $\omega_{\mathcal{K}_0}(g)$. □

The structured tangent cone generally contains directions whose straight lines leave the rank-constrained model. The following pathwise version is the form used after Appendix C constructs rank-preserving second-order corrections.

Corollary D.10 (Uniform expansion along corrected structured paths)**.** *Fix $R, C < \infty$. Suppose that, for every $u \in \mathcal{C}_0$ with $\|u\|_{\mathcal{I}_{0,w}} \le R$, a feasible path satisfies*

$$P_u(t) = P_0 + tu + e_u(t), \qquad \sup_{\|u\|_{\mathcal{I}_{0,w}} \le R} \|e_u(t)\|_{\mathcal{I}_{0,w}} \le Ct^2. \tag{187}$$

Define

$$M_{R,C} := \frac{L_T C}{\sqrt{w_{\min}}} + \frac{Q_T (R + C)^2}{w_{\min}}. \tag{188}$$

Then, uniformly over such u, $\pi \in \Pi_0$, and $0 \le t \le \tau_{R,C}^{\mathrm{Bell}}$,

$$|J^\star(P_u(t)) - J^\star(P_0) - t h_{\mathfrak{G}_0}(\mathsf{W}_0 u)| \le M_{R,C} t^2, \tag{189}$$

$$|J^\star(P_u(t)) - J^\pi(P_u(t)) - t\,[h_{\mathfrak{G}_0}(\mathsf{W}_0 u) - \langle \gamma_\pi, \mathsf{W}_0 u \rangle]| \le 2M_{R,C} t^2. \tag{190}$$

Proof. Set $h_u(t) := tu + e_u(t)$. By Lemma D.1, $\|h_u(t)\|_{\infty,1} \le w_{\min}^{-1/2}(Rt + Ct^2)$ uniformly. Moreover,

$$\left| DJ_{P_0}^\pi [e_u(t)] \right| \le L_T \|e_u(t)\|_{\infty,1} \le L_T C w_{\min}^{-1/2} t^2.$$

For $t \le 1$,

$$\|h_u(t)\|_{\infty,1} \le \frac{(R + C)t}{\sqrt{w_{\min}}}.$$

Corollary D.4 therefore bounds the quadratic Taylor remainder by $Q_T(R + C)^2 t^2 / w_{\min}$, while the linear contribution of $e_u(t)$ is at most $L_T C t^2 / \sqrt{w_{\min}}$. The radius $\tau_{R,C}^{\mathrm{Bell}}$ places the entire path in the active-face neighborhood by equation 180. Repeating the finite maximum argument from Theorem D.8 gives the first bound with constant $M_{R,C}$; subtracting the fixed-policy expansion gives the second bound with $2M_{R,C}$. □

When the paths are those of Theorem C.8, set

$$\tau_R^{\mathrm{all}} := \min\left\{\tau_R^{\mathrm{geom}}, \tau_{R,C_R^{\mathrm{geom}}}^{\mathrm{Bell}}\right\}. \tag{191}$$

Then geometric feasibility, strict positivity, active-policy localization, and the Bellman remainder bounds all hold simultaneously on $0 \le t \le \tau_R^{\mathrm{all}}$.

D.6 Statewise max–plus derivative for the complete deterministic policy class

This subsection has a different theorem contract: it specializes to the complete class of deterministic nonstationary Markov policies rather than the fixed candidate class used by the main minimax theorem. Let the optimal Bellman recursion at P_0 be

$$\begin{aligned} V_T^\star &:= r_T, \\ Q_t^\star(s,a) &:= r_t(s,a) + \sum_{s'} P_0(s' \mid s,a) V_{t+1}^\star(s'), \\ V_t^\star(s) &:= \max_{a\in\mathcal{A}} Q_t^\star(s,a), \qquad A_t^\star(s) := \arg\max_{a\in\mathcal{A}} Q_t^\star(s,a). \end{aligned} \tag{192}$$

Theorem D.11 (Max–plus Bellman derivative)**.** *For every row-sum-zero direction u, define recursively*

$$\begin{aligned} \dot V_T^u(s) &:= 0, \\ \dot Q_t^u(s,a) &:= \sum_{s'} u(s' \mid s,a) V_{t+1}^\star(s') + \sum_{s'} P_0(s' \mid s,a) \dot V_{t+1}^u(s'), \\ \dot V_t^u(s) &:= \max_{a\in A_t^\star(s)} \dot Q_t^u(s,a). \end{aligned} \tag{193}$$

Then the vector-valued map $P \mapsto V_t^\star(P)$ is Hadamard directionally differentiable at P_0, with derivative $\dot V_t^u$. In particular,

$$DJ_{P_0}^\star[u] = \langle \mu, \dot V_1^u \rangle. \tag{194}$$

Let

$$\Pi_0^{\mathrm{Bell}} := \{\pi : \pi_t(s) \in A_t^\star(s) \textit{ for every } t, s\}. \tag{195}$$

Then

$$\boxed{\langle \mu, \dot V_1^u \rangle = \max_{\pi\in\Pi_0^{\mathrm{Bell}}} \ell_\pi(u) = \max_{\pi\in\Pi_0} \ell_\pi(u).} \tag{196}$$

Proof. We argue backward. At $t = T$, the value is independent of P and the assertion is immediate. Suppose $P \mapsto V_{t+1}^\star(P)$ is Hadamard directionally differentiable with derivative $\dot V_{t+1}^u$. For each fixed (s,a), the map

$$P \longmapsto r_t(s,a) + \sum_{s'} P(s' \mid s,a) V_{t+1}^\star(P)(s')$$

is the composition of a continuously bilinear map with a Hadamard directionally differentiable map. Its directional derivative is therefore $\dot Q_t^u(s,a)$ in equation 193; this can also be verified directly by adding and subtracting $P_0 V_{t+1}^\star(P_0 + tu)$ and dividing by t.

For each state s, $V_t^\star(P)(s)$ is the maximum of finitely many action values. Repeating the finite-maximum argument from Theorem D.8 yields

$$DV_t^\star(P_0)[u](s) = \max_{a\in A_t^\star(s)} \dot Q_t^u(s,a),$$

which completes the induction and proves equation 194.

The recursion in equation 193 is itself a finite-horizon control problem with transition kernel P_0, admissible actions restricted to $A_t^\star(s)$, terminal reward zero, and stage reward

$$c_t^u(s,a) := \sum_{s'} u(s' \mid s,a) V_{t+1}^\star(s').$$

Dynamic programming therefore gives

$$\langle \mu, \dot{V}_1^u \rangle = \max_{\pi \in \Pi_0^{\mathrm{Bell}}} \mathbb{E}_{P_0}^{\pi} \left[\sum_{t=1}^{T-1} c_t^u(S_t, A_t) \right].$$

Every $\pi \in \Pi_0^{\mathrm{Bell}}$ satisfies $V_t^{\pi} = V_t^{\star}$ for all t, so the displayed expectation is exactly $\ell_\pi(u)$ by Proposition D.5. This proves the first equality in equation 196. The second follows because both sides equal the Hadamard derivative of the same optimized value: the left side by the statewise recursion just proved, and the right side by Theorem D.8 applied to the finite full policy class. □

Remark D.12 (What is new and what is imported). The fixed-policy derivative, sensitivity of optimized MDP values, and Hadamard calculus for finite maxima are established ideas (Kern et al., 2020; Shapiro, 1990). The role of this appendix is to place them in the exact coordinate and uniformity regime required by the paper: the constants are uniform over the finite policy class, the Fisher–Riesz gradient is explicitly projected onto the equivariant row-sum-zero tangent space, and the regret expansion remains valid along the second-order rank-preserving paths used by the singular local experiment. The new decision-theoretic claims begin only after this calculus is coupled to the structured tangent cone and the Gaussian minimax transfer.

E THE LOCAL GAUSSIAN EXPERIMENT

This appendix identifies the statistical limit of the structured transition experiment. Local asymptotic normality and comparison of experiments are classical tools due to Le Cam (Le Cam, 1960); we use the modern finite-dimensional formulations in van der Vaart (1998) and Le Cam & Yang (2000). Our contribution here is not the abstract LAN principle. It is the proof that the present deterministic-design, row-categorical experiment admits a *uniform* likelihood expansion over the nonconvex local model sets of Appendix C, together with the two interfaces needed later: convergence of every finite structured subexperiment to its Gaussian counterpart, and a uniform upper-risk transfer for Lipschitz Gaussian policy rules.

Throughout, P_0 is the strictly positive base kernel from Section 2. For each $x \in \mathcal{X}$, let $n_x = n_x(n)$ be deterministic, set

$$n := \sum_{x \in \mathcal{X}} n_x, \qquad w_{n,x} := \frac{n_x}{n}, \qquad \delta_n := \max_{x \in \mathcal{X}} |w_{n,x} - w_x|, \tag{197}$$

so that $\delta_n \to 0$ and $w_x \geq w_{\min} > 0$. The sample law is

$$\mathbb{P}_P^{(n)} := \bigotimes_{x \in \mathcal{X}} P(\cdot \mid x)^{\otimes n_x}. \tag{198}$$

All asymptotics are fixed dimensional and centered at the known kernel P_0. Unless displayed explicitly, constants in this appendix may depend on $(P_0, w, p_{\min}, w_{\min}, d, R)$ and the fixed finite row/state spaces, but not on n, the selected support/rank stratum, or the local kernel $P \in \Theta_{n,R}(P_0)$.

E.1 FISHER-ORTHONORMAL SCORES AND ROWWISE DIFFERENTIABILITY

Let $d := \dim \mathcal{H}_G^0$, let $(e_j)_{j=1}^d$ be the standard basis of $\mathbb{R}^d$, and define the Fisher-orthonormal perturbations

$$\psi_j := \mathsf{W}_0^{-1} e_j \in \mathcal{H}_G^0, \qquad \langle \psi_j, \psi_k \rangle_{\mathcal{I}_{0,w}} = \mathbf{1}\{j = k\}. \tag{199}$$

For $x \in \mathcal{X}$ and $s \in \mathcal{S}$, define the score vector

$$\xi_x(s) := \left(\frac{\psi_1(s \mid x)}{P_0(s \mid x)}, \ldots, \frac{\psi_d(s \mid x)}{P_0(s \mid x)} \right)^{\top} \in \mathbb{R}^d. \tag{200}$$

Lemma E.1 (Score identities and a uniform envelope). *For every $x \in \mathcal{X}$,*

$$\mathbb{E}_{Y \sim P_0(\cdot|x)}[\xi_x(Y)] = 0. \tag{201}$$

If

$$I_x := \mathbb{E}_{Y \sim P_0(\cdot|x)}[\xi_x(Y)\xi_x(Y)^{\top}], \tag{202}$$

then

$$\sum_{x\in\mathcal{X}} w_x I_x = I_d. \tag{203}$$

Moreover, with

$$B_\xi := \max_{x,s} \|\xi_x(s)\|_2, \tag{204}$$

one has the explicit bound

$$\boxed{B_\xi^2 \le \frac{d}{w_{\min} p_{\min}}.} \tag{205}$$

Proof. Because each ψ_j has zero row sum,

$$\mathbb{E}_{P_0(\cdot|x)}[\xi_{x,j}(Y)] = \sum_s \psi_j(s \mid x) = 0.$$

The (j,k) entry of the left side of equation 203 equals $\langle \psi_j, \psi_k \rangle_{\mathcal{I}_{0,w}} = \mathbf{1}\{j=k\}$. Finally, for every (x,s),

$$\begin{aligned} w_x P_0(s \mid x)\|\xi_x(s)\|_2^2 &= w_x \sum_{j=1}^{d} \frac{\psi_j(s \mid x)^2}{P_0(s \mid x)} \\ &\le \sum_{j=1}^{d} \sum_{x',s'} w_{x'} \frac{\psi_j(s' \mid x')^2}{P_0(s' \mid x')} = d. \end{aligned}$$

Use $w_x \ge w_{\min}$ and $P_0(s \mid x) \ge p_{\min}$. □

The next lemma records differentiability in quadratic mean. We include it to make the link with classical LAN theory explicit, although the stronger uniform logarithmic expansion is proved directly below.

Lemma E.2 (Uniform rowwise differentiability in quadratic mean)**.** *Fix $R < \infty$. Uniformly over $u \in \mathcal{H}_G^0$ satisfying $\|u\|_{\mathcal{I}_{0,w}} \le R$ and over rows $x \in \mathcal{X}$,*

$$\begin{aligned} \sum_{s\in\mathcal{S}} \Bigg[&\sqrt{P_0(s \mid x) + t u(s \mid x)} - \sqrt{P_0(s \mid x)} \\ &- \frac{t}{2} \frac{u(s \mid x)}{\sqrt{P_0(s \mid x)}} \Bigg]^2 = O_R(t^4) \end{aligned} \tag{206}$$

whenever $|t|$ is sufficiently small that the perturbed row is nonnegative. Thus the row score in direction u is $u(s \mid x)/P_0(s \mid x)$.

Proof. By Lemma D.1, Fisher-bounded u are uniformly bounded entrywise. Strict positivity of P_0 therefore places every $P_0(s \mid x) + t u(s \mid x)$ in $[p_{\min}/2, 1 + p_{\min}/2]$ for all sufficiently small $|t|$, uniformly in u and x. Taylor's theorem for $r \mapsto \sqrt{r}$ on this compact interval gives a coordinatewise remainder bounded by $C_R t^2 u(s \mid x)^2$. Squaring and summing over the finite state space yields equation 206. □

E.2 The observable central sequence and uniform LAN

Define the observable central sequence and empirical information matrix

$$Z_n := \frac{1}{\sqrt{n}} \sum_{x\in\mathcal{X}} \sum_{i=1}^{n_x} \xi_x(Y_{x,i}), \tag{207}$$

$$\widehat{I}_n := \frac{1}{n} \sum_{x\in\mathcal{X}} \sum_{i=1}^{n_x} \xi_x(Y_{x,i}) \xi_x(Y_{x,i})^\top, \tag{208}$$

and the deterministic design information

$$I_n := \sum_{x \in \mathcal{X}} w_{n,x} I_x. \tag{209}$$

Under P_0,

$$\mathbb{E}_{P_0}[Z_n] = 0, \qquad \mathrm{Cov}_{P_0}(Z_n) = I_n, \qquad \mathbb{E}_{P_0}[\widehat{I}_n] = I_n. \tag{210}$$

Furthermore,

$$\|I_n - I_d\|_{\mathrm{op}} \le |\mathcal{X}| B_\xi^2 \delta_n \longrightarrow 0, \tag{211}$$

and

$$\mathbb{E}_{P_0}\|\widehat{I}_n - I_n\|_{\mathrm{F}}^2 \le \frac{d^2 B_\xi^4}{n}. \tag{212}$$

For $R > 0$, define the actual structured local neighborhood and its whitened parameter set by

$$\Theta_{n,R}(P_0) := \left\{ P \in \Theta_{G,s,\boldsymbol{r}} : \left\| \mathsf{W}_0 \sqrt{n}(P - P_0) \right\|_2 \le R \right\}, \tag{213}$$

$$\mathfrak{K}_{n,R} := \left\{ \theta_n(P) := \mathsf{W}_0 \sqrt{n}(P - P_0) : P \in \Theta_{n,R}(P_0) \right\}. \tag{214}$$

Theorem E.3 (Uniform LAN over the structured local neighborhoods)**.** *Assume $P_0(s \mid x) \ge p_{\min} > 0$, the dimension d is fixed, and the deterministic row allocations satisfy $w_{n,x} \to w_x \ge w_{\min} > 0$. For every fixed $R < \infty$ and every*

$$n \ge 4R^2 B_\xi^2, \tag{215}$$

every $P \in \Theta_{n,R}(P_0)$ satisfies the exact representation

$$\log \frac{d\mathbb{P}_P^{(n)}}{d\mathbb{P}_{P_0}^{(n)}} = \theta_n(P)^\top Z_n - \frac{1}{2} \theta_n(P)^\top \widehat{I}_n \theta_n(P) + \rho_n(P), \tag{216}$$

where the remainder obeys the deterministic uniform bound

$$\sup_{P \in \Theta_{n,R}(P_0)} |\rho_n(P)| \le \frac{8R^3 B_\xi^3}{3\sqrt{n}}. \tag{217}$$

Consequently,

$$\boxed{\sup_{P \in \Theta_{n,R}(P_0)} \left| \log \frac{d\mathbb{P}_P^{(n)}}{d\mathbb{P}_{P_0}^{(n)}} - \theta_n(P)^\top Z_n + \frac{1}{2}\|\theta_n(P)\|_2^2 \right| \underset{P_0}{\longrightarrow} 0.} \tag{218}$$

Under $\mathbb{P}_{P_0}^{(n)}$,

$$Z_n \Longrightarrow N(0, I_d). \tag{219}$$

Proof. Fix $P \in \Theta_{n,R}(P_0)$, put $u := \sqrt{n}(P - P_0)$ and $\theta := \mathsf{W}_0 u$. Since $u = \sum_{j=1}^d \theta_j \psi_j$,

$$\frac{P(s \mid x)}{P_0(s \mid x)} = 1 + \frac{1}{\sqrt{n}} \theta^\top \xi_x(s). \tag{220}$$

The score envelope and $\|\theta\|_2 \le R$ ensure that the increment in equation 220 has absolute value at most $1/2$. Taylor's theorem and

$$\sup_{|z| \le 1/2} \left| \log(1+z) - z + \frac{z^2}{2} \right| \le \frac{8}{3}|z|^3$$

give equation 216. Summing the cubic envelope over the n observations gives equation 217. Therefore

$$\sup_{P \in \Theta_{n,R}(P_0)} \left| \log \frac{d\mathbb{P}_P^{(n)}}{d\mathbb{P}_{P_0}^{(n)}} - \theta_n(P)^\top Z_n + \frac{1}{2}\|\theta_n(P)\|_2^2 \right| \tag{221}$$

$$\le \frac{R^2}{2}\|\widehat{I}_n - I_n\|_{\mathrm{op}} + \frac{R^2}{2}\|I_n - I_d\|_{\mathrm{op}} + \frac{8R^3 B_\xi^3}{3\sqrt{n}}. \tag{222}$$

Equations equation 211 and equation 212 prove the uniform remainder claim. Finally, Z_n is a sum of independent mean-zero vectors, its covariance I_n converges to I_d, and every normalized summand has norm at most $B_\xi/\sqrt{n}$. The multivariate Lindeberg–Feller theorem yields equation 219. □

Remark E.4 (Why the expansion is genuinely uniform). The supremum in equation 218 ranges over the full nonconvex structured neighborhood, including changing block supports and rank-deficient strata. The proof does not differentiate a parametrization of that set. It uses only the ambient Fisher radius and strict positivity, while the geometry of the admissible means is supplied separately by Appendix C.

E.3 THE CENTRAL SEQUENCE UNDER LOCAL ALTERNATIVES

Theorem E.5 (Uniform shifted-central-sequence approximation)**.** *Assume the hypotheses of Theorem E.3 and fix $R < \infty$. For $P \in \Theta_{n,R}(P_0)$ and $\theta := \theta_n(P)$,*

$$\mathbb{E}_P[Z_n] = I_n\theta. \tag{223}$$

Moreover, with the explicit constant

$$C_R^{\mathrm{cov}} := \frac{B_\xi^2 R}{\sqrt{w_{\min}}} + B_\xi^4 R^2, \tag{224}$$

which is independent of n, P, and the selected structured stratum,

$$\sup_{P\in\Theta_{n,R}(P_0)} \|\mathrm{Cov}_P(Z_n) - I_n\|_{\mathrm{op}} \le \frac{C_R^{\mathrm{cov}}}{\sqrt{n}}. \tag{225}$$

Consequently, if $P_n \in \Theta_{n,R}(P_0)$ and $\theta_n(P_n) \to \theta$, then

$$Z_n \Longrightarrow N(\theta, I_d) \qquad \textit{under } \mathbb{P}_{P_n}^{(n)}. \tag{226}$$

In fact, writing

$$d_{\mathrm{BL}}(\mu,\nu) := \sup_{\substack{\|f\|_\infty\le 1\\ \mathrm{Lip}(f)\le 1}} \left| \int f\, d\mu - \int f\, d\nu \right|, \tag{227}$$

one has the uniform convergence

$$\boxed{\sup_{P\in\Theta_{n,R}(P_0)} d_{\mathrm{BL}}(\mathcal{L}_P(Z_n), N(\theta_n(P), I_d)) \longrightarrow 0.} \tag{228}$$

Proof. Let $u = \mathsf{W}_0^{-1}\theta$. From $P = P_0 + n^{-1/2}u$ and equation 200,

$$\mathbb{E}_P[\xi_x(Y)] = \frac{1}{\sqrt{n}} I_x\theta. \tag{229}$$

Summing with weights $w_{n,x}$ proves equation 223. The second moment changes from its P_0 value by at most $B_\xi^2\|P(\cdot \mid x) - P_0(\cdot \mid x)\|_1$. Lemma D.1 gives

$$\max_x \|P(\cdot \mid x) - P_0(\cdot \mid x)\|_1 \le \frac{R}{\sqrt{n w_{\min}}}.$$

Moreover, $\|I_x\|_{\mathrm{op}} \le B_\xi^2$, so the squared row means in equation 229 contribute at most $B_\xi^4 R^2/n$. Summing the second-moment perturbation with weights $w_{n,x}$ and using $\sum_x w_{n,x} = 1$ gives the first term in equation 224. Since $n^{-1} \le n^{-1/2}$, this proves equation 225.

For any local sequence, center each score summand by its mean under P_n. The normalized summands have norm at most $2B_\xi/\sqrt{n}$, their covariance converges to I_d, and hence the multivariate Lindeberg–Feller theorem gives a centered $N(0, I_d)$ limit. Equation equation 223 shifts the limit by θ.

To prove uniformity, suppose equation 228 fails. Then there are a subsequence and $P_n \in \Theta_{n,R}(P_0)$ for which the bounded-Lipschitz distance stays bounded away from zero. Compactness of $R\mathbb{B}_2$ gives a further subsequence with $\theta_n(P_n) \to \theta$. The preceding CLT implies $\mathcal{L}_{P_n}(Z_n) \Rightarrow N(\theta, I_d)$, while $N(\theta_n(P_n), I_d) \Rightarrow N(\theta, I_d)$, contradicting the triangle inequality for d_{BL}. □

Corollary E.6 (Local contiguity). *If $P_n \in \Theta_{n,R}(P_0)$ and $\theta_n(P_n) \to \theta$, then $\mathbb{P}_{P_n}^{(n)}$ and $\mathbb{P}_{P_0}^{(n)}$ are mutually contiguous. Under $\mathbb{P}_{P_0}^{(n)}$,*

$$\log \frac{d\mathbb{P}_{P_n}^{(n)}}{d\mathbb{P}_{P_0}^{(n)}} \Longrightarrow N\left(-\frac{1}{2}\|\theta\|_2^2, \|\theta\|_2^2\right), \tag{230}$$

and under $\mathbb{P}_{P_n}^{(n)}$ the same log likelihood ratio converges to

$$N\left(\frac{1}{2}\|\theta\|_2^2, \|\theta\|_2^2\right). \tag{231}$$

Proof. Combine Theorems E.3 and E.5. The first contiguity direction follows from Le Cam's first lemma because the limiting likelihood ratio is strictly positive and has expectation one. Contiguity transfers the LAN remainder to the local law; equation 226 then gives equation 231, and applying the first lemma to the inverse likelihood ratio yields the reverse direction (Le Cam & Yang, 2000; van der Vaart, 1998). □

E.4 The structured parameter cone and finite subexperiments

Proposition E.7 (Convergence of the local mean sets). *Let C_R^{geom} be the constant in Theorem C.9. For every fixed $R < \infty$ and all sufficiently large n,*

$$\boxed{\mathrm{d_H}(\mathfrak{K}_{n,R}, \mathcal{K}_0 \cap R\mathbb{B}_2) \le \frac{2C_R^{\mathrm{geom}}}{\sqrt{n}}.} \tag{232}$$

The Hausdorff distance is Euclidean.

Proof. Set $t = n^{-1/2}$ in Theorem C.9. The map W_0 is an isometry from the Fisher norm to the Euclidean norm, so it preserves Hausdorff distance and maps $\mathcal{C}_0 \cap R\mathbb{B}_{\mathcal{I}}$ to $\mathcal{K}_0 \cap R\mathbb{B}_2$. □

Lemma E.8 (Uniform second moments for a finite family of local paths). *Fix $m < \infty$. For $k \in [m]$, let $P_{n,k}$ be stochastic kernels dominated by P_0 and define*

$$u_{n,k} := \sqrt{n}\,(P_{n,k} - P_0). \tag{233}$$

Assume every $u_{n,k}$ has zero row sums and

$$B^2 := \sup_n \max_{1 \le k \le m} \|u_{n,k}\|_{\mathcal{I}_{0,w}}^2 < \infty. \tag{234}$$

Then, for all sufficiently large n,

$$\boxed{\max_{1 \le k \le m} \mathbb{E}_{P_0}^{(n)}\left[\left(\frac{d\mathbb{P}_{P_{n,k}}^{(n)}}{d\mathbb{P}_{P_0}^{(n)}}\right)^2\right] \le \exp(2B^2).} \tag{235}$$

Consequently, for each k, the likelihood-ratio sequence is uniformly integrable; because m is finite, this uniform integrability holds simultaneously over the embedded finite experiment.

Proof. Let

$$L_{n,k} := \frac{d\mathbb{P}_{P_{n,k}}^{(n)}}{d\mathbb{P}_{P_0}^{(n)}}.$$

For one observation from row x,

$$\frac{P_{n,k}(s \mid x)}{P_0(s \mid x)} = 1 + \frac{u_{n,k}(s \mid x)}{\sqrt{n}\,P_0(s \mid x)}.$$

Since $\sum_s u_{n,k}(s \mid x) = 0$,

$$\sum_s P_0(s \mid x)\left(1 + \frac{u_{n,k}(s \mid x)}{\sqrt{n}\, P_0(s \mid x)}\right)^2 = 1 + \frac{1}{n}\sum_s \frac{u_{n,k}(s \mid x)^2}{P_0(s \mid x)}.$$

Independence across observations and rows therefore gives the exact identity

$$\mathbb{E}^{(n)}_{P_0}[L^2_{n,k}] = \prod_x \left[1 + \frac{1}{n}\sum_s \frac{u_{n,k}(s \mid x)^2}{P_0(s \mid x)}\right]^{n_x} \tag{236}$$

$$\leq \exp\left\{\sum_x w_{n,x} \sum_s \frac{u_{n,k}(s \mid x)^2}{P_0(s \mid x)}\right\}, \tag{237}$$

where $w_{n,x} := n_x/n$ and $1 + z \leq e^z$ was used. Because the row set is finite and $w_{n,x}/w_x \to 1$ with $w_x > 0$, for all sufficiently large n,

$$\max_x \frac{w_{n,x}}{w_x} \leq 2.$$

The exponent in equation 237 is therefore at most $2\|u_{n,k}\|^2_{\mathcal{I}_{0,w}} \leq 2B^2$. The L^2 bound implies uniform integrability by Cauchy–Schwarz. □

Proposition E.9 (Embedding of finite Gaussian subexperiments)**.** *Fix $R < \infty$, $\eta > 0$, and a finite collection*

$$\theta_1, \ldots, \theta_m \in \mathcal{K}_0 \cap (R - \eta)\mathbb{B}_2. \tag{238}$$

Then there exist model sequences $P_{n,k} \in \Theta_{n,R}(P_0)$ such that

$$\theta_n(P_{n,k}) \longrightarrow \theta_k, \qquad k = 1, \ldots, m, \tag{239}$$

and, under $\mathbb{P}^{(n)}_{P_0}$,

$$\left(\log \frac{d\mathbb{P}^{(n)}_{P_{n,k}}}{d\mathbb{P}^{(n)}_{P_0}}\right)_{k=1}^m \Longrightarrow \left(\theta_k^\top Z - \frac{1}{2}\|\theta_k\|_2^2\right)_{k=1}^m, \qquad Z \sim N(0, I_d). \tag{240}$$

Consequently, the finite experiments indexed by $\{0, \theta_1, \ldots, \theta_m\}$ converge in Le Cam's weak sense to the corresponding Gaussian shift experiment.

Proof. Put $u_k := \mathsf{W}_0^{-1}\theta_k$. The path-lifting theorem C.8 gives

$$P_{n,k} := P_{u_k}(n^{-1/2}) = P_0 + n^{-1/2}u_k + O_{\mathcal{I}}(n^{-1}).$$

Hence equation 239 holds, and the strict radius margin η places $P_{n,k}$ in $\Theta_{n,R}(P_0)$ for all sufficiently large n. Moreover,

$$\sqrt{n}(P_{n,k} - P_0) = u_k + O_{\mathcal{I}}(n^{-1/2}),$$

so the finite family satisfies the uniform Fisher bound in Lemma E.8. Apply the uniform LAN expansion simultaneously to this finite collection and use $Z_n \Rightarrow N(0, I_d)$ to obtain equation 240. The limiting likelihood ratios have mean one, and Lemma E.8 supplies the uniform integrability needed by the likelihood-ratio criterion for weak convergence of finite experiments (Le Cam & Yang, 2000). □

E.5 Decision-specific transfer interfaces

For $q \in \Delta(\Pi_0)$, define

$$J^q(P) := \sum_{\pi \in \Pi_0} q(\pi) J^\pi(P), \qquad \bar{\gamma}_q := \sum_{\pi \in \Pi_0} q(\pi)\gamma_\pi, \tag{241}$$

and the Gaussian local-regret loss

$$L_0(\theta, q) := h_{\mathfrak{G}_0}(\theta) - \langle \bar{\gamma}_q, \theta \rangle \,. \tag{242}$$

Let

$$\Gamma_0 := \max_{\pi \in \Pi_0} \|\gamma_\pi\|_2. \tag{243}$$

Then, for $\|\theta\|_2 \le R$,

$$0 \le L_0(\theta, q) \le 2R\Gamma_0. \tag{244}$$

Proposition E.10 (Uniform reduction of candidate-policy regret to Gaussian loss)**.** *For every fixed $R < \infty$, all sufficiently large n, every $P \in \Theta_{n,R}(P_0)$, and every $q \in \Delta(\Pi_0)$,*

$$\boxed{\left|\sqrt{n}\big[J^\star(P) - J^q(P)\big] - L_0\big(\theta_n(P), q\big)\right| \le \frac{2Q_T R^2}{w_{\min}\sqrt{n}}.} \tag{245}$$

Proof. For $P \in \Theta_{n,R}(P_0)$, put $u := \sqrt{n}(P - P_0)$ and $t := n^{-1/2}$. Then $P = P_0 + tu$ and $\|u\|_{\mathcal{I}_{0,w}} \le R$. For all large n, Theorem D.8 applies and gives, uniformly over $\pi \in \Pi_0$,

$$\left|\sqrt{n}\big[J^\star(P) - J^\pi(P)\big] - L_0\big(\theta_n(P), \delta_\pi\big)\right| \le \frac{2Q_T R^2}{w_{\min}\sqrt{n}}.$$

Average this inequality with weights $q(\pi)$. □

Proposition E.11 (Upper-risk transfer for Lipschitz Gaussian candidate-policy rules)**.** *Let $\delta : \mathbb{R}^d \to \Delta(\Pi_0)$ be Borel and globally Lipschitz from Euclidean distance to ℓ_1 distance, with constant $\mathrm{Lip}(\delta) < \infty$. Use the observable data rule $q_n := \delta(Z_n)$. Then*

$$\boxed{\sup_{P \in \Theta_{n,R}(P_0)} \Big| \mathbb{E}_P\big[\sqrt{n}\big(J^\star(P) - J^{q_n}(P)\big)\big] - \mathbb{E}_{Z \sim N(\theta_n(P), I_d)}\left[L_0\big(\theta_n(P), \delta(Z)\big)\right] \Big| \longrightarrow 0.} \tag{246}$$

Proof. Proposition E.10 reduces the claim to the central-sequence loss. For fixed $\theta \in R\mathbb{B}_2$, the function

$$f_\theta(z) := L_0(\theta, \delta(z))$$

is uniformly bounded by $2R\Gamma_0$. Moreover,

$$\begin{aligned} |f_\theta(z) - f_\theta(z')| &\le \|\theta\|_2 \|\bar{\gamma}_{\delta(z)} - \bar{\gamma}_{\delta(z')}\|_2 \\ &\le R\Gamma_0 \|\delta(z) - \delta(z')\|_1 \\ &\le R\Gamma_0 \,\mathrm{Lip}(\delta) \|z - z'\|_2. \end{aligned}$$

Thus the family $(f_\theta)_{\|\theta\| \le R}$ is uniformly bounded and uniformly Lipschitz. More precisely, set

$$A_{R,\delta} := \max\{1, 2R\Gamma_0, R\Gamma_0 \,\mathrm{Lip}(\delta)\} \,.$$

Then $f_\theta / A_{R,\delta}$ has sup norm and Lipschitz constant at most one, uniformly in θ. Hence the absolute expectation difference is at most $A_{R,\delta}$ times the bounded–Lipschitz distance in equation 228, which converges to zero uniformly in the local kernel. □

Remark E.12 (Exact scope of the transfer proved here). Proposition E.9 is the lower-bound interface: every finite Gaussian subexperiment can be embedded by valid structured model paths. Proposition E.11 is the upper-bound interface for a fixed Lipschitz Gaussian rule. This appendix does *not* claim full deficiency convergence of the compact nonconvex experiments. Appendix F must still prove that Gaussian ε-minimax rules can be approximated by Lipschitz rules without increasing worst-case risk, and then combine the two interfaces to obtain the matching local minimax equality.

F FIXED-RADIUS LOCAL ASYMPTOTIC MINIMAX CANDIDATE-POLICY REGRET

This appendix completes the local decision-theoretic reduction. The classical Hájek–Le Cam local asymptotic minimax principle transfers lower bounds from a limit experiment to a sequence of statistical experiments (Le Cam & Yang, 2000; van der Vaart, 1998). The setting here requires two additional arguments. First, the limiting parameter set is a compact section of a generally nonconvex sparse determinantal cone, and the action is a randomized policy rather than an estimator in a linear space. Second, Appendix E proves only the two problem-specific transfer interfaces needed here, not full deficiency convergence. We therefore establish the minimax equality directly: finite Gaussian subexperiments yield the lower bound, while Gaussian mollification produces Lipschitz near-minimax rules to which the upper-risk transfer applies.

Throughout, fix the task $\mathfrak{T}$, the strictly positive center P_0, and a radius $R \in (0, \infty)$. All state/action spaces, dimensions, structural budgets, the candidate class, and the limiting design are fixed before $n \to \infty$. The local rule may use P_0, W_0, and Π_0. Unless displayed explicitly, constants may depend on these fixed objects and on R, but not on n, the local parameter, or the rule under evaluation. Write

$$\mathcal{K}_R := \mathcal{K}_0 \cap R\mathbb{B}_2, \qquad \Gamma_0 := \max_{\pi \in \Pi_0} \|\gamma_\pi\|_2. \tag{247}$$

By Theorem C.5, $\mathcal{K}_0$ is a finite union of closed cones; hence $\mathcal{K}_R$ is compact. For a Borel Gaussian rule $\delta : \mathbb{R}^d \to \Delta(\Pi_0)$, define

$$\mathcal{G}_\delta(\theta) := \mathbb{E}_{Z \sim N(\theta, I_d)} \left[L_0\big(\theta, \delta(Z)\big) \right], \qquad \theta \in \mathcal{K}_R, \tag{248}$$

and recall

$$\mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_0) = \inf_{\delta} \sup_{\theta \in \mathcal{K}_R} \mathcal{G}_\delta(\theta). \tag{249}$$

All infima over Gaussian rules are over Borel randomized Markov kernels into the candidate-policy simplex; deterministic Borel maps are included as a special case.

F.1 COMPACTNESS, SADDLE POINTS, AND FINITE PARAMETER REDUCTIONS

The first lemma provides a uniform modulus of continuity that is independent of the decision rule.

Lemma F.1 (Uniform regularity of Gaussian risks)**.** *For every Borel rule δ and every $\theta, \eta \in R\mathbb{B}_2$,*

$$0 \le \mathcal{G}_\delta(\theta) \le 2R\Gamma_0, \tag{250}$$

$$|\mathcal{G}_\delta(\theta) - \mathcal{G}_\delta(\eta)| \le C_R \|\theta - \eta\|_2, \qquad C_R := \Gamma_0(R + 2). \tag{251}$$

Consequently, the family $\{\mathcal{G}_\delta : \delta \text{ Borel}\}$ is uniformly bounded and equicontinuous on $R\mathbb{B}_2$.

Proof. The envelope is equation 244. Put

$$m_\delta(\theta) := \mathbb{E}_{N(\theta, I_d)} \big[\bar{\gamma}_{\delta(Z)} \big], \qquad \|m_\delta(\theta)\|_2 \le \Gamma_0.$$

Then

$$\mathcal{G}_\delta(\theta) = h_{\mathfrak{G}_0}(\theta) - \langle \theta, m_\delta(\theta) \rangle.$$

The support function is Γ_0-Lipschitz. Furthermore,

$$\begin{aligned} \|m_\delta(\theta) - m_\delta(\eta)\|_2 &\le \Gamma_0 \int_{\mathbb{R}^d} |\varphi_d(z - \theta) - \varphi_d(z - \eta)| \, dz \\ &\le \Gamma_0 \|\theta - \eta\|_2. \end{aligned}$$

The last inequality follows from Pinsker's inequality because $\mathrm{KL}(N(\theta, I_d), N(\eta, I_d)) = \|\theta - \eta\|_2^2/2$; it remains valid when $\|\theta - \eta\|_2 > 2$ because an L^1 distance is at most two. Therefore,

$$|\langle \theta, m_\delta(\theta) \rangle - \langle \eta, m_\delta(\eta) \rangle| \le \Gamma_0 \|\theta - \eta\|_2 + R\Gamma_0 \|\theta - \eta\|_2.$$

Adding the support-function difference proves equation 251. □

The next proposition is not required for the asymptotic transfer itself, but it clarifies the structure of the Gaussian game. In particular, nonconvexity of $\mathcal{K}_R$ causes no minimax obstruction once the adversary is allowed to randomize over parameters.

Proposition F.2 (Gaussian minimax rule and least-favorable prior)**.** *Let $\mathcal{P}(\mathcal{K}_R)$ denote the Borel probability measures on $\mathcal{K}_R$. Then*

$$\boxed{\mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_0) = \min_{\delta} \max_{\theta \in \mathcal{K}_R} \mathcal{G}_\delta(\theta) = \max_{\nu \in \mathcal{P}(\mathcal{K}_R)} \min_{\delta} \int_{\mathcal{K}_R} \mathcal{G}_\delta(\theta)\, \nu(d\theta).} \tag{252}$$

A minimax rule and a least-favorable prior both exist. For a fixed prior ν, a Bayes rule may be chosen to put all mass on

$$\arg\max_{\pi \in \Pi_0} \langle \gamma_\pi, m_\nu(z) \rangle, \qquad m_\nu(z) := \frac{\int_{\mathcal{K}_R} \theta e^{\theta^\top z - \|\theta\|_2^2/2}\, \nu(d\theta)}{\int_{\mathcal{K}_R} e^{\theta^\top z - \|\theta\|_2^2/2}\, \nu(d\theta)}. \tag{253}$$

Proof. Let $\mu_0 = N(0, I_d)$ and identify a decision rule with an $|\Pi_0|$-tuple in $L^\infty(\mu_0)$ satisfying nonnegativity and unit sum μ_0-almost everywhere. This set, denoted $\mathfrak{D}$, is convex and weak-star compact by Banach–Alaoglu. To see that the simplex constraints are weak-star closed, observe that $f \geq 0$ is equivalent to $\int fg\, d\mu_0 \geq 0$ for every nonnegative $g \in L^1(\mu_0)$, and the unit-sum constraint is the intersection of the weak-star closed hyperplanes $\int (\sum_\pi f_\pi - 1) g\, d\mu_0 = 0$, $g \in L^1(\mu_0)$. Every L^∞ equivalence class admits a Borel representative on the standard Borel space $\mathbb{R}^d$. Moreover, every Gaussian shift $N(\theta, I_d)$ is equivalent to μ_0, so one common μ_0-null-set convention defines the rule under all parameters.

For $\theta \in \mathcal{K}_R$, the density relative to μ_0 is

$$\Lambda_\theta(z) = \exp\left(\theta^\top z - \frac{1}{2}\|\theta\|_2^2\right),$$

which lies in $L^1(\mu_0)$. Hence, for fixed θ, $\delta \mapsto \mathcal{G}_\delta(\theta)$ is weak-star continuous and affine. The map $\delta \mapsto \sup_{\theta \in \mathcal{K}_R} \mathcal{G}_\delta(\theta)$ is therefore weak-star lower semicontinuous, so its minimum is attained on $\mathfrak{D}$.

For $\nu \in \mathcal{P}(\mathcal{K}_R)$, set

$$\Psi(\delta, \nu) := \int_{\mathcal{K}_R} \mathcal{G}_\delta(\theta)\, \nu(d\theta).$$

By Lemma F.1, Ψ is continuous affine in ν under weak convergence. For fixed ν, Fubini's theorem shows that Ψ is weak-star continuous affine in δ. Sion's minimax theorem therefore gives (Sion, 1958)

$$\min_{\delta \in \mathfrak{D}} \max_{\nu \in \mathcal{P}(\mathcal{K}_R)} \Psi(\delta, \nu) = \max_{\nu \in \mathcal{P}(\mathcal{K}_R)} \min_{\delta \in \mathfrak{D}} \Psi(\delta, \nu).$$

The maximum over priors on the left equals the maximum over point masses. The Bayes value is upper semicontinuous in ν, being the infimum of continuous functions, and $\mathcal{P}(\mathcal{K}_R)$ is compact; hence a least-favorable prior exists.

For the Bayes rule, condition on $Z = z$. The term $\mathbb{E}[h_{\mathfrak{G}_0}(\theta) \mid Z = z]$ is independent of the action, whereas the action-dependent term is $-\langle \bar{\gamma}_q, m_\nu(z) \rangle$. Maximizing this linear functional over the policy simplex yields equation 253. A Borel tie-breaking rule exists because the policy set is finite. □

For a finite set $F \subseteq \mathcal{K}_R$, let

$$\mathfrak{R}(F) := \inf_{\delta} \max_{\theta \in F} \mathcal{G}_\delta(\theta). \tag{254}$$

Corollary F.3 (Finite-subexperiment exhaustion)**.** *For every $R > 0$,*

$$\boxed{\mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_0) = \sup_{\substack{F \subseteq \mathcal{K}_R \\ F \text{ finite}}} \mathfrak{R}(F) = \sup_{0 < \eta < R} \sup_{\substack{F \subseteq \mathcal{K}_0 \cap (R-\eta)\mathbb{B}_2 \\ F \text{ finite}}} \mathfrak{R}(F).} \tag{255}$$

Proof. The inequality $\mathfrak{R}(F) \leq \mathfrak{R}_R$ is immediate. Let F_ε be a finite ε-net of $\mathcal{K}_R$. Lemma F.1 gives, uniformly over δ,

$$\sup_{\theta \in \mathcal{K}_R} \mathcal{G}_\delta(\theta) \leq \max_{\eta \in F_\varepsilon} \mathcal{G}_\delta(\eta) + C_R \varepsilon.$$

Taking the infimum over δ and then letting $\varepsilon \downarrow 0$ proves the first equality. For the second, scale a finite set $F \subseteq \mathcal{K}_R$ to $(1-a)F$ with $a \in (0,1)$. Because $\mathcal{K}_0$ is a cone, $(1-a)F \subseteq \mathcal{K}_0 \cap (R - aR)\mathbb{B}_2$, and the uniform Lipschitz bound gives $|\mathfrak{R}(F) - \mathfrak{R}((1-a)F)| \leq C_R a R$. Let $a \downarrow 0$. □

F.2 Lipschitz regularization of Gaussian decision rules

Appendix E transfers the upper risk for fixed globally Lipschitz Gaussian rules. The following result shows that this restriction does not alter the Gaussian minimax value.

Let $\phi_{\sigma,d}$ denote the density of $N(0,\sigma^2 I_d)$. For a Borel rule $\delta = (\delta_\pi)_{\pi\in\Pi_0}$ and $\sigma > 0$, define its Gaussian mollification componentwise by

$$\delta_\pi^{(\sigma)}(z) := \int_{\mathbb{R}^d} \delta_\pi(z-y)\phi_{\sigma,d}(y)\,dy. \tag{256}$$

Proposition F.4 (Risk-dense Lipschitz rules)**.** *For every Borel rule δ and every $\sigma > 0$:*

1. *$\delta^{(\sigma)}$ is a C^∞ map from $\mathbb{R}^d$ to $\Delta(\Pi_0)$;*
2. *it is globally Lipschitz from Euclidean distance to ℓ_1 distance, with*

$$\mathrm{Lip}(\delta^{(\sigma)}) \le |\Pi_0|\,\|\nabla\phi_{\sigma,d}\|_{L^1} = \frac{|\Pi_0|\,\mathbb{E}\|G\|_2}{\sigma}, \qquad G\sim N(0,I_d); \tag{257}$$

3. *its risk satisfies*

$$\boxed{\sup_{\theta\in\mathcal{K}_R} |\mathcal{G}_{\delta^{(\sigma)}}(\theta) - \mathcal{G}_\delta(\theta)| \le R\Gamma_0\sqrt{d\big[\sigma^2 - \log(1+\sigma^2)\big]} \xrightarrow[\sigma\downarrow 0]{} 0.} \tag{258}$$

Consequently,

$$\boxed{\mathfrak{R}_R(\mathcal{K}_0,\mathfrak{G}_0) = \inf_{\substack{\delta:\mathbb{R}^d\to\Delta(\Pi_0)\\ \delta \text{ globally Lipschitz}}} \sup_{\theta\in\mathcal{K}_R} \mathcal{G}_\delta(\theta).} \tag{259}$$

Proof. Convolution preserves nonnegativity and the identity $\sum_\pi \delta_\pi = 1$, proving that the mollified rule is simplex-valued. Standard differentiation under the integral gives smoothness and equation 257.

Let $g_\delta(z) := \bar\gamma_{\delta(z)}$, so that $\|g_\delta(z)\|_2 \le \Gamma_0$. If $Z\sim N(\theta, I_d)$ and $Y\sim N(0,\sigma^2 I_d)$ are independent, Fubini's theorem gives

$$\mathbb{E}_\theta[g_{\delta^{(\sigma)}}(Z)] = \mathbb{E}[g_\delta(Z+Y)],$$

where $Z+Y\sim N(\theta,(1+\sigma^2)I_d)$. The support-function term cancels, so

$$|\mathcal{G}_{\delta^{(\sigma)}}(\theta) - \mathcal{G}_\delta(\theta)| \le R\Gamma_0\left\|N(0,(1+\sigma^2)I_d) - N(0,I_d)\right\|_{L^1}.$$

Pinsker's inequality and

$$\mathrm{KL}\left(N(0,(1+\sigma^2)I_d)\middle\|N(0,I_d)\right) = \frac{d}{2}\left[\sigma^2 - \log(1+\sigma^2)\right]$$

prove equation 258. Given an arbitrary ε-minimax Borel rule, choose σ small enough that the right side is at most ε. This proves equation 259. □

F.3 Finite-experiment lower transfer

We use the following finite-experiment lemma. It is a decision-theoretic form of the likelihood-ratio criterion for weak convergence of experiments (Le Cam & Yang, 2000).

Lemma F.5 (Lower semicontinuity under finite experiment convergence)**.** *Let $\mathcal{A}$ be a compact metric action space. For each n, let $\{P_{n,k} : 0\le k\le m\}$ be a finite dominated experiment with likelihood ratios*

$$\Lambda_{n,k} := \frac{dP_{n,k}}{dP_{n,0}}, \qquad k\in[m].$$

Assume that, under $P_{n,0}$,

$$(\Lambda_{n,1},\dots,\Lambda_{n,m}) \Longrightarrow (\Lambda_1,\dots,\Lambda_m), \qquad \mathbb{E}[\Lambda_k] = 1, \tag{260}$$

and that $(\Lambda_{n,k})_n$ is uniformly integrable for every k. Let $\ell_{n,k}, \ell_k : \mathcal{A} \to [0, M]$ be continuous and satisfy

$$\max_{1\le k\le m} \sup_{a\in\mathcal{A}} |\ell_{n,k}(a) - \ell_k(a)| \longrightarrow 0. \tag{261}$$

Then

$$\boxed{\liminf_{n\to\infty} \inf_{\delta_n} \max_{1\le k\le m} \mathbb{E}_{P_{n,k}}[\ell_{n,k}(\delta_n)] \ge \inf_{\delta} \max_{1\le k\le m} \mathbb{E}_{Q_k}[\ell_k(\delta)],} \tag{262}$$

where Q_0 is the law of $\Lambda = (\Lambda_1, \dots, \Lambda_m)$, $dQ_k/dQ_0 = \Lambda_k$, and the infima are over randomized decision rules.

Proof. Take a subsequence realizing the left-side liminf and rules δ_n whose maximum risks are within $1/n$ of the corresponding infima. Realize any randomization on an enlarged sample space and let $A_n \in \mathcal{A}$ denote the resulting action. Tightness of Λ_n and compactness of $\mathcal{A}$ give a further subsequence such that, under $P_{n,0}$,

$$(\Lambda_n, A_n) \Longrightarrow (\Lambda, A).$$

For every k, uniform integrability of $\Lambda_{n,k}$, boundedness of the losses, and equation 261 yield

$$\mathbb{E}_{P_{n,k}}[\ell_{n,k}(A_n)] = \mathbb{E}_{P_{n,0}}[\Lambda_{n,k}\ell_{n,k}(A_n)] \longrightarrow \mathbb{E}[\Lambda_k \ell_k(A)]. \tag{263}$$

Let $K(da \mid \lambda)$ be a regular conditional law of A given $\Lambda = \lambda$. This kernel is a randomized rule in the canonical limit experiment, and its risk under Q_k equals the right side of equation 263. Its maximal risk is at least the minimax risk of that limit experiment. Taking the maximum over k and then the chosen liminf subsequence proves equation 262. □

F.4 The local asymptotic minimax theorem

A data rule is a Markov kernel from the observation array to the simplex of candidate policies. For a realized external lottery q, one candidate policy is drawn and executed for the full episode, and its conditional expected value is $J^q(P) = \sum_\pi q(\pi) J^\pi(P)$. Every expectation below includes both data randomness and internal randomization of the Markov kernel. Define the unrestricted and base-optimal local risks

$$\mathcal{R}^{\mathrm{all}}_{n,R}(P_0) := \inf_{\delta_n:\, \mathrm{data}\rightsquigarrow\Delta(\Pi)} \sup_{P\in\Theta_{n,R}(P_0)} \mathbb{E}_P\left[\sqrt{n}\big(J^\star(P) - J^{\delta_n}(P)\big)\right], \tag{264}$$

$$\mathcal{R}^{0}_{n,R}(P_0) := \inf_{\delta_n:\, \mathrm{data}\rightsquigarrow\Delta(\Pi_0)} \sup_{P\in\Theta_{n,R}(P_0)} \mathbb{E}_P\left[\sqrt{n}\big(J^\star(P) - J^{\delta_n}(P)\big)\right]. \tag{265}$$

Lemma F.6 (Policies outside the base-optimal face are locally dominated)**.** *For every fixed $R < \infty$, there exists $n_0(R) < \infty$ such that*

$$\mathcal{R}^{\mathrm{all}}_{n,R}(P_0) = \mathcal{R}^{0}_{n,R}(P_0), \qquad n \ge n_0(R). \tag{266}$$

Proof. If $\Pi = \Pi_0$, the claim is immediate. Otherwise fix $\pi^\circ \in \Pi_0$ and recall the strict gap Δ_0 from equation 175. For $P \in \Theta_{n,R}(P_0)$, Lemma D.1 and Corollary D.4 give, uniformly in π,

$$|J^\pi(P) - J^\pi(P_0)| \le \varepsilon_{n,R} := \frac{L_T R}{\sqrt{n w_{\min}}}.$$

Hence, for every $\pi \notin \Pi_0$,

$$J^{\pi^\circ}(P) - J^\pi(P) \ge \Delta_0 - 2\varepsilon_{n,R} > 0$$

for all sufficiently large n. Given any rule on Π, move all mass it assigns outside Π_0 to π°. This transformation weakly increases the selected policy value at every sample and every local model. Thus it cannot increase regret. The reverse inequality follows because rules on Π_0 form a subclass of rules on Π. □

Theorem F.7 (Fixed-radius local asymptotic minimax candidate-policy regret)**.** *Assume finite state/action spaces and horizon, bounded known rewards, a fixed finite candidate class, a known strictly positive center in the exact G-equivariant sparse low-rank transition class, deterministic row*

allocations with $w_{n,x} \to w_x > 0$, fixed dimension, and oracle access to P_0, W_0, and Π_0. Allow randomized external candidate-policy lotteries. Then, for every fixed $R \in (0, \infty)$,

$$\boxed{\lim_{n\to\infty} \mathcal{R}^{\mathrm{all}}_{n,R}(P_0) = \lim_{n\to\infty} \mathcal{R}^{0}_{n,R}(P_0) = \mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_0).} \tag{267}$$

More precisely:

1. *every sequence of data rules satisfies the Gaussian lower bound;*
2. *for every $\varepsilon > 0$, there exists a fixed globally Lipschitz Gaussian rule δ_ε such that the observable rule $\delta_\varepsilon(Z_n)$ has asymptotic worst-case local risk at most $\mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_0) + \varepsilon$;*
3. *there exists a single diagonal sequence of observable rules whose worst-case local risk converges to $\mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_0)$.*

The result is pointwise, fixed-radius, fixed-dimensional, and oracle-local. The diagonal attaining sequence is existential; it is not asserted to be a computable finite-sample algorithm.

Proof. By Lemma F.6, it suffices to analyze $\mathcal{R}^0_{n,R}$.

Lower bound. Fix $0 < \eta < R$ and a finite set $F = \{\theta_1, \dots, \theta_m\} \subseteq \mathcal{K}_0 \cap (R - \eta)\mathbb{B}_2$. Proposition E.9 gives model sequences $P_{n,k} \in \Theta_{n,R}(P_0)$ whose finite likelihood-ratio experiment converges to the Gaussian experiment at F. For $q \in \Delta(\Pi_0)$, put

$$\ell_{n,k}(q) := \sqrt{n}\big[J^\star(P_{n,k}) - J^q(P_{n,k})\big], \qquad \ell_k(q) := L_0(\theta_k, q).$$

Proposition E.10, convergence of the local means, and compactness of the policy simplex imply

$$\max_k \sup_q |\ell_{n,k}(q) - \ell_k(q)| \longrightarrow 0,$$

and the losses are uniformly bounded. The embedded finite family satisfies the uniform Fisher-radius hypothesis of Lemma E.8; hence its likelihood-ratio vector is uniformly integrable by the explicit second-moment bound equation 235. The canonical likelihood-ratio limit is equivalent to the finite Gaussian shift experiment on F, because the likelihood-ratio vector is a sufficient statistic for that finite family. Lemma F.5 therefore gives

$$\liminf_{n\to\infty} \mathcal{R}^0_{n,R}(P_0) \geq \mathfrak{R}(F).$$

Take the supremum over all such finite F and use Corollary F.3 to obtain

$$\liminf_{n\to\infty} \mathcal{R}^0_{n,R}(P_0) \geq \mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_0). \tag{268}$$

Upper bound. Fix $\varepsilon > 0$. Proposition F.4 provides a globally Lipschitz rule δ_ε satisfying

$$\sup_{\theta \in \mathcal{K}_R} \mathcal{G}_{\delta_\varepsilon}(\theta) \leq \mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_0) + \varepsilon. \tag{269}$$

Use the observable rule $q_n = \delta_\varepsilon(Z_n)$. Proposition E.11 gives

$$\sup_{P \in \Theta_{n,R}(P_0)} \mathbb{E}_P\left[\sqrt{n}\big(J^\star(P) - J^{q_n}(P)\big)\right] \leq \sup_{\theta \in \mathfrak{K}_{n,R}} \mathcal{G}_{\delta_\varepsilon}(\theta) + o(1).$$

By Proposition E.7 and the uniform Lipschitz bound in Lemma F.1,

$$\sup_{\theta \in \mathfrak{K}_{n,R}} \mathcal{G}_{\delta_\varepsilon}(\theta) \longrightarrow \sup_{\theta \in \mathcal{K}_R} \mathcal{G}_{\delta_\varepsilon}(\theta).$$

Therefore,

$$\limsup_{n\to\infty} \mathcal{R}^0_{n,R}(P_0) \leq \mathfrak{R}_R(\mathcal{K}_0, \mathfrak{G}_0) + \varepsilon. \tag{270}$$

Let $\varepsilon \downarrow 0$ and combine with equation 268.

For the diagonal sequence, choose Lipschitz rules δ_j with Gaussian risk at most $\mathfrak{R}_R + 1/j$. For each j, the preceding transfer supplies N_j such that the data rule $\delta_j(Z_n)$ has local risk at most $\mathfrak{R}_R + 2/j$ for all $n \geq N_j$. Taking (N_j) strictly increasing and using $\delta_{j(n)}(Z_n)$, where $j(n) := \max\{j : N_j \leq n\}$, gives a sequence with limiting upper risk at most $\mathfrak{R}_R$. Its limiting lower risk is at least $\mathfrak{R}_R$ because every rule dominates the finite-sample minimax value and that minimax value converges to $\mathfrak{R}_R$. This proves the final assertion. □

Remark F.8 (Why this is stronger than a lower-bound-only LAN statement). Uniform LAN alone would imply, at most, a lower bound after embedding finite Gaussian subexperiments. Equality in equation 267 additionally uses three facts that are specific to the present problem: uniform convergence of the structured local mean sets, risk-density of Lipschitz Gaussian policy rules, and the observable central-sequence upper transfer. No full deficiency convergence of the nonconvex compact experiments is asserted or required.

Remark F.9 (Interpretation). The theorem identifies the exact first-order decision problem at a joint model and policy singularity. The cone $\mathcal{K}_0$ determines which local transition perturbations are feasible, the Gaussian shift determines which are statistically distinguishable, and the support-function loss generated by $\mathfrak{G}_0$ determines which perturbations change the optimal decision. The limit is therefore an cone-alignment-sensitive statistical game, not a scalar parameter count.

G Symmetry reduction and task hierarchy

Symmetry-aware transition models and MDP homomorphisms motivate equivariant policies, invariant value functions, and value-preserving abstractions (van der Pol et al., 2020; Panangaden et al., 2024). Our question is more local: when the transition experiment is already restricted to the exact G-equivariant model of Section 2, which multiplicity blocks can influence a given task at first order? The answer is controlled by the fixed spaces of the task stabilizer subgroup. We work throughout under the real-Schur-type convention of Appendix B.

G.1 Subgroup projectors and fixed isotypic coordinates

Let $H \leq G$. For $\rho \in \widehat{G}_{\mathbb{R}}$, let $\rho : G \to O(V_\rho)$ denote the chosen orthogonal irreducible representation and define the subgroup Reynolds operator

$$Q_{\rho,H} := \frac{1}{|H|} \sum_{h \in H} \rho(h), \qquad \kappa_H(\rho) := \operatorname{rank}(Q_{\rho,H}). \tag{271}$$

Write

$$V_\rho^H := \{v \in V_\rho : \rho(h)v = v \; \forall h \in H\}.$$

Lemma G.1 (Subgroup Reynolds projector)**.** *The operator $Q_{\rho,H}$ is the orthogonal projector onto V_ρ^H. Consequently,*

$$\kappa_H(\rho) = \dim V_\rho^H = \frac{1}{|H|} \sum_{h \in H} \chi_\rho(h) = \left\langle \operatorname{Res}_H^G \chi_\rho, \mathbf{1}_H \right\rangle_H. \tag{272}$$

If $H_0 \supseteq H_1$, then

$$V_\rho^{H_0} \subseteq V_\rho^{H_1}, \qquad \kappa_{H_0}(\rho) \leq \kappa_{H_1}(\rho). \tag{273}$$

Proof. Since ρ is orthogonal,

$$Q_{\rho,H}^\top = \frac{1}{|H|} \sum_{h \in H} \rho(h^{-1}) = Q_{\rho,H}.$$

Moreover,

$$Q_{\rho,H}^2 = \frac{1}{|H|^2} \sum_{h,k \in H} \rho(hk) = \frac{1}{|H|} \sum_{g \in H} \rho(g) = Q_{\rho,H},$$

because every $g \in H$ has exactly $|H|$ representations as hk. Thus $Q_{\rho,H}$ is an orthogonal projector. For every $h_0 \in H$, $\rho(h_0)Q_{\rho,H} = Q_{\rho,H}$, so its image is contained in V_ρ^H. Conversely, $Q_{\rho,H}v = v$ for every $v \in V_\rho^H$. This proves the projector claim. Taking traces gives the character formula; the final identity is the definition of the character inner product on H. The inclusion in equation 273 is immediate from the definitions. □

Fix an orthonormal basis $f_{\rho,H,1},\dots,f_{\rho,H,\kappa_H(\rho)}$ of V_ρ^H. For every multiplicity m, the H-fixed vectors in the ρ-isotypic module satisfy

$$(V_\rho \otimes \mathbb{R}^m)^H = V_\rho^H \otimes \mathbb{R}^m. \tag{274}$$

Hence every z in this fixed space has a unique expansion

$$z = \sum_{a=1}^{\kappa_H(\rho)} f_{\rho,H,a} \otimes z_a, \qquad z_a \in \mathbb{R}^m, \qquad \|z\|_2^2 = \sum_a \|z_a\|_2^2. \tag{275}$$

These coefficient vectors are the only representation coordinates permitted by subgroup fixed-space information; a particular controlled process may realize a strict subset.

G.2 Propagation of task symmetry

Definition G.2 (H-symmetric task and policy)**.** A task is H-invariant if

$$\mu(hs) = \mu(s), \qquad r_t(hs,ha) = r_t(s,a) \quad (t<T), \qquad r_T(hs) = r_T(s) \tag{276}$$

for all $h \in H$. A deterministic nonstationary Markov policy is H-equivariant if

$$\pi_t(hs) = h\pi_t(s) \qquad \forall h \in H,\ s \in \mathcal{S},\ t < T. \tag{277}$$

The theorem below is conditional on the existence of such a deterministic policy; stabilizers of states can obstruct deterministic equivariant selection. The stochastic-policy analogue follows by replacing equation 277 by $\pi_t(ha \mid hs) = \pi_t(a \mid s)$, but is not needed by the core paper.

Proposition G.3 (Invariance of laws, occupancies, and continuation values)**.** *Assume P_0 is G-equivariant, the task is H-invariant, and π is H-equivariant. Then, for every stage t,*

$$\nu_t^\pi(hs) = \nu_t^\pi(s), \qquad d_t^\pi(hs,ha) = d_t^\pi(s,a), \tag{278}$$

and

$$V_t^\pi(hs) = V_t^\pi(s), \qquad Q_t^\pi(hs,ha) = Q_t^\pi(s,a). \tag{279}$$

Equivalently, ν_t^π and V_t^π are H-fixed vectors in $\mathbb{R}^{\mathcal{S}}$, while d_t^π and Q_t^π are H-fixed vectors in $\mathbb{R}^{\mathcal{X}}$.

Proof. The initial law is H-invariant by assumption. Suppose ν_t^π is H-invariant. Using deterministic policy equivariance and transition equivariance, for $h \in H$,

$$\begin{aligned}
\nu_{t+1}^\pi(hs') &= \sum_{s\in\mathcal{S}} \nu_t^\pi(s) P_0\big(hs' \mid s, \pi_t(s)\big) \\
&= \sum_{z\in\mathcal{S}} \nu_t^\pi(hz) P_0\big(hs' \mid hz, \pi_t(hz)\big) \\
&= \sum_{z\in\mathcal{S}} \nu_t^\pi(z) P_0\big(s' \mid z, \pi_t(z)\big) = \nu_{t+1}^\pi(s').
\end{aligned}$$

The occupancy identity follows from $d_t^\pi(s,a) = \nu_t^\pi(s)\mathbf{1}\{a = \pi_t(s)\}$.

For values, $V_T^\pi = r_T$ is H-invariant. If V_{t+1}^π is H-invariant, then a change of variables $s' = hz$ gives

$$\begin{aligned}
Q_t^\pi(hs,ha) &= r_t(hs,ha) + \sum_{s'} P_0(s' \mid hs, ha) V_{t+1}^\pi(s') \\
&= r_t(s,a) + \sum_z P_0(z \mid s,a) V_{t+1}^\pi(z) = Q_t^\pi(s,a).
\end{aligned}$$

Finally, $V_t^\pi(hs) = Q_t^\pi(hs, \pi_t(hs)) = Q_t^\pi(hs, h\pi_t(s)) = V_t^\pi(s)$. □

G.3 Block factorization of the transition derivative

Let $\Psi_{\mathcal{S}}$ and $\Psi_{\mathcal{X}}$ be the orthogonal isotypic maps from Appendix B. Write $\widehat{d}^{\pi}_{t,\rho}$ and $\widehat{v}^{\pi}_{t+1,\rho}$ for the ρ-isotypic coordinates of d^{π}_t and V^{π}_{t+1}, respectively. By Proposition G.3 and equation 274, there are unique matrices

$$D_{t,\pi,\rho} \in \mathbb{R}^{\kappa_H(\rho)\times m^{\mathcal{X}}_{\rho}}, \qquad W_{t+1,\pi,\rho} \in \mathbb{R}^{\kappa_H(\rho)\times m^{\mathcal{S}}_{\rho}} \tag{280}$$

such that, writing rows as coefficient vectors,

$$\widehat{d}^{\pi}_{t,\rho} = \sum_{a=1}^{\kappa_H(\rho)} f_{\rho,H,a} \otimes D_{t,\pi,\rho}[a,:]^{\top}, \tag{281}$$

$$\widehat{v}^{\pi}_{t+1,\rho} = \sum_{a=1}^{\kappa_H(\rho)} f_{\rho,H,a} \otimes W_{t+1,\pi,\rho}[a,:]^{\top}. \tag{282}$$

When $\kappa_H(\rho) = 0$, both matrices have zero rows and the displayed vectors vanish.

Theorem G.4 (Potentially visible raw derivative blocks and rank hierarchy)**.** *Under the assumptions of Proposition G.3, let $u = \mathcal{B}_G(H_{\mathbf{1}}, (H_\rho)_{\rho\in\mathfrak{R}}) \in \mathcal{H}^0_G$. Then the fixed-policy derivative admits the block representation*

$$\ell_\pi(u) = \ell_{\pi,\mathbf{1}}(H_{\mathbf{1}}) + \sum_{\rho\in\mathfrak{R}} \langle G_{\pi,\rho}, H_\rho\rangle_{\mathrm{F}}\,, \tag{283}$$

where the nontrivial raw gradient blocks are

$$\boxed{G_{\pi,\rho} = \sum_{t=1}^{T-1} D^{\top}_{t,\pi,\rho} W_{t+1,\pi,\rho}.} \tag{284}$$

Consequently,

$$\boxed{\operatorname{rank}(G_{\pi,\rho}) \le \min\left\{m^{\mathcal{X}}_{\rho}, m^{\mathcal{S}}_{\rho}, (T-1)\kappa_H(\rho)\right\}.} \tag{285}$$

In particular,

$$\kappa_H(\rho) = 0 \quad\Longrightarrow\quad G_{\pi,\rho} = 0. \tag{286}$$

For two H-equivariant policies π, π' under the same H-invariant task,

$$\boxed{\operatorname{rank}\big(G_{\pi,\rho} - G_{\pi',\rho}\big) \le \min\left\{m^{\mathcal{X}}_{\rho}, m^{\mathcal{S}}_{\rho}, 2(T-1)\kappa_H(\rho)\right\}.} \tag{287}$$

All zero-block and rank statements are invariant under orthogonal changes of multiplicity bases.

Proof. By Proposition D.5,

$$\ell_\pi(u) = \sum_{t=1}^{T-1} \left\langle d^{\pi}_t, K_u V^{\pi}_{t+1}\right\rangle_{\mathbb{R}^{\mathcal{X}}}.$$

In the orthogonal isotypic coordinates of Appendix B, K_u acts on the ρ-summand as $I_{V_\rho} \otimes H_\rho$. Substituting equation 281–equation 282 and using orthonormality of the $f_{\rho,H,a}$ gives the stage-t contribution

$$\begin{aligned}\left\langle \widehat{d}^{\pi}_{t,\rho}, (I_{V_\rho}\otimes H_\rho)\widehat{v}^{\pi}_{t+1,\rho}\right\rangle &= \sum_{a=1}^{\kappa_H(\rho)} D_{t,\pi,\rho}[a,:] H_\rho W_{t+1,\pi,\rho}[a,:]^{\top} \\ &= \left\langle D^{\top}_{t,\pi,\rho} W_{t+1,\pi,\rho}, H_\rho\right\rangle_{\mathrm{F}}.\end{aligned}$$

Summing over stages and irreducible sectors proves equation 283 and equation 284. Every summand in equation 284 has rank at most $\kappa_H(\rho)$, giving equation 285 by subadditivity of rank. Equation equation 286 is the zero-row case. Finally, $\operatorname{rank}(A-B) \le \operatorname{rank}(A) + \operatorname{rank}(B)$ gives equation 287. Orthogonal left-right changes of multiplicity bases preserve zero status and rank. □

Proposition G.5 (Algebraic sharpness of the rank envelope)**.** *Fix integers $m_X, m_S, L, \kappa \geq 1$. Every matrix $M \in \mathbb{R}^{m_X \times m_S}$ with*

$$\operatorname{rank}(M) \leq L\kappa$$

admits a factorization

$$M = \sum_{t=1}^{L} D_t^\top W_t, \qquad D_t \in \mathbb{R}^{\kappa \times m_X}, \quad W_t \in \mathbb{R}^{\kappa \times m_S}. \tag{288}$$

Hence the coefficient $(T-1)\kappa_H(\rho)$ in equation 285 cannot be improved using subgroup fixed spaces alone.

Proof. Take a singular-value decomposition $M = \sum_{j=1}^{q} \sigma_j u_j v_j^\top$ with $q \leq L\kappa$. Partition the q rank-one terms into L groups of size at most κ. For group t, place the vectors $\sqrt{\sigma_j} u_j^\top$ and $\sqrt{\sigma_j} v_j^\top$ in corresponding rows of D_t and W_t, padding unused rows by zero. Then $D_t^\top W_t$ is exactly the sum of the rank-one terms in that group. □

The sharpness proposition is algebraic: actual MDP occupancies and continuation values may impose additional dynamical restrictions.

G.4 A SUBGROUP HIERARCHY OF POTENTIALLY VISIBLE SECTORS

Define the potentially visible nontrivial sectors

$$\operatorname{Vis}(H) := \{\rho \in \mathfrak{R} : \kappa_H(\rho) > 0\}. \tag{289}$$

For one policy and for a pairwise gradient difference, respectively, define the representation-theoretic envelopes

$$\mathcal{E}_H^{(1)} := \{(G_\rho)_\rho : G_\rho = 0 \text{ if } \rho \notin \operatorname{Vis}(H),\ \operatorname{rank}(G_\rho) \leq (T-1)\kappa_H(\rho)\}, \tag{290}$$

$$\mathcal{E}_H^{(2)} := \{(G_\rho)_\rho : G_\rho = 0 \text{ if } \rho \notin \operatorname{Vis}(H),\ \operatorname{rank}(G_\rho) \leq 2(T-1)\kappa_H(\rho)\}. \tag{291}$$

Ambient row and column dimensions are understood in these definitions.

Corollary G.6 (Monotone task hierarchy)**.** *Every raw gradient generated by an H-invariant task and an H-equivariant policy belongs to $\mathcal{E}_H^{(1)}$, and every pairwise raw gradient difference belongs to $\mathcal{E}_H^{(2)}$. If $H_0 \supseteq H_1$, then*

$$\operatorname{Vis}(H_0) \subseteq \operatorname{Vis}(H_1), \qquad \mathcal{E}_{H_0}^{(j)} \subseteq \mathcal{E}_{H_1}^{(j)}, \quad j \in \{1, 2\}. \tag{292}$$

Moreover, the class of H_0-invariant tasks is contained in the class of H_1-invariant tasks, and every H_0-equivariant policy is H_1-equivariant.

Proof. The first claim is Theorem G.4. The inclusions follow from Lemma G.1. The final class inclusions are immediate because $H_1 \subseteq H_0$. □

Corollary G.7 (Full-symmetry collapse)**.** *If $H = G$, then*

$$G_{\pi,\rho} = 0 \qquad \forall \rho \in \mathfrak{R} \tag{293}$$

for every G-invariant task and G-equivariant policy. Thus every nontrivial transition block is raw first-order invisible; only the trivial-isotypic stochastic nuisance block can affect the fixed-policy value to first order.

Proof. For a nontrivial irreducible G-representation, the fixed subspace V_ρ^G is a G-invariant subspace. By irreducibility it is either zero or all of V_ρ; the latter would make ρ trivial. Hence $\kappa_G(\rho) = 0$ for every $\rho \in \mathfrak{R}$, and equation 293 follows from equation 286. □

The hierarchy is a statement about which raw derivative sectors are *permitted* to be potentially decision relevant. It does not imply that every sector with $\kappa_H(\rho) > 0$ is realized by a given dynamics, nor that local minimax risk is monotone when tasks with different rewards are compared.

G.5 Compatibility with Fisher whitening and structural width

Potential visibility is intrinsically a statement about the raw derivative functional in equation 283. Translating it to the Fisher–Riesz or whitened gradients requires additional structure on the Fisher metric.

Proposition G.8 (Visibility under separable block Fisher geometry)**.** *Suppose that, in the multiplicity coordinates of Appendix B, the Fisher form has no cross-block terms and, on each nontrivial block, has the separable form*

$$\langle H_\rho, K_\rho \rangle_{\mathcal{I}} = \mathrm{tr}\big(H_\rho^\top L_\rho^{\mathrm{F}} K_\rho R_\rho^{\mathrm{F}}\big), \qquad L_\rho^{\mathrm{F}} \succ 0, \quad R_\rho^{\mathrm{F}} \succ 0. \tag{294}$$

If the raw derivative block is $G_{\pi,\rho}$, then its Riesz and whitened representatives are

$$R_{\pi,\rho} = (L_\rho^{\mathrm{F}})^{-1} G_{\pi,\rho} (R_\rho^{\mathrm{F}})^{-1}, \qquad \widetilde{G}_{\pi,\rho} = (L_\rho^{\mathrm{F}})^{-1/2} G_{\pi,\rho} (R_\rho^{\mathrm{F}})^{-1/2}. \tag{295}$$

Consequently, zero blocks and ranks are preserved:

$$G_{\pi,\rho} = 0 \iff \widetilde{G}_{\pi,\rho} = 0, \qquad \mathrm{rank}(\widetilde{G}_{\pi,\rho}) = \mathrm{rank}(G_{\pi,\rho}). \tag{296}$$

Proof. For every perturbation H_ρ,

$$\mathrm{tr}\big(R_{\pi,\rho}^\top L_\rho^{\mathrm{F}} H_\rho R_\rho^{\mathrm{F}}\big) = \langle G_{\pi,\rho}, H_\rho \rangle_{\mathrm{F}}$$

if and only if $L_\rho^{\mathrm{F}} R_{\pi,\rho} R_\rho^{\mathrm{F}} = G_{\pi,\rho}$, proving the Riesz formula. Under the rank-preserving whitening $H_\rho \mapsto (L_\rho^{\mathrm{F}})^{1/2} H_\rho (R_\rho^{\mathrm{F}})^{1/2}$, the Euclidean gradient is $(L_\rho^{\mathrm{F}})^{-1/2} G_{\pi,\rho} (R_\rho^{\mathrm{F}})^{-1/2}$. Since the left and right factors are invertible, zero status and rank are preserved. □

Corollary G.9 (Task-conditioned sparse low-rank width)**.** *Assume the maximally singular setting and the separable Fisher geometry of Theorem H.8. Suppose a binary base tie $\{\pi_+, \pi_-\}$ consists of H-equivariant policies under an H-invariant task. Let*

$$G_\rho := \frac{1}{2}\left(G_{\pi_+,\rho} - G_{\pi_-,\rho}\right), \qquad \widetilde{G}_\rho := (L_\rho^{\mathrm{F}})^{-1/2} G_\rho (R_\rho^{\mathrm{F}})^{-1/2}, \tag{297}$$

and define the task-conditioned effective rank

$$r_{H,\rho}^{\mathrm{dec}} := \min\left\{r_\rho, 2(T-1)\kappa_H(\rho), m_\rho^{\mathcal{X}}, m_\rho^{\mathcal{S}}\right\}. \tag{298}$$

Then the nontrivial block scores in the exact structural-width formula are

$$b_{H,\rho}^2 := \sum_{j=1}^{r_{H,\rho}^{\mathrm{dec}}} \sigma_j(\widetilde{G}_\rho)^2, \qquad b_{H,\rho} = 0 \quad \textit{if } \rho \notin \mathrm{Vis}(H), \tag{299}$$

and

$$\boxed{\omega_{\mathcal{K}_0}(g)^2 = b_{\mathbf{1}}^2 + \sum_{\rho \in \mathrm{Top}_{\le s}(b_H)} b_{H,\rho}^2,} \tag{300}$$

where the top-s operation ranges only over $\mathrm{Vis}(H)$. In particular, task symmetry removes invisible sectors before block sparsity selects the largest positive task-conditioned block scores.

Proof. Theorem G.4 and Proposition G.8 give $\mathrm{rank}(\widetilde{G}_\rho) \le 2(T-1)\kappa_H(\rho)$ and zero blocks outside $\mathrm{Vis}(H)$. Substitute these facts into the exact separable metric formula of Theorem H.8. □

For $H = G$, all nontrivial scores in equation 300 vanish. If the centered trivial-isotypic nuisance gradient also vanishes, then the entire first-order Gaussian policy-regret game is degenerate and a higher-order analysis is necessary.

G.6 Why raw visibility is not whitened support in general

Example G.10 (Cross-block Fisher coupling)**.** Consider two scalar parameter blocks $z = (z_1, z_2)$ with Fisher inner product

$$\langle z, z' \rangle_{F_\eta} := z^\top \begin{bmatrix} 1 & \eta \\ \eta & 1 \end{bmatrix} z', \qquad |\eta| < 1. \tag{301}$$

Let the raw derivative functional be $L(z) = z_1$, so the second raw block is invisible. Its Fisher–Riesz representative is

$$F_\eta^{-1} e_1 = \frac{1}{1-\eta^2} \begin{bmatrix} 1 \\ -\eta \end{bmatrix}, \tag{302}$$

which has a nonzero second coordinate whenever $\eta \neq 0$. Thus an invisible raw block can acquire a nonzero Riesz or whitened coordinate when the Fisher metric couples blocks. This is why Theorem G.4 is stated for raw derivative blocks and Corollary G.9 explicitly assumes separable Fisher geometry.

Appendix B gives an explicit symmetric Markov experiment in which the Fisher form couples inequivalent representation blocks; group invariance of the sampling experiment alone is therefore insufficient.

G.7 Payoff-null coordinates can remain statistically useful

A further distinction is required. A coordinate that does not enter the policy loss directly need not be statistically disposable when the structured parameter set couples it to a decision-relevant coordinate.

Example G.11 (No projection onto the gradient span without product geometry)**.** Consider the Gaussian experiment

$$Z \sim N(\theta, I_2), \qquad \theta \in \mathcal{K} := \{t(1,1) : t \in \mathbb{R}\}, \tag{303}$$

with binary gradients $\{-e_1, e_1\}$. The loss depends only on θ_1; the second observation coordinate is payoff-null. Restrict the parameter set to the equal-prior pair

$$\theta_+ = a(1,1), \qquad \theta_- = -a(1,1), \qquad a > 0.$$

Using the full observation, the optimal testing error is

$$\Phi(-\sqrt{2}\, a), \tag{304}$$

whereas a rule restricted to Z_1 has optimal error

$$\Phi(-a). \tag{305}$$

Since selecting the wrong policy costs $2a$, the corresponding Bayes policy regrets are $2a\Phi(-\sqrt{2}a)$ and $2a\Phi(-a)$, respectively, with strict inequality. The second coordinate helps because the cone imposes $\theta_2 = \theta_1$, even though the payoff functional annihilates that coordinate.

Consequently, one may quotient or project away decision-null coordinates only after proving an appropriate product or sufficiency structure for the model cone and statistical experiment. The Bellman–Le Cam geometry keeps the complete structured cone for precisely this reason.

G.8 Scope of the symmetry reduction

The results above make three distinctions explicit. First, exact G-equivariance of the transition model does not itself produce a nontrivial G-action on its multiplicity parameters: every admissible transition operator is already an intertwiner. The hierarchy arises from the stabilizer H of the task and policy, through the restricted fixed spaces V_ρ^H. Second, the envelopes equation 290–equation 291 are necessary representation-theoretic constraints, not claims that all admissible blocks are dynamically realizable. Third, converting raw task visibility into a statement about whitened gradient support requires block-separable Fisher geometry. Subject to those qualifications, subgroup branching gives a canonical coarse-to-fine hierarchy: enlarging task symmetry removes transition sectors and lowers their maximum decision-gradient rank, while breaking symmetry can release additional sectors and singular directions.

H STRUCTURED GAUSSIAN CANDIDATE-POLICY GAMES

This appendix isolates the Gaussian decision problem that appears as the local limit of the structured reinforcement-learning experiment. We first define a finite-action Gaussian candidate-policy game, then derive a sharp binary lower bound in terms of a deterministic support functional. We next compute that functional exactly for unions of block-sparse, block-low-rank cones. The matrix calculation is based on the classical trace inequality of von Neumann (1937) and the singular-value truncation principles of Eckart & Young (1936); Mirsky (1960); we include complete proofs because the precise constrained support function is used later without loss. The binary testing calculation is an explicit specialization of the Neyman–Pearson likelihood-ratio argument (Neyman & Pearson, 1933).

Throughout, all Euclidean spaces are finite dimensional, all policy sets are finite, and all decision rules are Borel measurable. For a nonempty closed cone $\mathcal{K} \subseteq \mathbb{R}^d$, write

$$\mathcal{K}_R := \mathcal{K} \cap R\mathbb{B}_2, \qquad \mathbb{B}_2 := \{\theta \in \mathbb{R}^d : \|\theta\|_2 \le 1\}.$$

A cone is called *centrally symmetric* when $\mathcal{K} = -\mathcal{K}$. Convexity is not assumed; the sparse low-rank cones considered below are generally nonconvex.

H.1 FINITE-ACTION GAUSSIAN CANDIDATE-POLICY GAMES

Let $\mathcal{P}$ be a finite set of candidate policies and associate each $\pi \in \mathcal{P}$ with a vector $\gamma_\pi \in \mathbb{R}^d$. Define the Bellman-gradient polytope and its support function by

$$\mathfrak{G} := \operatorname{conv}\{\gamma_\pi : \pi \in \mathcal{P}\}, \qquad h_{\mathfrak{G}}(\theta) := \max_{\pi \in \mathcal{P}} \langle \gamma_\pi, \theta \rangle . \tag{306}$$

For $q \in \Delta(\mathcal{P})$, let

$$\bar{\gamma}_q := \sum_{\pi \in \mathcal{P}} q(\pi)\gamma_\pi, \qquad L_{\mathfrak{G}}(\theta, q) := h_{\mathfrak{G}}(\theta) - \langle \bar{\gamma}_q, \theta \rangle . \tag{307}$$

The loss is nonnegative because $\bar{\gamma}_q \in \mathfrak{G}$.

A randomized decision rule is a Borel map

$$\delta : \mathbb{R}^d \longrightarrow \Delta(\mathcal{P}).$$

Under parameter $\theta \in \mathcal{K}_R$, the observation follows

$$Z \sim \mathcal{N}(\theta, I_d). \tag{308}$$

The radius-R Gaussian candidate-policy risk is

$$\mathfrak{R}_R(\mathcal{K}, \mathfrak{G}) := \inf_{\delta} \sup_{\theta \in \mathcal{K}_R} \mathbb{E}_\theta \left[L_{\mathfrak{G}}(\theta, \delta(Z)) \right]. \tag{309}$$

Here and below, the infimum is over all Borel randomized decision rules. We do not require the infimum to be attained in this appendix.

Lemma H.1 (Elementary properties of the Gaussian game)**.** *Let $M_{\mathfrak{G}} := \max_{\pi \in \mathcal{P}} \|\gamma_\pi\|_2$. Then:*

1. $0 \le L_{\mathfrak{G}}(\theta, q) \le 2M_{\mathfrak{G}} \|\theta\|_2$ *for every $(\theta, q) \in \mathbb{R}^d \times \Delta(\mathcal{P})$; consequently,* $0 \le \mathfrak{R}_R(\mathcal{K}, \mathfrak{G}) \le 2RM_{\mathfrak{G}}$.
2. *If $v \in \mathbb{R}^d$ and every gradient is translated to $\gamma'_\pi = \gamma_\pi + v$, then*
$$L_{\mathfrak{G}+v}(\theta, q) = L_{\mathfrak{G}}(\theta, q), \qquad \mathfrak{R}_R(\mathcal{K}, \mathfrak{G} + v) = \mathfrak{R}_R(\mathcal{K}, \mathfrak{G}).$$
Thus the game depends only on pairwise gradient differences.
3. *Replacing the finite set $\{\gamma_\pi\}_{\pi \in \mathcal{P}}$ by its convex hull does not change the support function or the loss.*

Proof. The lower bound in part 1 follows from $\bar{\gamma}_q \in \mathfrak{G}$. For the upper bound,

$$h_{\mathfrak{G}}(\theta) \le M_{\mathfrak{G}} \|\theta\|_2 , \qquad |\langle \bar{\gamma}_q, \theta \rangle| \le M_{\mathfrak{G}} \|\theta\|_2 ,$$

because $\|\bar{\gamma}_q\|_2 \le M_{\mathfrak{G}}$. Taking the supremum over $\mathcal{K}_R$ proves the risk bound. For part 2,

$$h_{\mathfrak{G}+v}(\theta) = h_{\mathfrak{G}}(\theta) + \langle v, \theta \rangle , \qquad \overline{\gamma'}_q = \bar{\gamma}_q + v,$$

so the two additional linear terms cancel. Part 3 follows directly from the definition of a support function and from $\bar{\gamma}_q \in \mathfrak{G}$. □

H.2 Eliminating orthogonal product nuisances

The following lemma identifies the precise situation in which a payoff-null observation coordinate can be discarded. It does not contradict the coupled counterexample in Appendix G: the conclusion requires an orthogonal product cone and parameter-free nuisance noise after restricting nature to the zero-nuisance submodel.

Lemma H.2 (Orthogonal product-nuisance invariance)**.** *Let $E = E_{\mathrm{dec}} \oplus E_{\mathrm{nui}}$ be an orthogonal Euclidean direct sum, let $\mathcal{K}_{\mathrm{dec}} \subseteq E_{\mathrm{dec}}$ be a nonempty closed cone, and fix $R > 0$. Consider the parameter set*

$$\mathcal{K}_R^{\mathrm{prod}} := \left\{(\theta, \eta) : \theta \in \mathcal{K}_{\mathrm{dec}},\ \eta \in E_{\mathrm{nui}},\ \|\theta\|_2^2 + \|\eta\|_2^2 \le R^2\right\}. \tag{310}$$

Observe independent coordinates

$$Z_{\mathrm{dec}} \sim N(\theta, I_{E_{\mathrm{dec}}}), \qquad Z_{\mathrm{nui}} \sim N(\eta, I_{E_{\mathrm{nui}}}). \tag{311}$$

Let $\mathcal{P}$ be finite and let $L(\theta, q)$ be a bounded measurable loss on $(\mathcal{K}_{\mathrm{dec}} \cap R\mathbb{B}_{E_{\mathrm{dec}}}) \times \Delta(\mathcal{P})$ that is affine in q and independent of η. Then

$$\boxed{\begin{aligned}\inf_{\delta} \sup_{(\theta,\eta)\in\mathcal{K}_R^{\mathrm{prod}}} \mathbb{E}_{\theta,\eta} L\big(\theta, \delta(Z_{\mathrm{dec}}, Z_{\mathrm{nui}})\big) \\ = \inf_{\delta_{\mathrm{dec}}} \sup_{\substack{\theta\in\mathcal{K}_{\mathrm{dec}} \\ \|\theta\|_2\le R}} \mathbb{E}_{\theta} L\big(\theta, \delta_{\mathrm{dec}}(Z_{\mathrm{dec}})\big).\end{aligned}} \tag{312}$$

The first infimum is over Borel simplex-valued rules on E and the second over such rules on E_{dec}.

Proof. Let V_{full} and V_{dec} denote the two sides. Given a decision-only rule δ_{dec}, lift it to the full observation by ignoring Z_{nui}. Its risk depends on (θ, η) only through θ, and the projection of equation 310 onto the decision coordinate is exactly $\mathcal{K}_{\mathrm{dec}} \cap R\mathbb{B}_{E_{\mathrm{dec}}}$. Hence $V_{\mathrm{full}} \le V_{\mathrm{dec}}$.

Conversely, fix a full rule δ and restrict nature to the valid submodel $\eta = 0$. Under that submodel, $Z_{\mathrm{nui}} \sim N(0, I)$ is independent of both θ and Z_{dec}. Define, componentwise on the policy simplex,

$$\bar{\delta}(z) := \int_{E_{\mathrm{nui}}} \delta(z, y)\, \varphi_{E_{\mathrm{nui}}}(y)\, dy, \tag{313}$$

where $\varphi_{E_{\mathrm{nui}}}$ is the standard Gaussian density. The map $\bar{\delta}$ is Borel and simplex valued. Affinity of L in the lottery and Fubini's theorem give, for every admissible θ,

$$\mathbb{E}_{\theta,0} L\big(\theta, \delta(Z_{\mathrm{dec}}, Z_{\mathrm{nui}})\big) = \mathbb{E}_{\theta} L\big(\theta, \bar{\delta}(Z_{\mathrm{dec}})\big).$$

Therefore the worst-case risk of δ over the full product set is at least the decision-only worst-case risk of $\bar{\delta}$, and hence at least V_{dec}. Taking the infimum over δ proves the reverse inequality. □

H.3 Binary ties and the restricted decision width

Consider two tied policies with gradients γ_+ and γ_-. By Lemma H.1, subtracting their midpoint does not change the game. Hence we may write

$$\gamma_+ = g, \qquad \gamma_- = -g, \qquad g := \frac{\gamma_+ - \gamma_-}{2}. \tag{314}$$

The support function is $h_{\mathrm{conv}\{-g,g\}}(\theta) = |\langle g, \theta\rangle|$. If $q \in [0, 1]$ is the probability of selecting the $+$ policy, then

$$\begin{aligned} L_g(\theta, q) &:= |\langle g, \theta\rangle| - (2q - 1)\, \langle g, \theta\rangle \\ &= \begin{cases} 2\, \langle g, \theta\rangle\, (1 - q), & \langle g, \theta\rangle \ge 0, \\ 2\, |\langle g, \theta\rangle|\, q, & \langle g, \theta\rangle < 0. \end{cases} \end{aligned} \tag{315}$$

Thus regret is twice the magnitude of the local policy gap when the wrong policy is selected and zero when the correct policy is selected.

Definition H.3 (Restricted decision width)**.** For a nonempty closed centrally symmetric cone $\mathcal{K} \subseteq \mathbb{R}^d$, define

$$\omega_{\mathcal{K}}(g) := \sup_{\substack{\theta \in \mathcal{K} \\ \|\theta\|_2 \le 1}} |\langle g, \theta \rangle| . \tag{316}$$

This is a deterministic support functional. It should not be confused with the Gaussian width used in empirical-process theory.

Because $\mathcal{K} \cap \mathbb{B}_2$ is compact, the supremum in equation 316 is attained. If $\omega_{\mathcal{K}}(g) > 0$, the conic property implies that a maximizer can be chosen with unit norm.

Theorem H.4 (Binary Gaussian candidate-policy risk versus restricted decision width)**.** *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a nonempty closed centrally symmetric cone and let $R > 0$. For the structured transition model, central symmetry is supplied by Corollary C.6. In the binary Gaussian game with gradients $\{-g, g\}$,*

$$\boxed{c_R\, \omega_{\mathcal{K}}(g) \le \mathfrak{R}_R\big(\mathcal{K}, \operatorname{conv}\{-g, g\}\big) \le R\, \omega_{\mathcal{K}}(g),} \tag{317}$$

where Φ is the standard normal distribution function and

$$c_R := \sup_{0 \le t \le R} 2t\Phi(-t). \tag{318}$$

In particular, $c_R > 0$ for every $R > 0$.

Proof. Write $\omega := \omega_{\mathcal{K}}(g)$. If $\omega = 0$, then $\langle g, \theta \rangle = 0$ for every $\theta \in \mathcal{K}$, so the loss vanishes identically and both bounds hold. Assume henceforth that $\omega > 0$. By compactness and central symmetry, there exists $\theta^\star \in \mathcal{K}$ such that

$$\|\theta^\star\|_2 = 1, \qquad \langle g, \theta^\star \rangle = \omega. \tag{319}$$

Fix $t \in [0, R]$ and restrict the parameter set to $\{t\theta^\star, -t\theta^\star\}$. Let $q(z) \in [0, 1]$ denote the probability that an arbitrary randomized rule selects the $+$ policy after observing z. Under $P_+ := \mathcal{N}(t\theta^\star, I_d)$ and $P_- := \mathcal{N}(-t\theta^\star, I_d)$, its risks at the two points are

$$\mathcal{R}_+(q) = 2t\omega\, \mathbb{E}_{P_+}[1 - q(Z)], \tag{320}$$

$$\mathcal{R}_-(q) = 2t\omega\, \mathbb{E}_{P_-}[q(Z)]. \tag{321}$$

Therefore

$$\max\{\mathcal{R}_+(q), \mathcal{R}_-(q)\} \ge \frac{\mathcal{R}_+(q) + \mathcal{R}_-(q)}{2} \tag{322}$$

$$= t\omega \left(\mathbb{E}_{P_+}[1 - q(Z)] + \mathbb{E}_{P_-}[q(Z)]\right). \tag{323}$$

The parenthesized quantity is twice the equal-prior testing error of the randomized test q for P_+ versus P_-. The log-likelihood ratio is

$$\log \frac{dP_+}{dP_-}(z) = 2t \langle \theta^\star, z \rangle .$$

Hence the likelihood-ratio test selects P_+ exactly when $\langle \theta^\star, z \rangle \ge 0$. Under P_+, $\langle \theta^\star, Z \rangle \sim \mathcal{N}(t, 1)$, and under P_- it is $\mathcal{N}(-t, 1)$. The minimum equal-prior testing error is therefore $\Phi(-t)$, so every q satisfies

$$\frac{1}{2}\left(\mathbb{E}_{P_+}[1 - q(Z)] + \mathbb{E}_{P_-}[q(Z)]\right) \ge \Phi(-t).$$

Substituting into equation 323 yields

$$\sup_{\theta \in \mathcal{K}_R} \mathbb{E}_\theta L_g(\theta, q(Z)) \ge 2t\omega\Phi(-t).$$

Taking the infimum over q and then the supremum over $t \in [0, R]$ proves the lower bound.

For the upper bound, use the data-independent rule that selects each policy with probability $1/2$. Its centered gradient is zero, so its loss at θ is $|\langle g, \theta \rangle|$. If $\theta \in \mathcal{K}_R$ and $\theta \ne 0$, then $\theta / \|\theta\|_2 \in \mathcal{K} \cap \mathbb{B}_2$, and hence

$$|\langle g, \theta \rangle| \le \|\theta\|_2\, \omega \le R\omega.$$

The same conclusion is trivial at $\theta = 0$. This proves the upper bound. □

Corollary H.5 (Sharpness on a one-dimensional cone)**.** *Let $u \in \mathbb{R}^d$ satisfy $\|u\|_2 = 1$ and let $\mathcal{K} = \operatorname{span}\{u\}$. Then*

$$\boxed{\mathfrak{R}_R\big(\mathcal{K}, \operatorname{conv}\{-g, g\}\big) = c_R \, |\langle g, u\rangle| \, .} \tag{324}$$

Thus the lower constant c_R in Theorem H.4 is sharp over the class of closed centrally symmetric cones.

Proof. The lower bound follows from Theorem H.4. Replace u by $-u$ if necessary so that $\langle g, u\rangle \geq 0$. Consider the rule that selects the $+$ policy when $\langle u, Z\rangle \geq 0$ and the $-$ policy otherwise. At $\theta = tu$ with $t \in [-R, R]$, the rule selects the wrong policy with probability $\Phi(-|t|)$, and the regret conditional on an error is $2\,|t|\,\langle g, u\rangle$. Hence its worst-case risk equals

$$\sup_{0 \leq t \leq R} 2t\Phi(-t)\,\langle g, u\rangle = c_R\,|\langle g, u\rangle| \, .$$

□

H.4 Rank-constrained matrix support functions

For a matrix $G \in \mathbb{R}^{m\times n}$ with singular values $\sigma_1(G) \geq \sigma_2(G) \geq \cdots \geq 0$, define the truncated Schatten–2 functional

$$\|G\|_{(r),2} := \left(\sum_{j=1}^{\min\{r,m,n\}} \sigma_j(G)^2\right)^{1/2}, \qquad r \in \mathbb{N}_0. \tag{325}$$

We use the Frobenius pairing $\langle G, \Delta\rangle_{\mathrm{F}} := \operatorname{tr}(G^\top \Delta)$.

Lemma H.6 (Support function of a rank-constrained Frobenius ball)**.** *For every $G \in \mathbb{R}^{m\times n}$, $r \in \mathbb{N}_0$, and $t \geq 0$,*

$$\boxed{\sup_{\substack{\operatorname{rank}(\Delta)\leq r \\ \|\Delta\|_{\mathrm{F}} \leq t}} \langle G, \Delta\rangle_{\mathrm{F}} = t\,\|G\|_{(r),2}.} \tag{326}$$

The same value is obtained if the objective is replaced by $|\langle G, \Delta\rangle_{\mathrm{F}}|$.

Proof. Let $k := \min\{r, m, n\}$. For every feasible Δ, von Neumann's trace inequality gives

$$\langle G, \Delta\rangle_{\mathrm{F}} \leq \sum_{j=1}^{\min\{m,n\}} \sigma_j(G)\sigma_j(\Delta). \tag{327}$$

Because $\operatorname{rank}(\Delta) \leq r$, all singular values of Δ after the first k vanish. Cauchy–Schwarz therefore yields

$$\langle G, \Delta\rangle_{\mathrm{F}} \leq \left(\sum_{j=1}^{k} \sigma_j(G)^2\right)^{1/2} \left(\sum_{j=1}^{k} \sigma_j(\Delta)^2\right)^{1/2} \leq t\,\|G\|_{(r),2}.$$

To prove attainability, write a singular value decomposition $G = U\operatorname{diag}(\sigma_1, \ldots, \sigma_p)V^\top$, where $p = \min\{m, n\}$. If $\|G\|_{(r),2} = 0$, choose $\Delta = 0$. Otherwise set

$$\Delta^\star := \frac{t}{\|G\|_{(r),2}} U_{[:,1:k]} \operatorname{diag}(\sigma_1, \ldots, \sigma_k) V_{[:,1:k]}^\top. \tag{328}$$

Then $\operatorname{rank}(\Delta^\star) \leq r$, $\|\Delta^\star\|_{\mathrm{F}} = t$, and $\langle G, \Delta^\star\rangle_{\mathrm{F}} = t\|G\|_{(r),2}$. Finally, the feasible set is centrally symmetric, so the supremum of the absolute objective equals the supremum of the signed objective. □

Remark H.7. Lemma H.6 is a support-function identity, rather than a low-rank approximation statement. Its proof is nevertheless the dual form of the same singular-value truncation geometry that underlies the Eckart–Young–Mirsky theorem; see Eckart & Young (1936); Mirsky (1960). The trace inequality used in equation 327 originates with von Neumann (1937); a later concise treatment is given by Mirsky (1975).

H.5 Exact width of a block-sparse, block-low-rank cone

Let $\mathfrak{R}$ be a finite index set. For each $\rho \in \mathfrak{R}$, fix matrix dimensions m_ρ, n_ρ, a rank budget $0 \le r_\rho \le \min\{m_\rho, n_\rho\}$, and a weight $\alpha_\rho > 0$. Let $d_0 \ge 0$ and let $I_0 \in \mathbb{R}^{d_0 \times d_0}$ be symmetric positive definite when $d_0 > 0$. We use the convention that the $d_0 = 0$ block is absent.

A block vector is

$$\mathbf{\Delta} = (\Delta_0, (\Delta_\rho)_{\rho \in \mathfrak{R}}) \in \mathbb{R}^{d_0} \oplus \bigoplus_{\rho \in \mathfrak{R}} \mathbb{R}^{m_\rho \times n_\rho}.$$

Equip this space with the block-orthogonal norm

$$\|\mathbf{\Delta}\|_{\mathcal{I}}^2 := \Delta_0^\top I_0 \Delta_0 + \sum_{\rho \in \mathfrak{R}} \alpha_\rho \|\Delta_\rho\|_{\mathrm{F}}^2 . \tag{329}$$

For $0 \le s \le |\mathfrak{R}|$, define the closed centrally symmetric cone

$$\mathcal{C}_{s,\boldsymbol{r}} := \Big\{ \mathbf{\Delta} : |\{\rho : \Delta_\rho \neq 0\}| \le s, \quad \operatorname{rank}(\Delta_\rho) \le r_\rho \text{ for every } \rho \in \mathfrak{R} \Big\}. \tag{330}$$

The nuisance coordinate Δ_0 is unrestricted.

Let

$$\mathbf{\Gamma} = (g_{\mathbf{1}}, (G_\rho)_{\rho \in \mathfrak{R}})$$

define the linear functional

$$\Lambda_{\mathbf{\Gamma}}(\mathbf{\Delta}) := g_{\mathbf{1}}^\top \Delta_0 + \sum_{\rho \in \mathfrak{R}} \langle G_\rho, \Delta_\rho \rangle_{\mathrm{F}} . \tag{331}$$

Define the dual nuisance norm and nontrivial block scores by

$$b_0^2 := g_{\mathbf{1}}^\top I_0^{-1} g_{\mathbf{1}}, \qquad b_\rho^2 := \alpha_\rho^{-1} \|G_\rho\|_{(r_\rho),2}^2. \tag{332}$$

If $d_0 = 0$, set $b_0 = 0$. Let $\mathrm{Top}_{\le s}(b)$ denote any subset containing the at most s largest strictly positive block scores. When fewer than s scores are positive, all positive-score blocks are selected; when $s = 0$, the set is empty. Ties among equal positive scores may be broken arbitrarily, and adding zero-score blocks does not change any displayed value.

Theorem H.8 (Exact top-s/top-r support formula)**.** *For the cone equation 330 and norm equation 329,*

$$\boxed{\sup_{\substack{\mathbf{\Delta} \in \mathcal{C}_{s,\boldsymbol{r}} \\ \|\mathbf{\Delta}\|_{\mathcal{I}} \le 1}} |\Lambda_{\mathbf{\Gamma}}(\mathbf{\Delta})| = \left(b_0^2 + \sum_{\rho \in \mathrm{Top}_{\le s}(b)} b_\rho^2 \right)^{1/2}.} \tag{333}$$

Moreover, the supremum is attained.

Proof. The feasible set is centrally symmetric, so it suffices to maximize the signed functional $\Lambda_{\mathbf{\Gamma}}$. Fix a support set $J \subseteq \mathfrak{R}$ with $|J| \le s$ and restrict to block vectors with $\Delta_\rho = 0$ for $\rho \notin J$. Define

$$t_0 := \left(\Delta_0^\top I_0 \Delta_0\right)^{1/2}, \qquad t_\rho := \sqrt{\alpha_\rho}\, \|\Delta_\rho\|_{\mathrm{F}}, \quad \rho \in J.$$

Then

$$t_0^2 + \sum_{\rho \in J} t_\rho^2 \le 1. \tag{334}$$

By the dual Cauchy–Schwarz inequality in the nuisance block and by Lemma H.6 in every matrix block,

$$\Lambda_{\mathbf{\Gamma}}(\mathbf{\Delta}) \le b_0 t_0 + \sum_{\rho \in J} \alpha_\rho^{-1/2} \|G_\rho\|_{(r_\rho),2} t_\rho \tag{335}$$

$$= b_0 t_0 + \sum_{\rho \in J} b_\rho t_\rho \tag{336}$$

$$\le \left(b_0^2 + \sum_{\rho \in J} b_\rho^2 \right)^{1/2} \left(t_0^2 + \sum_{\rho \in J} t_\rho^2 \right)^{1/2} \tag{337}$$

$$\le \left(b_0^2 + \sum_{\rho \in J} b_\rho^2 \right)^{1/2}. \tag{338}$$

We now verify equality for each fixed J. Let

$$B_J := \left(b_0^2 + \sum_{\rho \in J} b_\rho^2 \right)^{1/2}.$$

If $B_J = 0$, the claimed value is zero. Otherwise choose

$$t_0 = \frac{b_0}{B_J}, \qquad t_\rho = \frac{b_\rho}{B_J}, \quad \rho \in J.$$

When $b_0 > 0$, set

$$\Delta_0^\star := \frac{t_0}{b_0} I_0^{-1} g_{\mathbf{1}}; \tag{339}$$

when $b_0 = 0$, set $\Delta_0^\star = 0$. For every $\rho \in J$ with $b_\rho > 0$, choose a rank-r_ρ support-function attainer $H_\rho^\star$ from Lemma H.6 with $\left\|H_\rho^\star\right\|_{\mathrm{F}} = 1$ and $\left\langle G_\rho, H_\rho^\star \right\rangle_{\mathrm{F}} = \|G_\rho\|_{(r_\rho),2}$, and set

$$\Delta_\rho^\star := \frac{t_\rho}{\sqrt{\alpha_\rho}} H_\rho^\star. \tag{340}$$

Set $\Delta_\rho^\star = 0$ when $b_\rho = 0$ or $\rho \notin J$. This block vector is feasible, saturates equation 334, and attains B_J in equation 338. Hence the optimal value for support J is exactly B_J.

It remains to maximize over J with $|J| \le s$. Because all terms are nonnegative, the maximum of $\sum_{\rho \in J} b_\rho^2$ is obtained by choosing at most s largest positive block scores; adding zero-score blocks cannot change the objective. This proves equation 333 and also constructs an attainer. □

Corollary H.9 (Whitening pullback)**.** *Let* W *be a linear isometry from the block space equipped with* $\|\cdot\|_{\mathcal{I}}$ *into a Euclidean space, let* $\mathcal{K} := \mathsf{W}\mathcal{C}_{s,\boldsymbol{r}}$*, and let* g *be the Euclidean Riesz vector of* $\Lambda_{\mathbf{\Gamma}}$*, meaning*

$$\langle g, \mathsf{W}\mathbf{\Delta} \rangle = \Lambda_{\mathbf{\Gamma}}(\mathbf{\Delta}) \quad \textit{for every } \mathbf{\Delta}. \tag{341}$$

Then

$$\omega_{\mathcal{K}}(g)^2 = b_0^2 + \sum_{\rho \in \mathrm{Top}_{\le s}(b)} b_\rho^2. \tag{342}$$

Proof. Because W is an isometry,

$$\omega_{\mathcal{K}}(g) = \sup_{\substack{\mathbf{\Delta} \in \mathcal{C}_{s,\boldsymbol{r}} \\ \|\mathbf{\Delta}\|_{\mathcal{I}} \le 1}} |\langle g, \mathsf{W}\mathbf{\Delta} \rangle| = \sup_{\substack{\mathbf{\Delta} \in \mathcal{C}_{s,\boldsymbol{r}} \\ \|\mathbf{\Delta}\|_{\mathcal{I}} \le 1}} |\Lambda_{\mathbf{\Gamma}}(\mathbf{\Delta})|.$$

Apply Theorem H.8. □

H.6 Rank-preserving anisotropic Fisher blocks

The scalar block weights in equation 329 are sufficient for the explicit separation constructed later, but the same calculation extends to any Fisher metric that acts by invertible left and right multiplication. This is a broad natural class for which whitening preserves matrix rank automatically.

Proposition H.10 (Kronecker-separable block metrics)**.** *For each $\rho \in \mathfrak{R}$, let $L_\rho^{\mathrm{F}} \in \mathbb{R}^{m_\rho \times m_\rho}$ and $R_\rho^{\mathrm{F}} \in \mathbb{R}^{n_\rho \times n_\rho}$ be symmetric positive definite, and replace the matrix-block contribution to the norm by*

$$\|\Delta_\rho\|^2_{L_\rho^{\mathrm{F}}, R_\rho^{\mathrm{F}}} := \operatorname{tr}\left(\Delta_\rho^\top L_\rho^{\mathrm{F}} \Delta_\rho R_\rho^{\mathrm{F}}\right) = \left\|(L_\rho^{\mathrm{F}})^{1/2} \Delta_\rho (R_\rho^{\mathrm{F}})^{1/2}\right\|_{\mathrm{F}}^2. \tag{343}$$

Then the block score in Theorem H.8 is replaced by

$$b_\rho^2 = \sum_{j=1}^{r_\rho} \sigma_j \left((L_\rho^{\mathrm{F}})^{-1/2} G_\rho (R_\rho^{\mathrm{F}})^{-1/2}\right)^2, \tag{344}$$

and the exact top-s/top-r formula remains valid.

Proof. Set

$$X_\rho := (L_\rho^{\mathrm{F}})^{1/2} \Delta_\rho (R_\rho^{\mathrm{F}})^{1/2}.$$

Since both factors are invertible, $\operatorname{rank}(X_\rho) = \operatorname{rank}(\Delta_\rho)$ and $\|X_\rho\|_{\mathrm{F}} = \|\Delta_\rho\|_{L_\rho^{\mathrm{F}}, R_\rho^{\mathrm{F}}}$. Moreover,

$$\begin{aligned}\langle G_\rho, \Delta_\rho \rangle_{\mathrm{F}} &= \operatorname{tr}\left(G_\rho^\top (L_\rho^{\mathrm{F}})^{-1/2} X_\rho (R_\rho^{\mathrm{F}})^{-1/2}\right) \\ &= \left\langle (L_\rho^{\mathrm{F}})^{-1/2} G_\rho (R_\rho^{\mathrm{F}})^{-1/2}, X_\rho \right\rangle_{\mathrm{F}}.\end{aligned}$$

Lemma H.6 applied to X_ρ gives equation 344; the remainder of the proof of Theorem H.8 is unchanged. □

Remark H.11 (Why arbitrary Fisher whitening is excluded). A general positive-definite operator on the vectorized matrix space need not factor as $\Delta \mapsto A^{1/2} \Delta B^{1/2}$ and need not preserve matrix rank. For example, on $\mathbb{R}^{2\times 2}$, the invertible linear map defined by $T(E_{11}) = I_2$ and $T(E_{ij}) = E_{ij}$ for $(i,j) \neq (1,1)$ maps the rank-one matrix E_{11} to the rank-two matrix I_2. The norm $\|\Delta\|_T := \|T\Delta\|_{\mathrm{F}}$ is induced by the positive-definite operator T^*T, yet its whitening map destroys rank. Consequently, the image of a rank-constrained cone under arbitrary Fisher whitening need not itself admit a rank constraint in the whitened coordinates. The exact singular-value formula above is therefore asserted only for scalar block weights or, more generally, for the separable metrics in Proposition H.10.

H.7 Abstract separations induced by rank and block diffusion

The following corollaries demonstrate that equal ambient dimensions, equal Frobenius norms, and equal structural budgets do not determine the binary Gaussian candidate-policy risk. They are statements about the limiting Gaussian game only. The realizability appendix separately constructs finite MDPs realizing the single-block separation.

Corollary H.12 (Single-block rank-diffusion separation)**.** *Fix integers $d \geq r \geq 1$ and let*

$$\mathcal{K}_{d,r} := \{\Theta \in \mathbb{R}^{d\times d} : \operatorname{rank}(\Theta) \leq r\}$$

with the Frobenius norm. Define

$$G_{d,r}^{\mathrm{conc}} := r^{-1/2} \operatorname{diag}(I_r, 0_{d-r}), \qquad G_d^{\mathrm{flat}} := d^{-1/2} I_d. \tag{345}$$

Both matrices have Frobenius norm one, but

$$\omega_{\mathcal{K}_{d,r}}(G_{d,r}^{\mathrm{conc}}) = 1, \qquad \omega_{\mathcal{K}_{d,r}}(G_d^{\mathrm{flat}}) = \sqrt{\frac{r}{d}}. \tag{346}$$

Consequently, for every $R > 0$,

$$\boxed{\frac{c_R}{R}\sqrt{\frac{d}{r}} \leq \frac{\mathfrak{R}_R\big(\mathcal{K}_{d,r}, \operatorname{conv}\{-G_{d,r}^{\mathrm{conc}}, G_{d,r}^{\mathrm{conc}}\}\big)}{\mathfrak{R}_R\big(\mathcal{K}_{d,r}, \operatorname{conv}\{-G_d^{\mathrm{flat}}, G_d^{\mathrm{flat}}\}\big)} \leq \frac{R}{c_R}\sqrt{\frac{d}{r}}.} \tag{347}$$

Proof. The singular values of $G_{d,r}^{\mathrm{conc}}$ are $r^{-1/2}$ with multiplicity r and zero thereafter, whereas all singular values of G_d^{flat} equal $d^{-1/2}$. Hence

$$\|G_{d,r}^{\mathrm{conc}}\|_{(r),2} = 1, \qquad \|G_d^{\mathrm{flat}}\|_{(r),2} = \sqrt{r/d}.$$

The width identities follow from Lemma H.6. Theorem H.4 gives

$$c_R \le \mathfrak{R}_R^{\mathrm{conc}} \le R, \qquad c_R\sqrt{r/d} \le \mathfrak{R}_R^{\mathrm{flat}} \le R\sqrt{r/d}.$$

Since $R > 0$ implies $c_R > 0$, division gives both sides of equation 347. □

Corollary H.13 (Joint block-sparsity and rank-diffusion separation)**.** *Fix $M \ge s \ge 1$ and $d \ge r \ge 1$. Consider M matrix blocks of size $d \times d$, unit block weights, support budget s, rank budget r in every block, and no nuisance block. Let $\mathbf{\Gamma}^{\mathrm{conc}}$ have the single nonzero block $r^{-1/2}\,\mathrm{diag}(I_r, 0_{d-r})$, and let $\mathbf{\Gamma}^{\mathrm{diff}}$ have every block equal to $(Md)^{-1/2}I_d$. Then*

$$\|\mathbf{\Gamma}^{\mathrm{conc}}\|_{\mathrm{F}} = \|\mathbf{\Gamma}^{\mathrm{diff}}\|_{\mathrm{F}} = 1, \tag{348}$$

while their restricted widths are

$$\omega_{\mathrm{conc}} = 1, \qquad \omega_{\mathrm{diff}} = \sqrt{\frac{sr}{Md}}. \tag{349}$$

Therefore

$$\boxed{\frac{\mathfrak{R}_R^{\mathrm{conc}}}{\mathfrak{R}_R^{\mathrm{diff}}} \ge \frac{c_R}{R}\sqrt{\frac{Md}{sr}}.} \tag{350}$$

Proof. The Frobenius-norm identity is immediate. For the concentrated functional, one block has score one and the rest have score zero, so Theorem H.8 gives width one. For the diffuse functional, every block has

$$b_\rho^2 = \sum_{j=1}^{r} \frac{1}{Md} = \frac{r}{Md}.$$

Selecting the s largest identical scores gives total squared width $sr/(Md)$. The risk ratio follows from Theorem H.4 exactly as in Corollary H.12. □

H.8 Summary of what is established at the Gaussian level

Theorem H.4 reduces binary local candidate-policy selection to a deterministic support functional of the feasible local parameter cone. Theorem H.8 computes that functional exactly for the block-sparse, block-low-rank geometry produced by a maximally singular structured transition model. Corollaries H.12 and H.13 show that model dimensions and norm budgets do not determine Gaussian decision difficulty: the concentration profile of the candidate-policy comparison functional relative to the feasible cone is decisive.

This appendix does not by itself establish an MDP theorem. The local Gaussian experiment is derived from the transition-sampling model in the local-experiment appendix, and the explicit finite-MDP realization of the single-block separation is proved in the realizability appendix.

I Realizable rank-diffusion separations

This appendix realizes the rank-diffusion phenomenon of Corollary H.12 by an explicit family of finite-horizon controlled processes. The transition model is exactly C_2-equivariant in the joint state–action sense of van der Pol et al. (2020); the two tasks deliberately use symmetry-breaking rewards in order to activate the nontrivial sign sector. Every transition, reward, policy, occupancy, continuation value, Fisher weight, and policy-gradient block is given in closed form. The only representation theory used below is the elementary decomposition of the permutation representation of C_2 into trivial and sign components; see, e.g., Serre (1977).

I.1 A BOUNDED-REWARD C2 CONSTRUCTION

Fix integers $d \geq r \geq 1$ and let

$$G = C_2 = \{e, \tau\}.$$

The state and action spaces are

$$\mathcal{S}_d := [d] \times \{-1, +1\}, \qquad \mathcal{A} := \{0, 1\}, \qquad \mathcal{X}_d := \mathcal{S}_d \times \mathcal{A}. \tag{351}$$

The nontrivial group element flips the sign coordinate and fixes the action:

$$\tau(i, \sigma) = (i, -\sigma), \qquad \tau(i, \sigma, a) = (i, -\sigma, a). \tag{352}$$

Thus $|\mathcal{S}_d| = 2d$ and $|\mathcal{X}_d| = 4d$. Let

$$P_0(s' \mid x) := \frac{1}{2d} \quad \forall (x, s') \in \mathcal{X}_d \times \mathcal{S}_d, \qquad \mu(s) := \frac{1}{2d}, \qquad w_x := \frac{1}{4d}. \tag{353}$$

The base transition, initial law, and generative design are therefore uniform and C_2-invariant. The horizon is

$$N := d^2, \qquad T := N + 1. \tag{354}$$

Index transition stages by pairs $(j, k) \in [d] \times [d]$ through

$$\vartheta(j, k) := (j - 1)d + k \in [N]. \tag{355}$$

We shall use a bounded coefficient-splitting map. For $z \in [-d, d]$, let

$$m(z) := \lfloor |z| \rfloor, \qquad \epsilon(z) := |z| - m(z), \qquad \operatorname{sgn}(0) := 0, \tag{356}$$

and, for $k \in [d]$, define

$$\psi_k(z) := \operatorname{sgn}(z) \left[\mathbf{1}\{k \leq m(z)\} + \epsilon(z) \mathbf{1}\{k = m(z) + 1\} \right], \tag{357}$$

where the second indicator is understood to vanish when $m(z) = d$. Then

$$|\psi_k(z)| \leq 1, \qquad \sum_{k=1}^{d} \psi_k(z) = z. \tag{358}$$

For a target diagonal vector

$$\lambda = (\lambda_1, \ldots, \lambda_d) \in [-1, 1]^d, \tag{359}$$

define

$$a^{\lambda}_{j,k} := \psi_k(d\lambda_j), \qquad |a^{\lambda}_{j,k}| \leq 1, \qquad \sum_{k=1}^{d} a^{\lambda}_{j,k} = d\lambda_j. \tag{360}$$

The candidate policy class contains exactly two deterministic nonstationary Markov policies,

$$\Pi_d := \{\pi^+, \pi^-\}, \tag{361}$$

with, at stage $t = \vartheta(j, k)$,

$$\begin{aligned} \pi^+_t(i, \sigma) &:= \mathbf{1}\{i = j,\ \sigma = +1\}, \\ \pi^-_t(i, \sigma) &:= \mathbf{1}\{i = j,\ \sigma = -1\}. \end{aligned} \tag{362}$$

Hence both policies choose action 0 except that, at stage $\vartheta(j, k)$, π^+ chooses action 1 at $(j, +1)$ whereas π^- chooses action 1 at $(j, -1)$.

The task indexed by λ has rewards bounded by one. Set

$$r^{\lambda}_1(x) := 1 \quad \forall x \in \mathcal{X}_d. \tag{363}$$

For $t = \vartheta(j, k) \in [N]$, define the mean-zero state function

$$\varrho^{\lambda}_t(i, \sigma) := a^{\lambda}_{j,k}\, \sigma\, \mathbf{1}\{i = j\}. \tag{364}$$

For $t < N$, use it as the next stage reward,

$$r^{\lambda}_{t+1}(i, \sigma, a) := \varrho^{\lambda}_t(i, \sigma), \tag{365}$$

and use the final signal as terminal reward,

$$r_T^\lambda(i,\sigma) := \varrho_N^\lambda(i,\sigma). \tag{366}$$

Equations equation 363–equation 366 are the complete reward table. They imply

$$\max_{t<T} \|r_t^\lambda\|_\infty \vee \|r_T^\lambda\|_\infty = 1. \tag{367}$$

The common constant reward at stage one is included only to make the reward bound exactly one for every task; it has no transition derivative.

Remark I.1 (What is and is not symmetric). The transition experiment equation 353 is exactly C_2-equivariant, but the signals equation 364 transform according to the sign representation and hence break task symmetry. This is necessary: an invariant task would not expose the sign block in first order. Accordingly, this appendix constructs transition-equivariant controlled processes with symmetry-breaking tasks; it does not claim that the complete reward-transition MDP is C_2-invariant.

I.2 Isotypic coordinates and the structured local cone

Let $e_{i,\sigma} \in \mathbb{R}^{\mathcal{S}_d}$ and $e_{i,\sigma,a} \in \mathbb{R}^{\mathcal{X}_d}$ denote the standard basis vectors. Define state-space trivial and sign vectors

$$v_j^{\mathrm{triv}} := \frac{e_{j,+1} + e_{j,-1}}{\sqrt{2}}, \qquad v_j^{\mathrm{sgn}} := \frac{e_{j,+1} - e_{j,-1}}{\sqrt{2}}, \qquad j \in [d], \tag{368}$$

and, for each action $a \in \{0,1\}$, define

$$u_{j,a}^{\mathrm{triv}} := \frac{e_{j,+1,a} + e_{j,-1,a}}{\sqrt{2}}, \qquad u_{j,a}^{\mathrm{sgn}} := \frac{e_{j,+1,a} - e_{j,-1,a}}{\sqrt{2}}. \tag{369}$$

The sign multiplicities are therefore

$$m_{\mathrm{sgn}}^{\mathcal{S}_d} = d, \qquad m_{\mathrm{sgn}}^{\mathcal{X}_d} = 2d. \tag{370}$$

Rotate the sign multiplicity basis on $\mathcal{X}_d$ by setting

$$p_j := \frac{u_{j,1}^{\mathrm{sgn}} - u_{j,0}^{\mathrm{sgn}}}{\sqrt{2}}, \qquad q_j := \frac{u_{j,1}^{\mathrm{sgn}} + u_{j,0}^{\mathrm{sgn}}}{\sqrt{2}}. \tag{371}$$

The vectors in equation 371 form an orthonormal basis of the sign-multiplicity space in $\mathbb{R}^{\mathcal{X}_d}$; the vectors $(v_j^{\mathrm{sgn}})_{j=1}^d$ form the corresponding orthonormal basis in $\mathbb{R}^{\mathcal{S}_d}$.

The uniform transition P_0 has zero sign block. Define the common structured transition class

$$\Theta_{d,r} := \left\{ P : \ P \text{ is stochastic and } C_2\text{-equivariant}, \quad \mathrm{rank}\big(B_{\mathrm{sgn}}(P)\big) \le r \right\}. \tag{372}$$

There is only one nontrivial real irreducible representation, so this is the support budget $s = 1$ specialization of Section 2. The trivial isotypic nuisance block remains unrestricted subject to stochastic normalization.

Proposition I.2 (Exact product tangent cone and block-orthogonal Fisher geometry)**.** *At the uniform base kernel P_0, the full contingent tangent cone is*

$$\boxed{\mathcal{C}_0 = \mathcal{B}_G\big(\mathcal{N}_{\mathbf{1}} \oplus \left\{\Delta \in \mathbb{R}^{2d\times d} : \mathrm{rank}(\Delta) \le r\right\}\big).} \tag{373}$$

In particular, its sign-block projection is

$$\boxed{\left\{\Delta \in \mathbb{R}^{2d\times d} : \mathrm{rank}(\Delta) \le r\right\}.} \tag{374}$$

Under the balanced design equation 353, the trivial and sign blocks are Fisher orthogonal and

$$\|\mathcal{B}_G(\Delta_{\mathbf{1}}, \Delta)\|_{\mathcal{I}_{0,w}}^2 = \frac{1}{2}\|\Delta_{\mathbf{1}}\|_{\mathrm{F}}^2 + \frac{1}{2}\|\Delta\|_{\mathrm{F}}^2. \tag{375}$$

Let $E_{\mathrm{nui},d}$ be the Euclidean image of the normalized trivial nuisance space under the block whitening map. Then the full whitened cone is the orthogonal product

$$\boxed{\mathcal{K}_0 = E_{\mathrm{nui},d} \oplus \mathcal{K}_{d,r}^{\mathrm{sgn}}, \qquad \mathcal{K}_{d,r}^{\mathrm{sgn}} := \left\{\Delta \in \mathbb{R}^{2d\times d} : \mathrm{rank}(\Delta) \le r\right\}.} \tag{376}$$

Consequently,

$$\mathcal{K}_0 \cap R\mathbb{B}_2 = \left\{(\eta, \Delta) : \eta \in E_{\mathrm{nui},d},\ \mathrm{rank}(\Delta) \le r,\ \|\eta\|_2^2 + \|\Delta\|_{\mathrm{F}}^2 \le R^2\right\}. \tag{377}$$

Proof. There is one nontrivial real irreducible representation, its base block is zero, and its rank budget is r. The exact structured tangent formula in Theorem C.5 therefore reduces to equation 373: the normalized trivial block is an unrestricted linear nuisance space, while the sign block is the rank-$\leq r$ cone. Equivalently, every rank-$\leq r$ sign matrix generates the exact block path $t\Delta$; strict positivity of the uniform base kernel makes its reconstruction feasible for all sufficiently small $|t|$, and closedness of the rank cone gives the converse.

Because $|\mathcal{S}_d| = 2d$, $|\mathcal{X}_d| = 4d$, and both irreducible representations of C_2 have dimension one, Proposition B.15 gives the common coefficient $|\mathcal{S}_d|/|\mathcal{X}_d| = 1/2$ and no cross term between the trivial and sign blocks. The block whitening map therefore scales each block by $2^{-1/2}$, which preserves the nuisance subspace and the conic rank constraint. This proves equation 375–equation 377. □

I.3 Exact values and occupancies

Lemma I.3 (Base values and tied optimality)**.** *For every $\lambda \in [-1,1]^d$ and both policies $\pi \in \Pi_d$,*

$$V_1^{\pi,\lambda} \equiv 1, \qquad V_{t+1}^{\pi,\lambda} = \varrho_t^\lambda \quad (t \in [N]), \qquad J^\pi(P_0) = 1. \tag{378}$$

In particular,

$$\Pi_0 = \Pi_d = \{\pi^+, \pi^-\}; \tag{379}$$

the two candidate policies are exactly tied at the base model.

Proof. Each signal ϱ_t^λ has zero mean under the uniform state law:

$$\frac{1}{2d}\sum_{i=1}^{d}\sum_{\sigma\in\{-1,+1\}} \varrho_t^\lambda(i,\sigma) = 0.$$

The uniform transition sends every state function to its uniform mean, independently of the current state and action. Backward induction therefore gives

$$V_{t+1}^{\pi,\lambda} = \varrho_t^\lambda + P_0 V_{t+2}^{\pi,\lambda} = \varrho_t^\lambda$$

for $t = N-1,\dots,1$, with the terminal case supplied by equation 366. Finally, $V_1^{\pi,\lambda} = r_1^\lambda + P_0 V_2^{\pi,\lambda} \equiv 1$. The result is independent of the policy because all rewards after time one are state-only. Since Π_d contains exactly two policies, both are base-optimal. □

Lemma I.4 (Exact occupancy difference)**.** *At every stage, the state distribution under either policy is uniform on $\mathcal{S}_d$. At stage $t = \vartheta(j,k)$, the state-action occupancy difference is*

$$\boxed{d_t^{\pi^+}(i,\sigma,a) - d_t^{\pi^-}(i,\sigma,a) = \frac{1}{2d}\,\mathbf{1}\{i=j\}\,\sigma\,(2a-1).} \tag{380}$$

Proof. The initial law is uniform, and every row of P_0 is the same uniform state law; hence the state distribution is uniform at every stage under every policy. At stage $\vartheta(j,k)$ the policies agree everywhere except at $(j,+1)$ and $(j,-1)$. Substituting their deterministic action choices from equation 362 into the uniform state occupancy gives equation 380 by inspection. □

I.4 Exact realization of the comparison-gradient spectral profile

Let

$$H^\lambda := H_{\pi^+} - H_{\pi^-} \in \mathbb{R}^{\mathcal{X}_d\times\mathcal{S}_d} \tag{381}$$

be the raw occupancy–continuation gradient difference from Proposition D.5. Let $\gamma_\lambda^+, \gamma_\lambda^-$ be the Fisher-whitened gradients of the two policies and define their centered difference

$$g^\lambda := \frac{\gamma_\lambda^+ - \gamma_\lambda^-}{2}. \tag{382}$$

Theorem I.5 (Exact diagonal realization)**.** *For every $\lambda \in [-1,1]^d$, the raw gradient difference is*

$$\boxed{H^\lambda = \sqrt{2}\sum_{j=1}^{d} \lambda_j\, p_j \big(v_j^{\mathrm{sgn}}\big)^\top.} \tag{383}$$

It is C_2-equivariant, has zero row sums, and has no trivial-isotypic component. In the sign multiplicity bases $(p_1,\dots,p_d,q_1,\dots,q_d)$ and $(v_1^{\mathrm{sgn}},\dots,v_d^{\mathrm{sgn}})$,

$$B_{\mathrm{sgn}}(H^\lambda) = \sqrt{2}\begin{bmatrix}\mathrm{diag}(\lambda)\\ 0_{d\times d}\end{bmatrix}. \tag{384}$$

Moreover, the centered Fisher-whitened gradient has zero trivial block and sign block

$$\boxed{B_{\mathrm{sgn}}(g^\lambda) = \begin{bmatrix}\mathrm{diag}(\lambda)\\ 0_{d\times d}\end{bmatrix}.} \tag{385}$$

Thus every diagonal matrix with entries in $[-1,1]$ is realized exactly, up to common zero padding, by a bounded-reward transition-equivariant finite MDP.

Proof. By Proposition D.5, Lemmas I.3 and I.4, and the stage indexing $t = \vartheta(j,k)$,

$$\begin{aligned} H^\lambda((i,\sigma,a),(i',\sigma')) &= \sum_{j=1}^{d}\sum_{k=1}^{d} \frac{1}{2d}\mathbf{1}\{i=j\}\sigma(2a-1)\cdot a_{j,k}^{\lambda}\sigma'\mathbf{1}\{i'=j\} \\ &= \frac{1}{2}\sum_{j=1}^{d}\lambda_j \mathbf{1}\{i=j=i'\}\sigma\sigma'(2a-1), \end{aligned}$$

where the second equality uses $\sum_k a_{j,k}^{\lambda} = d\lambda_j$. On the other hand, the basis definitions give

$$p_j(i,\sigma,a) = \frac{1}{2}\mathbf{1}\{i=j\}\sigma(2a-1), \qquad v_j^{\mathrm{sgn}}(i',\sigma') = \frac{1}{\sqrt{2}}\mathbf{1}\{i'=j\}\sigma'.$$

Hence the preceding entrywise expression is exactly equation 383. Each factor changes sign under τ, so their outer product is fixed by simultaneous conjugation and is therefore equivariant. Every v_j^{sgn} has zero coordinate sum, so $H^\lambda \mathbf{1}_{\mathcal{S}} = 0$. Equation equation 384 follows immediately from orthonormality of the chosen sign bases. The factorization contains only sign-to-sign terms, proving that all trivial and cross-isotypic components vanish.

Under equation 353, Proposition B.15 gives

$$\langle U, V\rangle_{\mathcal{I}_{0,w}} = \frac{1}{2}\langle U, V\rangle_{\mathrm{HS}} \qquad (U,V \in \mathcal{H}_G^0).$$

The centered derivative difference is

$$\frac{1}{2}\left(DJ_{P_0}^{\pi^+}[u] - DJ_{P_0}^{\pi^-}[u]\right) = \frac{1}{2}\left\langle H^\lambda, u\right\rangle_{\mathrm{HS}} = \left\langle H^\lambda, u\right\rangle_{\mathcal{I}_{0,w}}.$$

Thus H^λ itself is the centered Fisher–Riesz representative in the original coordinates. Fisher whitening multiplies the sign block by $2^{-1/2}$; applying this scalar to equation 384 proves equation 385. □

Remark I.6 (Why the d^2-stage schedule is used). A single visit to coordinate j contributes at most $1/d$ to the corresponding centered gradient coefficient when rewards are bounded by one, because the uniform occupancy of an individual state is $(2d)^{-1}$. Repeating each coordinate over d stages and splitting $d\lambda_j$ into bounded amplitudes via equation 357 realizes the desired coefficient exactly while keeping every reward in $[-1,1]$. The longer horizon is shared by the two instances and therefore cannot explain their risk ratio.

I.5 Concentrated and flat tasks

Define

$$\lambda_{d,r}^{\mathrm{conc}} := \frac{1}{\sqrt{r}}(\underbrace{1,\dots,1}_{r},0,\dots,0), \qquad \lambda_d^{\mathrm{flat}} := \frac{1}{\sqrt{d}}(1,\dots,1), \tag{386}$$

and let

$$M_{d,r}^{\mathrm{conc}} := M_{d,r}^{\lambda_{d,r}^{\mathrm{conc}}}, \qquad M_{d,r}^{\mathrm{flat}} := M_{d,r}^{\lambda_d^{\mathrm{flat}}} \tag{387}$$

be the corresponding tasks. By Theorem I.5, their centered whitened sign-block gradients are

$$\boxed{G_{d,r}^{\mathrm{conc}} = \begin{bmatrix} r^{-1/2}\,\mathrm{diag}(I_r,0_{d-r}) \\ 0_{d\times d}\end{bmatrix}, \qquad G_d^{\mathrm{flat}} = \begin{bmatrix} d^{-1/2}I_d \\ 0_{d\times d}\end{bmatrix}.} \tag{388}$$

The zero rows are common padding; the nonzero $d\times d$ parts are exactly the matrices in Corollary H.12.

Proposition I.7 (Matched structural summaries)**.** *For fixed (d,r), the two tasks in equation 387 have identical:*

1. *state space, action space, horizon $T = d^2+1$, group action, initial distribution, candidate policy class, and candidate policies;*
2. *strictly positive uniform base transition P_0, whose ordinary matrix rank is one;*
3. *balanced generative design and Fisher information;*
4. *structured transition class $\Theta_{d,r}$, consisting of one nontrivial sign block of size $2d\times d$ with rank budget r and support budget one;*
5. *two base-optimal policies with equal value one and zero pairwise value gap; because $\Pi_d = \Pi_0$, there is no inactive candidate policy;*
6. *reward envelope one;*
7. *centered gradient Frobenius norm one;*
8. *zero centered gradient in the complete trivial-isotypic nuisance block.*

They differ only in the singular-value concentration profile of the centered gradient relative to the same sign-block rank cone.

Proof. Items 1–6 follow directly from equation 351–equation 367 and Lemma I.3. The base transition has identical rows and is nonzero, hence has ordinary matrix rank one. Proposition I.2 gives the common structured and Fisher geometry. From equation 388,

$$\|G_{d,r}^{\mathrm{conc}}\|_{\mathrm{F}}^2 = r\cdot r^{-1} = 1, \qquad \|G_d^{\mathrm{flat}}\|_{\mathrm{F}}^2 = d\cdot d^{-1} = 1.$$

Theorem I.5 proves that both centered gradients have zero trivial block. Their singular values are, respectively, $r^{-1/2}$ with multiplicity r and $d^{-1/2}$ with multiplicity d, which is the only listed distinction. □

I.6 Realizable local-risk separation

For $\diamond\in\{\mathrm{conc},\mathrm{flat}\}$, let $\mathcal{R}_{n,R}^{\diamond}$ denote the local minimax candidate-policy regret of Section 2.2 for the common transition experiment $(\Theta_{d,r},P_0,w)$ and the task $M_{d,r}^{\diamond}$. Let $\mathfrak{R}_R^{\diamond}$ be its limiting binary Gaussian game.

Theorem I.8 (Realizable rank-diffusion separation)**.** *Consider the exactly two-policy candidate class $\Pi_d=\{\pi^+,\pi^-\}$, the strictly positive balanced C_2-equivariant transition experiment in equation 353, and the symmetry-breaking bounded-reward tasks in equation 365–equation 366. Assume the fixed-radius oracle-local candidate-policy minimax theorem of Section 2.2. For every fixed $d\ge r\ge 1$ and $R>0$,*

$$\lim_{n\to\infty}\mathcal{R}_{n,R}^{\diamond} = \mathfrak{R}_R^{\diamond}, \qquad \diamond\in\{\mathrm{conc},\mathrm{flat}\}, \tag{389}$$

and

$$\boxed{\frac{c_R}{R}\sqrt{\frac{d}{r}} \le \frac{\lim_{n\to\infty} \mathcal{R}^{\mathrm{conc}}_{n,R}}{\lim_{n\to\infty} \mathcal{R}^{\mathrm{flat}}_{n,R}} \le \frac{R}{c_R}\sqrt{\frac{d}{r}}.} \tag{390}$$

Thus, after taking the fixed-dimensional local limit for each (d, r), the limiting-risk ratio is $\Theta_R(\sqrt{d/r})$. The limit is severely nonuniform in d because

$$p_{\min} = \frac{1}{2d}, \qquad w_{\min} = \frac{1}{4d}, \qquad T = d^2 + 1.$$

Accordingly, no joint finite-n/growing-dimensional conclusion is claimed.

Proof. The fixed-radius local minimax theorem identifies each MDP risk limit with the Gaussian game generated by the common full whitened tangent cone and the corresponding centered gradient. Proposition I.2 shows that this cone is the orthogonal product of the unrestricted trivial nuisance space and the sign rank cone. Theorem I.5 shows that both centered gradients have zero trivial block. Therefore Lemma H.2 applies and proves that the full Gaussian minimax value equals the value of the sign-block experiment alone.

The concentrated sign gradient has singular values $r^{-1/2}$ with multiplicity r, so its restricted decision width on $\mathcal{K}^{\mathrm{sgn}}_{d,r}$ is one. The flat sign gradient has singular values $d^{-1/2}$ with multiplicity d, so its width is $\sqrt{r/d}$. This follows directly from Lemma H.6; the common zero rows do not alter singular values. The binary risk–width theorem then gives

$$c_R \le \mathfrak{R}^{\mathrm{conc}}_R \le R, \qquad c_R\sqrt{r/d} \le \mathfrak{R}^{\mathrm{flat}}_R \le R\sqrt{r/d}.$$

Since $R > 0$ implies $c_R > 0$, division yields both inequalities in equation 390. □

Remark I.9 (Order of limits). Theorem I.8 is a fixed-dimensional local asymptotic statement. One first fixes (d, r) and lets $n \to \infty$; only then may one compare the resulting limiting risks as $d/r \to \infty$. No growing-dimensional LAN claim is made.

Remark I.10 (Restricted candidate class). The construction studies selection between two explicitly specified candidate policies, exactly as allowed by the finite policy class in Section 2. It does not assert that these are the only optimal policies among the complete deterministic Markov policy class. Extending the gadget to enforce global uniqueness outside the two candidates would require an additional coupling mechanism and is not needed for the local binary separation proved here.

I.7 Summary of the realizability certificate

The construction closes the realizability gap left by the abstract Gaussian example:

1. the transition experiment is finite, strictly positive, uniformly sampled, and exactly C_2-equivariant;
2. the only nontrivial representation sector is a $2d \times d$ sign multiplicity block with rank budget r;
3. rewards are uniformly bounded by one and the horizon and candidate policies are identical across the two tasks;
4. the centered Fisher-whitened policy gradients are exactly the concentrated and flat matrices, up to common zero padding;
5. therefore the rank-diffusion factor $\sqrt{d/r}$ is realized by genuine finite MDP decision problems rather than by an unconstrained Gaussian matrix surrogate.

J Adaptive attainment on stable support–rank–policy strata

This appendix replaces the oracle-local result of Theorem 2.4 by a center-adaptive theorem on uniformly regular strata. The oracle theorem remains the benchmark and the lower-bound engine; the new content is that the center, Fisher metric, support/rank stratum, and tied candidate face can all be

estimated without first-order regret loss when they are separated from neighboring strata by fixed margins. The proof uses a vanishing pilot fraction to recover the discrete geometry, local constrained likelihood projections on the main sample to preserve the root-n policy-comparison signal, and a stability theorem for the Gaussian Bellman game. Constrained likelihood asymptotics on smooth sets are classical (Geyer, 1994; Shapiro, 2000); all uniformity and policy-regret transfers needed here are proved below in the finite categorical experiment.

J.1 A QUANTITATIVE STABLE-STRATUM CLASS

We now define the regular regime on which center adaptation is possible. The class is deliberately narrower than the full structured model: the nontrivial block support saturates the support budget, every active block saturates its rank budget, the base-optimal candidate face is separated from all inactive candidates, and the active value-gap map meets the model stratum transversally. These conditions exclude exactly the support, rank, and policy-face intersections at which Appendix K proves that uniform adaptation can fail.

Throughout this subsection the design vector $w = (w_x)_{x\in\mathcal{X}}$ is fixed, G-invariant, satisfies $\sum_x w_x = 1$, and obeys

$$w_x \geq w_\star > 0 \qquad (x \in \mathcal{X}). \tag{391}$$

The state and action spaces, horizon, group action, candidate class Π, rewards, support budget s, and rank budgets $(r_\rho)_{\rho\in\mathfrak{R}}$ are fixed.

J.1.1 REGULAR SUPPORT–RANK STRATA

Let

$$\mathfrak{R}_+ := \{\rho \in \mathfrak{R} : r_\rho \geq 1\}. \tag{392}$$

An *admissible saturated support pattern* is a set $\Lambda \subseteq \mathfrak{R}_+$ with $|\Lambda| = s$; for $s = 0$ the unique pattern is $\Lambda = \varnothing$. Write

$$\alpha := \big(\Lambda, (r_\rho)_{\rho\in\Lambda}\big). \tag{393}$$

Using the affine block coordinates of Appendix B, define

$$\begin{aligned}\mathcal{M}_\alpha := \Big\{ P \in \mathcal{A}_G : {} & P(s' \mid x) > 0 \ \forall (x, s'), \\ & B_\rho(P) = 0 \quad (\rho \notin \Lambda), \\ & \operatorname{rank} B_\rho(P) = r_\rho \quad (\rho \in \Lambda)\Big\},\end{aligned} \tag{394}$$

where $\mathcal{A}_G$ is the affine space of normalized G-equivariant kernels. The normalized trivial-isotypic nuisance coordinate is unrestricted. The support and rank constraints in equation 394 saturate the budgets of the ambient model $\Theta_{G,s,\boldsymbol{r}}$.

For $P \in \mathcal{M}_\alpha$, choose compact SVDs

$$B_\rho(P) = U_\rho \Sigma_\rho V_\rho^\top, \qquad \rho \in \Lambda, \tag{395}$$

and write $U_{\rho,\perp}, V_{\rho,\perp}$ for orthogonal complements. Set

$$\mathcal{L}_P := T_P \mathcal{M}_\alpha = \mathcal{B}_G \left(\mathcal{N}_{\mathbf{1}} \oplus \prod_{\rho\in\Lambda} \big\{ H_\rho : U_{\rho,\perp}^\top H_\rho V_{\rho,\perp} = 0 \big\} \times \prod_{\rho\notin\Lambda} \{0\} \right). \tag{396}$$

Let $d_{\mathbf{1}} := \dim \mathcal{N}_{\mathbf{1}}$ and

$$d_\alpha := d_{\mathbf{1}} + \sum_{\rho\in\Lambda} r_\rho \big(m_\rho^{\mathcal{X}} + m_\rho^{\mathcal{S}} - r_\rho \big). \tag{397}$$

Proposition J.1 (Smoothness and tangent space of a saturated stratum)**.** *For every admissible α, $\mathcal{M}_\alpha$ is a smooth embedded submanifold of $\mathcal{A}_G$ of dimension d_α, and its tangent space is given by equation 396.*

Proof. The affine reconstruction map $\mathcal{B}_G$ is a linear isomorphism between the normalized block-coordinate space and the translation space of $\mathcal{A}_G$. It is therefore enough to treat one active block. Let $X \in \mathbb{R}^{m\times n}$ have rank r and compact SVD $X = U\Sigma V^\top$. For a matrix Y near X, write

$$[U\ U_\perp]^\top Y [V\ V_\perp] = \begin{bmatrix} S & B \\ C & D \end{bmatrix}.$$

If S is invertible, then Y has rank r if and only if its Schur complement vanishes:

$$D - CS^{-1}B = 0.$$

Consequently, a neighborhood of X in the rank-r set is parametrized smoothly by

$$(A, B, C) \longmapsto [U\ U_\perp] \begin{bmatrix} \Sigma + A & B \\ C & C(\Sigma + A)^{-1}B \end{bmatrix} [V\ V_\perp]^\top, \tag{398}$$

for $\Sigma + A$ invertible. The derivative at the origin is

$$(A, B, C) \longmapsto [U\ U_\perp] \begin{bmatrix} A & B \\ C & 0 \end{bmatrix} [V\ V_\perp]^\top.$$

Hence the tangent space is $\{H : U_\perp^\top H V_\perp = 0\}$ and the dimension is $r^2 + r(n-r) + (m-r)r = r(m+n-r)$. Taking the finite product over active blocks, fixing inactive blocks at zero, adjoining the affine nuisance factor, and intersecting with the open positivity set proves the result. This agrees with the standard fixed-rank matrix-manifold description (Absil et al., 2008). □

J.1.2 Regular candidate-policy tie manifolds

Fix a candidate set $A \subseteq \Pi$ with $|A| \geq 2$ and a reference $\pi_A \in A$. Put $q_A := |A| - 1$ and define the active-gap map

$$F_A(P) := \big(J^\pi(P) - J^{\pi_A}(P)\big)_{\pi \in A \setminus \{\pi_A\}} \in \mathbb{R}^{q_A}. \tag{399}$$

For $P \in \mathcal{M}_\alpha$, let

$$\mathsf{A}_{A,P} := DF_A(P)|_{\mathcal{L}_P} : \mathcal{L}_P \to \mathbb{R}^{q_A}. \tag{400}$$

Equip $\mathcal{L}_P$ with the Fisher inner product $\langle \cdot, \cdot \rangle_{\mathcal{I}_{P,w}}$ and $\mathbb{R}^{q_A}$ with the Euclidean inner product. Let $\mathsf{A}_{A,P}^{*,P}$ be the corresponding adjoint and define the transversality Gram matrix

$$\mathsf{G}_A(P) := \mathsf{A}_{A,P}\mathsf{A}_{A,P}^{*,P} \in \mathbb{R}^{q_A \times q_A}. \tag{401}$$

The condition $\lambda_{\min}(\mathsf{G}_A(P)) > 0$ is invariant under the choice of reference policy: changing the reference applies an invertible linear transformation to F_A.

J.1.3 Definition of the stable class

Fix constants

$$p_\star, \beta_\star, \kappa_\star, \Delta_\star, \tau_\star, \Gamma_\star > 0. \tag{402}$$

For an admissible pair (α, A), define $\mathcal{P}_{\alpha,A}^{\mathrm{stab}}$ to be the set of all $P \in \mathcal{M}_\alpha$ satisfying:

$$\min_{x,s'} P(s' \mid x) \geq p_\star, \tag{403}$$

$$\min_{\rho \in \Lambda} \|B_\rho(P)\|_{\mathrm{F}} \geq \beta_\star, \qquad \min_{\rho \in \Lambda} \sigma_{r_\rho}(B_\rho(P)) \geq \kappa_\star, \tag{404}$$

$$F_A(P) = 0, \tag{405}$$

$$\min_{\pi \notin A} \big[J^{\pi_A}(P) - J^\pi(P)\big] \geq \Delta_\star, \tag{406}$$

$$\lambda_{\min}\big(\mathsf{G}_A(P)\big) \geq \tau_\star^2, \tag{407}$$

$$\max_{\pi \in A} \big\| DJ_P^\pi|_{\mathcal{L}_P} \big\|_{\mathcal{I}_{P,w},*} \leq \Gamma_\star. \tag{408}$$

The minimum over $\pi \notin A$ is $+\infty$ when $A = \Pi$; the minima over $\rho \in \Lambda$ are vacuous when $s = 0$.

Definition J.2 (Uniformly stable support–rank–policy class)**.** The stable-stratum class is the finite union

$$\boxed{\mathcal{P}_{\mathrm{stab}} := \bigcup_{(\alpha,A) \in \mathfrak{C}_{\mathrm{stab}}} \mathcal{P}_{\alpha,A}^{\mathrm{stab}},} \tag{409}$$

where $\mathfrak{C}_{\mathrm{stab}}$ is any fixed finite collection of admissible saturated support patterns and candidate sets with $|A| \geq 2$. Empty components are discarded.

The condition equation 407 is a full-row-rank condition. Thus the present theorem treats regular ties whose $|A| - 1$ relative value constraints are locally independent. Redundant multi-policy ties of lower constant rank are not included in the adaptive theorem.

J.1.4 COMPLETE STABLE-STRATUM THEOREM

For each nontrivial sector let $\mathsf{B}_\rho : \mathcal{H}_G^0 \to \mathbb{R}^{m_\rho^{\mathcal{X}} \times m_\rho^{\mathcal{S}}}$ denote the linear coordinate extraction map obtained from $\mathcal{B}_G^{-1}$. Define the finite operator constant

$$C_{\mathrm{blk}} := \max_{\rho \in \mathfrak{R}} \sup_{H \in \mathcal{H}_G^0 \setminus \{0\}} \frac{\|\mathsf{B}_\rho(H)\|_{\mathrm{F}}}{\|H\|_{\max}}, \tag{410}$$

with $C_{\mathrm{blk}} := 1$ if $\mathfrak{R} = \varnothing$. Set

$$\delta_{\mathrm{str}} := \min\left\{ \frac{p_\star}{2}, \frac{\beta_\star}{2C_{\mathrm{blk}}}, \frac{\kappa_\star}{2C_{\mathrm{blk}}} \right\}. \tag{411}$$

Theorem J.3 (Geometry and uniform isolation of the stable class)**.** *The class in Definition J.2 has the following properties.*

(i) Compactness and component separation. *Each nonempty $\mathcal{P}_{\alpha,A}^{\mathrm{stab}}$ is compact. Distinct components are disjoint, and because the index family is finite,*

$$\delta_{\mathrm{comp}} := \min_{(\alpha,A) \neq (\alpha',A')} \mathrm{dist}_{\max}\big(\mathcal{P}_{\alpha,A}^{\mathrm{stab}}, \mathcal{P}_{\alpha',A'}^{\mathrm{stab}}\big) > 0, \tag{412}$$

where the minimum is taken only over nonempty pairs.

(ii) Uniform local identification of the model stratum. *For every $P_0 \in \mathcal{P}_{\alpha,A}^{\mathrm{stab}}$,*

$$\boxed{\Theta_{G,s,\boldsymbol{r}} \cap B_{\max}(P_0, \delta_{\mathrm{str}}) = \mathcal{M}_\alpha \cap B_{\max}(P_0, \delta_{\mathrm{str}}).} \tag{413}$$

Consequently, for every fixed $R < \infty$, if

$$n \geq \frac{R^2}{w_\star \delta_{\mathrm{str}}^2}, \tag{414}$$

then every radius-R local alternative around every stable center belongs to the corresponding smooth stratum.

(iii) Uniform regular tie manifolds. *There exists an open neighborhood $\mathcal{U}_{\alpha,A}$ of $\mathcal{P}_{\alpha,A}^{\mathrm{stab}}$ in $\mathcal{M}_\alpha$ such that*

$$\mathcal{T}_{\alpha,A} := \{P \in \mathcal{U}_{\alpha,A} : F_A(P) = 0\} \tag{415}$$

is a smooth embedded submanifold of codimension $q_A = |A| - 1$. At every $P \in \mathcal{T}_{\alpha,A}$,

$$\mathcal{S}_P := T_P \mathcal{T}_{\alpha,A} = \ker \mathsf{A}_{A,P}, \qquad \dim \mathcal{S}_P = d_\alpha - q_A. \tag{416}$$

The Fisher-normal space is

$$\mathcal{N}_P := \mathcal{L}_P \cap \mathcal{S}_P^{\perp_{\mathcal{I}_{P,w}}} = \mathrm{range}(\mathsf{A}_{A,P}^{*,P}), \qquad \dim \mathcal{N}_P = q_A. \tag{417}$$

The Fisher-orthogonal normal projector is

$$\boxed{\Pi_{\mathcal{N}}(P) = \mathsf{A}_{A,P}^{*,P}\big(\mathsf{A}_{A,P}\mathsf{A}_{A,P}^{*,P}\big)^{-1}\mathsf{A}_{A,P}.} \tag{418}$$

After shrinking $\mathcal{U}_{\alpha,A}$ if necessary,

$$\lambda_{\min}\big(\mathsf{A}_{A,P}\mathsf{A}_{A,P}^{*,P}\big) \geq \frac{\tau_\star^2}{2} \qquad (P \in \mathcal{U}_{\alpha,A}). \tag{419}$$

(iv) Uniform candidate-face localization. *The neighborhoods $\mathcal{U}_{\alpha,A}$ may be chosen so that*

$$\max_{\pi \in A} J^\pi(Q) - \max_{\pi \notin A} J^\pi(Q) \geq \frac{\Delta_\star}{2} \qquad (Q \in \mathcal{U}_{\alpha,A}). \tag{420}$$

Hence every locally optimal candidate lies in A throughout the neighborhood, and

$$Q \in \mathcal{T}_{\alpha,A} \quad \Longrightarrow \quad \Pi_0(Q) = A. \tag{421}$$

(v) Uniform charts, tubular neighborhoods, and bundle regularity. *There exist positive constants* $r_{\mathrm{chart}}, r_{\mathrm{tub}}, C_{\mathrm{chart}}, C_{\mathrm{bun}}$*, independent of* $P_0 \in \mathcal{P}_{\mathrm{stab}}$*, such that every stable center admits a* C^3 *chart*

$$\varphi_{P_0} : B_{\mathbb{R}^{d_\alpha - q_A}}(0, r_{\mathrm{chart}}) \times B_{\mathbb{R}^{q_A}}(0, r_{\mathrm{chart}}) \longrightarrow \mathcal{M}_\alpha \tag{422}$$

with

$$\varphi_{P_0}(0,0) = P_0, \qquad \varphi_{P_0}(s,t) \in \mathcal{T}_{\alpha,A} \iff t = 0, \tag{423}$$

and the chart, inverse chart, and their first three derivatives bounded by C_{chart}*. The tie manifolds have a Fisher-normal tubular neighborhood of radius at least* r_{tub} *over the stable class. Moreover, on these neighborhoods the maps*

$$P \longmapsto \mathcal{L}_P, \quad \mathcal{S}_P, \quad \mathcal{N}_P, \quad \Pi_{\mathcal{N}}(P), \quad \mathcal{I}_{P,w}, \quad DJ_P^\pi|_{\mathcal{L}_P} \tag{424}$$

are C^2 *in local coordinates and have derivatives bounded by* C_{bun}*.*

Proof. Step 1: compactness. The set of stochastic kernels is compact in the ambient finite-dimensional space. Within a fixed component, the inactive-block equalities are closed. For every active block, the conditions $\operatorname{rank} B_\rho(P) \le r_\rho$ and $\sigma_{r_\rho}(B_\rho(P)) \ge \kappa_\star$ are closed and together imply $\operatorname{rank} B_\rho(P) = r_\rho$. The positivity, block-norm, tie, inactive-gap, transversality, and gradient-bound conditions are closed because the block maps, singular values, fixed-policy values and derivatives, tangent spaces of the fixed-rank manifolds, and Fisher form vary continuously on the indicated rank-saturated set. Thus each component is compact.

Two different support/rank indices cannot contain the same kernel. If the model index is the same but $A \neq A'$, choose $\pi \in A \triangle A'$. Membership in one component makes π tied for the maximum, while membership in the other either excludes it by the positive inactive gap or requires a different exact active set. Hence the components are disjoint. Finitely many disjoint compact sets have positive pairwise distance, proving equation 412.

Step 2: local model-stratum equality. Let $P_0 \in \mathcal{P}^{\mathrm{stab}}_{\alpha,A}$ and $Q \in \Theta_{G,s,\boldsymbol{r}}$ satisfy $\|Q - P_0\|_{\max} < \delta_{\mathrm{str}}$. For every active block,

$$\|B_\rho(Q) - B_\rho(P_0)\|_{\mathrm{F}} = \|\mathsf{B}_\rho(Q - P_0)\|_{\mathrm{F}} \le C_{\mathrm{blk}} \|Q - P_0\|_{\max}. \tag{425}$$

Therefore

$$\|B_\rho(Q)\|_{\mathrm{F}} \ge \beta_\star/2 > 0.$$

Weyl's singular-value inequality and $\|M\|_{\mathrm{op}} \le \|M\|_{\mathrm{F}}$ give

$$\sigma_{r_\rho}(B_\rho(Q)) \ge \kappa_\star/2 > 0.$$

The ambient model imposes $\operatorname{rank} B_\rho(Q) \le r_\rho$, so every active block has exact rank r_ρ. Since all s active blocks remain nonzero and the total support budget is at most s, every inactive block must remain zero. Finally, $Q(s' \mid x) \ge p_\star/2$, so $Q \in \mathcal{M}_\alpha$. The reverse inclusion is immediate from the definition of the saturated stratum. This proves equation 413.

For $H \in \mathcal{H}^0_G$ and $P_0(s' \mid x) \le 1$,

$$\|H\|^2_{\mathcal{I}_{P_0,w}} = \sum_{x,s'} w_x \frac{H(s' \mid x)^2}{P_0(s' \mid x)} \ge w_\star \|H\|^2_{\mathrm{F}} \ge w_\star \|H\|^2_{\max}. \tag{426}$$

Hence a radius-R local alternative satisfies $\|P - P_0\|_{\max} \le R/\sqrt{n w_\star}$, proving the sample threshold.

Step 3: regular tie manifold and projector. At every stable center, $\mathsf{G}_A(P)$ is positive definite by equation 407; hence $\mathsf{A}_{A,P}$ has full row rank q_A. Full row rank is an open condition. Continuity and compactness permit a common neighborhood on which equation 419 holds. The regular-level set theorem then makes $F_A^{-1}(0)$ a smooth embedded submanifold of codimension q_A, with tangent kernel equation 416.

For a surjective linear map A between finite-dimensional Hilbert spaces, $(\ker \mathsf{A})^\perp = \operatorname{range}(\mathsf{A}^*)$ and $\mathsf{A}\mathsf{A}^*$ is invertible. The operator in equation 418 is self-adjoint, idempotent, has range $\operatorname{range}(\mathsf{A}^*)$, and annihilates $\ker \mathsf{A}$; therefore it is the Fisher-orthogonal projector onto $\mathcal{N}_P$.

Step 4: candidate-face localization. The fixed-policy value maps are C^1. On a compact neighborhood of the stable class their derivatives are uniformly bounded, so there is $L_\star < \infty$ such that

$$|J^\pi(Q) - J^\pi(P)| \le L_\star \|Q - P\|_{\max} \qquad (\pi \in \Pi).$$

Choose the neighborhood radius at most $\Delta_\star/(4L_\star)$. If P is stable and Q lies in this neighborhood, then

$$\begin{aligned}\max_{\pi\in A} J^\pi(Q) - \max_{\pi\notin A} J^\pi(Q) &\geq \Delta_\star - 2L_\star\|Q-P\|_{\max}\\ &\geq \Delta_\star/2.\end{aligned}$$

On the tie manifold, all policies in A have the same value, so the exact base-optimal face is A.

Step 5: uniform charts and tubular neighborhoods. Proposition J.1 and the regular-level set theorem provide C^∞ charts that straighten the tie manifold near every stable center. The fixed-rank chart equation 398, the Fisher form, and the finite-horizon value maps are smooth. The inverse Gram matrix in equation 418 is smooth and uniformly bounded by $2/\tau_\star^2$ on the chosen neighborhood. Hence all bundle maps in equation 424 are C^2.

Each stable component is compact. A finite subcover of its local straightening charts yields common chart radii and derivative bounds. The tubular neighborhood theorem applied to the compact stable subset of each embedded tie manifold yields a positive normal radius; taking the minimum over the finite component family gives $r_{\mathrm{tub}} > 0$. This proves all remaining claims. The use of smooth fixed-rank manifolds and constrained smooth submodels is standard; see, respectively, Absil et al. (2008) and Geyer (1994); Shapiro (2000). □

J.1.5 THE STABLE CLASS IS EXACTLY THE QUANTITATIVE REGULAR LOCUS

Definition J.4 (Qualitatively regular tie center)**.** A kernel $P_0 \in \Theta_{G,s,\boldsymbol{r}}$ is a qualitatively regular tie center if:

1. P_0 is strictly positive;
2. its nontrivial support has size s and every active block has rank exactly its budget r_ρ;
3. its exact base-optimal candidate face $A = \Pi_0(P_0)$ has $|A| \geq 2$;
4. the gap differential $\mathsf{A}_{A,P_0} : \mathcal{L}_{P_0} \to \mathbb{R}^{|A|-1}$ is surjective.

Proposition J.5 (Stable classes exhaust the regular tie locus)**.** *Every point of $\mathcal{P}_{\mathrm{stab}}$ is a qualitatively regular tie center. Conversely, every qualitatively regular tie center belongs to a nonempty compact stable component for some positive constants $p_\star, \beta_\star, \kappa_\star, \Delta_\star, \tau_\star, \Gamma_\star$. Thus Definition J.2 is a quantitative compact exhaustion of the regular support–rank–policy locus.*

Proof. The forward implication is immediate from the definition. Conversely, let P_0 be qualitatively regular with model index α and active face A. Strict positivity, nonzero active blocks, positive smallest active singular values, finiteness of the candidate class, and surjectivity of the gap differential give positive pointwise margins $p_0, \beta_0, \kappa_0, \Delta_0, \tau_0$. The active derivative norms are finite. By continuity, all these inequalities persist—with half their pointwise margins—on a sufficiently small neighborhood of P_0 inside the regular tie manifold. Choose a closed coordinate ball in that neighborhood. It is compact, nonempty, contains P_0, and is a component of the form $\mathcal{P}^{\mathrm{stab}}_{\alpha,A}$ for suitable positive constants and some finite $\Gamma_\star$. □

J.2 THE ORACLE GAME ON THE FISHER NORMAL SPACE

Fix $P_0 \in \mathcal{P}^{\mathrm{stab}}_{\alpha,A}$ and let $k = q_A = |A| - 1$. Choose a Fisher-orthonormal frame $U_{P_0} : \mathbb{R}^k \to \mathcal{N}_{P_0}$, so that

$$U_{P_0}^* \mathcal{I}_{P_0,w} U_{P_0} = I_k. \tag{427}$$

For $\pi \in A$, define the normal derivative coordinate

$$b_\pi(P_0) := \left(D(J^\pi - J^{\pi_A})_{P_0}[U_{P_0} e_j]\right)_{j=1}^k, \qquad b_{\pi_A}(P_0) = 0. \tag{428}$$

Write $\mathbb{B}_R^k := \{z \in \mathbb{R}^k : \|z\|_2 \leq R\}$ and

$$L_B(\zeta, q) := \max_{\pi\in A} b_\pi^\top \zeta - \sum_{\pi\in A} q(\pi) b_\pi^\top \zeta, \qquad B = (b_\pi)_{\pi\in A}. \tag{429}$$

The normal Gaussian-game value is

$$V_R(B) := \inf_{\delta:\mathbb{R}^k\to\Delta(A)} \sup_{\zeta\in\mathbb{B}_R^k} \mathbb{E}_{Z\sim N(\zeta,I_k)} L_B(\zeta,\delta(Z)). \tag{430}$$

Proposition J.6 (Oracle reduction to the tie-normal game)**.** *For every stable center,*

$$\boxed{\mathfrak{R}_R\big(\mathcal{K}_0(P_0),\mathfrak{G}_{0,\mathfrak{T}}(P_0)\big) = V_R\big(B(P_0)\big).} \tag{431}$$

The equality is independent of the chosen Fisher-orthonormal frame.

Proof. On a stable stratum the contingent cone is the linear tangent space $\mathcal{L}_{P_0}$. If $h = s + U_{P_0}\zeta$ with $s \in \mathcal{S}_{P_0}$, then, by the definition of the tie tangent,

$$D(J^\pi - J^{\pi_A})_{P_0}[s] = 0 \qquad (\pi \in A).$$

Subtracting the common derivative of π_A from every active policy leaves policy regret unchanged and gives precisely $L_{B(P_0)}(\zeta, q)$. The Fisher ball in $\mathcal{L}_{P_0}$ is the orthogonal product

$$\{(s,\zeta) : \|s\|_{\mathcal{I}}^2 + \|\zeta\|_2^2 \le R^2\}.$$

The tangent observation similarly decomposes into independent standard Gaussian components on $\mathcal{S}_{P_0}$ and $\mathcal{N}_{P_0}$. The product-nuisance lemma, Lemma H.2, eliminates the tie-tangent component and yields equation 431. Replacing U_{P_0} by $U_{P_0}O$ rotates both $B(P_0)$ and the Gaussian observation by the same $O \in O(k)$, leaving the game value unchanged. □

J.3 Stability of finite-action Gaussian policy-selection games

This subsection proves the deterministic perturbation theorem needed by the adaptive procedure. The result is formulated modulo the three exact symmetries of the Gaussian policy-selection game: orthogonal changes of coordinates, relabeling of candidate policies, and addition of a common vector to every candidate gradient. The proof is self-contained and also yields certified finite-net and plug-in-rule bounds.

Let A be a nonempty finite action set and let $B = (b_a)_{a\in A} \subset \mathbb{R}^d$. For $q \in \Delta(A)$ and $\theta \in \mathbb{R}^d$, define

$$L_B(\theta, q) := \max_{a\in A}\langle b_a,\theta\rangle - \sum_{a\in A} q(a)\langle b_a,\theta\rangle. \tag{432}$$

For a Borel rule $\delta : \mathbb{R}^d \to \Delta(A)$ and a nonempty compact mean set $K \subseteq \mathbb{R}^d$, put

$$r_{B,\delta}(\theta) := \mathbb{E}_{Z\sim N(\theta,I_d)}\left[L_B\big(\theta,\delta(Z)\big)\right], \tag{433}$$

$$\mathcal{R}(K,B;\delta) := \sup_{\theta\in K} r_{B,\delta}(\theta), \tag{434}$$

$$V(K,B) := \inf_{\delta} \mathcal{R}(K,B;\delta), \tag{435}$$

where the infimum ranges over all Borel simplex-valued rules.

The loss is invariant under a common translation of the gradients. It is therefore natural to measure the size of B by its translation-invariant radius

$$\Gamma(B) := \min_{c\in\mathbb{R}^d}\max_{a\in A}\|b_a - c\|_2. \tag{436}$$

The minimum exists because the objective is continuous and coercive. For $R, \Gamma, t \ge 0$, define

$$\vartheta(t) := 2\Phi(t/2) - 1, \qquad \Omega_{R,\Gamma}(t) := 2\Gamma t + 2R\Gamma\vartheta(t). \tag{437}$$

Since $\vartheta'(t) = \varphi(t/2) \le (2\pi)^{-1/2}$,

$$\Omega_{R,\Gamma}(t) \le \left(2 + \sqrt{\frac{2}{\pi}}\,R\right)\Gamma t. \tag{438}$$

J.3.1 EXACT INVARIANCES

Lemma J.7 (Translation, orthogonal, and relabeling invariance)**.** *Let $c \in \mathbb{R}^d$. Then*

$$L_{(b_a+c)_a}(\theta, q) = L_B(\theta, q), \qquad V\big(K, (b_a + c)_a\big) = V(K, B). \tag{439}$$

Let $\widetilde{A}$ be a finite set with $|\widetilde{A}| = |A|$, let $\sigma : A \to \widetilde{A}$ be a bijection, and let $O \in O(d)$. Given $\widetilde{B} = (\widetilde{b}_{\widetilde{a}})_{\widetilde{a}\in\widetilde{A}}$, define the aligned configuration on A by

$$\widetilde{b}_a^{O,\sigma} := O\widetilde{b}_{\sigma(a)}. \tag{440}$$

For a rule $\widetilde{\delta} : \mathbb{R}^d \to \Delta(\widetilde{A})$, define

$$\big(\mathsf{T}_{O,\sigma}\widetilde{\delta}(z)\big)(a) := \widetilde{\delta}(O^\top z)\big(\sigma(a)\big). \tag{441}$$

Then

$$\mathcal{R}\left(O\widetilde{K}, \widetilde{B}^{O,\sigma}; \mathsf{T}_{O,\sigma}\widetilde{\delta}\right) = \mathcal{R}\left(\widetilde{K}, \widetilde{B}; \widetilde{\delta}\right), \tag{442}$$

$$V\left(O\widetilde{K}, \widetilde{B}^{O,\sigma}\right) = V\left(\widetilde{K}, \widetilde{B}\right). \tag{443}$$

Moreover,

$$\Gamma\left(\widetilde{B}^{O,\sigma}\right) = \Gamma(\widetilde{B}). \tag{444}$$

Proof. For the common translation,

$$\begin{aligned} L_{(b_a+c)_a}(\theta, q) &= \max_a \big\{\langle b_a, \theta\rangle + \langle c, \theta\rangle\big\} - \sum_a q(a)\big\{\langle b_a, \theta\rangle + \langle c, \theta\rangle\big\} \\ &= L_B(\theta, q), \end{aligned}$$

which proves equation 439.

For the coordinate and label transformation, write $\theta = O\widetilde{\theta}$ and $z = O\widetilde{z}$. If $q(a) = \widetilde{q}(\sigma(a))$, then

$$\begin{aligned} L_{\widetilde{B}^{O,\sigma}}(O\widetilde{\theta}, q) &= \max_{a\in A}\left\langle O\widetilde{b}_{\sigma(a)}, O\widetilde{\theta}\right\rangle - \sum_{a\in A} q(a)\left\langle O\widetilde{b}_{\sigma(a)}, O\widetilde{\theta}\right\rangle \\ &= L_{\widetilde{B}}(\widetilde{\theta}, \widetilde{q}). \end{aligned}$$

Also, if $Z \sim N(O\widetilde{\theta}, I_d)$, then $O^\top Z \sim N(\widetilde{\theta}, I_d)$. Applying these two identities to equation 441 proves equation 442; taking infima over the bijectively transported rule classes proves equation 443. Finally, Euclidean norms are orthogonally invariant and relabeling does not change a maximum, proving equation 444. □

J.3.2 GAUSSIAN TOTAL VARIATION AND A RULE-UNIFORM RISK MODULUS

Lemma J.8 (Exact total variation between equal-covariance Gaussian shifts)**.** *For every $\theta, \theta' \in \mathbb{R}^d$,*

$$\boxed{\mathrm{TV}\big(N(\theta, I_d), N(\theta', I_d)\big) = 2\Phi\left(\frac{\|\theta - \theta'\|_2}{2}\right) - 1.} \tag{445}$$

Consequently,

$$\mathrm{TV}\big(N(\theta, I_d), N(\theta', I_d)\big) \le \frac{\|\theta - \theta'\|_2}{\sqrt{2\pi}}. \tag{446}$$

Proof. The result is trivial when $\theta = \theta'$. Otherwise put $h := \theta - \theta'$ and $m := (\theta + \theta')/2$. The likelihood ratio of $N(\theta, I_d)$ to $N(\theta', I_d)$ is at least one precisely on the halfspace

$$H := \{z : \langle h, z - m\rangle \ge 0\}.$$

By the Neyman–Pearson characterization of total variation,

$$\mathrm{TV}\big(N(\theta, I_d), N(\theta', I_d)\big) = \mathbb{P}_\theta(H) - \mathbb{P}_{\theta'}(H).$$

Under $N(\theta, I_d)$, $\langle h, Z - m\rangle/\|h\|_2$ is $N(\|h\|_2/2, 1)$; under $N(\theta', I_d)$ it is $N(-\|h\|_2/2, 1)$. Hence the difference equals $2\Phi(\|h\|_2/2) - 1$, proving equation 445. Since the derivative of $t \mapsto 2\Phi(t/2) - 1$ is $\varphi(t/2) \le \varphi(0) = (2\pi)^{-1/2}$, equation 446 follows by integration from zero. □

Lemma J.9 (Rule-uniform modulus of Gaussian policy risk)**.** *Let $K \subseteq R\mathbb{B}_2^d$ and let $B = (b_a)_{a\in A}$ satisfy $\Gamma(B) \le \Gamma$. Then, for every Borel rule δ and every $\theta, \theta' \in R\mathbb{B}_2^d$,*

$$\boxed{|r_{B,\delta}(\theta) - r_{B,\delta}(\theta')| \le \Omega_{R,\Gamma}\big(\|\theta - \theta'\|_2\big).} \tag{447}$$

In particular,

$$|r_{B,\delta}(\theta) - r_{B,\delta}(\theta')| \le \left(2 + \sqrt{\frac{2}{\pi}}R\right)\Gamma\|\theta - \theta'\|_2. \tag{448}$$

The constants are independent of δ.

Proof. Choose c attaining equation 436 and replace every b_a by $b_a - c$. Lemma J.7 leaves the loss unchanged and gives $\max_a \|b_a\|_2 \le \Gamma$. Hence, for every $q \in \Delta(A)$,

$$0 \le L_B(\theta, q) \le 2\Gamma\|\theta\|_2, \tag{449}$$

$$|L_B(\theta, q) - L_B(\theta', q)| \le 2\Gamma\|\theta - \theta'\|_2. \tag{450}$$

The first inequality follows because a convex combination cannot exceed the maximum, and the upper bound follows by comparing the largest and smallest linear forms. The second follows because both the support function and every convexly averaged linear form are Γ-Lipschitz.

Let $P_\theta = N(\theta, I_d)$ and put $f_{\theta'}(z) := L_B(\theta', \delta(z))$. Add and subtract $\mathbb{E}_{P_\theta} f_{\theta'}(Z)$. By equation 450, the first difference is at most $2\Gamma\|\theta - \theta'\|_2$. By equation 449, $0 \le f_{\theta'} \le 2R\Gamma$. For any $0 \le f \le M$, $|\mathbb{E}_P f - \mathbb{E}_Q f| \le M\mathrm{TV}(P, Q)$; this follows, for example, from the layer-cake representation $f = \int_0^M \mathbf{1}\{f > t\}\, dt$. Lemma J.8 therefore bounds the second difference by

$$2R\Gamma\vartheta\big(\|\theta - \theta'\|_2\big).$$

This proves equation 447; equation 448 follows from equation 438. □

J.3.3 Perturbing the gradient configuration

Lemma J.10 (Uniform perturbation of the policy loss)**.** *Let $B = (b_a)_{a\in A}$ and $\widetilde{B} = (\widetilde{b}_a)_{a\in A}$ be two configurations in the same coordinate system and with the same action labels. Define the translation-quotient discrepancy*

$$d_{\mathrm{gr}}(B, \widetilde{B}) := \min_{c\in\mathbb{R}^d} \max_{a\in A} \|b_a - \widetilde{b}_a - c\|_2. \tag{451}$$

If $d_{\mathrm{gr}}(B, \widetilde{B}) \le \eta$, then for every $\theta \in R\mathbb{B}_2^d$ and every $q \in \Delta(A)$,

$$\boxed{\big|L_B(\theta, q) - L_{\widetilde{B}}(\theta, q)\big| \le 2R\eta.} \tag{452}$$

Consequently, for every rule δ,

$$\sup_{\theta\in R\mathbb{B}_2^d} \Big|r_{B,\delta}(\theta) - r_{\widetilde{B},\delta}(\theta)\Big| \le 2R\eta. \tag{453}$$

Proof. The minimum in equation 451 is attained because its objective is continuous and coercive. Choose c attaining it and replace $\widetilde{b}_a$ by $\widetilde{b}_a + c$. The game is unchanged by Lemma J.7, and now $\max_a \|b_a - \widetilde{b}_a\|_2 \le \eta$. Therefore

$$\Big|\max_a \langle b_a, \theta\rangle - \max_a \langle \widetilde{b}_a, \theta\rangle\Big| \le \eta\|\theta\|_2,$$

$$\left|\sum_a q(a)\langle b_a - \widetilde{b}_a, \theta\rangle\right| \le \eta\|\theta\|_2.$$

Adding the two bounds and using $\|\theta\|_2 \le R$ proves equation 452. Taking expectations gives equation 453. □

J.3.4 MAIN STABILITY THEOREM

Theorem J.11 (Lipschitz stability of finite-action Gaussian decision games)**.** *Let $K, \widetilde{K} \subseteq R\mathbb{B}_2^d$ be nonempty compact sets. Let $B = (b_a)_{a\in A}$ and $\widetilde{B} = (\widetilde{b}_{\widetilde{a}})_{\widetilde{a}\in\widetilde{A}}$ be finite gradient configurations with $|A| = |\widetilde{A}|$. Fix an orthogonal map $O \in O(d)$ and a bijection $\sigma : A \to \widetilde{A}$, and define*

$$\varepsilon_O := d_{\mathrm{H}}\left(K, O\widetilde{K}\right), \tag{454}$$

$$\eta_{O,\sigma} := \min_{c\in\mathbb{R}^d} \max_{a\in A} \left\| b_a - O\widetilde{b}_{\sigma(a)} - c \right\|_2, \tag{455}$$

$$\overline{\Gamma} := \max\left\{\Gamma(B), \Gamma(\widetilde{B})\right\}, \tag{456}$$

$$\Delta_{O,\sigma} := \Omega_{R,\overline{\Gamma}}(\varepsilon_O) + 2R\eta_{O,\sigma}. \tag{457}$$

Then the following statements hold.

1. *For every Borel rule $\widetilde{\delta} : \mathbb{R}^d \to \Delta(\widetilde{A})$,*

$$\boxed{\left|\mathcal{R}\left(K, B; \mathsf{T}_{O,\sigma}\widetilde{\delta}\right) - \mathcal{R}\left(\widetilde{K}, \widetilde{B}; \widetilde{\delta}\right)\right| \le \Delta_{O,\sigma}.} \tag{458}$$

2. *The minimax values satisfy*

$$\boxed{\left|V(K, B) - V(\widetilde{K}, \widetilde{B})\right| \le \Delta_{O,\sigma}.} \tag{459}$$

3. *In particular,*

$$\boxed{\left|V(K, B) - V(\widetilde{K}, \widetilde{B})\right| \le \left(2 + \sqrt{\frac{2}{\pi}}R\right)\overline{\Gamma}\,\varepsilon_O + 2R\eta_{O,\sigma} \le (R+2)\overline{\Gamma}\,\varepsilon_O + 2R\eta_{O,\sigma}.} \tag{460}$$

The same inequalities remain true after taking the infimum over all $O \in O(d)$ and all bijections $\sigma : A \to \widetilde{A}$.

Proof. By Lemma J.7, the approximate game and rule can be transported without changing their risks or value. We may therefore replace $(\widetilde{K}, \widetilde{B}, \widetilde{\delta})$ by

$$\left(O\widetilde{K}, (O\widetilde{b}_{\sigma(a)})_{a\in A}, \mathsf{T}_{O,\sigma}\widetilde{\delta}\right)$$

and assume from now on that both games have the same coordinates and action labels. By common-translation invariance, choose a vector c attaining equation 455 and translate all approximate gradients by c. Then

$$\max_{a\in A} \|b_a - \widetilde{b}_a\|_2 \le \eta_{O,\sigma}. \tag{461}$$

Fix any rule $\delta : \mathbb{R}^d \to \Delta(A)$. By Lemma J.9, $r_{B,\delta}$ has modulus $\Omega_{R,\overline{\Gamma}}$. Since the two mean sets are compact, every point of either set has a nearest point in the other. Therefore

$$\left|\sup_{\theta\in K} r_{B,\delta}(\theta) - \sup_{\widetilde{\theta}\in\widetilde{K}} r_{B,\delta}(\widetilde{\theta})\right| \le \Omega_{R,\overline{\Gamma}}(\varepsilon_O). \tag{462}$$

Lemma J.10 and equation 461 give

$$\left|\sup_{\widetilde{\theta}\in\widetilde{K}} r_{B,\delta}(\widetilde{\theta}) - \sup_{\widetilde{\theta}\in\widetilde{K}} r_{\widetilde{B},\delta}(\widetilde{\theta})\right| \le 2R\eta_{O,\sigma}. \tag{463}$$

The triangle inequality proves equation 458.

For any two real families $(f_\delta)_\delta$ and $(g_\delta)_\delta$ indexed by the same set,

$$\left|\inf_\delta f_\delta - \inf_\delta g_\delta\right| \le \sup_\delta |f_\delta - g_\delta|.$$

Applying this elementary inequality to the two worst-case risk families and using equation 458 proves equation 459. The linear estimate equation 460 follows from equation 438. Finally, because the bound holds for every alignment (O, σ), it also holds after taking the infimum over alignments. □

J.3.5 Consequences for approximate rules, finite nets, and plug-in games

Corollary J.12 (Transfer of a near-minimax rule)**.** *Under the assumptions of Theorem J.11, suppose $\widetilde{\delta}$ is ξ-minimax for the approximate game:*

$$\mathcal{R}\left(\widetilde{K}, \widetilde{B}; \widetilde{\delta}\right) \le V(\widetilde{K}, \widetilde{B}) + \xi. \tag{464}$$

Then its transported version satisfies

$$\boxed{\mathcal{R}\left(K, B; \mathsf{T}_{O,\sigma}\widetilde{\delta}\right) \le V(K, B) + \xi + 2\Delta_{O,\sigma}.} \tag{465}$$

Proof. By fixed-rule stability,

$$\mathcal{R}\left(K, B; \mathsf{T}_{O,\sigma}\widetilde{\delta}\right) \le \mathcal{R}\left(\widetilde{K}, \widetilde{B}; \widetilde{\delta}\right) + \Delta_{O,\sigma}.$$

Use equation 464 and then value stability: $V(\widetilde{K}, \widetilde{B}) \le V(K, B) + \Delta_{O,\sigma}$. □

Corollary J.13 (Certified finite-net approximation)**.** *Let $K \subseteq R\mathbb{B}_2^d$ be nonempty compact, let $F \subseteq K$ be a nonempty finite ε-net of K, and let B satisfy $\Gamma(B) \le \Gamma$. Then*

$$\boxed{0 \le V(K, B) - V(F, B) \le \Omega_{R,\Gamma}(\varepsilon).} \tag{466}$$

If δ_F is ξ-minimax for the finite game, then

$$\boxed{\mathcal{R}(K, B; \delta_F) \le V(K, B) + \xi + \Omega_{R,\Gamma}(\varepsilon).} \tag{467}$$

Proof. Because $F \subseteq K$, monotonicity in the parameter set gives $V(F, B) \le V(K, B)$. Since $d_{\mathrm{H}}(F, K) \le \varepsilon$, the stability theorem with identical gradients gives the upper bound in equation 466. Finally,

$$\mathcal{R}(K, B; \delta_F) \le \mathcal{R}(F, B; \delta_F) + \Omega_{R,\Gamma}(\varepsilon) \le V(F, B) + \xi + \Omega_{R,\Gamma}(\varepsilon),$$

and $V(F, B) \le V(K, B)$ proves equation 467. □

Corollary J.14 (Random plug-in games under sample splitting)**.** *Let (K, B) be a fixed true game. Let $\mathcal{H}_n$ be a pilot sigma-field, and conditional on $\mathcal{H}_n$ let the main observation have law $N(\theta, I_d)$ under every $\theta \in K$. Suppose $\widehat{K}_n, \widehat{B}_n, O_n, \sigma_n$, and a rule $\widehat{\delta}_n$ are $\mathcal{H}_n$-measurable, and $\widehat{\delta}_n$ is conditionally ξ_n-minimax for $(\widehat{K}_n, \widehat{B}_n)$. Define the pathwise errors*

$$\widehat{\varepsilon}_n := d_{\mathrm{H}}\left(K, O_n\widehat{K}_n\right), \tag{468}$$

$$\widehat{\eta}_n := \min_c \max_{a \in A} \left\| b_a - O_n \widehat{b}_{\sigma_n(a),n} - c \right\|_2, \tag{469}$$

$$\widehat{\Delta}_n := \Omega_{R,\overline{\Gamma}_n}(\widehat{\varepsilon}_n) + 2R\widehat{\eta}_n, \tag{470}$$

where $\overline{\Gamma}_n := \max\{\Gamma(B), \Gamma(\widehat{B}_n)\}$. Then the transported data rule obeys

$$\boxed{\sup_{\theta \in K} \mathbb{E}_\theta L_B\left(\theta, \mathsf{T}_{O_n,\sigma_n}\widehat{\delta}_n(Z)\right) \le V(K, B) + \sup_{\theta \in K} \mathbb{E}_\theta\left[\xi_n + 2\widehat{\Delta}_n\right].} \tag{471}$$

Consequently, if the rightmost supremum converges to zero, the plug-in rule attains the true Gaussian minimax value.

Proof. Condition on $\mathcal{H}_n$. For every realized pilot outcome, Corollary J.12 applies deterministically and its bound is uniform over $\theta \in K$. Take the conditional expectation under each θ, then the unconditional expectation, and finally the supremum in θ. □

Corollary J.15 (Continuity on a stable component)**.** *Fix one stable component $\mathcal{P}^{\mathrm{stab}}_{\alpha,A}$. Then*

$$P \longmapsto V_R\big(B(P)\big) \tag{472}$$

is continuous and hence uniformly continuous on that component. More precisely, if $P_n \to P$ and the normal frames are aligned by orthogonal maps O_n, then

$$|V_R(B(P_n)) - V_R(B(P))| \le 2R \min_c \max_{\pi\in A} \|b_\pi(P) - O_n b_\pi(P_n) - c\|_2 \longrightarrow 0. \tag{473}$$

Proof. On one stable component, the normal dimension and active candidate labels are fixed. The smooth bundle statement in Proposition J.3 gives local smooth Fisher-orthonormal frames; after orthogonal frame alignment, the normal derivative coordinates vary continuously. Both normal mean sets are the same ball $\mathbb{B}_R^{q_A}$, so Theorem J.11 applies with zero set error and yields equation 473. Compactness then upgrades continuity to uniform continuity. □

J.4 Construction of the adaptive estimator

We now construct one rule that does not use the unknown center, Fisher metric, support–rank stratum, or tied candidate face. The rule may use the fixed stable-class margins and the fixed candidate class. The construction has three ingredients: a vanishing pilot sample identifies the discrete component and estimates its normal geometry; the remaining sample produces an efficient normal statistic through unrestricted and tie-constrained likelihood projections; and a finite-net Gaussian decision rule converts that statistic into a policy lottery. Sample splitting is used only to make the selected rule fixed conditional on the pilot. The likelihood-projection expansions are proved below directly in the finite categorical model; they are consistent with classical constrained M-estimation theory (Geyer, 1994; Shapiro, 2000).

J.4.1 Explicit sample split and thresholds

For the total sample size n, define

$$\rho_n := n^{-1/3}, \qquad m_{n,x} := \lfloor \rho_n n_x \rfloor, \qquad N_{n,x} := n_x - m_{n,x}, \tag{474}$$

and put $m_n := \sum_x m_{n,x}$ and $N_n := \sum_x N_{n,x}$. Since $w_{n,x} := n_x/n \to w_x \ge w_\star$, uniformly over the finite row set,

$$m_n \asymp n^{2/3}, \qquad N_n/n \longrightarrow 1, \qquad \min_x m_{n,x} \asymp n^{2/3}. \tag{475}$$

Use the common deterministic threshold

$$t_n := n^{-1/6}. \tag{476}$$

Thus $t_n \downarrow 0$ and $m_n t_n^2 \asymp n^{1/3}$.

Let $\bar{P}_n^{\mathrm{pil}}$ be the rowwise empirical transition kernel based on the pilot observations. Define the Reynolds symmetrization

$$(\mathsf{R}_G Q)(s' \mid x) := \frac{1}{|G|} \sum_{g\in G} Q(g^{-1}s' \mid g^{-1}x), \tag{477}$$

and set

$$\widetilde{P}_n^{\mathrm{pil}} := \mathsf{R}_G \bar{P}_n^{\mathrm{pil}}. \tag{478}$$

The map R_G preserves nonnegativity and row stochasticity and is a contraction in maximum-entry norm. Since the true transition kernel is G-equivariant, it is fixed by R_G.

Define

$$\widehat{\Lambda}_n := \left\{ \rho \in \mathfrak{R} : \|B_\rho(\widetilde{P}_n^{\mathrm{pil}})\|_{\mathrm{F}} > t_n \right\}, \tag{479}$$

$$\widehat{r}_{\rho,n} := \#\left\{ j : \sigma_j(B_\rho(\widetilde{P}_n^{\mathrm{pil}})) > t_n \right\}, \qquad \rho \in \widehat{\Lambda}_n, \tag{480}$$

$$\widehat{A}_n := \left\{ \pi \in \Pi : \max_{\pi'\in\Pi} J^{\pi'}(\widetilde{P}_n^{\mathrm{pil}}) - J^{\pi}(\widetilde{P}_n^{\mathrm{pil}}) \le t_n \right\}. \tag{481}$$

If the resulting pair $(\widehat{\alpha}_n, \widehat{A}_n)$ is not in the fixed finite component list $\mathfrak{C}_{\mathrm{stab}}$, the procedure declares a pilot failure and uses a fixed fallback policy.

J.4.2 Compact safe sets

For an admissible support–rank index $\alpha = (\Lambda, (r_\rho)_{\rho\in\Lambda})$, define the compact safe model patch

$$\begin{aligned}\mathcal{M}_\alpha^{\mathrm{safe}} := \Big\{ Q \in \mathcal{A}_G :\ & \min_{x,s'} Q(s' \mid x) \ge p_\star/2, \\ & B_\rho(Q) = 0 \quad (\rho \notin \Lambda), \\ & \operatorname{rank} B_\rho(Q) \le r_\rho \quad (\rho \in \Lambda), \\ & \sigma_{r_\rho}(B_\rho(Q)) \ge \kappa_\star/2, \quad \|B_\rho(Q)\|_{\mathrm{F}} \ge \beta_\star/2 \quad (\rho \in \Lambda)\Big\}.\end{aligned} \tag{482}$$

The singular-value lower bound and the rank upper bound force exact rank r_ρ. Hence $\mathcal{M}_\alpha^{\mathrm{safe}}$ is a compact subset of the regular stratum $\mathcal{M}_\alpha$.

For a candidate face A, define the compact safe tie patch

$$\begin{aligned}\mathcal{T}_{\alpha,A}^{\mathrm{safe}} := \Big\{ Q \in \mathcal{M}_\alpha^{\mathrm{safe}} :\ & F_A(Q) = 0, \\ & \min_{\pi\notin A} [J^{\pi_A}(Q) - J^\pi(Q)] \ge \Delta_\star/2, \\ & \lambda_{\min}(\mathsf{G}_A(Q)) \ge \tau_\star^2/4, \\ & \max_{\pi\in A} \|DJ_Q^\pi|_{\mathcal{L}_Q}\|_{\mathcal{I}_{Q,w},*} \le 2\Gamma_\star\Big\}.\end{aligned} \tag{483}$$

The minimum over $\pi \notin A$ is $+\infty$ if $A = \Pi$. Every stable center in $\mathcal{P}_{\alpha,A}^{\mathrm{stab}}$ belongs to both safe patches.

On a valid pilot component, define $\widetilde{P}_{0,n}$ to be the lexicographically first minimizer, in fixed ambient coordinates, of

$$Q \longmapsto \|Q - \widetilde{P}_n^{\mathrm{pil}}\|_{\mathrm{F}} \quad \text{over } Q \in \mathcal{T}_{\widehat{\alpha}_n,\widehat{A}_n}^{\mathrm{safe}}. \tag{484}$$

Compactness gives existence. The lexicographic rule gives a Borel selection; equivalently, one may invoke the measurable maximum theorem. If the safe tie set is empty, the procedure declares a pilot failure.

Let

$$w_{N,n,x} := N_{n,x}/N_n \tag{485}$$

and let $\widehat{\mathcal{I}}_n := \mathcal{I}_{\widetilde{P}_{0,n},w_{N,n}}$. At $\widetilde{P}_{0,n}$, form the Fisher-normal projector using equation 418, with w replaced by $w_{N,n}$. A Borel Fisher-orthonormal frame

$$\widehat{U}_n : \mathbb{R}^{\widehat{k}_n} \longrightarrow \widehat{\mathcal{N}}_n \tag{486}$$

is obtained by projecting the fixed ambient coordinate vectors, selecting the lexicographically first linearly independent $\widehat{k}_n$-tuple, and applying Fisher Gram–Schmidt. This construction is Borel on every fixed-rank normal bundle.

For the lexicographically first reference $\widehat{\pi}_n \in \widehat{A}_n$, define

$$\widehat{b}_{\pi,n} := \Big(D(J^\pi - J^{\widehat{\pi}_n})_{\widetilde{P}_{0,n}}[\widehat{U}_n e_j]\Big)_{j=1}^{\widehat{k}_n}, \qquad \pi \in \widehat{A}_n. \tag{487}$$

J.4.3 Main-sample likelihood projections

Let

$$\ell_n^{\mathrm{main}}(Q) := \sum_{x\in\mathcal{X}} \sum_{i=m_{n,x}+1}^{n_x} \log Q(Y_{x,i} \mid x) \tag{488}$$

be the main-sample log likelihood. On a valid pilot component, define $\widehat{P}_{M,n}$ and $\widehat{P}_{T,n}$ as the lexicographically first maximizers of ℓ_n^{main} over

$$\mathcal{M}_{\widehat{\alpha}_n}^{\mathrm{safe}} \quad \text{and} \quad \mathcal{T}_{\widehat{\alpha}_n,\widehat{A}_n}^{\mathrm{safe}}, \tag{489}$$

respectively. These maximizers exist and are Borel because the parameter sets are compact and the log likelihood is continuous and finite on them.

Define the adaptive normal statistic

$$\boxed{\widehat{Z}_n := \widehat{U}_n^* \widehat{\mathcal{I}}_n \sqrt{N_n}(\widehat{P}_{M,n} - \widehat{P}_{T,n}).} \tag{490}$$

The pilot and main samples are independent, so conditional on the pilot the estimated geometry and the Gaussian rule chosen below are fixed.

J.5 Uniform recovery of the discrete geometry

Let $\mathcal{E}_n$ be the event that the support, all active ranks, and the candidate face are recovered correctly, the safe tie set is nonempty, and the projection equation 484 is defined.

Lemma J.16 (Exponentially accurate pilot recovery)**.** *For every fixed $R < \infty$, there exist constants $C, c > 0$, depending only on the stable class and the fixed model, such that, for all sufficiently large n,*

$$\boxed{\sup_{P_0 \in \mathcal{P}_{\mathrm{stab}}} \sup_{P \in \Theta_{n,R}(P_0)} \mathbb{P}_P(\mathcal{E}_n^c) \le C \exp(-cn^{1/3}).} \tag{491}$$

Moreover, on $\mathcal{E}_n$,

$$\sup_{P_0, P} \|\widetilde{P}_{0,n} - P_0\|_{\mathrm{F}} = O_{\mathbb{P}}(n^{-1/3}) \tag{492}$$

uniformly over the same local family.

Proof. Let $m_{n,\min} := \min_x m_{n,x}$. Hoeffding's inequality and a union bound give for every $u > 0$,

$$\sup_P \mathbb{P}_P\big(\|\bar{P}_n^{\mathrm{pil}} - P\|_{\max} > u\big) \le 2|\mathcal{X}||\mathcal{S}| \exp(-2m_{n,\min}u^2). \tag{493}$$

The Reynolds map is a maximum-norm contraction. Every block-extraction map and every fixed-policy value map is Lipschitz on the compact safe neighborhood; write C_{blk} and L_{val} for common constants. By Proposition J.3, every radius-R local alternative is on the same regular stratum for large n, and

$$\|P - P_0\|_{\max} \le R/(\sqrt{nw_\star}). \tag{494}$$

Choose a fixed $c_0 > 0$ small enough that $C_{\mathrm{blk}}c_0 t_n$ and $L_{\mathrm{val}}c_0 t_n$ are at most $t_n/4$. On the event in which the empirical error is at most $c_0 t_n$, all inactive blocks have estimated norm and singular values below t_n, while all active blocks remain above the threshold by the fixed margins $\beta_\star, \kappa_\star$. Weyl's inequality gives rank recovery. Every policy in $A = \Pi_0(P_0)$ has pilot estimated gap $o(t_n)$ because both the pilot error and the local displacement are $o(t_n)$; every policy outside A retains gap at least $\Delta_\star/2$ for large n. This proves correct face recovery. Since $m_{n,\min}t_n^2 \asymp n^{1/3}$, equation 493 yields equation 491.

On $\mathcal{E}_n$, P_0 belongs to the selected safe tie set. By optimality of the metric projection,

$$\|\widetilde{P}_{0,n} - \widetilde{P}_n^{\mathrm{pil}}\|_{\mathrm{F}} \le \|P_0 - \widetilde{P}_n^{\mathrm{pil}}\|_{\mathrm{F}}.$$

Therefore

$$\|\widetilde{P}_{0,n} - P_0\|_{\mathrm{F}} \le 2\|\widetilde{P}_n^{\mathrm{pil}} - P_0\|_{\mathrm{F}} = O_{\mathbb{P}}(m_n^{-1/2}) + O(n^{-1/2}),$$

uniformly, which is equation 492. □

J.6 Uniform likelihood projection on the stable manifolds

We next prove the asymptotic expansion of the two main-sample maximizers. The argument is stated in the adapted charts of Proposition J.3 and does not invoke an abstract black-box M-estimation theorem.

For a stable center $P_0 \in \mathcal{P}_{\alpha,A}^{\mathrm{stab}}$, let

$$\varphi_{P_0} : B_{r_0}^{p-k} \times B_{r_0}^{k} \longrightarrow \mathcal{M}_\alpha \tag{495}$$

be a uniform C^3 chart satisfying

$$\varphi_{P_0}(0,0) = P_0, \qquad \varphi_{P_0}(s,t) \in \mathcal{T}_{\alpha,A} \iff t = 0, \tag{496}$$

and normalized so that $D\varphi_{P_0}(0)$ is a Fisher isometry from $\mathbb{R}^p$ to $\mathcal{L}_{P_0}$, with the last k coordinates mapped to $\mathcal{N}_{P_0}$.

For the main sample, define the chart score

$$\Delta_{n,P_0} := \frac{1}{\sqrt{N_n}} \nabla_\theta \ell_n^{\mathrm{main}}(\varphi_{P_0}(\theta))\big|_{\theta=0} \in \mathbb{R}^p. \tag{497}$$

Write $\Delta_{n,P_0} = (\Delta_{n,s}, \Delta_{n,t})$ according to the adapted splitting.

Lemma J.17 (Uniform quadratic chart likelihood)**.** *Fix $R, M < \infty$. Uniformly over $P_0 \in \mathcal{P}_{\mathrm{stab}}$ and $P \in \Theta_{n,R}(P_0)$, the following hold.*

Let $h_n(P; P_0) := \sqrt{N_n}\, \varphi_{P_0}^{-1}(P)$. Then $\|h_n(P; P_0)\|_2 \le R + o(1)$ and

$$d_{\mathrm{BL}}\left(\mathcal{L}_P(\Delta_{n,P_0}), N(h_n(P; P_0), I_p)\right) \longrightarrow 0. \tag{498}$$

Moreover, if

$$Q_{n,P_0}(v) := \ell_n^{\mathrm{main}}(\varphi_{P_0}(v/\sqrt{N_n})) - \ell_n^{\mathrm{main}}(P_0), \tag{499}$$

then

$$\boxed{\begin{aligned} \sup_{\|v\|_2 \le M} \Big[& |Q_{n,P_0}(v) - v^\top \Delta_{n,P_0} + \tfrac{1}{2}\|v\|_2^2| \\ & + \|\nabla Q_{n,P_0}(v) - (\Delta_{n,P_0} - v)\|_2 \\ & + \|\nabla^2 Q_{n,P_0}(v) + I_p\|_{\mathrm{op}} \Big] \xrightarrow{P} 0. \end{aligned}} \tag{500}$$

Proof. On the safe patches every transition probability is at least $p_\star/2$. The uniform C^3 chart bounds in Proposition J.3 therefore imply a common bound on the first three derivatives, in θ, of each one-observation log likelihood $\log \varphi_{P_0}(\theta)(Y \mid x)$ for $\|\theta\| \le r_0$. Taylor's theorem at zero, summed over the independent main observations, gives uniformly for bounded v

$$Q_{n,P_0}(v) = v^\top \Delta_{n,P_0} - \frac{1}{2} v^\top I_{n,P_0} v + R_{n,P_0}(v), \tag{501}$$

where the deterministic third-order contribution is bounded by $C_M/\sqrt{N_n}$ and the empirical Hessian fluctuation is a finite sum of bounded independent matrices. Hoeffding's inequality and a finite-dimensional net yield

$$\sup_{P_0, P} \mathbb{P}_P\left(\|I_{n,P_0} - I_p\|_{\mathrm{op}} > \epsilon\right) \longrightarrow 0. \tag{502}$$

The same derivative bounds control the gradient and Hessian remainders, proving equation 500.

The score summands are uniformly bounded and have covariance tending to I_p. Under $P = P_0 + O(n^{-1/2})$, their mean equals $h_n(P; P_0) + o(1)$ by the same categorical likelihood expansion used in Appendix E. The finite-dimensional triangular array Lindeberg theorem, uniformly by compactness and a subsequence argument, gives equation 498. Finally, the chart derivative is a Fisher isometry and its quadratic remainder is uniform, so the Fisher-radius constraint implies $\|h_n(P; P_0)\| \le R + o(1)$. □

Lemma J.18 (Uniform localization of the global maximizers)**.** *For every $\epsilon > 0$,*

$$\begin{aligned} \sup_{P_0 \in \mathcal{P}_{\mathrm{stab}}} \sup_{P \in \Theta_{n,R}(P_0)} \mathbb{P}_P\Big(& \|\widehat{P}_{M,n} - P_0\|_{\mathrm{F}} > \epsilon \\ & \text{or } \|\widehat{P}_{T,n} - P_0\|_{\mathrm{F}} > \epsilon,\ \mathcal{E}_n \Big) \longrightarrow 0. \end{aligned} \tag{503}$$

Proof. All safe sets are compact and every candidate transition probability on them is at least $p_\star/2$. Hence their log likelihoods are uniformly bounded and continuous. Identifiability of the product categorical experiment and positive limiting row weights imply that, for every fixed $\epsilon > 0$,

$$\begin{aligned} c_\epsilon := \inf\Big\{ & \sum_x w_x D_{\mathrm{KL}}(P_0(\cdot \mid x) \| Q(\cdot \mid x)) :\ P_0 \in \mathcal{P}_{\mathrm{stab}}, \\ & Q \in \mathcal{M}_\alpha^{\mathrm{safe}},\ \|Q - P_0\|_{\mathrm{F}} \ge \epsilon \Big\} > 0, \end{aligned} \tag{504}$$

where the finite component index is matched to P_0. The same lower bound, possibly decreased, holds when Q is restricted to the safe tie set. The strict positivity follows by compactness and the fact that zero categorical KL divergence forces equality of every row.

The true local alternative satisfies $\|P - P_0\| = O(n^{-1/2})$, so replacing P_0 by P in the expectation changes the KL contrast by $o(1)$ uniformly. A finite-net uniform law of large numbers for the bounded log-likelihood class then implies that every point at distance at least ϵ has main log likelihood below that of P in the model problem and below that of P_0 in the tie-constrained problem, with probability tending to one uniformly. This proves localization of both global maximizers. □

Theorem J.19 (Uniform adaptive normal experiment)**.** *Fix $R < \infty$. On $\mathcal{E}_n$, there exist orthogonal matrices $O_n(P_0) \in O(q_A)$, used only for theoretical alignment, such that*

$$\boxed{\begin{aligned} &\sup_{P_0 \in \mathcal{P}^{\mathrm{stab}}_{\alpha,A}} \sup_{P \in \Theta_{n,R}(P_0)} \mathbb{E}_P\Big[\mathbf{1}_{\mathcal{E}_n} d_{\mathrm{BL}}\Big(\mathcal{L}_P\big(O_n(P_0)\widehat{Z}_n \mid \mathcal{F}_n^{\mathrm{pil}}\big), N(\zeta_n(P;P_0), I_{q_A})\Big)\Big] \longrightarrow 0, \\ &\sup_{P_0 \in \mathcal{P}^{\mathrm{stab}}_{\alpha,A}} \sup_{P \in \Theta_{n,R}(P_0)} \mathbb{E}_P\Big[\mathbf{1}_{\mathcal{E}_n} \max_{\pi \in A} \|O_n(P_0)\widehat{b}_{\pi,n} - b_\pi(P_0)\|_2\Big] \longrightarrow 0, \end{aligned}} \tag{505}$$

where

$$\zeta_n(P;P_0) := U_{P_0}^* \mathcal{I}_{P_0,w} \sqrt{n}(P - P_0). \tag{506}$$

Proof. Work on $\mathcal{E}_n$ and fix one stable component. By Lemma J.18, both main maximizers enter the uniform adapted chart with probability tending to one. Write their scaled coordinates as $\widehat{v}_M$ and $\widehat{v}_T$. The safe-margin constraints are inactive in that neighborhood. The model maximizer therefore satisfies the unconstrained score equation, while the tie maximizer has last k coordinates zero and satisfies the first $p-k$ score equations. The Hessian part of equation 500 is uniformly negative definite on every fixed local ball. The score at zero is $O_{\mathbb{P}}(1)$, and strong concavity first gives $\widehat{v}_M, \widehat{v}_T = O_{\mathbb{P}}(1)$ uniformly. The gradient expansion then yields

$$\widehat{v}_M = \Delta_{n,P_0} + o_{\mathbb{P}}(1), \qquad \widehat{v}_T = (\Delta_{n,s}, 0) + o_{\mathbb{P}}(1), \tag{507}$$

uniformly. Taylor expansion of the chart gives

$$\sqrt{N_n}(\widehat{P}_{M,n} - \widehat{P}_{T,n}) = U_{P_0}\Delta_{n,t} + o_{\mathbb{P}}(1), \tag{508}$$

where U_{P_0} is the chosen true Fisher-normal frame.

By Lemma J.16, $\widetilde{P}_{0,n} \to P_0$ uniformly in probability. The smooth-bundle statement equation 424 and $w_{N,n} \to w$ imply convergence of the pilot normal projector and Fisher form. For any two orthonormal frames of subspaces whose orthogonal projectors converge, the orthogonal Procrustes construction supplies an $O_n(P_0)$ such that

$$\|\widehat{U}_n O_n(P_0)^\top - U_{P_0}\|_{\mathrm{F}} \xrightarrow{P} 0 \tag{509}$$

uniformly. Applying the pilot projection in equation 490 to equation 508, and using $N_n/n \to 1$, gives

$$O_n(P_0)\widehat{Z}_n = \Delta_{n,t} + o_{\mathbb{P}}(1). \tag{510}$$

Because $D\varphi_{P_0}(0)$ is a Fisher isometry and the chart has a uniform quadratic remainder, the last k coordinates of $h_n(P;P_0)$ differ from $\sqrt{N_n/n}\,\zeta_n(P;P_0)$ by $o(1)$ uniformly. Since $N_n/n \to 1$, the Gaussian mean in equation 498 may therefore be replaced by $\zeta_n(P;P_0)$. The conditional bounded–Lipschitz statement follows from Lemma J.17; because the pilot and main samples are independent, the argument applies conditional on every pilot realization in a probability-one good set, and dominated subsequence convergence upgrades it to the displayed expectation.

Finally, the active value derivatives, the Fisher metric, and the normal bundle are continuous with uniformly bounded derivatives on the stable class. Combining this fact with equation 492 and equation 509 proves the gradient-coordinate convergence. □

J.7 A FINITE MEASURABLE FAMILY OF NEAR-MINIMAX GAUSSIAN RULES

For fixed A, k and $M := 4\Gamma_\star$, let

$$\mathfrak{B}_{k,M}(A) := \{B = (b_\pi)_{\pi\in A} : b_{\pi_A} = 0,\ \max_\pi \|b_\pi\|_2 \le M\}. \tag{511}$$

This is compact. For every integer $m \ge 1$, choose a finite m^{-1}-net $\mathcal{G}_m$ in the maximum gradient norm. For each $G \in \mathcal{G}_m$, Proposition F.4 supplies a globally Lipschitz rule $\delta_{m,G}$ satisfying

$$\sup_{\|\zeta\|\le R} \mathbb{E}_\zeta L_G(\zeta, \delta_{m,G}(Z)) \le V_R(G) + m^{-1}. \tag{512}$$

Since $\mathcal{G}_m$ is finite, the maximum of their Lipschitz constants is a finite deterministic number L_m. Let $\mathsf{Q}_m(B)$ be the lexicographically first nearest net point and define

$$\mathsf{D}_m(B) := \delta_{m,\mathsf{Q}_m(B)}. \tag{513}$$

Lemma J.20 (Uniform finite-net near-minimaxity)**.** *For every $B \in \mathfrak{B}_{k,M}(A)$,*

$$\boxed{\sup_{\|\zeta\|\le R} \mathbb{E}_\zeta L_B(\zeta, \mathsf{D}_m(B)(Z)) \le V_R(B) + \varepsilon_m, \qquad \varepsilon_m := \frac{1+4R}{m}.} \tag{514}$$

The map $B \mapsto \mathsf{D}_m(B)$ is Borel and every selected rule is L_m-Lipschitz.

Proof. Let $G = \mathsf{Q}_m(B)$. Then $\max_\pi \|b_\pi - g_\pi\|_2 \le m^{-1}$. The gradient perturbation part of Theorem J.11 gives both

$$|V_R(B) - V_R(G)| \le 2R/m \tag{515}$$

and, for the fixed rule $\delta_{m,G}$,

$$|\mathcal{R}_R(B; \delta_{m,G}) - \mathcal{R}_R(G; \delta_{m,G})| \le 2R/m. \tag{516}$$

Combining these two inequalities with equation 512 proves the displayed bound. Nearest-point selection in a finite set with lexicographic tie breaking is Borel. □

Choose a deterministic sequence $m(n) \uparrow \infty$ sufficiently slowly that all fixed-m uniform approximation errors in Theorem J.19, after multiplication by $1 + L_m$, converge to zero along $m = m(n)$. Such a sequence exists by a standard diagonal construction because the convergence is uniform for every fixed m.

On $\mathcal{E}_n$, define

$$\boxed{\widehat{\delta}_{n,R} := \mathsf{D}_{m(n)}(\widehat{B}_n)(\widehat{Z}_n),} \tag{517}$$

with zero mass assigned to candidates outside $\widehat{A}_n$. On $\mathcal{E}_n^c$, select one fixed candidate policy. The output of equation 517 is a policy lottery; an external draw from that lottery is used for the complete episode.

J.8 UNIFORM LOWER BOUND OVER THE STABLE CLASS

Lemma J.21 (Uniform oracle lower bound on stable strata)**.** *For every fixed $R < \infty$ and every sequence of data-dependent randomized candidate-policy rules δ_n,*

$$\boxed{\liminf_{n\to\infty} \inf_{P_0\in\mathcal{P}_{\mathrm{stab}}} \left[\sup_{P\in\Theta_{n,R}(P_0)} \mathbb{E}_P \sqrt{n}\big(J^\star(P) - J^{\delta_n}(P)\big) - V_R(B(P_0))\right] \ge 0.} \tag{518}$$

Proof. Suppose the conclusion fails. Then there are $\epsilon > 0$, a subsequence n_j, centers $P_{0,j} \in \mathcal{P}_{\mathrm{stab}}$, and rules δ_{n_j} such that

$$\sup_{P\in\Theta_{n_j,R}(P_{0,j})} \mathbb{E}_P \sqrt{n_j}\big(J^\star(P) - J^{\delta_{n_j}}(P)\big) \le V_R(B(P_{0,j})) - \epsilon. \tag{519}$$

Because the stable class is a finite union of compact components, pass to a further subsequence on one component with $P_{0,j} \to P_\infty$. Align the normal frames orthogonally. Smoothness gives $B(P_{0,j}) \to B(P_\infty)$ and Theorem J.11 gives $V_R(B(P_{0,j})) \to V_R(B(P_\infty))$.

Choose $0 < \eta < R$ and a finite set $F = \{\zeta_1, \dots, \zeta_M\} \subset (R-\eta)\mathbb{B}_2^{q_A}$ whose finite Gaussian-game value is at least $V_R(B(P_\infty)) - \epsilon/4$; finite exhaustion follows from the common risk modulus. In the uniform adapted chart at $P_{0,j}$, define the exact model points

$$P_{j,\ell} := \varphi_{P_{0,j}}\left(0, \frac{\zeta_\ell}{\sqrt{n_j}}\right). \tag{520}$$

The chart's quadratic remainder and the margin η imply $P_{j,\ell} \in \Theta_{n_j,R}(P_{0,j})$ for all large j. The uniform categorical likelihood expansion gives joint convergence of the finite likelihood-ratio vector to that of the Gaussian shifts $N(\zeta_\ell, I_{q_A})$, while the uniform Bellman expansion gives convergence of every policy-lottery loss to $L_{B(P_\infty)}(\zeta_\ell, q)$. By the uniform inactive-candidate gap, for all large j any rule may be pointwise modified by moving all mass outside A to the reference policy π_A; this weakly decreases regret over the complete local neighborhood. Hence the lower-bound action space may be restricted to the compact simplex $\Delta(A)$, on which the scaled local losses are uniformly bounded. The finite likelihood-ratio second-moment lemma supplies uniform integrability. Applying the finite-experiment lower-semicontinuity lemma from Appendix F yields

$$\liminf_j \max_{\ell \le M} \mathbb{E}_{P_{j,\ell}} \sqrt{n_j}\big(J^\star(P_{j,\ell}) - J^{\delta_{n_j}}(P_{j,\ell})\big) \ge V(F, B(P_\infty)). \tag{521}$$

The right side is at least $V_R(B(P_\infty)) - \epsilon/4$, contradicting equation 519 for large j. □

J.9 Adaptive attainment theorem

Theorem J.22 (Complete construction and adaptive attainment)**.** *Fix $R < \infty$ and the quantitative stable class of Definition J.2. The rule $\widehat{\delta}_{n,R}$ in equation 517 is a single Borel data-dependent randomized rule. It depends on the known model class, candidate class, design, radius, and stable-class margins, but not on the unknown center, its Fisher metric, its support–rank component, or its tied candidate face. It satisfies*

$$\boxed{\sup_{P_0 \in \mathcal{P}_{\mathrm{stab}}} \left| \sup_{P \in \Theta_{n,R}(P_0)} \mathbb{E}_P\Big[\sqrt{n}\big(J^\star(P) - J^{\widehat{\delta}_{n,R}}(P)\big)\Big] - \mathfrak{R}_R\big(\mathcal{K}_0(P_0), \mathfrak{G}_{0,\mathfrak{T}}(P_0)\big) \right| \longrightarrow 0.} \tag{522}$$

Thus estimating the center, Fisher normal geometry, support–rank stratum, and tied candidate face has no first-order policy-regret cost on uniformly stable regular strata.

Proof. By Proposition J.6, the oracle value at P_0 equals $V_R(B(P_0))$. We first prove the uniform upper bound.

On $\mathcal{E}_n$, the selected lottery is supported on the true active face A. The uniform corrected-path Bellman expansion, specialized to the smooth stratum and decomposed into tie-tangent and normal coordinates, gives

$$\sqrt{n}\big(J^\star(P) - J^q(P)\big) = L_{B(P_0)}(\zeta_n(P; P_0), q) + o(1) \tag{523}$$

uniformly over the stable class, local alternatives, and $q \in \Delta(A)$.

Condition on the pilot. The estimated configuration and selected rule are then fixed, while the main sample is independent. For every fixed resolution m, Theorem J.19 transfers the risk of the L_m-Lipschitz rule from $\widehat{Z}_n$ to its Gaussian shift uniformly. After the orthogonal frame alignment, let $B_0^{\mathrm{pil}} := O_n(P_0)^\top B(P_0)$. The same theorem gives

$$\eta_n := \max_{\pi \in A} \|\widehat{b}_{\pi,n} - b_{0,\pi}^{\mathrm{pil}}\|_2 \xrightarrow{L^1} 0 \tag{524}$$

uniformly. Lemma J.20 and the fixed-rule part of Theorem J.11 therefore imply

$$\sup_{\|\zeta\|\le R} \mathbb{E}_\zeta L_{B_0^{\mathrm{pil}}}\big(\zeta, \mathsf{D}_m(\widehat{B}_n)(Z)\big) \le V_R(B(P_0)) + \varepsilon_m + 4R\eta_n. \tag{525}$$

The diagonal choice $m = m(n)$ makes the conditional Gaussian-transfer error, $\varepsilon_{m(n)}$, and the gradient error all vanish uniformly. Combining equation 523– equation 525 gives

$$\limsup_n \sup_{P_0\in\mathcal{P}_{\mathrm{stab}}} \left[\sup_{P\in\Theta_{n,R}(P_0)} \mathbb{E}_P\sqrt{n}\big(J^\star(P) - J^{\widehat{\delta}_{n,R}}(P)\big) - V_R(B(P_0))\right] \le 0 \tag{526}$$

after restricting to $\mathcal{E}_n$.

On $\mathcal{E}_n^c$, finite-horizon bounded rewards give $0 \le J^\star(P) - J^\pi(P) \le 2TR_{\max}$ for every candidate policy. Hence the failure contribution is at most

$$2TR_{\max}\sqrt{n}\sup_{P_0,P}\mathbb{P}_P(\mathcal{E}_n^c) \le 2TR_{\max}C\sqrt{n}\,e^{-cn^{1/3}} \longrightarrow 0 \tag{527}$$

by Lemma J.16. This proves the uniform upper bound without conditioning.

The matching uniform lower bound for this rule is a special case of Lemma J.21. Combining upper and lower bounds and using equation 431 proves equation 522. □

Remark J.23 (Exact scope). The theorem is adaptive to the unknown center and to the discrete support–rank–face component, but it is not assumption free. The stable margins, support and rank budgets, finite candidate class, design, and local radius are treated as known. Support/rank-unsaturated points and singular tie intersections are excluded; Appendix K shows that a positive adaptation tax can occur there. No finite-n rate or polynomial-time algorithm for the Gaussian game is asserted.

K An unavoidable adaptation tax at a singular rank intersection

Appendix J proves that one center-independent procedure attains the oracle Bellman–Le Cam value uniformly on quantitative stable support–rank–policy strata. We now prove that the stable-margin restriction is substantive. At a singular intersection of two smooth rank-one branches, an oracle that knows the branch uses a one-dimensional Gaussian comparison coordinate. If the branch label is hidden, the two coordinates cannot in general be recovered without first-order loss. We first establish the exact Gaussian decision lower bound and then realize it by a strictly positive, one-step MDP with a C_2-equivariant rank-one transition model.

The proof uses only finite simple-hypothesis testing and finite-experiment likelihood-ratio convergence. These are standard components of Gaussian limit-experiment theory (Le Cam & Yang, 2000; van der Vaart, 1998); all inequalities specific to the present loss and all MDP calculations are proved below.

K.1 The hidden-direction Gaussian game

Fix $R > 0$, $\alpha > 0$, and unit vectors $u_1, u_2 \in \mathbb{R}^D$ satisfying

$$u_1 \notin \{u_2, -u_2\}. \tag{528}$$

For a branch label $j \in \{1, 2\}$ and scalar parameter $t \in [-R, R]$, observe

$$Z \sim N(tu_j, I_D). \tag{529}$$

The decision is a randomized sign $q \in \Delta(\{-1, +1\})$. Write $q_{-\,\mathrm{sgn}(t)}$ for the mass assigned to the wrong sign and set

$$L_\alpha(t, q) := 2\alpha|t|\, q_{-\,\mathrm{sgn}(t)}, \qquad L_\alpha(0, q) := 0. \tag{530}$$

A Borel rule is a map $\delta : \mathbb{R}^D \to \Delta(\{-1, +1\})$.

If the branch label is revealed, define

$$V_{R,j}^{\mathrm{or}} := \inf_\delta \sup_{|t|\le R} \mathbb{E}_{tu_j} L_\alpha(t, \delta(Z)). \tag{531}$$

If one rule must be used without observing j, define

$$V_R^{\mathrm{ad}} := \inf_{\delta} \max_{j\in\{1,2\}} \sup_{|t|\le R} \mathbb{E}_{tu_j} L_\alpha(t, \delta(Z)). \tag{532}$$

Finally, put

$$c_R := \max_{0\le t\le R} 2t\Phi(-t), \qquad t_R \in \arg\max_{0\le t\le R} 2t\Phi(-t). \tag{533}$$

Since $R > 0$, every maximizer t_R is strictly positive.

Lemma K.1 (Exact oracle line value)**.** *For each $j \in \{1, 2\}$,*

$$\boxed{V_{R,j}^{\mathrm{or}} = \alpha c_R.} \tag{534}$$

A minimax rule chooses $+1$ if and only if $u_j^\top Z \ge 0$.

Proof. Under $N(tu_j, I_D)$, the scalar $Y_j := u_j^\top Z$ is $N(t, 1)$. The sign rule therefore has wrong-sign probability $\Phi(-|t|)$ and risk $2\alpha|t|\Phi(-|t|)$, proving the upper bound αc_R.

For the lower bound, put equal prior mass on $t = \pm t_R$. The two induced laws are $N(\pm t_R u_j, I_D)$. By the Neyman–Pearson lemma, their minimum equal-prior testing error is $\Phi(-t_R)$, attained by the sign of $u_j^\top Z$. Every decision rule consequently has Bayes risk at least $2\alpha t_R\Phi(-t_R) = \alpha c_R$. A maximum risk dominates a Bayes risk, completing the proof. □

For $t > 0$, define the two sign-mixture laws

$$\begin{aligned} Q_{+,t} &:= \frac{1}{2}N(tu_1, I_D) + \frac{1}{2}N(tu_2, I_D), \\ Q_{-,t} &:= \frac{1}{2}N(-tu_1, I_D) + \frac{1}{2}N(-tu_2, I_D), \end{aligned} \tag{535}$$

and their equal-prior Bayes testing error

$$e_{\mathrm{mix}}(t; u_1, u_2) := \frac{1}{2}\left[1 - \mathrm{TV}(Q_{+,t}, Q_{-,t})\right]. \tag{536}$$

Lemma K.2 (Strict contraction after hiding a noncollinear direction)**.** *For every $t > 0$,*

$$\boxed{e_{\mathrm{mix}}(t; u_1, u_2) > \Phi(-t).} \tag{537}$$

More precisely, if

$$d_j(z) := \varphi_D(z - tu_j) - \varphi_D(z + tu_j), \tag{538}$$

then

$$\boxed{e_{\mathrm{mix}}(t; u_1, u_2) - \Phi(-t) = \frac{1}{8}\int_{\mathbb{R}^D} \big(|d_1| + |d_2| - |d_1 + d_2|\big)\, dz > 0.} \tag{539}$$

Proof. For either unit vector u_j,

$$\mathrm{TV}\big(N(tu_j, I_D), N(-tu_j, I_D)\big) = 1 - 2\Phi(-t). \tag{540}$$

Indeed, the likelihood-ratio decision boundary is the half-space $u_j^\top z = 0$. With the convention $\mathrm{TV}(P, Q) = \frac{1}{2}\int |p - q|$, the mixture difference has density $\frac{1}{2}(d_1 + d_2)$, and hence

$$\begin{aligned} &\big[1 - 2\Phi(-t)\big] - \mathrm{TV}(Q_{+,t}, Q_{-,t}) \\ &\quad = \frac{1}{4}\int_{\mathbb{R}^D} \big(|d_1| + |d_2| - |d_1 + d_2|\big)\, dz. \end{aligned} \tag{541}$$

Moreover,

$$\frac{\varphi_D(z - tu_j)}{\varphi_D(z + tu_j)} = \exp(2tu_j^\top z), \tag{542}$$

so $\operatorname{sgn} d_j(z) = \operatorname{sgn}(u_j^\top z)$ away from the hyperplane $u_j^\top z = 0$. Because u_1, u_2 are noncollinear, the open cone

$$\{z : u_1^\top z > 0,\ u_2^\top z < 0\} \tag{543}$$

is nonempty and has positive Lebesgue measure. On this cone, $d_1 d_2 < 0$ and both density differences are nonzero; the scalar triangle inequality is strict. Thus the integral in equation 541 is strictly positive. Dividing that identity by two proves equation 539. □

Theorem K.3 (Exact positive adaptation tax in the Gaussian experiment)**.** *The hidden-branch minimax value satisfies*

$$\boxed{V_R^{\mathrm{ad}} - \max_{j\in\{1,2\}} V_{R,j}^{\mathrm{or}} \geq \Delta_{R,\alpha}(u_1,u_2) > 0,} \tag{544}$$

where

$$\boxed{\Delta_{R,\alpha}(u_1,u_2) := 2\alpha t_R \left[e_{\mathrm{mix}}(t_R;u_1,u_2) - \Phi(-t_R)\right].} \tag{545}$$

Proof. Place prior mass $1/4$ on each of the four points $(j,t) \in \{1,2\}\times\{-t_R,+t_R\}$. Conditional on a positive sign, the observation law is Q_{+,t_R}; conditional on a negative sign, it is Q_{-,t_R}. Therefore every rule has prior-average wrong-sign probability at least $e_{\mathrm{mix}}(t_R;u_1,u_2)$, and hence prior-average loss at least

$$2\alpha t_R e_{\mathrm{mix}}(t_R;u_1,u_2). \tag{546}$$

The maximum of the two branch-wise worst-case risks dominates this Bayes average. Lemma K.1 gives the oracle value $2\alpha t_R\Phi(-t_R) = \alpha c_R$. Subtraction and Lemma K.2 prove the theorem. □

K.2 A FINITE-EXPERIMENT TRANSFER LEMMA

The finite-MDP proof below uses the following elementary consequence of joint likelihood-ratio convergence. It is included to make the lower transfer independent of any unproved full-experiment convergence claim.

Lemma K.4 (Total variation of finite mixtures under likelihood-ratio convergence)**.** *Let $P_{0,n}$ dominate probability measures $P_{1,n},\dots,P_{m,n}$, and put $L_{k,n} := dP_{k,n}/dP_{0,n}$. Suppose that under $P_{0,n}$,*

$$(L_{1,n},\dots,L_{m,n}) \Longrightarrow (L_1,\dots,L_m) \quad \textit{under } Q_0, \tag{547}$$

and

$$\sup_n \max_{k\leq m} \mathbb{E}_{P_{0,n}} L_{k,n}^2 < \infty. \tag{548}$$

Assume $L_k \geq 0$ and $\mathbb{E}_{Q_0} L_k = 1$. For any probability vectors $\lambda,\mu \in \Delta_m$, define

$$P_{\lambda,n} := \sum_{k=1}^m \lambda_k P_{k,n}, \qquad Q_\lambda := \sum_{k=1}^m \lambda_k Q_k, \qquad dQ_k := L_k\, dQ_0. \tag{549}$$

Then

$$\boxed{\mathrm{TV}(P_{\lambda,n},P_{\mu,n}) \longrightarrow \mathrm{TV}(Q_\lambda,Q_\mu).} \tag{550}$$

Proof. The mixture likelihood ratios relative to $P_{0,n}$ are

$$M_{\lambda,n} = \sum_k \lambda_k L_{k,n}, \qquad M_{\mu,n} = \sum_k \mu_k L_{k,n}. \tag{551}$$

Hence

$$\mathrm{TV}(P_{\lambda,n},P_{\mu,n}) = \frac{1}{2}\mathbb{E}_{P_{0,n}}|M_{\lambda,n} - M_{\mu,n}|. \tag{552}$$

Joint convergence of the likelihood-ratio vector implies convergence in law of the absolute difference. Moreover, by Cauchy–Schwarz and finiteness of m,

$$\begin{aligned}\mathbb{E}_{P_{0,n}}|M_{\lambda,n} - M_{\mu,n}|^2 &\leq \left(\sum_k |\lambda_k - \mu_k|\right)\sum_k |\lambda_k - \mu_k|\mathbb{E}L_{k,n}^2 \\ &\leq 4\max_{k\leq m}\mathbb{E}L_{k,n}^2.\end{aligned} \tag{553}$$

Thus the absolute differences are uniformly integrable, and their expectations converge. The limit equals $\frac{1}{2}\mathbb{E}_{Q_0}|\sum_k(\lambda_k-\mu_k)L_k|$, which is precisely $\mathrm{TV}(Q_\lambda,Q_\mu)$. □

Lemma K.5 (Uniform threshold probabilities from bounded–Lipschitz convergence)**.** *Let $S_{n,t}$ be real random variables indexed by $t \in [-R, R]$. If*

$$\varepsilon_n := \sup_{|t|\le R} d_{\mathrm{BL}}\big(\mathcal{L}(S_{n,t}), N(t,1)\big) \longrightarrow 0, \tag{554}$$

then

$$\boxed{\sup_{|t|\le R} |\mathbb{P}(S_{n,t} \le 0) - \Phi(-t)| \longrightarrow 0.} \tag{555}$$

Proof. For $\delta \in (0,1]$, let ψ_δ equal one on $(-\infty, 0]$, decrease linearly to zero on $[0, \delta]$, and vanish on $[\delta, \infty)$. Its envelope is one and its Lipschitz constant is $1/\delta$. After rescaling in the definition of d_{BL},

$$|\mathbb{E}\psi_\delta(S_{n,t}) - \mathbb{E}\psi_\delta(G_t)| \le \frac{\varepsilon_n}{\delta}, \qquad G_t \sim N(t,1). \tag{556}$$

The Gaussian probability of any interval of length δ is at most $\delta/\sqrt{2\pi}$. Applying the same construction from below yields

$$\sup_{|t|\le R} |\mathbb{P}(S_{n,t} \le 0) - \Phi(-t)| \le \frac{\varepsilon_n}{\delta} + \frac{\delta}{\sqrt{2\pi}}. \tag{557}$$

Take $\delta = \min\{1, \sqrt{\varepsilon_n}\}$. □

K.3 A ONE-STEP RANK-ONE MDP REALIZATION

Let $G = C_2 = \{e, \tau\}$, let

$$\mathcal{S} = [2] \times \{-1, +1\}, \qquad \mathcal{A} = \{0, 1\}, \qquad T = 2, \tag{558}$$

and let τ flip the sign coordinate and fix the action. Use the uniform initial law, transition kernel, and generative design

$$\mu(s) = \frac{1}{4}, \qquad P_0(s' \mid x) = \frac{1}{4}, \qquad w_x = \frac{1}{8}. \tag{559}$$

The candidate class is $\Pi = \{\pi^+, \pi^-\}$, where

$$\begin{aligned} \pi^+(i,\sigma) &:= \mathbf{1}\{i = 1,\ \sigma = +1\}, \\ \pi^-(i,\sigma) &:= \mathbf{1}\{i = 1,\ \sigma = -1\}, \end{aligned} \tag{560}$$

and both policies choose action zero elsewhere. Let $r_1 \equiv 0$ and

$$r_2(i,\sigma) := \sigma \mathbf{1}\{i = 1\}. \tag{561}$$

The transition class consists of stochastic C_2-equivariant kernels whose sign multiplicity block has rank at most one.

Use the orthonormal sign multiplicity bases (p_1, p_2, q_1, q_2) and $(v_1^{\mathrm{sgn}}, v_2^{\mathrm{sgn}})$ from Appendix I. Let E_{ab} denote the sign-block matrix unit in the $(p_a, v_b^{\mathrm{sgn}})$ coordinates.

Lemma K.6 (Exact candidate comparison functional)**.** *At the uniform transition experiment,*

$$B_{\mathrm{sgn}}(H^+ - H^-) = \frac{1}{\sqrt{2}} E_{11}. \tag{562}$$

Because the sign-block Fisher norm is one half of the raw Frobenius norm, the centered Fisher-whitened comparison gradient is

$$\boxed{g = \frac{1}{2} E_{11}.} \tag{563}$$

If a kernel has whitened sign block Θ, then its exact candidate value gap is

$$\boxed{J^{\pi^+}(P) - J^{\pi^-}(P) = \langle E_{11}, \Theta \rangle_{\mathrm{F}}.} \tag{564}$$

Proof. At the only decision stage, the state law is uniform. Directly from equation 560,

$$d^{\pi^+}(i,\sigma,a) - d^{\pi^-}(i,\sigma,a) = \frac{1}{4}\mathbf{1}\{i=1\}\sigma(2a-1). \tag{565}$$

The left side equals $\frac{1}{2}p_1$ in the sign row basis, while the terminal reward equals $\sqrt{2}v_1^{\mathrm{sgn}}$. Their outer product is therefore $E_{11}/\sqrt{2}$, proving equation 562. Under equation 559, the Fisher norm of a raw sign block H is $\|H\|_{\mathrm{F}}^2/2$; hence the whitened coordinate is $\Theta = H/\sqrt{2}$. Pairing the raw gradient difference with $H = \sqrt{2}\Theta$ gives equation 564, and halving the gradient difference gives equation 563. Since $T = 2$, policy values are affine in the transition kernel, so the displayed relation is exact rather than a first-order approximation. □

Fix $\vartheta \in (0, \pi/2)$ and set

$$q := \cos\vartheta, \qquad s := \sin\vartheta, \qquad v := qe_1 + se_2. \tag{566}$$

Embed the rank-one matrices

$$U_1 := e_1 v^\top, \qquad U_2 := v e_1^\top \tag{567}$$

in the upper-left 2×2 part of the 4×2 sign block. Then

$$\boxed{\begin{array}{c}\mathrm{rank}(U_j) = 1, \qquad \|U_j\|_{\mathrm{F}} = 1, \qquad \langle E_{11}, U_j\rangle_{\mathrm{F}} = q, \\ \langle U_1, U_2\rangle_{\mathrm{F}} = q^2 \in (0,1).\end{array}} \tag{568}$$

Thus U_1, U_2 are noncollinear unit directions.

For $j \in \{1,2\}$ and $|t| \le R$, define $P_{j,n,t}$ by assigning whitened sign block

$$\Theta_{j,n,t} := \frac{t}{\sqrt{n}} U_j, \tag{569}$$

setting the normalized trivial nuisance block to zero, and reconstructing the kernel in the affine equivariant chart. Its raw sign block is $\sqrt{2}\Theta_{j,n,t}$. Since maximum entry norm is dominated by Hilbert–Schmidt norm, for

$$n > 32R^2 \tag{570}$$

every such kernel is strictly positive. It is stochastic, C_2-equivariant, and has sign-block rank at most one. Both branches pass through the common rank-zero center P_0 at $t = 0$.

For a data rule δ_n taking values in $\Delta(\{\pi^+, \pi^-\})$, define the branch-oracle and hidden-branch risks

$$\mathcal{R}_{n,R}^{\mathrm{or},j} := \inf_{\delta_n} \sup_{|t|\le R} \mathbb{E}_{P_{j,n,t}} \left[\sqrt{n}\big(J^\star(P_{j,n,t}) - J^{\delta_n}(P_{j,n,t})\big)\right], \tag{571}$$

$$\mathcal{R}_{n,R}^{\mathrm{ad}} := \inf_{\delta_n} \max_{j\in\{1,2\}} \sup_{|t|\le R} \mathbb{E}_{P_{j,n,t}} \left[\sqrt{n}\big(J^\star(P_{j,n,t}) - J^{\delta_n}(P_{j,n,t})\big)\right]. \tag{572}$$

Here the oracle in equation 571 is told which one-dimensional branch model is in force; the adaptive rule in equation 572 is not.

Lemma K.7 (Exact finite-sample decision loss)**.** *For every j, n, t,*

$$\sqrt{n}\left[J^{\pi^+}(P_{j,n,t}) - J^{\pi^-}(P_{j,n,t})\right] = qt. \tag{573}$$

Consequently, the scaled regret of choosing the wrong candidate equals $q|t|$ exactly.

Proof. Apply equation 564, equation 569, and $\langle E_{11}, U_j\rangle = q$. □

K.4 Exact asymptotic branch-oracle values

Let Z_n be the observable Fisher-whitened central sequence at P_0, and set

$$S_{j,n} := \langle U_j, Z_n\rangle_{\mathrm{F}}. \tag{574}$$

Uniform shifted-central-sequence approximation gives

$$\sup_{|t|\le R} d_{\mathrm{BL}}\left(\mathcal{L}_{P_{j,n,t}}(S_{j,n}), N(t,1)\right) \longrightarrow 0. \tag{575}$$

Proposition K.8 (Exact branch-oracle limit)**.** *For $j = 1, 2$,*

$$\boxed{\lim_{n\to\infty} \mathcal{R}^{\mathrm{or},j}_{n,R} = \frac{q}{2} c_R.} \tag{576}$$

Proof. For the upper bound, use the rule choosing π^+ iff $S_{j,n} \geq 0$. Lemma K.5 and equation 575 imply that its worst-case risk converges to

$$\sup_{|t|\leq R} q|t|\Phi(-|t|) = \frac{q}{2} c_R. \tag{577}$$

For the lower bound, restrict to $t = \pm t_R$. Relative to the common base law, the two likelihood ratios converge jointly to those of $N(\pm t_R U_j, I)$. The finite-path second-moment bound from Lemma E.8 and Lemma K.4 therefore give

$$\mathrm{TV}\left(\mathbb{P}^{(n)}_{P_{j,n,t_R}}, \mathbb{P}^{(n)}_{P_{j,n,-t_R}}\right) \longrightarrow 1 - 2\Phi(-t_R). \tag{578}$$

Under the equal prior on these two points, every rule has asymptotic wrong-sign probability at least $\Phi(-t_R)$. By Lemma K.7, its maximum risk is asymptotically at least $qt_R\Phi(-t_R) = qc_R/2$. □

K.5 The finite-MDP adaptation lower bound

Theorem K.9 (Positive adaptation tax at the rank-zero branch intersection)**.** *For every $R > 0$ and every $\vartheta \in (0, \pi/2)$,*

$$\boxed{\liminf_{n\to\infty} \left[\mathcal{R}^{\mathrm{ad}}_{n,R} - \max_{j\in\{1,2\}} \mathcal{R}^{\mathrm{or},j}_{n,R}\right] \geq \Delta_{\vartheta,R} > 0,} \tag{579}$$

where

$$\boxed{\Delta_{\vartheta,R} := qt_R\left[e_{\mathrm{mix}}(t_R; U_1, U_2) - \Phi(-t_R)\right], \qquad q = \cos\vartheta.} \tag{580}$$

Thus no rule that is ignorant of the branch model can attain both branch-specific oracle values at their common singular intersection.

Proof. Restrict the adaptive parameter set to the four alternatives $(j, t) \in \{1, 2\} \times \{-t_R, +t_R\}$ and put the uniform prior on them. Let

$$Q_{+,n} := \frac{1}{2}\mathbb{P}^{(n)}_{P_{1,n,t_R}} + \frac{1}{2}\mathbb{P}^{(n)}_{P_{2,n,t_R}}, \tag{581}$$

$$Q_{-,n} := \frac{1}{2}\mathbb{P}^{(n)}_{P_{1,n,-t_R}} + \frac{1}{2}\mathbb{P}^{(n)}_{P_{2,n,-t_R}}. \tag{582}$$

For every rule, its maximum risk dominates its prior-average risk, which by Lemma K.7 is at least

$$qt_R\,\frac{1}{2}\left[1 - \mathrm{TV}(Q_{+,n}, Q_{-,n})\right]. \tag{583}$$

The four likelihood ratios relative to the common base experiment converge jointly to

$$\exp\left(\pm t_R\langle U_j, Z\rangle_{\mathrm{F}} - \frac{1}{2}t_R^2\right), \qquad Z \sim N(0, I), \tag{584}$$

and have uniformly bounded second moments by Lemma E.8. Therefore Lemma K.4 yields

$$\frac{1}{2}\left[1 - \mathrm{TV}(Q_{+,n}, Q_{-,n})\right] \longrightarrow e_{\mathrm{mix}}(t_R; U_1, U_2). \tag{585}$$

Hence

$$\liminf_n \mathcal{R}^{\mathrm{ad}}_{n,R} \geq qt_R e_{\mathrm{mix}}(t_R; U_1, U_2). \tag{586}$$

Subtract Proposition K.8 and use $c_R = 2t_R\Phi(-t_R)$ to obtain equation 579. Strict positivity follows from $\langle U_1, U_2\rangle = q^2 \in (0, 1)$ and Lemma K.2. □

K.6 SMOOTH RANK-ONE CENTERS COLLAPSING INTO THE INTERSECTION

The preceding theorem uses two one-dimensional branch models meeting at the rank-zero center. The same tax persists if each branch is recentered at a rank-one smooth point whose smallest nonzero singular value collapses faster than the root-n scale.

Define

$$X_1 := e_2 v^\top, \qquad X_2 := v e_2^\top. \tag{587}$$

Then

$$\langle E_{11}, X_j \rangle_{\mathrm{F}} = 0, \qquad X_1 + aU_1 = (e_2 + ae_1)v^\top, \qquad X_2 + aU_2 = v(e_2 + ae_1)^\top. \tag{588}$$

Let $\epsilon_n > 0$ satisfy

$$\epsilon_n \downarrow 0, \qquad \sqrt{n}\,\epsilon_n \longrightarrow 0, \tag{589}$$

and define $P^{\epsilon}_{j,n,t}$ by the whitened sign block

$$\Theta^{\epsilon}_{j,n,t} := \epsilon_n X_j + \frac{t}{\sqrt{n}} U_j. \tag{590}$$

Every such block has exact rank one, and $P^{\epsilon}_{j,n,0}$ is an exact candidate tie because $\langle E_{11}, X_j \rangle = 0$.

Corollary K.10 (The tax persists along collapsing smooth centers)**.** *Let $\mathcal{R}^{\mathrm{or},j,\epsilon}_{n,R}$ and $\mathcal{R}^{\mathrm{ad},\epsilon}_{n,R}$ be defined as in equation 571–equation 572, with $P_{j,n,t}$ replaced by $P^{\epsilon}_{j,n,t}$. Then*

$$\mathcal{R}^{\mathrm{or},j,\epsilon}_{n,R} - \mathcal{R}^{\mathrm{or},j}_{n,R} \longrightarrow 0, \qquad j = 1, 2, \tag{591}$$

and

$$\mathcal{R}^{\mathrm{ad},\epsilon}_{n,R} - \mathcal{R}^{\mathrm{ad}}_{n,R} \longrightarrow 0. \tag{592}$$

Consequently, the lower bound equation 579 remains valid with all risks replaced by the moving-center risks.

Proof. The exact scaled candidate gap is unchanged because $\langle E_{11}, X_j \rangle = 0$. For all large n, all branch probabilities are bounded below by a common positive constant. The two kernels $P^{\epsilon}_{j,n,t}$ and $P_{j,n,t}$ differ by the fixed perturbation with whitened block $\epsilon_n X_j$. The elementary categorical bound

$$D_{\mathrm{KL}}(p \| q) \le \sum_s \frac{(p_s - q_s)^2}{q_s} \tag{593}$$

therefore gives, uniformly in j and $|t| \le R$,

$$D_{\mathrm{KL}}\left(\mathbb{P}^{(n)}_{P^{\epsilon}_{j,n,t}} \middle\| \mathbb{P}^{(n)}_{P_{j,n,t}} \right) \le C n \epsilon_n^2 \longrightarrow 0. \tag{594}$$

Pinsker's inequality yields uniform total-variation convergence. Since the scaled regret is bounded by qR, the risk of every rule changes by $o(1)$, uniformly over the corresponding branch or union. Taking infima and suprema proves equation 591– equation 592. □

Remark K.11 (Interpretation and exact scope). The theorem is an adaptation lower bound between two smooth one-dimensional rank-one branch models. It does not assert that the branch-oracle value equals the oracle value for the entire rank-≤ 1 local model. Its conclusion is that no single rule can simultaneously attain the two smaller, branch-specific oracle benchmarks when the branch label is hidden at their singular intersection. The moving-center corollary shows that this obstruction persists for smooth rank-one centers whose rank margin vanishes faster than $n^{-1/2}$.

L FULL-POLICY REALIZATION BY A COMMITMENT LIFT

The realization in Appendix I uses a prescribed binary candidate class. This appendix removes that restriction. Starting from the same concentrated and diffuse comparison-gradient gadgets, we add a strictly positive commitment mechanism and bounded consistency penalties. In the resulting finite-horizon controlled process, the complete deterministic nonstationary Markov policy class has exactly two optimal policies, every other deterministic policy has a common positive base gap, and the Fisher-whitened comparison gradients retain the concentrated/diffuse singular-value profiles up to one common positive scalar. Hence the two-sided $\Theta_R(\sqrt{d/r})$ limiting-risk separation is intrinsic to full MDP policy optimization rather than to an externally supplied two-policy menu.

The transition action below is again the joint state–action symmetry used in Appendix I; related uses of group actions on joint state–action spaces appear in symmetry-aware reinforcement learning (van der Pol et al., 2020). All claims specific to the present commitment construction are proved directly.

L.1 COMMITMENT-LIFTED CONTROLLED PROCESS

Fix $d \geq r \geq 1$ and retain the signal state space $\mathcal{S}_d = [d] \times \{-1, +1\}$, action set $\mathcal{A} = \{0, 1\}$, signal horizon $N = d^2$, stage map ϑ, policies $\pi_k^{\pm}$, and mean-zero signal functions ϱ_k^{λ} from Appendix I. Write

$$m := |\mathcal{S}_d| = 2d, \qquad \mathsf{B} := \{-1, +1\}. \tag{595}$$

Introduce one commitment state $\star$ and define the augmented state space

$$\widetilde{\mathcal{S}}_d := \{\star\} \sqcup (\mathsf{B} \times \mathcal{S}_d), \qquad L := |\widetilde{\mathcal{S}}_d| = 1 + 2m = 1 + 4d. \tag{596}$$

The horizon is

$$\widetilde{T} := N + 2. \tag{597}$$

The nontrivial element of C_2 fixes $\star$, fixes the commitment mode $\omega \in \mathsf{B}$, and flips the sign coordinate of the signal state:

$$\tau\star = \star, \qquad \tau(\omega, (i, \sigma)) = (\omega, (i, -\sigma)), \qquad \tau a = a. \tag{598}$$

Put

$$\chi := \frac{1}{2}, \qquad \xi := \frac{1}{8(N+1)}, \qquad \rho := (1 - \xi)^2. \tag{599}$$

Let μ_d be the uniform law on $\mathcal{S}_d$. The strictly positive initial law is

$$\widetilde{\mu}(\star) = \chi, \qquad \widetilde{\mu}(\omega, s) = \frac{1 - \chi}{2m} \quad (\omega \in \mathsf{B},\ s \in \mathcal{S}_d). \tag{600}$$

For $a \in \{0, 1\}$, write

$$\omega(a) := 2a - 1 \in \mathsf{B}. \tag{601}$$

Define the commitment distribution on signal states by

$$C_a(\omega, s) := \left[(1 - \xi)\mathbf{1}\{\omega = \omega(a)\} + \frac{\xi}{2}\right] \mu_d(s). \tag{602}$$

For a stochastic kernel Q on $\mathcal{S}_d$, define the signal transition component

$$D_Q((\omega', s') \mid (\omega, s), a) := \frac{\xi}{2}\mu_d(s') + (1 - \xi)\mathbf{1}\{\omega' = \omega\}Q(s' \mid s, a). \tag{603}$$

The commitment-lifted transition kernel $\mathsf{L}_d(Q)$ on $\widetilde{\mathcal{S}}_d$ is

$$\mathsf{L}_d(Q)(z' \mid \star, a) := \frac{\xi}{L} + (1 - \xi)C_a(z')\mathbf{1}\{z' \neq \star\}, \tag{604}$$

$$\mathsf{L}_d(Q)(z' \mid (\omega, s), a) := \frac{\xi}{L} + (1 - \xi)D_Q(z' \mid (\omega, s), a)\mathbf{1}\{z' \neq \star\}. \tag{605}$$

Here $C_a(\star) = 0$ and $D_Q(\star \mid \cdot) = 0$ by convention. Every row sums to one and every transition probability is at least ξ/L.

Let Q_0 be the uniform signal kernel,

$$Q_0(s' \mid s, a) = \frac{1}{m}. \tag{606}$$

Let $\Theta_{d,r}$ be the C_2-equivariant rank-r signal-kernel class in equation 372, and define its commitment lift

$$\widetilde{\Theta}^{\mathrm{com}}_{d,r} := \{\mathsf{L}_d(Q) : Q \in \Theta_{d,r}\}. \tag{607}$$

The base transition is

$$\widetilde{P}_0 := \mathsf{L}_d(Q_0). \tag{608}$$

Lemma L.1 (Stochasticity, positivity, and equivariance)**.** *For every $Q \in \Theta_{d,r}$, the kernel $\mathsf{L}_d(Q)$ is stochastic, strictly positive, and C_2-equivariant. Moreover,*

$$\min_{z,a,z'} \mathsf{L}_d(Q)(z' \mid z, a) \geq \frac{\xi}{L}. \tag{609}$$

The map L_d is affine and injective.

Proof. The uniform term in every row has total mass ξ. The distribution C_a sums to one over $\mathsf{B} \times \mathcal{S}_d$, and so does D_Q: its first term has total mass ξ and its second term has total mass $1 - \xi$. This proves row stochasticity and equation 609. The uniform law, μ_d, and the event $\{\omega' = \omega\}$ are invariant under the sign flip. Hence C_a is invariant and D_Q is equivariant whenever Q is. The coefficient multiplying Q in equation 605 is $(1 - \xi)^2 > 0$, so the map is affine and injective. □

For the generative experiment, every row $(z, a) \in \widetilde{\mathcal{S}}_d \times \mathcal{A}$ receives limiting weight

$$\widetilde{w}_{z,a} := \frac{1}{2L}. \tag{610}$$

Samples from the commitment rows are ancillary because those rows do not depend on Q; retaining them makes every row weight strictly positive.

L.2 Rewards and the complete deterministic policy class

For $\omega \in \mathsf{B}$, identify

$$\pi^{\omega}_k := \begin{cases} \pi^{+}_k, & \omega = +1, \\ \pi^{-}_k, & \omega = -1. \end{cases} \tag{611}$$

Define the signal attenuation coefficients

$$c_k := \frac{1}{2}\rho^{N-k}, \qquad k \in [N]. \tag{612}$$

They satisfy $0 < c_k \leq 1/2$.

The stage-one reward is

$$\begin{aligned} \widetilde{r}^{\lambda}_1(\star, a) &:= 0, \\ \widetilde{r}^{\lambda}_1((\omega, s), a) &:= -\frac{1}{2}\mathbf{1}\{a \neq 0\}. \end{aligned} \tag{613}$$

For $k \in [N]$, at stage $t = k + 1$ set

$$\begin{aligned} \widetilde{r}^{\lambda}_{k+1}(\star, a) &:= -\frac{1}{2}\mathbf{1}\{a \neq 0\}, \\ \widetilde{r}^{\lambda}_{k+1}((\omega, s), a) &:= \mathbf{1}\{k \geq 2\} c_{k-1} \varrho^{\lambda}_{k-1}(s) - \frac{1}{2}\mathbf{1}\{a \neq \pi^{\omega}_k(s)\}. \end{aligned} \tag{614}$$

The terminal reward is

$$\widetilde{r}^{\lambda}_{\widetilde{T}}(\star) := 0, \qquad \widetilde{r}^{\lambda}_{\widetilde{T}}(\omega, s) := c_N \varrho^{\lambda}_N(s). \tag{615}$$

Since $\|\varrho^{\lambda}_k\|_\infty \leq 1$ and $c_k \leq 1/2$,

$$\max_{t < \widetilde{T}} \|\widetilde{r}^{\lambda}_t\|_\infty \vee \|\widetilde{r}^{\lambda}_{\widetilde{T}}\|_\infty \leq 1. \tag{616}$$

Let $\widetilde{\Pi}_d^{\mathrm{det}}$ denote the complete class of deterministic nonstationary Markov policies on $(\widetilde{\mathcal{S}}_d, \mathcal{A}, \widetilde{T})$. Define two policies $\widetilde{\pi}^+, \widetilde{\pi}^- \in \widetilde{\Pi}_d^{\mathrm{det}}$ by

$$\widetilde{\pi}_1^{\varsigma}(\star) := \mathbf{1}\{\varsigma = +\}, \qquad \varsigma \in \{+, -\}, \tag{617}$$

$$\widetilde{\pi}_1^{\varsigma}(\omega, s) := 0, \tag{618}$$

$$\widetilde{\pi}_{k+1}^{\varsigma}(\star) := 0, \qquad k \in [N], \tag{619}$$

$$\widetilde{\pi}_{k+1}^{\varsigma}(\omega, s) := \pi_k^{\omega}(s), \qquad k \in [N]. \tag{620}$$

The two policies differ only at the initial commitment state.

Lemma L.2 (Every signal has zero one-step mean at the base)**.** *Let $f : \mathcal{S}_d \to \mathbb{R}$ satisfy $\sum_s f(s) = 0$, and extend it to $\widetilde{f}$ on $\widetilde{\mathcal{S}}_d$ by*

$$\widetilde{f}(\star) = 0, \qquad \widetilde{f}(\omega, s) = f(s). \tag{621}$$

Then, for every $z \in \widetilde{\mathcal{S}}_d$ and $a \in \mathcal{A}$,

$$\sum_{z'} \widetilde{P}_0(z' \mid z, a)\widetilde{f}(z') = 0. \tag{622}$$

Consequently, under $\widetilde{P}_0$, the expected total contribution of all ϱ_k^{λ} rewards is zero for every deterministic or randomized policy.

Proof. The common uniform component has zero expectation because the two mode copies contribute $2\sum_s f(s) = 0$. In a commitment row, the remaining component has uniform signal coordinate s, so its expectation is zero. In a signal row, the internal mode-mixing component and the Q_0 component also have uniform signal coordinate. This proves equation 622. Every signal reward after stage one is evaluated at a state generated by one such row, so iterated expectation proves the final assertion. □

Define

$$\nu_d := \min\left\{\frac{1}{4m}, \frac{\xi}{L}\right\}, \qquad \Delta_d^{\mathrm{full}} := \frac{\nu_d}{2} > 0. \tag{623}$$

Theorem L.3 (Exactly two globally optimal deterministic policies)**.** *For every $\lambda \in [-1, 1]^d$, at the base transition $\widetilde{P}_0$,*

$$\boxed{\arg\max_{\pi \in \widetilde{\Pi}_d^{\mathrm{det}}} J^{\pi}(\widetilde{P}_0) = \{\widetilde{\pi}^+, \widetilde{\pi}^-\}.} \tag{624}$$

Both optimal policies have value zero. Every other deterministic policy satisfies

$$\boxed{J^{\widetilde{\pi}^+}(\widetilde{P}_0) - J^{\pi}(\widetilde{P}_0) = J^{\widetilde{\pi}^-}(\widetilde{P}_0) - J^{\pi}(\widetilde{P}_0) \geq \Delta_d^{\mathrm{full}}.} \tag{625}$$

In fact, the complete base-policy value table is independent of λ.

Proof. By Lemma L.2, every policy has zero expected signal reward. Its value is therefore exactly minus one half of its expected number of consistency violations. At stage one and state $\star$, both actions are allowed. At every other state–time pair there is one prescribed action: action zero at stage-one signal states and at later visits to $\star$, and $\pi_k^{\omega}(s)$ at signal state (ω, s) and stage $k + 1$. The two policies in equation 617–equation 620 incur no violation and hence have value zero.

Any other deterministic policy violates at least one prescribed action. At stage one, every signal state has probability $(1 - \chi)/(2m) = 1/(4m)$. At every later stage, every state has probability at least ξ/L, because every transition entry is at least ξ/L. A violation therefore incurs expected penalty at least $\nu_d/2 = \Delta_d^{\mathrm{full}}$. Finally, a policy with no violation may choose either action at $(t, z) = (1, \star)$ and is otherwise forced to coincide with the displayed canonical policy. Hence there are exactly two such policies. The value formula contains no λ-dependent term, proving the last assertion. □

L.3 Exact occupancies and continuation values

For $k \in [N]$, define

$$\delta_k := \chi \rho^k. \tag{626}$$

Lemma L.4 (Canonical continuation values)**.** *For either $\varsigma \in \{+, -\}$ and every $k \in [N]$,*

$$\begin{aligned} V_{k+2}^{\widetilde{\pi}^\varsigma, \lambda}(\star) &= 0, \\ V_{k+2}^{\widetilde{\pi}^\varsigma, \lambda}(\omega, s) &= c_k \varrho_k^\lambda(s). \end{aligned} \tag{627}$$

Moreover, $V_1^{\widetilde{\pi}^\varsigma, \lambda} = V_2^{\widetilde{\pi}^\varsigma, \lambda} \equiv 0$ on the states relevant at those stages.

Proof. The terminal statement is exactly equation 615. Suppose equation 627 holds at the next stage. Along either canonical policy every consistency penalty is zero. The conditional expectation of the next signal is zero by Lemma L.2. Hence the Bellman recursion leaves only the current state signal. Backward induction proves the claim. □

Lemma L.5 (Exact canonical occupancy difference)**.** *At signal stage k, namely time $t = k + 1$, the state-distribution difference between the two canonical policies is*

$$\boxed{\begin{aligned} \Pr_{\widetilde{\pi}^+}(S_{k+1} = \star) - \Pr_{\widetilde{\pi}^-}(S_{k+1} = \star) &= 0, \\ \Pr_{\widetilde{\pi}^+}(S_{k+1} = (\omega, s)) - \Pr_{\widetilde{\pi}^-}(S_{k+1} = (\omega, s)) &= \frac{\omega \delta_k}{m}. \end{aligned}} \tag{628}$$

Consequently, the state–action occupancy difference is

$$\boxed{\widetilde{d}_{k+1}^+(\omega, s, a) - \widetilde{d}_{k+1}^-(\omega, s, a) = \frac{\omega \delta_k}{m} \mathbf{1}\{a = \pi_k^\omega(s)\}.} \tag{629}$$

Proof. At time one the policies differ only at $\star$, whose mass is χ. The common uniform smoothing cancels in the difference, while the two commitment distributions differ by $(1 - \xi)\omega/m$. The outer coefficient $(1 - \xi)$ in equation 604 gives equation 628 for $k = 1$, with $\delta_1 = \chi(1 - \xi)^2 = \chi\rho$.

Assume the formula at stage k. The policies have identical actions after the initial commitment. The total signed mass is zero, so the common uniform term and the internal mode-uniform term cancel. The mode-preserving component multiplies the sign-mode difference by $(1 - \xi)^2 = \rho$. It creates no mass at $\star$. This proves the induction. Multiplication by the common deterministic action rule proves equation 629. □

L.4 Preservation of the structured cone and comparison gradient

For a row-sum-zero signal-kernel direction u, differentiation of L_d at Q_0 gives

$$D\mathsf{L}_d(Q_0)[u](z' \mid \star, a) = 0, \tag{630}$$

while, for signal rows,

$$\boxed{D\mathsf{L}_d(Q_0)[u]((\omega', s') \mid (\omega, s), a) = \rho \mathbf{1}\{\omega' = \omega\} u(s' \mid s, a),} \tag{631}$$

and the derivative toward $\star$ is zero.

Define

$$\alpha_d := \frac{\chi}{2} \rho^{N+1} = \frac{1}{4}(1 - \xi)^{2N+2}. \tag{632}$$

Theorem L.6 (No unwanted first-order gradient and exact profile transfer)**.** *Let H^λ be the raw comparison-gradient difference in Theorem I.5. For every admissible signal kernel direction u,*

$$\boxed{D\Big(J^{\widetilde{\pi}^+} - J^{\widetilde{\pi}^-}\Big)_{\widetilde{P}_0} [D\mathsf{L}_d(Q_0)[u]] = \alpha_d \left\langle H^\lambda, u \right\rangle_{\mathrm{HS}}.} \tag{633}$$

Thus the commitment rows and consistency penalties create no additional transition-gradient component. In pullback signal-kernel coordinates, the raw comparison gradient is exactly $\alpha_d H^\lambda$.

Proof. The unknown parameter occurs only in signal rows, so the commitment stage has zero transition derivative. Along the canonical policies every consistency penalty vanishes, and Lemma L.4 gives the continuation signal after transition stage k. Using equation 629 and equation 631, the stage-k contribution is

$$\frac{\delta_k \rho c_k}{m} \sum_{\omega \in \mathsf{B}} \sum_{s,a} \omega \mathbf{1}\{a = \pi_k^\omega(s)\} \sum_{s'} u(s' \mid s, a) \varrho_k^\lambda(s').$$

At $k = \vartheta(j, \ell)$,

$$\sum_{\omega \in \mathsf{B}} \omega \mathbf{1}\{a = \pi_k^\omega(i, \sigma)\} = \mathbf{1}\{i = j\} \sigma(2a - 1), \tag{634}$$

which is exactly the action factor in the original occupancy difference equation 380. Moreover,

$$\delta_k \rho c_k = \chi \rho^k \rho \frac{1}{2} \rho^{N-k} = \alpha_d \tag{635}$$

for every k. Summing over all stages therefore multiplies the original raw gradient difference by the common scalar α_d. □

Let

$$p_{\mathrm{same}} := \frac{\xi}{L} + (1 - \xi) \frac{1 - \xi/2}{m}. \tag{636}$$

Under the balanced design equation 610, define

$$\kappa_d := \frac{2\rho^2}{L p_{\mathrm{same}}}, \qquad \beta_d := \frac{\alpha_d}{\sqrt{\kappa_d}} > 0. \tag{637}$$

Proposition L.7 (Exact pullback Fisher scaling)**.** *For any two normalized equivariant signal-kernel directions u, v,*

$$\boxed{\langle D\mathsf{L}_d(Q_0)[u], D\mathsf{L}_d(Q_0)[v] \rangle_{\widetilde{\mathcal{I}}_0} = \kappa_d \langle u, v \rangle_{\mathcal{I}_{0,w}}.} \tag{638}$$

Consequently, the Fisher-whitened contingent cone of the lifted family is orthogonally congruent to the original cone in Appendix I, and the centered whitened comparison gradient is $\beta_d g^\lambda$. Its sign block is

$$\boxed{B_{\mathrm{sgn}}(\widetilde{g}^\lambda) = \beta_d \begin{bmatrix} \mathrm{diag}(\lambda) \\ 0_{d \times d} \end{bmatrix}.} \tag{639}$$

Proof. Only same-mode signal successors contribute to the derivative. Their base probability is p_{same}, and the derivative coefficient is ρ. There are two mode copies of every original state–action row. Therefore

$$\begin{aligned} &\langle D\mathsf{L}_d(Q_0)[u], D\mathsf{L}_d(Q_0)[v] \rangle_{\widetilde{\mathcal{I}}_0} \\ &\quad = \sum_{x \in \mathcal{X}_d} \frac{2}{2L} \sum_{s' \in \mathcal{S}_d} \frac{\rho^2 u(s' \mid x) v(s' \mid x)}{p_{\mathrm{same}}}. \end{aligned}$$

Under the original balanced experiment,

$$\langle u, v \rangle_{\mathcal{I}_{0,w}} = \sum_{x \in \mathcal{X}_d} \frac{1}{2m} \sum_{s'} \frac{u(s' \mid x) v(s' \mid x)}{1/m} = \frac{1}{2} \sum_{x, s'} u(s' \mid x) v(s' \mid x). \tag{640}$$

Comparison proves equation 638.

The affine injection L_d maps the original contingent cone onto the lifted contingent cone. If W_0 is an original Fisher isometry, then $D\mathsf{L}_d(Q_0)[u] \mapsto \sqrt{\kappa_d}\, \mathsf{W}_0 u$ is a lifted Fisher isometry. Because the tangent set is a cone, multiplication by the positive scalar $\sqrt{\kappa_d}$ does not change its Euclidean image as a set. Theorem L.6 and Riesz uniqueness show that the lifted whitened gradient is $\alpha_d / \sqrt{\kappa_d} = \beta_d$ times the original whitened gradient. Equation equation 639 then follows from equation 385. □

Proposition L.8 (Fixed-radius minimax transfer for the commitment lift)**.** *For every fixed d, r, R and every $\lambda \in [-1,1]^d$, let $\widetilde{\mathcal{R}}_{n,R}^{\lambda,\mathrm{full}}$ be the local minimax regret in the commitment-lifted model $\widetilde{\Theta}_{d,r}^{\mathrm{com}}$ when decisions range over the complete deterministic nonstationary policy class. Then*

$$\boxed{\lim_{n\to\infty} \widetilde{\mathcal{R}}_{n,R}^{\lambda,\mathrm{full}} = \widetilde{\mathfrak{R}}_{R}^{\lambda,\mathrm{full}},} \tag{641}$$

where the right side is the Gaussian Bellman game generated by the lifted Fisher-whitened tangent cone and the gradients of the two policies in equation 624.

Proof. The parameterization $Q \mapsto \mathsf{L}_d(Q)$ is affine and injective, so

$$T_{\widetilde{\Theta}_{d,r}^{\mathrm{com}}}(\widetilde{P}_0) = D\mathsf{L}_d(Q_0)\big[T_{\Theta_{d,r}}(Q_0)\big]. \tag{642}$$

Every rank-preserving path and every inverse local-set approximation in the signal-kernel model is carried by the affine map to a path or approximation in the lifted model, with its norm multiplied by the fixed operator norm of $D\mathsf{L}_d(Q_0)$. Proposition L.7 shows that the lifted Fisher metric is one positive scalar multiple of the original metric. Consequently, the bounded local mean sets converge to the lifted whitened cone with the same order as in Appendix C.

The lifted experiment is a finite product of categorical rows with a uniform positive probability margin ξ/L and deterministic positive row weights. The uniform likelihood expansion, shifted-central-sequence approximation, and finite-subexperiment embedding in Appendix E therefore apply directly after replacing the original score by the score of $D\mathsf{L}_d(Q_0)[u]$. The complete deterministic policy class is finite, rewards are bounded by one, and Theorem L.3 gives a positive base gap from the two active policies to every inactive policy. The Bellman expansion and active-face localization in Appendix D thus apply uniformly. Finally, the finite-experiment lower transfer and Lipschitz-rule upper transfer in Appendix F require only these verified ingredients. They prove equation 641. □

L.5 Full-policy concentrated/diffuse separation

Let $\widetilde{M}_{d,r}^{\mathrm{conc}}$ and $\widetilde{M}_{d,r}^{\mathrm{flat}}$ be the commitment-lifted tasks obtained from $\lambda_{d,r}^{\mathrm{conc}}$ and $\lambda_{d}^{\mathrm{flat}}$ in equation 386. Let $\widetilde{\mathcal{R}}_{n,R}^{\diamond,\mathrm{full}}$ be the local minimax regret when the action set of the statistical decision problem is the complete class $\widetilde{\Pi}_d^{\mathrm{det}}$, for $\diamond \in \{\mathrm{conc}, \mathrm{flat}\}$.

Proposition L.9 (Matched full-policy summaries)**.** *For fixed (d, r), the two lifted tasks have the same:*

1. *state and action spaces, horizon $d^2 + 2$, group action, strictly positive initial law, and complete deterministic nonstationary policy class;*
2. *commitment-lifted structured transition model $\widetilde{\Theta}_{d,r}^{\mathrm{com}}$, base transition $\widetilde{P}_0$, balanced design, and Fisher information;*
3. *transition-probability margin ξ/L and reward envelope one;*
4. *complete base-policy value table;*
5. *exactly two globally optimal policies $\{\widetilde{\pi}^+, \widetilde{\pi}^-\}$, with common inactive-policy gap at least Δ_d^{full};*
6. *centered whitened comparison-gradient Frobenius norm β_d.*

They differ only in the singular-value concentration profile of the centered comparison gradient relative to the common rank-constrained signal cone.

Proof. Everything except the task signal ϱ_k^λ is common by construction. Theorem L.3 proves that the complete base-policy value table and the optimal face are independent of λ. Proposition L.7 and equation 388 give Frobenius norm β_d for both centered gradients. Their nonzero singular values are respectively $\beta_d/\sqrt{r}$ with multiplicity r and $\beta_d/\sqrt{d}$ with multiplicity d. □

Theorem L.10 (Full-policy rank-diffusion separation)**.** *Fix $d \geq r \geq 1$ and $R > 0$. In each commitment-lifted task, minimize local regret over the complete deterministic nonstationary Markov policy class $\widetilde{\Pi}_d^{\mathrm{det}}$. Then*

$$\lim_{n\to\infty} \widetilde{\mathcal{R}}_{n,R}^{\diamond,\mathrm{full}} = \widetilde{\mathfrak{R}}_R^{\diamond,\mathrm{full}}, \qquad \diamond \in \{\mathrm{conc}, \mathrm{flat}\}, \tag{643}$$

with

$$\begin{aligned} c_R\beta_d \leq \widetilde{\mathfrak{R}}_R^{\mathrm{conc,full}} \leq R\beta_d, \\ c_R\beta_d\sqrt{r/d} \leq \widetilde{\mathfrak{R}}_R^{\mathrm{flat,full}} \leq R\beta_d\sqrt{r/d}. \end{aligned} \tag{644}$$

Consequently,

$$\boxed{\frac{c_R}{R}\sqrt{\frac{d}{r}} \leq \frac{\lim_{n\to\infty} \widetilde{\mathcal{R}}_{n,R}^{\mathrm{conc,full}}}{\lim_{n\to\infty} \widetilde{\mathcal{R}}_{n,R}^{\mathrm{flat,full}}} \leq \frac{R}{c_R}\sqrt{\frac{d}{r}}.} \tag{645}$$

Thus the rank-diffusion separation persists when the learner selects among all *deterministic nonstationary Markov policies. The fixed-dimensional limit $n \to \infty$ is taken first.*

Proof. The complete deterministic policy class is finite. By Theorem L.3, its base-optimal face consists of exactly $\widetilde{\pi}^+$ and $\widetilde{\pi}^-$, with a positive common gap to all other policies. Hence active-face localization reduces the local first-order decision problem to this binary face.

The affine parameterization $Q \mapsto \mathsf{L}_d(Q)$, together with Proposition L.7, identifies the local Gaussian mean cone with the original product cone from Appendix I up to an orthogonal coordinate change. Proposition L.7 identifies the centered gradients as β_d times the concentrated and flat blocks in equation 388. Their nuisance component remains zero, so Lemma H.2 removes the payoff-null trivial block. Proposition L.8 therefore gives equation 643.

The restricted widths of the two sign-block gradients are β_d and $\beta_d\sqrt{r/d}$. Applying the binary risk–width theorem gives equation 644; division yields equation 645. □

Remark L.11 (Exact scope). The commitment lift is time homogeneous, strictly positive, and uses a fixed finite action set. Its transition parameter is the affine image of the same C_2-equivariant rank-constrained signal kernel used in Appendix I. The augmented multiplicity block contains duplicated mode coordinates, but the free local parameter remains the common rank-r signal block; the equality coupling across modes is part of the model definition. The result is fixed-dimensional and generative. It does not assert a trajectory-sampling or growing-dimensional rate.

M Certified computation of Gaussian policy-selection games

This appendix turns the Gaussian benchmark into a certified finite computation. The parameter set may be nonconvex, disconnected, or a finite union of sparse determinantal cones. We prove four complementary results: (i) exact least-favorable-prior duality for every finite parameter set; (ii) a primal–dual projected-supergradient scheme with a nonasymptotic gap certificate; (iii) a completely finite linear program obtained by jointly discretizing the parameter and observation spaces, with an explicit error bound; and (iv) a semi-infinite exchange algorithm with finite termination for every prescribed tolerance. The minimax interchange uses the compact-convex principle of Sion (1958); every approximation and convergence bound needed below is proved directly. Exchange and discretization methods are classical paradigms in semi-infinite programming (Hettich & Kortanek, 1993), but the present certificates exploit the rule-uniform Gaussian-risk modulus specific to our decision problem.

We use the notation of Appendix J.3. Thus A is a nonempty finite candidate set, $B = (b_a)_{a\in A} \subset \mathbb{R}^d$, and

$$L_B(\theta, q) = \max_{a\in A}\langle b_a, \theta\rangle - \sum_{a\in A} q(a)\langle b_a, \theta\rangle. \tag{646}$$

For a compact $K \subseteq R\mathbb{B}_2^d$, write

$$V(K, B) = \inf_{\delta} \sup_{\theta\in K} \mathbb{E}_{Z\sim N(\theta, I_d)} L_B\big(\theta, \delta(Z)\big). \tag{647}$$

By common-gradient translation invariance, we may and do center B so that

$$\max_{a\in A}\|b_a\|_2 \le \Gamma, \qquad \Gamma := \Gamma(B). \tag{648}$$

Then $0 \le L_B \le 2\Gamma R$ on $R\mathbb{B}_2^d \times \Delta(A)$.

M.1 Finite parameter games and least-favorable priors

Let

$$F = \{\theta_1, \dots, \theta_m\} \subseteq R\mathbb{B}_2^d \tag{649}$$

be nonempty, and define

$$h_i := \max_{a\in A}\langle b_a, \theta_i\rangle, \qquad p_i(z) := \varphi_d(z-\theta_i). \tag{650}$$

For a rule δ, write

$$r_i(\delta) := \mathbb{E}_{\theta_i} L_B\big(\theta_i, \delta(Z)\big), \qquad r(\delta) := (r_1(\delta), \dots, r_m(\delta)). \tag{651}$$

Theorem M.1 (Finite-game duality and explicit Bayes rules)**.** *For $\lambda \in \Delta_m$, define*

$$\boxed{\beta_F(\lambda) := \sum_{i=1}^m \lambda_i h_i - \int_{\mathbb{R}^d} \max_{a\in A} \left\langle b_a, \sum_{i=1}^m \lambda_i p_i(z)\theta_i \right\rangle dz.} \tag{652}$$

Then:

1. *the finite Gaussian game satisfies the exact dual identity*

$$\boxed{V(F,B) = \max_{\lambda\in\Delta_m} \beta_F(\lambda);} \tag{653}$$

2. *β_F is concave and continuous, and*

$$|\beta_F(\lambda) - \beta_F(\mu)| \le 2\Gamma R\|\lambda - \mu\|_1; \tag{654}$$

3. *for every prior λ, a Bayes rule is obtained by the Borel tie-breaking selector*

$$a_\lambda(z) := \min \arg\max_{a\in A} \left\langle b_a, \sum_{i=1}^m \lambda_i p_i(z)\theta_i \right\rangle; \tag{655}$$

4. *if δ_λ is any Bayes rule for λ, then its risk vector is a supergradient of the concave dual objective:*

$$\boxed{\beta_F(\mu) \le \beta_F(\lambda) + \langle \mu - \lambda, r(\delta_\lambda)\rangle \quad \forall \mu \in \Delta_m;} \tag{656}$$

5. *a least-favorable prior $\lambda^\star$ and a minimax rule $\delta^\star$ exist. They may be chosen as a saddle pair, in which case*

$$r_i(\delta^\star) \le V(F,B) \quad \forall i, \qquad r_i(\delta^\star) = V(F,B) \quad \text{if } \lambda_i^\star > 0. \tag{657}$$

Proof. Let

$$\nu := \frac{1}{m}\sum_{i=1}^m N(\theta_i, I_d). \tag{658}$$

All $N(\theta_i, I_d)$ are equivalent to Lebesgue measure and hence to ν. Identify a randomized rule with $q = (q_a)_{a\in A} \in L^\infty(\nu)^{|A|}$ satisfying $q_a \ge 0$ and $\sum_a q_a = 1$ ν-almost surely. Denote this set by $\mathfrak{D}_F$. It is convex. It is weak-$*$ compact because it is a weak-$*$ closed subset of the product of the unit balls of $L^\infty(\nu)$: positivity is closed since $\int f q_a \, d\nu \ge 0$ for every nonnegative $f \in L^1(\nu)$, and the simplex equality is weak-$*$ closed for the same reason.

For every i, the map $q \mapsto r_i(q)$ is weak-$*$ continuous because its coefficient is an $L^1(\nu)$ function. Therefore

$$\Psi(q, \lambda) := \sum_{i=1}^{m} \lambda_i r_i(q) \tag{659}$$

is continuous and bilinear on the compact convex sets $\mathfrak{D}_F \times \Delta_m$. Since $\max_i r_i(q) = \max_{\lambda \in \Delta_m} \Psi(q, \lambda)$, Sion's theorem gives

$$\begin{aligned} V(F, B) &= \min_{q \in \mathfrak{D}_F} \max_{\lambda \in \Delta_m} \Psi(q, \lambda) \\ &= \max_{\lambda \in \Delta_m} \min_{q \in \mathfrak{D}_F} \Psi(q, \lambda). \end{aligned} \tag{660}$$

For fixed λ, expand the inner objective using Lebesgue densities:

$$\Psi(q, \lambda) = \sum_i \lambda_i h_i - \int_{\mathbb{R}^d} \sum_{a \in A} q_a(z) \left\langle b_a, \sum_i \lambda_i p_i(z) \theta_i \right\rangle dz.$$

The integrand is minimized pointwise by putting all mass on a maximizing action. Because A is finite and every score is continuous, the smallest-index maximizer in equation 655 is Borel. This proves the formula equation 652 and the dual identity.

The representation $\beta_F(\lambda) = \inf_q \langle \lambda, r(q) \rangle$ proves concavity. Since $0 \le r_i(q) \le 2\Gamma R$,

$$\beta_F(\lambda) - \beta_F(\mu) \le \sup_q \langle \lambda - \mu, r(q) \rangle \le 2\Gamma R \|\lambda - \mu\|_1,$$

and reversing λ, μ gives equation 654. Compactness therefore yields a maximizing prior.

If δ_λ is Bayes for λ, then $\beta_F(\lambda) = \langle \lambda, r(\delta_\lambda) \rangle$, whereas $\beta_F(\mu) \le \langle \mu, r(\delta_\lambda) \rangle$. Subtracting proves equation 656.

Finally, compactness of $\mathfrak{D}_F$ gives a minimax rule. Any minimax rule and least-favorable prior form a saddle pair: for all q, λ,

$$\Psi(\delta^\star, \lambda) \le V(F, B) \le \Psi(q, \lambda^\star).$$

Taking λ to be each simplex vertex gives $r_i(\delta^\star) \le V(F, B)$. Since $\sum_i \lambda_i^\star r_i(\delta^\star) = V(F, B)$, every index of positive prior mass must have equality. □

Remark M.2 (Standard-normal evaluation formula). Since $p_i(z) = \varphi_d(z) \exp\{\theta_i^\top z - \|\theta_i\|_2^2/2\}$, equation 652 is equivalently

$$\begin{aligned} \beta_F(\lambda) = &\sum_i \lambda_i h_i \\ &- \mathbb{E}_{G \sim N(0, I_d)} \max_{a \in A} \sum_i \lambda_i e^{\theta_i^\top G - \|\theta_i\|_2^2/2} \langle b_a, \theta_i \rangle. \end{aligned} \tag{661}$$

This expression uses one fixed reference Gaussian for every prior and is convenient for quadrature or Monte Carlo evaluation. The certified finite LP below avoids stochastic integration altogether.

M.2 A PRIMAL–DUAL ALGORITHM FOR FINITE GAMES

The dual objective is nonsmooth but has bounded supergradients. This yields a simple certified algorithm that simultaneously approximates a least-favorable prior and a minimax rule.

Projected-supergradient procedure. Given $T \ge 1$, initialize $\lambda_1 \in \Delta_m$. For $t = 1, \dots, T$:

1. compute a Bayes rule δ_t for λ_t;
2. compute its risk vector $g_t := r(\delta_t)$;
3. update

$$\lambda_{t+1} = \Pi_{\Delta_m}\big(\lambda_t + \eta g_t\big), \tag{662}$$

where Π_{Δ_m} is Euclidean projection.

Return the averaged rule

$$\overline{\delta}_T(z) := \frac{1}{T}\sum_{t=1}^{T}\delta_t(z), \tag{663}$$

and the certificates

$$\underline{V}_T := \max_{t\le T}\beta_F(\lambda_t), \qquad \overline{V}_T := \max_{i\le m} r_i(\overline{\delta}_T). \tag{664}$$

Theorem M.3 (Certified projected-supergradient solution of a finite game)**.** *Assume $\Gamma R > 0$ and set*

$$D := \sqrt{2}, \qquad G := 2\Gamma R\sqrt{m}, \qquad \eta := \frac{D}{G\sqrt{T}}. \tag{665}$$

Then

$$\boxed{\underline{V}_T \le V(F,B) \le \overline{V}_T, \qquad \overline{V}_T - \underline{V}_T \le \frac{2\sqrt{2}\,\Gamma R\sqrt{m}}{\sqrt{T}}.} \tag{666}$$

In particular, $T \ge 8\Gamma^2R^2m/\varepsilon^2$ guarantees a primal–dual gap at most ε. If $\Gamma R = 0$, the game value is zero and any rule is minimax.

Proof. By Theorem M.1, g_t is a supergradient of β_F at λ_t, and

$$\|g_t\|_2 \le 2\Gamma R\sqrt{m} = G. \tag{667}$$

The simplex has Euclidean diameter at most $D = \sqrt{2}$. Nonexpansiveness of Euclidean projection gives, for every $\lambda \in \Delta_m$,

$$\begin{aligned}\|\lambda_{t+1} - \lambda\|_2^2 &\le \|\lambda_t + \eta g_t - \lambda\|_2^2 \\ &= \|\lambda_t - \lambda\|_2^2 + 2\eta\langle g_t, \lambda_t - \lambda\rangle + \eta^2\|g_t\|_2^2.\end{aligned}$$

Summing and rearranging yields the standard regret inequality

$$\sum_{t=1}^{T}\langle \lambda - \lambda_t, g_t\rangle \le \frac{D^2}{2\eta} + \frac{\eta G^2 T}{2}. \tag{668}$$

Because δ_t is Bayes for λ_t,

$$\langle \lambda_t, g_t\rangle = \beta_F(\lambda_t). \tag{669}$$

Choose λ to maximize $\langle \lambda, T^{-1}\sum_t g_t\rangle$. Since risk is affine in the rule, $T^{-1}\sum_t g_t = r(\overline{\delta}_T)$, and hence

$$\begin{aligned}\overline{V}_T &= \max_{\lambda\in\Delta_m}\left\langle \lambda, \frac{1}{T}\sum_t g_t\right\rangle \\ &\le \frac{1}{T}\sum_t \beta_F(\lambda_t) + \frac{D^2}{2\eta T} + \frac{\eta G^2}{2} \\ &\le \underline{V}_T + \frac{D^2}{2\eta T} + \frac{\eta G^2}{2}.\end{aligned}$$

Weak duality gives $\underline{V}_T \le V(F,B)$, while $V(F,B) \le \overline{V}_T$ because $\overline{\delta}_T$ is an admissible rule. Substituting equation 665 gives $DG/\sqrt{T} = 2\sqrt{2}\Gamma R\sqrt{m/T}$. □

M.3 A COMPLETELY FINITE LINEAR PROGRAM

The preceding algorithm still assumes exact Gaussian integration when it computes Bayes risks and risk vectors. We now give a fully finite alternative whose coefficients are products of one-dimensional Gaussian cdf differences.

Fix $\sigma, h > 0$ and $M > R$. Partition the cube $[-M, M]^d$ into axis-aligned boxes $C_1, \dots, C_J$ of side length at most h, and put

$$C_\infty := \mathbb{R}^d \setminus [-M, M]^d. \tag{670}$$

For $i \in [m]$ and $j \in [J] \cup \{\infty\}$, define

$$\pi_{ij} := N(\theta_i, I_d)(C_j). \tag{671}$$

If $C_j = \prod_{\ell=1}^{d} [\ell_{j\ell}, u_{j\ell})$, then

$$\pi_{ij} = \prod_{\ell=1}^{d} \left[\Phi(u_{j\ell} - \theta_{i\ell}) - \Phi(\ell_{j\ell} - \theta_{i\ell})\right], \tag{672}$$

and $\pi_{i\infty} = 1 - \sum_{j=1}^{J} \pi_{ij}$.

A partition rule is specified by $q_j \in \Delta(A)$ for every $j \in [J] \cup \{\infty\}$ and uses q_j when $Z \in C_j$. Its finite-parameter risk is

$$R_i(q) := \sum_{j \in [J] \cup \{\infty\}} \pi_{ij} L_B(\theta_i, q_j). \tag{673}$$

The optimal partition value is the linear program

$$\boxed{\begin{aligned} W(F, B; M, h) := \min_{t, (q_j)} \quad & t \\ \text{subject to} \quad & R_i(q) \le t, && i = 1, \dots, m, \\ & q_j(a) \ge 0, && j \in [J] \cup \{\infty\},\ a \in A, \\ & \sum_{a \in A} q_j(a) = 1, && j \in [J] \cup \{\infty\}. \end{aligned}} \tag{674}$$

For later use define

$$S_{R,\Gamma,d}(\sigma) := \Gamma R \sqrt{d\,[\sigma^2 - \log(1+\sigma^2)]}, \tag{675}$$

$$G_{R,\Gamma,d}(\sigma, h) := \frac{\Gamma R\, h \sqrt{d}}{\sigma\sqrt{2\pi}}, \tag{676}$$

$$\tau_{d,R}(M) := \min\left\{1, 2d\Phi(-(M - R))\right\}, \tag{677}$$

$$T_{R,\Gamma,d}(M) := 2\Gamma R \tau_{d,R}(M), \tag{678}$$

$$D_{\mathrm{obs}}(\sigma, h, M) := S_{R,\Gamma,d}(\sigma) + G_{R,\Gamma,d}(\sigma, h) + T_{R,\Gamma,d}(M). \tag{679}$$

Lemma M.4 (Gaussian mollification and observation gridding)**.** *For every Borel rule δ, define*

$$\delta^{(\sigma)}(z) := \mathbb{E}_{G \sim N(0, I_d)} \delta(z + \sigma G). \tag{680}$$

Then:

1. *$\delta^{(\sigma)}$ is simplex valued and*

$$\sup_{\theta \in R\mathbb{B}_2^d} \left| r_{B,\delta^{(\sigma)}}(\theta) - r_{B,\delta}(\theta) \right| \le S_{R,\Gamma,d}(\sigma); \tag{681}$$

2. *it is globally Lipschitz in total output variation:*

$$\|\delta^{(\sigma)}(z) - \delta^{(\sigma)}(z')\|_1 \le \sqrt{\frac{2}{\pi}} \frac{\|z - z'\|_2}{\sigma}; \tag{682}$$

3. *if z_j is the center of C_j, define the partition rule $q_j := \delta^{(\sigma)}(z_j)$ on bounded cells and take any $q_\infty \in \Delta(A)$ on the tail cell. Then*

$$\sup_{\theta \in R\mathbb{B}_2^d} |r_{B,q}(\theta) - r_{B,\delta}(\theta)| \le D_{\mathrm{obs}}(\sigma, h, M). \tag{683}$$

Proof. The rule $\delta^{(\sigma)}$ is the output law obtained by passing $N(z, \sigma^2 I_d)$ through the Markov kernel δ, so it is Borel and simplex valued. If $Z \sim N(\theta, I_d)$ and G is independent, then $Z + \sigma G \sim N(\theta, (1+\sigma^2) I_d)$. The loss is bounded by $2\Gamma R$. Moreover,

$$D_{\mathrm{KL}}\left(N(\theta, (1+\sigma^2) I_d) \middle\| N(\theta, I_d)\right) = \frac{d}{2}\left[\sigma^2 - \log(1+\sigma^2)\right]. \tag{684}$$

Pinsker's inequality gives total variation at most one half of the square root in equation 681; multiplying by the loss envelope proves that bound.

Markov-kernel contraction of total variation and Lemma J.8 give

$$\begin{aligned}\|\delta^{(\sigma)}(z)-\delta^{(\sigma)}(z')\|_1 &\le 2\mathrm{TV}\left(N(z,\sigma^2 I_d),N(z',\sigma^2 I_d)\right)\\ &= 2\left[2\Phi\left(\frac{\|z-z'\|_2}{2\sigma}\right)-1\right]\\ &\le \sqrt{\frac{2}{\pi}}\frac{\|z-z'\|_2}{\sigma},\end{aligned}$$

which proves equation 682.

Inside a bounded box, $\|z-z_j\|_2 \le h\sqrt{d}/2$. Since the loss is ΓR-Lipschitz in q under the ℓ_1 norm, replacing $\delta^{(\sigma)}(z)$ by q_j costs at most $G_{R,\Gamma,d}(\sigma,h)$. Outside the cube, two losses in $[0,2\Gamma R]$ differ by at most $2\Gamma R$. For $Z\sim N(\theta,I_d)$ with $\|\theta\|_2\le R$, a union bound gives

$$\Pr_\theta\{Z\notin[-M,M]^d\} \le 2d\Phi(-(M-R)) \le \tau_{d,R}(M). \tag{685}$$

Adding the mollification, grid, and tail errors proves equation 683. □

Theorem M.5 (Certified finite linear-program approximation). *For every finite $F\subseteq R\mathbb{B}_2^d$,*

$$\boxed{0\le W(F,B;M,h)-V(F,B)\le D_{\mathrm{obs}}(\sigma,h,M).} \tag{686}$$

Now let $K\subseteq R\mathbb{B}_2^d$ be nonempty compact and let $F\subseteq K$ be a finite ε_K-net. If q^{LP} is an optimal solution of equation 674, then

$$\boxed{W-D_{\mathrm{obs}}\le V(K,B)\le W+\Omega_{R,\Gamma}(\varepsilon_K),} \tag{687}$$

where $W=W(F,B;M,h)$, and

$$\boxed{\mathcal{R}(K,B;q^{\mathrm{LP}})\le V(K,B)+D_{\mathrm{obs}}(\sigma,h,M)+\Omega_{R,\Gamma}(\varepsilon_K).} \tag{688}$$

Proof. Partition-measurable rules are a subset of all Borel rules, so $V(F,B)\le W$. Conversely, choose a rule with finite-game risk at most $V(F,B)+\xi$. Lemma M.4 constructs a partition rule with risk at most $V(F,B)+\xi+D_{\mathrm{obs}}$. Letting $\xi\downarrow 0$ proves equation 686.

Because $F\subseteq K$, Corollary J.13 gives

$$V(F,B)\le V(K,B)\le V(F,B)+\Omega_{R,\Gamma}(\varepsilon_K). \tag{689}$$

Combining this with $W-D_{\mathrm{obs}}\le V(F,B)\le W$ proves equation 687. Finally, the LP rule has risk W on F, so the rule-uniform modulus gives

$$\begin{aligned}\mathcal{R}(K,B;q^{\mathrm{LP}}) &\le W+\Omega_{R,\Gamma}(\varepsilon_K)\\ &\le V(K,B)+D_{\mathrm{obs}}+\Omega_{R,\Gamma}(\varepsilon_K),\end{aligned}$$

proving equation 688. □

Corollary M.6 (An explicit ε-accurate finite algorithm). *Assume $d\ge 1$ and $\Gamma R>0$, and let $\varepsilon>0$. Put*

$$C_{R,\Gamma}:=\left(2+\sqrt{\frac{2}{\pi}}R\right)\Gamma, \qquad \varepsilon_K:=\frac{\varepsilon}{4C_{R,\Gamma}}, \tag{690}$$

$$\sigma:=\min\left\{1,\sqrt{\frac{\varepsilon}{4\Gamma R\sqrt{d/2}}}\right\}, \qquad h:=\frac{\varepsilon\sigma\sqrt{2\pi}}{4\Gamma R\sqrt{d}}, \tag{691}$$

$$M:=R+\sqrt{2\log\left(\max\left\{e,\frac{8d\Gamma R}{\varepsilon}\right\}\right)}. \tag{692}$$

Let F be any finite ε_K-net of K, and solve the finite LP. Then

$$\mathcal{R}(K, B; q^{\mathrm{LP}}) \le V(K, B) + \varepsilon, \tag{693}$$

and the certified value interval in equation 687 has width at most ε. Moreover, one may choose

$$|F| \le \left(1 + \frac{2R}{\varepsilon_K}\right)^d, \qquad J \le \left(1 + \frac{2M}{h}\right)^d. \tag{694}$$

Thus the LP has at most $(J+1)|A| + 1$ scalar variables, $|F|$ risk constraints, and $J+1$ simplex equalities, in addition to nonnegativity. If $\Gamma R = 0$ or $d = 0$, the value is zero and no optimization is needed.

Proof. The elementary inequality $x - \log(1+x) \le x^2/2$ for $x \ge 0$ gives

$$S_{R,\Gamma,d}(\sigma) \le \Gamma R \sigma^2 \sqrt{d/2} \le \varepsilon/4. \tag{695}$$

The definition of h makes $G_{R,\Gamma,d}(\sigma, h) = \varepsilon/4$. The Gaussian tail bound $\Phi(-x) \le \frac{1}{2} e^{-x^2/2}$ gives

$$\tau_{d,R}(M) \le d e^{-(M-R)^2/2} \le \frac{\varepsilon}{8\Gamma R}, \tag{696}$$

so $T_{R,\Gamma,d}(M) \le \varepsilon/4$. Hence $D_{\mathrm{obs}} \le 3\varepsilon/4$. The linear risk modulus gives $\Omega_{R,\Gamma}(\varepsilon_K) \le \varepsilon/4$, proving the error claim.

For the covering bound, take a maximal ε_K-separated subset of K. The balls of radius $\varepsilon_K/2$ around its points are disjoint and lie in the ball of radius $R + \varepsilon_K/2$. Comparing Euclidean volumes gives the first inequality in equation 694. The box-count bound follows by partitioning each interval $[-M, M]$ into at most $1 + 2M/h$ subintervals. □

M.4 Exchange algorithm and finite termination

A global parameter net is exponentially expensive. An exchange method adds only parameters that are currently worst for the computed rule.

Exact exchange procedure. Choose a nonempty finite $F_1 \subseteq K$. At iteration m:

1. solve the finite game on F_m, obtaining a minimax rule δ_m and value
$$\underline{V}_m := V(F_m, B); \tag{697}$$
2. compute
$$\theta_m \in \arg\max_{\theta \in K} r_{B,\delta_m}(\theta), \qquad \overline{V}_m := r_{B,\delta_m}(\theta_m); \tag{698}$$
3. stop if $\overline{V}_m - \underline{V}_m \le \varepsilon$; otherwise set $F_{m+1} := F_m \cup \{\theta_m\}$.

Define the inverse modulus

$$\rho_\varepsilon := \inf\{t \ge 0 : \Omega_{R,\Gamma}(t) \ge \varepsilon\}. \tag{699}$$

If $\Gamma > 0$ and $\varepsilon > 0$, then $\rho_\varepsilon > 0$.

Theorem M.7 (Certified finite termination of the exchange algorithm)**.** *Assume exact finite-game and worst-risk oracles. At every iteration,*

$$\boxed{\underline{V}_m \le V(K, B) \le \overline{V}_m.} \tag{700}$$

For every $\varepsilon > 0$, the algorithm terminates after finitely many iterations. More precisely, before termination all newly added points are pairwise more than ρ_ε apart, so the number of additions is at most the packing number

$$\mathsf{Pack}(K, \rho_\varepsilon) \le \left(1 + \frac{2R}{\rho_\varepsilon}\right)^d. \tag{701}$$

At termination, δ_m is ε-minimax for the full game.

Proof. Since $F_m \subseteq K$, monotonicity gives $\underline{V}_m \le V(K,B)$. Since δ_m is an admissible full-game rule, $V(K,B) \le \mathcal{R}(K,B;\delta_m) = \overline{V}_m$, proving the certificate.

Suppose the algorithm does not stop at iteration m. For every $\theta \in F_m$, finite-game optimality gives

$$r_{B,\delta_m}(\theta) \le \underline{V}_m. \tag{702}$$

The newly added point satisfies

$$r_{B,\delta_m}(\theta_m) = \overline{V}_m > \underline{V}_m + \varepsilon. \tag{703}$$

The rule-uniform risk modulus therefore implies

$$\Omega_{R,\Gamma}\big(\|\theta_m - \theta\|_2\big) > \varepsilon, \qquad \theta \in F_m, \tag{704}$$

so $\|\theta_m - \theta\|_2 > \rho_\varepsilon$. Thus every nonterminal addition is separated from every earlier point by more than ρ_ε. Compactness rules out an infinite such sequence, and the volumetric packing bound gives equation 701. At termination,

$$\mathcal{R}(K,B;\delta_m) = \overline{V}_m \le \underline{V}_m + \varepsilon \le V(K,B) + \varepsilon. \tag{705}$$

□

Theorem M.8 (Inexact exchange with certified inner and separation errors)**.** *At iteration m, suppose the finite solver returns a number L_m and rule δ_m satisfying*

$$L_m \le V(F_m,B), \qquad \max_{\theta\in F_m} r_{B,\delta_m}(\theta) \le L_m + \alpha_m, \tag{706}$$

and the separation oracle returns $\theta_m \in K$ and $\beta_m \ge 0$ satisfying

$$\sup_{\theta\in K} r_{B,\delta_m}(\theta) \le r_{B,\delta_m}(\theta_m) + \beta_m. \tag{707}$$

Put

$$U_m := r_{B,\delta_m}(\theta_m) + \beta_m. \tag{708}$$

Then

$$L_m \le V(K,B) \le U_m. \tag{709}$$

If the method stops when $U_m - L_m \le \varepsilon$, its returned rule is ε-minimax. If, before termination,

$$\alpha_m + \beta_m \le \varepsilon/2, \tag{710}$$

then every new point is separated from F_m by more than $\rho_{\varepsilon/2}$. Consequently the method terminates after at most $\mathsf{Pack}(K,\rho_{\varepsilon/2})$ *additions.*

Proof. The lower certificate follows from $F_m \subseteq K$, and the upper certificate follows from equation 707. At termination,

$$\mathcal{R}(K,B;\delta_m) \le U_m \le L_m + \varepsilon \le V(K,B) + \varepsilon.$$

If the method has not terminated, then

$$r_{B,\delta_m}(\theta_m) \ge U_m - \beta_m > L_m + \varepsilon - \beta_m,$$

whereas every old point has risk at most $L_m + \alpha_m$. Therefore the risk difference exceeds $\varepsilon - \alpha_m - \beta_m \ge \varepsilon/2$, and the common modulus gives the separation. The packing argument is identical to Theorem M.7. □

Remark M.9 (Using the finite LP inside exchange). Theorem M.5 supplies exactly the inner certificate required by equation 706: for the current finite set F_m, take

$$L_m = W_m - D_{\mathrm{obs},m}, \qquad \alpha_m = D_{\mathrm{obs},m}. \tag{711}$$

Thus a prescribed exchange tolerance can be divided explicitly between observation discretization and approximate worst-risk separation.

M.5 Explicit nets for rank-constrained and block-sparse cones

The generic volumetric bound uses the ambient dimension. The next result exhibits a net whose exponent reflects rank.

Proposition M.10 (Covering a Frobenius rank cone)**.** *Let*

$$\mathcal{M}_{m,n,r}(R) := \left\{X \in \mathbb{R}^{m\times n} : \operatorname{rank}(X) \le r,\ \|X\|_{\mathrm{F}} \le R\right\}. \tag{712}$$

For $R > 0$, $1 \le r \le \min\{m, n\}$, and $\varepsilon > 0$, this set admits an internal ε-net in Frobenius norm of cardinality at most

$$\boxed{\left(1 + \frac{6R\sqrt{r}}{\varepsilon}\right)^{r(m+n)} \left(1 + \frac{6R}{\varepsilon}\right)^{r}.} \tag{713}$$

For $r = 0$ or $R = 0$, the set is the singleton $\{0\}$.

Proof. Pad every compact singular-value decomposition to length r:

$$X = U \operatorname{diag}(\sigma) V^{\top}, \qquad U \in \mathrm{St}(m, r), \quad V \in \mathrm{St}(n, r), \quad \sigma \in \mathbb{R}^{r}_{+}, \quad \|\sigma\|_2 \le R. \tag{714}$$

The Stiefel sets lie in Frobenius balls of radius $\sqrt{r}$ in ambient Euclidean dimensions mr and nr. The same maximal-packing volume argument used in Corollary M.6 gives a δ_U-net of $\mathrm{St}(m, r)$ of cardinality at most $(1 + 2\sqrt{r}/\delta_U)^{mr}$ and similarly for V. The radius-R singular value ball has a δ_σ-net of size at most $(1 + 2R/\delta_\sigma)^r$.

For two triples,

$$\|U \operatorname{diag}(\sigma) V^{\top} - \widetilde{U} \operatorname{diag}(\widetilde{\sigma}) \widetilde{V}^{\top}\|_{\mathrm{F}} \le R\|U - \widetilde{U}\|_{\mathrm{F}} + \|\sigma - \widetilde{\sigma}\|_2 + R\|V - \widetilde{V}\|_{\mathrm{F}}.$$

Choose $\delta_U = \delta_V = \varepsilon/(3R)$ and $\delta_\sigma = \varepsilon/3$. Multiplying the three net cardinalities gives equation 713. □

Corollary M.11 (Covering a block-sparse low-rank cone section)**.** *Consider the Euclidean product cone section*

$$\begin{aligned}\mathcal{K}_{\mathrm{blk}}(R) := \Big\{(x_0, (X_\rho)_{\rho\in\mathfrak{R}}) : \|x_0\|_2^2 + \sum_{\rho} \|X_\rho\|_{\mathrm{F}}^2 \le R^2,\\ |\operatorname{supp}(X)| \le s, \quad \operatorname{rank}(X_\rho) \le r_\rho\Big\},\end{aligned} \tag{715}$$

where $x_0 \in \mathbb{R}^{d_0}$ and $X_\rho \in \mathbb{R}^{m_\rho \times n_\rho}$. Put $\bar{\varepsilon} := \varepsilon/\sqrt{s+1}$. Then an ε-net exists with cardinality at most

$$\boxed{\left(1 + \frac{2R}{\bar{\varepsilon}}\right)^{d_0} \sum_{\substack{J\subseteq\mathfrak{R}\\ |J|\le s}} \prod_{\rho\in J} \left(1 + \frac{6R\sqrt{r_\rho}}{\bar{\varepsilon}}\right)^{r_\rho(m_\rho+n_\rho)} \left(1 + \frac{6R}{\bar{\varepsilon}}\right)^{r_\rho}.} \tag{716}$$

Factors with $r_\rho = 0$ are interpreted as one.

Proof. For each support set J, cover the nuisance ball and every active block with accuracy $\bar{\varepsilon}$, using Proposition M.10. Before enforcing the common radius constraint, the product approximation error is at most $\sqrt{|J|+1}\,\bar{\varepsilon} \le \varepsilon$. If a product net point has norm larger than R, project it radially onto the radius-R Euclidean ball. Euclidean ball projection is nonexpansive, fixes the target point, and common radial scaling preserves every support and rank constraint. The resulting centers therefore lie in $\mathcal{K}_{\mathrm{blk}}(R)$ with no increase in approximation error. Taking the union over all admissible supports gives the displayed bound. □

M.6 Computational scope

The preceding results give three levels of computation.

1. For a finite mean set, Theorem M.1 converts the infinite-dimensional decision problem into finite-dimensional least-favorable-prior optimization with explicit Bayes responses.

2. Corollary M.6 gives a completely finite LP and an explicit end-to-end error certificate for every compact mean set supplied with a finite net.
3. Theorems M.7 and M.8 replace a global net by active parameter generation and provide certified finite termination under exact or controlled-error oracles.

The algorithms are finite and globally certified, but the generic complexity is exponential in the Gaussian dimension and the exchange method requires a global worst-risk oracle. No polynomial-time claim is made for arbitrary nonconvex unions of sparse rank cones. For the block-sparse determinantal sections used in this paper, Corollary M.11 replaces ambient-dimensional covering by a rank-sensitive bound; designing a polynomial-time separation oracle for the general multi-policy risk remains a separate optimization problem.

N FINITE-SAMPLE EVALUATION AND REPRODUCIBILITY

This appendix gives the complete finite-sample evaluation protocol, proves the coverage guarantee for every reported interval, defines every simulated rule, and records the numerical results. All observations are generated from the finite categorical transition experiments in Appendices K and L. No Gaussian observation is simulated directly. The evaluation does not replace the analytical proofs; it measures how rapidly their finite-sample manifestations appear and compares geometry- aware rules with natural geometry-agnostic baselines.

N.1 PRE-SPECIFIED PROTOCOL AND EXACT SIMULTANEOUS INTERVALS

The executable protocol fixes the random seed, sample sizes, local parameters, number of Monte Carlo replications, policy rules, and figures before reading any output. Across the complete suite there are $K = 85$ reported Bernoulli error probabilities. Every scaled-regret observation has the form

$$Y_{\ell,m} = c_\ell E_{\ell,m}, \qquad E_{\ell,m} \in \{0,1\}, \tag{717}$$

where $c_\ell \geq 0$ is an exactly computed policy-regret cost and $E_{\ell,m}$ records a wrong decision in replication m.

For $0 < \alpha < 1$, let

$$\begin{aligned} L(k,M;\alpha) &:= \begin{cases} 0, & k = 0, \\ \mathrm{Beta}^{-1}(\alpha/2; k, M-k+1), & k > 0, \end{cases} \\ U(k,M;\alpha) &:= \begin{cases} 1, & k = M, \\ \mathrm{Beta}^{-1}(1-\alpha/2; k+1, M-k), & k < M. \end{cases} \end{aligned} \tag{718}$$

These are the exact Clopper–Pearson binomial limits (Clopper & Pearson, 1934).

Proposition N.1 (Simultaneous finite-sample coverage)**.** *For experiment index $\ell \in [K]$, let $E_{\ell,1}, \ldots, E_{\ell,M_\ell}$ be independent Bernoulli variables with mean p_ℓ, and put $K_\ell = \sum_m E_{\ell,m}$. Define*

$$I_\ell := \left[L\Big(K_\ell, M_\ell; \frac{\alpha}{K}\Big), U\Big(K_\ell, M_\ell; \frac{\alpha}{K}\Big) \right]. \tag{719}$$

Then

$$\boxed{\Pr\{p_\ell \in I_\ell \ \forall \ell \in [K]\}) \geq 1 - \alpha.} \tag{720}$$

Consequently, with the same simultaneous coverage, the risk $c_\ell p_\ell$ lies in $c_\ell I_\ell$. If two risks are positive and have simultaneous intervals $[L_1, U_1]$ and $[L_2, U_2]$ with $L_2 > 0$, then their ratio lies in

$$\left[\frac{L_1}{U_2}, \frac{U_1}{L_2} \right]. \tag{721}$$

Proof. For each ℓ, the Clopper–Pearson construction has coverage at least $1 - \alpha/K$ for every Bernoulli parameter. Hence

$$\Pr(p_\ell \notin I_\ell) \leq \alpha/K.$$

The union bound gives

$$\Pr(\exists \ell : p_\ell \notin I_\ell) \le \sum_{\ell=1}^{K} \alpha/K = \alpha.$$

Multiplication by a deterministic nonnegative cost preserves interval inclusion. The ratio statement follows by monotonicity on the positive orthant. □

All figures use $\alpha = 0.05$. The coverage claim is finite sample and does not rely on a normal approximation to the Monte Carlo estimator.

N.2 Experiment I: adaptation tax at the rank-zero intersection

We use the one-step transition experiment in Appendix K, with

$$R = 3, \qquad \vartheta = \pi/3, \qquad q = \cos\vartheta = 1/2. \tag{722}$$

The least-favorable line parameter is

$$t_R = 0.7517915169, \qquad \Phi(-t_R) = 0.2260882222. \tag{723}$$

The exact wrong-policy scaled-regret cost is

$$c_{\mathrm{wrong}} = q t_R = 0.3758957584. \tag{724}$$

For every replication, the branch $j \in \{1, 2\}$ and sign $\sigma \in \{-1, +1\}$ are drawn independently and uniformly, and transition samples are drawn independently from every state-action row of $P_{j,n,\sigma t_R}$.

We evaluate six rules.

1. *Exact branch-oracle LRT*: the branch j is revealed and the exact row-product likelihoods under $\sigma = +1$ and $\sigma = -1$ are compared.
2. *Exact hidden-branch mixture LRT*: compare the equal-branch mixture likelihoods for positive and negative signs.
3. *Gaussian branch score*: choose the sign of the exact finite- sample central score in the known direction U_j.
4. *Gaussian mixture score*: choose the sign of $\sinh(t_R S_1) + \sinh(t_R S_2)$.
5. *Empirical policy-gap score*: choose the sign of the score in the E_{11} policy-gradient direction.
6. *Uniform randomization*.

The first two rules are finite-sample Bayes-optimal for their corresponding known-branch two-point and hidden-branch four-point experiments by the Neyman–Pearson lemma. The Gaussian limits are

$$\begin{aligned} \mathcal{R}^G_{\mathrm{or}} &= q t_R \Phi(-t_R) = 0.0849856037, \\ \mathcal{R}^G_{\mathrm{mix}} &= q t_R e_{\mathrm{mix}}(t_R) = 0.1038003190. \end{aligned} \tag{725}$$

The numerical mixture integral is evaluated on a deterministic 1001×1001 grid over $[-7, 7]^2$; enlarging the grid or interval changes the displayed digits by less than 10^{-6}.

At $n = 8192$, using 12000 replications per plotted point, the exact branch oracle has risk 0.0835 with simultaneous interval $[0.0787, 0.0885]$, while the exact hidden-branch mixture rule has risk 0.1014 with interval $[0.0962, 0.1067]$. The empirical policy-gap score has risk 0.1309. Figure 3 shows that the finite-sample separation is visible throughout the evaluated range.

N.3 Experiment II: adaptive attainment on stable rank-one branches

We shift the two branch models to fixed rank-one tie centers

$$\Theta_{j,n,t} = 0.12 X_j + \frac{t}{\sqrt{n}} U_j. \tag{726}$$

The sample is split rowwise into

$$m_n = \lfloor n^{2/3} \rfloor \quad \text{pilot observations and} \quad N_n = n - m_n \quad \text{main observations.} \tag{727}$$

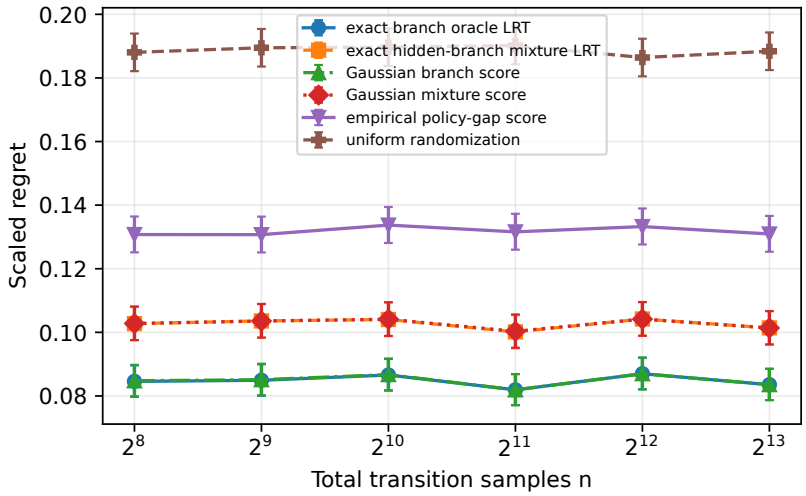

Figure 3: Finite-sample adaptation tax at the common rank-zero intersection. The exact hidden-branch LRT remains separated from the branch oracle and closely tracks the Gaussian-mixture prediction.

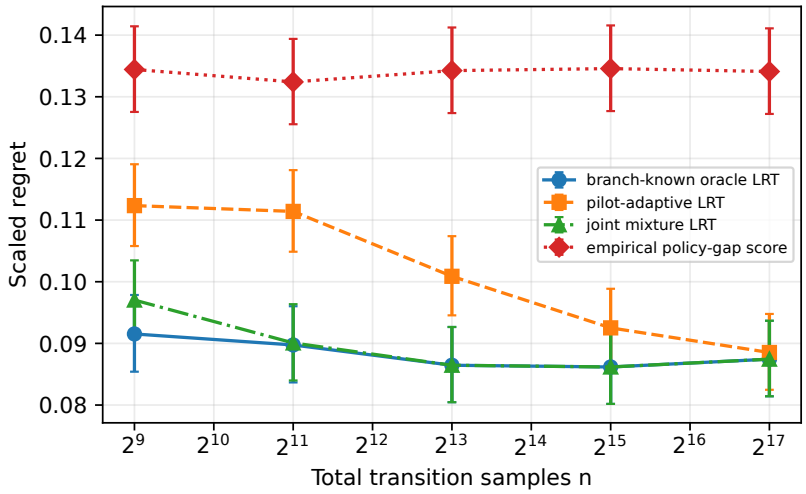

Figure 4: Adaptive attainment on stable rank-one branches. As the pilot identifies the branch, the pilot-adaptive LRT approaches the branch-known oracle.

The pilot performs a generalized likelihood-ratio comparison over both branch labels and a pre-specified 17-point center grid in $[0.04, 0.20]$. The main sample then applies the exact positive-versus-negative LRT within the selected branch and center. We compare it with the branch-known oracle, an exact joint mixture LRT using the true center, the empirical policy-gap score, and uniform randomization.

This rule is a computational specialization of the general adaptive procedure in Appendix J; no optimality claim is made for the finite pilot grid. Its branch-recovery probability is estimated directly. At $n = 131072$, based on 8000 replications, branch recovery is 0.9845, the pilot-adaptive risk is 0.0885 with interval $[0.0825, 0.0948]$, and the branch-known oracle risk is 0.0874 with interval $[0.0814, 0.0937]$. The empirical policy-gap rule remains at risk 0.1341.

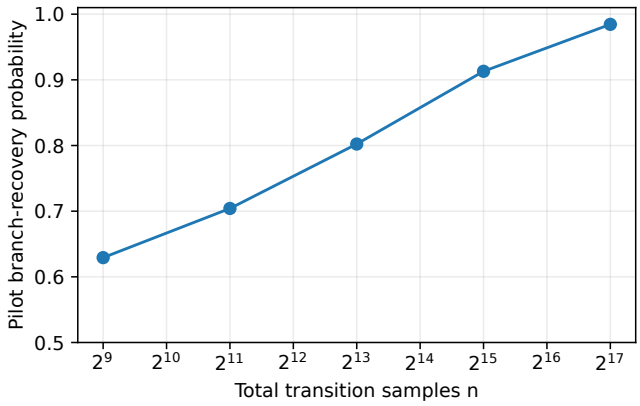

Figure 5: Pilot branch-recovery probability in the stable experiment.

Table 2: Finite-sample transition-data evaluation. Intervals are Bonferroni-adjusted exact binomial intervals, simultaneously valid over the pre-specified suite.

Setting and rule	Error probability	Scaled regret
singular: exact branch oracle LRT	0.222 [0.209,0.236]	0.084 [0.079,0.089]
singular: exact hidden-branch mixture LRT	0.270 [0.256,0.284]	0.101 [0.096,0.107]
singular: empirical policy-gap score	0.348 [0.333,0.363]	0.131 [0.125,0.137]
stable: branch-known oracle LRT	0.233 [0.217,0.249]	0.087 [0.081,0.094]
stable: pilot-adaptive LRT	0.235 [0.219,0.252]	0.089 [0.082,0.095]
stable: empirical policy-gap score	0.357 [0.338,0.375]	0.134 [0.127,0.141]

N.4 Experiment III: full-policy rank diffusion

We simulate the complete-policy commitment lift of Appendix L, not the earlier two-candidate gadget. We set $r = 1$, use $d \in \{2, 4, 8\}$, and take $n = 65536$ total transition samples under the balanced generative design. For each task, the two local alternatives are placed in the rank-one direction maximizing the corresponding restricted support functional. Their amplitude is normalized so that the lifted Fisher mean has Euclidean norm t_R.

The evaluated rules are:

1. the exact two-point likelihood-ratio test;
2. the score in the rank-cone maximizing direction;
3. the empirical policy-gradient score;
4. uniform randomization.

The exact LRT is Bayes-optimal for the equal-prior two-point subproblem. For each local alternative, dynamic programming verifies that one of the two canonical policies is globally optimal in the complete deterministic policy class. The plotted loss uses the exact finite-MDP value gap, rather than its first-order approximation.

The observed exact-LRT concentrated-to-diffuse risk ratios are

$$1.367\ [1.141, 1.639], \quad 1.917\ [1.603, 2.291], \quad 3.030\ [2.526, 3.636] \tag{728}$$

for $d = 2, 4, 8$, respectively, compared with $\sqrt{2}$, 2, and $\sqrt{8}$.

For the diffuse $d = 8$ task, the exact LRT and cone-line score have risks 0.02274 and 0.02274, whereas the empirical gradient score has risk 0.04121 and uniform randomization has risk 0.05271. This ablation isolates the value of respecting the feasible cone geometry.

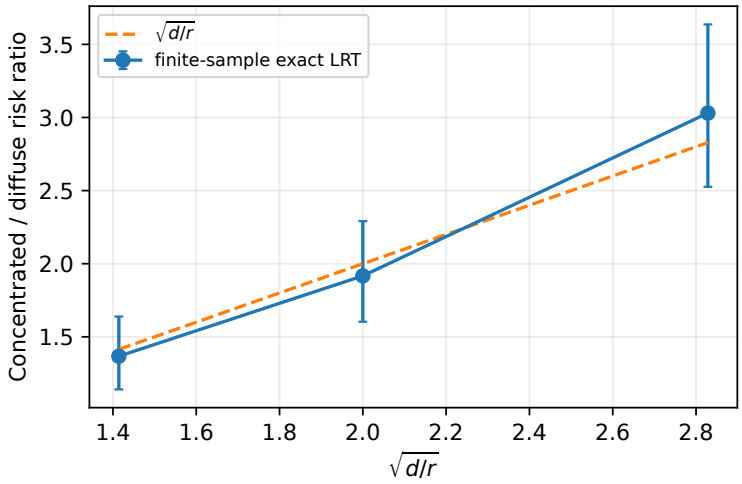

Figure 6: Finite-sample full-policy rank-diffusion ratio. The exact two-point LRT tracks the predicted $\sqrt{d/r}$ scaling.

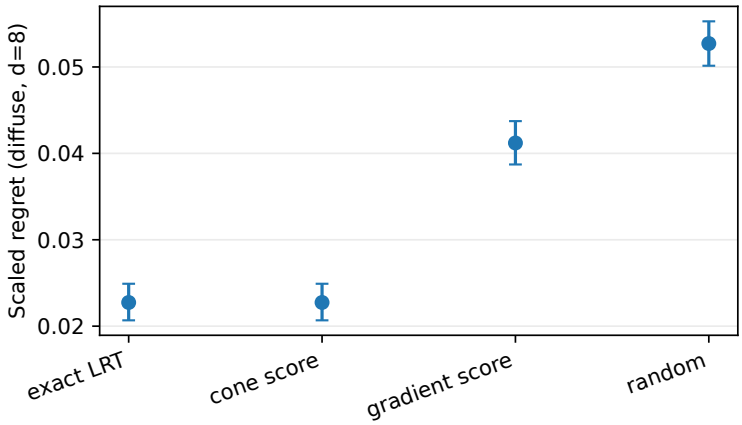

Figure 7: Geometry ablation in the diffuse $d = 8, r = 1$ full-policy task. The raw empirical policy-gradient score discards the rank-cone direction and incurs nearly twice the regret of the cone-aware score.

N.5 Reproducibility boundary

The supplemental directory contains:

- the complete simulation script;
- the raw CSV with all 85 estimates and intervals;
- the rank-ratio CSV;
- machine-readable metadata;
- all PDF and PNG figures;
- a requirements file and fixed random seed.

The default full run uses 12000 replications for every singular-intersection point, 8000 for every stable-stratum point, and 5000 for every full-policy rank-diffusion point. A separate quick mode reproduces the pipeline with smaller replication counts.

The evaluation establishes finite-sample performance only for the displayed constructed MDP families and rules. It does not prove a nonasymptotic minimax theorem over the complete structured model, a trajectory-sampling guarantee, or dimension-uniform convergence. Its purpose is to verify that the asymptotic oracle values, adaptation tax, and rank-diffusion effect are visible at finite sample sizes in the exact transition experiments used by the theory.

O COMPACT ASSUMPTION AND CLAIM AUDIT

This appendix records only the theorem dependencies, assumptions, and nonclaims needed to audit the oral-level result. Full proofs appear in the cited appendices; this section introduces no new claim.

O.1 DEPENDENCY GRAPH

The logical order is

$$B \to \{C, D\}, \qquad \{B, C, D\} \to E, \qquad \{D, E\} \to F,$$

$$\{B, D, F\} \to J, \qquad \{E, H, I\} \to K, \qquad \{B, D, F, H, I\} \to L, \qquad J \to M,$$

with Appendix N providing finite-sample evaluation of the constructions in Appendices K and L. Appendix H is the abstract Gaussian-game layer; Appendix M makes that layer computational.

O.2 ASSUMPTION LEDGER

Assumption	Role	Where used
Finite spaces, horizon, and policy class; bounded rewards	Smooth finite-dimensional Bellman values and uniform remainder bounds	Appendices D–F and J.
Known finite group; exact transition equivariance; real-Schur-type shared irreducibles	Affine intertwiner coordinates and block support/rank	Appendices B–C, G–I, L.
Strictly positive transition center and positive design masses	One dominated root-n categorical experiment and nondegenerate Fisher form	Appendices E–F and J–N.
Fixed radius and fixed dimension	Uniform local Gaussian approximation and compact Gaussian games	Appendices E–F, H, J–M.
Stable support/rank/policy margins	Uniform stratum and tied-face recovery and constrained-likelihood normal coordinates	Appendix J only.
Independent generative row samples	Product categorical likelihood and observable rowwise central sequence	Appendices E–F and J–N.

O.3 MAIN CLAIMS AND NONCLAIMS

Claim	**Proof location**	**Not claimed**
Oracle fixed-radius Gaussian decision limit	Appendices B–F	Full deficiency convergence, global PAC bounds, or growing-dimensional LAN.
Uniform adaptive oracle attainment on stable strata	Appendix J	Adaptation at support/rank/policy-face intersections or a finite-n minimax rate.
Positive adaptation tax at a rank intersection	Appendix K	A lower bound for every possible singular intersection.
Full-policy $\Theta_R(\sqrt{d/r})$ rank-diffusion family	Appendices I and L	A joint (n, d) theorem or trajectory-sampling result.
Certified Gaussian-game computation	Appendix M	Polynomial-time complexity for arbitrary nonconvex sparse-rank unions.
Finite-sample evidence with simultaneous exact intervals	Appendix N and code	A replacement for the analytical theorems.

O.4 IMPORTED INGREDIENTS AND PAPER-SPECIFIC CONTRIBUTION

Finite-group representation theory, determinantal tangent-cone identities, LAN, compact minimax, and binary Gaussian testing are imported with citation. The paper-specific contribution is their structured control synthesis: quantitative stochastic sparse-determinantal local geometry; Bellman policy regret; exact lower and upper transfers; adaptive attainment and an adaptation barrier; a full-policy rank-diffusion realization; and certified computation and finite-sample evaluation.