WHEN EQUILIBRIA BIFURCATE: STATISTICAL RESOLUTION LIMITS FOR LEARNING EQUILIBRIUM CORRESPONDENCES

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning in games often targets one profile with small exploitability. We show that this can be statistically much easier than learning the complete equilibrium correspondence: in a contiguous family of two-player zero-sum games, a data-independent profile has residual $O(T^{-1/2})$, while every adaptive T-query procedure incurs order-one minimax Hausdorff error for the exact correspondence. We therefore interpret approximation tolerance as statistical branch resolution. At a specified base game $A_\star$, we define a branch-survival cost and prove that, at every positive first-order resolution, approximate-equilibrium correspondences converge uniformly to its sublevel sets. The cost equals the duality gap of the perturbation game restricted to the base optimal-strategy polytopes and the support function of a critical-contrast polytope affinely equivalent to the complete base correspondence. Under local alternatives $A_\star + H/\sqrt{T}$, known-variance Gaussian payoff observations, and deterministic full-support sampling allocations, we obtain a finite-sample risk decomposition with an asymptotically vanishing strategic discrepancy and a local Gaussian random-set minimax limit. This yields a subcritical–critical–supercritical resolution transition and an adaptive lower bound for honest exact-set inference. The statistical comparison is fixed-dimensional and local oracle; complete Hausdorff convergence is proved at positive first-order resolution, while at exact resolution only an outer limit is universal.

1 INTRODUCTION

Learning in games is often formulated as a point problem: return one profile with small exploitability. The complete equilibrium correspondence is a different statistical object. An estimate may report only valid profiles and still omit strategically distant equilibria. Although all equilibria of a zero-sum game share the minimax value, their supports, action frequencies, perturbation robustness, and compatibility with external policy constraints can differ; downstream conclusions may therefore depend on the full optimal-strategy face rather than on one selected profile.

The distinction already appears in

$$A_\theta := \begin{pmatrix} 1 & 0 \\ 0 & \theta \end{pmatrix}. \tag{1.1}$$

At $\theta = 0$, the exact equilibrium correspondence is a line segment, whereas opposite-sign local perturbations select singleton branches converging to different endpoints. Nevertheless, a fixed profile has residual $O(|\theta|)$. Proposition 1 proves that over $\{-c/\sqrt{T}, 0, c/\sqrt{T}\}$, one $O(T^{-1/2})$-residual profile is available without data, while every adaptive T-query estimator has order-one minimax Hausdorff error for the complete exact set. Thus pointwise approximate validity does not imply branch completeness.

Definition 1.1 (Exact-correspondence bifurcation)**.** Equip payoff matrices with the entrywise ℓ_∞ norm and equilibrium sets with Hausdorff distance induced by the product ℓ_1 strategy norm. A game $A_\star$ is an exact-correspondence bifurcation point if $A \mapsto \mathcal{E}_0(A)$ is not locally continuous at $A_\star$ under these metrics.

The title refers to this Hausdorff discontinuity; positive first-order resolution regularizes it.

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

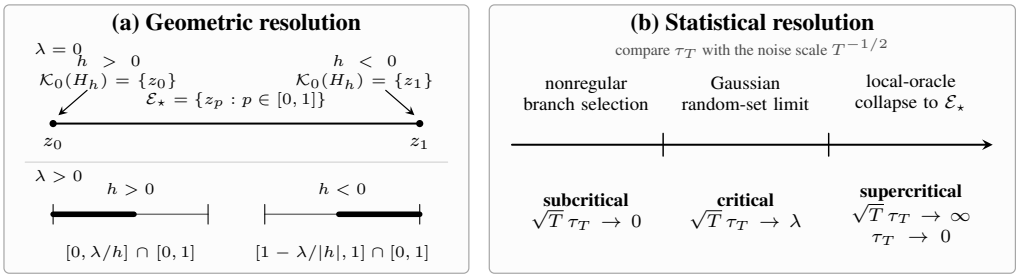

Figure 1: Guiding family. Opposite-sign exact perturbations converge to opposite endpoints of the base face. The sets $\mathcal{K}_0(H_h)$ are first-order outer envelopes, whereas positive resolution yields a clipped interval; comparing τ_T with $T^{-1/2}$ gives the three regimes in Corollary 5.

Fix a specified base game $A_\star$. A *limiting equilibrium profile* is a point $z \in \mathcal{E}_0(A_\star)$; we use *branch* only as shorthand. For a direction H, the branch-survival cost $\beta_\star(z; H)$ is the infimal first-order tolerance required for profiles converging to z to remain approximate equilibria of $A_\star + tH$. For $\lambda > 0$, its sublevel set $\mathcal{K}_\lambda(H)$ is the correspondence visible at resolution $t\lambda$. At $\lambda = 0$, $\mathcal{K}_0(H)$ is the zero-cost outer envelope and need not consist entirely of exact persistent branches. A direction H induces first-order branch selection when this outer envelope is a proper subset of $\mathcal{E}_0(A_\star)$. Theorem 2 proves complete uniform Hausdorff convergence at every positive resolution; at exact resolution only an outer limit is universal.

Our structural result identifies the game-specific derivative. If $\mathcal{X}_\star, \mathcal{Y}_\star$ are the base optimal-strategy polytopes and $z = (x, y)$, Theorem 3 gives

$$\beta_\star(z; H) = \max_{\mu \in \mathcal{X}_\star} \mu^\top H y - \min_{\nu \in \mathcal{Y}_\star} x^\top H \nu = \max_{C \in \mathcal{C}_\star(z)} \langle C, H \rangle_{\mathrm{F}}. \tag{1.2}$$

Moreover, $(\mu, \nu) \mapsto \mu y^\top - x\nu^\top$ is an affine bijection from the complete base correspondence onto $\mathcal{C}_\star(z)$. The contribution is therefore not merely the dualization of a tangent LP: equilibrium multiplicity is encoded exactly as the dual geometry governing first-order persistence.

Relation to prior theory. Classical matrix-game perturbation theory already identifies the one-sided derivative of the scalar value of $A_\star + tH$ through the perturbation game H restricted to the base optimal-strategy polytopes, and studies limits of exact perturbed optimal strategies (Mills, 1956; Altman et al., 2001). We do not claim novelty for that restricted game or scalar value derivative. The new content is a profile-wise epi-derivative of the saddle residual, uniform positive-resolution Hausdorff convergence, the affine critical-contrast representation, and its composition with a noisy payoff experiment under Hausdorff loss. Related work derives complete optimizer-set limits and confidence sets for random LPs and general minimax solution sets (Liu et al., 2023; Meitz, 2024; Meitz & Shapiro, 2025); noisy simulation-game, single-equilibrium, and complete-set methods study different observation models and estimands (Areyan Viqueira et al., 2019; 2020; Jiang & Zhang, 2026; Maiti et al., 2025; Feinstein et al., 2024). Appendix M gives the full comparison and records the corrigendum to Meitz (2024).

Under local alternatives $A_\star + H/\sqrt{T}$, known-variance Gaussian payoff observations, and deterministic full-support allocations, Theorem 4 compares the finite correspondence-learning risk with a Gaussian random-set decision problem. Corollary 5 identifies $T^{-1/2}$ as the critical resolution scale. Theorem 6 then shows that exact-set confidence procedures cannot adapt honestly across the guiding bifurcation, even under arbitrary adaptive querying.

Scope. The matching minimax comparison is fixed-dimensional, local oracle, known-variance, and deterministic-design. Predictable adaptive designs are used only for LAN and lower-bound statements. We do not prove unknown-center localization, a matching adaptive upper theorem, an operational rate for the strategic discrepancy, growing-action results, or arbitrary general-sum Nash correspondence theory.

Table 1: Selected closest neighboring theories and the theorem-level distinction.

Work	Established object	Distinction here
Mills; Altman et al.	Scalar value derivative and limits of exact perturbed optimal strategies	Profile-wise residual derivative, positive-resolution full-set limit, and statistical composition
Viqueira et al.; Jiang–Zhang	PAC precision–recall guarantees for all approximate equilibria of one fixed simulation game; or one selected equilibrium estimate under noisy feedback	Contiguous local alternatives crossing a bifurcation, exact-set Hausdorff loss, and tolerance at the $T^{-1/2}$ branch-resolution scale
Liu et al.; Meitz; Meitz–Shapiro	Nonunique optimizer/minimax solution sets and confidence or limit laws	Joint saddle correspondence and strategic contrast polytope affinely equivalent to the base set
Maiti et al.; Feinstein et al.	Exact-query recovery or deterministic inner–outer set approximation	Noise-scale tolerance $\lambda/\sqrt{T}$ as statistical branch resolution

2 Experiment and a Bifurcation Separation

Let $A \in [-1,1]^{m\times n}$ be the row player's payoff matrix in a finite zero-sum game, and let $\mathcal{Z} := \Delta_m \times \Delta_n$. The bounded domain is only a normalization for the global query model and the finite-sample example; the local variational results hold for every finite real payoff matrix. For any compact direction class $\mathcal{H}$, the local alternatives $A_\star + H/\sqrt{T}$ lie in one bounded neighborhood for all sufficiently large T. At round t, a randomized history-dependent design selects $I_t = (i_t, j_t)$ and observes

$$Y_t = A_{i_t j_t} + \xi_t, \qquad \xi_t \sim \mathcal{N}(0,\sigma^2), \tag{2.1}$$

with conditionally independent noise. Conditional on the observed history, the query kernel is parameter independent.

For $z = (x,y) \in \mathcal{Z}$, define the lower and upper security values, saddle residual, and complete resolution-τ correspondence by

$$\begin{aligned} \underline{v}_A(x) &:= \min_{j\in[n]} x^\top A e_j, & \overline{v}_A(y) &:= \max_{i\in[m]} e_i^\top A y, \\ g_A(x,y) &:= \overline{v}_A(y) - \underline{v}_A(x), & \mathcal{E}_\tau(A) &:= \{z \in \mathcal{Z} : g_A(z) \le \tau\}. \end{aligned} \tag{2.2}$$

Equip $\mathcal{Z}$ with $\|(x,y)\|_\vee := \|x\|_1 + \|y\|_1$, whose diameter is four. For nonempty compact sets $C, D \subseteq \mathcal{Z}$, write

$$e(C,D) := \sup_{z\in C} \operatorname{dist}(z,D), \qquad d_{\mathrm{H}}(C,D) := \max\{e(C,D), e(D,C)\}. \tag{2.3}$$

For an estimate $\widehat{C}$ of a target C, $e(\widehat{C}, C)$ is the validity error and $e(C,\widehat{C})$ is the completeness error.

For the family in (1.1), write $x(p) = (p, 1-p)^\top$, $y(q) = (q, 1-q)^\top$, and $z(p,q) = (x(p), y(q))$. Direct calculation gives

$$g_{A_\theta}(z(p,q)) = \max\{q, \theta(1-q)\} - \min\{p, \theta(1-p)\}, \tag{2.4}$$

and therefore

$$\mathcal{E}_0(A_\theta) = \begin{cases} \{z(1,0)\}, & \theta < 0, \\ \{z(p,0) : p \in [0,1]\}, & \theta = 0, \\ \{z(a_\theta, a_\theta)\}, & \theta > 0, \end{cases} \qquad a_\theta = \theta/(1+\theta). \tag{2.5}$$

The Hausdorff distance between the central segment and either limiting endpoint singleton equals two under the product ℓ_1 metric.

For $\Theta_T(c) := \{-c/\sqrt{T}, 0, c/\sqrt{T}\}$, let

$$\mathfrak{R}_T^{\mathrm{corr}}(c,\sigma) := \inf_{\mathsf{A},\widehat{C}_T} \sup_{\theta\in\Theta_T(c)} \mathbb{E}_\theta^{\mathsf{A}} d_{\mathrm{H}}\Big(\widehat{C}_T, \mathcal{E}_0(A_\theta)\Big), \tag{2.6}$$

where $\mathsf{A} = (\pi_t)_{t=1}^T$ ranges over randomized, history-dependent, parameter-independent query kernels, and $\widehat{C}_T$ ranges over randomized $\mathfrak{K}(\mathcal{Z})$-valued rules measurable with respect to the resulting transcript.

Proposition 1 (Finite-sample adaptive validity–completeness separation)**.** *For every $T \geq 1$, $\sigma > 0$, and $0 < c \leq \min\{1, \sigma\}$,*

$$1 - \frac{c}{2\sigma} \leq \mathfrak{R}_T^{\mathrm{corr}}(c, \sigma) \leq 1. \tag{2.7}$$

By contrast, the data-independent profile $\bar{z} = z(1/2, 0)$ satisfies

$$\sup_{\theta \in \Theta_T(c)} g_{A_\theta}(\bar{z}) = \frac{c}{2\sqrt{T}}, \qquad e(\{\bar{z}\}, \mathcal{E}_0(A_0)) = 0, \quad e(\mathcal{E}_0(A_0), \{\bar{z}\}) = 1. \tag{2.8}$$

Proof sketch. Use $\theta = 0$ and $\theta = c/\sqrt{T}$, whose targets are Hausdorff distance two. Only payoff $(2, 2)$ differs, so the adaptive KL chain rule gives

$$D_{\mathrm{KL}}(P_0 \| P_+) = \frac{c^2}{2\sigma^2 T} \mathbb{E}_0 N_{22}(T) \leq \frac{c^2}{2\sigma^2}. \tag{2.9}$$

Pinsker and data processing for randomized estimators imply total variation at most $c/(2\sigma)$. The metric two-point inequality then gives $\frac{1}{2} \cdot 2 \cdot (1 - c/(2\sigma))$, proving the lower bound. The constant segment midpoint gives the upper bound and (2.8). Appendix B proves the randomized metric lemma and the stronger two-sided endpoint statement. □

3 Strategic Epi-Resolution

Fix $A_\star \in \mathbb{R}^{m \times n}$, let $v_\star = \mathrm{val}(A_\star)$, and define

$$\begin{aligned} \mathcal{X}_\star &:= \{x \in \Delta_m : A_\star^\top x \geq v_\star \mathbf{1}_n\}, \qquad & \mathcal{Y}_\star &:= \{y \in \Delta_n : A_\star y \leq v_\star \mathbf{1}_m\}, \\ \mathcal{E}_\star &:= \mathcal{E}_0(A_\star) = \mathcal{X}_\star \times \mathcal{Y}_\star. \end{aligned} \tag{3.1}$$

A *limiting equilibrium profile* is a point $z \in \mathcal{E}_\star$; we use *branch* only as shorthand. For $\lambda > 0$, it is first-order persistent in direction H at resolution λ if there exist $t_k \downarrow 0$, $H_k \to H$, and $z_k \to z$ with $g_{A_\star + t_k H_k}(z_k) \leq t_k \lambda + o(t_k)$. Define the branch-survival cost and visible correspondence by

$$\beta_\star(z; H) := \inf_{\substack{t_k \downarrow 0, \ H_k \to H \\ z_k \to z}} \liminf_k \frac{g_{A_\star + t_k H_k}(z_k)}{t_k}, \qquad \mathcal{K}_\lambda(H) := \{z \in \mathcal{E}_\star : \beta_\star(z; H) \leq \lambda\}. \tag{3.2}$$

The limit is one-sided in t; both perturbation signs are represented by replacing H with $-H$. For $\lambda > 0$, points of $\mathcal{K}_\lambda(H)$ are called persistent or visible at resolution λ. At $\lambda = 0$, $\mathcal{K}_0(H)$ is the zero-cost outer envelope; its points need not all be exact persistent branches.

Theorem 2 (Positive-resolution strategic epi-resolution)**.** *Equip payoff directions with the entrywise ℓ_∞ norm and $\mathcal{Z}$ with the product ℓ_1 norm. For $t > 0$, set*

$$\mathcal{F}_t(H, z) := t^{-1} g_{A_\star + tH}(z), \qquad \mathcal{F}(H, z) := \begin{cases} \beta_\star(z; H), & z \in \mathcal{E}_\star, \\ +\infty, & z \notin \mathcal{E}_\star. \end{cases} \tag{3.3}$$

Then $\mathcal{F}$ is jointly lower semicontinuous and:

(a) $\mathcal{F}_t$ epi-converges jointly to $\mathcal{F}$ on $\mathbb{R}^{m \times n} \times \mathcal{Z}$ as $t \downarrow 0$.

(b) For every compact $\mathcal{H} \subset \mathbb{R}^{m \times n}$ and $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$,

$$\sup_{\substack{H \in \mathcal{H} \\ \lambda \in [\lambda_{\min}, \lambda_{\max}]}} d_{\mathrm{H}}(\mathcal{E}_{t\lambda}(A_\star + tH), \mathcal{K}_\lambda(H)) \longrightarrow 0. \tag{3.4}$$

(c) $(H, \lambda) \mapsto \mathcal{K}_\lambda(H)$ is Hausdorff continuous on $\mathbb{R}^{m \times n} \times (0, \infty)$.

(d) If $t_k \downarrow 0$ and $H_k \to H$, then

$$\limsup_k \mathcal{E}_0(A_\star + t_k H_k) \subseteq \mathcal{K}_0(H). \tag{3.5}$$

If $\mathcal{K}_0(H)$ is a singleton, the exact sets converge to it in Hausdorff distance.

The operational identity

$$\beta_\star(z; H) = \inf\left\{\lambda \geq 0 : z \in \liminf_{t \downarrow 0} \mathcal{E}_{t\lambda}(A_\star + tH)\right\} \tag{3.6}$$

is an infimal threshold: when it equals zero, exact-level persistence need not occur.

Proof architecture. Four ingredients isolate the nonregular point. First, the fixed-base Hoffman bound gives, uniformly for bounded H,

$$e(\mathcal{E}_{t\lambda}(A_\star + tH), \mathcal{E}_\star) \le t\kappa_\star(\lambda + 2\|H\|_\infty). \tag{3.7}$$

Second, for arbitrary $z_k \to z \in \mathcal{E}_\star$—even when $(z_k - z)/t_k$ is unbounded—the active-set support inequality yields

$$\frac{g_{A_\star + t_k H_k}(z_k)}{t_k} \ge \beta_\star(z; H_k) - \|H_k\|_\infty \|z_k - z\|_{\mathsf{V}}. \tag{3.8}$$

Third, approximate tangent minimizers are pasted over growing index blocks to obtain recovery along every prescribed sequence $t_k \downarrow 0$, $H_k \to H$. Finally, $z \mapsto \beta_\star(z; H)$ is continuous and convex on $\mathcal{E}_\star$ and has a zero-cost point. Interpolation with such a point makes strict λ-sublevels dense in closed λ-sublevels for every $\lambda > 0$; this last step fails at $\lambda = 0$. Appendices C–F prove these statements and the compact-uniform upgrade.

For $z = (x, y) \in \mathcal{E}_\star$, let $I_\star(y) = \arg\max_i e_i^\top A_\star y$ and $J_\star(x) = \arg\min_j x^\top A_\star e_j$. The Bouligand simplex tangent cone is

$$T_{\Delta_m}(x) = \big\{u : \mathbf{1}_m^\top u = 0,\ u_i \ge 0 \text{ whenever } x_i = 0\big\}, \tag{3.9}$$

with the analogous formula for y. Using auxiliary epigraph variables, the tangent problem is the finite LP

$$\begin{aligned} \min_{u,v,a,b} \quad & a + b \\ \text{s.t.} \quad & u \in T_{\Delta_m}(x), \quad v \in T_{\Delta_n}(y), \\ & e_i^\top A_\star v + e_i^\top H y \le a, \qquad i \in I_\star(y), \\ & -u^\top A_\star e_j - x^\top H e_j \le b, \qquad j \in J_\star(x). \end{aligned} \tag{3.10}$$

We use $K^+ := \{w : \langle w, d\rangle \ge 0\ \forall d \in K\}$ and $K^\circ = -K^+$. The dual probability variables satisfy

$$\mu \in \Delta_m(I_\star(y)), \quad \nu \in \Delta_n(J_\star(x)), \quad A_\star^\top \mu \in T_{\Delta_n}(y)^+, \quad -A_\star \nu \in T_{\Delta_m}(x)^+. \tag{3.11}$$

These certificate sets turn out to be the full optimal-strategy polytopes.

Theorem 3 (Tangent programs and critical strategic contrasts)**.** *For every $z = (x, y) \in \mathcal{E}_\star$ and $H \in \mathbb{R}^{m\times n}$,*

$$\begin{aligned} \beta_\star(z; H) &= \min_{\substack{u \in T_{\Delta_m}(x) \\ v \in T_{\Delta_n}(y)}} \left[\max_{i \in I_\star(y)} \{e_i^\top H y + e_i^\top A_\star v\} - \min_{j \in J_\star(x)} \{x^\top H e_j + u^\top A_\star e_j\}\right] \\ &= \max_{\mu \in \mathcal{X}_\star} \mu^\top H y - \min_{\nu \in \mathcal{Y}_\star} x^\top H \nu = \max_{C \in \mathcal{C}_\star(z)} \langle C, H\rangle_{\mathrm{F}}, \end{aligned} \tag{3.12}$$

where

$$\mathcal{C}_\star(z) := \big\{\mu y^\top - x\nu^\top : \mu \in \mathcal{X}_\star,\ \nu \in \mathcal{Y}_\star\big\}. \tag{3.13}$$

All three optimization problems attain their extrema; $\mathcal{C}_\star(z)$ is a compact polytope containing zero. The displayed map is an affine bijection from $\mathcal{E}_\star$ onto $\mathcal{C}_\star(z)$, with inverse

$$\mu = x + C\mathbf{1}_n, \qquad \nu = y - C^\top \mathbf{1}_m. \tag{3.14}$$

Consequently,

$$\dim \mathcal{C}_\star(z) = \dim \mathcal{E}_\star, \qquad \operatorname{rank}(C) \le 2, \qquad \|C\|_{1,\text{entrywise}} \le 2. \tag{3.15}$$

The restricted first-order game in the second line of (3.12) is classical in scalar value-sensitivity theory (Mills, 1956; Altman et al., 2001). The new content is its identification as the complete profile-wise branch-cost field, the positive-resolution correspondence limit, and the affine critical-contrast geometry.

Table 2: Computational status of the retained local-oracle objects.

Object	Representation and basic operation	Status
$\mathcal{E}_\star$	Product of two optimal-strategy polytopes; LP feasibility and optimization	Polynomial LP oracle; complete vertex output may be exponential
$\beta_\star(z;H)$	Tangent LP or two restricted-game LPs	Polynomial time for fixed rational $(A_\star, z, H)$
$\mathcal{K}_\lambda(H)$	Lifted polyhedron in (3.17)	Polynomial membership, separation, and linear-optimization oracles
Gaussian near-minimax rule	Decision-theoretic optimization over measurable rules with compact action space $\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$	Existential finite-range approximation; no polynomial-time global rule claimed

Computational status. For fixed $A_\star, z, H$, $\beta_\star(z;H)$ is obtained either from (3.10) or from two LPs over $\mathcal{X}_\star, \mathcal{Y}_\star$. For fixed H, λ, write

$$U_H(y) := \max_{\mu\in\mathcal{X}_\star} \mu^\top H y, \qquad L_H(x) := \min_{\nu\in\mathcal{Y}_\star} x^\top H \nu. \tag{3.16}$$

Then $\mathcal{K}_\lambda(H) = \{(x,y) \in \mathcal{E}_\star : U_H(y) - L_H(x) \le \lambda\}$. Strong LP duality gives the explicit lifted formulation

$$\mathcal{K}_\lambda(H) = \mathrm{proj}_{x,y} \left\{ (x,y,s,r,\alpha,\gamma) : \begin{array}{l} x \in \mathcal{X}_\star,\ y \in \mathcal{Y}_\star,\ s \ge 0,\ r \ge 0, \\ \alpha \mathbf{1}_m - A_\star s \ge H y, \\ \gamma \mathbf{1}_n - A_\star^\top r \le H^\top x, \\ \alpha - v_\star \mathbf{1}_n^\top s - \gamma + v_\star \mathbf{1}_m^\top r \le \lambda \end{array} \right\}. \tag{3.17}$$

For rational input data under standard LP bit-complexity assumptions, membership and separation require two restricted-game LPs, and linear optimization is an LP in the lifted variables. An explicit projected vertex or facet description can nevertheless be output-exponential.

Example 3.1 (Resolved branches in the guiding bifurcation)**.** For $A_\star = A_0$, $H_h = h e_2 e_2^\top$, and $z_p = (x(p), e_2)$,

$$\beta_\star(z_p; H_h) = h_+ p + h_-(1-p), \tag{3.18}$$

so

$$\mathcal{K}_\lambda(H_h) = \begin{cases} \{z_p : 0 \le p \le \min(1, \lambda/h)\}, & h > 0, \\ \mathcal{E}_\star, & h = 0, \\ \{z_p : \max(0, 1 - \lambda/|h|) \le p \le 1\}, & h < 0. \end{cases} \tag{3.19}$$

At $\lambda = 0$, the zero-cost outer envelopes for opposite signs are the opposite endpoints and inner convergence fails at $h = 0$; these are not the finite-t exact singleton locations. Every fixed $\lambda > 0$ yields Hausdorff continuity. Thus the positive-resolution restriction in Theorem 2 is sharp.

4 Gaussian Comparison and the Resolution Trichotomy

Fix a compact direction class $\mathcal{H} \subset \mathbb{R}^{m\times n}$, set $d = mn$, and consider local games

$$A_{T,H} := A_\star + H/\sqrt{T}, \qquad H \in \mathcal{H}. \tag{4.1}$$

For a deterministic sampling allocation $q_T \in \mathrm{ri}\,\Delta_d$, fix a canonical parameter-independent query order realizing its coordinate counts and assume

$$q_{a,T} > 0, \qquad T q_{a,T} \in \mathbb{N}, \qquad \sum_{a=1}^{d} T q_{a,T} = T. \tag{4.2}$$

If $\widehat{A}_T$ contains the entrywise sample means, then

$$W_T := \sqrt{T}\ \mathrm{vec}(\widehat{A}_T - A_\star) \sim \mathcal{N}(\mathrm{vec}(H), \Sigma(q_T)), \qquad \Sigma(q) := \sigma^2\, \mathrm{diag}(q)^{-1}. \tag{4.3}$$

Appendix G proves exact finite-sample equivalence between the complete deterministic-design transcript and W_T: the forward kernel takes sample means, and a parameter-free reverse kernel restores the centered Gaussian residuals around those means.

For $\lambda > 0$, define the finite and limiting targets

$$\mathcal{T}_{T,\lambda}(H) := \mathcal{E}_{\lambda/\sqrt{T}}(A_{T,H}), \qquad \mathcal{T}_{\lambda}(H) := \mathcal{K}_{\lambda}(H). \tag{4.4}$$

For a Borel randomized decision rule δ with values in $\mathfrak{K}(\mathcal{Z})$, let

$$\mathfrak{R}_{T,q_T}(\mathcal{H},\lambda) := \inf_{\delta} \sup_{H\in\mathcal{H}} \mathbb{E}_H^{q_T} d_{\mathrm{H}}(\delta(W_T), \mathcal{T}_{T,\lambda}(H)), \tag{4.5}$$

$$\mathfrak{R}_{q}^{\mathrm{G}}(\mathcal{H},\lambda) := \inf_{\delta} \sup_{H\in\mathcal{H}} \mathbb{E}_H^{q} d_{\mathrm{H}}(\delta(W), \mathcal{T}_{\lambda}(H)), \quad W \sim \mathcal{N}(\mathrm{vec}(H), \Sigma(q)). \tag{4.6}$$

The decisions may be randomized, but the targets are compact and convex. Introduce the strategic and experiment discrepancies

$$\omega_T(\mathcal{H},\lambda) := \sup_{H\in\mathcal{H}} d_{\mathrm{H}}\big(\mathcal{T}_{T,\lambda}(H), \mathcal{T}_{\lambda}(H)\big), \tag{4.7}$$

$$\zeta_T(q_T, q) := d_{\mathrm{TV}}(\mathcal{N}(0,\Sigma(q_T)), \mathcal{N}(0,\Sigma(q))). \tag{4.8}$$

Theorem 2 gives $\omega_T \to 0$ uniformly on compact $\mathcal{H}$, but no finite-sample rate is asserted. Gaussian covariance comparison gives the explicit bound

$$\zeta_T(q_T, q) \le \frac{1}{2}\left[\sum_{a=1}^{d} \phi\left(\frac{q_a}{q_{a,T}}\right)\right]^{1/2}, \qquad \phi(r) := r - 1 - \log r. \tag{4.9}$$

Theorem 4 (Finite-sample Gaussian comparison and local minimax limit)**.** *Assume fixed finite m, n, a specified base game $A_\star$, known σ^2, compact $\mathcal{H}$, fixed $\lambda > 0$, deterministic integer counts satisfying* (4.2)*, and full-support allocations $q_T, q \in \mathrm{ri}\,\Delta_d$ with $q_T \to q$. Then*

$$\mathfrak{R}_{T,q_T}(\mathcal{H},\lambda) \ge \mathfrak{R}_{q}^{\mathrm{G}}(\mathcal{H},\lambda) - \omega_T(\mathcal{H},\lambda) - 4\zeta_T(q_T,q), \tag{4.10}$$

$$\mathfrak{R}_{T,q_T}(\mathcal{H},\lambda) \le \mathfrak{R}_{q}^{\mathrm{G}}(\mathcal{H},\lambda) + \omega_T(\mathcal{H},\lambda) + 4\zeta_T(q_T,q). \tag{4.11}$$

Consequently,

$$\mathfrak{R}_{T,q_T}(\mathcal{H},\lambda) \longrightarrow \mathfrak{R}_{q}^{\mathrm{G}}(\mathcal{H},\lambda). \tag{4.12}$$

A sequence of Borel deterministic finite-range rules with values in $\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$ approaches the Gaussian benchmark up to $o(1)$. The factor four is the diameter of $\mathcal{Z} = \Delta_m \times \Delta_n$ under the product ℓ_1 metric and hence an upper bound on Hausdorff loss. Balanced rounding—flooring Tq_a and assigning the remaining samples to the largest fractional parts—supplies (4.2) *for every $q \in \mathrm{ri}\,\Delta_d$ and satisfies $\|q_T - q\|_\infty \le T^{-1}$.*

Proof sketch. Exact sufficiency first replaces the transcript by W_T without loss. Target replacement changes every loss by at most ω_T. Translation invariance of total variation and the loss bound four change every risk by at most $4\zeta_T$. These two transfers yield the lower inequality. For the upper inequality, random compact actions are convexified, conditional support functions are averaged to obtain a deterministic compact-convex barycenter, and a finite Hausdorff net gives an η_T-minimax finite-range rule with $\eta_T \downarrow 0$. Appendices G– I prove the reverse kernel, covariance comparison, measurability, and derandomization steps. □

Why the experiment reduction is exact. Let $N_{a,T} = Tq_{a,T}$ be the number of observations of payoff coordinate a. The sample mean is sufficient for its Gaussian mean. Conversely, from $W_{T,a}$ one reconstructs a transcript block by setting

$$\widetilde{Y}_a = \left(A_{\star,a} + \frac{W_{T,a}}{\sqrt{T}}\right)\mathbf{1}_{N_{a,T}} + \left(I - \frac{1}{N_{a,T}}\mathbf{1}\mathbf{1}^\top\right) Z_a, \qquad Z_a \sim \mathcal{N}(0,\sigma^2 I). \tag{4.13}$$

The mean component has covariance $\sigma^2\mathbf{1}\mathbf{1}^\top/N_{a,T}$ and the residual component has covariance $\sigma^2(I - \mathbf{1}\mathbf{1}^\top/N_{a,T})$; their sum is $\sigma^2 I$. This parameter-free reverse kernel proves zero Le Cam distance for every T. The only limiting operations in Theorem 4 are therefore the strategic target replacement and $q_T \to q$.

Why randomization is unnecessary asymptotically. After convexifying a random reported set, let η_w be its conditional law given the Gaussian observation w. Define a compact convex barycenter by its support function

$$h_{\bar{C}(w)}(u) := \int h_C(u)\, \eta_w(dC). \tag{4.14}$$

The Hausdorff support-function identity and Jensen's inequality give, for every compact convex target D,

$$d_{\mathrm{H}}(\bar{C}(w), D) \le \int d_{\mathrm{H}}(C, D)\, \eta_w(dC). \tag{4.15}$$

A measurable projection onto a finite Hausdorff net of $\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$ then produces the deterministic finite-range sequence in the theorem. This is an existence construction for decision theory, not an efficient algorithm for a globally minimax rule.

This is a decision-theoretic achievability theorem, not a polynomial-time construction of a globally minimax random-set rule. The sample means, $\beta_\star(z; H)$, and LP membership in $\mathcal{K}_\lambda(H)$ are computable; a near-minimax Gaussian random-set rule is existential. The matching upper and lower comparison is deterministic-design. Predictable adaptive designs receive LAN and one-sided lower-bound results only when their normalized information converges to a stable deterministic full-support limit.

Let $\tau_T \downarrow 0$ and $\lambda_T = \sqrt{T}\,\tau_T$. Define the Mahalanobis separation

$$\mathfrak{d}_q(H_0, H_1)^2 := \mathrm{vec}(H_1 - H_0)^\top \sigma^{-2}\, \mathrm{diag}(q)\, \mathrm{vec}(H_1 - H_0). \tag{4.16}$$

Corollary 5 (Statistical resolution trichotomy)**.** *Let $q_T \to q \in \mathrm{ri}\,\Delta_d$ and let $\mathcal{H}$ be compact.*

(i) If $\lambda_T \to 0$ and $\mathcal{H}$ contains directions H_0, H_1 with separated singleton zero-cost targets $\mathcal{K}_0(H_r) = \{z_r\}$, then

$$\liminf_T \mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) \ge \|z_0 - z_1\|_\vee \Phi(-\mathfrak{d}_q(H_0, H_1)/2) > 0. \tag{4.17}$$

(ii) If $\lambda_T \to \lambda \in (0, \infty)$, then

$$\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) \longrightarrow \mathfrak{R}^{\mathrm{G}}_q(\mathcal{H}, \lambda). \tag{4.18}$$

(iii) If $\tau_T \to 0$ and $\lambda_T \to \infty$, then

$$\sup_{H \in \mathcal{H}} d_{\mathrm{H}}\Big(\mathcal{E}_{\tau_T}(A_\star + H/\sqrt{T}), \mathcal{E}_\star\Big) \le \kappa_\star \left(\tau_T + 2R_{\mathcal{H}}/\sqrt{T}\right) \longrightarrow 0, \tag{4.19}$$

where $R_{\mathcal{H}} = \sup_{H \in \mathcal{H}} \|H\|_\infty$.

Part (i) is conditional: it is not a universal conclusion at every nonunique game. In the guiding family, if there exists a transverse direction H_0 such that $\{\pm a H_0 : 0 < a \le a_0\} \subseteq \mathcal{H}$, then the visible singletons remain distance two apart while their Mahalanobis separation tends to zero as $a \downarrow 0$, strengthening the lower bound to one. Its proof combines convergence of the two finite targets to $\{z_0\}$ and $\{z_1\}$ with the exact common-covariance Gaussian identity

$$1 - d_{\mathrm{TV}}\left(\mathcal{N}(H_0, \Sigma(q)), \mathcal{N}(H_1, \Sigma(q))\right) = 2\Phi(-\mathfrak{d}_q(H_0, H_1)/2)\,. \tag{4.20}$$

Part (ii) uses Theorem 2 uniformly on a compact positive interval containing all sufficiently large λ_T. In part (iii), eventually every $z \in \mathcal{E}_\star$ has perturbed residual at most $2R_{\mathcal{H}}/\sqrt{T} \le \tau_T$, while the reverse excess follows from the fixed-base Hoffman bound. Thus the conclusion is uniform local-oracle collapse to the known base correspondence, not global recovery of an unknown game. Appendix J contains the complete testing, varying-level, and localization arguments.

5 Exact-Set Honest Nonadaptivity

The unknown parameter is one compact equilibrium set in $\mathcal{S} := \mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$. A confidence procedure returns a random compact collection $\mathfrak{C}_T \in \mathfrak{K}(\mathcal{S})$; its epistemic diameter is

$$\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T) := \sup_{C, D \in \mathfrak{C}_T} d_{\mathrm{H}}(C, D). \tag{5.1}$$

Thus the intrinsic diameter of one known equilibrium face is not charged; only uncertainty among candidate sets is. The two hyperspace levels are

$$\underbrace{S_T(\theta) = \mathcal{E}_0(A_\theta) \in \mathcal{S}}_{\text{one set-valued parameter}}, \qquad \underbrace{\mathfrak{C}_T \in \mathfrak{K}(\mathcal{S})}_{\text{a collection of candidate sets}}, \tag{5.2}$$

and honesty means that the true compact set belongs to the reported collection with high probability. This distinction matters at $\theta = 0$: the true equilibrium segment is intrinsically large, but a singleton confidence collection containing that known segment would have epistemic diameter zero.

Theorem 6 (Finite-sample adaptive lower bound for exact-set honesty)**.** *For the family A_θ in* (1.1)*, fix $0 < \bar{c} \le 1$, $0 < \alpha < 1/2$, and $\Theta_T(\bar{c}) = \{\theta : |\theta| \le \bar{c}/\sqrt{T}\}$. Under any randomized adaptive query strategy, every measurable confidence collection satisfying*

$$\inf_{\theta \in \Theta_T(\bar{c})} P_{T,\theta}(\mathcal{E}_0(A_\theta) \in \mathfrak{C}_T) \ge 1 - \alpha \tag{5.3}$$

obeys, under A_0,

$$P_{T,0}(\operatorname{diam}_{\mathrm{H}}(\mathfrak{C}_T) \ge 2) \ge 1 - 2\alpha, \qquad \mathbb{E}_{T,0} \operatorname{diam}_{\mathrm{H}}(\mathfrak{C}_T) \ge 2(1 - 2\alpha). \tag{5.4}$$

Proof sketch. For $\theta_{T,\pm} = \pm c/\sqrt{T}$, the exact targets are Hausdorff distance two, while each adaptive transcript law is within $c/(2\sigma)$ total variation of the base law. Let $E_\pm$ be the two coverage events. Total-variation transport gives $P_{T,0}(E_\pm^c) \le \alpha + c/(2\sigma)$, and therefore

$$P_{T,0}(E_+ \cap E_-) \ge 1 - 2\alpha - c/\sigma. \tag{5.5}$$

On this intersection, $\mathfrak{C}_T$ contains two sets at Hausdorff distance two, so its epistemic diameter is at least two. The event on the left does not depend on the auxiliary c; letting $c \downarrow 0$ proves the probability bound, and nonnegativity gives the expectation bound. Appendix K proves all hyperspace measurability statements. □

Why the obstruction is exact-set specific. The result concerns exact correspondences in the displayed family and uniform honesty over a $T^{-1/2}$ neighborhood crossing its bifurcation. It is not a generic impossibility theorem for every nonunique game. The exact target jumps by Hausdorff distance two under contiguous alternatives, while for every fixed $\lambda > 0$ the visible map $(H, \lambda) \mapsto \mathcal{K}_\lambda(H)$ is Hausdorff continuous. Accordingly, Appendix J constructs simultaneous finite-sample inner–outer confidence corridors for positive approximation levels. Appendix K shows that the exact-set lower bound also forces any measurable Hausdorff confidence ball in $\mathcal{S}$ to have order-one radius at the bifurcation. Positive resolution does not erase statistical uncertainty; it changes the estimand to a topologically stable one.

6 Conclusion

We established a first-order resolution theory for complete zero-sum equilibrium correspondences near exact-set bifurcations. A data-independent $O(T^{-1/2})$-residual profile can coexist with order-one Hausdorff difficulty for the exact set, while positive-resolution correspondences converge to sublevel sets of a branch-survival cost. Its critical-contrast polytope is an affine copy of the base correspondence, converting multiplicity into the dual geometry that determines statistical visibility. Under the stated fixed-dimensional, known-variance, deterministic-design local experiment, this geometry yields the Gaussian random-set minimax limit and the resolution trichotomy; exact-set honest adaptation fails across the exhibited family even under adaptive querying. We do not prove global localization, a matching adaptive upper theorem, a rate for ω_T, unknown-variance or non-Gaussian equivalence, growing-action guarantees, or arbitrary general-sum Nash results. Natural next problems are globally implementable positive-resolution inference, computationally explicit near-minimax rules, higher-order silent directions, and extensions beyond polyhedral zero-sum correspondences.

AI USE STATEMENT

Generative AI tools were used for literature discovery, conceptual stress testing, formulation and critique of candidate mathematical claims, proof drafting, manuscript editing, reference formatting, LaTeX preparation, and preparation of regression-test code. Retained AI-suggested citations were checked for substantive relevance; the closest-work, recent-work, and corrected bibliographic records were audited against primary or publisher sources. The anonymous regression suite and rendered-PDF audit were rerun after the final source changes. These checks are counterexample searches and consistency checks, not formal proofs and not substitutes for expert mathematical review. No claim or proof step is retained solely on the basis of model output. The authors take responsibility for the complete submission.

REPRODUCIBILITY STATEMENT

Complete proofs of Proposition 1, Theorems 2–4 and 6, and Corollary 5 appear in Appendices B–K. Appendix A fixes all norms, topologies, cone conventions, hyperspace measurability structures, and modes of convergence. The anonymous supplementary package contains deterministic symbolic and numerical regression tests for the guiding bifurcation, tangent–dual equality, positive-resolution convergence, Gaussian comparison constants, and support-function barycenter inequalities. These tests search for counterexamples and implementation inconsistencies; they are not formal proofs. The paper contains no empirical datasets or experimental claims.

REFERENCES

Eitan Altman, Eugene A. Feinberg, Jerzy A. Filar, and Vladimir A. Gaitsgory. Perturbed zero-sum games with applications to stochastic and repeated games. In Eitan Altman and Odile Pourtallier (eds.), *Advances in Dynamic Games and Applications*, volume 6 of *Annals of the International Society of Dynamic Games*, pp. 165–181. Birkhäuser, 2001. doi: 10.1007/978-1-4612-0155-7_11.

Enrique Areyan Viqueira, Cyrus Cousins, Eli Upfal, and Amy Greenwald. Learning equilibria of simulation-based games. *arXiv preprint arXiv:1905.13379*, 2019. doi: 10.48550/arXiv.1905.13379.

Enrique Areyan Viqueira, Cyrus Cousins, and Amy Greenwald. Improved algorithms for learning equilibria in simulation-based games. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems*, pp. 79–87. International Foundation for Autonomous Agents and Multiagent Systems, 2020.

Annalisa Barbara, Riccardo Poiani, Martino Bernasconi, and Andrea Celli. Optimal rates for feasible payoff set estimation in games. *arXiv preprint arXiv:2602.04397*, 2026. doi: 10.48550/arXiv.2602.04397.

J. Frédéric Bonnans and Alexander Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, New York, 2000. doi: 10.1007/978-1-4612-1394-9.

Clément L. Canonne. A short note on an inequality between KL and TV. *arXiv preprint arXiv:2202.07198*, 2022. doi: 10.48550/arXiv.2202.07198.

John M. Danskin. *The Theory of Max-Min and Its Application to Weapons Allocation Problems*, volume 5 of *Econometrics and Operations Research*. Springer, Berlin, Heidelberg, 1967. doi: 10.1007/978-3-642-46092-0.

Ruoyu Diao, Yu-Hong Dai, and Liwei Zhang. Stability for Nash equilibrium problems. *Mathematics of Operations Research*, 51(3):2017–2040, 2025. doi: 10.1287/moor.2024.0609.

Zachary Feinstein, Niklas Hey, and Birgit Rudloff. Approximating the set of Nash equilibria for convex games. *Operations Research*, 73(5):2729–2743, 2024. doi: 10.1287/opre.2023.0541.

Peter Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics. Academic Press, New York, 1980. ISBN 978-0-12-319350-6.

Alan J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49(4):263–265, 1952. doi: 10.6028/JRES.049.027.

Matthew O. Jackson, Tomás Rodríguez-Barraquer, and Xu Tan. Epsilon-equilibria of perturbed games. *Games and Economic Behavior*, 75(1):198–216, 2012. doi: 10.1016/j.geb.2011.09.007.

Jiashuo Jiang and Mengxiao Zhang. An LP-based approach for bilinear saddle point problem with instance-dependent guarantee and noisy feedback. *arXiv preprint arXiv:2602.12513*, 2026. doi: 10.48550/arXiv.2602.12513.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016. URL https://jmlr.org/papers/v17/kaufman16a.html.

Marcel Klatt, Axel Munk, and Yoav Zemel. Limit laws for empirical optimal solutions in random linear programs. *Annals of Operations Research*, 315(1):251–278, 2022. doi: 10.1007/s10479-022-04698-0.

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.

Shuyu Liu, Florentina Bunea, and Jonathan Niles-Weed. Asymptotic confidence sets for random linear programs. In *Proceedings of the 36th Annual Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 3919–3940. PMLR, 2023. URL https://proceedings.mlr.press/v195/liu23d.html.

Mark G. Low. On nonparametric confidence intervals. *The Annals of Statistics*, 25(6):2547–2554, 1997. doi: 10.1214/aos/1030741084.

Shu Lu and Amarjit Budhiraja. Confidence regions for stochastic variational inequalities. *Mathematics of Operations Research*, 38(3):545–568, 2013. doi: 10.1287/moor.1120.0579.

Arnab Maiti, Ross Boczar, Kevin Jamieson, and Lillian J. Ratliff. Query-efficient algorithm to find all Nash equilibria in a two-player zero-sum matrix game. *ACM Transactions on Economics and Computation*, 13(3):1–18, 2025. doi: 10.1145/3732787.

Mika Meitz. Statistical inference for generative adversarial networks and other minimax problems. *Scandinavian Journal of Statistics*, 51(3):1323–1356, 2024. doi: 10.1111/sjos.12710.

Mika Meitz. Corrigendum to "statistical inference for generative adversarial networks and other minimax problems". *Scandinavian Journal of Statistics*, 52(3):1477–1478, 2025. doi: 10.1111/sjos.12793.

Mika Meitz and Alexander Shapiro. Minimax asymptotics. *Electronic Journal of Statistics*, 19(2):5117–5146, 2025. doi: 10.1214/25-EJS2457.

Harlan D. Mills. Marginal values of matrix games and linear programs. In Harold W. Kuhn and Albert W. Tucker (eds.), *Linear Inequalities and Related Systems*, volume 38 of *Annals of Mathematics Studies*, pp. 183–193. Princeton University Press, Princeton, NJ, 1956.

Ilya Molchanov. *Theory of Random Sets*, volume 87 of *Probability Theory and Stochastic Modelling*. Springer, London, 2 edition, 2017. doi: 10.1007/978-1-4471-7349-6.

Javier F. Peña, Juan C. Vera, and Luis F. Zuluaga. New characterizations of Hoffman constants for systems of linear constraints. *Mathematical Programming*, 187(1–2):79–109, 2021. doi: 10.1007/s10107-020-01473-6.

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.

Rolf Schneider. *Convex Bodies: The Brunn–Minkowski Theory*, volume 151 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2 edition, 2014. doi: 10.1017/CBO9781139003858.

Alexander Schrijver. *Theory of Linear and Integer Programming*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Chichester, 1986. ISBN 978-0-471-90854-8.

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

Kyungchul Song. Local asymptotic minimax estimation of nonregular parameters with translation-scale equivariant maps. *Journal of Multivariate Analysis*, 125:136–158, 2014. doi: 10.1016/j.jmva.2013.12.003.

Kenta Takatsu and Arun Kumar Kuchibhotla. Generalized van trees inequality: Local minimax bounds for non-smooth functionals and irregular statistical models. *arXiv preprint arXiv:2405.06437*, 2024. doi: 10.48550/arXiv.2405.06437.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. doi: 10.1007/b13794.

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

John von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928. doi: 10.1007/BF01448847.

Liner Xiang, Yixin Wang, and Hengrui Cai. Policy optimization and statistical inference for online contextual matrix games. *arXiv preprint arXiv:2608.17173*, 2026. doi: 10.48550/arXiv.2608.17173.

APPENDIX ROADMAP AND THEOREM DEPENDENCIES

The appendices are ordered by proof dependency. Appendix A fixes all norms, topologies, cones, and hyperspace measurability conventions; later appendices do not change them.

Block	Mathematical role	Main result
A	Norms, hyperspaces, epi-convergence, cones, measurability	Common infrastructure
B	Guiding-family geometry, adaptive KL, metric testing	Proposition 1
C–D	Hoffman localization, perturbation, directional liminf and recovery	Analytic foundation of Theorem 2
E	Tangent LP, strong duality, equilibrium certificates, affine contrasts	Theorem 3
F	Joint epi-limit, positive sublevels, exact-level boundary	Theorem 2
G–I	Gaussian sufficiency, lower transfer, derandomization, upper transfer	Theorem 4
J	Resolution regimes and confidence corridors	Corollary 5
K	Hyperspace confidence collections and coverage transport	Theorem 6
L	Dependency and verification notes	Proof map and regression-test description
M	Related work and boundary cases	Novelty and scope

$$\begin{aligned}&\text{residual perturbation} \longrightarrow \text{fixed-base Hoffman localization} \longrightarrow \text{directional residual calculus}\\ &\quad\longrightarrow \text{unbounded-velocity liminf and prescribed recovery} \longrightarrow \text{tangent primal–dual identity}\\ &\quad\longrightarrow \text{critical contrasts} \longrightarrow \text{joint epi-resolution} \longrightarrow \text{positive-level Hausdorff convergence}\\ &\quad\longrightarrow \text{Gaussian experiment comparison} \longrightarrow \text{minimax limit and resolution trichotomy.}\end{aligned} \tag{.1}$$

Positive-resolution set convergence additionally uses strict-sublevel density and compactness; epi-convergence alone is not used as a substitute for level-set regularity. The exact-resolution outer limit is proved separately. The matching minimax comparison is deterministic-design and known-variance; adaptive querying is used for lower bounds and honest nonadaptivity.

A NOTATION AND VARIATIONAL PRELIMINARIES

This appendix fixes the ambient spaces, norms, set-convergence conventions, and measurability structures used throughout. Our conventions for Painlevé–Kuratowski and epigraphical convergence, tangent and normal cones, and set-valued mappings follow Rockafellar & Wets (1998); our random-compact-set conventions follow Molchanov (2017). We include the elementary finite-dimensional arguments needed later so that no topological, metric, or cone-duality convention remains implicit.

A.1 NOTATION SUMMARY

Symbol	Meaning
$g_A(z)$	saddle residual of game A at profile z
$\mathcal{E}_\tau(A)$	complete resolution-τ correspondence
$\mathcal{E}_\star$	exact correspondence of the specified base game $A_\star$
$\mathcal{X}_\star, \mathcal{Y}_\star$	row and column optimal-strategy polytopes
$\mathcal{V}_X^\star, \mathcal{V}_Y^\star$	their vertex sets, when used
$\beta_\star(z; H)$	branch-survival cost in direction H
$\mathcal{K}_\lambda(H)$	visible correspondence at first-order resolution λ
$\mathcal{C}_\star(z)$	critical-contrast polytope at base profile z
q, q_T	limiting and finite deterministic sampling allocations
$\mathfrak{R}_{T,q_T}, \mathfrak{R}_q^{\mathrm{G}}$	finite and Gaussian minimax risks

Throughout, H denotes one payoff direction and $\mathcal{H}$ a class of directions. A polytope and its vertex set are never denoted by the same symbol.

A.2 AMBIENT SPACES, NORMS, AND MATRIX GAMES

For a positive integer r, write

$$[r] := \{1, \dots, r\},$$

let $e_i \in \mathbb{R}^r$ denote the ith standard basis vector, and let $\mathbf{1}_r \in \mathbb{R}^r$ denote the all-ones vector. The probability simplex is

$$\Delta_r := \{p \in \mathbb{R}^r_+ : \mathbf{1}_r^\top p = 1\}. \tag{A.1}$$

We work on the finite-dimensional vector spaces

$$\mathsf{V} := \mathbb{R}^m \times \mathbb{R}^n, \qquad \mathsf{M} := \mathbb{R}^{m\times n}, \qquad \mathcal{Z} := \Delta_m \times \Delta_n. \tag{A.2}$$

The bilinear pairings on V and M are

$$\langle (a,b),(x,y)\rangle_{\mathsf{V}} := a^\top x + b^\top y, \qquad \langle C, H\rangle_{\mathrm{F}} := \operatorname{tr}(C^\top H). \tag{A.3}$$

We equip V with the product ℓ_1 norm and M with the entrywise maximum norm,

$$\|(x,y)\|_{\mathsf{V}} := \|x\|_1 + \|y\|_1, \qquad \|A\|_\infty := \max_{\substack{i\in[m]\\ j\in[n]}} |A_{ij}|. \tag{A.4}$$

Unless explicitly stated otherwise, all distances on $\mathcal{Z}$ are induced by $\|\cdot\|_{\mathsf{V}}$.

Lemma A.1 (Dual of the product norm)**.** *The dual norm of $\|\cdot\|_{\mathsf{V}}$ is*

$$\|(a,b)\|_{\mathsf{V},*} = \max\{\|a\|_\infty, \|b\|_\infty\}. \tag{A.5}$$

Proof. For every $(x,y) \in \mathsf{V}$, Hölder's inequality gives

$$\left|a^\top x + b^\top y\right| \le \|a\|_\infty \|x\|_1 + \|b\|_\infty \|y\|_1 \le \max\{\|a\|_\infty, \|b\|_\infty\}\|(x,y)\|_{\mathsf{V}}.$$

Equality is attained by allocating the entire unit-norm budget to a signed coordinate attaining the larger of $\|a\|_\infty$ and $\|b\|_\infty$. □

A matrix $A \in \mathsf{M}$ specifies a two-player zero-sum game: the row player chooses $x \in \Delta_m$ and maximizes $x^\top Ay$, while the column player chooses $y \in \Delta_n$ and minimizes the same quantity. The finite-game minimax theorem (von Neumann, 1928) gives

$$\mathrm{val}(A) := \max_{x\in\Delta_m} \min_{y\in\Delta_n} x^\top Ay = \min_{y\in\Delta_n} \max_{x\in\Delta_m} x^\top Ay. \tag{A.6}$$

For $z = (x, y) \in \mathcal{Z}$, define the saddle residual

$$\begin{aligned} g_A(z) &:= \max_{u\in\Delta_m} u^\top Ay - \min_{v\in\Delta_n} x^\top Av \\ &= \max_{i\in[m]} e_i^\top Ay - \min_{j\in[n]} x^\top Ae_j. \end{aligned} \tag{A.7}$$

For $\tau \in \mathbb{R}$, the resolution-τ equilibrium correspondence is

$$\mathcal{E}_\tau(A) := \{z \in \mathcal{Z} : g_A(z) \le \tau\}, \tag{A.8}$$

with the convention

$$\mathcal{E}_\tau(A) = \varnothing \qquad \text{when } \tau < 0. \tag{A.9}$$

Proposition A.2 (Elementary structure of the equilibrium correspondence)**.** *For every $A \in \mathsf{M}$, the following statements hold.*

1. *The function g_A is finite, nonnegative, continuous, and convex on $\mathcal{Z}$.*
2. *A pair $(x, y) \in \mathcal{Z}$ belongs to $\mathcal{E}_0(A)$ if and only if it is a saddle point of A.*
3. *For every $\tau \ge 0$, the set $\mathcal{E}_\tau(A)$ is nonempty, compact, and convex.*
4. *Let*

$$\mathcal{X}^\star(A) := \arg\max_{x\in\Delta_m} \min_{y\in\Delta_n} x^\top Ay$$

and

$$\mathcal{Y}^\star(A) := \arg\min_{y\in\Delta_n} \max_{x\in\Delta_m} x^\top Ay.$$

Then

$$\mathcal{E}_0(A) = \mathcal{X}^\star(A) \times \mathcal{Y}^\star(A). \tag{A.10}$$

Proof. The second equality in (A.7) follows because a linear functional attains its extrema over a simplex at an extreme point. Moreover,

$$\max_{u\in\Delta_m} u^\top Ay \ge x^\top Ay \ge \min_{v\in\Delta_n} x^\top Av,$$

so $g_A \ge 0$.

The first term of (A.7) is a finite maximum of affine functions of y. The negative of the second term is a finite maximum of affine functions of x. Consequently, g_A is continuous and convex in the joint variable (x, y).

The equality $g_A(x, y) = 0$ holds exactly when

$$\max_{u\in\Delta_m} u^\top Ay = x^\top Ay = \min_{v\in\Delta_n} x^\top Av,$$

which is equivalent to

$$u^\top Ay \le x^\top Ay \le x^\top Av \qquad \text{for all } u \in \Delta_m,\ v \in \Delta_n.$$

These are precisely the saddle inequalities.

Existence of a saddle point follows from (A.6), so $\mathcal{E}_0(A)$ is nonempty. For every $\tau \ge 0$, the set $\mathcal{E}_\tau(A)$ is therefore a nonempty closed convex subset of the compact set $\mathcal{Z}$, and is hence compact.

Finally, if $x \in \mathcal{X}^\star(A)$ and $y \in \mathcal{Y}^\star(A)$, then

$$\max_{u\in\Delta_m} u^\top Ay = \mathrm{val}(A) = \min_{v\in\Delta_n} x^\top Av,$$

so $g_A(x, y) = 0$. Conversely, any saddle pair consists of optimal maximin and minimax strategies. □

A.3 SET DISTANCES, SUPPORT FUNCTIONS, AND COMPACT HYPERSPACES

Let (X, d) be a metric space. For $x \in \mathsf{X}$ and a nonempty set $C \subseteq \mathsf{X}$, write

$$\operatorname{dist}(x, C) := \inf_{c \in C} d(x, c). \tag{A.11}$$

For nonempty sets $C, D \subseteq \mathsf{X}$, define the directed excess and Hausdorff distance by

$$e(C, D) := \sup_{c \in C} \operatorname{dist}(c, D), \qquad d_{\mathrm{H}}(C, D) := \max\{e(C, D), e(D, C)\}. \tag{A.12}$$

When empty sets arise, we use the extended conventions

$$e(\varnothing, D) := 0, \qquad e(C, \varnothing) := +\infty \quad \text{for } C \neq \varnothing. \tag{A.13}$$

The Hausdorff metric is used only between nonempty compact sets.

Denote by $\mathfrak{K}(\mathsf{X})$ the family of nonempty compact subsets of X. If X is a convex subset of a finite-dimensional vector space, denote by $\mathfrak{K}_{\mathrm{cvx}}(\mathsf{X})$ the family of its nonempty compact convex subsets.

Lemma A.3 (Distance-function representation)**.** *For $C, D \in \mathfrak{K}(\mathsf{X})$,*

$$d_{\mathrm{H}}(C, D) = \sup_{x \in \mathsf{X}} |\operatorname{dist}(x, C) - \operatorname{dist}(x, D)|. \tag{A.14}$$

In particular, when X is compact,

$$C \longmapsto \operatorname{dist}(\cdot, C)$$

is an isometric embedding of $(\mathfrak{K}(\mathsf{X}), d_{\mathrm{H}})$ into $C(\mathsf{X})$ equipped with the supremum norm.

Proof. For every $x \in \mathsf{X}$,

$$\operatorname{dist}(x, D) \leq \operatorname{dist}(x, C) + e(C, D),$$

and the symmetric inequality holds after interchanging C and D. Consequently, the right-hand side of (A.14) is at most $d_{\mathrm{H}}(C, D)$.

Conversely, evaluating at $x = c \in C$ gives

$$\operatorname{dist}(c, D) = |\operatorname{dist}(c, C) - \operatorname{dist}(c, D)|,$$

so $e(C, D)$ is at most the right-hand side. Interchanging C and D gives the same conclusion for the reverse excess. □

Proposition A.4 (Compactness of the action hyperspace)**.** *If X is compact, then*

$$\big(\mathfrak{K}(\mathsf{X}), d_{\mathrm{H}}\big)$$

is compact. If, in addition, X is a compact convex subset of a finite-dimensional vector space, then $\mathfrak{K}_{\mathrm{cvx}}(\mathsf{X})$ is a closed, and hence compact, subspace of $\mathfrak{K}(\mathsf{X})$.

Proof. For $C \in \mathfrak{K}(\mathsf{X})$, the distance function

$$d_C := \operatorname{dist}(\cdot, C)$$

is 1-Lipschitz and takes values in $[0, \operatorname{diam}(\mathsf{X})]$. Given a sequence $(C_k) \subseteq \mathfrak{K}(\mathsf{X})$, the Arzelà–Ascoli theorem yields a subsequence, not relabeled, for which d_{C_k} converges uniformly to a nonnegative 1-Lipschitz function f.

Choose $c_k \in C_k$. By compactness, pass to a further subsequence such that $c_k \to c$. Uniform convergence gives $f(c) = 0$, and hence the closed set

$$C := \{x \in \mathsf{X} : f(x) = 0\}$$

is nonempty.

We claim that $f = d_C$. Since f is 1-Lipschitz and vanishes on C,

$$f(x) \leq d_C(x) \qquad \text{for every } x \in \mathsf{X}.$$

Conversely, for fixed x, select $c_k^x \in C_k$ such that

$$d_{C_k}(x) = d(x, c_k^x)$$

and pass to a convergent subsequence $c_k^x \to c^x$. Uniform convergence implies $f(c^x) = 0$, so $c^x \in C$, and therefore

$$f(x) = \lim_k d(x, c_k^x) = d(x, c^x) \geq d_C(x).$$

Thus $f = d_C$. Lemma A.3 now gives

$$d_{\mathrm{H}}(C_k, C) \to 0,$$

proving compactness.

It remains to prove closedness of the convex subspace. Suppose $C_k \in \mathfrak{K}_{\mathrm{cvx}}(\mathsf{X})$ and $C_k \to C$ in Hausdorff distance. For $x, x' \in C$ and $\alpha \in [0,1]$, choose $x_k, x_k' \in C_k$ with

$$x_k \to x, \qquad x_k' \to x'.$$

Convexity gives

$$\alpha x_k + (1-\alpha)x_k' \in C_k.$$

Hausdorff convergence then implies

$$\alpha x + (1-\alpha)x' \in C.$$

Hence C is convex. □

For a nonempty compact set $C \subseteq \mathsf{V}$, its support function is

$$h_C(w) := \sup_{z \in C} \langle w, z \rangle_{\mathsf{V}}, \qquad w \in \mathsf{V}. \tag{A.15}$$

The following norm-general representation is standard in convex geometry; see, for example, Schneider (2014).

Lemma A.5 (Hausdorff distance through support functions)**.** *Let $C, D \in \mathfrak{K}_{\mathrm{cvx}}(\mathsf{V})$. Then*

$$e(C, D) = \sup_{\|w\|_{\mathsf{V},*} \leq 1} \big(h_C(w) - h_D(w)\big), \tag{A.16}$$

$$d_{\mathrm{H}}(C, D) = \sup_{\|w\|_{\mathsf{V},*} \leq 1} |h_C(w) - h_D(w)|. \tag{A.17}$$

Proof. Let $\mathbb{B}_{\mathsf{V}}$ be the closed unit ball of $\|\cdot\|_{\mathsf{V}}$. By definition of the excess,

$$e(C, D) = \inf\{r \geq 0 : C \subseteq D + r\mathbb{B}_{\mathsf{V}}\}.$$

For closed convex sets, the inclusion

$$C \subseteq D + r\mathbb{B}_{\mathsf{V}}$$

is equivalent, by finite-dimensional separation, to

$$h_C(w) \leq h_D(w) + r\|w\|_{\mathsf{V},*} \qquad \text{for every } w \in \mathsf{V}.$$

Taking the smallest admissible r and normalizing every nonzero w yields (A.16). The choice $w = 0$ ensures that the supremum is nonnegative. Applying the same identity after interchanging C and D, and then taking the maximum of the two directed excesses, proves (A.17). □

A.4 Painlevé–Kuratowski and epigraphical convergence

We use sequential definitions, which are equivalent to the standard topological definitions in the finite-dimensional metric spaces considered here (Rockafellar & Wets, 1998). For a family of sets $(C_t)_{t>0} \subseteq \mathsf{X}$, define

$$\limsup_{t \downarrow 0} C_t := \left\{ x : \begin{array}{l} \exists\, t_k \downarrow 0,\ \exists\, x_k \in C_{t_k} \\ \text{such that } x_k \to x \end{array} \right\}, \tag{A.18}$$

$$\liminf_{t \downarrow 0} C_t := \left\{ x : \begin{array}{l} \text{for every } t_k \downarrow 0,\ \exists\, x_k \in C_{t_k} \\ \text{such that } x_k \to x \end{array} \right\}. \tag{A.19}$$

We say that C_t converges to C in the Painlevé–Kuratowski sense when

$$\limsup_{t \downarrow 0} C_t = C = \liminf_{t \downarrow 0} C_t.$$

Lemma A.6 (Hausdorff versus Painlevé–Kuratowski convergence)**.** *Let* X *be compact and let* $C_t, C \in \mathfrak{K}(\mathsf{X})$*. Then*

$$d_{\mathrm{H}}(C_t, C) \longrightarrow 0 \quad \Longleftrightarrow \quad \limsup_{t\downarrow 0} C_t \subseteq C \subseteq \liminf_{t\downarrow 0} C_t. \tag{A.20}$$

Proof. Hausdorff convergence immediately gives both set inclusions.

Conversely, suppose that the inclusions hold but Hausdorff convergence fails. Along a sequence $t_k \downarrow 0$, either

$$e(C_{t_k}, C) \geq \varepsilon$$

or

$$e(C, C_{t_k}) \geq \varepsilon$$

for some $\varepsilon > 0$.

In the first case, choose $x_k \in C_{t_k}$ such that

$$\operatorname{dist}(x_k, C) \geq \varepsilon$$

and extract a convergent subsequence. Its limit belongs to the outer limit and hence to C, a contradiction.

In the second case, choose $x_k \in C$ such that

$$\operatorname{dist}(x_k, C_{t_k}) \geq \varepsilon$$

and extract a subsequence $x_k \to x \in C$. Since $C \subseteq \liminf_{t\downarrow 0} C_t$, there exist $y_k \in C_{t_k}$ such that $y_k \to x$. Hence

$$\operatorname{dist}(x_k, C_{t_k}) \leq d(x_k, y_k) \longrightarrow 0,$$

again a contradiction. □

Let

$$\mathbb{R}_\infty := \mathbb{R} \cup \{+\infty\}.$$

A family of extended-real-valued functions

$$f_t : \mathsf{X} \to \mathbb{R}_\infty$$

epi-converges to $f : \mathsf{X} \to \mathbb{R}_\infty$ as $t \downarrow 0$, written

$$f_t \xrightarrow{\text{epi}} f,$$

if the following two conditions hold.

1. For every $t_k \downarrow 0$ and every $x_k \to x$,
$$\liminf_{k\to\infty} f_{t_k}(x_k) \geq f(x). \tag{A.21}$$
2. For every $x \in \mathsf{X}$ and every $t_k \downarrow 0$, there exists $x_k \to x$ such that
$$\limsup_{k\to\infty} f_{t_k}(x_k) \leq f(x). \tag{A.22}$$

A sequence satisfying the second condition is called a *recovery sequence*.

Proposition A.7 (Sublevel sets under epi-convergence)**.** *Let* X *be compact, let*

$$f_t, f : \mathsf{X} \to \mathbb{R}_\infty$$

be lower semicontinuous, and suppose $f_t \xrightarrow{\text{epi}} f$*. For* $\lambda \in \mathbb{R}$*, define*

$$S_t(\lambda) := \{x \in \mathsf{X} : f_t(x) \leq \lambda\}, \qquad S(\lambda) := \{x \in \mathsf{X} : f(x) \leq \lambda\}.$$

Then

$$\{x \in \mathsf{X} : f(x) < \lambda\} \subseteq \liminf_{t\downarrow 0} S_t(\lambda) \subseteq \limsup_{t\downarrow 0} S_t(\lambda) \subseteq S(\lambda). \tag{A.23}$$

If $S(\lambda)$ *is nonempty and*

$$\operatorname{cl}\{x \in \mathsf{X} : f(x) < \lambda\} = S(\lambda), \tag{A.24}$$

then $S_t(\lambda)$ *is nonempty for all sufficiently small* t*, and*

$$d_{\mathrm{H}}\big(S_t(\lambda), S(\lambda)\big) \longrightarrow 0. \tag{A.25}$$

Proof. If $f(x) < \lambda$, then along every $t_k \downarrow 0$, a recovery sequence satisfies

$$f_{t_k}(x_k) < \lambda$$

for all sufficiently large k. This proves the first inclusion in (A.23).

If x belongs to the outer limit, there exist $t_k \downarrow 0$ and $x_k \in S_{t_k}(\lambda)$ such that $x_k \to x$. The epi-liminf inequality gives

$$f(x) \le \liminf_{k\to\infty} f_{t_k}(x_k) \le \lambda,$$

which proves the final inclusion.

Under (A.24), the first inclusion and the closedness of the Painlevé–Kuratowski inner limit give

$$S(\lambda) \subseteq \liminf_{t\downarrow 0} S_t(\lambda).$$

The Hausdorff conclusion now follows from Lemma A.6. Eventual nonemptiness follows from the same inner-limit inclusion. □

Remark A.8 (Parameter-dependent epi-convergence). When the functions also depend on a direction parameter H, the phrase *joint epi-convergence* means epi-convergence of

$$(H, z) \longmapsto f_t(H, z)$$

on the product space. Thus, the epi-liminf condition is tested along arbitrary sequences $H_t \to H$ and $z_t \to z$. Joint epi-convergence and uniform Hausdorff convergence of sublevel sets are distinct assertions; the latter is always stated and proved separately.

A.5 Tangent, normal, and positive-dual cones

Let C be a closed subset of a finite-dimensional vector space and let $z \in C$. The Bouligand tangent cone is

$$T_C(z) := \left\{ d : \begin{array}{l} \exists\, t_k \downarrow 0,\ \exists\, z_k \in C \\ \text{such that } (z_k - z)/t_k \to d \end{array} \right\}. \tag{A.26}$$

If C is convex, then

$$T_C(z) = \mathrm{cl}\,\mathrm{cone}(C - z). \tag{A.27}$$

The convex-analytic normal cone is

$$N_C(z) := \{w : \langle w, c - z\rangle \le 0 \text{ for every } c \in C\}. \tag{A.28}$$

For a cone K, define its polar and positive dual by

$$K^\circ := \{w : \langle w, d\rangle \le 0 \text{ for every } d \in K\}, \qquad K^+ := \{w : \langle w, d\rangle \ge 0 \text{ for every } d \in K\}. \tag{A.29}$$

Thus,

$$K^+ = -K^\circ. \tag{A.30}$$

For a closed convex set C,

$$N_C(z) = T_C(z)^\circ = -T_C(z)^+. \tag{A.31}$$

The explicit sign convention in (A.29) is used throughout the tangent-program duality calculations.

Lemma A.9 (Simplex tangent and positive-dual cones)**.** *Let $p \in \Delta_r$ and let*

$$S := \mathrm{supp}(p) = \{i \in [r] : p_i > 0\}.$$

Then

$$T_{\Delta_r}(p) = \big\{ d \in \mathbb{R}^r : \mathbf{1}_r^\top d = 0,\ d_i \ge 0 \text{ for every } i \notin S \big\}, \tag{A.32}$$

$$T_{\Delta_r}(p)^+ = \left\{ a \in \mathbb{R}^r : \begin{array}{l} \exists\, \lambda \in \mathbb{R} \text{ such that} \\ a_i = \lambda \text{ for every } i \in S, \\ a_i \ge \lambda \text{ for every } i \notin S \end{array} \right\}. \tag{A.33}$$

Moreover, every $d \in T_{\Delta_r}(p)$ is realized by the affine feasible path

$$p_t = p + td$$

for all sufficiently small $t > 0$.

Proof. If d is a tangent direction, the simplex equality constraint gives

$$\mathbf{1}_r^\top d = 0.$$

If $p_i = 0$, feasibility of the approximating sequence implies $d_i \geq 0$. This proves necessity in (A.32).

Conversely, suppose that the displayed conditions hold. The positive coordinates of p remain nonnegative for all sufficiently small $t > 0$, while every zero coordinate becomes $td_i \geq 0$, and

$$\mathbf{1}_r^\top (p + td) = 1.$$

Thus $p + td \in \Delta_r$, proving sufficiency and the affine-path claim.

Now let a belong to the positive dual. For $i, j \in S$, both $e_i - e_j$ and $e_j - e_i$ belong to the tangent cone, so

$$a_i = a_j.$$

Let their common value be λ. For $i \notin S$ and $j \in S$, the direction $e_i - e_j$ is tangent, so

$$a_i \geq a_j = \lambda.$$

This proves necessity in (A.33).

Conversely, if a has the displayed form, then every tangent direction d satisfies

$$a^\top d = \sum_{i \notin S} (a_i - \lambda) d_i + \lambda \mathbf{1}_r^\top d \geq 0.$$

Hence a belongs to the positive dual. □

Corollary A.10 (Product tangent geometry)**.** *For $z = (x, y) \in \mathcal{Z}$,*

$$T_{\mathcal{Z}}(z) = T_{\Delta_m}(x) \times T_{\Delta_n}(y), \tag{A.34}$$

$$T_{\mathcal{Z}}(z)^+ = T_{\Delta_m}(x)^+ \times T_{\Delta_n}(y)^+. \tag{A.35}$$

Proof. The tangent identity follows directly from the sequential tangent definition applied coordinatewise. The positive-dual identity follows from the separability of the product pairing. □

A.6 Measurable compact-set-valued maps

Let $(\Omega, \mathcal{F})$ be a measurable space and let X be a compact metric space. A compact-set-valued random element is an

$$\mathcal{F}/\mathcal{B}(\mathfrak{K}(\mathsf{X}))$$

measurable map

$$\Gamma : \Omega \to \mathfrak{K}(\mathsf{X}),$$

where the hyperspace carries the Borel sigma-field induced by the Hausdorff metric. This is the notion of measurability used for set-valued estimators; see Molchanov (2017).

Proposition A.11 (Distance characterization of hyperspace measurability)**.** *Let*

$$D = \{x_1, x_2, \ldots\}$$

be a countable dense subset of X*. A map*

$$\Gamma : \Omega \to \mathfrak{K}(\mathsf{X})$$

is measurable if and only if

$$\omega \longmapsto \operatorname{dist}(x_\ell, \Gamma(\omega)) \tag{A.36}$$

is $\mathcal{F}$-measurable for every $\ell \geq 1$.

Proof. If Γ is hyperspace-measurable, then

$$C \longmapsto \operatorname{dist}(x_\ell, C)$$

is 1-Lipschitz under the Hausdorff metric by Lemma A.3. Therefore (A.36) is measurable.

Conversely, continuity of distance functions and density of D imply

$$d_{\mathrm{H}}(C, K) = \sup_{\ell \geq 1} |\mathrm{dist}(x_\ell, C) - \mathrm{dist}(x_\ell, K)| \tag{A.37}$$

for every $C, K \in \mathfrak{K}(\mathsf{X})$. Hence

$$\omega \longmapsto d_{\mathrm{H}}(\Gamma(\omega), K)$$

is measurable for each fixed K.

By Proposition A.4, $\mathfrak{K}(\mathsf{X})$ is a compact metric space and is therefore separable. Choose a countable dense family of centers and rational radii; the associated open balls form a countable base. Their inverse images are measurable by (A.37), proving that Γ is Borel measurable. □

For a correspondence

$$\Gamma : \Theta \rightrightarrows \mathsf{X}$$

with nonempty compact values, define its graph by

$$\mathrm{gph}\, \Gamma := \{(\theta, x) \in \Theta \times \mathsf{X} : x \in \Gamma(\theta)\}. \tag{A.38}$$

At $\theta \in \Theta$, the correspondence Γ is *upper hemicontinuous* if, whenever

$$\theta_k \to \theta, \qquad x_k \in \Gamma(\theta_k), \qquad x_k \to x,$$

one has $x \in \Gamma(\theta)$. It is *lower hemicontinuous* if, whenever $\theta_k \to \theta$ and $x \in \Gamma(\theta)$, there exist

$$x_k \in \Gamma(\theta_k) \qquad \text{such that} \qquad x_k \to x.$$

Corollary A.12 (Hausdorff continuity of compact-valued correspondences)**.** *Let Θ be a metric space, let X be compact, and let*

$$\Gamma : \Theta \to \mathfrak{K}(\mathsf{X}).$$

Then Γ is continuous as a map into

$$\big(\mathfrak{K}(\mathsf{X}), d_{\mathrm{H}}\big)$$

if and only if it is both upper and lower hemicontinuous.

Proof. Apply Lemma A.6 along every sequence $\theta_k \to \theta$. Upper hemicontinuity is the outer-limit inclusion, while lower hemicontinuity is the inner-limit inclusion. □

Finally, because $\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$ is compact, the second-order hyperspace

$$\mathfrak{K}(\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})) \tag{A.39}$$

is itself a compact metric space under the Hausdorff metric induced by d_{H}. For a nonempty compact confidence collection

$$\mathfrak{C} \subseteq \mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z}),$$

define its hyperspace diameter by

$$\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}) := \sup_{C, D \in \mathfrak{C}} d_{\mathrm{H}}(C, D). \tag{A.40}$$

This is the uncertainty measure used in the honest-adaptation lower bound.

Remark A.13 (Scope of Appendix A). No regularity, uniqueness, or continuity of an equilibrium selector is assumed anywhere in this appendix. All statistical decisions later in the paper take values in the compact hyperspace $\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$, so equilibrium multiplicity is retained rather than removed by an arbitrary selection rule.

B Proof of the Validity–Completeness Separation

This appendix proves a finite-sample separation between two statistical objectives that are often conflated in equilibrium learning:

1. producing one strategy profile with small saddle residual; and
2. estimating the complete equilibrium correspondence in Hausdorff distance.

The separation already occurs in a one-parameter family of 2×2 zero-sum games. It is not caused by inadequate exploration: the information lower bound permits arbitrary randomized, history-dependent payoff queries and remains valid even when the learner queries the unique informative entry at every round.

B.1 EXACT EQUILIBRIUM CORRESPONDENCE

For $\theta \in [-1, 1]$, consider the zero-sum game

$$A_\theta := \begin{pmatrix} 1 & 0 \\ 0 & \theta \end{pmatrix}. \tag{B.1}$$

For $p, q \in [0, 1]$, write

$$x(p) := \begin{pmatrix} p \\ 1-p \end{pmatrix}, \qquad y(q) := \begin{pmatrix} q \\ 1-q \end{pmatrix}, \qquad z(p, q) := (x(p), y(q)). \tag{B.2}$$

Lemma B.1 (Exact residual)**.** *For every $\theta \in [-1, 1]$ and $p, q \in [0, 1]$,*

$$g_{A_\theta}\big(z(p, q)\big) = \max\{q,\, \theta(1-q)\} - \min\{p,\, \theta(1-p)\}. \tag{B.3}$$

In particular,

$$g_{A_\theta}\big(z(p, q)\big) = \begin{cases} q - \theta(1-p), & \theta < 0, \\ q, & \theta = 0. \end{cases} \tag{B.4}$$

Proof. A direct calculation gives

$$A_\theta y(q) = \begin{pmatrix} q \\ \theta(1-q) \end{pmatrix}, \qquad x(p)^\top A_\theta = \begin{pmatrix} p & \theta(1-p) \end{pmatrix}.$$

Substituting these expressions into (A.7) proves (B.3).

If $\theta < 0$, then

$$\theta(1-q) \le 0 \le q, \qquad \theta(1-p) \le 0 \le p.$$

Consequently, the maximum and minimum in (B.3) equal q and $\theta(1-p)$, respectively. The case $\theta = 0$ follows immediately. □

Proposition B.2 (Exact equilibrium sets)**.** *The exact equilibrium correspondence of* (B.1) *is*

$$\mathcal{E}_0(A_\theta) = \begin{cases} \{z(1, 0)\}, & \theta < 0, \\ \{z(p, 0) : p \in [0, 1]\}, & \theta = 0, \\ \{z(a_\theta, a_\theta)\}, & \theta > 0, \end{cases} \tag{B.5}$$

where

$$a_\theta := \frac{\theta}{1+\theta} \qquad \textit{for } \theta > 0. \tag{B.6}$$

Thus, the positive-dimensional equilibrium face at $\theta = 0$ collapses to different singleton branches under positive and negative perturbations.

Proof. If $\theta < 0$, Lemma B.1 gives

$$g_{A_\theta}\big(z(p, q)\big) = q + (-\theta)(1-p).$$

Both terms are nonnegative. Hence the residual vanishes if and only if

$$q = 0 \qquad \text{and} \qquad p = 1.$$

If $\theta = 0$, the residual equals q. It therefore vanishes exactly when $q = 0$, with no restriction on p.

Now suppose $\theta > 0$. Define

$$r_\theta(q) := \max\{q, \theta(1-q)\}, \qquad s_\theta(p) := \min\{p, \theta(1-p)\}.$$

The two affine functions in the definition of r_θ intersect at

$$q = a_\theta = \frac{\theta}{1+\theta},$$

and their common value is a_θ. Therefore,

$$r_\theta(q) \geq a_\theta, \tag{B.7}$$

with equality if and only if $q = a_\theta$.

Similarly, the affine functions defining s_θ intersect at $p = a_\theta$, and

$$s_\theta(p) \leq a_\theta, \tag{B.8}$$

with equality if and only if $p = a_\theta$.

Since

$$g_{A_\theta}\big(z(p,q)\big) = r_\theta(q) - s_\theta(p),$$

the residual is zero if and only if equality holds simultaneously in (B.7) and (B.8). Hence

$$p = q = a_\theta,$$

which proves the positive-θ case. □

The product ℓ_1 metric admits a particularly simple expression under the parametrization (B.2):

$$\|z(p,q) - z(p',q')\|_{\mathsf{V}} = 2|p-p'| + 2|q-q'|. \tag{B.9}$$

Fix $T \geq 1$ and $c \in (0,1]$, and define the three local parameters

$$\theta_{T,-} := -\frac{c}{\sqrt{T}}, \qquad \theta_{T,0} := 0, \qquad \theta_{T,+} := \frac{c}{\sqrt{T}}. \tag{B.10}$$

Write

$$S_{T,s} := \mathcal{E}_0(A_{\theta_{T,s}}), \qquad s \in \{-,0,+\}, \tag{B.11}$$

and define

$$a_T := a_{\theta_{T,+}} = \frac{c}{\sqrt{T}+c}. \tag{B.12}$$

Since $c \leq 1$ and $T \geq 1$,

$$0 < a_T \leq \frac{1}{2}. \tag{B.13}$$

Proposition B.3 (Exact directed and Hausdorff geometry)**.** *The local equilibrium sets are*

$$S_{T,-} = \{z(1,0)\}, \qquad S_{T,0} = \{z(p,0) : p \in [0,1]\}, \qquad S_{T,+} = \{z(a_T,a_T)\}. \tag{B.14}$$

Their directed excesses satisfy

$$e(S_{T,-}, S_{T,0}) = 0, \qquad e(S_{T,0}, S_{T,-}) = 2, \tag{B.15}$$

$$e(S_{T,+}, S_{T,0}) = 2a_T, \qquad e(S_{T,0}, S_{T,+}) = 2, \tag{B.16}$$

and

$$d_{\mathrm{H}}(S_{T,r}, S_{T,s}) = 2 \qquad \textit{for every distinct } r,s \in \{-,0,+\}. \tag{B.17}$$

Proof. The set representations follow from Proposition B.2.

Since $z(1,0) \in S_{T,0}$,

$$e(S_{T,-}, S_{T,0}) = 0.$$

Conversely,

$$e(S_{T,0}, S_{T,-}) = \sup_{p\in[0,1]} 2|p-1| = 2.$$

For the positive perturbation,

$$\begin{aligned} e(S_{T,+}, S_{T,0}) &= \inf_{p\in[0,1]} \left(2|a_T - p| + 2a_T\right) \\ &= 2a_T, \end{aligned} \tag{B.18}$$

where the minimum is attained at $p = a_T$.

In the reverse direction,

$$e(S_{T,0}, S_{T,+}) = \sup_{p\in[0,1]} (2|p - a_T| + 2a_T)\,. \tag{B.19}$$

Because $a_T \le 1/2$, the maximum is attained at $p = 1$, giving

$$2(1 - a_T) + 2a_T = 2.$$

Finally,

$$\begin{aligned} d_{\mathrm{H}}(S_{T,+}, S_{T,-}) &= \|z(a_T, a_T) - z(1, 0)\|_{\mathsf{V}} \\ &= 2(1 - a_T) + 2a_T \\ &= 2. \end{aligned}$$

Combining these identities proves (B.17). □

Remark B.4 (One-sided continuity and branch loss). Equation (B.16) exhibits the exact validity–completeness asymmetry:

$$e(S_{T,+}, S_{T,0}) = \frac{2c}{\sqrt{T} + c} \longrightarrow 0, \qquad e(S_{T,0}, S_{T,+}) = 2.$$

Thus, the perturbed singleton branch lies asymptotically close to the unperturbed equilibrium face, but it fails to represent an order-one portion of that face. This is a failure of lower hemicontinuity of the exact equilibrium correspondence, not a failure of pointwise approximate optimality.

Define the deterministic profile

$$\bar{z} := z\left(\frac{1}{2}, 0\right) = \left(\begin{pmatrix}1/2\\1/2\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right). \tag{B.20}$$

Lemma B.5 (A no-data low-residual profile)**.** *For every $\theta \in [-1, 1]$,*

$$g_{A_\theta}(\bar{z}) = \frac{|\theta|}{2}. \tag{B.21}$$

Moreover,

$$d_{\mathrm{H}}\left(\{\bar{z}\}, \mathcal{E}_0(A_\theta)\right) = 1 \qquad \textit{for every } \theta \in [-1, 1]. \tag{B.22}$$

Proof. Substituting $p = 1/2$ and $q = 0$ into (B.3) gives

$$g_{A_\theta}(\bar{z}) = \max\{0, \theta\} - \min\left\{\frac{1}{2}, \frac{\theta}{2}\right\}.$$

Since $\theta \le 1$, the minimum equals $\theta/2$, and (B.21) follows.

If $\theta < 0$, the equilibrium set is the singleton $\{z(1, 0)\}$, and

$$\|\bar{z} - z(1, 0)\|_{\mathsf{V}} = 1.$$

If $\theta = 0$, then $\bar{z} \in \mathcal{E}_0(A_0)$, while

$$\sup_{p\in[0,1]} \|z(p, 0) - \bar{z}\|_{\mathsf{V}} = \sup_{p\in[0,1]} 2\left|p - \frac{1}{2}\right| = 1.$$

If $\theta > 0$, write $a = a_\theta \le 1/2$. Then

$$\|\bar{z} - z(a, a)\|_{\mathsf{V}} = 2\left(\frac{1}{2} - a\right) + 2a = 1.$$

□

B.2 ADAPTIVE INFORMATION BOUND

We now formalize the sequential observation model. Fix a horizon $T \in \mathbb{N}$ and noise level $\sigma > 0$. At round t, a possibly randomized sampling rule chooses

$$I_t \in [2] \times [2]$$

as a function of the preceding transcript

$$\mathsf{H}_{t-1} := (I_1, Y_1, \ldots, I_{t-1}, Y_{t-1}).$$

Conditionally on H_{t-1} and $I_t = (i, j)$, the learner observes

$$Y_t = (A_\theta)_{ij} + \xi_t, \qquad \xi_t \sim \mathcal{N}(0, \sigma^2), \tag{B.23}$$

where the noise variables are conditionally independent.

Formally, the sampling rule is a sequence of Markov kernels

$$\pi_t(\cdot \mid h_{t-1})$$

on $[2] \times [2]$, identical under every value of θ. Let P_θ^T denote the induced distribution of the transcript H_T. Define the number of queries to entry (i, j) by

$$N_{ij}(T) := \sum_{t=1}^{T} \mathbf{1}\{I_t = (i, j)\}. \tag{B.24}$$

The following identity is the fixed-horizon Gaussian specialization of the adaptive change-of-measure decompositions used in sequential bandit lower bounds; see, for example, Kaufmann et al. (2016). We give the complete proof.

Lemma B.6 (Adaptive KL decomposition)**.** *For any $\theta, \theta' \in [-1, 1]$,*

$$D_{\mathrm{KL}}\left(P_\theta^T \,\middle\|\, P_{\theta'}^T\right) = \frac{1}{2\sigma^2} \sum_{i=1}^{2} \sum_{j=1}^{2} \mathbb{E}_\theta\left[N_{ij}(T)\right] \left((A_\theta)_{ij} - (A_{\theta'})_{ij}\right)^2. \tag{B.25}$$

For the family (B.1)*, this reduces to*

$$D_{\mathrm{KL}}\left(P_\theta^T \,\middle\|\, P_{\theta'}^T\right) = \frac{(\theta - \theta')^2}{2\sigma^2} \mathbb{E}_\theta\left[N_{22}(T)\right] \le \frac{T(\theta - \theta')^2}{2\sigma^2}. \tag{B.26}$$

Proof. The chain rule for relative entropy gives

$$D_{\mathrm{KL}}\left(P_\theta^T \,\middle\|\, P_{\theta'}^T\right) = \sum_{t=1}^{T} \mathbb{E}_\theta\left[D_{\mathrm{KL}}\left(P_\theta\left(I_t, Y_t \mid \mathsf{H}_{t-1}\right) \,\middle\|\, P_{\theta'}\left(I_t, Y_t \mid \mathsf{H}_{t-1}\right)\right)\right]. \tag{B.27}$$

Conditioned on $\mathsf{H}_{t-1} = h_{t-1}$, the action kernel $\pi_t(\cdot \mid h_{t-1})$ is identical under the two models. Consequently, the conditional divergence in (B.27) equals

$$\sum_{i=1}^{2} \sum_{j=1}^{2} \pi_t\big((i, j) \mid h_{t-1}\big) D_{\mathrm{KL}}\left(\mathcal{N}\big((A_\theta)_{ij}, \sigma^2\big) \,\middle\|\, \mathcal{N}\big((A_{\theta'})_{ij}, \sigma^2\big)\right). \tag{B.28}$$

For Gaussians with a common variance,

$$D_{\mathrm{KL}}\left(\mathcal{N}(\mu, \sigma^2) \,\middle\|\, \mathcal{N}(\mu', \sigma^2)\right) = \frac{(\mu - \mu')^2}{2\sigma^2}. \tag{B.29}$$

Substitution into (B.27), followed by the tower property, yields (B.25).

In the family (B.1), the models differ only at entry $(2, 2)$. This proves the equality in (B.26). The inequality follows from $N_{22}(T) \le T$. □

For probability measures P, Q on a common measurable space, define

$$d_{\mathrm{TV}}(P, Q) := \sup_B |P(B) - Q(B)|. \tag{B.30}$$

Pinsker's inequality (see, e.g., Canonne, 2022) gives

$$d_{\mathrm{TV}}(P, Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)}. \tag{B.31}$$

Corollary B.7 (Local statistical indistinguishability)**.** *For the local parameters in* (B.10)*,*

$$D_{\mathrm{KL}}\left(P^T_{\theta_{T,0}} \,\middle\|\, P^T_{\theta_{T,+}}\right) \le \frac{c^2}{2\sigma^2}, \qquad d_{\mathrm{TV}}\left(P^T_{\theta_{T,0}}, P^T_{\theta_{T,+}}\right) \le \frac{c}{2\sigma}, \tag{B.32}$$

$$D_{\mathrm{KL}}\left(P^T_{\theta_{T,+}} \,\middle\|\, P^T_{\theta_{T,-}}\right) \le \frac{2c^2}{\sigma^2}, \qquad d_{\mathrm{TV}}\left(P^T_{\theta_{T,+}}, P^T_{\theta_{T,-}}\right) \le \frac{c}{\sigma}. \tag{B.33}$$

These bounds hold for every adaptive sampling strategy.

Proof. Apply (B.26) and (B.31), noting that

$$\theta_{T,+} - \theta_{T,0} = \frac{c}{\sqrt{T}}$$

and

$$\theta_{T,+} - \theta_{T,-} = \frac{2c}{\sqrt{T}}.$$

□

Remark B.8 (The lower bound is not an exploration failure). The upper bound in (B.26) is attained by querying entry $(2, 2)$ at every round. Hence no adaptive design can obtain more order-wise information about θ in this family. The forthcoming impossibility result persists under the information-maximizing query allocation.

B.3 A METRIC FORM OF LE CAM'S TWO-POINT METHOD

We require a two-point inequality for parameters taking values in a general metric space. It is a direct metric strengthening of the usual testing formulation of Le Cam's method (Le Cam & Yang, 2000; van der Vaart, 1998).

Lemma B.9 (Metric two-point inequality)**.** *Let* (A, d) *be a metric space. Let* $a_0, a_1 \in \mathsf{A}$ *be two target values, and let* P_0, P_1 *be two observation laws. For any possibly randomized estimator*

$$\widehat{a} : \textit{observations} \rightsquigarrow \mathsf{A},$$

one has

$$\max_{r \in \{0,1\}} \mathbb{E}_r\left[d(\widehat{a}, a_r)\right] \ge \frac{d(a_0, a_1)}{2}\left(1 - d_{\mathrm{TV}}(P_0, P_1)\right). \tag{B.34}$$

Proof. Let Q_r denote the law of the randomized estimator output under P_r. Since Q_r is obtained by applying the same Markov kernel to P_r, the data-processing inequality for total variation gives

$$d_{\mathrm{TV}}(Q_0, Q_1) \le d_{\mathrm{TV}}(P_0, P_1). \tag{B.35}$$

Let $\nu := Q_0 + Q_1$, and write

$$q_r := \frac{dQ_r}{d\nu}.$$

By the triangle inequality,

$$d(a, a_0) + d(a, a_1) \ge d(a_0, a_1) \qquad \text{for every } a \in \mathsf{A}.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}_0\left[d(\widehat{a}, a_0)\right] + \mathbb{E}_1\left[d(\widehat{a}, a_1)\right] \\
&\quad = \int_{\mathsf{A}} \left[d(a, a_0) q_0(a) + d(a, a_1) q_1(a)\right] d\nu(a) \\
&\quad \geq \int_{\mathsf{A}} \min\{q_0(a), q_1(a)\} \left[d(a, a_0) + d(a, a_1)\right] d\nu(a) \\
&\quad \geq d(a_0, a_1) \int_{\mathsf{A}} \min\{q_0(a), q_1(a)\} d\nu(a) \\
&\quad = d(a_0, a_1)\left(1 - d_{\mathrm{TV}}(Q_0, Q_1)\right) \\
&\quad \geq d(a_0, a_1)\left(1 - d_{\mathrm{TV}}(P_0, P_1)\right). \qquad \text{(B.36)}
\end{aligned}$$

At least one of the two risks is at least one half of their sum. □

We now apply the metric inequality with action space

$$\mathfrak{K}(\mathcal{Z})$$

and metric d_{H}. A set-valued estimator may be randomized and may depend on the entire adaptive transcript.

Define the local parameter set

$$\Theta_T(c) := \left\{-cT^{-1/2}, 0, cT^{-1/2}\right\}. \qquad \text{(B.37)}$$

Let

$$\mathfrak{R}_T^{\mathrm{corr}}(c, \sigma) := \inf_{\mathsf{A}, \widehat{S}_T} \sup_{\theta \in \Theta_T(c)} \mathbb{E}_\theta^{\mathsf{A}}\left[d_{\mathrm{H}}\left(\widehat{S}_T, \mathcal{E}_0(A_\theta)\right)\right], \qquad \text{(B.38)}$$

where the infimum ranges over all randomized adaptive T-query strategies A and all measurable estimators

$$\widehat{S}_T \in \mathfrak{K}(\mathcal{Z}).$$

Proposition B.10 (Finite-sample validity–completeness separation)**.** *Fix $\sigma > 0$, $T \geq 1$, and*

$$0 < c \leq \min\{1, \sigma\}. \qquad \text{(B.39)}$$

Then

$$1 - \frac{c}{2\sigma} \leq \mathfrak{R}_T^{\mathrm{corr}}(c, \sigma) \leq 1. \qquad \text{(B.40)}$$

In particular,

$$\mathfrak{R}_T^{\mathrm{corr}}(c, \sigma) \geq \frac{1}{2} \qquad \textit{for every } T \geq 1. \qquad \text{(B.41)}$$

By contrast, the deterministic profile $\bar{z}$, which uses no observations, satisfies

$$\sup_{\theta \in \Theta_T(c)} g_{A_\theta}(\bar{z}) = \frac{c}{2\sqrt{T}}. \qquad \text{(B.42)}$$

Thus, pointwise saddle residual can vanish at the statistical noise scale even though minimax estimation of the complete exact equilibrium correspondence has order-one risk.

Proof. For the lower bound, restrict the supremum in (B.38) to

$$\theta_{T,0} = 0 \qquad \text{and} \qquad \theta_{T,+} = \frac{c}{\sqrt{T}}.$$

By Proposition B.3,

$$d_{\mathrm{H}}(S_{T,0}, S_{T,+}) = 2.$$

Applying Lemma B.9 and then Corollary B.7 gives, for every adaptive algorithm and estimator,

$$\begin{aligned}
\max_{s \in \{0,+\}} \mathbb{E}_{\theta_{T,s}}\left[d_{\mathrm{H}}(\widehat{S}_T, S_{T,s})\right] &\geq 1 - d_{\mathrm{TV}}\left(P_{\theta_{T,0}}^T, P_{\theta_{T,+}}^T\right) \\
&\geq 1 - \frac{c}{2\sigma}. \qquad \text{(B.43)}
\end{aligned}$$

Taking the infimum over procedures proves the first inequality in (B.40).

For the upper bound, use any sampling strategy and return the deterministic singleton

$$\widehat{S}_T := \{\bar{z}\}.$$

Lemma B.5 gives

$$d_{\mathrm{H}}\left(\{\bar{z}\}, \mathcal{E}_0(A_\theta)\right) = 1$$

for every $\theta \in \Theta_T(c)$. Thus

$$\mathfrak{R}_T^{\mathrm{corr}}(c, \sigma) \le 1.$$

The condition $c \le \sigma$ gives (B.41).

Finally, Lemma B.5 implies

$$\sup_{\theta \in \Theta_T(c)} g_{A_\theta}(\bar{z}) = \frac{1}{2} \sup_{\theta \in \Theta_T(c)} |\theta| = \frac{c}{2\sqrt{T}}.$$

□

Corollary B.11 (Two-sided branch uncertainty)**.** *If*

$$0 < c \le \min\left\{1, \frac{\sigma}{2}\right\},$$

then every adaptive estimator satisfies

$$\max_{s \in \{-,+\}} \mathbb{E}_{\theta_{T,s}}\left[d_{\mathrm{H}}(\widehat{S}_T, S_{T,s})\right] \ge 1 - \frac{c}{\sigma} \ge \frac{1}{2}. \tag{B.44}$$

Proof. Apply Lemma B.9 to $S_{T,-}$ and $S_{T,+}$, whose Hausdorff distance is 2, and use (B.33). □

B.4 Validity and completeness losses

For a reported nonempty compact set C and a target set S, define

$$L_{\mathrm{val}}(C; S) := e(C, S), \qquad L_{\mathrm{comp}}(C; S) := e(S, C). \tag{B.45}$$

The first quantity penalizes reported points that are far from the target. The second penalizes target branches that the report fails to represent. By definition,

$$d_{\mathrm{H}}(C, S) = \max\{L_{\mathrm{val}}(C; S), L_{\mathrm{comp}}(C; S)\}. \tag{B.46}$$

Proposition B.12 (Exact directed-loss separation)**.** *At the bifurcation game A_0, the singleton report $\{\bar{z}\}$ satisfies*

$$L_{\mathrm{val}}\left(\{\bar{z}\}; \mathcal{E}_0(A_0)\right) = 0, \qquad L_{\mathrm{comp}}\left(\{\bar{z}\}; \mathcal{E}_0(A_0)\right) = 1. \tag{B.47}$$

For the positive local branch,

$$L_{\mathrm{val}}\left(S_{T,+}; S_{T,0}\right) = \frac{2c}{\sqrt{T} + c} \longrightarrow 0, \qquad L_{\mathrm{comp}}\left(S_{T,+}; S_{T,0}\right) = 2. \tag{B.48}$$

Proof. Because

$$\bar{z} = z\left(\frac{1}{2}, 0\right) \in \mathcal{E}_0(A_0),$$

the validity loss in (B.47) is zero. The completeness loss is

$$\sup_{p \in [0,1]} 2\left|p - \frac{1}{2}\right| = 1.$$

The second assertion is precisely (B.16), together with (B.12). □

Remark B.13 (Interpretation). The two separations above are logically distinct.

First, (B.47) shows that exact validity of a selected equilibrium does not imply completeness: an output may contain no invalid point and nevertheless omit equilibrium branches at order-one distance.

Second, (B.48) shows that even the exact equilibrium set of a statistically contiguous perturbed game may be asymptotically valid relative to the bifurcation game while remaining permanently incomplete. Hence the obstruction is intrinsic to the set-valued parameter and cannot be repaired by improving the optimization accuracy of a point-valued learner.

C Polyhedral Error Bounds and Localization

This appendix establishes the deterministic localization principle used in all subsequent first-order arguments. For every fixed matrix game A, the saddle residual controls the distance to the complete equilibrium correspondence through a global polyhedral error bound,

$$\operatorname{dist}\big(z, \mathcal{E}_0(A)\big) \le \kappa_{\mathrm{eq}}(A)\, g_A(z).$$

The constant $\kappa_{\mathrm{eq}}(A)$ is finite for every finite game but is intrinsically instance dependent and may diverge along sequences of games approaching an equilibrium bifurcation. Combining this error bound with a sharp perturbation inequality for g_A yields uniform $O(t)$ localization of all $O(t)$-approximate equilibria of $A_\star + tH$ around $\mathcal{E}_0(A_\star)$, uniformly over bounded direction classes. The polyhedral error-bound argument is an application of Hoffman's theorem (Hoffman, 1952); see also Peña et al. (2021) for modern relative-polyhedron formulations and sharp Hoffman constants.

C.1 Linear representation of saddle points

For $A \in \mathbb{R}^{m\times n}$, define the upper and lower security levels

$$\overline{v}_A(y) := \max_{i\in[m]} e_i^\top A y, \qquad \underline{v}_A(x) := \min_{j\in[n]} x^\top A e_j, \tag{C.1}$$

and write

$$v_A := \operatorname{val}(A). \tag{C.2}$$

Thus,

$$g_A(x,y) = \overline{v}_A(y) - \underline{v}_A(x). \tag{C.3}$$

Lemma C.1 (Weak duality and residual decomposition)**.** *For every $x \in \Delta_m$ and $y \in \Delta_n$,*

$$\underline{v}_A(x) \le v_A \le \overline{v}_A(y). \tag{C.4}$$

Consequently,

$$g_A(x,y) = \big[v_A - \underline{v}_A(x)\big] + \big[\overline{v}_A(y) - v_A\big], \tag{C.5}$$

where both terms on the right-hand side are nonnegative.

Proof. For every fixed $x \in \Delta_m$,

$$\underline{v}_A(x) = \min_{y'\in\Delta_n} x^\top A y' \le \max_{x'\in\Delta_m} \min_{y'\in\Delta_n} x'^\top A y' = v_A.$$

The other inequality follows similarly from

$$v_A = \min_{y'\in\Delta_n} \max_{x'\in\Delta_m} x'^\top A y' \le \max_{x'\in\Delta_m} x'^\top A y = \overline{v}_A(y).$$

Subtracting and adding v_A in (C.3) proves (C.5). □

Define the row and column optimal-strategy polytopes

$$\mathcal{X}^\star(A) := \big\{x \in \Delta_m : A^\top x \ge v_A \mathbf{1}_n\big\}, \tag{C.6}$$

$$\mathcal{Y}^\star(A) := \{y \in \Delta_n : A y \le v_A \mathbf{1}_m\}. \tag{C.7}$$

All vector inequalities are interpreted componentwise.

Proposition C.2 (Polyhedral representation of the equilibrium correspondence)**.** *For every $A \in \mathbb{R}^{m\times n}$,*

$$\mathcal{E}_0(A) = \mathcal{X}^\star(A) \times \mathcal{Y}^\star(A). \tag{C.8}$$

Moreover, the lifted saddle-point polyhedron

$$\mathcal{Q}(A) := \big\{(x,y,v) \in \Delta_m \times \Delta_n \times \mathbb{R} : A y \le v\mathbf{1}_m,\ A^\top x \ge v\mathbf{1}_n\big\} \tag{C.9}$$

satisfies

$$\mathcal{Q}(A) = \mathcal{E}_0(A) \times \{v_A\}. \tag{C.10}$$

Proof. If $x \in \mathcal{X}^\star(A)$, then

$$\underline{v}_A(x) = \min_{j\in[n]} x^\top A e_j \geq v_A.$$

Weak duality in Lemma C.1 gives the reverse inequality, hence $\underline{v}_A(x) = v_A$. Thus $\mathcal{X}^\star(A)$ is exactly the maximin-optimal strategy set. The same argument shows that $\mathcal{Y}^\star(A)$ is exactly the minimax-optimal strategy set. Equation (C.8) now follows from Proposition A.2.

Let $(x, y, v) \in \mathcal{Q}(A)$. The defining inequalities imply

$$\overline{v}_A(y) \leq v \leq \underline{v}_A(x).$$

Combining this with (C.4) yields

$$v_A \leq \overline{v}_A(y) \leq v \leq \underline{v}_A(x) \leq v_A.$$

All quantities are therefore equal to v_A, and $(x, y) \in \mathcal{E}_0(A)$. This proves $\mathcal{Q}(A) \subseteq \mathcal{E}_0(A) \times \{v_A\}$. The reverse inclusion follows immediately from (C.6)– (C.7). □

C.2 A HOFFMAN ERROR BOUND FOR THE EQUILIBRIUM RESIDUAL

For a vector r, let r_+ denote its componentwise positive part. We use the following norm-specific form of Hoffman's error bound.

Lemma C.3 (Hoffman error bound)**.** *Let $B \in \mathbb{R}^{p\times d}$. There exists a finite constant $H_{1,\infty}(B) \geq 0$, depending only on B and the chosen ℓ_1/ℓ_∞ norms, such that, for every $b \in \mathbb{R}^p$ for which*

$$P(B, b) := \{u \in \mathbb{R}^d : Bu \leq b\}$$

is nonempty and every $u \in \mathbb{R}^d$,

$$\operatorname{dist}_{\ell_1}\big(u, P(B, b)\big) \leq H_{1,\infty}(B)\|(Bu - b)_+\|_\infty. \tag{C.11}$$

Proof. This is Hoffman's theorem for systems of linear inequalities (Hoffman, 1952), specialized to the ℓ_1 distance and the ℓ_∞ residual norm. The norm-general finite-dimensional form follows equivalently by norm equivalence; modern characterizations of the sharp constant are given by Peña et al. (2021). □

To apply Lemma C.3, define

$$B_X(A) := \begin{bmatrix} -I_m \\ \mathbf{1}_m^\top \\ -\mathbf{1}_m^\top \\ -A^\top \end{bmatrix}, \qquad b_X(A) := \begin{bmatrix} 0_m \\ 1 \\ -1 \\ -v_A\mathbf{1}_n \end{bmatrix}, \tag{C.12}$$

$$B_Y(A) := \begin{bmatrix} -I_n \\ \mathbf{1}_n^\top \\ -\mathbf{1}_n^\top \\ A \end{bmatrix}, \qquad b_Y(A) := \begin{bmatrix} 0_n \\ 1 \\ -1 \\ v_A\mathbf{1}_m \end{bmatrix}. \tag{C.13}$$

Then

$$\mathcal{X}^\star(A) = P\big(B_X(A), b_X(A)\big), \qquad \mathcal{Y}^\star(A) = P\big(B_Y(A), b_Y(A)\big). \tag{C.14}$$

Set

$$H_X(A) := H_{1,\infty}\big(B_X(A)\big), \qquad H_Y(A) := H_{1,\infty}\big(B_Y(A)\big). \tag{C.15}$$

Lemma C.4 (Playerwise error bounds)**.** *For every $x \in \Delta_m$ and $y \in \Delta_n$,*

$$\operatorname{dist}_{\ell_1}\big(x, \mathcal{X}^\star(A)\big) \leq H_X(A)\big[v_A - \underline{v}_A(x)\big], \tag{C.16}$$

$$\operatorname{dist}_{\ell_1}\big(y, \mathcal{Y}^\star(A)\big) \leq H_Y(A)\big[\overline{v}_A(y) - v_A\big]. \tag{C.17}$$

Proof. For $x \in \Delta_m$, all constraint blocks in (C.12) except the final block are satisfied exactly. Therefore,

$$\begin{aligned}\Big\|\big(B_X(A)x - b_X(A)\big)_+\Big\|_\infty &= \Big\|\big(v_A\mathbf{1}_n - A^\top x\big)_+\Big\|_\infty \\ &= \max_{j\in[n]}\big[v_A - x^\top Ae_j\big]_+ \\ &= v_A - \min_{j\in[n]} x^\top Ae_j \\ &= v_A - \underline{v}_A(x),\end{aligned}$$

where the penultimate equality uses $\underline{v}_A(x) \le v_A$. Applying Lemma C.3 proves (C.16).

The proof of (C.17) is identical: for $y \in \Delta_n$,

$$\begin{aligned}\Big\|\big(B_Y(A)y - b_Y(A)\big)_+\Big\|_\infty &= \Big\|\big(Ay - v_A\mathbf{1}_m\big)_+\Big\|_\infty \\ &= \max_{i\in[m]}\big[e_i^\top Ay - v_A\big]_+ \\ &= \overline{v}_A(y) - v_A,\end{aligned}$$

using $v_A \le \overline{v}_A(y)$. □

Proposition C.5 (Global equilibrium error bound)**.** *For every finite zero-sum game A, there exists a finite constant $\kappa_A < \infty$ such that*

$$\operatorname{dist}\big(z, \mathcal{E}_0(A)\big) \le \kappa_A\, g_A(z) \qquad \textit{for every } z \in \mathcal{Z}, \tag{C.18}$$

where the distance is induced by the product norm (A.4)*. One admissible choice is*

$$\kappa_A := \max\{H_X(A), H_Y(A)\}. \tag{C.19}$$

Proof. Because the equilibrium correspondence is the Cartesian product in (C.8), separability of the product ℓ_1 norm gives

$$\begin{aligned}\operatorname{dist}\big((x,y), \mathcal{E}_0(A)\big) &= \inf_{\substack{x^\star\in\mathcal{X}^\star(A)\\ y^\star\in\mathcal{Y}^\star(A)}} \big(\|x - x^\star\|_1 + \|y - y^\star\|_1\big) \\ &= \operatorname{dist}_{\ell_1}\big(x, \mathcal{X}^\star(A)\big) + \operatorname{dist}_{\ell_1}\big(y, \mathcal{Y}^\star(A)\big).\end{aligned} \tag{C.20}$$

Applying Lemma C.4 and then (C.5) yields

$$\begin{aligned}\operatorname{dist}\big((x,y), \mathcal{E}_0(A)\big) &\le H_X(A)\big[v_A - \underline{v}_A(x)\big] + H_Y(A)\big[\overline{v}_A(y) - v_A\big] \\ &\le \kappa_A\,\big(v_A - \underline{v}_A(x) + \overline{v}_A(y) - v_A\big) \\ &= \kappa_A g_A(x,y).\end{aligned}$$

□

The smallest valid constant in (C.18) is the equilibrium error-bound modulus

$$\kappa_{\mathrm{eq}}(A) := \sup_{z\in\mathcal{Z}\setminus\mathcal{E}_0(A)} \frac{\operatorname{dist}\big(z, \mathcal{E}_0(A)\big)}{g_A(z)}, \tag{C.21}$$

with the convention that the supremum of the empty set is zero. Proposition C.5 proves

$$0 \le \kappa_{\mathrm{eq}}(A) \le \kappa_A < \infty. \tag{C.22}$$

Remark C.6 (Degeneration at the guiding bifurcation)*.* For the family A_θ in (B.1),

$$\kappa_{\mathrm{eq}}(A_0) = 2, \qquad \kappa_{\mathrm{eq}}(A_\theta) \ge \frac{2}{|\theta|} \quad \text{for every } \theta \ne 0. \tag{C.23}$$

Indeed, at $\theta = 0$,

$$g_{A_0}\big(z(p,q)\big) = q, \qquad \operatorname{dist}\big(z(p,q), \mathcal{E}_0(A_0)\big) = 2q,$$

so the modulus equals 2. If $\theta > 0$, use $z(1,0)$, whose distance to the unique equilibrium is 2 and whose residual is θ. If $\theta < 0$, use $z(0,0)$, whose distance to the unique equilibrium is 2 and whose residual is $|\theta|$. Hence the local conditioning of the neighboring singleton equilibria deteriorates at least as $|\theta|^{-1}$, even though the base game itself has a finite error-bound modulus.

C.3 PAYOFF PERTURBATION INEQUALITIES

Lemma C.7 (Lipschitz properties of value and residual)**.** *For every $A, B \in \mathbb{R}^{m\times n}$ and every $z = (x, y), z' = (x', y') \in \mathcal{Z}$,*

$$|v_A - v_B| \le \|A - B\|_\infty, \tag{C.24}$$

$$|g_A(z) - g_B(z)| \le 2\|A - B\|_\infty, \tag{C.25}$$

$$|g_A(z) - g_A(z')| \le \|A\|_\infty \|z - z'\|_{\mathsf{V}}. \tag{C.26}$$

Consequently,

$$|g_A(z) - g_B(z')| \le 2\|A - B\|_\infty + \max\{\|A\|_\infty, \|B\|_\infty\}\|z - z'\|_{\mathsf{V}}. \tag{C.27}$$

Proof. For every $x \in \Delta_m$ and $y \in \Delta_n$,

$$\left|x^\top (A - B)y\right| \le \|A - B\|_\infty. \tag{C.28}$$

Taking maximin values on the two sides of

$$x^\top By - \|A - B\|_\infty \le x^\top Ay \le x^\top By + \|A - B\|_\infty$$

proves (C.24).

For fixed y,

$$\begin{aligned}\left|\max_i e_i^\top Ay - \max_i e_i^\top By\right| &\le \max_i \left|e_i^\top (A - B)y\right| \\ &\le \|A - B\|_\infty.\end{aligned}$$

Likewise,

$$\left|\min_j x^\top Ae_j - \min_j x^\top Be_j\right| \le \|A - B\|_\infty.$$

Their sum proves (C.25).

For fixed A,

$$\begin{aligned}\left|\max_i e_i^\top Ay - \max_i e_i^\top Ay'\right| &\le \|A(y - y')\|_\infty \\ &\le \|A\|_\infty \|y - y'\|_1,\end{aligned}$$

and similarly

$$\left|\min_j x^\top Ae_j - \min_j x'^\top Ae_j\right| \le \|A\|_\infty \|x - x'\|_1.$$

This proves (C.26). Finally, apply the triangle inequality through either $g_A(z')$ or $g_B(z)$ and use the larger of $\|A\|_\infty$ and $\|B\|_\infty$. □

Corollary C.8 (Perturbation sandwich for approximate equilibria)**.** *Let $A, B \in \mathbb{R}^{m\times n}$, let $\delta := \|A - B\|_\infty$, and let $\tau \ge 0$. Then*

$$\mathcal{E}_{\tau-2\delta}(B) \subseteq \mathcal{E}_\tau(A) \subseteq \mathcal{E}_{\tau+2\delta}(B), \tag{C.29}$$

where the negative-tolerance convention (A.9) *is in force. Moreover,*

$$e\big(\mathcal{E}_\tau(B), \mathcal{E}_0(A)\big) \le \kappa_{\mathrm{eq}}(A)\big(\tau + 2\delta\big). \tag{C.30}$$

In particular,

$$e\big(\mathcal{E}_0(B), \mathcal{E}_0(A)\big) \le 2\kappa_{\mathrm{eq}}(A)\|A - B\|_\infty. \tag{C.31}$$

Proof. The two inclusions in (C.29) follow immediately from (C.25). If $z \in \mathcal{E}_\tau(B)$, then

$$g_A(z) \le g_B(z) + 2\delta \le \tau + 2\delta.$$

The definition of $\kappa_{\mathrm{eq}}(A)$ therefore gives

$$\mathrm{dist}\big(z, \mathcal{E}_0(A)\big) \le \kappa_{\mathrm{eq}}(A)\big(\tau + 2\delta\big).$$

Taking the supremum over $z \in \mathcal{E}_\tau(B)$ proves (C.30); setting $\tau = 0$ proves (C.31). □

Remark C.9 (Why the bound is directed). The constant in (C.31) is evaluated at the target base game A. Interchanging A and B gives the reverse excess with $\kappa_{\mathrm{eq}}(B)$, but no uniform two-sided bound follows when the neighboring moduli diverge. In the guiding family, $\kappa_{\mathrm{eq}}(A_0) = 2$ while $\kappa_{\mathrm{eq}}(A_\theta) \gtrsim |\theta|^{-1}$. This is exactly the geometry behind the asymmetry

$$e\big(\mathcal{E}_0(A_\theta), \mathcal{E}_0(A_0)\big) \longrightarrow 0, \qquad e\big(\mathcal{E}_0(A_0), \mathcal{E}_0(A_\theta)\big) \not\longrightarrow 0.$$

Thus a polyhedral residual error bound guarantees first-order validity relative to a fixed base game, but not completeness across a bifurcation.

C.4 Uniform localization near the equilibrium set

Fix a base game $A_\star \in \mathbb{R}^{m\times n}$, and abbreviate

$$\mathcal{E}_\star := \mathcal{E}_0(A_\star), \qquad \kappa_\star := \kappa_{\mathrm{eq}}(A_\star). \tag{C.32}$$

Let $\mathcal{H} \subseteq \mathbb{R}^{m\times n}$ be nonempty and bounded and define

$$R_{\mathcal{H}} := \sup_{H\in\mathcal{H}} \|H\|_\infty < \infty. \tag{C.33}$$

Proposition C.10 (Uniform first-order localization)**.** *For every $t \geq 0$, every $H \in \mathcal{H}$, every $\tau \geq 0$, and every $z \in \mathcal{E}_\tau(A_\star + tH)$,*

$$\operatorname{dist}(z, \mathcal{E}_\star) \leq \kappa_\star \left(\tau + 2tR_{\mathcal{H}}\right). \tag{C.34}$$

Equivalently,

$$e\left(\mathcal{E}_\tau(A_\star + tH), \mathcal{E}_\star\right) \leq \kappa_\star \left(\tau + 2tR_{\mathcal{H}}\right) \tag{C.35}$$

uniformly in $H \in \mathcal{H}$.

Proof. Apply (C.30) with $A = A_\star$, $B = A_\star + tH$, and

$$\|A - B\|_\infty = t\|H\|_\infty \leq tR_{\mathcal{H}}.$$

□

Corollary C.11 (Localization of critical-resolution sequences)**.** *Let $t_k \downarrow 0$, let $H_k \in \mathcal{H}$, and let $z_k \in \mathcal{Z}$. Suppose that, for some finite M,*

$$g_{A_\star + t_k H_k}(z_k) \leq Mt_k \qquad \textit{for every } k. \tag{C.36}$$

Then there exist points $\bar{z}_k \in \mathcal{E}_\star$ such that

$$\|z_k - \bar{z}_k\|_{\mathsf{V}} \leq \kappa_\star\big(M + 2R_{\mathcal{H}}\big)t_k. \tag{C.37}$$

Consequently, every accumulation point of (z_k) belongs to $\mathcal{E}_\star$. More precisely, every subsequence admits a further subsequence along which

$$\bar{z}_k \to z \in \mathcal{E}_\star, \qquad z_k \to z. \tag{C.38}$$

Proof. Condition (C.36) means that $z_k \in \mathcal{E}_{Mt_k}(A_\star + t_k H_k)$. Proposition C.10 yields

$$\operatorname{dist}(z_k, \mathcal{E}_\star) \leq \kappa_\star\big(M + 2R_{\mathcal{H}}\big)t_k.$$

Because $\mathcal{E}_\star$ is compact, the distance is attained by some $\bar{z}_k \in \mathcal{E}_\star$, proving (C.37). Compactness of $\mathcal{E}_\star$ gives a convergent subsequence of $(\bar{z}_k)$; the distance estimate and $t_k \downarrow 0$ force (z_k) to have the same limit. □

Corollary C.12 (Uniform outer localization of perturbed equilibria)**.** *For every bounded direction class $\mathcal{H}$,*

$$\sup_{H\in\mathcal{H}} e\left(\mathcal{E}_0(A_\star + tH), \mathcal{E}_0(A_\star)\right) \leq 2\kappa_\star R_{\mathcal{H}} t. \tag{C.39}$$

In particular,

$$\left\{ z \in \mathcal{Z} : \begin{array}{l} \exists\, t_k \downarrow 0,\ \exists\, H_k \in \mathcal{H},\ \exists\, z_k \in \mathcal{E}_0(A_\star + t_k H_k) \\ \textit{such that } z_k \to z \end{array} \right\} \subseteq \mathcal{E}_0(A_\star). \tag{C.40}$$

Proof. Set $\tau = 0$ in (C.35). The outer-limit statement then follows directly from the definition (A.18). □

Remark C.13 (Localization does not remove tangential drift). Corollary C.11 controls the distance from z_k to the *set* $\mathcal{E}_\star$ at scale t_k. If $\mathcal{E}_\star$ has positive dimension, it does not imply that $(z_k - z)/t_k$ is bounded for a fixed $z \in \mathcal{E}_\star$: the nearest points $\bar{z}_k$ may drift along the equilibrium face at a scale larger than t_k. Subsequent directional arguments must therefore account explicitly for motion of the base equilibrium point; this issue cannot be bypassed by the Hoffman bound alone.

D Directional Calculus and Feasible Tangent Paths

This appendix develops the first-order calculus underlying the branch-survival geometry. The saddle residual is a finite max–min of bilinear forms, and hence is Hadamard directionally differentiable under simultaneous perturbations of the payoff matrix and both mixed strategies. The fixed-base formula is a finite-dimensional instance of Danskin-type envelope calculus (Danskin, 1967; Bonnans & Shapiro, 2000); we give complete proofs because later arguments require a one-sided expansion that remains valid for possibly unbounded tangent velocities. Our tangent and normal-cone conventions agree with Rockafellar & Wets (1998).

The main conclusion is an exact sequential characterization of the branch-survival cost. If $z \in \mathcal{E}_0(A)$, then the smallest possible first-order residual under a payoff direction H is obtained by minimizing the directional saddle-gap formula over the entire tangent cone $T_{\mathcal{Z}}(z)$. Crucially, the lower bound is stable even when $(z_t - z)/t$ diverges because of tangential drift along a positive-dimensional equilibrium face: its error depends on the actual displacement $\|z_t - z\|_{\mathsf{V}}$, not on the magnitude of the normalized velocity.

D.1 Active best responses and complementary support

For $A \in \mathbb{R}^{m\times n}$, $x \in \Delta_m$, and $y \in \Delta_n$, define the active row and column sets

$$I_A(y) := \arg\max_{i\in[m]} e_i^\top Ay, \tag{D.1}$$

$$J_A(x) := \arg\min_{j\in[n]} x^\top Ae_j. \tag{D.2}$$

Define the corresponding nonnegative slacks

$$\rho_i^A(y) := \max_{k\in[m]} e_k^\top Ay - e_i^\top Ay, \qquad i \in [m], \tag{D.3}$$

$$\sigma_j^A(x) := x^\top Ae_j - \min_{\ell\in[n]} x^\top Ae_\ell, \qquad j \in [n]. \tag{D.4}$$

Thus,

$$I_A(y) = \big\{i : \rho_i^A(y) = 0\big\}, \qquad J_A(x) = \big\{j : \sigma_j^A(x) = 0\big\}. \tag{D.5}$$

For later uniformity statements, set

$$\delta_I(A, y) := \min_{i\notin I_A(y)} \rho_i^A(y), \tag{D.6}$$

$$\delta_J(A, x) := \min_{j\notin J_A(x)} \sigma_j^A(x), \tag{D.7}$$

with the convention that the minimum of the empty set is $+\infty$. Because the action sets are finite, every finite active margin is strictly positive.

Lemma D.1 (Complementary support at a saddle point)**.** *Let $z = (x, y) \in \mathcal{E}_0(A)$. Then*

$$\max_{i\in[m]} e_i^\top Ay = x^\top Ay = \min_{j\in[n]} x^\top Ae_j = v_A, \tag{D.8}$$

and

$$\operatorname{supp}(x) \subseteq I_A(y), \qquad \operatorname{supp}(y) \subseteq J_A(x). \tag{D.9}$$

Proof. The saddle inequalities imply

$$\max_i e_i^\top Ay = \min_j x^\top Ae_j,$$

and weak duality from Lemma C.1 forces their common value to be v_A. Since

$$x^\top Ay = \sum_{i=1}^m x_i e_i^\top Ay = v_A$$

while every term satisfies $e_i^\top Ay \le v_A$, each index with $x_i > 0$ must attain v_A. This proves $\mathrm{supp}(x) \subseteq I_A(y)$. Similarly,

$$x^\top Ay = \sum_{j=1}^n y_j x^\top Ae_j = v_A$$

while every term satisfies $x^\top Ae_j \ge v_A$, so every index with $y_j > 0$ belongs to $J_A(x)$. □

D.2 Directional derivative of the saddle residual

For $A, H \in \mathbb{R}^{m\times n}$, $z = (x, y) \in \mathcal{Z}$, and $d = (u, v) \in \mathsf{V}$, define the joint first-order form

$$\begin{aligned}\dot g_{A,z}(H; u, v) := &\max_{i\in I_A(y)} \left\{e_i^\top Hy + e_i^\top Av\right\} \\ &- \min_{j\in J_A(x)} \left\{x^\top He_j + u^\top Ae_j\right\}. \end{aligned} \tag{D.10}$$

The notation emphasizes that H is a payoff direction and (u, v) is a strategy direction.

Lemma D.2 (Finite-extremum expansion)**.** *Let $a \in \mathbb{R}^r$, let $t_k \downarrow 0$, and suppose $b_k \to b$ and c_k is bounded in $\mathbb{R}^r$. If*

$$I(a) := \arg\max_{i\in[r]} a_i,$$

then

$$\frac{\max_{i\in[r]} \left(a_i + t_k b_{k,i} + t_k^2 c_{k,i}\right) - \max_{i\in[r]} a_i}{t_k} \longrightarrow \max_{i\in I(a)} b_i. \tag{D.11}$$

If

$$J(a) := \arg\min_{i\in[r]} a_i,$$

then

$$\frac{\min_{i\in[r]} \left(a_i + t_k b_{k,i} + t_k^2 c_{k,i}\right) - \min_{i\in[r]} a_i}{t_k} \longrightarrow \min_{i\in J(a)} b_i. \tag{D.12}$$

Proof. We prove the maximum statement; the minimum statement follows by applying it to $-a$. Let

$$M := \max_i a_i.$$

If $I(a) \ne [r]$, define

$$\delta := \min_{i\notin I(a)} (M - a_i) > 0;$$

otherwise set $\delta = +\infty$. Since b_k and c_k are bounded, there is a finite B such that

$$|b_{k,i} + t_k c_{k,i}| \le B$$

for all sufficiently large k and every i. For such k, if $2t_k B < \delta$, no inactive index can maximize: every inactive value is at most $M - \delta + t_k B$, while every active value is at least $M - t_k B$. Hence, eventually,

$$\begin{aligned}&\max_{i\in[r]} \left(a_i + t_k b_{k,i} + t_k^2 c_{k,i}\right) \\ &\qquad = M + t_k \max_{i\in I(a)} \left(b_{k,i} + t_k c_{k,i}\right).\end{aligned}$$

The desired limit follows from finiteness of $I(a)$, convergence $b_k \to b$, and boundedness of c_k. □

Proposition D.3 (Hadamard directional differentiability)**.** *Fix $A \in \mathbb{R}^{m\times n}$ and $z = (x, y) \in \mathcal{Z}$. Let $t_k \downarrow 0$, $H_k \to H$, $u_k \to u$, and $v_k \to v$. Then*

$$\frac{g_{A+t_kH_k}\big(x + t_ku_k, y + t_kv_k\big) - g_A(x, y)}{t_k} \longrightarrow \dot{g}_{A,z}(H; u, v). \tag{D.13}$$

The conclusion is algebraic and does not require feasibility of the perturbed strategies. If the perturbed strategies belong to the simplices, it is the one-sided Hadamard directional derivative of the saddle residual restricted to $\mathcal{Z}$.

Proof. For each row index i,

$$e_i^\top (A + t_kH_k)(y + t_kv_k) = e_i^\top Ay + t_k\left(e_i^\top H_ky + e_i^\top Av_k\right) + t_k^2 e_i^\top H_kv_k. \tag{D.14}$$

The vectors of first-order coefficients converge to

$$\left(e_i^\top Hy + e_i^\top Av\right)_{i=1}^m,$$

and the second-order coefficients are bounded. Applying Lemma D.2 gives

$$\frac{\max_i e_i^\top (A + t_kH_k)(y + t_kv_k) - \max_i e_i^\top Ay}{t_k} \longrightarrow \max_{i\in I_A(y)} \left\{e_i^\top Hy + e_i^\top Av\right\}. \tag{D.15}$$

Similarly, for each column index j,

$$(x + t_ku_k)^\top (A + t_kH_k)e_j = x^\top Ae_j + t_k\left(x^\top H_ke_j + u_k^\top Ae_j\right) + t_k^2 u_k^\top H_ke_j, \tag{D.16}$$

and the finite-minimum part of Lemma D.2 yields the corresponding limit over $J_A(x)$. Subtracting the two limits proves (D.13). □

The preceding limit is locally uniform on bounded direction sets once the base point is fixed. We record an explicit version because it is used in the recovery-sequence construction.

Proposition D.4 (Uniform fixed-base expansion)**.** *Fix $A \in \mathbb{R}^{m\times n}$, $z = (x, y) \in \mathcal{Z}$, and finite constants $R_H, R_u, R_v \geq 0$. Define*

$$B_y := R_H + \|A\|_\infty R_v + R_HR_v, \tag{D.17}$$

$$B_x := R_H + \|A\|_\infty R_u + R_HR_u. \tag{D.18}$$

There exists $\bar{t} > 0$, depending only on A, z, R_H, R_u, R_v, such that for every $t \in (0, \bar{t}]$, every H, u, v satisfying

$$\|H\|_\infty \leq R_H, \qquad \|u\|_1 \leq R_u, \qquad \|v\|_1 \leq R_v,$$

one has

$$|g_{A+tH}(x + tu, y + tv) - g_A(x, y) - t\dot{g}_{A,z}(H; u, v)| \leq t^2R_H(R_u + R_v). \tag{D.19}$$

One admissible choice is any $\bar{t} \leq 1$ satisfying

$$2\bar{t}B_y < \delta_I(A, y) \quad \text{and} \quad 2\bar{t}B_x < \delta_J(A, x), \tag{D.20}$$

with inequalities involving $+\infty$ interpreted as vacuous.

Proof. For any row index i, the absolute change in $e_i^\top Ay$ caused by the simultaneous perturbation is at most

$$\begin{aligned} &t|e_i^\top Hy| + t|e_i^\top Av| + t^2|e_i^\top Hv| \\ &\quad \leq t\left(R_H + \|A\|_\infty R_v + R_HR_v\right) = tB_y, \end{aligned}$$

where $t \leq 1$ was used. Hence (D.20) ensures that no row outside $I_A(y)$ can overtake an active row. The perturbed row maximum is therefore attained within $I_A(y)$. An identical argument using B_x shows that the perturbed column minimum is attained within $J_A(x)$. Consequently,

$$\begin{aligned} &g_{A+tH}(x + tu, y + tv) - g_A(x, y) \\ &= t \max_{i\in I_A(y)} \left\{e_i^\top Hy + e_i^\top Av + te_i^\top Hv\right\} \\ &\quad - t \min_{j\in J_A(x)} \left\{x^\top He_j + u^\top Ae_j + tu^\top He_j\right\}. \end{aligned}$$

The maximum changes by at most $t\|H\|_\infty\|v\|_1$ after deleting its $te_i^\top Hv$ terms, while the minimum changes by at most $t\|H\|_\infty\|u\|_1$. Multiplying by the outer factor t proves (D.19). □

D.3 Feasible paths in the product simplex

For $p \in \Delta_r$ and $d \in \mathbb{R}^r$, define the maximal affine step

$$\tau_p(d) := \min_{\substack{i \in [r] \\ d_i < 0}} \frac{p_i}{-d_i}, \tag{D.21}$$

with the minimum of the empty set equal to $+\infty$.

Lemma D.5 (Exact affine realization of simplex tangents)**.** *Let $p \in \Delta_r$ and $d \in \mathbb{R}^r$. Then*

$$d \in T_{\Delta_r}(p) \quad \Longleftrightarrow \quad \tau_p(d) > 0 \text{ and } p + td \in \Delta_r \text{ for every } t \in [0, \tau_p(d)]. \tag{D.22}$$

If $d \notin T_{\Delta_r}(p)$, no sequence $t_k \downarrow 0$ can satisfy $p + t_k d \in \Delta_r$ for all k.

Proof. By Lemma A.9, membership in the tangent cone is equivalent to

$$\mathbf{1}_r^\top d = 0, \qquad d_i \geq 0 \quad \text{whenever } p_i = 0.$$

Thus every index with $d_i < 0$ has $p_i > 0$, so $\tau_p(d) > 0$. For $0 \leq t \leq \tau_p(d)$, each coordinate of $p + td$ is nonnegative and

$$\mathbf{1}_r^\top (p + td) = 1.$$

This proves feasibility. Conversely, if $p + t_k d \in \Delta_r$ along a sequence $t_k \downarrow 0$, then the equality constraint gives $\mathbf{1}_r^\top d = 0$, and any coordinate with $p_i = 0$ must have $d_i \geq 0$. Hence d is tangent. □

For $z = (x, y) \in \mathcal{Z}$ and $d = (u, v) \in \mathsf{V}$, define

$$\tau_z(d) := \min\{\tau_x(u), \tau_y(v)\}. \tag{D.23}$$

Corollary D.6 (Product-simplex paths and arbitrary feasible increments)**.** *For $z \in \mathcal{Z}$ and $d \in \mathsf{V}$,*

$$d \in T_{\mathcal{Z}}(z) \quad \Longleftrightarrow \quad z + td \in \mathcal{Z} \text{ for every } t \in [0, \tau_z(d)]. \tag{D.24}$$

Moreover, if $t > 0$ and $z_t \in \mathcal{Z}$, then

$$d_t := \frac{z_t - z}{t} \in T_{\mathcal{Z}}(z). \tag{D.25}$$

This implication holds without any boundedness assumption on d_t.

Proof. The first assertion follows coordinatewise from Lemma D.5 and Corollary A.10. For the second, convexity of $\mathcal{Z}$ implies

$$d_t = \frac{1}{t}(z_t - z) \in \operatorname{cone}(\mathcal{Z} - z) \subseteq T_{\mathcal{Z}}(z).$$

□

Lemma D.7 (Uniform feasible radius on bounded tangent sets)**.** *Fix $p \in \Delta_r$, and let $D \subseteq T_{\Delta_r}(p)$ satisfy*

$$R_D := \sup_{d \in D} \|d\|_\infty < \infty.$$

If $R_D = 0$, then $p + td = p$ for every $d \in D$ and every $t \geq 0$. If $R_D > 0$, define

$$p_{\min} := \min_{i \in \operatorname{supp}(p)} p_i > 0. \tag{D.26}$$

Then

$$p + td \in \Delta_r \qquad \text{for every } d \in D \text{ and every } 0 \leq t \leq \frac{p_{\min}}{R_D}. \tag{D.27}$$

Proof. A tangent direction can be negative only on $\operatorname{supp}(p)$. Therefore, for $i \in \operatorname{supp}(p)$,

$$p_i + td_i \geq p_{\min} - tR_D \geq 0,$$

while coordinates outside the support remain nonnegative because their tangent components are nonnegative. The simplex equality is preserved because every tangent direction sums to zero. □

D.4 THE BRANCH-SURVIVAL DIRECTIONAL VALUE

For $z \in \mathcal{E}_0(A)$ and $H \in \mathbb{R}^{m\times n}$, define the branch-survival directional value

$$\beta_A(z;H) := \inf_{(u,v)\in T_{\mathcal{Z}}(z)} \dot{g}_{A,z}(H;u,v). \tag{D.28}$$

For $z \notin \mathcal{E}_0(A)$, set

$$\beta_A(z;H) := +\infty. \tag{D.29}$$

The tangent-duality appendix derives a finite-dimensional primal–dual linear program for (D.28). Here we establish the analytic properties needed for the epi-resolution theorem.

Proposition D.8 (Basic properties of the branch-survival value)**.** *Fix A and $z \in \mathcal{E}_0(A)$. Then:*

1. *for every H,*
$$0 \le \beta_A(z;H) \le 2\|H\|_\infty; \tag{D.30}$$
2. *the map $H \mapsto \beta_A(z;H)$ is 2-Lipschitz under $\|\cdot\|_\infty$:*
$$|\beta_A(z;H) - \beta_A(z;H')| \le 2\|H - H'\|_\infty; \tag{D.31}$$
3. *the map $H \mapsto \beta_A(z;H)$ is convex and positively homogeneous, hence sublinear.*

Proof. Let $d = (u,v) \in T_{\mathcal{Z}}(z)$. By Corollary D.6, the path $z + td$ is feasible for all sufficiently small $t > 0$. Since the saddle residual is nonnegative and $g_A(z) = 0$, Proposition D.3 gives

$$\dot{g}_{A,z}(H;u,v) = \lim_{t\downarrow 0} \frac{g_{A+tH}(z+td)}{t} \ge 0.$$

Taking the infimum proves the lower bound in (D.30).

For the upper bound, take $u = v = 0$. By Lemma D.1,

$$\max_{i\in I_A(y)} e_i^\top H y \ge x^\top H y \ge \min_{j\in J_A(x)} x^\top H e_j,$$

so the resulting directional value is nonnegative, and each of its two terms has absolute value at most $\|H\|_\infty$. Hence it is at most $2\|H\|_\infty$.

For any fixed tangent direction d, the finite max–min formula gives

$$|\dot{g}_{A,z}(H;d) - \dot{g}_{A,z}(H';d)| \le 2\|H - H'\|_\infty. \tag{D.32}$$

Taking approximate minimizers under H and H' and applying (D.32) in both directions proves (D.31).

Finally, $(H,d) \mapsto \dot{g}_{A,z}(H;d)$ is convex because it is the sum of two finite maxima of linear functions. To verify convexity of the value function directly, fix H_1, H_2, $\lambda \in [0,1]$, and $\varepsilon > 0$. Choose d_1, d_2 in the tangent cone with

$$\dot{g}_{A,z}(H_r;d_r) \le \beta_A(z;H_r) + \varepsilon, \qquad r \in \{1,2\}.$$

The convex combination $d_\lambda := \lambda d_1 + (1-\lambda)d_2$ remains tangent, and joint convexity gives

$$\begin{aligned}\beta_A\big(z;\lambda H_1 + (1-\lambda)H_2\big) &\le \dot{g}_{A,z}\big(\lambda H_1 + (1-\lambda)H_2; d_\lambda\big)\\ &\le \lambda\beta_A(z;H_1) + (1-\lambda)\beta_A(z;H_2) + \varepsilon.\end{aligned}$$

Letting $\varepsilon \downarrow 0$ proves convexity. For $\alpha > 0$, conicity of the tangent cone and joint positive homogeneity give

$$\begin{aligned}\beta_A(z;\alpha H) &= \inf_d \dot{g}_{A,z}(\alpha H; d)\\ &= \inf_{d'} \dot{g}_{A,z}(\alpha H; \alpha d')\\ &= \alpha \inf_{d'} \dot{g}_{A,z}(H; d')\\ &= \alpha\beta_A(z;H).\end{aligned}$$

The case $\alpha = 0$ follows from (D.30). □

The next inequality is the key device for handling tangential drift. Unlike the two-sided expansion in Proposition D.4, it requires no active-set margin and no bound on the normalized tangent direction.

Proposition D.9 (One-sided support inequality)**.** *Let $z = (x, y) \in \mathcal{E}_0(A)$, let $H \in \mathbb{R}^{m\times n}$, let $d = (u, v) \in T_{\mathcal{Z}}(z)$, and let $t \in (0, \tau_z(d)]$. Then*

$$\frac{g_{A+tH}(z+td)}{t} \geq \dot{g}_{A,z}(H; u, v) - t\|H\|_\infty \left(\|u\|_1 + \|v\|_1\right). \tag{D.33}$$

Equivalently,

$$\frac{g_{A+tH}(z+td)}{t} \geq \dot{g}_{A,z}(H; u, v) - \|H\|_\infty \|td\|_{\mathsf{V}}. \tag{D.34}$$

Proof. Let v_A denote the game value. For every active row $i \in I_A(y)$, Lemma D.1 gives $e_i^\top Ay = v_A$. Hence

$$\begin{aligned}
&\max_{k\in[m]} e_k^\top (A + tH)(y + tv) \\
&\quad \geq v_A + t \max_{i\in I_A(y)} \left\{e_i^\top Hy + e_i^\top Av + te_i^\top Hv\right\} \\
&\quad \geq v_A + t \max_{i\in I_A(y)} \left\{e_i^\top Hy + e_i^\top Av\right\} - t^2\|H\|_\infty\|v\|_1.
\end{aligned} \tag{D.35}$$

Likewise, every active column $j \in J_A(x)$ satisfies $x^\top Ae_j = v_A$, and therefore

$$\begin{aligned}
&\min_{\ell\in[n]} (x + tu)^\top (A + tH)e_\ell \\
&\quad \leq v_A + t \min_{j\in J_A(x)} \left\{x^\top He_j + u^\top Ae_j + tu^\top He_j\right\} \\
&\quad \leq v_A + t \min_{j\in J_A(x)} \left\{x^\top He_j + u^\top Ae_j\right\} + t^2\|H\|_\infty\|u\|_1.
\end{aligned} \tag{D.36}$$

Subtracting (D.36) from (D.35) and dividing by t proves the claim. □

Proposition D.10 (Sequential liminf bound with unbounded velocities)**.** *Fix A, H, and $z \in \mathcal{E}_0(A)$. Let $t_k \downarrow 0$, $H_k \to H$, and $z_k \in \mathcal{Z}$ satisfy $z_k \to z$. Then*

$$\liminf_{k\to\infty} \frac{g_{A+t_kH_k}(z_k)}{t_k} \geq \beta_A(z; H). \tag{D.37}$$

No boundedness assumption is imposed on $(z_k - z)/t_k$.

Proof. Set

$$d_k := \frac{z_k - z}{t_k}.$$

Corollary D.6 gives $d_k \in T_{\mathcal{Z}}(z)$. Applying Proposition D.9 yields

$$\begin{aligned}
\frac{g_{A+t_kH_k}(z_k)}{t_k} &\geq \dot{g}_{A,z}(H_k; d_k) - \|H_k\|_\infty\|z_k - z\|_{\mathsf{V}} \\
&\geq \beta_A(z; H_k) - \|H_k\|_\infty\|z_k - z\|_{\mathsf{V}}.
\end{aligned} \tag{D.38}$$

The final term vanishes because (H_k) is bounded and $z_k \to z$, while Proposition D.8 gives

$$\beta_A(z; H_k) \longrightarrow \beta_A(z; H).$$

Taking lower limits proves (D.37). □

Proposition D.11 (Recovery along arbitrary perturbation scales)**.** *Fix A, H, and $z \in \mathcal{E}_0(A)$. For every pair of sequences $t_k \downarrow 0$ and $H_k \to H$, there exists a sequence $z_k \in \mathcal{Z}$ with $z_k \to z$ such that*

$$\limsup_{k\to\infty} \frac{g_{A+t_kH_k}(z_k)}{t_k} \leq \beta_A(z; H). \tag{D.39}$$

Proof. Let $\varepsilon_\ell := 1/\ell$. For each $\ell \geq 1$, choose $d^\ell = (u^\ell, v^\ell) \in T_{\mathcal{Z}}(z)$ such that

$$\dot{g}_{A,z}(H; d^\ell) \leq \beta_A(z; H) + \varepsilon_\ell. \tag{D.40}$$

Such a direction exists by the definition of the infimum.

For fixed ℓ, Proposition D.3, continuity in H, and Corollary D.6 imply the existence of $\eta_\ell > 0$ such that whenever

$$0 < t \leq \eta_\ell, \qquad \|H' - H\|_\infty \leq \eta_\ell,$$

one has $z + td^\ell \in \mathcal{Z}$ and

$$\frac{g_{A+tH'}(z + td^\ell)}{t} \leq \dot{g}_{A,z}(H; d^\ell) + \varepsilon_\ell. \tag{D.41}$$

By replacing η_ℓ with $\min_{r \leq \ell} \eta_r$, assume without loss of generality that (η_ℓ) is nonincreasing.

Since $t_k \to 0$ and $H_k \to H$, choose increasing integers K_ℓ such that for every $k \geq K_\ell$,

$$t_k \leq \eta_\ell, \qquad \|H_k - H\|_\infty \leq \eta_\ell, \qquad t_k \left\|d^\ell\right\|_{\mathsf{V}} \leq \varepsilon_\ell. \tag{D.42}$$

For $K_\ell \leq k < K_{\ell+1}$, set

$$z_k := z + t_k d^\ell.$$

Then $z_k \in \mathcal{Z}$, and (D.42) implies $z_k \to z$. Combining (D.40) and (D.41) gives

$$\frac{g_{A+t_k H_k}(z_k)}{t_k} \leq \beta_A(z; H) + 2\varepsilon_\ell$$

on the ℓth block. Since $\ell \to \infty$ with k, the desired limsup bound follows. □

Proposition D.12 (Exact sequential characterization)**.** *For every A, H and $z \in \mathcal{Z}$,*

$$\beta_A(z; H) = \inf_{\substack{t_k \downarrow 0,\ H_k \to H,\\ z_k \in \mathcal{Z},\ z_k \to z}} \liminf_{k \to \infty} \frac{g_{A+t_k H_k}(z_k)}{t_k}. \tag{D.43}$$

Here both sides are $+\infty$ when $z \notin \mathcal{E}_0(A)$.

Proof. If $z \in \mathcal{E}_0(A)$, the lower bound for every admissible sequence is Proposition D.10, and Proposition D.11 provides a sequence attaining the reverse inequality.

If $z \notin \mathcal{E}_0(A)$, continuity of $(A', z') \mapsto g_{A'}(z')$ and positivity of $g_A(z)$ imply that, for every $H_k \to H$ and $z_k \to z$,

$$g_{A+t_k H_k}(z_k) \longrightarrow g_A(z) > 0.$$

Division by $t_k \downarrow 0$ yields $+\infty$, agreeing with (D.29). □

Remark D.13 (Resolution of tangential drift). Appendix C showed that an $O(t)$-residual sequence is only known to lie within $O(t)$ of the equilibrium *set*; its motion along a positive-dimensional equilibrium face can be much larger than t. Proposition D.10 does not require a bounded normalized velocity. Indeed, the error in (D.34) is

$$\|H_k\|_\infty \|z_k - z\|_{\mathsf{V}} = o(1),$$

even when $\|(z_k - z)/t_k\|_{\mathsf{V}} \to \infty$. Thus the entire tangent cone in (D.28), including its recession directions, is the correct first-order object.

Remark D.14 (What remains for the duality appendix). Proposition D.12 identifies the branch-survival cost as the exact lower epi-derivative of the saddle residual. The tangent-duality appendix next converts its tangent-cone infimum into a finite-dimensional linear program and derives a dual representation as the support function of a compact polytope of critical strategic contrasts.

E TANGENT LINEAR PROGRAMS AND DUAL STRATEGIC CONTRASTS

This appendix converts the branch-survival value introduced in Appendix D into an explicit finite-dimensional linear program and derives its exact dual. Strong duality follows from the finite-dimensional minimax theorem of Sion (1958); all primal and dual optima are attained because the primal is a feasible linear program with finite value and the dual feasible region is compact. Standard polyhedral and linear-programming facts used below may be found, for example, in Schrijver (1986).

The dual variables admit a sharper game-theoretic interpretation. Their feasible sets coincide exactly with the row and column optimal-strategy polytopes of the base game. Consequently, the branch-survival value is the support function of a compact polytope of *critical strategic contrasts*. This contrast polytope is not an auxiliary relaxation: it is affinely isomorphic to the complete equilibrium correspondence itself. Thus the first-order statistical obstruction at a selected equilibrium branch is an exact linear image of equilibrium multiplicity.

E.1 PRIMAL TANGENT PROGRAM

Fix a matrix game $A \in \mathbb{R}^{m\times n}$, an equilibrium $z = (x, y) \in \mathcal{E}_0(A)$, and a payoff direction $H \in \mathbb{R}^{m\times n}$. Throughout this appendix, abbreviate

$$I := I_A(y), \qquad J := J_A(x), \qquad S_x := \operatorname{supp}(x), \qquad S_y := \operatorname{supp}(y). \tag{E.1}$$

For a nonempty index set $K \subseteq [r]$, define the corresponding face of the simplex by

$$\Delta_r(K) := \{p \in \Delta_r : p_i = 0 \text{ for every } i \notin K\}. \tag{E.2}$$

Consider the linear program

$$\begin{aligned}
\underset{u\in\mathbb{R}^m,\, v\in\mathbb{R}^n,\, a,b\in\mathbb{R}}{\text{minimize}} \quad & a + b && && \text{(E.3a)}\\
\text{subject to} \quad & e_i^\top Hy + e_i^\top Av \le a, && i \in I, && \text{(E.3b)}\\
& -x^\top He_j - u^\top Ae_j \le b, && j \in J, && \text{(E.3c)}\\
& \mathbf{1}_m^\top u = 0, \qquad u_i \ge 0, && i \notin S_x, && \text{(E.3d)}\\
& \mathbf{1}_n^\top v = 0, \qquad v_j \ge 0, && j \notin S_y. && \text{(E.3e)}
\end{aligned}$$

Denote its optimal value by $\operatorname{val}(\mathsf{P}_z(H))$.

Proposition E.1 (Exact primal representation and attainment)**.** *For every A, $z \in \mathcal{E}_0(A)$, and H, the program* (E.3) *is feasible, has a finite attained optimum, and*

$$\operatorname{val}(\mathsf{P}_z(H)) = \beta_A(z; H). \tag{E.4}$$

More precisely, if $(u^\star, v^\star)$ minimizes the tangent expression, then an optimal epigraph pair is

$$a^\star = \max_{i\in I} \left\{ e_i^\top Hy + e_i^\top Av^\star \right\}, \tag{E.5}$$

$$b^\star = \max_{j\in J} \left\{ -x^\top He_j - (u^\star)^\top Ae_j \right\}. \tag{E.6}$$

Proof. By Lemma A.9, constraints (E.3d)– (E.3e) are exactly

$$u \in T_{\Delta_m}(x), \qquad v \in T_{\Delta_n}(y).$$

For any such pair (u, v), minimizing over a and b gives

$$a = \max_{i\in I} \left\{ e_i^\top Hy + e_i^\top Av \right\},$$

$$b = \max_{j\in J} \left\{ -x^\top He_j - u^\top Ae_j \right\}.$$

Their sum is exactly $\dot g_{A,z}(H; u, v)$ from (D.10). Taking the infimum over the product tangent cone proves equality of the two infimal values.

The program is feasible, for example with $u = v = 0$ and sufficiently large a, b. Its value is finite because Proposition D.8 gives

$$0 \le \beta_A(z; H) \le 2\|H\|_\infty.$$

The feasible region in (E.3) is a nonempty polyhedron and the objective is linear. A feasible finite-dimensional linear program with finite optimal value attains its optimum; equivalently, this follows from the Minkowski–Weyl representation of its feasible polyhedron. The formulas (E.5)– (E.6) then follow from epigraph minimality. □

E.2 Dual program and strong duality

Define the row and column certificate sets

$$\mathcal{M}_A(z) := \{\mu \in \Delta_m(I) : A^\top \mu \in T_{\Delta_n}(y)^+\}, \tag{E.7}$$

$$\mathcal{N}_A(z) := \{\nu \in \Delta_n(J) : -A\nu \in T_{\Delta_m}(x)^+\}. \tag{E.8}$$

The corresponding dual program is

$$\underset{\substack{\mu\in\mathcal{M}_A(z)\\ \nu\in\mathcal{N}_A(z)}}{\text{maximize}} \quad \mu^\top H y - x^\top H \nu. \tag{$\mathsf{D}_z(H)$}$$

Lemma E.2 (Nonemptiness of the dual certificate sets)**.** *The selected equilibrium strategies satisfy*

$$x \in \mathcal{M}_A(z), \qquad y \in \mathcal{N}_A(z). \tag{E.9}$$

Consequently, the dual feasible region is nonempty and contains a feasible point of objective value zero.

Proof. Complementary support from Lemma D.1 gives

$$x \in \Delta_m(I), \qquad y \in \Delta_n(J).$$

Because x is row optimal,

$$A^\top x \geq v_A \mathbf{1}_n.$$

Every coordinate in S_y equals v_A, since $S_y \subseteq J_A(x)$. The simplex positive-dual characterization (A.33) therefore gives

$$A^\top x \in T_{\Delta_n}(y)^+.$$

Thus $x \in \mathcal{M}_A(z)$.

Similarly, column optimality gives

$$Ay \leq v_A \mathbf{1}_m,$$

with equality on $S_x \subseteq I_A(y)$. Hence

$$-Ay \in T_{\Delta_m}(x)^+,$$

so $y \in \mathcal{N}_A(z)$. At $(\mu, \nu) = (x, y)$, the dual objective is $x^\top H y - x^\top H y = 0$. □

Proposition E.3 (Strong tangent-program duality)**.** *For every A, $z \in \mathcal{E}_0(A)$, and H, the primal and dual programs attain their optima and have the common value*

$$\beta_A(z; H) = \min \mathsf{P}_z(H) = \max \mathsf{D}_z(H). \tag{E.10}$$

Equivalently,

$$\beta_A(z; H) = \max_{\substack{\mu\in\mathcal{M}_A(z)\\ \nu\in\mathcal{N}_A(z)}} \left\{\mu^\top H y - x^\top H \nu\right\}. \tag{E.11}$$

Proof. For $\mu \in \Delta_m(I)$, $\nu \in \Delta_n(J)$, and tangent directions $u \in T_{\Delta_m}(x)$, $v \in T_{\Delta_n}(y)$, define

$$\mathcal{L}_H(u, v; \mu, \nu) := \mu^\top H y - x^\top H \nu + \langle A^\top \mu, v\rangle + \langle -A\nu, u\rangle. \tag{E.12}$$

The elementary simplex identities

$$\max_{i\in I} r_i = \max_{\mu\in\Delta_m(I)} \mu^\top r, \qquad \max_{j\in J} s_j = \max_{\nu\in\Delta_n(J)} \nu^\top s$$

show that

$$\dot{g}_{A,z}(H; u, v) = \max_{\substack{\mu\in\Delta_m(I)\\ \nu\in\Delta_n(J)}} \mathcal{L}_H(u, v; \mu, \nu). \tag{E.13}$$

The set $\Delta_m(I) \times \Delta_n(J)$ is compact and convex, the product tangent cone is convex, and $\mathcal{L}_H$ is continuous and affine in each block of variables. Sion's minimax theorem therefore yields

$$\begin{aligned}\beta_A(z;H) &= \inf_{\substack{u\in T_{\Delta_m}(x)\\ v\in T_{\Delta_n}(y)}} \max_{\substack{\mu\in\Delta_m(I)\\ \nu\in\Delta_n(J)}} \mathcal{L}_H(u,v;\mu,\nu)\\ &= \max_{\substack{\mu\in\Delta_m(I)\\ \nu\in\Delta_n(J)}} \inf_{\substack{u\in T_{\Delta_m}(x)\\ v\in T_{\Delta_n}(y)}} \mathcal{L}_H(u,v;\mu,\nu). \end{aligned} \tag{E.14}$$

For a cone K and a vector w,

$$\inf_{d\in K}\langle w,d\rangle = \begin{cases} 0, & w\in K^+,\\ -\infty, & w\notin K^+. \end{cases} \tag{E.15}$$

Indeed, the value is zero at $d=0$ when w is nonnegative on K; otherwise a direction with negative pairing can be scaled arbitrarily. Applying (E.15) separately to u and v shows that the inner infimum in (E.14) equals $\mu^\top Hy - x^\top H\nu$ when $A^\top\mu \in T_{\Delta_n}(y)^+$ and $-A\nu \in T_{\Delta_m}(x)^+$, and equals $-\infty$ otherwise. This is exactly the dual program ($\mathsf{D}_z(H)$).

The dual feasible region is a closed subset of the compact set $\Delta_m(I)\times\Delta_n(J)$, and is nonempty by Lemma E.2; hence its maximum is attained. Primal attainment was established in Proposition E.1. □

Remark E.4 (Lagrange-multiplier interpretation). The variables μ and ν are the nonnegative multipliers of the active row and column epigraph constraints in (E.3). Minimization over the free epigraph variables forces $\mathbf{1}_m^\top\mu = \mathbf{1}_n^\top\nu = 1$, while minimization over the tangent directions produces the positive-dual cone constraints in (E.7)– (E.8). Thus ($\mathsf{D}_z(H)$) is also the ordinary finite-dimensional linear-programming dual of (E.3); the minimax proof above records the same duality without introducing auxiliary multipliers for the tangent-cone representation.

E.3 Equilibrium certificates and critical strategic contrasts

The dual certificate sets admit an exact simplification: they are the full optimal-strategy polytopes of the base game, independently of the selected equilibrium branch.

Proposition E.5 (Dual certificates are equilibrium strategies)**.** *For every $z=(x,y)\in\mathcal{E}_0(A)$,*

$$\mathcal{M}_A(z) = \mathcal{X}^\star(A), \qquad \mathcal{N}_A(z) = \mathcal{Y}^\star(A), \tag{E.16}$$

where $\mathcal{X}^\star(A)$ and $\mathcal{Y}^\star(A)$ are defined in (C.6)– (C.7)*.*

Proof. We first prove $\mathcal{M}_A(z)\subseteq\mathcal{X}^\star(A)$. Let $\mu\in\mathcal{M}_A(z)$. Since $\mu\in\Delta_m(I_A(y))$ and every active row earns v_A against y,

$$\mu^\top Ay = v_A. \tag{E.17}$$

By Lemma A.9, the condition

$$A^\top\mu \in T_{\Delta_n}(y)^+$$

means that there exists $\lambda\in\mathbb{R}$ such that

$$(A^\top\mu)_j = \lambda \quad (j\in S_y), \qquad (A^\top\mu)_j \ge \lambda \quad (j\notin S_y). \tag{E.18}$$

Averaging over y, whose support is S_y, and using (E.17) gives $\lambda = v_A$. Therefore

$$A^\top\mu \ge v_A\mathbf{1}_n,$$

which is precisely $\mu\in\mathcal{X}^\star(A)$.

Conversely, let $\mu\in\mathcal{X}^\star(A)$. Since $(\mu,y)\in\mathcal{X}^\star(A)\times\mathcal{Y}^\star(A)=\mathcal{E}_0(A)$,

$$\mu^\top Ay = v_A.$$

Every row payoff against the optimal strategy y is at most v_A. Thus a positive component of μ can occur only on an active row, which proves $\mu\in\Delta_m(I_A(y))$. Moreover, $A^\top\mu\ge v_A\mathbf{1}_n$. Its average under y equals v_A, so every coordinate in S_y equals v_A. The simplex positive-dual characterization again gives

$$A^\top\mu \in T_{\Delta_n}(y)^+.$$

Hence $\mu \in \mathcal{M}_A(z)$, proving the first equality.

The proof for the column player is symmetric. If $\nu \in \mathcal{N}_A(z)$, then support on $J_A(x)$ gives $x^\top A\nu = v_A$, while $-A\nu \in T_{\Delta_m}(x)^+$ implies that $A\nu$ is constant on S_x and no larger outside. The constant must be v_A, so $A\nu \le v_A \mathbf{1}_m$ and $\nu \in \mathcal{Y}^\star(A)$. Conversely, every $\nu \in \mathcal{Y}^\star(A)$ is supported on $J_A(x)$, satisfies $A\nu \le v_A \mathbf{1}_m$, and has equality on S_x; hence it belongs to $\mathcal{N}_A(z)$. □

Definition E.6 (Critical strategic contrast polytope)**.** For $z = (x, y) \in \mathcal{E}_0(A)$, define

$$\mathcal{C}_A(z) := \left\{\mu y^\top - x\nu^\top : \mu \in \mathcal{X}^\star(A),\ \nu \in \mathcal{Y}^\star(A)\right\} \subseteq \mathbb{R}^{m\times n}. \tag{E.19}$$

Each element compares the selected branch (x, y) with an alternative equilibrium pair (μ, ν): the first rank-one term changes the row marginal while holding y fixed, and the second changes the column marginal while holding x fixed.

Proposition E.7 (Support-function representation)**.** *For every $z = (x, y) \in \mathcal{E}_0(A)$ and every payoff direction H,*

$$\begin{aligned}
\beta_A(z; H) &= \max_{C\in\mathcal{C}_A(z)} \langle C, H\rangle_{\mathrm{F}} && \text{(E.20)}\\
&= \max_{\substack{\mu\in\mathcal{X}^\star(A)\\ \nu\in\mathcal{Y}^\star(A)}} \left\{\mu^\top H y - x^\top H\nu\right\} && \text{(E.21)}\\
&= \max_{\mu\in\mathcal{X}^\star(A)} (\mu - x)^\top H y + \max_{\nu\in\mathcal{Y}^\star(A)} x^\top H(y - \nu) && \text{(E.22)}\\
&= \max_{\mu\in\mathcal{X}^\star(A)} \mu^\top H y - \min_{\nu\in\mathcal{Y}^\star(A)} x^\top H\nu. && \text{(E.23)}
\end{aligned}$$

In particular, $\beta_A(z; \cdot)$ is the support function of the critical contrast polytope under the Frobenius pairing.

Proof. Substitute Proposition E.5 into the dual identity (E.11). The equality

$$\left\langle \mu y^\top - x\nu^\top, H\right\rangle_{\mathrm{F}} = \mu^\top H y - x^\top H\nu$$

proves (E.20) and (E.21). Adding and subtracting $x^\top H y$ gives

$$\mu^\top H y - x^\top H\nu = (\mu - x)^\top H y + x^\top H(y - \nu).$$

The feasible set is the Cartesian product $\mathcal{X}^\star(A) \times \mathcal{Y}^\star(A)$, so maximizing the two separated terms proves (E.22). The final expression is an equivalent rearrangement. □

Proposition E.8 (The contrast polytope is an affine copy of the equilibrium correspondence)**.** *Define*

$$\Phi_z : \mathcal{X}^\star(A) \times \mathcal{Y}^\star(A) \longrightarrow \mathbb{R}^{m\times n}, \qquad \Phi_z(\mu, \nu) := \mu y^\top - x\nu^\top. \tag{E.24}$$

Then Φ_z is an affine bijection from $\mathcal{E}_0(A) = \mathcal{X}^\star(A) \times \mathcal{Y}^\star(A)$ onto $\mathcal{C}_A(z)$. Its inverse is

$$C \longmapsto \left(x + C\mathbf{1}_n,\ y - C^\top \mathbf{1}_m\right). \tag{E.25}$$

Consequently, $\mathcal{C}_A(z)$ is a nonempty compact convex polytope, contains the origin, and

$$\dim \mathcal{C}_A(z) = \dim \mathcal{X}^\star(A) + \dim \mathcal{Y}^\star(A) = \dim \mathcal{E}_0(A), \tag{E.26}$$

where dimensions are affine dimensions.

Moreover, every $C \in \mathcal{C}_A(z)$ satisfies

$$\begin{aligned}
C\mathbf{1}_n &= \mu - x, && \text{(E.27)}\\
C^\top \mathbf{1}_m &= y - \nu, && \text{(E.28)}\\
\mathbf{1}_m^\top C \mathbf{1}_n &= 0, && \text{(E.29)}\\
\operatorname{rank}(C) &\le 2, && \text{(E.30)}\\
\sum_{i=1}^{m}\sum_{j=1}^{n} |C_{ij}| &\le 2. && \text{(E.31)}
\end{aligned}$$

Proof. Surjectivity is the definition of $\mathcal{C}_A(z)$. If $C = \Phi_z(\mu, \nu)$, then the probability normalizations give

$$C\mathbf{1}_n = \mu(y^\top \mathbf{1}_n) - x(\nu^\top \mathbf{1}_n) = \mu - x,$$
$$C^\top \mathbf{1}_m = y(\mu^\top \mathbf{1}_m) - \nu(x^\top \mathbf{1}_m) = y - \nu.$$

Thus (μ, ν) is recovered by (E.25), proving injectivity and the marginal identities. Since the domain is a compact convex polytope and Φ_z is affine, its image is a compact convex polytope. Taking $(\mu, \nu) = (x, y)$ gives $0 \in \mathcal{C}_A(z)$. Affine injectivity preserves affine dimension, proving (E.26).

Equation (E.29) follows from either marginal identity. Since C is the difference of two rank-one matrices, $\operatorname{rank}(C) \le 2$. Finally,

$$\sum_{i,j} |C_{ij}| \le \sum_{i,j} \mu_i y_j + \sum_{i,j} x_i \nu_j$$
$$= 1 + 1 = 2.$$

□

Corollary E.9 (Equilibrium multiplicity is exactly first-order contrast geometry)**.** *For any selected equilibrium $z \in \mathcal{E}_0(A)$, the following are equivalent:*

1. $\mathcal{E}_0(A) = \{z\}$;
2. $\mathcal{C}_A(z) = \{0\}$;
3. $\beta_A(z; H) = 0$ *for every* $H \in \mathbb{R}^{m \times n}$.

If the equilibrium correspondence is non-singleton, then there exists a payoff direction H for which $\beta_A(z; H) > 0$.

Proof. The equivalence of the first two statements follows from the affine bijection in Proposition E.8. If the contrast polytope is $\{0\}$, its support function vanishes identically. Conversely, if it contains a nonzero matrix C_0, then choosing $H = C_0$ gives

$$\beta_A(z; C_0) \ge \langle C_0, C_0 \rangle_{\mathrm{F}} > 0.$$

□

Remark E.10 (Residual resolution versus strategy displacement). Corollary E.9 concerns the first-order *residual budget* required for an equilibrium branch to survive. At a unique but singular equilibrium, the selected equilibrium may move under perturbation at a nonclassical or even non-Lipschitz rate, while its exact residual remains zero. Thus $\beta_A(z; H) = 0$ does not assert root-t or linear displacement of the equilibrium strategy; it asserts only that no positive first-order gap tolerance is required to retain an exact branch.

Corollary E.11 (First-order survival cone and critical subgradients)**.** *Define the exact first-order survival cone*

$$\mathfrak{S}_A(z) := \{H \in \mathbb{R}^{m \times n} : \beta_A(z; H) = 0\}. \tag{E.32}$$

Then

$$\mathfrak{S}_A(z) = \{H : \langle C, H \rangle_{\mathrm{F}} \le 0 \text{ for every } C \in \mathcal{C}_A(z)\}, \tag{E.33}$$

so $\mathfrak{S}_A(z)$ is a closed convex cone. Furthermore, the convex subdifferential with respect to H is

$$\partial_H \beta_A(z; H) = \operatorname*{arg\,max}_{C \in \mathcal{C}_A(z)} \langle C, H \rangle_{\mathrm{F}}, \tag{E.34}$$

and, in particular,

$$\partial_H \beta_A(z; 0) = \mathcal{C}_A(z). \tag{E.35}$$

Proof. Because $0 \in \mathcal{C}_A(z)$, its support function is nonnegative. It equals zero exactly when every contrast has nonpositive pairing with H, proving (E.33). The subdifferential identity is the standard exposed-face formula for the support function of a nonempty compact convex set; it also follows directly from the subgradient inequality and finite-dimensional separation. At $H = 0$, every contrast maximizes the zero linear functional. □

E.4 CLOSED-FORM EXAMPLES

Example E.1 (The guiding two-action bifurcation)**.** Let

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad z_p := \big(x(p), e_2\big), \qquad p \in [0,1].$$

Appendix B showed that

$$\mathcal{X}^\star(A_0) = \{x(r) : r \in [0,1]\}, \qquad \mathcal{Y}^\star(A_0) = \{e_2\}.$$

Therefore

$$\begin{aligned} \mathcal{C}_{A_0}(z_p) &= \big\{\big(x(r) - x(p)\big)e_2^\top : r \in [0,1]\big\} \\ &= \left\{(r-p)\begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} : r \in [0,1]\right\}. \end{aligned} \tag{E.36}$$

For an arbitrary direction $H \in \mathbb{R}^{2\times 2}$, set

$$d_H := (e_1 - e_2)^\top H e_2 = H_{12} - H_{22}.$$

The support-function formula gives

$$\beta_{A_0}(z_p; H) = (1-p)[d_H]_+ + p[-d_H]_+. \tag{E.37}$$

In particular, for the scalar perturbation $H = he_2e_2^\top$,

$$\beta_{A_0}\left(z_p; he_2e_2^\top\right) = h_+ p + h_-(1-p), \tag{E.38}$$

where $h_+ := \max\{h, 0\}$ and $h_- := \max\{-h, 0\}$. Hence positive perturbations preserve the $p = 0$ endpoint at zero first-order cost, negative perturbations preserve the $p = 1$ endpoint, and interior points require a strictly positive resolution budget in either transverse direction.

Example E.2 (The zero game)**.** If $A = 0$, then every strategy pair is an equilibrium:

$$\mathcal{X}^\star(0) = \Delta_m, \qquad \mathcal{Y}^\star(0) = \Delta_n.$$

For any selected $z = (x, y)$,

$$\mathcal{C}_0(z) = \big\{\mu y^\top - x\nu^\top : \mu \in \Delta_m,\ \nu \in \Delta_n\big\}, \tag{E.39}$$

and

$$\begin{aligned} \beta_0(z; H) &= \max_{\mu\in\Delta_m} \mu^\top H y - \min_{\nu\in\Delta_n} x^\top H \nu \\ &= \max_{i\in[m]} e_i^\top H y - \min_{j\in[n]} x^\top H e_j \\ &= g_H(x, y). \end{aligned} \tag{E.40}$$

Thus, at a maximally degenerate base game, the first-order branch-survival problem is itself the saddle-residual problem of the perturbation game.

Proposition E.12 (Lifted representation of the visible correspondence)**.** *Fix a base game A, write $v_A = \mathrm{val}(A)$, and fix a direction H and $\lambda \geq 0$. The set*

$$\{(x, y) \in \mathcal{X}^\star(A) \times \mathcal{Y}^\star(A) : \beta_A((x, y); H) \leq \lambda\} \tag{E.41}$$

is the projection onto (x, y) of the polyhedron in variables $(x, y, s, r, \alpha, \gamma)$ satisfying

$$x \in \mathcal{X}^\star(A), \qquad y \in \mathcal{Y}^\star(A), \quad s \in \mathbb{R}^n_+, \quad r \in \mathbb{R}^m_+, \tag{E.42}$$

$$\alpha \mathbf{1}_m - As \geq Hy, \qquad \gamma \mathbf{1}_n - A^\top r \leq H^\top x, \tag{E.43}$$

$$\alpha - v_A \mathbf{1}_n^\top s - \gamma + v_A \mathbf{1}_m^\top r \leq \lambda. \tag{E.44}$$

For fixed (x, y), membership can be decided by the two restricted-game linear programs in Proposition E.7. If membership fails and $\mu^\star, \nu^\star$ solve those programs, then

$$(\mu^\star)^\top H y' - x'^\top H \nu^\star \leq \lambda \tag{E.45}$$

is valid for every visible (x', y') and strictly separates the rejected point. Thus the visible correspondence has polynomial-size lifted LP, membership, separation, and linear-optimization oracles for fixed rational input data.

Proof. For fixed y, strong LP duality gives

$$\max_{\mu\in\mathcal{X}^\star(A)} \mu^\top H y = \min_{\substack{s\geq 0,\ \alpha\in\mathbb{R} \\ \alpha\mathbf{1}_m - As \geq Hy}} \left(\alpha - v_A \mathbf{1}_n^\top s\right). \tag{E.46}$$

Similarly,

$$\min_{\nu\in\mathcal{Y}^\star(A)} x^\top H \nu = \max_{\substack{r\geq 0,\ \gamma\in\mathbb{R} \\ \gamma\mathbf{1}_n - A^\top r \leq H^\top x}} \left(\gamma - v_A \mathbf{1}_m^\top r\right). \tag{E.47}$$

The lifted constraints therefore imply that the restricted-game gap is at most λ; conversely, optimal dual solutions yield feasible lifted variables whenever the gap is at most λ. For separation, every visible (x', y') satisfies

$$\max_{\mu\in\mathcal{X}^\star(A)} \mu^\top H y' - \min_{\nu\in\mathcal{Y}^\star(A)} {x'}^\top H \nu \leq \lambda,$$

which implies (E.45); at the rejected point it equals the violated restricted gap. □

Remark E.13 (Role of the support representation). The support-function representation has two uses in the retained paper: it converts the tangent epi-derivative into an explicit polyhedral object for the positive-resolution limit, and it identifies the payoff directions that govern statistical branch visibility.

F Proof of the Strategic Epi-Resolution Theorem

This appendix proves the set-valued first-order limit that underlies the statistical resolution theory. Fix a base game $A_\star \in \mathbb{R}^{m\times n}$. For $t > 0$, a payoff direction $H \in \mathbb{R}^{m\times n}$, and $z \in \mathcal{Z}$, define the rescaled saddle residual

$$\mathcal{F}_t(H, z) := \frac{g_{A_\star + tH}(z)}{t}. \tag{F.1}$$

Its candidate first-order limit is

$$\mathcal{F}(H, z) := \beta_{A_\star}(z; H), \tag{F.2}$$

where the branch-survival value is finite on $\mathcal{E}_0(A_\star)$ and equals $+\infty$ outside that set by (D.29).

The main result is stronger than pointwise epi-convergence. It allows the payoff direction to vary with t, identifies the complete positive-level sublevel-set limit, proves Hausdorff continuity of the limiting visible correspondence, and yields convergence uniformly over compact direction classes and compact resolution intervals bounded away from zero. Our set- and epi-convergence conventions follow Rockafellar & Wets (1998); all ingredients needed for the present finite-dimensional argument are proved explicitly below.

F.1 The restricted perturbation game

Abbreviate the row and column optimal-strategy polytopes of the base game by

$$\mathcal{X}_\star := \mathcal{X}^\star(A_\star), \qquad \mathcal{Y}_\star := \mathcal{Y}^\star(A_\star), \qquad \mathcal{E}_\star := \mathcal{E}_0(A_\star) = \mathcal{X}_\star \times \mathcal{Y}_\star. \tag{F.3}$$

For a direction $H \in \mathbb{R}^{m\times n}$ and $z = (x, y) \in \mathcal{E}_\star$, define the duality gap of the *restricted first-order game*

$$g^{(1)}_{\star,H}(x, y) := \max_{\mu\in\mathcal{X}_\star} \mu^\top H y - \min_{\nu\in\mathcal{Y}_\star} x^\top H \nu. \tag{F.4}$$

The strategy sets in this auxiliary game are the complete optimal-strategy polytopes of $A_\star$, while its payoff matrix is the perturbation direction H.

Proposition F.1 (Branch cost as a restricted-game gap)**.** *For every $H \in \mathbb{R}^{m\times n}$ and $z = (x, y) \in \mathcal{E}_\star$,*

$$\beta_{A_\star}(z; H) = g^{(1)}_{\star,H}(x, y). \tag{F.5}$$

Consequently, if

$$v^{(1)}_\star(H) := \max_{x\in\mathcal{X}_\star} \min_{y\in\mathcal{Y}_\star} x^\top H y = \min_{y\in\mathcal{Y}_\star} \max_{x\in\mathcal{X}_\star} x^\top H y, \tag{F.6}$$

then the zero-cost correspondence is the saddle-point set of the restricted game:

$$\{z \in \mathcal{E}_\star : \beta_{A_\star}(z; H) = 0\} = \mathcal{X}^{(1)}_\star(H) \times \mathcal{Y}^{(1)}_\star(H), \tag{F.7}$$

where

$$\mathcal{X}_\star^{(1)}(H) := \arg\max_{x\in\mathcal{X}_\star} \min_{y\in\mathcal{Y}_\star} x^\top H y, \tag{F.8}$$

$$\mathcal{Y}_\star^{(1)}(H) := \arg\min_{y\in\mathcal{Y}_\star} \max_{x\in\mathcal{X}_\star} x^\top H y. \tag{F.9}$$

In particular, the zero-cost correspondence is nonempty, compact, and convex for every H.

Proof. Equation (F.5) is precisely the max–min representation (E.23) with $A = A_\star$. The sets $\mathcal{X}_\star$ and $\mathcal{Y}_\star$ are nonempty compact convex polytopes. Sion's minimax theorem therefore gives (F.6). The same saddle-point argument as in Proposition A.2, now with strategy sets $\mathcal{X}_\star$ and $\mathcal{Y}_\star$, shows that the restricted gap vanishes exactly on $\mathcal{X}_\star^{(1)}(H) \times \mathcal{Y}_\star^{(1)}(H)$. □

Definition F.2 (Visible equilibrium correspondence)**.** For $\lambda \geq 0$ and $H \in \mathbb{R}^{m\times n}$, define

$$\mathcal{K}_\lambda(H) := \{z \in \mathcal{E}_\star : \beta_{A_\star}(z; H) \leq \lambda\}. \tag{F.10}$$

Equivalently,

$$\mathcal{K}_\lambda(H) = \Big\{z \in \mathcal{X}_\star \times \mathcal{Y}_\star : g^{(1)}_{\star,H}(z) \leq \lambda\Big\}. \tag{F.11}$$

Thus $\mathcal{K}_\lambda(H)$ is the complete λ-approximate saddle correspondence of the restricted first-order game.

Proposition F.3 (Geometry and continuity of the branch cost)**.** *The following statements hold.*

1. *For every H, the map*
$$z \longmapsto \beta_{A_\star}(z; H) \tag{F.12}$$
is finite, nonnegative, continuous, and convex on $\mathcal{E}_\star$.

2. *For all $z, z' \in \mathcal{E}_\star$ and all $H, H' \in \mathbb{R}^{m\times n}$,*
$$\begin{aligned}&|\beta_{A_\star}(z; H) - \beta_{A_\star}(z'; H')| \\ &\quad\leq 2\|H - H'\|_\infty + \max\{\|H\|_\infty, \|H'\|_\infty\}\|z - z'\|_{\mathsf{V}}.\end{aligned} \tag{F.13}$$

3. *For every $\lambda \geq 0$, the set $\mathcal{K}_\lambda(H)$ is nonempty, compact, and convex.*

4. *For every $\lambda > 0$,*
$$\mathrm{cl}\{z \in \mathcal{E}_\star : \beta_{A_\star}(z; H) < \lambda\} = \mathcal{K}_\lambda(H). \tag{F.14}$$

Proof. The restricted-gap representation (F.5) writes the branch cost as

$$\max_{\mu\in\mathcal{X}_\star} \mu^\top H y + \max_{\nu\in\mathcal{Y}_\star}(-x^\top H \nu). \tag{F.15}$$

For fixed H, this is a sum of two support functions composed with linear maps. It is therefore finite, continuous, and convex in (x, y). Nonnegativity follows from weak duality for the restricted game.

To prove the joint estimate, for $\mu \in \mathcal{X}_\star$ and $\nu \in \mathcal{Y}_\star$, define

$$\phi_{\mu,\nu}(z, H) := \mu^\top H y - x^\top H \nu.$$

Since x, y, μ, ν are probability vectors,

$$\begin{aligned}&|\phi_{\mu,\nu}(z, H) - \phi_{\mu,\nu}(z', H')| \\ &\quad\leq \big|\mu^\top (H - H')y\big| + \big|\mu^\top H'(y - y')\big| + \big|(x - x')^\top H'\nu\big| + \big|x^\top (H - H')\nu\big| \\ &\quad\leq 2\|H - H'\|_\infty + \|H'\|_\infty\|z - z'\|_{\mathsf{V}}.\end{aligned}$$

Interchanging (z, H) and (z', H') yields the same inequality with $\|H\|_\infty$ in place of $\|H'\|_\infty$. Taking the supremum over the fixed compact set $\mathcal{X}_\star \times \mathcal{Y}_\star$ proves (F.13).

Compactness and convexity of the visible correspondence now follow from compactness and convexity of $\mathcal{E}_\star$, continuity and convexity of the branch cost, and nonemptiness of the zero-cost set from Proposition F.1.

It remains to prove (F.14). Fix $z \in \mathcal{K}_\lambda(H)$, and choose $z_0 \in \mathcal{K}_0(H)$. For $\varepsilon \in (0,1]$, set

$$z_\varepsilon := (1-\varepsilon)z + \varepsilon z_0.$$

Then $z_\varepsilon \in \mathcal{E}_\star$, $z_\varepsilon \to z$, and convexity gives

$$\beta_{A_\star}(z_\varepsilon; H) \le (1-\varepsilon)\beta_{A_\star}(z; H) \le (1-\varepsilon)\lambda < \lambda.$$

This proves that every point of the closed sublevel set is a limit of strict sublevel points. The reverse inclusion follows from continuity. □

F.2 Joint epi-convergence of the rescaled residual

We next prove the function-level convergence. The result is stated sequentially because all spaces are finite dimensional and hence first countable.

Proposition F.4 (Joint strategic epi-limit)**.** *Let $t_k \downarrow 0$, $H_k \to H$, and $z_k \to z$ with $z_k, z \in \mathcal{Z}$. Then*

$$\liminf_{k\to\infty} \mathcal{F}_{t_k}(H_k, z_k) \ge \mathcal{F}(H, z). \tag{F.16}$$

Moreover, if $z \in \mathcal{E}_\star$, then for every pair of sequences $t_k \downarrow 0$ and $H_k \to H$, there exists $\widetilde{z}_k \in \mathcal{Z}$ with $\widetilde{z}_k \to z$ such that

$$\limsup_{k\to\infty} \mathcal{F}_{t_k}(H_k, \widetilde{z}_k) \le \mathcal{F}(H, z). \tag{F.17}$$

Consequently,

$$\mathcal{F}_t \xrightarrow{\text{epi}} \mathcal{F} \qquad \text{on } \mathbb{R}^{m\times n} \times \mathcal{Z} \text{ as } t \downarrow 0. \tag{F.18}$$

The recovery statement (F.17) *is stronger than ordinary joint epi-convergence because it holds for every prescribed sequence $H_k \to H$.*

Proof. If $z \in \mathcal{E}_\star$, the liminf inequality is exactly Proposition D.10, and the recovery statement is Proposition D.11, both with $A = A_\star$.

If $z \notin \mathcal{E}_\star$, then $g_{A_\star}(z) > 0$. Joint continuity of the saddle residual, or alternatively the perturbation bounds in Lemma C.7, gives

$$g_{A_\star + t_k H_k}(z_k) \longrightarrow g_{A_\star}(z) > 0.$$

Hence

$$\mathcal{F}_{t_k}(H_k, z_k) = \frac{g_{A_\star + t_k H_k}(z_k)}{t_k} \longrightarrow +\infty = \mathcal{F}(H, z).$$

The recovery requirement at a point where the limit function equals $+\infty$ is vacuous. This proves joint epi-convergence. □

The branch cost therefore has the following exact operational meaning.

Proposition F.5 (Minimal first-order resolution of a branch)**.** *For every $H \in \mathbb{R}^{m\times n}$ and every $z \in \mathcal{E}_\star$,*

$$\beta_{A_\star}(z; H) = \inf\left\{\lambda \ge 0 : z \in \liminf_{t\downarrow 0} \mathcal{E}_{t\lambda}(A_\star + tH)\right\}. \tag{F.19}$$

More precisely:

1. *if $\lambda > \beta_{A_\star}(z; H)$, then*
$$z \in \liminf_{t\downarrow 0} \mathcal{E}_{t\lambda}(A_\star + tH); \tag{F.20}$$
2. *if $\lambda < \beta_{A_\star}(z; H)$, then*
$$z \notin \limsup_{t\downarrow 0} \mathcal{E}_{t\lambda}(A_\star + tH). \tag{F.21}$$

No assertion is made at the boundary $\lambda = \beta_{A_\star}(z; H)$ without an additional recovery condition.

Proof. Suppose first that $\lambda > \beta_{A_\star}(z; H)$, and let $t_k \downarrow 0$ be arbitrary. Apply the recovery statement (F.17) with $H_k = H$. There exists $z_k \to z$ such that

$$\limsup_{k\to\infty} \frac{g_{A_\star + t_k H}(z_k)}{t_k} \le \beta_{A_\star}(z; H) < \lambda.$$

Thus $z_k \in \mathcal{E}_{t_k\lambda}(A_\star + t_k H)$ for all sufficiently large k. Since the sequence (t_k) was arbitrary, (F.20) follows.

Conversely, if $z \in \limsup_{t\downarrow 0} \mathcal{E}_{t\lambda}(A_\star + tH)$, then there exist $t_k \downarrow 0$ and $z_k \to z$ satisfying

$$\mathcal{F}_{t_k}(H, z_k) \le \lambda.$$

The liminf inequality (F.16) gives

$$\beta_{A_\star}(z; H) \le \lambda.$$

This proves (F.21) and hence the infimum identity. □

F.3 Positive-resolution set convergence

For $t > 0$, $H \in \mathbb{R}^{m\times n}$, and $\lambda \ge 0$, set

$$\mathcal{S}_t(H, \lambda) := \mathcal{E}_{t\lambda}(A_\star + tH) = \{z \in \mathcal{Z} : \mathcal{F}_t(H, z) \le \lambda\}. \tag{F.22}$$

Every $\mathcal{S}_t(H, \lambda)$ is nonempty, compact, and convex, because it contains the exact equilibrium correspondence of $A_\star + tH$.

Proposition F.6 (Parameterized positive-resolution convergence)**.** *Let*

$$t_k \downarrow 0, \qquad H_k \to H, \qquad \lambda_k \to \lambda > 0. \tag{F.23}$$

Then

$$d_{\mathrm{H}}\left(\mathcal{S}_{t_k}(H_k, \lambda_k), \mathcal{K}_\lambda(H)\right) \longrightarrow 0. \tag{F.24}$$

Equivalently,

$$\mathcal{E}_{t_k\lambda_k}(A_\star + t_k H_k) \underset{d_{\mathrm{H}}}{\longrightarrow} \mathcal{K}_\lambda(H). \tag{F.25}$$

Proof. Define functions on the compact space $\mathcal{Z}$ by

$$f_k(z) := \mathcal{F}_{t_k}(H_k, z) - \lambda_k, \qquad f(z) := \mathcal{F}(H, z) - \lambda. \tag{F.26}$$

Each f_k is continuous. The function f is lower semicontinuous because $\mathcal{E}_\star$ is closed, the branch cost is continuous on $\mathcal{E}_\star$, and $\mathcal{F}(H, z) = +\infty$ outside $\mathcal{E}_\star$. Proposition F.4, including its robust recovery statement for the prescribed sequence (H_k), implies

$$f_k \xrightarrow{\mathrm{epi}} f.$$

Their zero sublevel sets are

$$\{z : f_k(z) \le 0\} = \mathcal{S}_{t_k}(H_k, \lambda_k)$$

and

$$\{z : f(z) \le 0\} = \mathcal{K}_\lambda(H).$$

By (F.14),

$$\mathrm{cl}\{z : f(z) < 0\} = \{z : f(z) \le 0\}.$$

Proposition A.7 therefore gives Hausdorff convergence of the zero sublevel sets, which is (F.24). □

Corollary F.7 (Branchwise positive-resolution completeness)**.** *Fix H, $\lambda > 0$, and $z \in \mathcal{K}_\lambda(H)$. For every sequence $t_k \downarrow 0$, there exist*

$$z_k \in \mathcal{E}_{t_k\lambda}(A_\star + t_k H) \tag{F.27}$$

such that $z_k \to z$. Conversely, every cluster point of a sequence

$$z_k \in \mathcal{E}_{t_k\lambda}(A_\star + t_k H) \tag{F.28}$$

belongs to $\mathcal{K}_\lambda(H)$.

Proof. This is the inner- and outer-limit content of Proposition F.6 with $H_k = H$ and $\lambda_k = \lambda$. □

The visible correspondence is itself stable at every positive resolution.

Proposition F.8 (Hausdorff continuity of the visible correspondence)**.** *The map*

$$(H, \lambda) \longmapsto \mathcal{K}_\lambda(H) \tag{F.29}$$

from $\mathbb{R}^{m\times n} \times (0,\infty)$ *into* $\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$ *is continuous under the Hausdorff metric. Equivalently, if*

$$H_k \to H, \qquad \lambda_k \to \lambda > 0, \tag{F.30}$$

then

$$d_{\mathrm{H}}\left(\mathcal{K}_{\lambda_k}(H_k), \mathcal{K}_\lambda(H)\right) \longrightarrow 0. \tag{F.31}$$

Proof. On the compact set $\mathcal{E}_\star$, define

$$q_k(z) := \beta_{A_\star}(z; H_k) - \lambda_k, \qquad q(z) := \beta_{A_\star}(z; H) - \lambda.$$

The Lipschitz estimate (D.31) gives uniform convergence $q_k \to q$ on $\mathcal{E}_\star$. Uniform convergence implies epi-convergence. The zero sublevel set of q is $\mathcal{K}_\lambda(H)$, and its strict zero sublevel is dense in it by (F.14). Applying Proposition A.7 on the compact ambient space $\mathcal{E}_\star$ proves (F.31). □

We can now upgrade sequential convergence to a uniform theorem.

Proposition F.9 (Uniform strategic epi-resolution)**.** *Let* $\mathcal{H} \subseteq \mathbb{R}^{m\times n}$ *be compact, and let*

$$0 < \lambda_{\min} \le \lambda_{\max} < \infty. \tag{F.32}$$

Then

$$\lim_{t\downarrow 0} \sup_{\substack{H\in\mathcal{H} \\ \lambda\in[\lambda_{\min},\lambda_{\max}]}} d_{\mathrm{H}}\left(\mathcal{E}_{t\lambda}(A_\star + tH), \mathcal{K}_\lambda(H)\right) = 0. \tag{F.33}$$

In particular, for every fixed $\lambda > 0$*,*

$$\sup_{H\in\mathcal{H}} d_{\mathrm{H}}\left(\mathcal{E}_{t\lambda}(A_\star + tH), \mathcal{K}_\lambda(H)\right) \longrightarrow 0. \tag{F.34}$$

Proof. Suppose (F.33) fails. Then there exist $\varepsilon > 0$, a sequence $t_k \downarrow 0$, directions $H_k \in \mathcal{H}$, and resolutions $\lambda_k \in [\lambda_{\min}, \lambda_{\max}]$ such that

$$d_{\mathrm{H}}\left(\mathcal{E}_{t_k\lambda_k}(A_\star + t_k H_k), \mathcal{K}_{\lambda_k}(H_k)\right) \ge \varepsilon \tag{F.35}$$

for every k. By compactness, pass to a subsequence such that

$$H_k \to H \in \mathcal{H}, \qquad \lambda_k \to \lambda \in [\lambda_{\min}, \lambda_{\max}].$$

Proposition F.6 gives

$$d_{\mathrm{H}}\left(\mathcal{E}_{t_k\lambda_k}(A_\star + t_k H_k), \mathcal{K}_\lambda(H)\right) \longrightarrow 0, \tag{F.36}$$

while Proposition F.8 gives

$$d_{\mathrm{H}}\left(\mathcal{K}_{\lambda_k}(H_k), \mathcal{K}_\lambda(H)\right) \longrightarrow 0. \tag{F.37}$$

The triangle inequality contradicts (F.35). □

F.4 The exact-resolution boundary

At $\lambda = 0$, the strict-sublevel density property used above is unavailable because the branch cost is nonnegative. The joint epi-limit still yields a universal outer approximation, but exact branch completeness requires additional structure.

Proposition F.10 (Universal exact-resolution outer limit)**.** *Let* $t_k \downarrow 0$*,* $H_k \to H$*, and choose*

$$z_k \in \mathcal{E}_0(A_\star + t_k H_k). \tag{F.38}$$

Every cluster point z *of* (z_k) *belongs to* $\mathcal{K}_0(H)$*. Equivalently,*

$$\limsup_{k\to\infty} \mathcal{E}_0(A_\star + t_k H_k) \subseteq \mathcal{K}_0(H). \tag{F.39}$$

Proof. Along any convergent subsequence $z_k \to z$, exact optimality gives

$$\mathcal{F}_{t_k}(H_k, z_k) = 0.$$

The liminf inequality (F.16) implies

$$0 \geq \mathcal{F}(H, z) = \beta_{A_\star}(z; H).$$

The branch cost is nonnegative, so it equals zero and $z \in \mathcal{K}_0(H)$. □

Corollary F.11 (Exact convergence under a unique first-order branch)**.** *Suppose*

$$\mathcal{K}_0(H) = \{z_H\} \tag{F.40}$$

for some H. Then for every $t_k \downarrow 0$ and every $H_k \to H$,

$$d_{\mathrm{H}}\left(\mathcal{E}_0(A_\star + t_k H_k), \{z_H\}\right) \longrightarrow 0. \tag{F.41}$$

In particular,

$$\mathcal{E}_0(A_\star + tH) \underset{d_{\mathrm{H}}}{\longrightarrow} \{z_H\} \qquad \text{as } t \downarrow 0. \tag{F.42}$$

Proof. Every perturbed exact equilibrium set is nonempty and contained in the compact space $\mathcal{Z}$. If the directed excess from $\mathcal{E}_0(A_\star + t_k H_k)$ to $\{z_H\}$ did not converge to zero, one could select points z_k at a fixed positive distance from z_H, extract a convergent subsequence, and contradict Proposition F.10. Hence

$$e\left(\mathcal{E}_0(A_\star + t_k H_k), \{z_H\}\right) \longrightarrow 0.$$

Nonemptiness of the perturbed sets then also gives

$$e\left(\{z_H\}, \mathcal{E}_0(A_\star + t_k H_k)\right) \longrightarrow 0,$$

which proves Hausdorff convergence. □

The guiding bifurcation makes clear why positive resolution is qualitatively more stable than exact resolution.

Example F.1 (Exact-resolution discontinuity and positive-resolution repair)**.** Let

$$A_\star = A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad H_h := h e_2 e_2^\top, \qquad h \in \mathbb{R}. \tag{F.43}$$

Write

$$z_p := \big(x(p), e_2\big), \qquad p \in [0, 1].$$

Equation (E.38) gives

$$\beta_{A_0}(z_p; H_h) = h_+ p + h_-(1 - p). \tag{F.44}$$

Therefore,

$$\mathcal{K}_0(H_h) = \begin{cases} \{z_0\}, & h > 0, \\ \{z_p : p \in [0, 1]\}, & h = 0, \\ \{z_1\}, & h < 0. \end{cases} \tag{F.45}$$

The exact-resolution target is thus discontinuous at $h = 0$.

For every $\lambda > 0$, however,

$$\mathcal{K}_\lambda(H_h) = \begin{cases} \{z_p : 0 \leq p \leq \min\{1, \lambda/h\}\}, & h > 0, \\ \{z_p : p \in [0, 1]\}, & h = 0, \\ \{z_p : \max\{0, 1 - \lambda/|h|\} \leq p \leq 1\}, & h < 0. \end{cases} \tag{F.46}$$

This correspondence is Hausdorff continuous in h, and Proposition F.9 gives, for every compact interval $\mathcal{H}_0 \subseteq \mathbb{R}$ and every $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$,

$$\sup_{\substack{h \in \mathcal{H}_0 \\ \lambda \in [\lambda_{\min}, \lambda_{\max}]}} d_{\mathrm{H}}\left(\mathcal{E}_{t\lambda}(A_0 + tH_h), \mathcal{K}_\lambda(H_h)\right) \longrightarrow 0. \tag{F.47}$$

Remark F.12 (Why the principal theorem is stated at positive resolution). At $\lambda > 0$, convexity and the existence of a zero-cost branch imply the strict-sublevel density (F.14); this is exactly the condition that upgrades epi-convergence to complete Hausdorff convergence of sublevel sets. At $\lambda = 0$, the target map can be discontinuous, as (F.45) shows, and Proposition F.10 supplies only the unconditional outer inclusion. The later statistical minimax theorem therefore treats $\lambda > 0$ as the regular critical-resolution regime, while exact correspondence inference is handled separately through nonregular lower bounds.

Remark F.13 (Input to the local statistical theory). Proposition F.9 is the geometric input required for local asymptotic decision theory. Under local alternatives $A_{T,H} = A_\star + H/\sqrt{T}$ and tolerance $\tau_T = \lambda/\sqrt{T}$, set $t = T^{-1/2}$ to obtain

$$\sup_{H \in \mathcal{H}} d_{\mathrm{H}} \left(\mathcal{E}_{\lambda/\sqrt{T}} \left(A_\star + \frac{H}{\sqrt{T}} \right), \mathcal{K}_\lambda(H) \right) \longrightarrow 0 \tag{F.48}$$

for every compact $\mathcal{H}$ and every fixed $\lambda > 0$. The subsequent statistical appendices combine this uniform target convergence with convergence of the noisy payoff experiment to a Gaussian shift experiment.

G Gaussian Payoff Experiments

This appendix identifies the exact statistical experiment generated by Gaussian payoff queries. Under a deterministic allocation, entrywise sample means are not merely asymptotically sufficient: the complete payoff transcript is exactly equivalent, in Le Cam's sense, to a finite-dimensional Gaussian shift experiment. We then quantify the convergence of the finite-sample covariance to its limiting information geometry. Finally, we treat predictable adaptive allocations. In that setting the experiment is generally not exactly Gaussian, but it is locally asymptotically normal whenever the empirical query frequencies converge to a deterministic full-support design. Our conventions for deficiencies, comparison of experiments, and local asymptotic normality follow Le Cam & Yang (2000) and van der Vaart (1998); the martingale central limit argument follows Hall & Heyde (1980).

G.1 Statistical experiments and comparison

A statistical experiment indexed by a parameter set Θ is a triple

$$\mathfrak{E} = \big(\mathsf{X}, \mathcal{X}, \{P_\theta : \theta \in \Theta\}\big), \tag{G.1}$$

where $(\mathsf{X}, \mathcal{X})$ is a measurable sample space. A Markov kernel K from $(\mathsf{X}, \mathcal{X})$ to another measurable space $(\mathsf{Y}, \mathcal{Y})$ acts on a probability measure P by

$$(KP)(B) := \int_{\mathsf{X}} K(x, B)\, P(dx), \qquad B \in \mathcal{Y}. \tag{G.2}$$

For experiments

$$\mathfrak{E} = \big(\mathsf{X}, \mathcal{X}, \{P_\theta : \theta \in \Theta\}\big), \qquad \mathfrak{F} = \big(\mathsf{Y}, \mathcal{Y}, \{Q_\theta : \theta \in \Theta\}\big) \tag{G.3}$$

with a common parameter set, define the deficiency of $\mathfrak{E}$ with respect to $\mathfrak{F}$ by

$$\delta(\mathfrak{E}, \mathfrak{F}) := \inf_K \sup_{\theta \in \Theta} d_{\mathrm{TV}}\big(KP_\theta, Q_\theta\big), \tag{G.4}$$

where the infimum ranges over all Markov kernels from X to Y. The Le Cam distance is

$$\Delta(\mathfrak{E}, \mathfrak{F}) := \max\{\delta(\mathfrak{E}, \mathfrak{F}), \delta(\mathfrak{F}, \mathfrak{E})\}. \tag{G.5}$$

Experiments with zero Le Cam distance are statistically equivalent.

Let D be a measurable decision space, let $L : \Theta \times \mathsf{D} \to [0, M]$ be a bounded loss, and let a decision rule be a Markov kernel from the observation space to D. Define its minimax risk in experiment $\mathfrak{E}$ by

$$\mathcal{R}^\star(\mathfrak{E}, L) := \inf_\rho \sup_{\theta \in \Theta} \int_{\mathsf{X}} \int_{\mathsf{D}} L(\theta, a)\, \rho(x, da)\, P_\theta(dx). \tag{G.6}$$

Lemma G.1 (Bounded-risk transfer)**.** *For experiments with a common parameter set and every loss taking values in* $[0, M]$*,*

$$|\mathcal{R}^\star(\mathfrak{E}, L) - \mathcal{R}^\star(\mathfrak{F}, L)| \le M\Delta(\mathfrak{E}, \mathfrak{F}). \tag{G.7}$$

In particular, under the product ℓ_1 *norm used in this paper, Hausdorff loss between nonempty compact subsets of* $\mathcal{Z}$ *is bounded by* 4*, so its minimax risks differ by at most* $4\Delta(\mathfrak{E}, \mathfrak{F})$*.*

Proof. Fix $\varepsilon > \delta(\mathfrak{E}, \mathfrak{F})$, and choose a kernel K such that

$$\sup_{\theta\in\Theta} d_{\mathrm{TV}}(KP_\theta, Q_\theta) \le \varepsilon.$$

For any decision rule ρ in $\mathfrak{F}$, the composition ρK is a decision rule in $\mathfrak{E}$. The distribution of its decision under P_θ differs in total variation by at most ε from the distribution of the decision produced by ρ under Q_θ, by data processing for total variation. Since the loss is nonnegative and bounded by M,

$$\sup_{\theta\in\Theta} R_\theta(\rho K; \mathfrak{E}) \le \sup_{\theta\in\Theta} R_\theta(\rho; \mathfrak{F}) + M\varepsilon.$$

Taking the infimum over ρ and then $\varepsilon \downarrow \delta(\mathfrak{E}, \mathfrak{F})$ yields

$$\mathcal{R}^\star(\mathfrak{E}, L) \le \mathcal{R}^\star(\mathfrak{F}, L) + M\delta(\mathfrak{E}, \mathfrak{F}).$$

Interchanging the experiments proves (G.7). Finally, each simplex has ℓ_1-diameter at most 2, so $\mathcal{Z}$ has product-norm diameter at most 4; the same bound holds for Hausdorff distance between its nonempty compact subsets. □

G.2 Sufficiency and exact finite-sample laws

Set $d := mn$, fix a bijection

$$\iota : [m] \times [n] \longrightarrow [d], \tag{G.8}$$

and let $\mathrm{vec} : \mathbb{R}^{m\times n} \to \mathbb{R}^d$ be the corresponding vectorization. Write

$$a_\star := \mathrm{vec}(A_\star), \qquad h := \mathrm{vec}(H). \tag{G.9}$$

For $q \in \mathbb{R}^d_+$, define

$$D(q) := \mathrm{diag}(q_1, \dots, q_d), \qquad \mathsf{I}(q) := \sigma^{-2}D(q), \qquad \Sigma(q) := \mathsf{I}(q)^{-1} = \sigma^2 D(q)^{-1}, \tag{G.10}$$

whenever every coordinate of q is positive.

Fix deterministic counts and a canonical parameter-independent query order that realizes them:

$$N_{a,T} \in \mathbb{N}, \qquad \sum_{a=1}^{d} N_{a,T} = T, \qquad q_{a,T} := \frac{N_{a,T}}{T} > 0. \tag{G.11}$$

Under the local parameter $h \in \Theta \subseteq \mathbb{R}^d$, observe mutually independent variables

$$Y_{a,\ell} = a_{\star,a} + \frac{h_a}{\sqrt{T}} + \xi_{a,\ell}, \qquad \xi_{a,\ell} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \qquad \ell \in [N_{a,T}]. \tag{G.12}$$

Let $\mathfrak{E}^{\mathrm{det}}_T(q_T, \Theta)$ denote this full-data experiment. Because the coordinate labels and their order are deterministic and parameter independent, reconstructing the within-coordinate Gaussian blocks reconstructs the complete transcript experiment. Define the entrywise sample means and rescaled local statistic by

$$\overline{Y}_{a,T} := \frac{1}{N_{a,T}} \sum_{\ell=1}^{N_{a,T}} Y_{a,\ell}, \qquad W_{T,a} := \sqrt{T}\big(\overline{Y}_{a,T} - a_{\star,a}\big). \tag{G.13}$$

Let

$$\mathfrak{G}_T(q_T, \Theta) := \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \{\mathcal{N}\big(h, \Sigma(q_T)\big) : h \in \Theta\}\big) \tag{G.14}$$

be the associated finite-sample Gaussian shift experiment.

Proposition G.2 (Exact Gaussian reduction)**.** *For every T, every positive deterministic allocation q_T, and every parameter set $\Theta \subseteq \mathbb{R}^d$,*

$$W_T \sim \mathcal{N}\big(h, \Sigma(q_T)\big) \qquad \textit{under parameter } h, \tag{G.15}$$

and

$$\Delta\left(\mathfrak{E}_T^{\det}(q_T, \Theta), \mathfrak{G}_T(q_T, \Theta)\right) = 0. \tag{G.16}$$

The log-likelihood ratio in the Gaussian experiment relative to $h = 0$ is exactly

$$\log \frac{d\mathcal{N}(h, \Sigma(q_T))}{d\mathcal{N}(0, \Sigma(q_T))}(w) = h^\top \mathsf{I}(q_T) w - \frac{1}{2} h^\top \mathsf{I}(q_T) h. \tag{G.17}$$

Proof. By independence of the observations across entries,

$$W_{T,a} = h_a + \sqrt{T}\, \frac{1}{N_{a,T}} \sum_{\ell=1}^{N_{a,T}} \xi_{a,\ell}$$

are independent Gaussian variables with

$$\mathrm{Var}_h(W_{T,a}) = \frac{T\sigma^2}{N_{a,T}} = \frac{\sigma^2}{q_{a,T}}.$$

This proves (G.15).

The statistic W_T is a measurable function of the full data, so a deterministic Markov kernel maps $\mathfrak{E}_T^{\det}(q_T, \Theta)$ to $\mathfrak{G}_T(q_T, \Theta)$ exactly. We construct a parameter-free reverse kernel. Given $w \in \mathbb{R}^d$, set

$$\overline{y}_a(w) := a_{\star,a} + \frac{w_a}{\sqrt{T}}. \tag{G.18}$$

For each entry a, independently draw

$$Z_a \sim \mathcal{N}(0, \sigma^2 I_{N_{a,T}}),$$

let

$$P_a := I_{N_{a,T}} - \frac{1}{N_{a,T}} \mathbf{1}_{N_{a,T}} \mathbf{1}_{N_{a,T}}^\top,$$

and output

$$\widetilde{Y}_a := \overline{y}_a(w) \mathbf{1}_{N_{a,T}} + P_a Z_a. \tag{G.19}$$

Under $w \sim \mathcal{N}(h, \Sigma(q_T))$, the vector $\overline{y}_a(w)\mathbf{1}_{N_{a,T}}$ has mean $(a_{\star,a} + h_a/\sqrt{T})\mathbf{1}_{N_{a,T}}$ and covariance

$$\frac{\sigma^2}{N_{a,T}} \mathbf{1}_{N_{a,T}} \mathbf{1}_{N_{a,T}}^\top.$$

It is independent of $P_a Z_a$, whose covariance is $\sigma^2 P_a$. Their sum therefore has covariance

$$\frac{\sigma^2}{N_{a,T}} \mathbf{1}\mathbf{1}^\top + \sigma^2 \left(I - \frac{1}{N_{a,T}} \mathbf{1}\mathbf{1}^\top \right) = \sigma^2 I_{N_{a,T}}.$$

Hence $\widetilde{Y}_a$ has exactly the distribution in (G.12). This reconstruction kernel does not depend on h, proving the reverse deficiency is zero and hence (G.16).

Finally, the likelihood-ratio formula (G.17) is the standard common-covariance Gaussian identity, using

$$\Sigma(q_T)^{-1} = \mathsf{I}(q_T).$$

□

Remark G.3 (Exact sufficiency, not merely asymptotic sufficiency). The residual vectors

$$\big(Y_{a,\ell} - \overline{Y}_{a,T}\big)_{\ell=1}^{N_{a,T}}$$

are ancillary and independent of the entrywise sample means. Proposition G.2 therefore identifies the complete finite-sample decision problem with a d-dimensional Gaussian shift. No central limit theorem is used in the deterministic-design reduction.

G.3 Convergence of Gaussian experiments

Fix a full-support design

$$q \in \operatorname{ri}\Delta_d, \qquad q_{\min} := \min_{a\in[d]} q_a > 0, \tag{G.20}$$

and define the limiting Gaussian shift

$$\mathfrak{G}(q,\Theta) := \left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \left\{\mathcal{N}\big(h, \Sigma(q)\big) : h \in \Theta\right\}\right). \tag{G.21}$$

Let

$$\phi(r) := r - 1 - \log r, \qquad r > 0. \tag{G.22}$$

Proposition G.4 (Quantitative convergence to the limiting shift)**.** *Suppose $q_T \in \operatorname{ri}\Delta_d$. Then, for every parameter set $\Theta \subseteq \mathbb{R}^d$,*

$$\begin{aligned}\Delta\left(\mathfrak{E}_T^{\det}(q_T,\Theta), \mathfrak{G}(q,\Theta)\right) &\le d_{\mathrm{TV}}\left(\mathcal{N}(0,\Sigma(q_T)), \mathcal{N}(0,\Sigma(q))\right)\\ &\le \frac{1}{2}\left[\sum_{a=1}^{d}\phi\left(\frac{q_a}{q_{a,T}}\right)\right]^{1/2}.\end{aligned} \tag{G.23}$$

Consequently, if $q_T \to q$, then

$$\Delta\left(\mathfrak{E}_T^{\det}(q_T,\Theta), \mathfrak{G}(q,\Theta)\right) \longrightarrow 0, \tag{G.24}$$

uniformly over arbitrary, possibly unbounded, parameter sets Θ.

More quantitatively, if

$$\eta_T := \max_{a\in[d]}\left|\frac{q_{a,T}}{q_a} - 1\right| \le \frac{1}{2}, \tag{G.25}$$

then

$$\Delta\left(\mathfrak{E}_T^{\det}(q_T,\Theta), \mathfrak{G}(q,\Theta)\right) \le \frac{3}{2\sqrt{2}}\sqrt{d}\,\eta_T. \tag{G.26}$$

Proof. By Proposition G.2, composing the exact reduction and reconstruction kernels with any comparison kernel for the Gaussian shifts preserves the same total-variation error. It is therefore enough to compare the two Gaussian shift experiments. The identity kernel on $\mathbb{R}^d$ gives

$$\begin{aligned}\Delta\left(\mathfrak{G}_T(q_T,\Theta), \mathfrak{G}(q,\Theta)\right) &\le \sup_{h\in\Theta} d_{\mathrm{TV}}\left(\mathcal{N}(h,\Sigma(q_T)), \mathcal{N}(h,\Sigma(q))\right)\\ &= d_{\mathrm{TV}}\left(\mathcal{N}(0,\Sigma(q_T)), \mathcal{N}(0,\Sigma(q))\right),\end{aligned}$$

where the equality follows from translation invariance of total variation.

The Gaussian relative entropy formula gives

$$\begin{aligned}&D_{\mathrm{KL}}\left(\mathcal{N}(0,\Sigma(q_T)) \,\|\, \mathcal{N}(0,\Sigma(q))\right)\\ &\quad= \frac{1}{2}\left[\operatorname{tr}\left(\Sigma(q)^{-1}\Sigma(q_T)\right) - d - \log\det\left(\Sigma(q)^{-1}\Sigma(q_T)\right)\right]\\ &\quad= \frac{1}{2}\sum_{a=1}^{d}\phi\left(\frac{q_a}{q_{a,T}}\right).\end{aligned} \tag{G.27}$$

Pinsker's inequality yields the second line of (G.23). Coordinatewise convergence of q_T to q and continuity of ϕ at 1 prove (G.24).

For the explicit bound, write

$$q_{a,T} = q_a(1+\varepsilon_{a,T}), \qquad |\varepsilon_{a,T}| \le \eta_T.$$

Then

$$r_{a,T} := \frac{q_a}{q_{a,T}} = \frac{1}{1+\varepsilon_{a,T}} \in \left[\frac{2}{3}, 2\right], \qquad |r_{a,T} - 1| \le 2\eta_T.$$

Since $\phi(1) = \phi'(1) = 0$ and $\phi''(r) = r^{-2} \le 9/4$ on $[2/3, 2]$, Taylor's theorem gives

$$\phi(r_{a,T}) \le \frac{9}{8}(r_{a,T} - 1)^2 \le \frac{9}{2}\eta_T^2.$$

Substituting this into (G.23) proves (G.26). □

Corollary G.5 (Rounded deterministic designs)**.** *For every $q \in \operatorname{ri}\Delta_d$ and every sufficiently large T, there exist positive integer counts $(N_{a,T})_{a=1}^d$ summing to T such that*

$$\max_{a\in[d]}\left|\frac{N_{a,T}}{T} - q_a\right| < \frac{1}{T}. \tag{G.28}$$

For such a sequence, whenever $Tq_{\min} \geq 2$,

$$\Delta\left(\mathfrak{E}_T^{\det}(q_T,\Theta), \mathfrak{G}(q,\Theta)\right) \leq \frac{3\sqrt{d}}{2\sqrt{2}\,Tq_{\min}}. \tag{G.29}$$

Proof. Set $N_{a,T}^{(0)} = \lfloor Tq_a \rfloor$, let $R_T = T - \sum_a N_{a,T}^{(0)}$, and add one observation to the R_T coordinates with largest fractional parts $Tq_a - \lfloor Tq_a \rfloor$. The resulting counts sum to T and differ from Tq_a by less than one, so $\|q_T - q\|_\infty < T^{-1}$. They are positive when $Tq_{\min} \geq 2$. Thus

$$\eta_T \leq \frac{1}{Tq_{\min}} \leq \frac{1}{2},$$

and (G.29) follows from (G.26). □

Corollary G.6 (Decision-theoretic transfer for deterministic designs)**.** *Let L_T be any possibly T-dependent loss taking values in $[0, M]$ on a common parameter and decision space. Under the assumptions of Proposition G.4,*

$$\left|\mathcal{R}^\star\left(\mathfrak{E}_T^{\det}(q_T,\Theta), L_T\right) - \mathcal{R}^\star\left(\mathfrak{G}(q,\Theta), L_T\right)\right| \leq M\varepsilon_T, \tag{G.30}$$

where ε_T may be taken to be either upper bound in (G.23)*. For Hausdorff loss on $\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$, one may take $M = 4$.*

Proof. Combine Lemma G.1 with Proposition G.4. □

Remark G.7 (Why full support is imposed)*.* If a limiting allocation assigns zero mass to an entry, the corresponding local payoff direction is not estimable at the $T^{-1/2}$ scale and $\Sigma(q)$ is singular. Such designs can be treated on the identifiable subspace, but the local experiment is then degenerate. The core minimax theory uses $q \in \operatorname{ri}\Delta_d$ so that the Gaussian shift is nondegenerate on the entire payoff space.

G.4 Predictable adaptive allocations

We next record the precise sense in which a regular adaptive design has the same Gaussian local limit. This extension is not needed for the exact deterministic-design comparison above, but it clarifies which adaptive sampling rules preserve the limiting information geometry.

For each T, let A_T be a nonanticipating query rule. We may, without loss of generality, enlarge the initial sigma-field to contain all internal random seeds of the rule; the queried coordinate $I_{T,t} \in [d]$ is then predictable with respect to a filtration $(\mathcal{F}_{T,t})_{t=0}^T$. Under local parameter $h \in \mathbb{R}^d$, observe

$$Y_{T,t} = a_{\star,I_{T,t}} + \frac{h_{I_{T,t}}}{\sqrt{T}} + \xi_{T,t}, \qquad \xi_{T,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2), \tag{G.31}$$

where $\xi_{T,t}$ is independent of $\mathcal{F}_{T,t-1}$, and $\mathcal{F}_{T,t}$ contains the observation at round t. Let $P_{T,h}^{\mathsf{A}}$ denote the law of the complete transcript. Define

$$\widehat{q}_{T,a} := \frac{1}{T}\sum_{t=1}^T \mathbf{1}\{I_{T,t} = a\}, \qquad \mathsf{I}_T := \mathsf{I}(\widehat{q}_T) = \frac{1}{\sigma^2 T}\sum_{t=1}^T e_{I_{T,t}} e_{I_{T,t}}^\top, \tag{G.32}$$

$$\Delta_T := \frac{1}{\sigma^2\sqrt{T}}\sum_{t=1}^T e_{I_{T,t}}\left(Y_{T,t} - a_{\star,I_{T,t}}\right). \tag{G.33}$$

Proposition G.8 (Exact adaptive likelihood ratio)**.** *For every $h \in \mathbb{R}^d$,*

$$\log\frac{dP_{T,h}^{\mathsf{A}}}{dP_{T,0}^{\mathsf{A}}} = h^\top\Delta_T - \frac{1}{2}h^\top\mathsf{I}_T h, \qquad P_{T,0}^{\mathsf{A}}\text{-almost surely}. \tag{G.34}$$

Proof. Conditionally on the past and the predictable query $I_{T,t} = a$, the two models differ only through the Gaussian mean shift $h_a/\sqrt{T}$. The one-step log-density ratio is

$$\frac{h_a}{\sigma^2\sqrt{T}}\big(Y_{T,t} - a_{\star,a}\big) - \frac{h_a^2}{2\sigma^2 T}.$$

The conditional query rule is identical under all parameters and therefore cancels from the transcript likelihood ratio. Summing the conditional log-density ratios proves (G.34). □

Assumption G.9 (Stable full-support allocation)**.** There exists a deterministic $q \in \mathrm{ri}\,\Delta_d$ such that, under the base law,

$$\widehat{q}_T \xrightarrow{\mathbb{P}} q \qquad \text{under } P^{\mathsf{A}}_{T,0}. \tag{G.35}$$

Proposition G.10 (Martingale central sequence)**.** *Under Assumption G.9,*

$$\Delta_T \xrightarrow{\mathrm{d}} \mathcal{N}\big(0, \mathsf{I}(q)\big) \qquad \textit{under } P^{\mathsf{A}}_{T,0}. \tag{G.36}$$

Proof. Under the base law,

$$Y_{T,t} - a_{\star,I_{T,t}} = \xi_{T,t},$$

so Δ_T is a vector martingale sum. Fix $c \in \mathbb{R}^d$. The scalar increments

$$X_{T,t}(c) := \frac{c_{I_{T,t}}\xi_{T,t}}{\sigma^2\sqrt{T}} \tag{G.37}$$

form a martingale-difference array, and their predictable quadratic variation is

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{E}^{\mathsf{A}}_0\left[X_{T,t}(c)^2 \,\middle|\, \mathcal{F}_{T,t-1}\right] &= \frac{1}{\sigma^2 T}\sum_{t=1}^{T} c^2_{I_{T,t}} \\ &= c^\top \mathsf{I}_T c \xrightarrow{\mathbb{P}} c^\top \mathsf{I}(q) c. \end{aligned} \tag{G.38}$$

For every $\varepsilon > 0$, if $c \neq 0$,

$$\begin{aligned}&\sum_{t=1}^{T} \mathbb{E}^{\mathsf{A}}_0\left[X_{T,t}(c)^2 \mathbf{1}\{|X_{T,t}(c)| > \varepsilon\} \,\middle|\, \mathcal{F}_{T,t-1}\right] \\ &\qquad \le \frac{\|c\|_\infty^2}{\sigma^4}\mathbb{E}\left[\xi^2 \mathbf{1}\left\{|\xi| > \frac{\varepsilon\sigma^2\sqrt{T}}{\|c\|_\infty}\right\}\right] \longrightarrow 0, \end{aligned} \tag{G.39}$$

where $\xi \sim \mathcal{N}(0, \sigma^2)$. The scalar martingale central limit theorem therefore yields

$$c^\top \Delta_T \xrightarrow{\mathrm{d}} \mathcal{N}\big(0, c^\top \mathsf{I}(q) c\big).$$

The vector conclusion follows from the Cramér–Wold device. □

Proposition G.11 (Adaptive local asymptotic normality)**.** *Under Assumption G.9, the adaptive experiment is locally asymptotically normal at $A_\star$ with information matrix $\mathsf{I}(q)$. More precisely, for every compact $K \subseteq \mathbb{R}^d$,*

$$\sup_{h\in K}\left|\log\frac{dP^{\mathsf{A}}_{T,h}}{dP^{\mathsf{A}}_{T,0}} - h^\top\Delta_T + \frac{1}{2}h^\top\mathsf{I}(q)h\right| \xrightarrow{\mathbb{P}} 0 \qquad \textit{under } P^{\mathsf{A}}_{T,0}, \tag{G.40}$$

together with (G.36).

For every fixed $h \in \mathbb{R}^d$, the sequences $P^{\mathsf{A}}_{T,h}$ and $P^{\mathsf{A}}_{T,0}$ are mutually contiguous, and

$$\Delta_T \xrightarrow{\mathrm{d}} \mathcal{N}\big(\mathsf{I}(q)h, \mathsf{I}(q)\big) \qquad \textit{under } P^{\mathsf{A}}_{T,h}. \tag{G.41}$$

Consequently,

$$\mathsf{I}(q)^{-1}\Delta_T \xrightarrow{\mathrm{d}} \mathcal{N}\big(h, \Sigma(q)\big) \qquad \textit{under } P^{\mathsf{A}}_{T,h}. \tag{G.42}$$

Moreover, $\widehat{q}_T \xrightarrow{\mathbb{P}} q$ also holds under every fixed local alternative $P^{\mathsf{A}}_{T,h}$.

Proof. The exact likelihood-ratio identity gives

$$\sup_{h\in K}\left|\log\frac{dP^{\mathsf{A}}_{T,h}}{dP^{\mathsf{A}}_{T,0}}-h^\top\Delta_T+\frac{1}{2}h^\top \mathsf{I}(q)h\right| \le \frac{1}{2}\sup_{h\in K}\|h\|_2^2\|\mathsf{I}_T-\mathsf{I}(q)\|_{\mathrm{op}},$$

which converges to zero in probability by (G.35). This proves the uniform LAN expansion.

By Proposition G.10, the log-likelihood ratio for fixed h converges under the base law to

$$h^\top Z-\frac{1}{2}h^\top \mathsf{I}(q)h, \qquad Z\sim\mathcal{N}(0,\mathsf{I}(q)).$$

Its exponential has expectation one. Le Cam's first lemma gives mutual contiguity, and Le Cam's third lemma exponentially tilts the Gaussian limit, yielding (G.41). Equation (G.42) follows by the continuous mapping theorem. Finally, contiguity transfers convergence in probability of $\widehat{q}_T$ from the base law to each fixed local alternative. □

Corollary G.12 (Finite adaptive subexperiments)**.** *Let $\Theta_0\subset\mathbb{R}^d$ be finite and contain 0. Under Assumption G.9, the vector of likelihood ratios*

$$\left(\frac{dP^{\mathsf{A}}_{T,h}}{dP^{\mathsf{A}}_{T,0}}\right)_{h\in\Theta_0} \tag{G.43}$$

converges in distribution under $P^{\mathsf{A}}_{T,0}$ to the likelihood ratio vector of the Gaussian shift experiment $\mathfrak{G}(q,\Theta_0)$ under its parameter-zero law. Consequently, the finite adaptive subexperiments converge weakly in Le Cam's sense to $\mathfrak{G}(q,\Theta_0)$.

Proof. For every $h\in\Theta_0$, the exact likelihood-ratio representation and Assumption G.9 give

$$\frac{dP^{\mathsf{A}}_{T,h}}{dP^{\mathsf{A}}_{T,0}}=\exp\left(h^\top\Delta_T-\frac{1}{2}h^\top\mathsf{I}_T h\right).$$

Joint convergence of this finite vector follows from Proposition G.10 and the continuous mapping theorem. Under the parameter-zero law of the Gaussian shift, if $W\sim\mathcal{N}(0,\Sigma(q))$, its likelihood ratio at parameter h is

$$\exp\left(h^\top\mathsf{I}(q)W-\frac{1}{2}h^\top\mathsf{I}(q)h\right),$$

and $\mathsf{I}(q)W\sim\mathcal{N}(0,\mathsf{I}(q))$. The limit vectors are therefore identical. The likelihood-ratio characterization of weak convergence of finite experiments (Le Cam & Yang, 2000) gives the conclusion. □

Remark G.13 (Deterministic versus adaptive comparison). For deterministic allocations, Proposition G.2 gives exact finite-sample equivalence, and Proposition G.4 gives an explicit uniform Le Cam-distance bound over arbitrary parameter sets. Under adaptive allocations, Proposition G.11 instead gives a local likelihood expansion and Gaussian limits for fixed or finite local parameter sets. We do not infer a uniform deficiency bound over an uncountable parameter class from LAN alone. The exact local minimax equivalence in the main theorem is therefore proved first under deterministic asymptotic designs; adaptive design is invoked only where the corresponding additional uniformity is established.

Remark G.14 (Random information limits). If $\widehat{q}_T$ converges to a nondegenerate random limit rather than a deterministic q, Assumption G.9 fails. Under the additional stable martingale-convergence hypotheses used in locally asymptotically mixed normal theory, the corresponding likelihood limit is mixed Gaussian with random information. Without such joint convergence, no universal local limit follows. A deterministic full-support information limit is therefore a substantive regularity condition, not a notational convenience.

H LOCAL MINIMAX LOWER BOUND

This appendix derives the minimax lower theory associated with the geometric limit in Appendix F. The argument has three layers. First, under a deterministic asymptotic allocation, the original payoff experiment is compared quantitatively with its limiting Gaussian shift by the Le Cam distance. Second, the finite-sample equilibrium correspondence is replaced by the visible first-order correspondence using the uniform Hausdorff limit from Theorem 2. Third, metric testing reductions give explicit two-point and packing lower bounds for the resulting Gaussian random-set problem. We finally show that the same Gaussian minimax risk is a lower bound for every predictable adaptive design whose empirical allocation converges to a deterministic full-support limit. Our decision- theoretic conventions follow Le Cam & Yang (2000) and van der Vaart (1998); the testing reductions are in the standard form developed, for example, in Tsybakov (2009).

H.1 LOCAL SET-VALUED DECISION PROBLEMS

Fix a nonempty compact local direction class

$$\mathcal{H} \subseteq \mathbb{R}^{m\times n}, \qquad \Theta := \mathrm{vec}(\mathcal{H}) \subseteq \mathbb{R}^d, \qquad d := mn, \tag{H.1}$$

and fix a positive resolution level $\lambda > 0$. For $H \in \mathcal{H}$, define the finite-sample and limiting targets

$$\mathcal{T}_{T,\lambda}(H) := \mathcal{E}_{\lambda/\sqrt{T}}\left(A_\star + \frac{H}{\sqrt{T}}\right), \tag{H.2}$$

$$\mathcal{T}_\lambda(H) := \mathcal{K}_\lambda(H). \tag{H.3}$$

Their uniform geometric discrepancy is

$$\omega_T(\mathcal{H},\lambda) := \sup_{H\in\mathcal{H}} d_{\mathrm{H}}\left(\mathcal{T}_{T,\lambda}(H), \mathcal{T}_\lambda(H)\right). \tag{H.4}$$

The uniform strategic epi-resolution theorem gives

$$\omega_T(\mathcal{H},\lambda) \longrightarrow 0. \tag{H.5}$$

Let

$$\mathsf{D} := \mathfrak{K}(\mathcal{Z}) \tag{H.6}$$

be the compact metric space of nonempty compact subsets of $\mathcal{Z}$, equipped with d_{H}. We permit randomized set-valued estimators, represented by Markov kernels into D. This convention can only strengthen a lower bound and makes the decision problem closed under randomization.

Let $q_T \in \mathrm{ri}\,\Delta_d$ be a deterministic allocation and let $P_{T,H}^{q_T}$ denote the full-data law in $\mathfrak{E}_T^{\mathrm{det}}(q_T,\Theta)$ at parameter $h = \mathrm{vec}(H)$. Define the finite local minimax risk

$$\mathfrak{R}_{T,q_T}(\mathcal{H},\lambda) := \inf_{\rho_T}\sup_{H\in\mathcal{H}} \int\int d_{\mathrm{H}}\left(C, \mathcal{T}_{T,\lambda}(H)\right)\rho_T(x,dC)P_{T,H}^{q_T}(dx), \tag{H.7}$$

where the infimum ranges over all Markov kernels from the payoff sample space to D.

For a full-support limit design $q \in \mathrm{ri}\,\Delta_d$, let

$$Q_H^q := \mathcal{N}\left(\mathrm{vec}(H), \Sigma(q)\right). \tag{H.8}$$

The Gaussian random-set minimax risk is

$$\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H},\lambda) := \inf_{\delta}\sup_{H\in\mathcal{H}} \int\int d_{\mathrm{H}}\left(C, \mathcal{T}_\lambda(H)\right)\delta(w,dC)Q_H^q(dw), \tag{H.9}$$

where δ ranges over all Markov kernels from $\mathbb{R}^d$ to D. Since $\mathcal{Z}$ has product-ℓ_1 diameter at most four, both losses in (H.7)– (H.9) take values in $[0,4]$.

Lemma H.1 (Perturbation of a Hausdorff target)**.** *For every metric space (X,d), every $A,B,C \in \mathfrak{K}(\mathsf{X})$,*

$$|d_{\mathrm{H}}(C,A) - d_{\mathrm{H}}(C,B)| \le d_{\mathrm{H}}(A,B). \tag{H.10}$$

Consequently, if two compact-valued target maps $S, S' : \Theta \to \mathfrak{K}(\mathsf{X})$ satisfy

$$\sup_{\theta\in\Theta} d_{\mathrm{H}}\big(S(\theta), S'(\theta)\big) \le \varepsilon, \tag{H.11}$$

then their minimax risks in any common statistical experiment differ by at most ε.

Proof. The triangle inequality for the Hausdorff metric gives

$$d_{\mathrm{H}}(C, A) \le d_{\mathrm{H}}(C, B) + d_{\mathrm{H}}(B, A).$$

Interchanging A and B proves (H.10). Integrating this inequality under an arbitrary decision rule, taking the supremum over the parameter, and then taking the infimum over decision rules proves the minimax-risk claim. □

H.2 Lower transfer to the Gaussian shift

Define the experiment discrepancy

$$\varepsilon_T(q_T, q; \Theta) := \Delta\left(\mathfrak{E}_T^{\det}(q_T, \Theta), \mathfrak{G}(q, \Theta)\right). \tag{H.12}$$

Appendix G proves the explicit bound

$$\varepsilon_T(q_T, q; \Theta) \le \frac{1}{2}\left[\sum_{a=1}^{d} \phi\left(\frac{q_a}{q_{a,T}}\right)\right]^{1/2}, \qquad \phi(r) := r - 1 - \log r. \tag{H.13}$$

Proposition H.2 (Finite-sample deterministic-design lower transfer)**.** *Let $\mathcal{H} \subseteq \mathbb{R}^{m\times n}$ be compact, let $\lambda > 0$, and let $q_T, q \in \operatorname{ri}\Delta_d$. Then*

$$\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda) \ge \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) - \omega_T(\mathcal{H}, \lambda) - 4\varepsilon_T(q_T, q; \Theta). \tag{H.14}$$

In particular, if $q_T \to q$, then

$$\liminf_{T\to\infty} \mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda) \ge \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda). \tag{H.15}$$

For the rounded designs of Corollary G.5, whenever $Tq_{\min} \ge 2$,

$$\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda) \ge \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) - \omega_T(\mathcal{H}, \lambda) - \frac{3\sqrt{2d}}{Tq_{\min}}. \tag{H.16}$$

Proof. Let $L_T(H, C) := d_{\mathrm{H}}(C, \mathcal{T}_{T,\lambda}(H))$ and $L(H, C) := d_{\mathrm{H}}(C, \mathcal{T}_\lambda(H))$. By Lemma H.1,

$$\left|\mathcal{R}^\star\left(\mathfrak{E}_T^{\det}(q_T, \Theta), L_T\right) - \mathcal{R}^\star\left(\mathfrak{E}_T^{\det}(q_T, \Theta), L\right)\right| \le \omega_T(\mathcal{H}, \lambda). \tag{H.17}$$

The bounded-risk transfer lemma in Appendix G, applied to $L \in [0, 4]$, gives

$$\mathcal{R}^\star\left(\mathfrak{E}_T^{\det}(q_T, \Theta), L\right) \ge \mathcal{R}^\star\left(\mathfrak{G}(q, \Theta), L\right) - 4\varepsilon_T(q_T, q; \Theta). \tag{H.18}$$

The three risks in (H.17)– (H.18) are exactly the risks in (H.7) and (H.9), proving (H.14). Equations (H.5) and (G.24) prove the asymptotic statement. Finally, multiplying (G.29) by four gives the last term in (H.16). □

Remark H.3 (What is, and is not, used in the lower transfer). Proposition H.2 uses neither an asymptotic normal approximation nor a point estimator of the payoff matrix. The deterministic payoff experiment is exactly equivalent to a Gaussian shift at every T, and only the covariance discretization and the set-valued target incur approximation error. The lower bound therefore applies to arbitrary randomized correspondence estimators.

H.3 Metric testing reductions in the Gaussian shift

For $H_0, H_1 \in \mathbb{R}^{m\times n}$, define their information distance under allocation q by

$$\mathfrak{d}_q(H_0, H_1) := \|\operatorname{vec}(H_1 - H_0)\|_{\mathsf{I}(q)} := \left[\operatorname{vec}(H_1 - H_0)^\top \mathsf{I}(q) \operatorname{vec}(H_1 - H_0)\right]^{1/2}. \tag{H.19}$$

Lemma H.4 (Exact total variation between common-covariance Gaussians)**.** *Let $P_0 = \mathcal{N}(\mu_0, \Sigma)$ and $P_1 = \mathcal{N}(\mu_1, \Sigma)$, where Σ is positive definite, and set*

$$\delta := \|\mu_1 - \mu_0\|_{\Sigma^{-1}}. \tag{H.20}$$

Then

$$d_{\mathrm{TV}}(P_0, P_1) = 2\Phi\left(\frac{\delta}{2}\right) - 1, \qquad 1 - d_{\mathrm{TV}}(P_0, P_1) = 2\Phi\left(-\frac{\delta}{2}\right), \tag{H.21}$$

where Φ is the standard normal distribution function.

Proof. The log-likelihood ratio $\ell = \log(dP_1/dP_0)$ satisfies

$$\ell \sim \mathcal{N}\left(-\frac{\delta^2}{2}, \delta^2\right) \quad \text{under } P_0, \qquad \ell \sim \mathcal{N}\left(\frac{\delta^2}{2}, \delta^2\right) \quad \text{under } P_1.$$

The Neyman–Pearson set attaining total variation is $B = \{\ell \geq 0\}$. Hence

$$P_1(B) - P_0(B) = \Phi\left(\frac{\delta}{2}\right) - \Phi\left(-\frac{\delta}{2}\right) = 2\Phi\left(\frac{\delta}{2}\right) - 1.$$

□

Proposition H.5 (Pairwise Gaussian random-set lower bound)**.** *For every $H_0, H_1 \in \mathcal{H}$,*

$$\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) \geq d_{\mathrm{H}}\left(\mathcal{K}_\lambda(H_0), \mathcal{K}_\lambda(H_1)\right) \Phi\left(-\frac{\mathfrak{d}_q(H_0, H_1)}{2}\right). \tag{H.22}$$

Moreover, for every deterministic finite-sample allocation q_T,

$$\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda) \geq d_{\mathrm{H}}\left(\mathcal{T}_{T,\lambda}(H_0), \mathcal{T}_{T,\lambda}(H_1)\right) \Phi\left(-\frac{\mathfrak{d}_{q_T}(H_0, H_1)}{2}\right). \tag{H.23}$$

Proof. Restrict the parameter class to $\{H_0, H_1\}$. Apply the metric Le Cam inequality, Lemma B.9, with action metric d_{H} and target values $\mathcal{K}_\lambda(H_0)$ and $\mathcal{K}_\lambda(H_1)$. Lemma H.4 gives

$$\frac{1}{2}\left(1 - d_{\mathrm{TV}}(Q_{H_0}^q, Q_{H_1}^q)\right) = \Phi\left(-\frac{\mathfrak{d}_q(H_0, H_1)}{2}\right),$$

which proves (H.22). For the finite-sample statement, Proposition G.2 identifies the full payoff experiment exactly with the Gaussian shift having covariance $\Sigma(q_T)$; repeat the same argument with the finite targets. □

Proposition H.6 (Finite packing lower bound)**.** *Let $M \geq 3$, let $H_1, \ldots, H_M \in \mathcal{H}$, and suppose that, for some $r > 0$,*

$$d_{\mathrm{H}}\left(\mathcal{K}_\lambda(H_j), \mathcal{K}_\lambda(H_k)\right) \geq 2r \qquad \textit{whenever } j \neq k. \tag{H.24}$$

Then, for every reference vector $h_\circ \in \mathbb{R}^d$,

$$\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) \geq r\left[1 - \frac{\frac{1}{2M}\sum_{j=1}^M \|\mathrm{vec}(H_j) - h_\circ\|_{\mathrm{I}(q)}^2 + \log 2}{\log M}\right]_+. \tag{H.25}$$

Proof. Let J be uniform on $[M]$, and conditionally on $J = j$, draw $W \sim Q_{H_j}^q$. Given a set-valued estimate $\widehat{C}$, define $\widehat{J}$ as a nearest target index, with deterministic tie breaking:

$$\widehat{J} \in \operatorname*{arg\,min}_{k \in [M]} d_{\mathrm{H}}\left(\widehat{C}, \mathcal{K}_\lambda(H_k)\right).$$

If $d_{\mathrm{H}}(\widehat{C}, \mathcal{K}_\lambda(H_J)) < r$, then the triangle inequality and (H.24) imply $\widehat{J} = J$. Hence

$$\frac{1}{M}\sum_{j=1}^M \mathbb{E}_{H_j} d_{\mathrm{H}}\left(\widehat{C}, \mathcal{K}_\lambda(H_j)\right) \geq r\mathbb{P}(\widehat{J} \neq J).$$

Fano's inequality gives

$$\mathbb{P}(\widehat{J} \neq J) \geq 1 - \frac{I(J; W) + \log 2}{\log M}.$$

If $Q_\circ := \mathcal{N}(h_\circ, \Sigma(q))$, the information-radius inequality yields

$$I(J; W) \leq \frac{1}{M}\sum_{j=1}^M D_{\mathrm{KL}}(Q_{H_j}^q \| Q_\circ) = \frac{1}{2M}\sum_{j=1}^M \|\mathrm{vec}(H_j) - h_\circ\|_{\mathrm{I}(q)}^2.$$

Taking the infimum over decision rules proves (H.25). □

H.4 An explicit bifurcation lower bound

Return to the guiding 2×2 base game

$$A_\star = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad H_a^+ := a e_2 e_2^\top, \qquad H_a^- := -a e_2 e_2^\top. \tag{H.26}$$

Write $q_{22} := q_{\iota(2,2)}$.

Proposition H.7 (Positive-resolution bifurcation lower bound)**.** *Let $a > \lambda > 0$, and suppose that $H_a^+, H_a^- \in \mathcal{H}$. Then*

$$\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) \geq 2 \left(1 - \frac{\lambda}{a}\right) \Phi\left(-\frac{a\sqrt{q_{22}}}{\sigma}\right). \tag{H.27}$$

Consequently, if $H_{2\lambda}^+, H_{2\lambda}^- \in \mathcal{H}$,

$$\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) \geq \Phi\left(-\frac{2\lambda\sqrt{q_{22}}}{\sigma}\right) > 0. \tag{H.28}$$

Every deterministic design sequence $q_T \to q$ therefore satisfies the same bounds with $\mathfrak{R}_q^{\mathrm{G}}$ replaced by the corresponding $\liminf$ *finite-sample risk.*

Proof. Example 3.1, proved in Appendix F, gives

$$\mathcal{K}_\lambda(H_a^+) = \{z_p : 0 \leq p \leq \lambda/a\},$$
$$\mathcal{K}_\lambda(H_a^-) = \{z_p : 1 - \lambda/a \leq p \leq 1\}.$$

Since $\|z_p - z_{p'}\|_{\mathsf{V}} = 2|p - p'|$,

$$d_{\mathrm{H}}\left(\mathcal{K}_\lambda(H_a^+), \mathcal{K}_\lambda(H_a^-)\right) = 2\left(1 - \frac{\lambda}{a}\right). \tag{H.29}$$

Moreover,

$$\mathfrak{d}_q(H_a^+, H_a^-) = \frac{2a\sqrt{q_{22}}}{\sigma}. \tag{H.30}$$

Substitution into Proposition H.5 proves (H.27). Taking $a = 2\lambda$ gives (H.28). The deterministic finite-sample conclusion follows from Proposition H.2. □

Corollary H.8 (Exact-resolution discontinuity forces order-one risk)**.** *Suppose that the local direction class contains*

$$\left\{a e_2 e_2^\top, -a e_2 e_2^\top : 0 < a \leq a_0\right\} \tag{H.31}$$

for some $a_0 > 0$. For the exact-resolution Gaussian problem,

$$\inf_\delta \sup_{H \in \mathcal{H}} \mathbb{E}_H d_{\mathrm{H}}\left(\delta(W), \mathcal{K}_0(H)\right) \geq 1. \tag{H.32}$$

Proof. For every $a \in (0, a_0]$, the exact-resolution target sets under H_a^+ and H_a^- are opposite singleton endpoints and therefore have Hausdorff distance two. Proposition H.5 at $\lambda = 0$ gives the lower bound

$$2\Phi\left(-\frac{a\sqrt{q_{22}}}{\sigma}\right).$$

Letting $a \downarrow 0$ proves (H.32). □

H.5 Adaptive finite-subexperiment lower bounds

We finally record the local minimax consequence of the adaptive LAN result in Appendix G. Fix a sequence of predictable query rules (A_T) satisfying Assumption G.9 with deterministic full-support limit q. For a compact direction class $\mathcal{H}$, define

$$\mathfrak{R}_T^{\mathsf{A}}(\mathcal{H}, \lambda) := \inf_{\rho_T} \sup_{H \in \mathcal{H}} \mathbb{E}_{T,H}^{\mathsf{A}, \rho_T} d_{\mathrm{H}}\left(\widehat{C}_T, \mathcal{T}_{T,\lambda}(H)\right), \tag{H.33}$$

where the expectation is under the joint law induced by the local game, the fixed adaptive query rule, and the terminal decision kernel. The infimum is only over the terminal set-valued decision rule; the adaptive sampling rules A_T are fixed.

Lemma H.9 (Finite-class lower semicontinuity under weak convergence)**.** *Let Θ_0 be finite. Suppose experiments $\mathfrak{E}_T$ converge weakly in Le Cam's sense to $\mathfrak{E}$ on Θ_0, the decision space is compact metric, and bounded continuous losses $\ell_T(\theta,\cdot)$ converge uniformly to $\ell(\theta,\cdot)$ for $\theta\in\Theta_0$. Then*

$$\liminf_{T\to\infty}\mathcal{R}^\star(\mathfrak{E}_T,\ell_T)\geq\mathcal{R}^\star(\mathfrak{E},\ell). \tag{H.34}$$

Proof. Fix a prior π on Θ_0, augmenting the parameter set by a zero-weight base point if necessary. Relative to the base law, let Λ_T^θ and Λ^θ denote the likelihood-ratio coordinates. Weak convergence of finite experiments is equivalent to joint convergence of these vectors. The Bayes risk can be written as

$$B_T(\pi)=\mathbb{E}_0\left[\min_a\sum_{\theta\in\Theta_0}\pi_\theta\Lambda_T^\theta\ell_T(\theta,a)\right].$$

Compactness of the action space and continuity of the losses justify the pointwise minimum and a measurable minimizing selector. Uniform convergence of the losses changes the displayed expectation by at most the uniform loss error, because $\mathbb{E}_0\Lambda_T^\theta=1$. The likelihood ratios are nonnegative, converge in distribution, and have limiting expectations equal to one; they are therefore uniformly integrable. The minimum in the display is continuous in the likelihood-ratio vector and is dominated by a constant times $\sum_\theta\pi_\theta\Lambda_T^\theta$. Hence $B_T(\pi)\to B(\pi)$.

Minimax risk dominates Bayes risk, so

$$\liminf_T\mathcal{R}^\star(\mathfrak{E}_T,\ell_T)\geq B(\pi).$$

Taking the supremum over priors and applying the finite-parameter minimax theorem for randomized rules, compact actions, and bounded continuous losses gives (H.34). □

Lemma H.10 (Finite subsets approximate the Gaussian minimax risk)**.** *For compact $\mathcal{H}$ and fixed $\lambda>0$,*

$$\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H},\lambda)=\sup_{\substack{\mathcal{H}_0\subseteq\mathcal{H}\\ \mathcal{H}_0\ \text{finite}}}\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}_0,\lambda). \tag{H.35}$$

Proof. The right-hand side is at most the left-hand side. For the reverse inequality, fix a Gaussian decision rule δ, and let $R_\delta(H)$ denote its risk at direction H. For any $H,H'\in\mathcal{H}$, bounded-risk transfer under the identity kernel and Lemma H.1 give

$$|R_\delta(H)-R_\delta(H')|\leq 4d_{\mathrm{TV}}(Q_H^q,Q_{H'}^q)+d_{\mathrm{H}}\left(\mathcal{K}_\lambda(H),\mathcal{K}_\lambda(H')\right). \tag{H.36}$$

The right-hand side tends to zero uniformly as $\|H-H'\|_\infty\to 0$, by the exact Gaussian total-variation formula and the Hausdorff continuity of the visible correspondence on positive resolution levels. This modulus is independent of δ. If $\mathcal{H}_\varepsilon$ is a finite ε-net, then

$$\sup_{H\in\mathcal{H}}R_\delta(H)\leq\max_{H\in\mathcal{H}_\varepsilon}R_\delta(H)+o_\varepsilon(1).$$

Taking the infimum over δ and then $\varepsilon\downarrow 0$ proves the result. □

Proposition H.11 (Adaptive local asymptotic minimax lower bound)**.** *Under Assumption G.9, for every compact $\mathcal{H}\subseteq\mathbb{R}^{m\times n}$ and every $\lambda>0$,*

$$\liminf_{T\to\infty}\mathfrak{R}_T^{\mathsf{A}}(\mathcal{H},\lambda)\geq\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H},\lambda). \tag{H.37}$$

Proof. Fix a finite subset $\mathcal{H}_0\subseteq\mathcal{H}$, and augment it by the zero direction when necessary. Corollary G.12 gives weak convergence of the corresponding adaptive finite subexperiments to the Gaussian shift. The uniform target convergence in Appendix F gives uniform convergence of the bounded continuous Hausdorff losses on this finite class. Lemma H.9 therefore yields

$$\liminf_{T\to\infty}\mathfrak{R}_T^{\mathsf{A}}(\mathcal{H},\lambda)\geq\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}_0,\lambda).$$

Take the supremum over finite subsets and apply Lemma H.10. □

Remark H.12 (Lower bounds versus asymptotic equivalence). For deterministic allocations, the exact Gaussian reduction and quantitative Le Cam comparison yield the finite-sample lower transfer (H.14). For predictable adaptive allocations, LAN and weak convergence of finite subexperiments suffice for the one-sided minimax bound (H.37), but do not by themselves provide a uniform reverse randomization kernel over an uncountable direction class. The matching deterministic-design upper bound is established separately in Appendix I.

I Local Minimax Achievability

This appendix proves the matching upper bound for the deterministic-design local minimax problem introduced in Appendix H. The argument is decision theoretic rather than estimator specific. First, because every target correspondence is compact and convex, arbitrary compact-set-valued actions can be convexified without increasing Hausdorff loss. Second, randomization over compact convex actions can be removed pointwise by taking a support-function barycenter. Hence the Gaussian random-set minimax risk can be approached by Borel measurable, deterministic, finite-range rules. Finally, we apply such a Gaussian rule to the exact sufficient statistic from Appendix G; the only losses are the covariance-comparison error between the finite and limiting Gaussian shifts and the geometric target error from Appendix F. Together with the lower bound in Appendix H, this yields the exact local asymptotic minimax equivalence under deterministic full-support allocations.

Our use of support functions and compact convex random sets follows Schneider (2014); Molchanov (2017); comparison and transfer of statistical experiments follow Le Cam & Yang (2000).

I.1 Convexification and deterministic set-valued rules

Recall the compact action spaces

$$\mathsf{D} := \mathfrak{K}(\mathcal{Z}), \qquad \mathsf{D}_{\mathrm{cvx}} := \mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z}), \tag{I.1}$$

equipped with the Hausdorff metric. Every finite and limiting target in (H.2)–(H.3) belongs to $\mathsf{D}_{\mathrm{cvx}}$.

Lemma I.1 (Convexification is a Hausdorff contraction)**.** *Let $C, C' \in \mathfrak{K}(\mathcal{Z})$, and let $D \in \mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$. Then*

$$d_{\mathrm{H}}\big(\operatorname{conv} C, \operatorname{conv} C'\big) \le d_{\mathrm{H}}(C, C'), \tag{I.2}$$

$$d_{\mathrm{H}}\big(\operatorname{conv} C, D\big) \le d_{\mathrm{H}}(C, D). \tag{I.3}$$

In particular, the convex-hull map $C \mapsto \operatorname{conv} C$ is continuous on D.

Proof. Let $r := d_{\mathrm{H}}(C, C')$, and let $\mathbb{B}_{\mathsf{V}}$ denote the closed unit ball of $\|\cdot\|_{\mathsf{V}}$. The definition of the Hausdorff distance gives

$$C \subseteq C' + r\mathbb{B}_{\mathsf{V}}, \qquad C' \subseteq C + r\mathbb{B}_{\mathsf{V}}.$$

Because Minkowski addition with the convex set $r\mathbb{B}_{\mathsf{V}}$ commutes with taking convex hulls,

$$\operatorname{conv} C \subseteq \operatorname{conv} C' + r\mathbb{B}_{\mathsf{V}}, \qquad \operatorname{conv} C' \subseteq \operatorname{conv} C + r\mathbb{B}_{\mathsf{V}}.$$

This proves (I.2). If D is convex, then $\operatorname{conv} D = D$, so (I.3) follows by setting $C' = D$. □

For a probability measure η on $\mathsf{D}_{\mathrm{cvx}}$, define its support-function barycenter as follows. For $u \in \mathsf{V}$, let

$$\overline{h}_\eta(u) := \int_{\mathsf{D}_{\mathrm{cvx}}} h_C(u)\, \eta(dC), \tag{I.4}$$

where h_C is defined in (A.15). The function $\overline{h}_\eta$ is finite, continuous, positively homogeneous, and subadditive. Hence it is the support function of a unique nonempty compact convex set, which we denote by

$$\mathfrak{B}(\eta) \in \mathfrak{K}_{\mathrm{cvx}}(\mathsf{V}). \tag{I.5}$$

Since $C \subseteq \mathcal{Z}$ for η-almost every C,

$$\overline{h}_\eta(u) \le h_{\mathcal{Z}}(u) \qquad \text{for every } u \in \mathsf{V}.$$

The support-function inclusion criterion therefore implies

$$\mathfrak{B}(\eta) \subseteq \mathcal{Z}, \tag{I.6}$$

so in fact $\mathfrak{B}(\eta) \in \mathsf{D}_{\mathrm{cvx}}$. This construction is the compact-convex Minkowski expectation written in support-function form.

Lemma I.2 (Barycentric Jensen inequality for Hausdorff loss)**.** *For every probability measure η on $\mathsf{D}_{\mathrm{cvx}}$ and every $D \in \mathsf{D}_{\mathrm{cvx}}$,*

$$d_{\mathrm{H}}\big(\mathfrak{B}(\eta), D\big) \le \int_{\mathsf{D}_{\mathrm{cvx}}} d_{\mathrm{H}}(C, D)\, \eta(dC). \tag{I.7}$$

Proof. By the support-function representation (A.17),

$$\begin{aligned} d_{\mathrm{H}}\big(\mathfrak{B}(\eta), D\big) &= \sup_{\|u\|_{\mathsf{V},*}\le 1} \big|\overline{h}_{\eta}(u) - h_D(u)\big| \\ &= \sup_{\|u\|_{\mathsf{V},*}\le 1} \left| \int \big(h_C(u) - h_D(u)\big)\, \eta(dC) \right| \\ &\le \int \sup_{\|u\|_{\mathsf{V},*}\le 1} |h_C(u) - h_D(u)|\, \eta(dC) \\ &= \int d_{\mathrm{H}}(C, D)\, \eta(dC). \end{aligned}$$

□

Lemma I.3 (Measurability of conditional support barycenters)**.** *Let $(\mathsf{X}, \mathcal{X})$ be a measurable space and let ρ be a Markov kernel from X to $\mathsf{D}_{\mathrm{cvx}}$. Then the map*

$$\bar{\rho}(x) := \mathfrak{B}\big(\rho(x, \cdot)\big) \tag{I.8}$$

is a Borel measurable map from X to $\mathsf{D}_{\mathrm{cvx}}$.

Proof. Let $\{u_\ell\}_{\ell\ge 1}$ be a countable dense subset of the dual unit ball $\{u : \|u\|_{\mathsf{V},*} \le 1\}$, and let $\{D_j\}_{j\ge 1}$ be a countable dense subset of the compact metric space $\mathsf{D}_{\mathrm{cvx}}$. For each fixed u_ℓ, the map

$$x \longmapsto \int h_C(u_\ell)\, \rho(x, dC)$$

is measurable because $C \mapsto h_C(u_\ell)$ is continuous. By continuity of support functions in the direction variable,

$$d_{\mathrm{H}}\big(\bar{\rho}(x), D_j\big) = \sup_{\ell\ge 1} \left| \int h_C(u_\ell)\, \rho(x, dC) - h_{D_j}(u_\ell) \right|. \tag{I.9}$$

The right-hand side is a countable supremum of measurable functions and is therefore measurable. Distances to a countable dense family generate the Borel sigma-field of a separable metric space, proving measurability of $\bar{\rho}$. □

Proposition I.4 (Convex deterministic rules are complete)**.** *Consider any statistical experiment and any compact-convex target map $S(\theta) \in \mathsf{D}_{\mathrm{cvx}}$, with loss $d_{\mathrm{H}}(C, S(\theta))$. Then:*

1. *restricting actions from D to $\mathsf{D}_{\mathrm{cvx}}$ does not increase the minimax risk; and*
2. *restricting randomized Markov decision rules to Borel measurable deterministic rules with values in $\mathsf{D}_{\mathrm{cvx}}$ does not increase the minimax risk.*

Consequently, all minimax risks in (H.7) *and* (H.9) *are unchanged if their infima are restricted to deterministic compact-convex-valued estimators.*

Proof. Given an arbitrary compact-valued output C, replace it by $\operatorname{conv} C$. Lemma I.1 shows that this cannot increase loss against a convex target, and the convex-hull map is continuous and hence measurable.

Now let $\rho(x, dC)$ be a randomized rule taking values in $\mathsf{D}_{\mathrm{cvx}}$, and replace it by the deterministic map $\bar{\rho}$ in (I.8). For every observation x and every target $D \in \mathsf{D}_{\mathrm{cvx}}$, Lemma I.2 gives

$$d_{\mathrm{H}}\big(\bar{\rho}(x), D\big) \le \int d_{\mathrm{H}}(C, D)\, \rho(x, dC).$$

Lemma I.3 makes $\bar{\rho}$ an admissible deterministic rule. Integrating the displayed inequality under every parameter law proves the claim. □

I.2 Finite-range Gaussian near-minimax rules

The next result removes any hidden measurable-selection issue from the achievability construction. Exact attainment of the Gaussian minimax infimum is unnecessary: finite-range deterministic rules can approach it arbitrarily closely.

Lemma I.5 (Measurable finite-net projection). *For every $r > 0$, there exist $C_1, \dots, C_N \in \mathsf{D}_{\mathrm{cvx}}$ and a Borel measurable map*

$$\Pi_r : \mathsf{D}_{\mathrm{cvx}} \longrightarrow \{C_1, \dots, C_N\} \tag{I.10}$$

such that

$$d_{\mathrm{H}}\big(\Pi_r(C), C\big) \le r \qquad \text{for every } C \in \mathsf{D}_{\mathrm{cvx}}. \tag{I.11}$$

Proof. Compactness of $\mathsf{D}_{\mathrm{cvx}}$ gives a finite r-net $\{C_1, \dots, C_N\}$. Let $\Pi_r(C)$ be a nearest net point, with ties broken by the smallest index. The tie-broken Voronoi cells are Borel because $C \mapsto d_{\mathrm{H}}(C, C_j)$ is continuous. This proves measurability and (I.11). □

Proposition I.6 (Finite-range Gaussian near-minimax rules). *Fix compact $\mathcal{H}$, $\lambda > 0$, and $q \in \mathrm{ri}\,\Delta_d$. For every $\eta > 0$, there exists a Borel measurable deterministic rule*

$$\delta_\eta : \mathbb{R}^d \longrightarrow \mathsf{D}_{\mathrm{cvx}} \tag{I.12}$$

with finite range such that

$$\sup_{H \in \mathcal{H}} \mathbb{E}_{Q_H^q} d_{\mathrm{H}}\left(\delta_\eta(W), \mathcal{K}_\lambda(H)\right) \le \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) + \eta. \tag{I.13}$$

Proof. By the definition of the infimum in (H.9), there exists a randomized rule whose worst-case risk is at most $\mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) + \eta/2$. Proposition I.4 replaces it by a Borel deterministic compact-convex-valued rule without increasing risk. Compose that rule with the projection $\Pi_{\eta/2}$ from Lemma I.5. The triangle inequality for d_{H} increases every pointwise loss by at most $\eta/2$, proving (I.13). □

Remark I.7 (No exact minimizer is assumed). Proposition I.6 uses only the definition of an infimum, compactness of the action hyperspace, and the barycentric derandomization above. It does not assume that an exact minimax rule exists, nor does it invoke an unverified measurable selector for an argmin correspondence.

I.3 Transfer from the Gaussian experiment

For full-support designs $q_T, q \in \mathrm{ri}\,\Delta_d$, define the covariance-comparison error

$$\zeta_T(q_T, q) := d_{\mathrm{TV}}\left(\mathcal{N}(0, \Sigma(q_T)), \mathcal{N}(0, \Sigma(q))\right). \tag{I.14}$$

By Proposition G.4,

$$\zeta_T(q_T, q) \le \frac{1}{2}\left[\sum_{a=1}^{d} \phi\left(\frac{q_a}{q_{a,T}}\right)\right]^{1/2}, \qquad \phi(r) = r - 1 - \log r. \tag{I.15}$$

Let W_T be the exact Gaussian sufficient statistic from (G.13). For a Gaussian rule δ_η, define the finite-sample estimator

$$\widehat{C}_{T,\eta} := \delta_\eta(W_T). \tag{I.16}$$

This is a deterministic, finite-range, compact-convex-valued estimator of the entire equilibrium correspondence. We write $\mathbb{E}_H^{q_T}$ for expectation under the deterministic-design local law $P_{T,H}^{q_T}$.

Proposition I.8 (Finite-sample deterministic-design upper transfer). *Let $\mathcal{H} \subseteq \mathbb{R}^{m \times n}$ be compact, let $\lambda > 0$, and let $q_T, q \in \mathrm{ri}\,\Delta_d$. For every $\eta > 0$, there exists an estimator of the form* (I.16) *satisfying*

$$\sup_{H \in \mathcal{H}} \mathbb{E}_H^{q_T} d_{\mathrm{H}}\left(\widehat{C}_{T,\eta}, \mathcal{T}_{T,\lambda}(H)\right) \le \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) + \eta + \omega_T(\mathcal{H}, \lambda) + 4\zeta_T(q_T, q). \tag{I.17}$$

Consequently,

$$\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda) \le \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda) + \omega_T(\mathcal{H}, \lambda) + 4\zeta_T(q_T, q). \tag{I.18}$$

Proof. Choose δ_η according to Proposition I.6. Under local direction H, Proposition G.2 gives

$$W_T \sim \mathcal{N}\big(\mathrm{vec}(H), \Sigma(q_T)\big).$$

By the target triangle inequality,

$$\begin{aligned}&\mathbb{E}_H^{q_T} d_{\mathrm{H}}\left(\delta_\eta(W_T), \mathcal{T}_{T,\lambda}(H)\right)\\ &\quad\le \mathbb{E}_H^{q_T} d_{\mathrm{H}}\left(\delta_\eta(W_T), \mathcal{K}_\lambda(H)\right) + d_{\mathrm{H}}\left(\mathcal{T}_{T,\lambda}(H), \mathcal{K}_\lambda(H)\right).\end{aligned} \tag{I.19}$$

The second term is at most $\omega_T(\mathcal{H}, \lambda)$. The first integrand takes values in $[0, 4]$. Translation invariance of total variation and data processing through the measurable rule δ_η therefore imply

$$\begin{aligned}&\mathbb{E}_{\mathcal{N}(\mathrm{vec}(H),\Sigma(q_T))} d_{\mathrm{H}}\left(\delta_\eta(W), \mathcal{K}_\lambda(H)\right)\\ &\quad\le \mathbb{E}_{\mathcal{N}(\mathrm{vec}(H),\Sigma(q))} d_{\mathrm{H}}\left(\delta_\eta(W), \mathcal{K}_\lambda(H)\right) + 4\zeta_T(q_T, q).\end{aligned} \tag{I.20}$$

Taking the supremum over H, applying (I.13), and combining with (I.19) proves (I.17). Taking the infimum over all finite-sample estimators and then letting $\eta \downarrow 0$ gives (I.18). □

Remark I.9 (An abstract deficiency-kernel version). The preceding proof uses the explicit sufficient statistic and the identity comparison between the two common-mean Gaussian shifts. Equivalently, one may compose a Gaussian near-minimax rule with any randomization kernel whose deficiency from $\mathfrak{E}_T^{\det}(q_T, \Theta)$ to $\mathfrak{G}(q, \Theta)$ is within $o(1)$ of optimal. Lemma G.1 then gives the same conclusion with $4\Delta(\mathfrak{E}_T^{\det}, \mathfrak{G})$ in place of $4\zeta_T$. The statistic-based construction is retained because it is explicit and deterministic.

I.4 Completion of the minimax equivalence

Combining the upper transfer with Proposition H.2 yields a quantitative finite-sample sandwich.

Proposition I.10 (Quantitative minimax sandwich)**.** *Under the assumptions of Proposition I.8,*

$$\left|\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda) - \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda)\right| \le \omega_T(\mathcal{H}, \lambda) + 4\zeta_T(q_T, q). \tag{I.21}$$

For the rounded deterministic designs of Corollary G.5, whenever $Tq_{\min} \ge 2$,

$$\left|\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda) - \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda)\right| \le \omega_T(\mathcal{H}, \lambda) + \frac{3\sqrt{2d}}{Tq_{\min}}. \tag{I.22}$$

Proof. The upper inequality is (I.18). For the lower inequality, Proposition H.2 gives

$$\mathfrak{R}_{T,q_T} \ge \mathfrak{R}_q^{\mathrm{G}} - \omega_T - 4\varepsilon_T,$$

where $\varepsilon_T = \Delta(\mathfrak{E}_T^{\det}(q_T, \Theta), \mathfrak{G}(q, \Theta))$. Proposition G.4 gives $\varepsilon_T \le \zeta_T(q_T, q)$, proving (I.21). For rounded designs, the covariance bound (G.23) together with the rounding argument in Corollary G.5 yields

$$4\zeta_T(q_T, q) \le \frac{3\sqrt{2d}}{Tq_{\min}},$$

which proves (I.22). □

Proposition I.11 (Exact local asymptotic minimax equivalence)**.** *Let $\mathcal{H} \subseteq \mathbb{R}^{m\times n}$ be compact, let $\lambda > 0$, and let $q_T, q \in \mathrm{ri}\,\Delta_d$ satisfy $q_T \to q$. Then*

$$\boxed{\lim_{T\to\infty} \mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda) = \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda).} \tag{I.23}$$

Moreover, for every deterministic sequence $\eta_T \downarrow 0$, there exist Borel measurable, deterministic, finite-range estimators $\widehat{C}_T \in \mathsf{D}_{\mathrm{cvx}}$ such that

$$\limsup_{T\to\infty} \sup_{H\in\mathcal{H}} \mathbb{E}_H^{q_T} d_{\mathrm{H}}\left(\widehat{C}_T, \mathcal{T}_{T,\lambda}(H)\right) \le \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda). \tag{I.24}$$

Proof. Theorem 2 gives $\omega_T(\mathcal{H}, \lambda) \to 0$, while Proposition G.4 gives $\zeta_T(q_T, q) \to 0$. Equation (I.23) follows from Proposition I.10.

For each T, choose the finite-range Gaussian rule δ_{η_T} from Proposition I.6 and set $\widehat{C}_T = \delta_{\eta_T}(W_T)$. Proposition I.8 yields

$$\sup_{H \in \mathcal{H}} \mathbb{E}_H^{q_T} d_{\mathrm{H}} \left(\widehat{C}_T, \mathcal{T}_{T,\lambda}(H) \right) \leq \mathfrak{R}_q^{\mathrm{G}} + \eta_T + \omega_T + 4\zeta_T.$$

Taking the limit superior proves (I.24). □

Remark I.12 (Local-oracle character of the theorem). The statistic W_T in (G.13) is centered at the fixed base game $A_\star$. Accordingly, Proposition I.11 is a local-oracle benchmark in the standard sense of local asymptotic decision theory: the center and limiting design are fixed when the local experiment is defined. The theorem does not claim that one globally implementable procedure attains all local oracle risks simultaneously. Appendix K shows that such cost-free uniform adaptation is impossible across regular and bifurcating games.

Remark I.13 (What has been achieved). The lower theory in Appendix H and the present upper theory match under identical assumptions: compact local direction class, positive resolution, fixed finite game dimension, known Gaussian noise, and deterministic full-support allocations converging to q. No explicit closed form for the optimal Gaussian decision rule is asserted. What is proved is stronger than existence of a favorable plug-in estimator: every Gaussian near-minimax risk can be transferred to the original payoff experiment with the decomposition $\omega_T + 4\zeta_T$, where only the experiment term is explicitly quantified. A sequence of deterministic finite-range correspondence estimators approaches the resulting asymptotic benchmark.

J Resolution Trichotomy and Confidence Corridors

This appendix synthesizes the geometric and statistical results of Appendices F–I. The first part identifies three qualitatively different regimes according to the normalized approximation tolerance

$$\lambda_T := \sqrt{T}\, \tau_T.$$

At a strategic bifurcation, subcritical resolution $\lambda_T \to 0$ leaves order-one branch uncertainty; critical resolution $\lambda_T \to \lambda \in (0, \infty)$ converges to the nondegenerate Gaussian random-set decision problem; and supercritical resolution $\lambda_T \to \infty$, subject to $\tau_T \to 0$, collapses uniformly to the base equilibrium correspondence. The second part constructs finite-sample inner–outer confidence corridors for the complete approximate equilibrium correspondence. The construction is nonasymptotic and remains valid at bifurcations, where confidence regions based on a differentiable single-valued equilibrium selector are unavailable. Our set-convergence conventions follow Rockafellar & Wets (1998); the testing arguments use the decision-theoretic bounds established in Appendix H; and the corridor-width bound is a second application of Hoffman's polyhedral error theorem (Hoffman, 1952; Peña et al., 2021).

Throughout, let

$$t_T := T^{-1/2}, \qquad \lambda_T := \sqrt{T}\, \tau_T, \qquad \mathcal{T}_T^{\lambda_T}(H) := \mathcal{E}_{t_T \lambda_T}(A_\star + t_T H). \tag{J.1}$$

For a deterministic full-support allocation q_T, we use the notation

$$\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) \tag{J.2}$$

for the minimax risk in (H.7), with the fixed normalized resolution λ there replaced by the possibly T-dependent value λ_T.

J.1 Subcritical resolution

We first record a consequence of the joint epi-limit that is specific to a unique zero-cost first-order branch. It permits the approximation level to converge to zero and therefore lies outside the positive-resolution theorem stated in Theorem 2.

Proposition J.1 (Vanishing-resolution convergence under a unique first-order branch)**.** *Let*

$$t_k \downarrow 0, \qquad H_k \to H, \qquad \lambda_k \to 0, \qquad \lambda_k \geq 0.$$

If

$$\mathcal{K}_0(H) = \{z_H\}, \tag{J.3}$$

then

$$d_{\mathrm{H}}\left(\mathcal{E}_{t_k\lambda_k}(A_\star + t_k H_k), \{z_H\}\right) \longrightarrow 0. \tag{J.4}$$

Proof. The target sets are nonempty because each perturbed finite game has an exact saddle point. Suppose that (J.4) fails. Then, after passing to a subsequence, there exist

$$z_k \in \mathcal{E}_{t_k\lambda_k}(A_\star + t_k H_k)$$

and $\varepsilon > 0$ such that

$$\|z_k - z_H\|_{\mathsf{V}} \geq \varepsilon \qquad \text{for every } k. \tag{J.5}$$

Compactness of $\mathcal{Z}$ yields a further subsequence, not relabeled, with $z_k \to z$. Since

$$\mathcal{F}_{t_k}(H_k, z_k) = \frac{g_{A_\star + t_k H_k}(z_k)}{t_k} \leq \lambda_k,$$

the joint epi-liminf inequality (F.16) gives

$$\beta_{A_\star}(z; H) \leq \liminf_{k\to\infty} \lambda_k = 0.$$

The branch cost is nonnegative; hence $z \in \mathcal{K}_0(H) = \{z_H\}$, contradicting (J.5). Therefore

$$\sup_{z \in \mathcal{E}_{t_k\lambda_k}(A_\star + t_k H_k)} \|z - z_H\|_{\mathsf{V}} \longrightarrow 0.$$

Since each target set is nonempty, this supremum also dominates the distance from z_H to the target set, proving Hausdorff convergence. □

The next result turns separated first-order branches into a local minimax lower bound. Recall the information distance $\mathfrak{d}_q$ from (H.19).

Proposition J.2 (Subcritical bifurcation lower bound)**.** *Let $\mathcal{H} \subseteq \mathbb{R}^{m\times n}$ be compact, let $q_T, q \in \operatorname{ri}\Delta_d$ satisfy $q_T \to q$, and let $\lambda_T \to 0$ with $\lambda_T \geq 0$. Suppose that there exist $H_0, H_1 \in \mathcal{H}$ and distinct points $z_0, z_1 \in \mathcal{Z}$ such that*

$$\mathcal{K}_0(H_r) = \{z_r\}, \qquad r \in \{0, 1\}. \tag{J.6}$$

Then

$$\liminf_{T\to\infty} \mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) \geq \|z_0 - z_1\|_{\mathsf{V}} \Phi\left(-\frac{\mathfrak{d}_q(H_0, H_1)}{2}\right) > 0. \tag{J.7}$$

Proof. Proposition J.1, applied with $t_T = T^{-1/2}$, gives

$$\mathcal{T}_T^{\lambda_T}(H_r) \underset{d_{\mathrm{H}}}{\longrightarrow} \{z_r\}, \qquad r \in \{0, 1\}.$$

Consequently,

$$d_{\mathrm{H}}\left(\mathcal{T}_T^{\lambda_T}(H_0), \mathcal{T}_T^{\lambda_T}(H_1)\right) \longrightarrow \|z_0 - z_1\|_{\mathsf{V}}. \tag{J.8}$$

The exact finite-sample pairwise lower bound (H.23) gives

$$\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) \geq d_{\mathrm{H}}\left(\mathcal{T}_T^{\lambda_T}(H_0), \mathcal{T}_T^{\lambda_T}(H_1)\right) \times \Phi\left(-\frac{\mathfrak{d}_{q_T}(H_0, H_1)}{2}\right).$$

Since $q_T \to q$, the information distance converges. Taking the limit inferior and using (J.8) proves (J.7). □

Corollary J.3 (Sharp order-one uncertainty in the guiding bifurcation)**.** *Let*

$$A_\star = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and suppose that, for some $a_0 > 0$, the local direction class contains

$$\left\{ae_2e_2^\top, -ae_2e_2^\top : 0 < a \le a_0\right\}. \tag{J.9}$$

If $q_T \to q \in \operatorname{ri}\Delta_4$ and $\lambda_T \to 0$, then

$$\liminf_{T\to\infty} \mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) \ge 1. \tag{J.10}$$

Proof. For every fixed $a \in (0, a_0]$, the two zero-cost visible correspondences under $H_a^+ = ae_2e_2^\top$ and $H_a^- = -ae_2e_2^\top$ are the opposite endpoints $\{z_0\}$ and $\{z_1\}$, whose product-ℓ_1 distance is two. Moreover,

$$\mathfrak{d}_q(H_a^+, H_a^-) = \frac{2a\sqrt{q_{22}}}{\sigma}.$$

Proposition J.2 therefore gives

$$\liminf_T \mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) \ge 2\Phi\left(-\frac{a\sqrt{q_{22}}}{\sigma}\right)$$

for every $a \in (0, a_0]$. Letting $a \downarrow 0$ proves (J.10). □

J.2 Critical resolution

Let $\lambda_T \to \lambda \in (0, \infty)$, and define the varying-level geometric discrepancy

$$\omega_T^{\mathrm{crit}}(\mathcal{H}; \lambda_T, \lambda) := \sup_{H\in\mathcal{H}} d_{\mathrm{H}}\left(\mathcal{E}_{\lambda_T/\sqrt{T}}\left(A_\star + \frac{H}{\sqrt{T}}\right), \mathcal{K}_\lambda(H)\right). \tag{J.11}$$

The parameterized positive-resolution theorem (F.24) and its uniform version imply

$$\omega_T^{\mathrm{crit}}(\mathcal{H}; \lambda_T, \lambda) \longrightarrow 0 \qquad \text{for every compact } \mathcal{H}. \tag{J.12}$$

Proposition J.4 (Critical-resolution minimax convergence)**.** *Let $\mathcal{H} \subseteq \mathbb{R}^{m\times n}$ be compact, let $q_T, q \in \operatorname{ri}\Delta_d$ satisfy $q_T \to q$, and let $\lambda_T \to \lambda \in (0, \infty)$. Then*

$$\left|\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) - \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda)\right| \le \omega_T^{\mathrm{crit}}(\mathcal{H}; \lambda_T, \lambda) + 4\zeta_T(q_T, q), \tag{J.13}$$

and therefore

$$\boxed{\lim_{T\to\infty} \mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) = \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda).} \tag{J.14}$$

For the balanced rounded designs of Corollary G.5, whenever $Tq_{\min} \ge 2$,

$$\left|\mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) - \mathfrak{R}_q^{\mathrm{G}}(\mathcal{H}, \lambda)\right| \le \omega_T^{\mathrm{crit}}(\mathcal{H}; \lambda_T, \lambda) + \frac{3\sqrt{2d}}{Tq_{\min}}. \tag{J.15}$$

Proof. Replace the finite target in Appendices H and I by

$$H \longmapsto \mathcal{E}_{\lambda_T/\sqrt{T}}\left(A_\star + \frac{H}{\sqrt{T}}\right)$$

and the limiting target by $H \mapsto \mathcal{K}_\lambda(H)$. The target-perturbation lemma (H.10) changes every risk by at most ω_T^{crit}. The experiment comparison is unchanged, so the lower and upper transfers incur at most $4\zeta_T(q_T, q)$. This proves (J.13). Equation (J.14) follows from (J.12) and $\zeta_T(q_T, q) \to 0$. The rounded-design bound follows from (G.29) exactly as in Proposition I.10. □

J.3 SUPERCRITICAL RESOLUTION

The supercritical regime requires two conditions:

$$\tau_T \longrightarrow 0, \qquad \sqrt{T}\,\tau_T = \lambda_T \longrightarrow \infty. \tag{J.16}$$

The first keeps the target asymptotically exact, while the second makes its tolerance dominate the local payoff noise.

Proposition J.5 (Uniform supercritical collapse)**.** *Let $\mathcal{H} \subseteq \mathbb{R}^{m\times n}$ be bounded, and set*

$$R_{\mathcal{H}} := \sup_{H\in\mathcal{H}} \|H\|_\infty. \tag{J.17}$$

Suppose (J.16) *holds. Then, for all sufficiently large T,*

$$\mathcal{E}_0(A_\star) \subseteq \mathcal{E}_{\tau_T}\left(A_\star + \frac{H}{\sqrt{T}}\right) \qquad \text{for every } H \in \mathcal{H}, \tag{J.18}$$

and

$$\sup_{H\in\mathcal{H}} d_{\mathrm{H}}\left(\mathcal{E}_{\tau_T}\left(A_\star + \frac{H}{\sqrt{T}}\right), \mathcal{E}_0(A_\star)\right) \le \kappa_\star\left(\tau_T + \frac{2R_{\mathcal{H}}}{\sqrt{T}}\right) \longrightarrow 0. \tag{J.19}$$

Consequently, for every sequence of deterministic designs q_T,

$$0 \le \mathfrak{R}_{T,q_T}(\mathcal{H}, \lambda_T) \le \kappa_\star\left(\tau_T + \frac{2R_{\mathcal{H}}}{\sqrt{T}}\right) \longrightarrow 0. \tag{J.20}$$

Proof. For every $z \in \mathcal{E}_0(A_\star)$, the payoff perturbation bound (C.25) gives

$$g_{A_\star + H/\sqrt{T}}(z) \le \frac{2\|H\|_\infty}{\sqrt{T}} \le \frac{2R_{\mathcal{H}}}{\sqrt{T}}.$$

Since $\sqrt{T}\,\tau_T \to \infty$, the final quantity is at most τ_T for all sufficiently large T, proving (J.18).

In the reverse direction, Proposition C.10 gives

$$e\left(\mathcal{E}_{\tau_T}\left(A_\star + \frac{H}{\sqrt{T}}\right), \mathcal{E}_0(A_\star)\right) \le \kappa_\star\left(\tau_T + \frac{2R_{\mathcal{H}}}{\sqrt{T}}\right).$$

The reverse directed excess is zero by (J.18), proving the Hausdorff bound. Finally, the constant estimator $\widehat{C}_T \equiv \mathcal{E}_0(A_\star)$ attains the right-hand side of (J.19), which proves (J.20). □

Corollary J.6 (Statistical resolution trichotomy)**.** *Under deterministic full-support designs converging to q, the local correspondence problem has the following three regimes.*

1. Subcritical resolution: *if $\lambda_T \to 0$, any pair of separated singleton zero-cost branches gives the positive lower bound* (J.7)*; in the guiding two-sided bifurcation, the lower bound is at least one.*
2. Critical resolution: *if $\lambda_T \to \lambda \in (0,\infty)$, the minimax risk converges to the nondegenerate Gaussian random-set risk in* (J.14)*.*
3. Supercritical resolution: *if $\lambda_T \to \infty$ and $\lambda_T/\sqrt{T} \to 0$, the minimax risk converges to zero by* (J.20)*.*

Thus $T^{-1/2}$ is the first-order statistical scale at which branch multiplicity remains nontrivial.

J.4 FINITE-SAMPLE CONFIDENCE CORRIDORS

We now construct simultaneous inner and outer confidence sets for the entire resolution-τ equilibrium correspondence. The construction is based only on a uniform payoff confidence event and therefore does not require uniqueness, strict complementarity, or differentiability of an equilibrium selector.

Proposition J.7 (Deterministic correspondence sandwich)**.** *Let $A, \widehat{A} \in \mathbb{R}^{m\times n}$, let $r \geq 0$, and suppose that*

$$\|\widehat{A} - A\|_\infty \leq r. \tag{J.21}$$

Then, simultaneously for every $\tau \geq 0$,

$$\boxed{\mathcal{E}_{\tau-2r}(\widehat{A}) \subseteq \mathcal{E}_\tau(A) \subseteq \mathcal{E}_{\tau+2r}(\widehat{A}),} \tag{J.22}$$

where $\mathcal{E}_s(\widehat{A}) = \varnothing$ for $s < 0$.

Proof. The residual perturbation inequality (C.25) gives

$$\left|g_{\widehat{A}}(z) - g_A(z)\right| \leq 2r \qquad \text{for every } z \in \mathcal{Z}.$$

If $g_{\widehat{A}}(z) \leq \tau - 2r$, then $g_A(z) \leq \tau$, proving the inner inclusion. If $g_A(z) \leq \tau$, then $g_{\widehat{A}}(z) \leq \tau + 2r$, proving the outer inclusion. Because the same event controls all profiles, both inclusions hold simultaneously for every $\tau \geq 0$. □

J.4.1 Coordinatewise payoff uncertainty

Let $R = (r_{ij}) \in \mathbb{R}_+^{m\times n}$ and define, for $z = (x, y)$,

$$\eta_R(x, y) := \max_i e_i^\top R y + \max_j x^\top R e_j. \tag{J.23}$$

Lemma J.8 (Coordinatewise residual perturbation)**.** *If $|\widehat{A}_{ij} - A_{ij}| \leq r_{ij}$ for every (i, j), then*

$$\left|g_{\widehat{A}}(z) - g_A(z)\right| \leq \eta_R(z) \leq 2r_\infty, \qquad r_\infty := \max_{i,j} r_{ij}. \tag{J.24}$$

Consequently, simultaneously for every $\tau \geq 0$,

$$\left\{z : g_{\widehat{A}}(z) + \eta_R(z) \leq \tau\right\} \subseteq \mathcal{E}_\tau(A) \subseteq \left\{z : g_{\widehat{A}}(z) - \eta_R(z) \leq \tau\right\}. \tag{J.25}$$

Proof. For the row best-response value,

$$\left|\max_i e_i^\top \widehat{A} y - \max_i e_i^\top A y\right| \leq \max_i \sum_j r_{ij} y_j.$$

The column term is bounded by $\max_j \sum_i x_i r_{ij}$. Adding proves the first inequality; probability normalization proves the scalar envelope. The two set inclusions follow by adding or subtracting the profile-dependent bound. □

For deterministic Gaussian counts, choose numbers $\delta_{ij} > 0$ with $\sum_{i,j} \delta_{ij} \leq \delta$ and set

$$r_{ij,T} := \sigma\sqrt{\frac{2\log(2/\delta_{ij})}{N_{ij,T}}}. \tag{J.26}$$

A union bound gives the coordinatewise event with probability at least $1 - \delta$, and therefore the simultaneous sandwich (J.25). Unlike the scalar corridor, this retains the allocation-dependent uncertainty seen by the particular profile.

Consider now a deterministic Gaussian payoff design with entry counts $N_{ij,T} \geq 1$ and entrywise sample means

$$\widehat{A}_{ij,T} := \frac{1}{N_{ij,T}} \sum_{\ell=1}^{N_{ij,T}} Y_{ij,\ell}. \tag{J.27}$$

For $\delta \in (0, 1)$, define

$$r_T(\delta) := \sigma \max_{\substack{i\in[m]\\ j\in[n]}} \sqrt{\frac{2\log(2mn/\delta)}{N_{ij,T}}}. \tag{J.28}$$

Proposition J.9 (Simultaneous finite-sample confidence corridor)**.** *For every payoff matrix A,*

$$\mathbb{P}_A\left(\|\widehat{A}_T - A\|_\infty \le r_T(\delta)\right) \ge 1-\delta. \tag{J.29}$$

Consequently, with probability at least $1-\delta$, the inclusions

$$\mathcal{E}_{\tau-2r_T(\delta)}(\widehat{A}_T) \subseteq \mathcal{E}_\tau(A) \subseteq \mathcal{E}_{\tau+2r_T(\delta)}(\widehat{A}_T) \tag{J.30}$$

hold simultaneously for every $\tau \ge 0$.

Proof. For each entry,

$$\widehat{A}_{ij,T} - A_{ij} \sim \mathcal{N}\left(0, \frac{\sigma^2}{N_{ij,T}}\right).$$

The standard Gaussian tail bound gives

$$\mathbb{P}_A\left(|\widehat{A}_{ij,T} - A_{ij}| > r_T(\delta)\right) \le 2\exp\left(-\frac{N_{ij,T} r_T(\delta)^2}{2\sigma^2}\right) \le \frac{\delta}{mn}.$$

A union bound proves (J.29); Proposition J.7 then gives (J.30). □

We next quantify the width of the corridor. This requires a level-set version of the Hoffman error bound that is uniform in the resolution level. For $s \ge 0$, define the lifted polyhedron

$$\mathcal{Q}_s(A) := \left\{(x,y,v) \in \Delta_m \times \Delta_n \times \mathbb{R} : Ay \le v\mathbf{1}_m,\ -A^\top x + v\mathbf{1}_n \le s\mathbf{1}_n\right\}. \tag{J.31}$$

Its projection onto (x,y) is exactly $\mathcal{E}_s(A)$. Equip the lifted space with

$$\|(x,y,v)\|_\sharp := \|x\|_1 + \|y\|_1 + |v|. \tag{J.32}$$

Because the coefficient matrix of (J.31) is independent of s, Hoffman's theorem yields a finite constant $H_{\mathrm{lev}}(A)$ valid simultaneously for all right-hand sides $s \ge 0$. Set

$$\kappa_{\mathrm{lev}}(A) := \frac{1}{2} H_{\mathrm{lev}}(A). \tag{J.33}$$

Proposition J.10 (Uniform error bound for approximate-equilibrium levels)**.** *For every $A \in \mathbb{R}^{m\times n}$, every $s \ge 0$, and every $z \in \mathcal{Z}$,*

$$\operatorname{dist}\left(z, \mathcal{E}_s(A)\right) \le \kappa_{\mathrm{lev}}(A)\left[g_A(z) - s\right]_+. \tag{J.34}$$

Proof. Let

$$\overline{u} := \max_i e_i^\top A y, \qquad \underline{u} := \min_j x^\top A e_j,$$

and choose

$$v_s(z) := \frac{\overline{u} + \underline{u} + s}{2}. \tag{J.35}$$

The simplex constraints are satisfied exactly. The largest violation of $Ay \le v_s(z)\mathbf{1}_m$ is

$$\left[\overline{u} - v_s(z)\right]_+ = \frac{1}{2}\left[g_A(z) - s\right]_+,$$

and the largest violation of $-A^\top x + v_s(z)\mathbf{1}_n \le s\mathbf{1}_n$ is

$$\left[v_s(z) - s - \underline{u}\right]_+ = \frac{1}{2}\left[g_A(z) - s\right]_+.$$

Applying Hoffman's theorem to the lifted linear system gives a point $(x', y', v') \in \mathcal{Q}_s(A)$ satisfying

$$\|(x,y,v_s(z)) - (x',y',v')\|_\sharp \le \frac{1}{2} H_{\mathrm{lev}}(A)\left[g_A(z) - s\right]_+.$$

Since $(x', y') \in \mathcal{E}_s(A)$, projecting and using (J.33) proves (J.34). □

Define the inner and outer corridor sets

$$\underline{\mathcal{E}}(\widehat{A};\tau,r) := \mathcal{E}_{\tau-2r}(\widehat{A}), \qquad \overline{\mathcal{E}}(\widehat{A};\tau,r) := \mathcal{E}_{\tau+2r}(\widehat{A}). \tag{J.36}$$

When $\tau \geq 2r$, both sets are nonempty and the inner set is contained in the outer set. Define the corresponding collection of admissible correspondences

$$\mathfrak{C}(\widehat{A};\tau,r) := \left\{ C \in \mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z}) : \underline{\mathcal{E}}(\widehat{A};\tau,r) \subseteq C \subseteq \overline{\mathcal{E}}(\widehat{A};\tau,r) \right\}. \tag{J.37}$$

Proposition J.11 (Corridor width)**.** *If $\tau \geq 2r$, then*

$$d_{\mathrm{H}}\left(\underline{\mathcal{E}}(\widehat{A};\tau,r), \overline{\mathcal{E}}(\widehat{A};\tau,r)\right) \leq 4\kappa_{\mathrm{lev}}(\widehat{A})r, \tag{J.38}$$

$$\sup_{C,D\in\mathfrak{C}(\widehat{A};\tau,r)} d_{\mathrm{H}}(C,D) \leq 4\kappa_{\mathrm{lev}}(\widehat{A})r. \tag{J.39}$$

On the event (J.21)*,*

$$\mathcal{E}_\tau(A) \in \mathfrak{C}(\widehat{A};\tau,r). \tag{J.40}$$

Proof. The inner set is contained in the outer set. For every $z \in \overline{\mathcal{E}}(\widehat{A};\tau,r)$,

$$g_{\widehat{A}}(z) \leq \tau + 2r.$$

Applying Proposition J.10 with $s = \tau - 2r$ gives

$$\mathrm{dist}\left(z, \underline{\mathcal{E}}(\widehat{A};\tau,r)\right) \leq 4\kappa_{\mathrm{lev}}(\widehat{A})r.$$

Taking the supremum over the outer set proves (J.38).

Now let $C, D \in \mathfrak{C}(\widehat{A};\tau,r)$. Since the inner set is contained in D, every $c \in C \subseteq \overline{\mathcal{E}}$ satisfies

$$\mathrm{dist}(c,D) \leq \mathrm{dist}\left(c, \underline{\mathcal{E}}\right) \leq 4\kappa_{\mathrm{lev}}(\widehat{A})r.$$

The same argument with C and D interchanged proves (J.39). Finally, (J.40) is exactly the sandwich in Proposition J.7. □

Remark J.12 (Operational status of the width bound). The coverage statement depends only on the payoff-confidence event. The width certificate is data dependent through $\kappa_{\mathrm{lev}}(\widehat{A})$. For rational input, the lifted coefficient matrix in (J.31) is explicit, so finite algorithms for Hoffman constants can compute the exact constant or a computable upper bound (Peña et al., 2021). The constant may be large near an ill-conditioned game; no uniform conditioning claim is made. Thus the width bound is directly evaluable for the observed system but may be conservative near degeneracy.

Corollary J.13 (Critical-scale confidence corridor)**.** *Let $N_{a,T} = Tq_{a,T}$ under a deterministic full-support design and define*

$$q_{\min,T} := \min_{a\in[d]} q_{a,T}, \qquad \rho_T(\delta) := \sqrt{T}\, r_T(\delta) = \sigma\sqrt{\frac{2\log(2d/\delta)}{q_{\min,T}}}. \tag{J.41}$$

Then, with probability at least $1-\delta$, simultaneously for every $\lambda \geq 0$,

$$\mathcal{E}_{(\lambda-2\rho_T(\delta))/\sqrt{T}}(\widehat{A}_T) \subseteq \mathcal{E}_{\lambda/\sqrt{T}}(A) \subseteq \mathcal{E}_{(\lambda+2\rho_T(\delta))/\sqrt{T}}(\widehat{A}_T). \tag{J.42}$$

The inner set is empty when its displayed resolution is negative.

Proof. Substitute $\tau = \lambda/\sqrt{T}$ and $r = r_T(\delta) = \rho_T(\delta)/\sqrt{T}$ into (J.30). □

The final result identifies the random-set limit of this finite-sample corridor. Let $q_T \to q \in \mathrm{ri}\,\Delta_d$, set

$$\rho_q(\delta) := \sigma\sqrt{\frac{2\log(2d/\delta)}{q_{\min}}}, \qquad q_{\min} := \min_{a\in[d]} q_a, \tag{J.43}$$

and, under the local game $A_{T,H} = A_\star + H/\sqrt{T}$, define the matrix-valued statistic

$$\widehat{H}_T := \sqrt{T}(\widehat{A}_T - A_\star). \tag{J.44}$$

Let $W_H \in \mathbb{R}^{m\times n}$ be the Gaussian random matrix satisfying

$$\mathrm{vec}(W_H) \sim \mathcal{N}\left(\mathrm{vec}(H), \Sigma(q)\right). \tag{J.45}$$

Proposition J.14 (Gaussian limit of the critical confidence corridor). *Fix $H \in \mathbb{R}^{m\times n}$, $\delta \in (0,1)$, and*

$$\lambda > 2\rho_q(\delta). \tag{J.46}$$

Under the local law generated by $A_{T,H}$, define

$$\underline{\mathcal{E}}_T := \mathcal{E}_{(\lambda-2\rho_T(\delta))/\sqrt{T}}(\widehat{A}_T), \tag{J.47}$$

$$\overline{\mathcal{E}}_T := \mathcal{E}_{(\lambda+2\rho_T(\delta))/\sqrt{T}}(\widehat{A}_T). \tag{J.48}$$

Then, in the product Hausdorff space,

$$\left(\underline{\mathcal{E}}_T, \overline{\mathcal{E}}_T\right) \xrightarrow{\mathrm{d}} \left(\mathcal{K}_{\lambda-2\rho_q(\delta)}(W_H), \mathcal{K}_{\lambda+2\rho_q(\delta)}(W_H)\right). \tag{J.49}$$

Moreover,

$$\mathbb{P}\Big(\mathcal{K}_{\lambda-2\rho_q(\delta)}(W_H) \subseteq \mathcal{K}_\lambda(H) \subseteq \mathcal{K}_{\lambda+2\rho_q(\delta)}(W_H)\Big) \geq 1-\delta. \tag{J.50}$$

Proof. Under the local law,

$$\mathrm{vec}(\widehat{H}_T) \sim \mathcal{N}\left(\mathrm{vec}(H), \Sigma(q_T)\right),$$

so $\widehat{H}_T \xrightarrow{\mathrm{d}} W_H$, and $\rho_T(\delta) \to \rho_q(\delta)$. The sequence $(\widehat{H}_T)$ is tight. On every compact set of directions, the uniform strategic epi-resolution theorem applies simultaneously to the two resolution levels, which remain bounded away from zero by (J.46). Therefore,

$$d_{\mathrm{H}}\left(\underline{\mathcal{E}}_T, \mathcal{K}_{\lambda-2\rho_T(\delta)}(\widehat{H}_T)\right) \xrightarrow{\mathbb{P}} 0,$$

$$d_{\mathrm{H}}\left(\overline{\mathcal{E}}_T, \mathcal{K}_{\lambda+2\rho_T(\delta)}(\widehat{H}_T)\right) \xrightarrow{\mathbb{P}} 0.$$

The continuity of $(G,s) \mapsto \mathcal{K}_s(G)$ on positive levels, proved in Proposition F.8, and the continuous mapping theorem now give (J.49).

For the coverage statement, the Gaussian union bound used in Proposition J.9 gives

$$\mathbb{P}\left(\|W_H - H\|_\infty \leq \rho_q(\delta)\right) \geq 1-\delta. \tag{J.51}$$

On this event, the direction-Lipschitz bound (F.13), with the strategy fixed, gives

$$|\beta_{A_\star}(z; W_H) - \beta_{A_\star}(z; H)| \leq 2\rho_q(\delta) \qquad \text{for every } z \in \mathcal{E}_\star.$$

The two set inclusions in (J.50) follow immediately. □

Remark J.15 (Statistical meaning of corridor width). At critical resolution, the normalized confidence half-width $\rho_T(\delta)$ converges to a nonzero constant, and the limiting corridor in (J.49) generally retains nontrivial branch uncertainty. In the supercritical regime $\sqrt{T}\,\tau_T \to \infty$, the same payoff uncertainty is negligible relative to the requested approximation tolerance, and Proposition J.5 forces the entire target to collapse to $\mathcal{E}_0(A_\star)$. The confidence corridor therefore exhibits the same statistical resolution transition as the minimax risk.

K Exact-Set Honest Nonadaptivity

This appendix proves that exact equilibrium-correspondence inference cannot adapt honestly across a strategic bifurcation. The obstruction is not the spatial diameter of a nonunique equilibrium set: even if the true correspondence were known exactly, that set could have positive diameter. Accordingly, uncertainty is measured in the *second-order hyperspace* of candidate equilibrium correspondences. We show that every confidence collection that is honest over a statistically indistinguishable local neighborhood of the bifurcation must contain mutually distant candidate correspondences with nonvanishing probability. The proof is finite-sample, permits arbitrary adaptive payoff queries, and uses only total-variation transport of coverage events. It is analogous in spirit to classical no-free-adaptation results for honest confidence procedures (Low, 1997), but the mechanism here is a discontinuous, set-valued strategic parameter rather than a nonparametric smoothness scale.

K.1 Confidence collections on the equilibrium hyperspace

Let

$$\mathsf{S} := \mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z}) \tag{K.1}$$

be the compact metric space of nonempty compact convex subsets of $\mathcal{Z}$, equipped with the Hausdorff metric d_{H}. A confidence procedure for the complete equilibrium correspondence is a random element

$$\mathfrak{C}_T \in \mathfrak{K}(\mathsf{S}). \tag{K.2}$$

Thus, $\mathfrak{C}_T$ is a nonempty compact collection of candidate compact-convex equilibrium sets. Its uncertainty diameter is

$$\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T) := \sup_{C,D\in\mathfrak{C}_T} d_{\mathrm{H}}(C,D). \tag{K.3}$$

All external randomization is understood to be parameter independent and is included in the underlying observation space.

Lemma K.1 (Measurability and continuity of hyperspace diameter)**.** *For $\mathfrak{C},\mathfrak{D} \in \mathfrak{K}(\mathsf{S})$, let $d_{\mathrm{H},\mathsf{S}}$ denote the Hausdorff metric induced by d_{H} on S. Then*

$$|\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}) - \mathrm{diam}_{\mathrm{H}}(\mathfrak{D})| \le 2d_{\mathrm{H},\mathsf{S}}(\mathfrak{C},\mathfrak{D}). \tag{K.4}$$

In particular, $\mathrm{diam}_{\mathrm{H}}$ is Borel measurable. Moreover, for every fixed $S \in \mathsf{S}$, the membership event

$$\{\mathfrak{C} \in \mathfrak{K}(\mathsf{S}) : S \in \mathfrak{C}\} \tag{K.5}$$

is closed, and hence Borel, in $\mathfrak{K}(\mathsf{S})$.

Proof. Let $\varepsilon > d_{\mathrm{H},\mathsf{S}}(\mathfrak{C},\mathfrak{D})$. For arbitrary $C_1, C_2 \in \mathfrak{C}$, choose $D_1, D_2 \in \mathfrak{D}$ satisfying $d_{\mathrm{H}}(C_r, D_r) < \varepsilon$ for $r \in \{1,2\}$. The triangle inequality gives

$$d_{\mathrm{H}}(C_1,C_2) \le d_{\mathrm{H}}(D_1,D_2) + 2\varepsilon \le \mathrm{diam}_{\mathrm{H}}(\mathfrak{D}) + 2\varepsilon.$$

Taking the supremum over C_1, C_2, then letting $\varepsilon \downarrow d_{\mathrm{H},\mathsf{S}}(\mathfrak{C},\mathfrak{D})$, yields

$$\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}) \le \mathrm{diam}_{\mathrm{H}}(\mathfrak{D}) + 2d_{\mathrm{H},\mathsf{S}}(\mathfrak{C},\mathfrak{D}).$$

Interchanging $\mathfrak{C}$ and $\mathfrak{D}$ proves (K.4).

For fixed $S \in \mathsf{S}$, the map

$$\mathfrak{C} \longmapsto \mathrm{dist}_{\mathsf{S}}(S,\mathfrak{C}) := \inf_{C\in\mathfrak{C}} d_{\mathrm{H}}(S,C)$$

is 1-Lipschitz under $d_{\mathrm{H},\mathsf{S}}$. Since

$$S \in \mathfrak{C} \quad\Longleftrightarrow\quad \mathrm{dist}_{\mathsf{S}}(S,\mathfrak{C}) = 0,$$

the membership set in (K.5) is closed. □

Definition K.2 (Honest confidence collection)**.** Let $\{P_{T,\vartheta} : \vartheta \in \Theta_T\}$ be a statistical experiment and let $S_T(\vartheta) \in \mathsf{S}$ be its set-valued target. A random collection $\mathfrak{C}_T \in \mathfrak{K}(\mathsf{S})$ is $(1-\alpha)$-honest over Θ_T if

$$\inf_{\vartheta\in\Theta_T} P_{T,\vartheta}\left(S_T(\vartheta) \in \mathfrak{C}_T\right) \ge 1-\alpha. \tag{K.6}$$

Remark K.3 (Why a second-order confidence object is necessary). An outer region $U_T \subseteq \mathcal{Z}$ satisfying $\mathcal{E}_0(A) \subseteq U_T$ measures both epistemic uncertainty and the intrinsic spatial extent of the true equilibrium set. At the guiding bifurcation, $\mathcal{E}_0(A_0)$ itself has product-ℓ_1 diameter 2, even if A_0 is known. By contrast, the singleton confidence collection $\{\mathcal{E}_0(A_0)\}$ has hyperspace diameter zero. The quantity $\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T)$ therefore measures uncertainty about *which equilibrium correspondence is true*, rather than multiplicity within a known correspondence.

K.2 Separated statistically indistinguishable alternatives

The next lemma is the basic transport principle. It is stated for an arbitrary metric-valued target and therefore applies beyond games.

Lemma K.4 (Transport of honest coverage)**.** *Let (A, d) be a metric space, let $a_0, a_1 \in \mathsf{A}$, and let $P_\star, P_0, P_1$ be probability measures on a common measurable space. Let $\mathfrak{C}$ be a random nonempty compact subset of A. Suppose*

$$P_r(a_r \in \mathfrak{C}) \geq 1 - \alpha_r, \qquad r \in \{0, 1\}. \tag{K.7}$$

Then

$$P_\star\left(\mathrm{diam}_d(\mathfrak{C}) \geq d(a_0, a_1)\right) \geq \left[1 - \alpha_0 - \alpha_1 - d_{\mathrm{TV}}(P_\star, P_0) - d_{\mathrm{TV}}(P_\star, P_1)\right]_+, \tag{K.8}$$

$$\mathbb{E}_{P_\star}\left[\mathrm{diam}_d(\mathfrak{C})\right] \geq d(a_0, a_1)\left[1 - \alpha_0 - \alpha_1 - d_{\mathrm{TV}}(P_\star, P_0) - d_{\mathrm{TV}}(P_\star, P_1)\right]_+. \tag{K.9}$$

Proof. Let $E_r := \{a_r \in \mathfrak{C}\}$. By the definition of total variation,

$$P_\star(E_r^c) \leq P_r(E_r^c) + d_{\mathrm{TV}}(P_\star, P_r) \leq \alpha_r + d_{\mathrm{TV}}(P_\star, P_r).$$

Hence, by the union bound,

$$\begin{aligned} P_\star(E_0 \cap E_1) &\geq 1 - P_\star(E_0^c) - P_\star(E_1^c) \\ &\geq 1 - \alpha_0 - \alpha_1 - d_{\mathrm{TV}}(P_\star, P_0) - d_{\mathrm{TV}}(P_\star, P_1). \end{aligned}$$

On $E_0 \cap E_1$, the collection $\mathfrak{C}$ contains both a_0 and a_1, so its diameter is at least $d(a_0, a_1)$. This proves (K.8); multiplying the event indicator by $d(a_0, a_1)$ gives (K.9). □

Definition K.5 (Statistically invisible bifurcation)**.** Consider a sequence of experiments $\{P_{T,\vartheta} : \vartheta \in \Theta_T\}$ with targets $S_T(\vartheta) \in (\mathsf{A}, d)$, and a reference law $P_{T,\star}$. We say that the target has a Δ*-statistically invisible bifurcation* at $P_{T,\star}$ if $\Delta > 0$ and, for every $\varepsilon > 0$, there exist alternatives $\vartheta_{T,0}^{(\varepsilon)}, \vartheta_{T,1}^{(\varepsilon)} \in \Theta_T$ such that

$$\limsup_{T\to\infty} \max_{r\in\{0,1\}} d_{\mathrm{TV}}\left(P_{T,\star}, P_{T,\vartheta_{T,r}^{(\varepsilon)}}\right) \leq \varepsilon, \tag{K.10}$$

$$\liminf_{T\to\infty} d\left(S_T(\vartheta_{T,0}^{(\varepsilon)}), S_T(\vartheta_{T,1}^{(\varepsilon)})\right) \geq \Delta. \tag{K.11}$$

Proposition K.6 (Generic no-free-adaptation principle)**.** *Suppose the target has a Δ-statistically invisible bifurcation at $P_{T,\star}$, and let $\mathfrak{C}_T$ be $(1-\alpha)$-honest over Θ_T, with $0 < \alpha < 1/2$. Then, for every $0 < \eta < \Delta$,*

$$\liminf_{T\to\infty} P_{T,\star}\left(\mathrm{diam}_d(\mathfrak{C}_T) \geq \Delta - \eta\right) \geq 1 - 2\alpha, \tag{K.12}$$

and

$$\liminf_{T\to\infty} \mathbb{E}_{T,\star}\left[\mathrm{diam}_d(\mathfrak{C}_T)\right] \geq \Delta(1 - 2\alpha). \tag{K.13}$$

Proof. Fix $\varepsilon > 0$. By (K.11), the two target values are separated by at least $\Delta - \eta$ for all sufficiently large T. Applying Lemma K.4 with $\alpha_0 = \alpha_1 = \alpha$ and using (K.10) gives

$$\liminf_{T\to\infty} P_{T,\star}\left(\mathrm{diam}_d(\mathfrak{C}_T) \geq \Delta - \eta\right) \geq 1 - 2\alpha - 2\varepsilon.$$

Letting $\varepsilon \downarrow 0$ proves (K.12). The expectation bound follows first with $\Delta - \eta$ in place of Δ, and then by letting $\eta \downarrow 0$. □

K.3 The hyperspace-diameter lower bound

We now specialize the generic principle to the guiding bifurcation. Recall

$$A_\theta = \begin{pmatrix} 1 & 0 \\ 0 & \theta \end{pmatrix}, \qquad \theta \in [-1, 1], \tag{K.14}$$

and let $P_{T,\theta}^{\mathsf{A}}$ be the transcript law induced by an arbitrary randomized adaptive T-query strategy A. For $0 < \bar{c} \leq 1$, define the shrinking local neighborhood

$$\Theta_T(\bar{c}) := \left\{\theta \in \mathbb{R} : |\theta| \leq \frac{\bar{c}}{\sqrt{T}}\right\}. \tag{K.15}$$

Proposition K.7 (Finite-sample impossibility at the bifurcation)**.** *Fix $T \geq 1$, $\sigma > 0$, $0 < \bar{c} \leq 1$, and $0 < \alpha < 1/2$. Let $\mathfrak{C}_T \in \mathfrak{K}(\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z}))$ be any confidence collection satisfying*

$$\inf_{\theta \in \Theta_T(\bar{c})} P^{\mathsf{A}}_{T,\theta}\left(\mathcal{E}_0(A_\theta) \in \mathfrak{C}_T\right) \geq 1 - \alpha. \tag{K.16}$$

Then, under the bifurcation game A_0,

$$P^{\mathsf{A}}_{T,0}\left(\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T) \geq 2\right) \geq 1 - 2\alpha, \tag{K.17}$$

$$\mathbb{E}^{\mathsf{A}}_{T,0}\left[\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T)\right] \geq 2(1 - 2\alpha). \tag{K.18}$$

These inequalities hold for every adaptive sampling strategy, including the information-maximizing strategy that queries entry $(2, 2)$ at every round.

Proof. Fix an arbitrary $c \in (0, \bar{c}]$, and set

$$\theta_{T,+} := \frac{c}{\sqrt{T}}, \qquad \theta_{T,-} := -\frac{c}{\sqrt{T}}.$$

By Proposition B.3,

$$d_{\mathrm{H}}\left(\mathcal{E}_0(A_{\theta_{T,+}}), \mathcal{E}_0(A_{\theta_{T,-}})\right) = 2. \tag{K.19}$$

Corollary B.7 and its symmetric negative-direction counterpart give

$$d_{\mathrm{TV}}\left(P^{\mathsf{A}}_{T,0}, P^{\mathsf{A}}_{T,\theta_{T,+}}\right) \leq \frac{c}{2\sigma}, \qquad d_{\mathrm{TV}}\left(P^{\mathsf{A}}_{T,0}, P^{\mathsf{A}}_{T,\theta_{T,-}}\right) \leq \frac{c}{2\sigma}. \tag{K.20}$$

Applying Lemma K.4 in the metric space $(\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z}), d_{\mathrm{H}})$ yields

$$P^{\mathsf{A}}_{T,0}\left(\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T) \geq 2\right) \geq 1 - 2\alpha - \frac{c}{\sigma}, \tag{K.21}$$

$$\mathbb{E}^{\mathsf{A}}_{T,0}\left[\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T)\right] \geq 2\left(1 - 2\alpha - \frac{c}{\sigma}\right)_+. \tag{K.22}$$

The left-hand sides do not depend on c. Since (K.21) holds for every $c \in (0, \bar{c}]$, letting $c \downarrow 0$ proves (K.17). The same limiting argument in (K.22) proves (K.18). □

Corollary K.8 (Honest Hausdorff balls have order-one radius)**.** *Under the assumptions of Proposition K.7, suppose $\widehat{C}_T \in \mathsf{S}$ and $r_T \geq 0$ are measurable and*

$$\mathfrak{C}_T = \left\{C \in \mathsf{S} : d_{\mathrm{H}}(C, \widehat{C}_T) \leq r_T\right\}. \tag{K.23}$$

Then, under A_0,

$$P^{\mathsf{A}}_{T,0}(r_T \geq 1) \geq 1 - 2\alpha, \qquad \mathbb{E}^{\mathsf{A}}_{T,0}[r_T] \geq 1 - 2\alpha. \tag{K.24}$$

Proof. For every realization, the triangle inequality gives $\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T) \leq 2r_T$. Hence $\{\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T) \geq 2\} \subseteq \{r_T \geq 1\}$, and the probability bound follows from (K.17). Finally $\mathbb{E}r_T \geq P(r_T \geq 1)$. □

Corollary K.9 (No shrinking honest confidence collection)**.** *Under the assumptions of Proposition K.7, no sequence of honest confidence collections can satisfy*

$$\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T) \xrightarrow[P^{\mathsf{A}}_{T,0}]{} 0. \tag{K.25}$$

More precisely, for every deterministic sequence $r_T < 2$,

$$\inf_{T \geq 1} P^{\mathsf{A}}_{T,0}\left(\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T) > r_T\right) \geq 1 - 2\alpha. \tag{K.26}$$

If the miscoverage levels satisfy $\alpha_T \to 0$, then

$$P^{\mathsf{A}}_{T,0}\left(\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T) \geq 2\right) \longrightarrow 1. \tag{K.27}$$

Proof. The first two claims follow directly from (K.17). Applying the same proposition at level α_T yields (K.27). □

Remark K.10 (The lower bound is stronger than a testing-rate statement). For each fixed nonzero c, a two-point argument gives the weaker bound $1 - 2\alpha - c/\sigma$. Uniform honesty over the entire $T^{-1/2}$-neighborhood allows $c \downarrow 0$ while the two exact equilibrium correspondences remain Hausdorff-separated by 2. This removes the testing penalty completely and yields the testing-penalty-free confidence obstruction $1 - 2\alpha$. The phenomenon is therefore driven by a statistically invisible discontinuity of the exact set-valued target.

K.4 Comparison with the strongly regular regime

The preceding impossibility is specific to a neighborhood crossing a game at which the exact correspondence is not locally Hausdorff continuous. We next show that ordinary shrinking confidence collections are available on a locally Hausdorff-Lipschitz stratum.

Assumption K.11 (Complete-correspondence regularity)**.** A game $A^\circ \in \mathbb{R}^{m\times n}$ is (ρ, L)-regular for the complete equilibrium correspondence if $\rho, L > 0$ and

$$d_{\mathrm{H}}\left(\mathcal{E}_0(B), \mathcal{E}_0(B')\right) \le L\|B - B'\|_\infty \tag{K.28}$$

holds for every $B, B' \in \mathbb{B}_\infty(A^\circ, \rho)$, where

$$\mathbb{B}_\infty(A, r) := \left\{B \in \mathbb{R}^{m\times n} : \|B - A\|_\infty \le r\right\}. \tag{K.29}$$

Strong regularity of the equilibrium generalized equation yields a locally single-valued Lipschitz localization; to obtain Assumption K.11 for the *complete* correspondence, that local branch must additionally exhaust all equilibria in the neighborhood. Stability and differentiable single-valued localizations for Nash KKT mappings are studied by Diao et al. (2025).

Proposition K.12 (Shrinking honest collections on a regular stratum)**.** *Assume A° is (ρ, L)-regular. Let $\widehat{A}_T$ be a measurable payoff estimator and let $r_T > 0$ satisfy*

$$\inf_{A\in\mathbb{B}_\infty(A^\circ,\rho/2)} P_A\left(\|\widehat{A}_T - A\|_\infty \le r_T\right) \ge 1 - \alpha, \qquad r_T \le \frac{\rho}{2}. \tag{K.30}$$

Define the measurable center

$$\widehat{S}_T := \begin{cases} \mathcal{E}_0(\widehat{A}_T), & \widehat{A}_T \in \mathbb{B}_\infty(A^\circ, \rho), \\ \mathcal{E}_0(A^\circ), & \textit{otherwise}, \end{cases} \tag{K.31}$$

and the confidence collection

$$\mathfrak{C}_T^{\mathrm{reg}} := \left\{C \in \mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z}) : d_{\mathrm{H}}(C, \widehat{S}_T) \le L r_T\right\}. \tag{K.32}$$

Then $\mathfrak{C}_T^{\mathrm{reg}}$ is a random nonempty compact collection and

$$\inf_{A\in\mathbb{B}_\infty(A^\circ,\rho/2)} P_A\left(\mathcal{E}_0(A) \in \mathfrak{C}_T^{\mathrm{reg}}\right) \ge 1 - \alpha, \tag{K.33}$$

$$\mathrm{diam}_{\mathrm{H}}(\mathfrak{C}_T^{\mathrm{reg}}) \le 2Lr_T \qquad \textit{almost surely}. \tag{K.34}$$

Proof. On the closed ball $\mathbb{B}_\infty(A^\circ, \rho)$, the map $B \mapsto \mathcal{E}_0(B)$ is Lipschitz by assumption and hence continuous. The piecewise definition (K.31) is therefore Borel measurable. Since $\mathfrak{K}_{\mathrm{cvx}}(\mathcal{Z})$ is compact, the closed metric ball in (K.32) is a nonempty compact subset of that space. Its graph as a function of the center is closed because d_{H} is continuous; standard compact-valued measurability, or the distance characterization in Proposition A.11, shows that $\mathfrak{C}_T^{\mathrm{reg}}$ is a random compact collection.

Fix $A \in \mathbb{B}_\infty(A^\circ, \rho/2)$ and consider the event $\|\widehat{A}_T - A\|_\infty \le r_T$. Because $r_T \le \rho/2$, this event implies $\widehat{A}_T \in \mathbb{B}_\infty(A^\circ, \rho)$. Assumption K.11 then gives

$$d_{\mathrm{H}}\left(\mathcal{E}_0(A), \widehat{S}_T\right) = d_{\mathrm{H}}\left(\mathcal{E}_0(A), \mathcal{E}_0(\widehat{A}_T)\right) \le L r_T.$$

Thus the true correspondence belongs to $\mathfrak{C}_T^{\mathrm{reg}}$ on the payoff-confidence event, proving (K.33).

Finally, for arbitrary $C, D \in \mathfrak{C}_T^{\mathrm{reg}}$, the triangle inequality yields

$$d_{\mathrm{H}}(C, D) \le d_{\mathrm{H}}(C, \widehat{S}_T) + d_{\mathrm{H}}(\widehat{S}_T, D) \le 2Lr_T.$$

Taking the supremum proves (K.34). □

Corollary K.13 (Root-T confidence diameter under Gaussian sampling)**.** *Under deterministic Gaussian payoff sampling, let $q_{\min,T} := \min_{a\in[d]} q_{a,T}$, and set*

$$r_T(\alpha) := \sigma\sqrt{\frac{2\log(2d/\alpha)}{T q_{\min,T}}}. \tag{K.35}$$

If A° is (ρ, L)-regular and $r_T(\alpha) \le \rho/2$, then the collection in (K.32) *is uniformly $(1-\alpha)$-honest over $\mathbb{B}_\infty(A^\circ, \rho/2)$ and satisfies*

$$\operatorname{diam}_{\mathrm{H}}(\mathfrak{C}_T^{\mathrm{reg}}) \le 2L\sigma\sqrt{\frac{2\log(2d/\alpha)}{T q_{\min,T}}} \qquad \textit{almost surely.} \tag{K.36}$$

In particular, if $\inf_T q_{\min,T} > 0$, the confidence diameter is $O(T^{-1/2})$ for fixed α.

Proof. The simultaneous Gaussian payoff bound in Proposition J.9 gives (K.30) with radius (K.35), uniformly over the true payoff matrix. Proposition K.12 then applies. □

Remark K.14 (Exact resolution is the source of nonadaptivity). On a regular stratum, the exact correspondence admits confidence collections with root-T hyperspace diameter. At the bifurcation, every collection honest over an arbitrarily small $T^{-1/2}$-neighborhood has order-one diameter with probability at least $1 - 2\alpha$. Positive first-order resolution removes the discontinuity: Appendix F proves continuity of $(H, \lambda) \mapsto \mathcal{K}_\lambda(H)$ for $\lambda > 0$, and Appendix J constructs honest resolution-dependent confidence corridors. Thus the obstruction is not to inference per se, but to demanding exact-set honesty below the statistically resolvable branch scale.

L Proof Dependency and Verification Notes

This appendix records the proof dependencies and the supplementary counterexample-search suite. It does not repeat the proofs in Appendices B–K, and the computational checks are not formal proofs.

L.1 Headline claim map

Result	Proof blocks	Delicate obligations
Proposition 1	A–B	product-ℓ_1 normalization; adaptive KL orientation; randomized metric two-point bound
Theorem 2	A, C–F	directed fixed-base Hoffman bound; unbounded velocities; complete-sequence recovery; positive-level density; compact uniformization
Theorem 3	A, D–E	active-set signs; Sion orientation; certificate sets; affine inverse; attainment
Theorem 4	A, F–I	exact reverse Gaussian kernel; covariance comparison; loss diameter four; barycentric measurability; finite-net projection
Corollary 5	C, F–J	separated singleton condition; varying positive levels; both supercritical conditions
Theorem 6	B, K	second-order measurability; coverage transport; the $c \downarrow 0$ argument

L.2 Supplementary regression tests

The anonymous directory `verification/` contains deterministic tests for: the exact 2×2 residual and equilibrium geometry; tangent-LP versus restricted-game duality on zero, unique, duplicated-action, and face-valued games; affine contrast marginals and inverse; positive-resolution convergence and exact-resolution inner-limit failure; Gaussian KL, total variation, loss diameters, and reverse-kernel covariance; support-function barycenter contraction; and static labels, references, citations, anonymity, and post-build rendered-PDF text checks. Exact or symbolic arithmetic is used for the guiding game and affine identities, with floating-point LP tests reserved for larger examples. The suite uses fixed seeds and explicit tolerances and exits nonzero on failure. These programs are counterexample searches and consistency checks; the universal statements are proved in the appendices.

L.3 Scope boundaries

The retained results do not claim an adaptive matching upper theorem; ω_T is not assigned a rate; exact resolution is not described as complete Hausdorff convergence; the supercritical result is local

Table 3: Selected closest neighboring theories.

Work	Experiment	Estimand and loss	Difference here
Mills; Altman et al. (Mills, 1956; Altman et al., 2001)	Deterministic linear perturbation	Value derivative and limits of exact optimal strategies	Profile-wise residual derivative, complete positive-resolution set limit, and statistical composition
Viqueira et al.; Jiang–Zhang (Areyan Viqueira et al., 2019; 2020; Jiang & Zhang, 2026)	Noisy samples from one fixed game	All-equilibria PAC precision–recall guarantees, or one selected equilibrium estimate	Contiguous bifurcation, exact-set Hausdorff loss, and $T^{-1/2}$ branch resolution
Liu et al.; Klatt et al. (Liu et al., 2023; Klatt et al., 2022)	Random LP objectives or constraints	Entire optimizer set; weak limits and confidence sets	Joint primal–dual saddle set and rank-two strategic contrast geometry
Meitz; Meitz–Shapiro (Meitz, 2024; Meitz & Shapiro, 2025)	Sample minimax objective	Complete solution set, values, or selected solutions	Payoff-coordinate local experiment and a resolution-dependent target
Maiti et al.; Feinstein et al. (Maiti et al., 2025; Feinstein et al., 2024)	Exact queries or deterministic convex-game access	Exact recovery or inner–outer approximation of all equilibria	Noisy local alternatives and tolerance $\lambda/\sqrt{T}$ as branch resolution
This paper	Gaussian coordinates under $A_\star + H/\sqrt{T}$	Complete visible correspondence under product-ℓ_1 Hausdorff loss	Affine-copy critical contrasts and the resolution trichotomy

oracle; and no arbitrary general-sum theorem is advertised. The supplementary static check verifies theorem numbering, cross-references, bibliography keys, anonymity-sensitive source strings, and malformed compounds in extracted PDF text.

M Extended Related Work and Boundary Cases

The comparison below separates the statistical experiment, estimand, perturbation regime, loss, and structural conclusion. We do not claim novelty for epi-derivatives, Hoffman bounds, LP duality, Gaussian sufficiency, or set-valued inference in isolation.

M.1 Theorem-level comparison

Classical perturbation boundary. Classical sensitivity theory proves

$$\frac{d}{dt^+}\operatorname{val}(A_\star + tH)\Big|_{t=0} = \operatorname{val}_{\mathcal{X}_\star,\mathcal{Y}_\star}(H), \tag{M.1}$$

and constrains accumulation points of exact perturbed optimal strategies by restricted or lexicographic perturbation games (Mills, 1956; Altman et al., 2001). The present theorem instead identifies the complete profile-wise field

$$\beta_\star((x,y);H) = \max_{\mu\in\mathcal{X}_\star} \mu^\top H y - \min_{\nu\in\mathcal{Y}_\star} x^\top H \nu, \tag{M.2}$$

and proves, uniformly for positive resolution,

$$\mathcal{E}_{t\lambda}(A_\star + tH) \xrightarrow[t\downarrow 0]{d_{\mathrm{H}}} \mathcal{K}_\lambda(H). \tag{M.3}$$

The restricted scalar game is classical; the profile-wise residual derivative, affine-copy contrast geometry, complete positive-resolution set limit, and statistical composition are the new statements.

Random LP and minimax inference. Random-LP work provides weak limits for the entire nonunique optimizer set and computational confidence sets through auxiliary LPs; general minimax inference provides Hausdorff consistency and confidence sets for multiple solutions (Liu et al., 2023; Klatt et al., 2022; Meitz, 2024). The 2025 corrigendum (Meitz, 2025) strengthens assumptions behind a directional- differentiability statement; none of our derivative claims relies on the uncorrected result. The additional structure here is explicit: the target is the joint primal–dual saddle correspondence; each profile has its own residual threshold; the derivative is supported by rank-two matrices $C =$

$\mu y^\top - x\nu^\top$; payoff coordinates are sampled directly; tolerance is the statistical branch-resolution parameter; and loss is Hausdorff loss on strategy correspondences. The contrast marginals recover both equilibrium strategies, which yields the affine-copy theorem rather than only a generic optimizer-set limit.

Regular and inverse game inference. Stochastic-VI confidence theory and Nash stability study regular or selected solutions (Lu & Budhiraja, 2013; Diao et al., 2025); recent contextual matrix-game work targets payoff parameters, selected equilibria, values, and regret (Xiang et al., 2026). Inverse-game work estimates payoff sets compatible with observed behavior (Barbara et al., 2026). Our map is forward, from noisy payoffs to a complete singular correspondence.

Robustness and nonregular decision theory. Approximate-equilibrium persistence under perturbed games is known qualitatively (Jackson et al., 2012). The branch cost is the exact first-order profile-specific threshold and is composed with a statistical experiment. The decision-theoretic lower-bound machinery is classical; the nonregular compact-set target is game specific (Song, 2014; Takatsu & Kuchibhotla, 2024).

M.2 Boundary cases

Proposition M.1 (Singleton base correspondence)**.** *If $\mathcal{E}_\star = \{z_\star\}$, then for every direction H and every $\lambda \geq 0$,*

$$\mathcal{C}_\star(z_\star) = \{0\}, \qquad \beta_\star(z_\star; H) = 0, \qquad \mathcal{K}_\lambda(H) = \{z_\star\}. \tag{M.4}$$

Every local approximate-equilibrium correspondence with vanishing tolerance converges to $\{z_\star\}$.

Proof. Theorem 3 makes the contrast polytope an affine copy of the singleton base correspondence, hence $\{0\}$. The support function and visible-set identities follow. The positive-resolution conclusion uses Theorem 2; the singleton exact-resolution clause gives the zero-level statement. □

This first-order conclusion does not imply a differentiable or Lipschitz selected equilibrium map. Singular unique branches may move at a fractional-power scale.

Proposition M.2 (Zero resolution is the topological boundary)**.** *For the guiding direction $H_h = he_2e_2^\top$, $\mathcal{K}_0(H_h)$ equals the left endpoint for $h > 0$, the full segment for $h = 0$, and the right endpoint for $h < 0$. Each one-sided limit is at Hausdorff distance two from $\mathcal{K}_0(0)$. For every fixed $\lambda > 0$, however, $h \mapsto \mathcal{K}_\lambda(H_h)$ is Hausdorff continuous and equals the full segment whenever $|h| \leq \lambda$.*

Proof. Equation (3.18) gives $\beta_\star(z_p; H_h) = h_+p + h_-(1-p)$. Taking its zero and positive sublevels yields the stated sets; the product-ℓ_1 distance from either endpoint singleton to the full segment is two. □

The comparison supports the stated theorem-level distinction but is not a formal proof of global priority.