The Feedback-Operator Theorem: Identifying Welfare Orders in Reinforcement Learning from World Feedback

Manoj Saravanan*

Abstract

We study reinforcement learning from world feedback through a consequence-space formulation in which each policy induces a consequence vector and heterogeneous feedback channels act on attainable consequence differences. Rather than asking whether a latent scalar reward is identifiable, we ask whether the welfare preorder on policies is identifiable. Our first result is a feedback-operator factorization theorem: exact identifiability holds if and only if the welfare label factors through the restricted feedback operator on the attainable difference set. In the full-dimensional polyhedral regime, this becomes a facet-observability theorem: exact identifiability is equivalent to observability of every order-separating facet normal of the welfare cone. Under Gaussian paired-difference feedback, each facet carries an intrinsic noise level $\sigma_j^2 = w_j^\top (A_V^* \Sigma^{-1} A_V)^\dagger w_j$, and the hardest observable facet governs the minimax difficulty of welfare-label recovery. We derive an exact two-point lower bound and a matching observable cone test, optimal up to the logarithmic factor required to control all facet inequalities simultaneously. Finally, we prove that exact order identifiability can hold even when no scalar Markov reward represents the same policy preorder. These results reframe world-feedback RL as an inverse problem over welfare orders rather than rewards.

1 Introduction

Much of modern alignment and preference-learning theory is reward-centric: noisy feedback is treated as evidence about a latent scalar objective to be recovered, identified, or optimized. This viewpoint underlies recent theory for heterogeneous feedback aggregation, multi-source imperfect preferences, general preference models, and latent-reward inference from multiple feedback modalities [Park et al., 2024, Shi et al., 2026, Ye et al., 2024, Baur et al., 2026]. It also aligns with a refined theory of reward ambiguity and invariance in inverse reward learning [Skalse et al., 2023, Metelli et al., 2023]. But many signals in real systems are not naturally evidence about a single reward: they are measurements of world consequences such as latency, energy, safety violations, service failure, or economic loss.

This mismatch is central for reinforcement learning from world feedback. There, the primitive object is often a *consequence vector*, not a scalar utility. A policy may simultaneously affect throughput, delay, risk, and constraint violation probability, so the relevant question is not necessarily "which reward generated the feedback?" but rather "which welfare relation over policies is identified by the available feedback channels?" The distinction is substantive: scalar Markov rewards cannot, in general, represent broad classes of multi-objective, risk-sensitive, and modal tasks [Skalse and Abate, 2023]. Thus perfect scalar-reward recovery may still miss the normative object.

We develop a theory of *order observability* for reinforcement learning from world feedback. Each policy π induces a consequence vector c^π, and welfare is encoded by a cone:

$$\pi \succeq \pi' \quad \Longleftrightarrow \quad c^\pi - c^{\pi'} \in C.$$

*Virginia Tech, Blacksburg, VA, USA. Correspondence to: Manoj Saravanan <manoj663@vt.edu>.

ICML 2026 Workshop on Reinforcement Learning from World Feedback (RLxF).

Heterogeneous feedback channels are modeled as observation operators on attainable consequence differences. The central problem is therefore an inverse problem over policy orderings: when do the induced observation laws determine the welfare label of every attainable policy pair?

Our contributions are fourfold. First, we prove a *feedback-operator factorization theorem*: exact identifiability is equivalent to the welfare label factoring through the restricted feedback operator on the attainable difference set. Second, in the full-dimensional polyhedral regime, we sharpen this into a *facet-observability theorem*: exact identifiability holds if and only if every order-separating facet normal lies in the row space of the fused operator. Third, under Gaussian paired-difference feedback, we derive a finite-sample minimax theory. Each facet carries an intrinsic noise scale $\sigma_j^2 = w_j^\top (A_V^* \Sigma^{-1} A_V)^\dagger w_j$, the hardest facet governs the minimax lower bound, and the observable cone test is optimal up to the logarithmic factor needed to protect all facet inequalities. Fourth, we prove a separation theorem: exact welfare-order identifiability can hold even when no scalar Markov reward represents the same preorder.

Our paper is related to, but distinct from, recent work on heterogeneous human feedback, multi-source imperfect preferences, general-preference RLHF, and multiple-feedback latent-reward models [Park et al., 2024, Shi et al., 2026, Ye et al., 2024, Baur et al., 2026]. Those lines remain centered on latent rewards or generic preference games. Our object is different: exact identifiability of a welfare preorder from heterogeneous consequence-observation operators. The closest conceptual neighbors are therefore recent theories of reward invariance, partial identifiability, and limitations of scalar Markov reward expressivity [Skalse et al., 2023, Metelli et al., 2023, Skalse and Abate, 2023].

Section 2 formalizes the model, Section 3 develops the exact identifiability theory, Section 4 derives the finite-sample minimax bounds and observable cone test, and Section 5 proves the separation from scalar-reward representability. The appendices extend the theory to robustness, non-Gaussian channels, and Hilbert consequence spaces.

2 Framework: policies, consequences, welfare, and feedback

We work in a finite-dimensional linear core. Fix a finite-horizon MDP

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, H, \nu_1, \{P_h\}_{h=1}^{H-1}),$$

with finite state and action spaces, horizon $H \in \mathbb{N}$, initial distribution ν_1, and transition kernels P_h. Let Π be a nonempty class of randomized Markov policies. For each $\pi \in \Pi$, let $x^\pi \in \mathbb{R}^n$, $n := H|\mathcal{S}||\mathcal{A}|$, denote the stage-wise state-action occupancy vector,

$$x_h^\pi(s,a) := \mathbb{P}_\pi(s_h = s, a_h = a), \qquad X := \{x^\pi : \pi \in \Pi\}.$$

The primitive object is a linear *consequence map*

$$B : \mathbb{R}^n \to \mathbb{R}^d, \qquad c^\pi := Bx^\pi.$$

The coordinates of c^π encode world consequences of executing π: latency, energy, failure probability, resource use, or any other quantity linear in the occupancy representation.

Fix a closed, pointed, polyhedral cone $C \subset \mathbb{R}^d$, and define welfare by

$$\pi \succeq \pi' \quad \Longleftrightarrow \quad c^\pi - c^{\pi'} \in C.$$

Define the policy-difference set and its effective span by

$$\Delta_\Pi := \{c^\pi - c^{\pi'} : \pi, \pi' \in \Pi\}, \qquad V := \operatorname{span}(\Delta_\Pi),$$

and the restricted cone $C_V := C \cap V$. We fix an irredundant facet representation

$$C_V = \{v \in V : w_j^\top v \geq 0,\ j = 1, \dots, r\}.$$

Heterogeneous feedback is modeled by linear observation operators $A_m : \mathbb{R}^d \to \mathbb{R}^{q_m}$ and the fused operator

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_M \end{bmatrix} : \mathbb{R}^d \to \mathbb{R}^q, \qquad q := \sum_{m=1}^M q_m.$$

For the exact and finite-sample theory in the main text we study the Gaussian paired-difference experiment

$$Y^{\pi,\pi'} = A(c^{\pi} - c^{\pi'}) + \xi, \qquad \xi \sim \mathcal{N}(0, \Sigma),\ \Sigma \succ 0.$$

Definition 2.1 (Exact identifiability)**.** Let $\mathsf{P}_{\pi,\pi'}$ denote the law of $Y^{\pi,\pi'}$. The welfare preorder is *exactly identifiable from A on Π* if there exists a measurable map $\phi : \mathcal{P}(\mathbb{R}^q) \to \{0, 1\}$ such that

$$\phi(\mathsf{P}_{\pi,\pi'}) = \mathbb{I}\{\pi \succeq \pi'\}, \qquad \forall\, \pi, \pi' \in \Pi.$$

Equivalently, identical observation laws must induce identical welfare labels.

Assumption 2.2 (Main-text regime)**.** The main results assume: (i) X is a nonempty compact polytope and B is linear; (ii) C_V is a closed, pointed, polyhedral cone with irredundant facet normals $w_1, \ldots, w_r$; (iii) every active facet intersects $\mathrm{relint}(\Delta_\Pi)$, so attainable differences can cross each order-separating boundary inside V; and (iv) the paired-difference model is Gaussian with fixed covariance $\Sigma \succ 0$.

3 Exact identifiability from feedback operators

By Theorem A.7, exact identifiability depends only on the effective ordered vector space (V, C_V) and the restricted feedback operator

$$A_V := A\restriction_V : V \to \mathbb{R}^q.$$

Accordingly, throughout this section we work entirely on V. In the reward-learning literature, partial identifiability is typically phrased in terms of observationally invisible rewards or value functionals [Skalse et al., 2023, Metelli et al., 2023]. Our first theorem gives the order-theoretic analogue for reinforcement learning from world feedback: exact identifiability is equivalent to a factorization of the welfare label through the restricted feedback operator on the attainable difference set. We then show that, in the full-dimensional polyhedral regime of the main text, this abstract factorization reduces to a concrete linear-algebraic condition: every order-separating facet normal must be observable through A_V.

For any linear map $T : V \to \mathbb{R}^q$, we write

$$\mathrm{row}(T) := \mathrm{range}(T^*) \subseteq V$$

for its row space, where $T^* : \mathbb{R}^q \to V$ is the Euclidean adjoint.

3.1 Order factorization through the feedback operator

Define the welfare label map on attainable consequence differences by

$$\psi_V : \Delta_\Pi \to \{0, 1\}, \qquad \psi_V(d) := \mathbb{I}\{d \in C_V\}.$$

At the most abstract level, exact identifiability means that the fibers of A_V do not mix the two labels of ψ_V. The next theorem formalizes this observation.

Theorem 3.1 (Feedback-operator factorization)**.** *Work under the Gaussian paired-difference model of Section 2. Then the following are equivalent.*

1. *The welfare preorder $\succeq$ is exactly identifiable from A on Π.*
2. *The fibers of A_V on Δ_Π are order-monochromatic: for all $d, d' \in \Delta_\Pi$,*
$$A_V d = A_V d' \quad \Longrightarrow \quad \big(d \in C_V \iff d' \in C_V\big).$$
3. *There exists a compact set*
$$\overline{\mathcal{C}} \subseteq A_V(\Delta_\Pi)$$
such that, for every $d \in \Delta_\Pi$,
$$d \in C_V \quad \iff \quad A_V d \in \overline{\mathcal{C}}.$$

Equivalently, there exists a Borel map

$$\bar{\psi} : A_V(\Delta_\Pi) \to \{0, 1\}$$

such that

$$\psi_V = \bar{\psi} \circ A_V \qquad \text{on } \Delta_\Pi.$$

Proof sketch. The implication $(1) \Rightarrow (2)$ is immediate: under the Gaussian paired-difference model, two policy pairs induce the same observation law if and only if they induce the same mean $A_V d$. Therefore exact identifiability forces each fiber $A_V^{-1}(y) \cap \Delta_\Pi$ to carry a single welfare label.

For $(2) \Rightarrow (3)$, define the *observable welfare region*

$$\overline{\mathcal{C}} := A_V(C_V \cap \Delta_\Pi).$$

Because Δ_Π is compact by Theorem A.3, C_V is closed, and A_V is linear, $\overline{\mathcal{C}}$ is compact. Order-monochromaticity of the fibers then implies

$$d \in C_V \iff A_V d \in \overline{\mathcal{C}}, \qquad d \in \Delta_\Pi.$$

For $(3) \Rightarrow (1)$, the welfare label is already a deterministic function of the feedback mean $A_V d$. Since $y \mapsto \mathcal{N}(y, \Sigma)$ is a continuous injection on the compact set $A_V(\Delta_\Pi)$, it is a homeomorphism onto its image. Hence the factorization through $A_V d$ lifts to a measurable factorization through the observation law, yielding exact identifiability in the sense of Theorem 2.1. The full quotient-space proof is given in the appendix. □

Remark 3.2 (Observable welfare region). The set $\overline{\mathcal{C}}$ in Theorem 3.1 should be viewed as the *observable welfare region* in feedback space. In general, $\overline{\mathcal{C}}$ need not be a cone, because Δ_Π is a compact difference body rather than a linear subspace closed under positive scaling. Thus Theorem 3.1 is an exact statement about the attainable difference geometry, not yet about the global cone structure of V.

Remark 3.3 (Why the factorization theorem is abstract but not vacuous). Theorem 3.1 gives the correct abstract notion of exact identifiability, but it does not by itself yield a checkable condition in terms of the linear-algebraic structure of A. The content of the next subsection is that, in the full-dimensional polyhedral regime and under the local-richness condition, this abstract factorization collapses to a concrete criterion: every order-separating facet normal must lie in the row space of A_V.

3.2 Facet observability

Assume now that C_V is full-dimensional in V, and recall the irredundant facet representation

$$C_V = \{v \in V : w_j^\top v \geq 0, \;\; j = 1, \ldots, r\}.$$

The main result of this subsection is that exact identifiability is equivalent to observability of these facet normals through the restricted feedback operator.

Theorem 3.4 (Facet observability)**.** *Work under the Gaussian paired-difference model of Section 2. Assume that C_V is full-dimensional in V, and that Assumption 2.2(3) holds. Then the following are equivalent.*

1. *The welfare preorder $\succeq$ is exactly identifiable from A on Π.*
2. *Every facet normal is observable through A_V, i.e.,*
$$w_j \in \mathrm{row}(A_V) = \mathrm{range}(A_V^*), \qquad j = 1, \ldots, r.$$
3. *There exist vectors $u_1, \ldots, u_r \in \mathbb{R}^q$ such that*
$$A_V^* u_j = w_j, \qquad j = 1, \ldots, r.$$
Equivalently, for every $d \in \Delta_\Pi$,
$$d \in C_V \iff u_j^\top A_V d \geq 0, \qquad j = 1, \ldots, r.$$

Proof sketch. The equivalence of (2) and (3) is immediate from $\mathrm{row}(A_V) = \mathrm{range}(A_V^*)$. For $(3) \Rightarrow (1)$, if $A_V^* u_j = w_j$ for all j, then

$$w_j^\top d = u_j^\top A_V d \qquad \forall d \in \Delta_\Pi,$$

so cone membership is a deterministic function of $A_V d$. Exact identifiability then follows from Theorem 3.1.

For $(1) \Rightarrow (2)$, argue by contradiction. Suppose $w_j \notin \mathrm{row}(A_V)$ for some facet j. By the row-space / kernel orthogonality identity, there exists an invisible direction $h \in \ker(A_V) \subseteq V$ with $w_j^\top h \neq 0$. By the local-richness assumption, the policy-difference set intersects the relative interior of the facet $F_j = \{v \in C_V : w_j^\top v = 0\}$ with nontrivial thickness. Using the facet-crossing lemma from Appendix A, one can construct two attainable differences

$$d_+,\ d_- \in \Delta_\Pi$$

such that

$$A_V d_+ = A_V d_-, \qquad d_+ \in \mathrm{relint}(C_V), \qquad d_- \notin C_V.$$

These two differences induce identical Gaussian observation laws but opposite welfare labels, contradicting exact identifiability. The full proof is given in Appendix C. □

Remark 3.5 (Observable facet certificates). Any collection $u_1, \ldots, u_r$ satisfying $A_V^* u_j = w_j$ will be called a set of *observable facet certificates*. Theorem 3.4 reduces exact order recovery to testing the signs of the linear functionals $u_j^\top A_V d$. In the statistical section, we will choose certificates with minimum variance under the Gaussian experiment.

Corollary 3.6 (Heterogeneous fusion can create exact identifiability)**.** *There exist a policy class Π, a consequence map B, a welfare cone C, and two feedback channels A_1, A_2 such that the induced welfare preorder is not exactly identifiable from A_1 alone and not exactly identifiable from A_2 alone, but is exactly identifiable from the fused operator*

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}.$$

Proof. Consider the one-step randomized model with four actions whose consequence vectors are the corners of the unit square:

$$(0,0),\ (0,1),\ (1,0),\ (1,1).$$

Then

$$\mathcal{K}_\Pi = [0,1]^2, \qquad \Delta_\Pi = [-1,1]^2, \qquad V = \mathbb{R}^2.$$

Let the welfare cone be

$$C_V = \mathbb{R}^2_+ = \{v \in \mathbb{R}^2 : e_1^\top v \geq 0,\ e_2^\top v \geq 0\},$$

so the facet normals are $w_1 = e_1$ and $w_2 = e_2$. Let

$$A_1 = [1 \quad 0], \qquad A_2 = [0 \quad 1].$$

Then

$$\mathrm{range}(A_1^*) = \mathrm{span}(e_1), \qquad \mathrm{range}(A_2^*) = \mathrm{span}(e_2),$$

so neither A_1 nor A_2 observes both facet normals. By Theorem 3.4, exact identifiability fails for each channel separately. For the fused operator

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2,$$

one has $\mathrm{range}(A^*) = \mathbb{R}^2$, so both facet normals are observable. Another application of Theorem 3.4 gives exact identifiability for the fused system. □

Corollary 3.7 (Necessary observation rank)**.** *Let*

$$L := \mathrm{span}\{w_1, \ldots, w_r\} \subseteq V.$$

If the welfare preorder is exactly identifiable from A on Π, then

$$\mathrm{rank}(A_V) = \dim \mathrm{range}(A_V^*) \geq \dim L.$$

In particular, any exact feedback system must expose at least $\dim L$ linearly independent observable directions on V.

Proof. If exact identifiability holds, then Theorem 3.4 implies $w_j \in \mathrm{range}(A_V^*)$ for every j. Therefore

$$L \subseteq \mathrm{range}(A_V^*),$$

and hence

$$\dim L \leq \dim \mathrm{range}(A_V^*) = \mathrm{rank}(A_V).$$ □

Takeaway. Theorem 3.1 identifies the exact abstract obstruction to order recovery: feedback fibers must be welfare-monochromatic on the attainable difference set. Theorem 3.4 then converts that obstruction into a sharp linear-algebraic criterion: in the full-dimensional polyhedral regime, exact RLxF identifiability holds if and only if every order-separating facet normal is visible through the restricted feedback operator.

4 Statistical complexity and optimal testing

The previous section characterized *whether* exact welfare-order recovery is possible. We now study the quantitative question: given exact identifiability, what is the minimax difficulty of deciding the welfare label from n noisy feedback samples, and what test achieves that rate? Throughout this section, we assume the equivalent conditions of Theorem 3.4, so in particular every facet normal $w_1, \dots, w_r$ is observable through the restricted feedback operator A_V.

Fix an unknown attainable consequence difference $d \in \Delta_\Pi$, and suppose we observe

$$Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(A_V d, \Sigma), \qquad \Sigma \succ 0,$$

with sample mean

$$\bar{Y}_n := \frac{1}{n} \sum_{k=1}^{n} Y_k \sim \mathcal{N}\Big(A_V d, \frac{1}{n}\Sigma\Big).$$

Our statistical target is the binary welfare label

$$\psi_V(d) = \mathbb{I}\{d \in C_V\}.$$

The observable geometry. Define the self-adjoint positive semidefinite operator

$$G := A_V^* \Sigma^{-1} A_V : V \to V.$$

Since exact identifiability holds, Theorem 3.4 implies

$$w_j \in \operatorname{row}(A_V) = \operatorname{range}(A_V^*) = \operatorname{range}(G), \qquad j = 1, \dots, r.$$

Hence the Moore–Penrose pseudoinverse $G^\dagger$ is well-defined on every facet normal. For each j, define the *minimum-variance observable certificate*

$$u_j^\star := \Sigma^{-1} A_V G^\dagger w_j \in \mathbb{R}^q \tag{4.1}$$

and its variance

$$\sigma_j^2 := w_j^\top G^\dagger w_j. \tag{4.2}$$

Since $w_j \neq 0$ and $w_j \in \operatorname{range}(G)$, one has $\sigma_j^2 > 0$ for every j.

Proposition 4.1 (Minimum-variance observable certificates)**.** *For each facet j, the vector $u_j^\star$ satisfies*

$$A_V^* u_j^\star = w_j.$$

Moreover, it uniquely solves the quadratic program

$$\min_{u \in \mathbb{R}^q} u^\top \Sigma u \qquad \textit{subject to} \qquad A_V^* u = w_j,$$

and the optimal value is σ_j^2. Equivalently,

$$\sigma_j^2 = \inf\{u^\top \Sigma u : A_V^* u = w_j\} = w_j^\top G^\dagger w_j.$$

Proof sketch. Because $w_j \in \operatorname{range}(G)$, one has

$$A_V^* u_j^\star = A_V^* \Sigma^{-1} A_V G^\dagger w_j = G G^\dagger w_j = w_j.$$

Thus $u_j^\star$ is a valid certificate. The quadratic program is a linearly constrained strictly convex minimization problem. Its Lagrange conditions yield the unique optimizer in (4.1), and substituting back gives the value in (4.2). The full derivation is given in Appendix D. □

Margin-separated alternatives. For each facet j, fix a witness

$$d_j^\circ \in \operatorname{relint}(F_j) \cap \operatorname{relint}(\Delta_\Pi)$$

guaranteed by Assumption 2.2(3), where

$$F_j := \{v \in C_V : w_j^\top v = 0\}.$$

Define the *least-observable unit-crossing direction*

$$h_j^\star := \frac{G^\dagger w_j}{w_j^\top G^\dagger w_j} = \frac{G^\dagger w_j}{\sigma_j^2}. \tag{4.3}$$

Then $w_j^\top h_j^\star = 1$, and by Theorem A.13 together with $d_j^\circ \in \operatorname{relint}(\Delta_\Pi)$, there exists a radius $\bar{\gamma}_j > 0$ such that, for every $0 < \gamma \le \bar{\gamma}_j$,

$$d_j^\circ + \gamma h_j^\star \in \Delta_\Pi \cap C_V, \qquad d_j^\circ - \gamma h_j^\star \in \Delta_\Pi \setminus C_V,$$

and, moreover, the positive-side alternative has margin at least γ on every facet. Let

$$\bar{\gamma} := \min_{1 \le j \le r} \bar{\gamma}_j > 0.$$

For $0 < \gamma \le \bar{\gamma}$, define the γ-margin class

$$\mathcal{D}_\gamma := \underbrace{\{d \in \Delta_\Pi \cap C_V : \min_{1 \le j \le r} w_j^\top d \ge \gamma\}}_{\text{inside the cone with margin } \gamma} \cup \underbrace{\{d \in \Delta_\Pi \setminus C_V : \max_{j : w_j^\top d < 0} (-w_j^\top d) \ge \gamma\}}_{\text{outside the cone with at least one violated facet of size } \gamma}.$$

Thus $\mathcal{D}_\gamma$ excludes differences that lie arbitrarily close to the welfare boundary.

Theorem 4.2 (Minimax lower bound)**.** *Assume exact identifiability, and let*

$$\sigma_* := \max_{1 \le j \le r} \sigma_j.$$

Then, for every $0 < \gamma \le \bar{\gamma}$ and every $n \in \mathbb{N}$,

$$\inf_{\hat{\psi}_n} \sup_{d \in \mathcal{D}_\gamma} \mathbb{P}_d\big(\hat{\psi}_n(Y_1, \dots, Y_n) \ne \psi_V(d)\big) \;\ge\; \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big),$$

where the infimum is over all measurable tests $\hat{\psi}_n : (\mathbb{R}^q)^n \to \{0, 1\}$, and Φ is the standard normal cdf.

Proof sketch. Choose a hardest facet $j_* \in \arg\max_j \sigma_j$, and consider the two local alternatives

$$d_+ := d_{j_*}^\circ + \gamma h_{j_*}^\star, \qquad d_- := d_{j_*}^\circ - \gamma h_{j_*}^\star.$$

By the definition of $\bar{\gamma}$, both lie in $\mathcal{D}_\gamma$, with

$$d_+ \in C_V, \qquad d_- \notin C_V.$$

Under $d_\pm$, the scalar statistic

$$Z_n := u_{j_*}^{\star\top} \bar{Y}_n$$

is Gaussian with

$$Z_n \sim \mathcal{N}(\pm\gamma, \sigma_{j_*}^2/n).$$

Thus the binary testing problem between d_+ and d_- reduces to the classical one-dimensional Gaussian location problem, whose minimax error is exactly

$$\Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_{j_*}}\Big).$$

Since $\sigma_{j_*} = \sigma_*$, and since the full class $\mathcal{D}_\gamma$ contains this two-point subproblem, the stated lower bound follows. The full proof is given in Appendix D. □

The observable cone test. The factorization and facet-observability theorems suggest a canonical plug-in decision rule: evaluate the observable facet certificates on the sample mean and declare the welfare label "inside the cone" if and only if every certificate is nonnegative. For $j = 1, \dots, r$, define

$$T_{j,n} := u_j^{\star\top} \bar{Y}_n.$$

Since $A_V^* u_j^\star = w_j$ and $u_j^{\star\top} \Sigma u_j^\star = \sigma_j^2$, one has, under a generic difference $d \in \Delta_\Pi$,

$$T_{j,n} \sim \mathcal{N}\Big(w_j^\top d, \frac{\sigma_j^2}{n}\Big). \tag{4.4}$$

We therefore define the *observable cone test*

$$\hat{\psi}_n^{\mathrm{OCT}} := \mathbb{I}\{T_{j,n} \geq 0, \;\; j = 1, \dots, r\} = \mathbb{I}\{u_j^{\star\top} \bar{Y}_n \geq 0, \;\; j = 1, \dots, r\}. \tag{4.5}$$

Theorem 4.3 (Upper bound for the observable cone test)**.** *Assume exact identifiability, and let $0 < \gamma \leq \bar{\gamma}$. Then the observable cone test satisfies the following one-sided error bounds:*

$$\sup_{d \in \mathcal{D}_\gamma \cap C_V} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 0\big) \leq \sum_{j=1}^r \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_j}\Big), \tag{4.6}$$

$$\sup_{d \in \mathcal{D}_\gamma \setminus C_V} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 1\big) \leq \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big). \tag{4.7}$$

Consequently,

$$\sup_{d \in \mathcal{D}_\gamma} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} \neq \psi_V(d)\big) \leq \sum_{j=1}^r \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_j}\Big) \leq r\,\Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big). \tag{4.8}$$

In particular, using the standard Gaussian tail bound $\Phi(-x) \leq e^{-x^2/2}$, it suffices to take

$$n \geq \frac{2\sigma_*^2}{\gamma^2} \log \frac{r}{\delta} \tag{4.9}$$

to guarantee

$$\sup_{d \in \mathcal{D}_\gamma} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} \neq \psi_V(d)\big) \leq \delta.$$

Proof sketch. Fix $d \in \mathcal{D}_\gamma$.

Inside-cone case. Suppose $d \in \mathcal{D}_\gamma \cap C_V$. By definition of $\mathcal{D}_\gamma$,

$$w_j^\top d \geq \gamma, \qquad j = 1, \dots, r.$$

Hence, by (4.4),

$$\mathbb{P}_d(T_{j,n} < 0) = \Phi\Big(-\frac{\sqrt{n}\, w_j^\top d}{\sigma_j}\Big) \leq \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_j}\Big).$$

The error event $\{\hat{\psi}_n^{\mathrm{OCT}} = 0\}$ is the union of the events $\{T_{j,n} < 0\}$, so a union bound yields (4.6).

Outside-cone case. Suppose $d \in \mathcal{D}_\gamma \setminus C_V$. Then, by definition of $\mathcal{D}_\gamma$, there exists at least one violated facet $j(d) \in \{1, \dots, r\}$ such that

$$w_{j(d)}^\top d \leq -\gamma.$$

If $\hat{\psi}_n^{\mathrm{OCT}} = 1$, then all facet statistics are nonnegative, and in particular

$$T_{j(d),n} \geq 0.$$

Therefore, again using (4.4),

$$\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 1\big) \leq \mathbb{P}_d\big(T_{j(d),n} \geq 0\big) = \Phi\Big(\frac{\sqrt{n}\, w_{j(d)}^\top d}{\sigma_{j(d)}}\Big) \leq \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_{j(d)}}\Big) \leq \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big),$$

which proves (4.7).

The uniform bound (4.8) follows because the inside-cone bound dominates the outside-cone bound:

$$\Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big) \le \sum_{j=1}^{r} \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_j}\Big).$$

Finally, (4.9) is immediate from (4.8) and the bound $\Phi(-x) \le e^{-x^2/2}$. The full proof is given in Appendix E. □

Corollary 4.4 (Finite-policy dominance-graph recovery)**.** *Let $\Pi_N = \{\pi_1, \dots, \pi_N\} \subseteq \Pi$ be a finite family of policies, and assume that every ordered pairwise difference*

$$c^{\pi_a} - c^{\pi_b}, \qquad a \neq b,$$

belongs to $\mathcal{D}_\gamma$. Suppose that for each ordered pair (π_a, π_b), $a \neq b$, one collects n i.i.d. paired-difference samples and applies the observable cone test (4.5) *independently to that pair. Then the full pairwise welfare relation on Π_N is recovered with probability at least $1-\delta$ whenever*

$$n \ge \frac{2\sigma_*^2}{\gamma^2} \log \frac{rN(N-1)}{\delta}. \tag{4.10}$$

In particular, under (4.10)*, the maximal-policy set under $\succeq$ is also recovered with probability at least $1-\delta$.*

Proof. Immediate from Theorem 4.3 and a union bound over the $N(N-1)$ ordered pairs. □

5 Separation from scalar-reward representability

The preceding results characterize when world feedback identifies a welfare order. The next theorem shows that this is strictly more general than scalar reward recovery. Call a preorder on Π *scalar-Markov representable* if there exists a stage-wise reward r such that

$$\pi \succeq \pi' \iff J_r(\pi) \ge J_r(\pi').$$

Theorem 5.1 (Exact order identifiability without scalar-reward representability)**.** *There exist a finite-horizon MDP $\mathcal{M}$, a policy class Π, a linear consequence map B, a closed pointed polyhedral cone C, and a feedback operator A such that (i) the induced welfare preorder*

$$\pi \succeq \pi' \iff c^{\pi} - c^{\pi'} \in C$$

is exactly identifiable from A on Π, but (ii) the same preorder is not scalar-Markov representable on Π.

Proof. Take the one-state, one-step MDP with actions a, b, c, the full randomized policy simplex $x^\pi \in \Delta(\{a, b, c\})$, consequence map

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \qquad C = \mathbb{R}^2_+, \qquad A = I_2.$$

Then welfare is Pareto dominance and Gaussian paired-difference feedback makes it exactly identifiable. But the deterministic policies choosing a and b have consequences $(1, 0)$ and $(0, 1)$ and are therefore incomparable, whereas any scalar Markov reward induces a complete preorder on the randomized policy simplex. Contradiction. □

6 Discussion and scope

Our results should be read as an observability theory for welfare orders. The main text isolates the finite-dimensional polyhedral regime in which exact identifiability becomes a row-space condition and testing is governed by the observable Gram operator $G = A_V^* \Sigma^{-1} A_V$. The appendices show that this is not fundamental: approximate observability yields controlled bias degradation, regular non-Gaussian channels replace G by a pullback Fisher operator, and Hilbert consequence spaces lead to the observable dual subspace $\overline{\mathrm{range}(A_\Delta^*)} = \ker(A_\Delta)^\perp$. Natural next steps are active query design, joint learning of B and welfare order, and a full exploration theory coupling order observability with online control.

References

Raphaël Baur, Yannick Metz, Maria Gkoulta, Mennatallah El-Assady, Giorgia Ramponi, and Thomas Kleine Buening. MAVRL: Learning reward functions from multiple feedback types with amortized variational inference, 2026. URL https://arxiv.org/abs/2602.15206.

Alberto Maria Metelli, Filippo Lazzati, and Marcello Restelli. Towards theoretical understanding of inverse reinforcement learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 24555–24591. PMLR, 2023. URL https://proceedings.mlr.press/v202/metelli23a.html.

Chanwoo Park, Mingyang Liu, Dingwen Kong, Kaiqing Zhang, and Asuman Ozdaglar. RLHF from heterogeneous feedback via personalization and preference aggregation, 2024. URL https://arxiv.org/abs/2405.00254.

Ming Shi, Yingbin Liang, Ness B. Shroff, and Ananthram Swami. Regret bounds for reinforcement learning from multi-source imperfect preferences, 2026. URL https://arxiv.org/abs/2603.20453.

Joar Skalse and Alessandro Abate. On the limitations of Markovian rewards to express multi-objective, risk-sensitive, and modal tasks. In Robin J. Evans and Ilya Shpitser, editors, *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 1974–1984. PMLR, 2023. URL https://proceedings.mlr.press/v216/skalse23a.html.

Joar Max Viktor Skalse, Matthew Farrugia-Roberts, Stuart Russell, Alessandro Abate, and Adam Gleave. Invariance in policy optimisation and partial identifiability in reward learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 32033–32058. PMLR, 2023. URL https://proceedings.mlr.press/v202/skalse23a.html.

Chenlu Ye, Wei Xiong, Yuheng Zhang, Hanze Dong, Nan Jiang, and Tong Zhang. Online iterative reinforcement learning from human feedback with general preference model. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-2598. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/94d13c2401fe119e57ba325b6fe526e0-Abstract-Conference.html.

A Ordered-vector-space and occupancy preliminaries

This appendix develops the geometric backbone behind the main text. In spirit, it is closest to recent work on reward-learning invariances and partial identifiability [Skalse et al., 2023, Metelli et al., 2023], but the primitive object here is not a latent scalar reward. Instead, the central objects are the consequence image, the policy-difference set, and the welfare cone restricted to the effective span of attainable differences. The final separation from scalar-reward representability is motivated by the expressivity limits established by Skalse and Abate [2023].

A.1 Occupancy images and consequence images

We begin with the standard tabular flow characterization of finite-horizon occupancy measures. Although classical, we include it for self-containment and to make clear which aspects of the later theory depend on the controlled system and which depend only on the induced difference geometry.

For the unrestricted tabular policy class, define the flow polytope

$$X_{\text{flow}} := \left\{ x \in \mathbb{R}^n_{\geq 0} : \begin{array}{l} \sum_{a\in\mathcal{A}} x_1(s,a) = \nu_1(s), \qquad \forall s \in \mathcal{S}, \\ \sum_{a\in\mathcal{A}} x_h(s,a) = \sum_{s'\in\mathcal{S}} \sum_{a'\in\mathcal{A}} P_{h-1}(s \mid s',a')\, x_{h-1}(s',a'), \\ \qquad\qquad \forall h \in \{2,\dots,H\},\ \forall s \in \mathcal{S} \end{array} \right\}.$$

Proposition A.1 (Flow characterization of occupancies)**.** *Suppose* Π *contains all randomized Markov policies on* $\mathcal{M}$*. Then*

$$X = X_{\text{flow}}.$$

In particular, X *is a nonempty compact convex polytope.*

Proof. Fix $\pi \in \Pi$. By definition,

$$x_h^\pi(s,a) = \mathbb{P}_\pi(s_h = s, a_h = a) \geq 0.$$

At stage $h = 1$,

$$\sum_{a\in\mathcal{A}} x_1^\pi(s,a) = \mathbb{P}_\pi(s_1 = s) = \nu_1(s), \qquad \forall s \in \mathcal{S}.$$

For $h \geq 2$, the law of total probability gives

$$\begin{aligned}\sum_{a\in\mathcal{A}} x_h^\pi(s,a) &= \mathbb{P}_\pi(s_h = s)\\ &= \sum_{s'\in\mathcal{S}}\sum_{a'\in\mathcal{A}} \mathbb{P}_\pi(s_{h-1} = s', a_{h-1} = a')\, P_{h-1}(s \mid s', a')\\ &= \sum_{s'\in\mathcal{S}}\sum_{a'\in\mathcal{A}} P_{h-1}(s \mid s', a')\, x_{h-1}^\pi(s', a').\end{aligned}$$

Hence $x^\pi \in X_{\text{flow}}$, proving $X \subseteq X_{\text{flow}}$.

For the reverse inclusion, fix $x \in X_{\text{flow}}$. Define the state marginals

$$d_h(s) := \sum_{a\in\mathcal{A}} x_h(s,a), \qquad h = 1, \ldots, H.$$

Construct a randomized Markov policy π by

$$\pi_h(a \mid s) := \begin{cases} \dfrac{x_h(s,a)}{d_h(s)}, & d_h(s) > 0,\\ \text{arbitrary in } \Delta(\mathcal{A}), & d_h(s) = 0.\end{cases}$$

We show by induction on h that $x_h^\pi(s,a) = x_h(s,a)$ for all s, a.

For $h = 1$, the flow constraint gives $d_1(s) = \nu_1(s)$, so

$$x_1^\pi(s,a) = \mathbb{P}_\pi(s_1 = s)\, \pi_1(a \mid s) = \nu_1(s)\, \pi_1(a \mid s) = x_1(s,a).$$

Assume now $x_{h-1}^\pi(s', a') = x_{h-1}(s', a')$ for all s', a'. Then

$$\begin{aligned}\mathbb{P}_\pi(s_h = s) &= \sum_{s'\in\mathcal{S}}\sum_{a'\in\mathcal{A}} \mathbb{P}_\pi(s_{h-1} = s', a_{h-1} = a')\, P_{h-1}(s \mid s', a')\\ &= \sum_{s'\in\mathcal{S}}\sum_{a'\in\mathcal{A}} x_{h-1}(s', a')\, P_{h-1}(s \mid s', a')\\ &= d_h(s),\end{aligned}$$

where the last equality is the flow constraint for x. Therefore

$$x_h^\pi(s,a) = \mathbb{P}_\pi(s_h = s)\, \pi_h(a \mid s) = d_h(s)\, \pi_h(a \mid s) = x_h(s,a).$$

Thus $x = x^\pi \in X$, proving $X_{\text{flow}} \subseteq X$.

Since X_{flow} is defined by finitely many linear equalities and inequalities, it is a polytope. Nonemptiness follows, for instance, from any deterministic Markov policy. □

Remark A.2 (Restricted policy classes). The later theory does *not* require Π to contain all randomized Markov policies. Once X is fixed, every subsequent result depends on the controlled system only through the geometric objects

$$X, \qquad \mathcal{K}_\Pi := B(X), \qquad \Delta_\Pi, \qquad V := \operatorname{span}(\Delta_\Pi).$$

Accordingly, Theorem A.1 is used only as a canonical reference model showing that Assumption 2.2(1) is natural in the unrestricted tabular case.

Define the consequence image

$$\mathcal{K}_\Pi := \{c^\pi : \pi \in \Pi\} = B(X) \subset \mathbb{R}^d.$$

Lemma A.3 (Consequence image and difference polytope)**.** *Under Assumption 2.2(1), the following hold.*

1. *$\mathcal{K}_\Pi$ is a nonempty compact convex polytope.*
2. $$\Delta_\Pi = \mathcal{K}_\Pi - \mathcal{K}_\Pi := \{u - v : u, v \in \mathcal{K}_\Pi\}.$$
3. *Δ_Π is a compact convex centrally symmetric polytope.*

Proof. Since X is a nonempty compact convex polytope and B is linear, its image

$$\mathcal{K}_\Pi = B(X)$$

is again a nonempty compact convex polytope.

By definition,

$$\Delta_\Pi = \{c^\pi - c^{\pi'} : \pi, \pi' \in \Pi\}.$$

Since $c^\pi \in \mathcal{K}_\Pi$ for every π, this is exactly the Minkowski difference

$$\mathcal{K}_\Pi - \mathcal{K}_\Pi = \mathcal{K}_\Pi + (-\mathcal{K}_\Pi).$$

Because $\mathcal{K}_\Pi$ is a compact convex polytope, so is $-\mathcal{K}_\Pi$, and the Minkowski sum of compact convex polytopes is again a compact convex polytope. Central symmetry follows from the identity

$$d \in \Delta_\Pi \quad \Longrightarrow \quad -d \in \Delta_\Pi,$$

since $d = u - v$ implies $-d = v - u$. □

Lemma A.4 (Span, affine hull, and interior around the origin)**.** *Let $V := \mathrm{span}(\Delta_\Pi)$. Then*

$$\mathrm{aff}(\Delta_\Pi) = V, \qquad 0 \in \mathrm{relint}(\Delta_\Pi),$$

where the relative interior is taken with respect to the Euclidean topology on V.

Proof. Because $\pi = \pi'$ is allowed in the definition of Δ_Π, one has $0 \in \Delta_\Pi$. Hence

$$\mathrm{aff}(\Delta_\Pi) = \mathrm{span}(\Delta_\Pi) = V.$$

It remains to show $0 \in \mathrm{relint}(\Delta_\Pi)$. Suppose, for contradiction, that $0 \notin \mathrm{relint}(\Delta_\Pi)$ relative to V. Since $\Delta_\Pi \subset V$ is nonempty, compact, convex, and contains the origin, the separating hyperplane theorem in the ambient space V implies that there exists a nonzero linear functional $\ell \in V^*$ such that

$$\ell(d) \le 0, \qquad \forall d \in \Delta_\Pi.$$

Identifying V^* with V via the Euclidean inner product, write $\ell(d) = g^\top d$ for some nonzero $g \in V$. Since Δ_Π is centrally symmetric by Theorem A.3, one also has $-d \in \Delta_\Pi$ whenever $d \in \Delta_\Pi$, and therefore

$$g^\top(-d) \le 0 \quad \Longrightarrow \quad g^\top d \ge 0 \qquad \forall d \in \Delta_\Pi.$$

Combining the two inequalities yields $g^\top d = 0$ for every $d \in \Delta_\Pi$. Thus g is orthogonal to $\mathrm{span}(\Delta_\Pi) = V$, contradicting $g \in V \setminus \{0\}$. Therefore $0 \in \mathrm{relint}(\Delta_\Pi)$. □

A.2 Restriction to the effective ordered space

The welfare relation of the main text depends only on whether a policy difference lies in the welfare cone. Since all attainable differences lie in $V = \mathrm{span}(\Delta_\Pi)$, the ambient space $\mathbb{R}^d$ is larger than necessary.

Define the restricted cone

$$C_V := C \cap V.$$

Proposition A.5 (Restriction of the welfare cone)**.** *Under Assumption 2.2(2), the set C_V is a closed, pointed, polyhedral cone in the Euclidean space V. Moreover, for every $d \in \Delta_\Pi$,*

$$d \in C \quad \Longleftrightarrow \quad d \in C_V.$$

Consequently, the preorder induced by $(\mathbb{R}^d, C)$ on policy pairs coincides with the preorder induced by (V, C_V).

Proof. Because V is a linear subspace of $\mathbb{R}^d$, the intersection $C_V = C \cap V$ of a closed polyhedral cone with a subspace is again a closed polyhedral cone in V. Pointedness is inherited: if $v \in C_V \cap (-C_V)$, then $v \in C \cap (-C)$, so $v = 0$ because C is pointed.

Now fix $d \in \Delta_\Pi$. Since $\Delta_\Pi \subset V$, one has

$$d \in C \quad \Longleftrightarrow \quad d \in C \cap V = C_V.$$

The final claim is simply a restatement of the definition of the preorder. □

Definition A.6 (Effective ordered vector space)**.** The pair (V, C_V) is called the *effective ordered vector space* induced by Π, B, and C.

Lemma A.7 (Reduction to the effective span)**.** *Let $A, \widetilde{A} : \mathbb{R}^d \to \mathbb{R}^q$ be linear operators with*

$$A\restriction_V = \widetilde{A}\restriction_V .$$

Then, for every policy pair (π, π'),

$$A(c^\pi - c^{\pi'}) = \widetilde{A}(c^\pi - c^{\pi'}).$$

In particular, exact identifiability on Π depends on the feedback operator only through its restriction to V, and depends on the welfare cone only through C_V.

Proof. For every policy pair,

$$c^\pi - c^{\pi'} \in \Delta_\Pi \subset V.$$

Hence

$$A(c^\pi - c^{\pi'}) = \widetilde{A}(c^\pi - c^{\pi'})$$

by equality of the restrictions on V. Therefore the paired-difference observation laws coincide under A and $\widetilde{A}$. The statement about the cone follows from Theorem A.5. □

Assumption A.8 (Nondegenerate ordered span)**.** For the facet-based exact-identifiability theory developed later, the restricted cone C_V is full-dimensional in V.

Remark A.9 (Why Theorem A.8 is natural). If C_V is not full-dimensional in V, then the induced order is degenerate on a positive-codimension subspace of attainable differences. In that regime, the later facet-normal characterization must be augmented with equality constraints describing $\mathrm{span}(C_V)$. Since the central theorems of the paper are designed to isolate *observability of order-separating directions*, the full-dimensional regime is the conceptually clean and statistically nontrivial case.

Proposition A.10 (Facet representation inside the effective span)**.** *Suppose Theorem A.8 holds. Then there exist nonzero vectors $w_1, \ldots, w_r \in V$ such that*

$$C_V = \{v \in V : w_j^\top v \geq 0, \;\; j = 1, \ldots, r\},$$

and the representation is irredundant in the sense that, for each j, the face

$$F_j := \{v \in C_V : w_j^\top v = 0\}$$

is a facet of C_V, i.e., a proper face of codimension one in V.

Proof. Let $m := \dim(V)$, and choose an orthonormal basis $u_1, \ldots, u_m$ of V. Let $U \in \mathbb{R}^{d \times m}$ be the matrix whose columns are $u_1, \ldots, u_m$. The map $y \mapsto Uy$ is an isometric isomorphism from $\mathbb{R}^m$ onto V.

Define the cone

$$\widehat{C} := \{y \in \mathbb{R}^m : Uy \in C_V\}.$$

Because C_V is a full-dimensional pointed polyhedral cone in V, $\widehat{C}$ is a full-dimensional pointed polyhedral cone in $\mathbb{R}^m$. By finite-dimensional polyhedral geometry, $\widehat{C}$ admits an irredundant facet representation

$$\widehat{C} = \{y \in \mathbb{R}^m : \widehat{w}_j^\top y \geq 0, \;\; j = 1, \ldots, r\},$$

where each

$$\widehat{F}_j = \{y \in \widehat{C} : \widehat{w}_j^\top y = 0\}$$

is a facet of $\widehat{C}$.

Set

$$w_j := U\widehat{w}_j \in V, \qquad j = 1, \ldots, r.$$

Then, for every $v \in V$, there exists a unique $y \in \mathbb{R}^m$ with $v = Uy$, and

$$w_j^\top v = (U\widehat{w}_j)^\top U y = \widehat{w}_j^\top y.$$

Hence

$$v \in C_V \quad \Longleftrightarrow \quad y \in \widehat{C} \quad \Longleftrightarrow \quad \widehat{w}_j^\top y \geq 0 \;\; \forall j \quad \Longleftrightarrow \quad w_j^\top v \geq 0 \;\; \forall j.$$

Irredundancy and the facet property transfer under the linear isomorphism $y \mapsto Uy$. □

A.3 Local richness and facet crossing

We now restate the local-richness condition from Assumption 2.2(3) in a way that makes its geometric role transparent.

Proposition A.11 (Equivalent form of local richness)**.** *Assume Theorem A.8, and let $F_1, \ldots, F_r$ be the facets of C_V from Theorem A.10. Then Assumption 2.2(3) is equivalent to*

$$\operatorname{relint}(F_j) \cap \operatorname{relint}(\Delta_\Pi) \neq \varnothing, \qquad j = 1, \ldots, r,$$

where $\operatorname{relint}(F_j)$ is taken relative to the affine hull of F_j, and $\operatorname{relint}(\Delta_\Pi)$ is taken relative to V.

Proof. By definition, Assumption 2.2(3) states that for each j, there exists $d_j \in \operatorname{relint}(F_j) \cap \Delta_\Pi$ and $\eta_j > 0$ such that

$$\{v \in V : \|v - d_j\|_2 < \eta_j\} \subset \Delta_\Pi.$$

But, by definition of relative interior in the ambient space V, the latter ball-inclusion statement is equivalent to $d_j \in \operatorname{relint}(\Delta_\Pi)$. Hence Assumption 2.2(3) holds exactly when

$$\operatorname{relint}(F_j) \cap \operatorname{relint}(\Delta_\Pi) \neq \varnothing$$

for every j. □

Remark A.12 (Why mere nonemptiness is vacuous). Because $0 \in \Delta_\Pi$ by construction and every cone facet contains the origin, the condition $F_j \cap \Delta_\Pi \neq \varnothing$ carries essentially no information. The substantive requirement is that Δ_Π intersect each facet *away from the apex* and with nontrivial transverse thickness inside V.

The next lemma is the geometric crossing statement that will later allow us to produce welfare-sign flips along observationally invisible directions.

Lemma A.13 (Small transverse perturbations across a facet)**.** *Assume Theorem A.8, and fix $j \in \{1, \ldots, r\}$. Let $d \in \operatorname{relint}(F_j)$. Then there exists $\eta > 0$ such that, for every $h \in V$ with $\|h\|_2 < \eta$, the following hold:*

1. *for every $\ell \neq j$,*
$$w_\ell^\top (d + h) > 0 \quad \textit{and} \quad w_\ell^\top (d - h) > 0;$$
2. *if $w_j^\top h > 0$, then*
$$d + h \in \operatorname{relint}(C_V) \qquad \textit{and} \qquad d - h \notin C_V;$$
3. *if $w_j^\top h < 0$, then*
$$d - h \in \operatorname{relint}(C_V) \qquad \textit{and} \qquad d + h \notin C_V.$$

Proof. Because $d \in \mathrm{relint}(F_j)$, one has

$$w_j^\top d = 0 \qquad \text{and} \qquad w_\ell^\top d > 0 \ \text{ for all } \ell \neq j.$$

Set

$$m_j := \min_{\ell \neq j} w_\ell^\top d > 0.$$

If $r = 1$, the minimum over $\ell \neq j$ is vacuous and the first statement is empty; in that case any sufficiently small $\eta > 0$ will do. Assume henceforth that $r \geq 2$. Choose

$$\eta < \frac{m_j}{2 \max_{\ell \neq j} \|w_\ell\|_2}.$$

Then for every $h \in V$ with $\|h\|_2 < \eta$ and every $\ell \neq j$,

$$|w_\ell^\top h| \leq \|w_\ell\|_2 \, \|h\|_2 < \frac{m_j}{2}.$$

Therefore

$$w_\ell^\top (d \pm h) = w_\ell^\top d \pm w_\ell^\top h \geq m_j - |w_\ell^\top h| > \frac{m_j}{2} > 0,$$

which proves the first claim.

Since $w_j^\top d = 0$, one has

$$w_j^\top (d + h) = w_j^\top h, \qquad w_j^\top (d - h) = -w_j^\top h.$$

If $w_j^\top h > 0$, then all facet inequalities are strictly positive at $d + h$, so $d + h \in \mathrm{relint}(C_V)$, while the j-th inequality is violated at $d - h$, so $d - h \notin C_V$. The case $w_j^\top h < 0$ is symmetric. □

Remark A.14 (What Appendix A buys us). Appendix A isolates the three geometric facts that power the later theory:

1. only the effective span $V = \mathrm{span}(\Delta_\Pi)$ matters;
2. the policy-difference set Δ_Π is a centrally symmetric compact convex body in V with $0 \in \mathrm{relint}(\Delta_\Pi)$;
3. the later exact-identifiability arguments reduce to understanding which facet normals of C_V are observable through the restricted operator $A\restriction_V$.

Everything in the main exact-identifiability theorem is a consequence of this ordered-vector-space reduction.

B Proof of the factorization theorem

This appendix gives the full proof of Theorem 3.1. The argument has two layers. First, we isolate the abstract quotient induced by observational indistinguishability under the restricted operator A_V. Second, we show that exact identifiability is equivalent to the welfare label descending to this quotient, or equivalently to the feedback-image set $A_V(\Delta_\Pi)$. In contrast to the facet-based result of Appendix C, no polyhedral structure is required here: the theorem is a purely order-theoretic statement about factorization through an observation operator. This is the policy-order analogue of the equivalence-class viewpoint that underlies partial identifiability in reward learning [Skalse et al., 2023, Metelli et al., 2023].

B.1 Indistinguishability and the quotient induced by A_V

Recall that

$$A_V := A\restriction_V : V \to \mathbb{R}^q$$

is the restricted feedback operator, and

$$\psi_V : \Delta_\Pi \to \{0, 1\}, \qquad \psi_V(d) := \mathbb{I}\{d \in C_V\}$$

is the welfare label on attainable consequence differences.

Definition B.1 (Indistinguishability relation)**.** Define a relation $\sim_{A_V}$ on Δ_Π by

$$d \sim_{A_V} d' \quad \Longleftrightarrow \quad A_V d = A_V d'.$$

We write

$$X_{A_V} := \Delta_\Pi / \sim_{A_V}$$

for the quotient set of equivalence classes, and

$$q_{A_V} : \Delta_\Pi \to X_{A_V}, \qquad q_{A_V}(d) := [d]_{A_V}$$

for the quotient map.

Lemma B.2 (Basic properties of the indistinguishability quotient)**.** *The relation $\sim_{A_V}$ is an equivalence relation on Δ_Π. If X_{A_V} is equipped with the quotient topology induced by q_{A_V}, then:*

1. *q_{A_V} is continuous and surjective;*
2. *X_{A_V} is compact;*
3. *the map*
$$\Theta_{A_V} : X_{A_V} \to A_V(\Delta_\Pi), \qquad \Theta_{A_V}([d]_{A_V}) := A_V d,$$
is well-defined and is a homeomorphism.

Proof. Reflexivity, symmetry, and transitivity of $\sim_{A_V}$ are immediate from the definition $d \sim_{A_V} d' \iff A_V d = A_V d'$, so it is an equivalence relation.

By construction, q_{A_V} is surjective. Endowing X_{A_V} with the quotient topology means precisely that a set $U \subseteq X_{A_V}$ is open iff $q_{A_V}^{-1}(U)$ is open in Δ_Π. Therefore q_{A_V} is continuous by definition.

Since Δ_Π is compact by Theorem A.3, and q_{A_V} is a continuous surjection, the quotient space X_{A_V} is compact.

We now prove the claims about Θ_{A_V}. First, Θ_{A_V} is well-defined: if $[d]_{A_V} = [d']_{A_V}$, then $d \sim_{A_V} d'$, so $A_V d = A_V d'$. Hence Θ_{A_V} does not depend on the choice of representative. It is clearly surjective onto $A_V(\Delta_\Pi)$. It is also injective: if $\Theta_{A_V}([d]_{A_V}) = \Theta_{A_V}([d']_{A_V})$, then $A_V d = A_V d'$, so $d \sim_{A_V} d'$, hence $[d]_{A_V} = [d']_{A_V}$.

To prove continuity, observe that

$$A_V = \Theta_{A_V} \circ q_{A_V} \qquad \text{on } \Delta_\Pi.$$

Since A_V is continuous and q_{A_V} is a quotient map, the universal property of quotient spaces implies that Θ_{A_V} is continuous.

Finally, $A_V(\Delta_\Pi)$ is compact as the continuous image of the compact set Δ_Π, hence it is Hausdorff as a compact subset of $\mathbb{R}^q$. Therefore Θ_{A_V} is a continuous bijection from the compact space X_{A_V} to the Hausdorff space $A_V(\Delta_\Pi)$, and is thus a homeomorphism. □

Remark B.3 (Quotient versus feedback image). The quotient X_{A_V} is the intrinsic space of observational equivalence classes, while $A_V(\Delta_\Pi)$ is its concrete realization inside $\mathbb{R}^q$. By Theorem B.2, these are canonically homeomorphic. Thus the factorization theorem may be stated either on the quotient X_{A_V} or directly on the compact image $A_V(\Delta_\Pi)$.

B.2 Monochromatic fibers and descent of the welfare label

The next lemma formalizes the elementary principle that a label descends through a map if and only if it is constant on the fibers of that map.

Lemma B.4 (Monochromatic fibers are equivalent to factorization)**.** *The following are equivalent.*

1. *The fibers of A_V on Δ_Π are order-monochromatic:*
$$A_V d = A_V d' \quad \Longrightarrow \quad \big(\psi_V(d) = \psi_V(d')\big), \qquad \forall\, d, d' \in \Delta_\Pi.$$
2. *There exists a unique map*
$$\widetilde{\psi}_V : X_{A_V} \to \{0, 1\}$$
such that
$$\psi_V = \widetilde{\psi}_V \circ q_{A_V} \qquad \text{on } \Delta_\Pi.$$

3. *There exists a unique map*
$$\bar{\psi}_V : A_V(\Delta_\Pi) \to \{0,1\}$$
such that
$$\psi_V = \bar{\psi}_V \circ A_V \qquad \text{on } \Delta_\Pi.$$

Moreover, if these equivalent conditions hold, then
$$\overline{\mathcal{C}} := \bar{\psi}_V^{-1}(1) = A_V(C_V \cap \Delta_\Pi) \subseteq A_V(\Delta_\Pi)$$
is compact, and
$$d \in C_V \iff A_V d \in \overline{\mathcal{C}}, \qquad \forall\, d \in \Delta_\Pi.$$

Proof. We prove $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1)$, and then the final claims about $\overline{\mathcal{C}}$.

Implication $(1) \Rightarrow (2)$**.** Assume (1). Define
$$\widetilde{\psi}_V([d]_{A_V}) := \psi_V(d), \qquad [d]_{A_V} \in X_{A_V}.$$
This is well-defined precisely because ψ_V is constant on each fiber of A_V, equivalently on each equivalence class of $\sim_{A_V}$. By construction,
$$(\widetilde{\psi}_V \circ q_{A_V})(d) = \widetilde{\psi}_V([d]_{A_V}) = \psi_V(d), \qquad d \in \Delta_\Pi.$$
Uniqueness follows from surjectivity of q_{A_V}.

Implication $(2) \Rightarrow (3)$**.** Assume (2). Since $\Theta_{A_V} : X_{A_V} \to A_V(\Delta_\Pi)$ is a bijection by Theorem B.2, define
$$\bar{\psi}_V := \widetilde{\psi}_V \circ \Theta_{A_V}^{-1}.$$
Then for every $d \in \Delta_\Pi$,
$$\bar{\psi}_V(A_V d) = \widetilde{\psi}_V\big(\Theta_{A_V}^{-1}(A_V d)\big) = \widetilde{\psi}_V([d]_{A_V}) = \psi_V(d),$$
so
$$\psi_V = \bar{\psi}_V \circ A_V \qquad \text{on } \Delta_\Pi.$$
Uniqueness follows because $A_V : \Delta_\Pi \to A_V(\Delta_\Pi)$ is surjective.

Implication $(3) \Rightarrow (1)$**.** Assume (3). If $A_V d = A_V d'$, then
$$\psi_V(d) = \bar{\psi}_V(A_V d) = \bar{\psi}_V(A_V d') = \psi_V(d'),$$
so the fibers of A_V are order-monochromatic.

Compactness and characterization of $\overline{\mathcal{C}}$. Assume any of the equivalent conditions holds. Since
$$C_V \cap \Delta_\Pi$$
is compact—being the intersection of the compact set Δ_Π with the closed cone $C_V \subseteq V$—its image
$$A_V(C_V \cap \Delta_\Pi)$$
is compact. Now fix $y \in A_V(\Delta_\Pi)$. By surjectivity of A_V onto its image, $y = A_V d$ for some $d \in \Delta_\Pi$. Then
$$y \in A_V(C_V \cap \Delta_\Pi) \iff \exists\, \tilde{d} \in C_V \cap \Delta_\Pi \text{ with } A_V \tilde{d} = y.$$
By monochromaticity of the fiber, this is equivalent to $d \in C_V$, i.e. to $\psi_V(d) = 1$, which is equivalent to $\bar{\psi}_V(y) = 1$. Hence
$$A_V(C_V \cap \Delta_\Pi) = \bar{\psi}_V^{-1}(1) = \overline{\mathcal{C}}.$$
The pointwise equivalence
$$d \in C_V \iff A_V d \in \overline{\mathcal{C}}$$
is then immediate from $\psi_V = \bar{\psi}_V \circ A_V$. □

Remark B.5 (Measurability of the descended label). The codomain $\{0,1\}$ is endowed with the discrete σ-algebra. Because $\overline{\mathcal{C}} \subseteq A_V(\Delta_\Pi)$ is compact, the map
$$\bar{\psi}_V(y) = \mathbb{I}\{y \in \overline{\mathcal{C}}\}$$
is Borel measurable on $A_V(\Delta_\Pi)$. Thus the factorization supplied by Theorem B.4 is measurable, not merely set-theoretic.

B.3 From feedback means to observation laws

The main theorem is stated in terms of exact identifiability from the *observation law*, rather than from the mean $A_V d$. We therefore record the basic topological properties of the Gaussian location family with fixed covariance.

Equip $\mathcal{P}(\mathbb{R}^q)$ with the weak topology and its Borel σ-algebra. Define

$$\Gamma_\Sigma : \mathbb{R}^q \to \mathcal{P}(\mathbb{R}^q), \qquad \Gamma_\Sigma(y) := \mathcal{N}(y, \Sigma),$$

where $\Sigma \succ 0$ is fixed.

Lemma B.6 (Gaussian location embedding)**.** *Let $S \subseteq \mathbb{R}^q$ be compact. Then:*

1. *$\Gamma_\Sigma|_S : S \to \Gamma_\Sigma(S)$ is continuous and injective;*
2. *$\Gamma_\Sigma(S)$ is compact, hence Borel, in $\mathcal{P}(\mathbb{R}^q)$;*
3. *$\Gamma_\Sigma|_S$ is a homeomorphism from S onto $\Gamma_\Sigma(S)$.*

In particular, for every $y, y' \in S$,

$$\Gamma_\Sigma(y) = \Gamma_\Sigma(y') \quad \Longleftrightarrow \quad y = y'.$$

Proof. Continuity of $y \mapsto \mathcal{N}(y, \Sigma)$ in the weak topology is standard: if $y_n \to y$ in $\mathbb{R}^q$, then for every bounded continuous test function f,

$$\int f \, d\mathcal{N}(y_n, \Sigma) \to \int f \, d\mathcal{N}(y, \Sigma).$$

Equivalently, if $Z \sim \mathcal{N}(0, \Sigma)$, then

$$\int f \, d\mathcal{N}(y_n, \Sigma) = \mathbb{E}[f(y_n + Z)] \to \mathbb{E}[f(y + Z)] = \int f \, d\mathcal{N}(y, \Sigma)$$

by dominated convergence, because f is bounded and $y_n + Z \to y + Z$ almost surely.

Injectivity follows because a Gaussian law with fixed covariance is determined by its mean. Indeed, if

$$\mathcal{N}(y, \Sigma) = \mathcal{N}(y', \Sigma),$$

then the two distributions have the same expectation, so $y = y'$.

Since S is compact and $\Gamma_\Sigma|_S$ is continuous, $\Gamma_\Sigma(S)$ is compact. The space $\mathcal{P}(\mathbb{R}^q)$ endowed with the weak topology is Hausdorff, so a continuous bijection from compact S onto its image $\Gamma_\Sigma(S)$ is a homeomorphism. This proves the third claim. □

B.4 Full proof of Theorem 3.1

We can now prove the factorization theorem exactly as stated in the main text.

Proof of Theorem 3.1. We prove $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1)$.

Implication $(1) \Rightarrow (2)$**.** Assume the welfare preorder is exactly identifiable from A on Π. Let $d, d' \in \Delta_\Pi$ satisfy

$$A_V d = A_V d'.$$

Choose policy pairs (π, π') and (ρ, ρ') such that

$$d = c^\pi - c^{\pi'}, \qquad d' = c^\rho - c^{\rho'}.$$

Under the Gaussian paired-difference model,

$$\mathsf{P}_{\pi,\pi'} = \Gamma_\Sigma(A_V d), \qquad \mathsf{P}_{\rho,\rho'} = \Gamma_\Sigma(A_V d').$$

Since $A_V d = A_V d'$, the two observation laws are identical. Exact identifiability therefore implies

$$\pi \succeq \pi' \iff \rho \succeq \rho',$$

which is equivalent to

$$d \in C_V \iff d' \in C_V.$$

Hence the fibers of A_V are order-monochromatic on Δ_Π.

Implication $(2) \Rightarrow (3)$**.** Assume the fibers of A_V are order-monochromatic. By Theorem B.4, there exists a unique map

$$\bar{\psi}_V : A_V(\Delta_\Pi) \to \{0,1\}$$

such that

$$\psi_V = \bar{\psi}_V \circ A_V \qquad \text{on } \Delta_\Pi.$$

Define

$$\overline{\mathcal{C}} := \bar{\psi}_V^{-1}(1).$$

By Theorem B.4, $\overline{\mathcal{C}} = A_V(C_V \cap \Delta_\Pi)$, hence $\overline{\mathcal{C}}$ is compact. Moreover, for every $d \in \Delta_\Pi$,

$$d \in C_V \iff \psi_V(d) = 1 \iff \bar{\psi}_V(A_V d) = 1 \iff A_V d \in \overline{\mathcal{C}}.$$

This is exactly statement (3).

Implication $(3) \Rightarrow (1)$**.** Assume there exists a compact set

$$\overline{\mathcal{C}} \subseteq A_V(\Delta_\Pi)$$

such that

$$d \in C_V \iff A_V d \in \overline{\mathcal{C}}, \qquad \forall\, d \in \Delta_\Pi.$$

Define the compact set of Gaussian laws

$$\mathcal{G}_{\overline{\mathcal{C}}} := \Gamma_\Sigma(\overline{\mathcal{C}}) = \{\mathcal{N}(y, \Sigma) : y \in \overline{\mathcal{C}}\} \subseteq \mathcal{P}(\mathbb{R}^q).$$

By Theorem B.6, $\mathcal{G}_{\overline{\mathcal{C}}}$ is compact, hence Borel. Define

$$\phi : \mathcal{P}(\mathbb{R}^q) \to \{0,1\}, \qquad \phi(\mu) := \mathbb{I}\{\mu \in \mathcal{G}_{\overline{\mathcal{C}}}\}.$$

Then ϕ is measurable.

Fix any policy pair (π, π'), and set

$$d := c^\pi - c^{\pi'} \in \Delta_\Pi.$$

Its observation law is

$$\mathsf{P}_{\pi,\pi'} = \Gamma_\Sigma(A_V d).$$

Therefore

$$\begin{aligned}
&\phi(\mathsf{P}_{\pi,\pi'}) = 1 \\
&\qquad \iff \mathsf{P}_{\pi,\pi'} \in \mathcal{G}_{\overline{\mathcal{C}}} \\
&\qquad \iff A_V d \in \overline{\mathcal{C}} \qquad \text{(by injectivity of } \Gamma_\Sigma \text{ on } A_V(\Delta_\Pi)) \\
&\qquad \iff d \in C_V \\
&\qquad \iff \pi \succeq \pi'.
\end{aligned}$$

Thus ϕ exactly identifies the welfare preorder from the observation law, which is precisely Theorem 2.1. This proves (1).

The final equivalent statement in Theorem 3.1 is exactly the map $\bar{\psi}_V$ supplied by Theorem B.4. □

Remark B.7 (Scope of the factorization theorem). The proof above uses only three structural inputs:

1. compactness of the attainable difference set Δ_Π;
2. injectivity of the map from feedback means $A_V d$ to observation laws;
3. the fact that the welfare label is binary.

Thus Theorem 3.1 extends verbatim to any observation family $\{\mathsf{P}_y : y \in A_V(\Delta_\Pi)\}$ for which $y \mapsto \mathsf{P}_y$ is a measurable embedding on the compact set $A_V(\Delta_\Pi)$. The Gaussian model is used in the main text because it yields a clean minimax statistical theory in Section 4.

C Proof of the facet observability theorem

This appendix gives the full proof of Theorem 3.4 and of its corollaries. The structure of the argument is intentionally modular. First, we isolate the linear-algebraic relation between unobservable directions and the kernel of the restricted feedback operator. Second, we combine that relation with the local-richness geometry of Appendix A to produce an *invisible sign flip across a facet*. This is the order-theoretic analogue of the invisible directions that arise in reward identifiability analyses [Skalse et al., 2023, Metelli et al., 2023], but the present argument works directly at the level of policy-order observability.

Throughout this appendix, we write

$$A_V := A\restriction_V \colon V \to \mathbb{R}^q$$

for the restriction of the fused feedback operator to the effective span $V = \mathrm{span}(\Delta_\Pi)$, and we work in the nondegenerate regime where C_V is full-dimensional in V.

C.1 Auxiliary linear-algebraic and statistical lemmas

Lemma C.1 (Row-space / kernel orthogonality)**.** *Let $T : V \to \mathbb{R}^q$ be linear, where V is a finite-dimensional Euclidean space. Then*

$$\mathrm{row}(T) = \mathrm{range}(T^*) = \ker(T)^\perp.$$

In particular, if $g \notin \mathrm{row}(T)$, then there exists $h \in \ker(T)$ such that

$$g^\top h \neq 0.$$

Proof. The identity

$$\mathrm{range}(T^*) = \ker(T)^\perp$$

is standard in finite-dimensional inner-product spaces. For completeness, we prove both inclusions.

If $u = T^* y \in \mathrm{range}(T^*)$, then for every $h \in \ker(T)$,

$$u^\top h = (T^* y)^\top h = y^\top T h = 0,$$

so $u \in \ker(T)^\perp$. Hence

$$\mathrm{range}(T^*) \subseteq \ker(T)^\perp.$$

Conversely, both spaces have the same dimension:

$$\dim \mathrm{range}(T^*) = \mathrm{rank}(T^*) = \mathrm{rank}(T),$$

while by rank–nullity,

$$\dim \ker(T)^\perp = \dim V - \dim \ker(T) = \mathrm{rank}(T).$$

Thus $\mathrm{range}(T^*)$ is a subspace of $\ker(T)^\perp$ with the same dimension, and therefore the two spaces are equal.

The final claim follows immediately: if $g \notin \mathrm{row}(T) = \ker(T)^\perp$, then g is not orthogonal to $\ker(T)$, so there exists $h \in \ker(T)$ with $g^\top h \neq 0$. □

Lemma C.2 (Equality of Gaussian paired-difference laws)**.** *Suppose*

$$Y_d \sim \mathcal{N}(A_V d, \Sigma), \qquad Y_{d'} \sim \mathcal{N}(A_V d', \Sigma), \qquad \Sigma \succ 0.$$

Then

$$\mathcal{L}(Y_d) = \mathcal{L}(Y_{d'}) \quad \Longleftrightarrow \quad A_V d = A_V d'.$$

Proof. If $A_V d = A_V d'$, then the two Gaussian laws are identical by definition.

Conversely, if $\mathcal{L}(Y_d) = \mathcal{L}(Y_{d'})$, then the two laws have the same expectation. Since

$$\mathbb{E}[Y_d] = A_V d, \qquad \mathbb{E}[Y_{d'}] = A_V d',$$

it follows that $A_V d = A_V d'$. □

Lemma C.3 (Observable certificates characterize cone membership)**.** *Suppose there exist vectors $u_1, \dots, u_r \in \mathbb{R}^q$ such that*

$$A_V^* u_j = w_j, \qquad j = 1, \dots, r.$$

Then for every $d \in V$,

$$d \in C_V \quad \Longleftrightarrow \quad u_j^\top A_V d \geq 0, \qquad j = 1, \dots, r.$$

In particular, the membership label $\mathbb{I}\{d \in C_V\}$ is a deterministic function of $A_V d$.

Proof. By Theorem A.10,

$$C_V = \{v \in V : w_j^\top v \geq 0, \;\; j = 1, \dots, r\}.$$

Hence, for $d \in V$,

$$d \in C_V \quad \Longleftrightarrow \quad w_j^\top d \geq 0, \;\; j = 1, \dots, r.$$

Using $A_V^* u_j = w_j$, we obtain

$$w_j^\top d = (A_V^* u_j)^\top d = u_j^\top A_V d.$$

Substituting this identity into the facet inequalities proves the claim. □

C.2 The invisible sign-flip lemma

The next lemma is the geometric core of the necessity proof. It shows that any unobservable facet normal gives rise to two attainable consequence differences that lie on opposite sides of the welfare boundary but induce exactly the same feedback.

Lemma C.4 (Invisible sign flip across a facet)**.** *Assume C_V is full-dimensional in V, and suppose Assumption 2.2(3) holds. Fix $j \in \{1, \dots, r\}$. If*

$$w_j \notin \mathrm{row}(A_V),$$

then there exist $d_+, d_- \in \Delta_\Pi$ such that

$$A_V d_+ = A_V d_-, \qquad d_+ \in \mathrm{relint}(C_V), \qquad d_- \notin C_V.$$

Proof. Since $w_j \notin \mathrm{row}(A_V)$, Theorem C.1 gives $h \in \ker(A_V)$ such that

$$w_j^\top h \neq 0.$$

Replacing h by $-h$ if necessary, we may assume

$$w_j^\top h > 0.$$

By Theorem A.11, Assumption 2.2(3) is equivalent to

$$\mathrm{relint}(F_j) \cap \mathrm{relint}(\Delta_\Pi) \neq \varnothing.$$

Choose

$$d_j \in \mathrm{relint}(F_j) \cap \mathrm{relint}(\Delta_\Pi).$$

Since $d_j \in \mathrm{relint}(\Delta_\Pi)$ relative to V, there exists $\eta_\Delta > 0$ such that

$$\{v \in V : \|v - d_j\|_2 < \eta_\Delta\} \subset \Delta_\Pi.$$

Since $d_j \in \mathrm{relint}(F_j)$, Theorem A.13 yields $\eta_C > 0$ such that, for all $g \in V$ with $\|g\|_2 < \eta_C$,

1. if $w_j^\top g > 0$, then
$$d_j + g \in \mathrm{relint}(C_V), \qquad d_j - g \notin C_V;$$
2. if $w_j^\top g < 0$, then
$$d_j - g \in \mathrm{relint}(C_V), \qquad d_j + g \notin C_V.$$

Choose

$$0 < t < \min\{\frac{\eta_\Delta}{\|h\|_2}, \frac{\eta_C}{\|h\|_2}\}.$$

Set

$$d_+ := d_j + th, \qquad d_- := d_j - th.$$

Then

$$\|d_\pm - d_j\|_2 = t\|h\|_2 < \eta_\Delta,$$

so $d_+, d_- \in \Delta_\Pi$. Moreover,

$$\|th\|_2 < \eta_C \qquad \text{and} \qquad w_j^\top(th) = t\, w_j^\top h > 0,$$

so Theorem A.13 implies

$$d_+ \in \operatorname{relint}(C_V), \qquad d_- \notin C_V.$$

Finally, because $h \in \ker(A_V)$,

$$A_V d_+ - A_V d_- = A_V(d_j + th) - A_V(d_j - th) = 2t\, A_V h = 0,$$

hence

$$A_V d_+ = A_V d_-.$$

This proves the claim. □

C.3 Full proof of Theorem 3.4

We now prove the main exact-identifiability theorem in full.

Proof of Theorem 3.4. We prove the cyclic chain

$$(2) \iff (3) \implies (1) \implies (2).$$

Equivalence of (2) and (3). By definition,

$$\operatorname{row}(A_V) = \operatorname{range}(A_V^*).$$

Hence

$$w_j \in \operatorname{row}(A_V) \quad \iff \quad \exists\, u_j \in \mathbb{R}^q \text{ such that } A_V^* u_j = w_j.$$

This proves $(2) \iff (3)$.

Implication $(3) \implies (1)$**.** Assume there exist $u_1, \ldots, u_r \in \mathbb{R}^q$ satisfying

$$A_V^* u_j = w_j, \qquad j = 1, \ldots, r.$$

Fix two policy pairs (π, π') and (ρ, ρ'), and define

$$d := c^\pi - c^{\pi'}, \qquad d' := c^\rho - c^{\rho'}.$$

Suppose the two observation laws are equal:

$$\mathsf{P}_{\pi,\pi'} = \mathsf{P}_{\rho,\rho'}.$$

Under the Gaussian paired-difference model, Theorem C.2 implies

$$A_V d = A_V d'.$$

By Theorem C.3,

$$d \in C_V \quad \iff \quad u_j^\top A_V d \geq 0, \;\; j = 1, \ldots, r,$$

and similarly for d'. Since $A_V d = A_V d'$, the right-hand sides are identical, so

$$d \in C_V \quad \iff \quad d' \in C_V.$$

Because membership in C_V is exactly the welfare label,

$$\pi \succeq \pi' \quad \iff \quad \rho \succeq \rho'.$$

By the equivalent characterization in Theorem 2.1, this proves exact identifiability.

Implication $(1) \Longrightarrow (2)$**.** Assume exact identifiability holds, but suppose for contradiction that

$$w_j \notin \mathrm{row}(A_V)$$

for some $j \in \{1, \dots, r\}$. By Theorem C.4, there exist $d_+, d_- \in \Delta_\Pi$ such that

$$A_V d_+ = A_V d_-, \qquad d_+ \in \mathrm{relint}(C_V), \qquad d_- \notin C_V.$$

Since $d_+, d_- \in \Delta_\Pi$, there exist policy pairs (π_+, π'_+) and (π_-, π'_-) such that

$$d_+ = c^{\pi_+} - c^{\pi'_+}, \qquad d_- = c^{\pi_-} - c^{\pi'_-}.$$

The equality $A_V d_+ = A_V d_-$, together with Theorem C.2, implies

$$\mathsf{P}_{\pi_+, \pi'_+} = \mathsf{P}_{\pi_-, \pi'_-}.$$

However,

$$d_+ \in C_V \qquad \text{and} \qquad d_- \notin C_V,$$

so the two policy pairs carry different welfare labels. This contradicts exact identifiability in the sense of Theorem 2.1. Therefore every facet normal must belong to $\mathrm{row}(A_V)$.

This completes the proof. □

C.4 Proofs of the corollaries

We now prove the two corollaries stated in the main text.

Proof of Theorem 3.6. Consider the one-state, one-step randomized model with four actions $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$, and let the consequence map be

$$B = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

If $p = (p_1, p_2, p_3, p_4) \in \Delta(\mathcal{A})$, then

$$Bp = (p_3 + p_4,\ p_2 + p_4).$$

Hence the consequence image is

$$\mathcal{K}_\Pi = [0, 1]^2,$$

and therefore

$$\Delta_\Pi = [-1, 1]^2, \qquad V = \mathbb{R}^2.$$

Let the welfare cone be the Pareto cone

$$C_V = \mathbb{R}^2_+ = \{v \in \mathbb{R}^2 : e_1^\top v \geq 0,\ e_2^\top v \geq 0\}.$$

Its two facet normals are

$$w_1 = e_1, \qquad w_2 = e_2.$$

The cone is full-dimensional in $V = \mathbb{R}^2$. Moreover, the local-richness assumption holds: for instance,

$$(0, 1/2) \in \mathrm{relint}(F_1) \cap \mathrm{relint}(\Delta_\Pi), \qquad (1/2, 0) \in \mathrm{relint}(F_2) \cap \mathrm{relint}(\Delta_\Pi),$$

where $F_1 = \{(0, y) : y \geq 0\}$ and $F_2 = \{(x, 0) : x \geq 0\}$.

Now define

$$A_1 = [1 \quad 0], \qquad A_2 = [0 \quad 1].$$

Then

$$\mathrm{row}(A_1) = \mathrm{span}(e_1), \qquad \mathrm{row}(A_2) = \mathrm{span}(e_2).$$

Hence $w_2 \notin \mathrm{row}(A_1)$ and $w_1 \notin \mathrm{row}(A_2)$. By Theorem 3.4, the welfare preorder is not exactly identifiable from A_1 alone and not exactly identifiable from A_2 alone.

For the fused operator

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2,$$

one has

$$\mathrm{row}(A) = \mathbb{R}^2.$$

Therefore both facet normals w_1, w_2 are observable, and another application of Theorem 3.4 shows that the welfare preorder is exactly identifiable from the fused system. □

Proof of Theorem 3.7. Let

$$L := \operatorname{span}\{w_1, \dots, w_r\} \subseteq V.$$

If the welfare preorder is exactly identifiable, then Theorem 3.4 implies

$$w_j \in \operatorname{row}(A_V) \qquad \text{for every } j = 1, \dots, r.$$

Therefore

$$L \subseteq \operatorname{row}(A_V).$$

Taking dimensions gives

$$\dim L \le \dim \operatorname{row}(A_V).$$

Finally,

$$\dim \operatorname{row}(A_V) = \operatorname{rank}(A_V),$$

which yields

$$\operatorname{rank}(A_V) \ge \dim L.$$

□

Remark C.5 (Interpretation of the rank lower bound). Theorem 3.7 is purely algebraic but already has a concrete meaning for RLxF design. The quantity $\dim L$ is the number of linearly independent order-separating directions that the welfare cone induces on the attainable difference span. Exact identifiability is impossible unless the fused feedback system exposes at least that many independent directions.

D Statistical lower bounds

This appendix proves the lower-bound half of Section 4. The argument has three components. First, we solve the quadratic programs underlying the observable facet certificates and the least-observable crossing directions. Second, we show that each facet admits a pair of γ-separated local alternatives that remain inside the attainable difference body Δ_Π. Third, we reduce the resulting two-point problem to an exact Gaussian hypothesis test, which yields the minimax lower bound in closed form.

Throughout, we assume the hypotheses of Theorem 3.4. In particular, every facet normal w_j lies in $\operatorname{row}(A_V) = \operatorname{range}(A_V^*)$, and we write

$$G := A_V^* \Sigma^{-1} A_V : V \to V.$$

D.1 Quadratic variational formulas

The matrix/operator G plays the role of the observable Fisher geometry on the effective ordered space V. The following lemma collects its basic algebraic properties.

Lemma D.1 (Observable Gram geometry)**.** *The operator G is self-adjoint and positive semidefinite on V. Moreover,*

$$\ker(G) = \ker(A_V), \qquad \operatorname{range}(G) = \operatorname{range}(A_V^*) = \operatorname{row}(A_V).$$

In particular, if exact identifiability holds, then

$$w_j \in \operatorname{range}(G), \qquad j = 1, \dots, r.$$

Proof. Self-adjointness follows from

$$G^* = (A_V^* \Sigma^{-1} A_V)^* = A_V^* \Sigma^{-1} A_V = G,$$

because Σ^{-1} is symmetric. Positive semidefiniteness follows from

$$h^\top G h = h^\top A_V^* \Sigma^{-1} A_V h = (A_V h)^\top \Sigma^{-1} (A_V h) \ge 0, \qquad \forall\, h \in V,$$

since $\Sigma^{-1} \succ 0$.

Now $h \in \ker(G)$ if and only if

$$0 = h^\top G h = (A_V h)^\top \Sigma^{-1} (A_V h),$$

which, by positive definiteness of Σ^{-1}, is equivalent to $A_V h = 0$. Hence

$$\ker(G) = \ker(A_V).$$

By Theorem C.1,

$$\operatorname{range}(G) = \ker(G)^\perp = \ker(A_V)^\perp = \operatorname{range}(A_V^*).$$

The final claim then follows from Theorem 3.4. □

We now prove the certificate formula stated in the main text.

Proof of Theorem 4.1. Fix a facet j. Since $w_j \in \operatorname{range}(G)$ by Theorem D.1, the vector

$$u_j^\star := \Sigma^{-1} A_V G^\dagger w_j$$

is well-defined. First,

$$A_V^* u_j^\star = A_V^* \Sigma^{-1} A_V G^\dagger w_j = G G^\dagger w_j = w_j,$$

because $GG^\dagger$ is the orthogonal projector onto $\operatorname{range}(G)$, which contains w_j.

Now let $u \in \mathbb{R}^q$ be any feasible vector satisfying $A_V^* u = w_j$. Since $\Sigma \succ 0$,

$$0 \le \left\| \Sigma^{1/2} u - \Sigma^{-1/2} A_V G^\dagger w_j \right\|_2^2.$$

Expanding gives

$$\begin{aligned} 0 &\le u^\top \Sigma u - 2u^\top A_V G^\dagger w_j + w_j^\top G^\dagger A_V^* \Sigma^{-1} A_V G^\dagger w_j \\ &= u^\top \Sigma u - 2(A_V^* u)^\top G^\dagger w_j + w_j^\top G^\dagger G G^\dagger w_j \\ &= u^\top \Sigma u - 2 w_j^\top G^\dagger w_j + w_j^\top G^\dagger w_j \\ &= u^\top \Sigma u - w_j^\top G^\dagger w_j. \end{aligned}$$

Therefore every feasible u satisfies

$$u^\top \Sigma u \ge w_j^\top G^\dagger w_j.$$

Equality holds for $u = u_j^\star$, because the squared norm above vanishes identically at $u_j^\star$. Hence $u_j^\star$ is optimal and the optimal value is

$$\sigma_j^2 := w_j^\top G^\dagger w_j.$$

Finally, uniqueness follows because the feasible set $\{u \in \mathbb{R}^q : A_V^* u = w_j\}$ is affine and the objective $u \mapsto u^\top \Sigma u$ is strictly convex when $\Sigma \succ 0$. □

The next variational problem produces the canonical least-observable direction that crosses a given facet by one unit.

Proposition D.2 (Canonical least-observable unit crossing)**.** *For each facet j, define*

$$h_j^\star := \frac{G^\dagger w_j}{w_j^\top G^\dagger w_j} = \frac{G^\dagger w_j}{\sigma_j^2}.$$

Then:

1. $w_j^\top h_j^\star = 1$*;*
2. $h_j^\star$ *solves*
$$\min_{h \in V} h^\top G h \qquad \textit{subject to} \qquad w_j^\top h = 1; \tag{D.1}$$
3. *every feasible h in* (D.1) *satisfies*
$$h^\top G h \ge \frac{1}{\sigma_j^2},$$
and the optimal value is exactly $1/\sigma_j^2$;

4. *the full set of minimizers is*

$$h_j^\star + \ker(G) = h_j^\star + \ker(A_V).$$

Proof. Because $w_j \in \mathrm{range}(G)$, one has

$$GG^\dagger w_j = w_j.$$

Hence

$$w_j^\top h_j^\star = \frac{w_j^\top G^\dagger w_j}{\sigma_j^2} = 1,$$

proving feasibility.

Now let $h \in V$ satisfy $w_j^\top h = 1$. Since G is self-adjoint and

$$Gh_j^\star = \frac{GG^\dagger w_j}{\sigma_j^2} = \frac{w_j}{\sigma_j^2},$$

we obtain

$$\begin{aligned}
(h - h_j^\star)^\top G(h - h_j^\star) &= h^\top Gh - 2h^\top Gh_j^\star + h_j^{\star\top} Gh_j^\star \\
&= h^\top Gh - \frac{2}{\sigma_j^2} h^\top w_j + \frac{1}{\sigma_j^4} w_j^\top G^\dagger GG^\dagger w_j \\
&= h^\top Gh - \frac{2}{\sigma_j^2} + \frac{1}{\sigma_j^4} w_j^\top G^\dagger w_j \\
&= h^\top Gh - \frac{1}{\sigma_j^2}.
\end{aligned}$$

Since $G \succeq 0$, the left-hand side is nonnegative, so

$$h^\top Gh \geq \frac{1}{\sigma_j^2}.$$

At $h = h_j^\star$, equality holds, proving optimality and the optimal value.

Finally, equality occurs if and only if

$$(h - h_j^\star)^\top G(h - h_j^\star) = 0,$$

which is equivalent to $h - h_j^\star \in \ker(G)$. By Theorem D.1, $\ker(G) = \ker(A_V)$. □

Remark D.3 (Interpretation of $h_j^\star$). The direction $h_j^\star$ is the least-observable unit crossing of the j-th facet. Among all perturbations that move one unit in the normal direction w_j, it minimizes the observable quadratic energy $h^\top Gh$, equivalently the Mahalanobis norm of the induced mean shift $A_V h$ under noise covariance Σ.

D.2 Margin-separated local alternatives

We now justify the γ-local alternative construction from the main text.

Fix a facet j. By Assumption 2.2(3), there exists

$$d_j^\circ \in \mathrm{relint}(F_j) \cap \mathrm{relint}(\Delta_\Pi),$$

where

$$F_j := \{v \in C_V : w_j^\top v = 0\}.$$

Set

$$a_{j\ell} := w_\ell^\top d_j^\circ, \qquad b_{j\ell} := w_\ell^\top h_j^\star, \qquad \ell = 1, \ldots, r.$$

Then

$$a_{jj} = 0, \qquad a_{j\ell} > 0 \text{ for all } \ell \neq j,$$

because $d_j^\circ \in \mathrm{relint}(F_j)$.

Lemma D.4 (Facet-centered γ-alternatives)**.** *For each facet j, there exists a constant $\bar{\gamma}_j > 0$ such that, for every $0 < \gamma \le \bar{\gamma}_j$, the two points*

$$d_{j,+}^{(\gamma)} := d_j^\circ + \gamma h_j^\star, \qquad d_{j,-}^{(\gamma)} := d_j^\circ - \gamma h_j^\star$$

satisfy:

1. $d_{j,+}^{(\gamma)}, d_{j,-}^{(\gamma)} \in \Delta_\Pi$;
2. $d_{j,+}^{(\gamma)} \in C_V$ *and*
$$\min_{1 \le \ell \le r} w_\ell^\top d_{j,+}^{(\gamma)} \ge \gamma;$$
3. $d_{j,-}^{(\gamma)} \notin C_V$ *and*
$$w_j^\top d_{j,-}^{(\gamma)} = -\gamma.$$

In particular,

$$d_{j,+}^{(\gamma)},\, d_{j,-}^{(\gamma)} \in \mathcal{D}_\gamma.$$

Proof. Because $d_j^\circ \in \operatorname{relint}(\Delta_\Pi)$, there exists $\rho_j > 0$ such that

$$\{v \in V : \|v - d_j^\circ\|_2 < \rho_j\} \subset \Delta_\Pi.$$

Define

$$\gamma_{j,\Delta} := \frac{\rho_j}{\|h_j^\star\|_2}.$$

Then for every $0 < \gamma < \gamma_{j,\Delta}$,

$$\|d_{j,\pm}^{(\gamma)} - d_j^\circ\|_2 = \gamma \|h_j^\star\|_2 < \rho_j,$$

hence $d_{j,\pm}^{(\gamma)} \in \Delta_\Pi$.

We next enforce the margin conditions on the positive-side alternative. For the active facet j,

$$w_j^\top d_{j,+}^{(\gamma)} = w_j^\top d_j^\circ + \gamma w_j^\top h_j^\star = 0 + \gamma = \gamma.$$

For any $\ell \ne j$,

$$w_\ell^\top d_{j,+}^{(\gamma)} = a_{j\ell} + \gamma b_{j\ell}.$$

If $b_{j\ell} \ge 1$, then

$$a_{j\ell} + \gamma b_{j\ell} \ge a_{j\ell} + \gamma > \gamma,$$

since $a_{j\ell} > 0$. If $b_{j\ell} < 1$, then

$$a_{j\ell} + \gamma b_{j\ell} \ge \gamma \quad \Longleftrightarrow \quad \gamma \le \frac{a_{j\ell}}{1 - b_{j\ell}}.$$

Define

$$\gamma_{j,\mathrm{marg}} := \min_{\ell \ne j:\, b_{j\ell} < 1} \frac{a_{j\ell}}{1 - b_{j\ell}},$$

with the convention that the minimum over the empty set is $+\infty$. Since each $a_{j\ell} > 0$, one has $\gamma_{j,\mathrm{marg}} > 0$. Therefore, for all $0 < \gamma \le \gamma_{j,\mathrm{marg}}$,

$$w_\ell^\top d_{j,+}^{(\gamma)} \ge \gamma, \qquad \ell = 1, \dots, r.$$

Thus $d_{j,+}^{(\gamma)} \in C_V$ and has margin at least γ on every facet.

Finally,

$$w_j^\top d_{j,-}^{(\gamma)} = w_j^\top d_j^\circ - \gamma w_j^\top h_j^\star = -\gamma < 0,$$

so $d_{j,-}^{(\gamma)} \notin C_V$, and the violated j-th facet has magnitude exactly γ.

Setting

$$\bar{\gamma}_j := \min\{\gamma_{j,\Delta},\, \gamma_{j,\mathrm{marg}}\}$$

proves all claims. □

D.3 Exact risk of the Gaussian two-point problem

We now solve the binary testing problem induced by the local alternatives from Theorem D.4. The result is exact and avoids any loss from generic information-theoretic inequalities.

Lemma D.5 (Exact minimax risk for a symmetric Gaussian pair)**.** *Let $m, a \in \mathbb{R}^q$, let $\Sigma \succ 0$, and let $\gamma > 0$. Consider the two simple hypotheses*

$$H_+ : \quad Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m + \gamma a, \Sigma),$$
$$H_- : \quad Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m - \gamma a, \Sigma).$$

Then

$$\inf_{\phi:(\mathbb{R}^q)^n \to \{0,1\}} \max\Big\{\mathbb{P}_+(\phi = 0), \mathbb{P}_-(\phi = 1)\Big\} = \Phi\Big(-\gamma\sqrt{n\, a^\top \Sigma^{-1} a}\Big),$$

where $\phi = 1$ means acceptance of H_+. The optimum is attained by the likelihood-ratio threshold test

$$\phi^\star(Y_1, \dots, Y_n) = \mathbb{I}\{a^\top \Sigma^{-1}(\bar{Y}_n - m) \geq 0\}.$$

Proof. The joint log-likelihood ratio is

$$\begin{aligned}
\log \frac{d\mathbb{P}_+}{d\mathbb{P}_-}(Y_1, \dots, Y_n) &= -\frac{1}{2}\sum_{k=1}^n \Big[(Y_k - m - \gamma a)^\top \Sigma^{-1}(Y_k - m - \gamma a) \\
&\qquad - (Y_k - m + \gamma a)^\top \Sigma^{-1}(Y_k - m + \gamma a)\Big] \\
&= 2\gamma \sum_{k=1}^n a^\top \Sigma^{-1}(Y_k - m) \\
&= 2n\gamma\, a^\top \Sigma^{-1}(\bar{Y}_n - m).
\end{aligned}$$

Hence the Neyman–Pearson optimal test is

$$\phi^\star = \mathbb{I}\{a^\top \Sigma^{-1}(\bar{Y}_n - m) \geq 0\}.$$

Define the scalar statistic

$$T_n := \frac{a^\top \Sigma^{-1}(\bar{Y}_n - m)}{a^\top \Sigma^{-1} a}.$$

Under $H_\pm$, since

$$\bar{Y}_n \sim \mathcal{N}\Big(m \pm \gamma a, \frac{1}{n}\Sigma\Big),$$

we have

$$T_n \sim \mathcal{N}\Big(\pm\gamma, \frac{1}{n\, a^\top \Sigma^{-1} a}\Big).$$

Therefore, under H_+,

$$\mathbb{P}_+(\phi^\star = 0) = \mathbb{P}_+(T_n < 0) = \Phi\Big(-\gamma\sqrt{n\, a^\top \Sigma^{-1} a}\Big),$$

and symmetrically under H_-,

$$\mathbb{P}_-(\phi^\star = 1) = \mathbb{P}_-(T_n \geq 0) = \Phi\Big(-\gamma\sqrt{n\, a^\top \Sigma^{-1} a}\Big).$$

Thus the worst-case error of $\phi^\star$ equals the claimed value.

To prove optimality in the minimax sense, note that for any test ϕ,

$$\max\{\mathbb{P}_+(\phi = 0), \mathbb{P}_-(\phi = 1)\} \geq \frac{1}{2}\Big(\mathbb{P}_+(\phi = 0) + \mathbb{P}_-(\phi = 1)\Big).$$

The right-hand side is the Bayes risk under equal priors. By the Neyman–Pearson lemma, $\phi^\star$ minimizes this Bayes risk among all tests for the simple-vs-simple problem. Since $\phi^\star$ also equalizes the two error probabilities by symmetry, it minimizes the worst-case error as well. □

Remark D.6 (Equivalent Mahalanobis form). If $\Delta := (m + \gamma a) - (m - \gamma a) = 2\gamma a$ and $S := \Sigma/n$, then the exact minimax error in Theorem D.5 can be written as

$$\Phi\Big(-\frac{1}{2}\sqrt{\Delta^\top S^{-1} \Delta}\Big).$$

Thus the difficulty of the two-point problem is governed exactly by the Mahalanobis separation of the two means.

D.4 Proof of the minimax lower bound

We now prove the main lower bound from Section 4.

Proof of Theorem 4.2. Fix $0 < \gamma \le \bar{\gamma}$, where

$$\bar{\gamma} := \min_{1 \le j \le r} \bar{\gamma}_j$$

and the constants $\bar{\gamma}_j$ are given by Theorem D.4. Let

$$j_\star \in \arg\max_{1 \le j \le r} \sigma_j, \qquad \sigma_\star := \sigma_{j_\star}.$$

Define the two alternatives

$$d_+ := d^{(\gamma)}_{j_\star,+} = d^\circ_{j_\star} + \gamma h^\star_{j_\star}, \qquad d_- := d^{(\gamma)}_{j_\star,-} = d^\circ_{j_\star} - \gamma h^\star_{j_\star}.$$

By Theorem D.4,

$$d_+, d_- \in \mathcal{D}_\gamma, \qquad \psi_V(d_+) = 1, \qquad \psi_V(d_-) = 0.$$

Under a generic difference d, the observation model is

$$Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(A_V d, \Sigma).$$

Thus, under the two alternatives $d_\pm$,

$$Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(A_V d^\circ_{j_\star} \pm \gamma A_V h^\star_{j_\star}, \Sigma).$$

Apply Theorem D.5 with

$$m = A_V d^\circ_{j_\star}, \qquad a = A_V h^\star_{j_\star}.$$

This yields

$$\inf_\phi \max\left\{\mathbb{P}_{d_+}(\phi = 0), \mathbb{P}_{d_-}(\phi = 1)\right\} = \Phi\Big(-\gamma\sqrt{n\, h^{\star\top}_{j_\star} G h^\star_{j_\star}}\Big).$$

By Theorem D.2,

$$h^{\star\top}_{j_\star} G h^\star_{j_\star} = \frac{1}{\sigma^2_{j_\star}} = \frac{1}{\sigma^2_\star}.$$

Hence

$$\inf_\phi \max\left\{\mathbb{P}_{d_+}(\phi = 0), \mathbb{P}_{d_-}(\phi = 1)\right\} = \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_\star}\Big).$$

Now let $\hat{\psi}_n : (\mathbb{R}^q)^n \to \{0, 1\}$ be any measurable test for the welfare label. Since $\{d_+, d_-\} \subseteq \mathcal{D}_\gamma$,

$$\sup_{d \in \mathcal{D}_\gamma} \mathbb{P}_d\big(\hat{\psi}_n \ne \psi_V(d)\big) \ge \max\Big\{\mathbb{P}_{d_+}\big(\hat{\psi}_n = 0\big), \mathbb{P}_{d_-}\big(\hat{\psi}_n = 1\big)\Big\}.$$

Taking the infimum over all measurable $\hat{\psi}_n$ and using the exact two-point minimax value above yields

$$\inf_{\hat{\psi}_n} \sup_{d \in \mathcal{D}_\gamma} \mathbb{P}_d\big(\hat{\psi}_n \ne \psi_V(d)\big) \ge \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_\star}\Big).$$

This proves the theorem. □

Remark D.7 (Why the lower bound is exact). The proof does not use a generic information inequality such as Le Cam or Fano; instead it embeds an exactly solvable two-point Gaussian testing problem into the class $\mathcal{D}_\gamma$. As a result, the lower bound is exact at the level of the Gaussian cdf and immediately matches the upper bound of the observable cone test up to the logarithmic multiplicity factor incurred by simultaneously enforcing all facet inequalities.

E Statistical upper bounds

This appendix proves the constructive half of Section 4. The main object is the observable cone test, which applies the minimum-variance observable facet certificates $u_1^\star, \dots, u_r^\star$ from Theorem 4.1 to the sample mean. The proof has two layers. First, we derive the exact Gaussian law of the certificate statistics. Second, we bound the two one-sided errors by exploiting the cone structure: false negatives inside the cone are controlled by a union bound over all facets, while false positives outside the cone are controlled by a single violated facet.

Throughout, we assume the hypotheses of Theorems 3.4 and 4.3, and we write

$$\hat{\psi}_n^{\mathrm{OCT}} = \mathbb{I}\{u_j^{\star\top}\bar{Y}_n \geq 0, \;\; j = 1, \dots, r\}.$$

E.1 Certificate statistics under the Gaussian experiment

We first make explicit the Gaussian structure induced by the observable facet certificates.

Lemma E.1 (Certificate statistics are jointly Gaussian)**.** *Fix $d \in \Delta_\Pi$. Define, for $j = 1, \dots, r$,*

$$T_{j,n} := u_j^{\star\top}\bar{Y}_n.$$

Then the vector

$$T_n := (T_{1,n}, \dots, T_{r,n})^\top \in \mathbb{R}^r$$

is Gaussian with mean vector

$$m(d) := \big(w_1^\top d, \dots, w_r^\top d\big)^\top$$

and covariance matrix

$$\frac{1}{n}\Omega, \qquad \Omega_{jk} := u_j^{\star\top}\Sigma u_k^\star.$$

In particular, each marginal satisfies

$$T_{j,n} \sim \mathcal{N}\Big(w_j^\top d, \frac{\sigma_j^2}{n}\Big), \qquad j = 1, \dots, r.$$

Proof. Since

$$\bar{Y}_n \sim \mathcal{N}\Big(A_V d, \frac{1}{n}\Sigma\Big),$$

any linear image of $\bar{Y}_n$ is Gaussian. Therefore T_n is jointly Gaussian. Its mean is

$$\mathbb{E}_d[T_{j,n}] = u_j^{\star\top}\mathbb{E}_d[\bar{Y}_n] = u_j^{\star\top}A_V d = d^\top A_V^* u_j^\star = w_j^\top d,$$

because $A_V^* u_j^\star = w_j$ by Theorem 4.1. Its covariance matrix is

$$\mathrm{Cov}_d(T_n) = \frac{1}{n}\begin{bmatrix} u_1^{\star\top}\Sigma u_1^\star & \cdots & u_1^{\star\top}\Sigma u_r^\star \\ \vdots & \ddots & \vdots \\ u_r^{\star\top}\Sigma u_1^\star & \cdots & u_r^{\star\top}\Sigma u_r^\star \end{bmatrix} = \frac{1}{n}\Omega.$$

Taking the diagonal entries and using $u_j^{\star\top}\Sigma u_j^\star = \sigma_j^2$ from Theorem 4.1 yields the marginal law. □

Remark E.2 (Only marginal Gaussian tails are needed). The coordinates $T_{1,n}, \dots, T_{r,n}$ need not be independent. The upper bounds below use only one-dimensional Gaussian tail probabilities and the union bound, so no independence assumption is needed.

E.2 Pointwise risk bounds for the observable cone test

The next proposition gives a pointwise version of the main-text theorem. It is strictly stronger than Theorem 4.3, which follows by specializing to the margin class $\mathcal{D}_\gamma$.

Proposition E.3 (Pointwise one-sided error bounds)**.** *Fix $d \in \Delta_\Pi$.*

1. *If $d \in C_V$, then*

$$\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 0\big) \leq \sum_{j=1}^r \Phi\Big(-\frac{\sqrt{n}\, w_j^\top d}{\sigma_j}\Big). \tag{E.1}$$

2. *If $d \notin C_V$, then*

$$\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 1\big) \le \min_{j:\, w_j^\top d<0} \Phi\Big(\frac{\sqrt{n}\, w_j^\top d}{\sigma_j}\Big). \tag{E.2}$$

Proof. Assume first that $d \in C_V$. By the facet representation of C_V,

$$w_j^\top d \ge 0, \qquad j = 1, \dots, r.$$

The event $\hat{\psi}_n^{\mathrm{OCT}} = 0$ means that at least one facet statistic is negative:

$$\{\hat{\psi}_n^{\mathrm{OCT}} = 0\} = \bigcup_{j=1}^{r} \{T_{j,n} < 0\}.$$

Therefore, by the union bound and Theorem E.1,

$$\begin{aligned}\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 0\big) &\le \sum_{j=1}^{r} \mathbb{P}_d(T_{j,n} < 0) \\ &= \sum_{j=1}^{r} \Phi\Big(-\frac{\sqrt{n}\, w_j^\top d}{\sigma_j}\Big),\end{aligned}$$

which proves (E.1).

Now assume $d \notin C_V$. Then at least one facet is violated, i.e. the set

$$J_-(d) := \{j \in \{1, \dots, r\} : w_j^\top d < 0\}$$

is nonempty. If $\hat{\psi}_n^{\mathrm{OCT}} = 1$, then all facet statistics are nonnegative, so in particular

$$T_{j,n} \ge 0 \qquad \forall j \in J_-(d).$$

Hence, for each $j \in J_-(d)$,

$$\{\hat{\psi}_n^{\mathrm{OCT}} = 1\} \subseteq \{T_{j,n} \ge 0\},$$

and therefore

$$\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 1\big) \le \mathbb{P}_d(T_{j,n} \ge 0).$$

Using Theorem E.1,

$$\mathbb{P}_d(T_{j,n} \ge 0) = \Phi\Big(\frac{\sqrt{n}\, w_j^\top d}{\sigma_j}\Big).$$

Since this holds for every $j \in J_-(d)$, taking the minimum over violated facets proves (E.2). □

Remark E.4 (Asymmetry of the two one-sided errors). The false-negative error inside the cone is inherently a simultaneous event: every facet inequality must be satisfied, so a union bound over all facets is natural. In contrast, a false positive outside the cone requires missing at least one violated facet, so one violated facet suffices to control the risk. This asymmetry explains why the outside-cone bound carries no factor r.

E.3 Proof of the upper bound theorem

We now deduce the main-text upper bound from the pointwise bounds.

Proof of Theorem 4.3. Fix $0 < \gamma \le \bar{\gamma}$.

Inside-cone alternatives. Let $d \in \mathcal{D}_\gamma \cap C_V$. By definition of $\mathcal{D}_\gamma$,

$$w_j^\top d \ge \gamma, \qquad j = 1, \dots, r.$$

Applying Theorem E.3,

$$\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 0\big) \le \sum_{j=1}^{r} \Phi\Big(-\frac{\sqrt{n}\, w_j^\top d}{\sigma_j}\Big) \le \sum_{j=1}^{r} \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_j}\Big),$$

because Φ is monotone increasing.

Taking the supremum over $d \in \mathcal{D}_\gamma \cap C_V$ yields

$$\sup_{d \in \mathcal{D}_\gamma \cap C_V} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 0\big) \le \sum_{j=1}^{r} \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_j}\Big). \tag{E.3}$$

Outside-cone alternatives. Let $d \in \mathcal{D}_\gamma \setminus C_V$. By definition of $\mathcal{D}_\gamma$, there exists at least one violated facet $j(d) \in \{1, \dots, r\}$ such that

$$w_{j(d)}^\top d \le -\gamma.$$

Applying Theorem E.3,

$$\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 1\big) \le \Phi\Big(\frac{\sqrt{n}\, w_{j(d)}^\top d}{\sigma_{j(d)}}\Big) \le \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_{j(d)}}\Big) \le \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big),$$

where

$$\sigma_* := \max_{1 \le j \le r} \sigma_j,$$

and the last step uses $\sigma_{j(d)} \le \sigma_*$ together with monotonicity of Φ.

Taking the supremum over $d \in \mathcal{D}_\gamma \setminus C_V$ gives

$$\sup_{d \in \mathcal{D}_\gamma \setminus C_V} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} = 1\big) \le \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big). \tag{E.4}$$

Uniform bound. Combining (E.3) and (E.4) yields the one-sided bounds stated in the theorem. Since

$$\Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big) \le \sum_{j=1}^{r} \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_j}\Big),$$

the worst-case error over the full class $\mathcal{D}_\gamma$ satisfies

$$\sup_{d \in \mathcal{D}_\gamma} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{OCT}} \ne \psi_V(d)\big) \le \sum_{j=1}^{r} \Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_j}\Big) \le r\,\Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big).$$

This proves (4.8).

Finally, the standard Gaussian tail inequality

$$\Phi(-x) \le e^{-x^2/2}, \qquad x \ge 0,$$

implies

$$r\,\Phi\Big(-\frac{\sqrt{n}\,\gamma}{\sigma_*}\Big) \le r\,\exp\Big(-\frac{n\gamma^2}{2\sigma_*^2}\Big).$$

Thus the right-hand side is at most δ whenever

$$n \ge \frac{2\sigma_*^2}{\gamma^2} \log \frac{r}{\delta},$$

which is exactly (4.9). □

Remark E.5 (When the upper bound is exact). If $r = 1$, or more generally if the active violated facet is known in advance, the observable cone test reduces to a single Gaussian threshold test and attains the lower bound of Theorem 4.2 exactly. The only gap between the minimax lower bound and Theorem 4.3 comes from simultaneously protecting against all r facet inequalities.

E.4 Proof of finite-policy dominance-graph recovery

We now prove the finite-policy corollary.

Proof of Theorem 4.4. For each ordered pair (a, b) with $a \ne b$, define the pairwise difference

$$d_{ab} := c^{\pi_a} - c^{\pi_b}.$$

By assumption,

$$d_{ab} \in \mathcal{D}_\gamma \qquad \forall\, a \ne b.$$

Apply the observable cone test to the n paired-difference samples collected for each ordered pair. By Theorem 4.3, each pairwise test has error probability at most

$$r\,\exp\Big(-\frac{n\gamma^2}{2\sigma_*^2}\Big).$$

There are exactly $N(N-1)$ ordered pairs (a,b) with $a \neq b$. Therefore a union bound gives

$$\mathbb{P}\big(\text{at least one pairwise decision is wrong}\big) \leq N(N-1)\,r\,\exp\Big(-\frac{n\gamma^2}{2\sigma_*^2}\Big).$$

Thus the full pairwise welfare relation is recovered with probability at least $1-\delta$ whenever

$$N(N-1)\,r\,\exp\Big(-\frac{n\gamma^2}{2\sigma_*^2}\Big) \leq \delta,$$

which is equivalent to

$$n \geq \frac{2\sigma_*^2}{\gamma^2}\log\frac{rN(N-1)}{\delta}.$$

If the full pairwise welfare relation on Π_N is recovered exactly, then the maximal-policy set is recovered exactly as well, since maximality is a deterministic function of the pairwise relation. □

F Approximate observability and robustness

The exact-identifiability theory in Section 3 is sharp, but it is also brittle: it requires every facet normal w_j to lie exactly in $\mathrm{row}(A_V)$. In practice, one may only have *approximate* observable certificates. This appendix shows that the entire testing theory degrades gracefully in that regime. The only price paid is a deterministic bias term that shrinks the effective margin.

The results below are deliberately stated in a way that isolates the geometry of approximate observability from the distributional assumptions on the noise. First, we develop a deterministic bias calculus for approximate certificates. Second, we derive Gaussian upper bounds. Third, we extend those bounds to a general sub-Gaussian observation model.

F.1 Approximate observable certificates

Fix arbitrary vectors $u_1,\dots,u_r \in \mathbb{R}^q$, not necessarily satisfying $A_V^* u_j = w_j$. Define the residuals

$$r_j := w_j - A_V^* u_j \in V, \qquad j=1,\dots,r. \tag{F.1}$$

Since $\Delta_\Pi \subset V$ is compact and $r_j^\top d$ is continuous in d, the quantity

$$\varepsilon_j := \sup_{d\in\Delta_\Pi} |r_j^\top d| \tag{F.2}$$

is finite and attained. We interpret ε_j as the worst-case deterministic bias of the j-th approximate certificate over the attainable difference set. Also define the noise scale

$$\tau_j^2 := u_j^\top \Sigma u_j, \qquad j=1,\dots,r. \tag{F.3}$$

Finally, let

$$\varepsilon_* := \max_{1\leq j\leq r}\varepsilon_j, \qquad \tau_* := \max_{1\leq j\leq r}\tau_j.$$

Given $u_1,\dots,u_r$, define the *approximate observable cone test*

$$\hat{\psi}_n^{\mathrm{aOCT}} := \mathbb{I}\{u_j^\top \bar{Y}_n \geq 0,\ \ j=1,\dots,r\}. \tag{F.4}$$

The next proposition gives the exact deterministic relation between the approximate certificate scores and the true facet functionals.

Proposition F.1 (Deterministic bias bounds)**.** *For every $d \in \Delta_\Pi$ and every $j \in \{1, \dots, r\}$,*

$$u_j^\top A_V d = w_j^\top d - r_j^\top d. \tag{F.5}$$

Consequently,

$$|u_j^\top A_V d - w_j^\top d| \le \varepsilon_j, \qquad d \in \Delta_\Pi. \tag{F.6}$$

In particular:

1. *if $w_j^\top d \ge \gamma$, then*
$$u_j^\top A_V d \ge \gamma - \varepsilon_j;$$
2. *if $w_j^\top d \le -\gamma$, then*
$$u_j^\top A_V d \le -\gamma + \varepsilon_j.$$

Proof. By definition of r_j,

$$A_V^* u_j = w_j - r_j.$$

Hence, for every $d \in \Delta_\Pi \subset V$,

$$u_j^\top A_V d = d^\top A_V^* u_j = d^\top (w_j - r_j) = w_j^\top d - r_j^\top d,$$

proving (F.5). Taking absolute values and using the definition of ε_j gives (F.6). The final two claims are immediate. □

Remark F.2 (Support-function interpretation). Because Δ_Π is a compact centrally symmetric convex body in V, ε_j can be written as the support function of Δ_Π:

$$\varepsilon_j = \sup_{d \in \Delta_\Pi} r_j^\top d.$$

Thus approximate observability is controlled by how large the residual direction r_j is when measured against the attainable consequence-difference geometry.

F.2 Gaussian robustness bounds

We first work under the Gaussian model of the main text. For $j = 1, \dots, r$, define the approximate certificate statistics

$$S_{j,n} := u_j^\top \bar{Y}_n.$$

Lemma F.3 (Gaussian law of the approximate certificate statistics)**.** *Under*

$$Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(A_V d, \Sigma),$$

one has

$$S_{j,n} \sim \mathcal{N}\Big(u_j^\top A_V d, \frac{\tau_j^2}{n}\Big), \qquad j = 1, \dots, r.$$

Proof. Since

$$\bar{Y}_n \sim \mathcal{N}\Big(A_V d, \frac{1}{n}\Sigma\Big),$$

the scalar $S_{j,n} = u_j^\top \bar{Y}_n$ is Gaussian with mean $u_j^\top A_V d$ and variance $u_j^\top \Sigma u_j / n = \tau_j^2 / n$. □

The next theorem is the main robustness statement for the Gaussian model.

Theorem F.4 (Robust Gaussian upper bound under approximate observability)**.** *Assume the Gaussian paired-difference model, and let $0 < \gamma \le \bar{\gamma}$ satisfy*

$$\gamma > \varepsilon_*.$$

Then the approximate observable cone test (F.4) *satisfies:*

$$\sup_{d \in \mathcal{D}_\gamma \cap C_V} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} = 0\big) \le \sum_{j=1}^r \Phi\Big(-\frac{\sqrt{n}\,(\gamma - \varepsilon_j)}{\tau_j}\Big), \tag{F.7}$$

$$\sup_{d \in \mathcal{D}_\gamma \setminus C_V} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} = 1\big) \le \Phi\Big(-\frac{\sqrt{n}\,(\gamma - \varepsilon_*)}{\tau_*}\Big). \tag{F.8}$$

Hence

$$\sup_{d\in\mathcal{D}_\gamma} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} \neq \psi_V(d)\big) \le \max\left\{\sum_{j=1}^{r} \Phi\Big(-\frac{\sqrt{n}\,(\gamma-\varepsilon_j)}{\tau_j}\Big),\ \Phi\Big(-\frac{\sqrt{n}\,(\gamma-\varepsilon_*)}{\tau_*}\Big)\right\}. \tag{F.9}$$

Proof. Fix $d \in \Delta_\Pi$.

Inside-cone alternatives. Suppose $d \in \mathcal{D}_\gamma \cap C_V$. Then

$$w_j^\top d \ge \gamma, \qquad j = 1, \dots, r.$$

By Theorem F.1,

$$u_j^\top A_V d \ge \gamma - \varepsilon_j.$$

By Theorem F.3,

$$S_{j,n} \sim \mathcal{N}\Big(u_j^\top A_V d, \frac{\tau_j^2}{n}\Big).$$

Therefore

$$\mathbb{P}_d(S_{j,n} < 0) = \Phi\Big(-\frac{\sqrt{n}\,u_j^\top A_V d}{\tau_j}\Big) \le \Phi\Big(-\frac{\sqrt{n}\,(\gamma-\varepsilon_j)}{\tau_j}\Big).$$

Since

$$\{\hat{\psi}_n^{\mathrm{aOCT}} = 0\} = \bigcup_{j=1}^{r} \{S_{j,n} < 0\},$$

a union bound yields

$$\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} = 0\big) \le \sum_{j=1}^{r} \Phi\Big(-\frac{\sqrt{n}\,(\gamma-\varepsilon_j)}{\tau_j}\Big).$$

Taking the supremum over $d \in \mathcal{D}_\gamma \cap C_V$ proves (F.7).

Outside-cone alternatives. Now suppose $d \in \mathcal{D}_\gamma \setminus C_V$. By definition of $\mathcal{D}_\gamma$, there exists at least one violated facet $j(d)$ such that

$$w_{j(d)}^\top d \le -\gamma.$$

By Theorem F.1,

$$u_{j(d)}^\top A_V d \le -\gamma + \varepsilon_{j(d)} \le -\gamma + \varepsilon_* < 0.$$

Since the event $\hat{\psi}_n^{\mathrm{aOCT}} = 1$ implies $S_{j(d),n} \ge 0$, Theorem F.3 gives

$$\begin{aligned}
\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} = 1\big) &\le \mathbb{P}_d\big(S_{j(d),n} \ge 0\big) \\
&= \Phi\Big(\frac{\sqrt{n}\,u_{j(d)}^\top A_V d}{\tau_{j(d)}}\Big) \\
&\le \Phi\Big(-\frac{\sqrt{n}\,(\gamma-\varepsilon_{j(d)})}{\tau_{j(d)}}\Big) \\
&\le \Phi\Big(-\frac{\sqrt{n}\,(\gamma-\varepsilon_*)}{\tau_*}\Big),
\end{aligned}$$

since $\varepsilon_{j(d)} \le \varepsilon_*$, $\tau_{j(d)} \le \tau_*$, and Φ is monotone increasing. Taking the supremum over $d \in \mathcal{D}_\gamma \setminus C_V$ proves (F.8).

The uniform bound (F.9) is immediate. □

Corollary F.5 (Exact observability is recovered as the zero-bias case)**.** *Suppose $u_j = u_j^\star$ for all j, where $u_j^\star$ is the minimum-variance exact certificate from Theorem 4.1. Then*

$$r_j = 0, \qquad \varepsilon_j = 0, \qquad \tau_j = \sigma_j, \qquad j = 1, \dots, r.$$

Consequently, Theorem F.4 reduces to the one-sided bounds in Theorem 4.3.

Proof. Since $A_V^* u_j^\star = w_j$, the residual $r_j = w_j - A_V^* u_j^\star$ vanishes, hence $\varepsilon_j = 0$. Also,

$$\tau_j^2 = u_j^{\star\top} \Sigma u_j^\star = \sigma_j^2$$

by Theorem 4.1. Substituting into Theorem F.4 yields the claimed specialization. □

F.3 Sub-Gaussian robustness

We now show that the same bias degradation survives beyond the Gaussian model.

Definition F.6 (Sub-Gaussian sample-mean noise model)**.** We say that the sample mean noise is Σ/n-sub-Gaussian on Δ_Π if for every $d \in \Delta_\Pi$, every $u \in \mathbb{R}^q$, and every $\lambda \in \mathbb{R}$,

$$\mathbb{E}_d\Big[\exp\Big(\lambda\, u^\top(\bar{Y}_n - A_V d)\Big)\Big] \le \exp\Big(\frac{\lambda^2}{2n} u^\top \Sigma u\Big). \tag{F.10}$$

Lemma F.7 (One-dimensional sub-Gaussian tails for approximate certificates)**.** *Under Theorem F.6, for every $d \in \Delta_\Pi$, every $j \in \{1, \dots, r\}$, and every $t > 0$,*

$$\mathbb{P}_d\big(S_{j,n} - u_j^\top A_V d \le -t\big) \le \exp\Big(-\frac{nt^2}{2\tau_j^2}\Big), \tag{F.11}$$

$$\mathbb{P}_d\big(S_{j,n} - u_j^\top A_V d \ge t\big) \le \exp\Big(-\frac{nt^2}{2\tau_j^2}\Big). \tag{F.12}$$

Proof. Fix d, j, and set

$$X := S_{j,n} - u_j^\top A_V d = u_j^\top(\bar{Y}_n - A_V d).$$

By Theorem F.6,

$$\mathbb{E}_d[e^{\lambda X}] \le \exp\Big(\frac{\lambda^2 \tau_j^2}{2n}\Big), \qquad \lambda \in \mathbb{R}.$$

The upper tail bound follows from Chernoff's inequality: for $\lambda > 0$,

$$\mathbb{P}_d(X \ge t) \le e^{-\lambda t}\mathbb{E}_d[e^{\lambda X}] \le \exp\Big(-\lambda t + \frac{\lambda^2 \tau_j^2}{2n}\Big).$$

Optimizing at $\lambda = nt/\tau_j^2$ gives (F.12). The lower-tail bound follows by applying the same argument to $-X$. □

Theorem F.8 (Robust sub-Gaussian upper bound)**.** *Assume the Σ/n-sub-Gaussian sample-mean noise model of Theorem F.6, and let $0 < \gamma \le \bar{\gamma}$ satisfy*

$$\gamma > \varepsilon_*.$$

Then the approximate observable cone test (F.4) *satisfies*

$$\sup_{d \in \mathcal{D}_\gamma \cap C_V} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} = 0\big) \le \sum_{j=1}^{r} \exp\Big(-\frac{n(\gamma - \varepsilon_j)^2}{2\tau_j^2}\Big), \tag{F.13}$$

$$\sup_{d \in \mathcal{D}_\gamma \setminus C_V} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} = 1\big) \le \exp\Big(-\frac{n(\gamma - \varepsilon_*)^2}{2\tau_*^2}\Big). \tag{F.14}$$

Hence

$$\sup_{d \in \mathcal{D}_\gamma} \mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} \ne \psi_V(d)\big) \le \max\left\{\sum_{j=1}^{r} \exp\Big(-\frac{n(\gamma - \varepsilon_j)^2}{2\tau_j^2}\Big),\ \exp\Big(-\frac{n(\gamma - \varepsilon_*)^2}{2\tau_*^2}\Big)\right\}. \tag{F.15}$$

Proof. The proof is identical in structure to that of Theorem F.4, with Theorem F.7 replacing the exact Gaussian tail computation.

If $d \in \mathcal{D}_\gamma \cap C_V$, then $u_j^\top A_V d \ge \gamma - \varepsilon_j$ by Theorem F.1, so

$$\mathbb{P}_d(S_{j,n} < 0) = \mathbb{P}_d\big(S_{j,n} - u_j^\top A_V d < -u_j^\top A_V d\big) \le \mathbb{P}_d\big(S_{j,n} - u_j^\top A_V d \le -(\gamma - \varepsilon_j)\big),$$

and Theorem F.7 gives

$$\mathbb{P}_d(S_{j,n} < 0) \le \exp\Big(-\frac{n(\gamma - \varepsilon_j)^2}{2\tau_j^2}\Big).$$

A union bound proves (F.13).

If $d \in \mathcal{D}_\gamma \setminus C_V$, choose a violated facet $j(d)$ with $w_{j(d)}^\top d \le -\gamma$. Then $u_{j(d)}^\top A_V d \le -\gamma + \varepsilon_{j(d)} \le -\gamma + \varepsilon_*$, so

$$\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} = 1\big) \le \mathbb{P}_d\big(S_{j(d),n} \ge 0\big) = \mathbb{P}_d\big(S_{j(d),n} - u_{j(d)}^\top A_V d \ge -u_{j(d)}^\top A_V d\big)$$

and therefore

$$\mathbb{P}_d\big(\hat{\psi}_n^{\mathrm{aOCT}} = 1\big) \le \exp\Big(-\frac{n(\gamma - \varepsilon_{j(d)})^2}{2\tau_{j(d)}^2}\Big) \le \exp\Big(-\frac{n(\gamma - \varepsilon_*)^2}{2\tau_*^2}\Big).$$

Taking suprema yields (F.14), and the uniform bound follows. □

Remark F.9 (Robustness interpretation). Approximate observability does not change the qualitative structure of the testing problem. The effect is purely quantitative: each facet margin γ is replaced by an effective margin $\gamma - \varepsilon_j$. Thus the exact theory is stable under certificate misspecification as long as the welfare signal remains separated from the boundary by more than the deterministic bias.

G Full proof of the separation theorem and variants

This appendix expands the separation result from Theorem 5.1. The main theorem of Section 5 shows that exact order identifiability from feedback operators does not require scalar Markov reward representability. Here we isolate the algebraic reason for that separation, prove the main construction in full detail, and record two broader variants. The first variant remains within the cone-based, translation-invariant world of the main paper; the second treats lexicographic welfare, which lies outside the closed-polyhedral regime of Assumption 2.2(2) but illustrates that the same obstruction is more general. These constructions are consistent with the broader expressivity limitations of scalar Markov rewards identified by Skalse and Abate [2023] and with the partial-identifiability viewpoint of reward learning in Skalse et al. [2023], Metelli et al. [2023].

G.1 A representation lemma for randomized one-step models

We first isolate a simple but important fact: on a one-step MDP with all randomized policies available, any scalar Markov reward that represents a preorder depending only on consequences must factor through an affine functional of the consequence vector.

Fix $m \in \mathbb{N}$, and let

$$\Delta_m := \{p \in \mathbb{R}_{\ge 0}^m : \mathbf{1}^\top p = 1\}$$

be the probability simplex over m actions. Let

$$B \in \mathbb{R}^{d \times m}, \qquad c(p) := Bp \in \mathbb{R}^d,$$

and let $K := B\Delta_m \subset \mathbb{R}^d$ denote the induced consequence set.

Lemma G.1 (Affine reduction of scalar-reward representations)**.** *Let $\succeq$ be a preorder on Δ_m satisfying*

$$Bp = Bq \quad \Longrightarrow \quad \big(p \succeq q \text{ and } q \succeq p\big) \qquad \forall\, p, q \in \Delta_m. \tag{G.1}$$

Suppose there exists $r \in \mathbb{R}^m$ such that

$$p \succeq q \quad \Longleftrightarrow \quad r^\top p \ge r^\top q \qquad \forall\, p, q \in \Delta_m. \tag{G.2}$$

Then there exist $w \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$ such that

$$r = B^\top w + \beta \mathbf{1}.$$

Consequently,

$$r^\top p = w^\top c(p) + \beta \qquad \forall\, p \in \Delta_m,$$

and $\succeq$ is induced on K by the affine functional $c \mapsto w^\top c + \beta$. In particular, pairwise comparisons are equivalently determined by the linear score $w^\top c$.

Proof. Define the linear subspace

$$H := \{h \in \mathbb{R}^m : Bh = 0,\ \mathbf{1}^\top h = 0\} = \ker(B) \cap \mathbf{1}^\perp.$$

We first prove that

$$r^\top h = 0 \qquad \forall\, h \in H. \tag{G.3}$$

Fix $h \in H$. Since $\mathbf{1}^\top h = 0$, the vector h has both positive and negative components unless $h = 0$, in which case (G.3) is trivial. Assume $h \neq 0$. Choose any point $q \in \operatorname{relint}(\Delta_m)$, so $q_i > 0$ for every i. Since q is strictly positive, there exists $t > 0$ such that

$$p := q + \frac{1}{t} h \in \Delta_m.$$

Indeed, $\mathbf{1}^\top p = \mathbf{1}^\top q + t^{-1}\mathbf{1}^\top h = 1$, and for t sufficiently large every coordinate of p remains nonnegative.

Now

$$Bp = Bq + \frac{1}{t} Bh = Bq,$$

so by (G.1) one has

$$p \succeq q \qquad \text{and} \qquad q \succeq p.$$

Applying the representation (G.2) yields

$$r^\top p = r^\top q.$$

Substituting $p = q + t^{-1}h$ gives $t^{-1} r^\top h = 0$, proving (G.3).

Therefore $r \in H^\perp$. Since

$$H = \ker(B) \cap \mathbf{1}^\perp,$$

the orthogonal-complement identity $(U \cap W)^\perp = U^\perp + W^\perp$ for linear subspaces gives

$$H^\perp = \ker(B)^\perp + (\mathbf{1}^\perp)^\perp = \operatorname{row}(B) + \operatorname{span}(\mathbf{1}) = \{B^\top w + \beta \mathbf{1} : w \in \mathbb{R}^d,\ \beta \in \mathbb{R}\}.$$

Hence

$$r = B^\top w + \beta \mathbf{1}$$

for some w, β, as claimed. Finally, for any $p \in \Delta_m$,

$$r^\top p = (B^\top w + \beta \mathbf{1})^\top p = w^\top B p + \beta\, \mathbf{1}^\top p = w^\top c(p) + \beta,$$

since $\mathbf{1}^\top p = 1$. This proves the factorization through an affine functional on K. Because the additive constant β cancels in pairwise comparisons, the represented preorder is equivalently determined by the linear score $w^\top c$. □

Corollary G.2 (Complete-preorder obstruction)**.** *If a preorder $\succeq$ on Δ_m has an incomparable pair $p, q \in \Delta_m$, then $\succeq$ cannot be represented by any scalar Markov reward of the form $p \mapsto r^\top p$.*

Proof. Every real-valued score $p \mapsto r^\top p$ induces a complete preorder: for any $p, q \in \Delta_m$, either $r^\top p \geq r^\top q$ or $r^\top q \geq r^\top p$ (or both). Hence no such score can represent a preorder with an incomparable pair. □

Lemma G.3 (Exact identifiability from full consequence-difference observation)**.** *Let $\psi : \mathbb{R}^d \to \{0, 1\}$ be any Borel-measurable map, and suppose the paired feedback experiment is*

$$Y^{p,q} \sim \mathcal{N}\big(B(p - q), \Sigma\big), \qquad \Sigma \succ 0.$$

Define a preorder on Δ_m by

$$p \succeq q \quad \Longleftrightarrow \quad \psi\big(B(p - q)\big) = 1.$$

Then $\succeq$ is exactly identifiable from the feedback experiment in the sense of Theorem 2.1.

Proof. In a Gaussian location family with fixed full-rank covariance Σ,

$$\mathcal{N}(\mu, \Sigma) = \mathcal{N}(\mu', \Sigma) \quad \Longleftrightarrow \quad \mu = \mu'.$$

Hence equality of observation laws is equivalent to equality of consequence-difference means:

$$\mathsf{P}_{p,q} = \mathsf{P}_{p',q'} \quad \Longleftrightarrow \quad B(p - q) = B(p' - q').$$

Since the welfare label is, by assumption, the measurable function $\psi(B(p - q))$, equal observation laws imply equal welfare labels. Thus the preorder is exactly identifiable by Theorem 2.1. □

G.2 Full proof of Theorem 5.1

We now reprove Theorem 5.1 in a form that makes the geometry of the construction explicit.

Proposition G.4 (The consequence geometry of the main separation construction)**.** *Consider the one-state, one-step MDP with action set $\mathcal{A} = \{a, b, c\}$, policy class $\Pi = \Delta(\mathcal{A})$, and consequence map*

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Then the consequence set is the triangle

$$K = B\Delta(\mathcal{A}) = \operatorname{conv}\{(1,0),(0,1),(1,1)\} = \{(x,y) \in [0,1]^2 : x + y \geq 1\}.$$

Proof. Let $p = (p_a, p_b, p_c) \in \Delta(\mathcal{A})$. Then

$$c(p) = Bp = (p_a + p_c,\ p_b + p_c).$$

Since $p_a, p_b, p_c \geq 0$ and $p_a + p_b + p_c = 1$, one has

$$0 \leq p_a + p_c \leq 1, \qquad 0 \leq p_b + p_c \leq 1, \qquad (p_a + p_c) + (p_b + p_c) = 1 + p_c \geq 1.$$

Thus $c(p) \in [0,1]^2$ and $c_1(p) + c_2(p) \geq 1$, proving $K \subseteq \{(x,y) \in [0,1]^2 : x + y \geq 1\}$.

Conversely, fix $(x,y) \in [0,1]^2$ with $x + y \geq 1$. Define

$$p_c := x + y - 1, \qquad p_a := 1 - y, \qquad p_b := 1 - x.$$

Then $p_a, p_b, p_c \geq 0$ and

$$p_a + p_b + p_c = (1 - y) + (1 - x) + (x + y - 1) = 1.$$

Hence $p = (p_a, p_b, p_c) \in \Delta(\mathcal{A})$, and

$$c(p) = (p_a + p_c,\ p_b + p_c) = (x, y).$$

This proves the reverse inclusion. The convex-hull description follows because the image of a simplex under a linear map is the convex hull of the images of its vertices. □

Full proof of Theorem 5.1. Consider the construction from Theorem G.4, and equip $\mathbb{R}^2$ with the Pareto cone

$$C = \mathbb{R}^2_+.$$

Thus the welfare preorder is

$$p \succeq p' \quad \Longleftrightarrow \quad c(p) - c(p') \in \mathbb{R}^2_+.$$

Step 1: exact identifiability. Take $A = I_2$, and let the paired feedback law be

$$Y^{p,p'} \sim \mathcal{N}\big(c(p) - c(p'), \Sigma\big), \qquad \Sigma \succ 0.$$

Define

$$\psi(d) := \mathbb{I}\{d \in \mathbb{R}^2_+\}, \qquad d \in \mathbb{R}^2.$$

Then

$$p \succeq p' \quad \Longleftrightarrow \quad \psi\big(c(p) - c(p')\big) = 1.$$

By Theorem G.3, the preorder is exactly identifiable from the feedback experiment.

Step 2: existence of incomparable policies. Let $p^a, p^b \in \Delta(\mathcal{A})$ denote the deterministic policies that choose actions a and b, respectively. Then

$$c(p^a) = (1,0), \qquad c(p^b) = (0,1).$$

Hence

$$c(p^a) - c(p^b) = (1,-1) \notin \mathbb{R}^2_+, \qquad c(p^b) - c(p^a) = (-1,1) \notin \mathbb{R}^2_+.$$

Therefore p^a and p^b are incomparable under $\succeq$.

Step 3: failure of scalar-reward representability. By Theorem G.2, no scalar reward $p \mapsto r^\top p$ can represent a preorder with an incomparable pair. Therefore $\succeq$ is not representable by any scalar Markov reward in the sense used in Section 5.

Combining Steps 1–3 proves the theorem. □

G.3 A general cone-based separation criterion

The previous construction is just the simplest member of a broad family of examples. Any consequence-defined cone order with at least one attainable incomparable pair yields the same logical separation.

Proposition G.5 (General cone-based separation criterion)**.** *Fix a one-step randomized model with action simplex Δ_m, consequence matrix $B \in \mathbb{R}^{d\times m}$, and consequence set $K = B\Delta_m$. Let $C \subset \mathbb{R}^d$ be any cone, and define the preorder*

$$p \succeq_C q \quad \Longleftrightarrow \quad B(p-q) \in C.$$

Assume there exist $u, v \in K$ such that

$$u - v \notin C \qquad \textit{and} \qquad v - u \notin C.$$

Then:

1. *the preorder $\succeq_C$ is exactly identifiable from the Gaussian experiment*

$$Y^{p,q} \sim \mathcal{N}\big(B(p-q), \Sigma\big), \qquad \Sigma \succ 0;$$

2. *$\succeq_C$ is not representable by any scalar Markov reward on the full randomized policy class Δ_m.*

Proof. The order label is the measurable function

$$\psi(d) := \mathbb{I}\{d \in C\}, \qquad d \in \mathbb{R}^d.$$

Hence exact identifiability follows from Theorem G.3.

Now choose $p, q \in \Delta_m$ such that $Bp = u$ and $Bq = v$. By assumption,

$$u - v \notin C \qquad \text{and} \qquad v - u \notin C,$$

so neither $p \succeq_C q$ nor $q \succeq_C p$ holds. Thus the preorder has an incomparable pair, and nonrepresentability follows from Theorem G.2. □

Corollary G.6 (Dimension-free Pareto separations)**.** *For every $d \geq 2$, there exists a one-step randomized MDP and a consequence map $B : \mathbb{R}^d \to \mathbb{R}^d$ such that the Pareto preorder*

$$p \succeq q \quad \Longleftrightarrow \quad B(p-q) \in \mathbb{R}^d_+$$

is exactly identifiable from full consequence-difference feedback but is not representable by any scalar Markov reward on the full randomized policy class.

Proof. Take $m = d$, $B = I_d$, and Δ_d as the full randomized policy class. Then $K = \Delta_d \subset \mathbb{R}^d$. The deterministic policies e_i and e_j for $i \neq j$ satisfy

$$e_i - e_j \notin \mathbb{R}^d_+, \qquad e_j - e_i \notin \mathbb{R}^d_+,$$

so they are incomparable under the Pareto order. Apply Theorem G.5 with $C = \mathbb{R}^d_+$. □

G.4 Lexicographic variant beyond the main polyhedral regime

The next variant lies outside Assumption 2.2(2): lexicographic welfare is translation invariant, but its associated positive cone is not closed in the Euclidean topology. It is nevertheless a useful stress test, because it shows that the failure of scalar-reward representability is not merely a consequence of incomparability.

Consider the one-step randomized model with four actions corresponding to the corners of the unit square:

$$c_1 = (0,0), \qquad c_2 = (0,1), \qquad c_3 = (1,0), \qquad c_4 = (1,1).$$

Equivalently, let

$$B = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \qquad K = B\Delta_4 = [0,1]^2.$$

Define the weak lexicographic preorder on K by

$$u \succeq_{\text{lex}} v \quad \Longleftrightarrow \quad \big(u_1 > v_1\big) \text{ or } \big(u_1 = v_1 \text{ and } u_2 \geq v_2\big).$$

Transport this preorder to policies by

$$p \succeq_{\text{lex}} q \quad \Longleftrightarrow \quad c(p) \succeq_{\text{lex}} c(q).$$

Proposition G.7 (Lexicographic welfare is exactly identifiable but not scalar-representable)**.** *For the square consequence set $K = [0,1]^2$ defined above:*

1. *the lexicographic preorder is exactly identifiable from the Gaussian difference experiment*
$$Y^{p,q} \sim \mathcal{N}\big(B(p-q), \Sigma\big), \qquad \Sigma \succ 0;$$
2. *the lexicographic preorder is not representable by any scalar Markov reward on the full randomized policy class Δ_4.*

Proof. For the first claim, define

$$\psi_{\text{lex}}(d) := \mathbb{I}\{d_1 > 0 \text{ or } (d_1 = 0 \text{ and } d_2 \geq 0)\}, \qquad d \in \mathbb{R}^2.$$

Then

$$p \succeq_{\text{lex}} q \quad \Longleftrightarrow \quad \psi_{\text{lex}}\big(B(p-q)\big) = 1.$$

Since ψ_{lex} is Borel measurable, exact identifiability follows from Theorem G.3.

For the second claim, suppose, for contradiction, that there exists a scalar reward vector $r \in \mathbb{R}^4$ representing $\succeq_{\text{lex}}$ on Δ_4. Because the preorder depends only on consequences, the fiber consistency condition (G.1) holds, and Theorem G.1 yields $w = (\alpha, \beta) \in \mathbb{R}^2$ and $\gamma \in \mathbb{R}$ such that

$$r^\top p = \alpha\, c_1(p) + \beta\, c_2(p) + \gamma \qquad \forall\, p \in \Delta_4.$$

Thus the represented preorder on K must coincide with the weak order induced by the affine score

$$s(u) = \alpha u_1 + \beta u_2 + \gamma.$$

Since the additive constant γ is irrelevant for comparisons, we may equivalently work with $u \mapsto \alpha u_1 + \beta u_2$.

Now compare the points $u = (0,1)$ and $v = (0,0)$. Lexicographically,

$$u \succeq_{\text{lex}} v \qquad \text{but} \qquad v \not\succeq_{\text{lex}} u.$$

Hence the affine score must satisfy

$$\alpha \cdot 0 + \beta \cdot 1 > \alpha \cdot 0 + \beta \cdot 0,$$

so necessarily

$$\beta > 0.$$

Next, for any $\varepsilon \in (0,1]$, consider the points

$$u_\varepsilon := (\varepsilon, 0), \qquad v := (0,1).$$

Since $u_{\varepsilon,1} > v_1$, lexicographic order gives

$$u_\varepsilon \succeq_{\text{lex}} v \qquad \text{and} \qquad v \not\succeq_{\text{lex}} u_\varepsilon.$$

Therefore the affine score must satisfy

$$\alpha\varepsilon > \beta.$$

Because this must hold for every $\varepsilon \in (0,1]$, letting $\varepsilon \downarrow 0$ forces $\beta \leq 0$, contradicting $\beta > 0$.

Hence no scalar Markov reward represents the lexicographic preorder. □

Remark G.8 (Why the lexicographic variant is conceptually useful). Theorem G.7 shows that the separation from scalar reward representability is not merely a byproduct of partial orders with incomparability. Even a *complete* translation-invariant welfare order can fail to collapse to a one-dimensional scalar reward on a rich randomized policy class. What breaks is not completeness, but affine representability.

Remark G.9 (Beyond the paired-difference model). The same proof strategy extends to richer world-feedback models in which the observable object is the policy consequence $c(\pi)$ itself rather than only the consequence difference $c(\pi) - c(\pi')$. In such models, one can build additional nonrepresentability examples based on threshold-first or constraint-priority welfare rules. Since those examples fall outside the exact paired-difference framework of the main paper, we do not include them here.

H Non-Gaussian and general statistical channels

The main text develops a finite-sample theory under a Gaussian paired-difference experiment. This appendix shows that the same order-observability perspective extends well beyond that setting. In particular, Bernoulli comparison channels, count channels, threshold alarms, and other world-feedback modalities can be treated under a common local information geometry. Such channels arise naturally in heterogeneous-feedback and general-preference formulations [Park et al., 2024, Ye et al., 2024].

The key message is that the Gaussian observable Gram operator

$$G = A_V^* \Sigma^{-1} A_V$$

from Section 4 is replaced, in a regular non-Gaussian channel, by a *pullback Fisher information operator*. The resulting local criterion is the exact analogue of facet observability: a facet normal is locally observable if and only if it lies in the range of the pullback Fisher operator.

H.1 A general law-map formulation

We begin with the most abstract level of generality. Let $(\mathcal{O}, \mathcal{B})$ be a standard Borel observation space, and suppose that each attainable difference $d \in \Delta_\Pi$ induces a one-sample law

$$\Lambda(d) \in \mathcal{P}(\mathcal{O}).$$

The associated n-sample experiment is then $\Lambda_n(d) := \Lambda(d)^{\otimes n}$.

Proposition H.1 (General exact-identifiability criterion)**.** *For an arbitrary measurable law map $\Lambda : \Delta_\Pi \to \mathcal{P}(\mathcal{O})$, the following are equivalent.*

1. *The welfare label $\psi_V(d) = \mathbb{I}\{d \in C_V\}$ is exactly identifiable from Λ.*
2. *The fibers of Λ are order-monochromatic:*
$$\Lambda(d) = \Lambda(d') \quad \Longrightarrow \quad \big(d \in C_V \iff d' \in C_V\big), \qquad d, d' \in \Delta_\Pi.$$
3. *There exists a measurable map*
$$\bar{\psi}_\Lambda : \Lambda(\Delta_\Pi) \to \{0, 1\}$$
such that
$$\psi_V = \bar{\psi}_\Lambda \circ \Lambda \qquad \text{on } \Delta_\Pi.$$

Proof. The proof is identical in structure to that of Theorem 3.1. The implication $(1) \Rightarrow (2)$ is immediate from the definition of exact identifiability. The implication $(2) \Rightarrow (3)$ is obtained by defining

$$\bar{\psi}_\Lambda(\Lambda(d)) := \psi_V(d),$$

which is well-defined precisely because the fibers are order-monochromatic. The implication $(3) \Rightarrow (1)$ is immediate by composition. □

Remark H.2 (Why the appendix focuses on local theory). Theorem H.1 shows that the order-theoretic core of the paper is not tied to Gaussianity. What *is* special to the Gaussian model is the finite-sample quadratic geometry that yields exact minimax constants. For general channels, the appropriate replacement is a *local asymptotic* theory governed by Fisher information.

H.2 Regular exponential-family channels

We now specialize to a regular parametric channel family rich enough to cover Bernoulli comparison channels, count channels, threshold alarms, and many multi-channel constructions.

Fix a point

$$d_\circ \in \Delta_\Pi.$$

Assume there exists an open neighborhood $U \subset V$ of $d_\circ$, an open parameter set $\Theta \subset \mathbb{R}^m$, a twice continuously differentiable map

$$\eta : U \to \Theta,$$

a measurable sufficient statistic $S : \mathcal{O} \to \mathbb{R}^m$, a σ-finite base measure ν, and a C^3 log-partition function $A : \Theta \to \mathbb{R}$ such that, for every $d \in U$, the one-sample law P_d has density

$$p_d(y) = \exp\Big(\eta(d)^\top S(y) - A(\eta(d))\Big)\, h(y) \tag{H.1}$$

with respect to ν, where h is a nonnegative measurable carrier.

Write

$$\eta_\circ := \eta(d_\circ), \qquad \mu_\circ := \nabla A(\eta_\circ), \qquad H_\circ := \nabla^2 A(\eta_\circ).$$

Since the family is regular, $H_\circ$ is positive definite. Let

$$D\eta_\circ : V \to \mathbb{R}^m$$

denote the Fréchet derivative of η at $d_\circ$, and define the *pullback Fisher information operator*

$$\mathcal{J}_\circ := D\eta_\circ^* H_\circ D\eta_\circ : V \to V. \tag{H.2}$$

Lemma H.3 (Basic geometry of the pullback Fisher operator)**.** *The operator $\mathcal{J}_\circ$ is self-adjoint and positive semidefinite. Moreover,*

$$\ker(\mathcal{J}_\circ) = \ker(D\eta_\circ), \qquad \operatorname{range}(\mathcal{J}_\circ) = \operatorname{range}(D\eta_\circ^*).$$

Proof. Self-adjointness and positive semidefiniteness are immediate from (H.2). For $h \in V$,

$$h^\top \mathcal{J}_\circ h = (D\eta_\circ h)^\top H_\circ (D\eta_\circ h).$$

Since $H_\circ \succ 0$, this vanishes if and only if $D\eta_\circ h = 0$, proving

$$\ker(\mathcal{J}_\circ) = \ker(D\eta_\circ).$$

Because $\mathcal{J}_\circ$ is self-adjoint, its range is the orthogonal complement of its kernel. Hence

$$\operatorname{range}(\mathcal{J}_\circ) = \ker(\mathcal{J}_\circ)^\perp = \ker(D\eta_\circ)^\perp = \operatorname{range}(D\eta_\circ^*). \qquad \square$$

H.3 Local asymptotic normality on consequence space

We now derive the local log-likelihood expansion on the effective ordered space V.

For $h \in V$, define the local alternative

$$d_{\circ,n}(h) := d_\circ + \frac{h}{\sqrt{n}},$$

whenever this point belongs to U. Let

$$P_\circ^{(n)} := P_{d_\circ}^{\otimes n}, \qquad P_h^{(n)} := P_{d_{\circ,n}(h)}^{\otimes n}.$$

Define the *central sequence*

$$\Delta_{n,\circ} := \frac{1}{\sqrt{n}} \sum_{i=1}^n D\eta_\circ^* \big(S(Y_i) - \mu_\circ\big) \in V. \tag{H.3}$$

Proposition H.4 (Local asymptotic normality)**.** *Fix $h \in V$. Under $P_\circ^{(n)}$,*

$$\Delta_{n,\circ} \Rightarrow \mathcal{N}(0, \mathcal{J}_\circ)$$

as $n \to \infty$, and the log-likelihood ratio admits the expansion

$$\log \frac{dP_h^{(n)}}{dP_\circ^{(n)}} = h^\top \Delta_{n,\circ} - \frac{1}{2} h^\top \mathcal{J}_\circ h + o_{P_\circ^{(n)}}(1). \tag{H.4}$$

Proof. Let

$$\delta_n := \eta(d_{\circ,n}(h)) - \eta_\circ.$$

Since η is C^2,

$$\delta_n = \frac{1}{\sqrt{n}} D\eta_\circ h + r_n, \qquad \|r_n\| = O\Big(\frac{1}{n}\Big).$$

Now

$$\log \frac{dP_h^{(n)}}{dP_\circ^{(n)}} = \sum_{i=1}^n \Big(\delta_n^\top S(Y_i) - A(\eta_\circ + \delta_n) + A(\eta_\circ)\Big).$$

Using the Taylor expansion of A at $\eta_\circ$,

$$A(\eta_\circ + \delta_n) = A(\eta_\circ) + \mu_\circ^\top \delta_n + \frac{1}{2}\delta_n^\top H_\circ \delta_n + R_n,$$

where $R_n = O(\|\delta_n\|^3) = O(n^{-3/2})$. Hence

$$\log \frac{dP_h^{(n)}}{dP_\circ^{(n)}} = \delta_n^\top \sum_{i=1}^n \big(S(Y_i) - \mu_\circ\big) - \frac{n}{2}\delta_n^\top H_\circ \delta_n - nR_n.$$

Because

$$\sum_{i=1}^n \big(S(Y_i) - \mu_\circ\big) = O_{P_\circ^{(n)}}(\sqrt{n})$$

by the multivariate central limit theorem, and because

$$\delta_n - \frac{1}{\sqrt{n}} D\eta_\circ h = O\Big(\frac{1}{n}\Big),$$

the first term becomes

$$\delta_n^\top \sum_{i=1}^n \big(S(Y_i) - \mu_\circ\big) = h^\top \Delta_{n,\circ} + o_{P_\circ^{(n)}}(1).$$

Similarly,

$$n\,\delta_n^\top H_\circ \delta_n = h^\top D\eta_\circ^* H_\circ D\eta_\circ h + o(1) = h^\top \mathcal{J}_\circ h + o(1),$$

and

$$nR_n = O(n^{-1/2}) = o(1).$$

Substituting these estimates proves (H.4).

Finally, under $P_\circ^{(n)}$, the summands

$$D\eta_\circ^*\big(S(Y_i) - \mu_\circ\big)$$

are i.i.d., centered, and have covariance

$$D\eta_\circ^* H_\circ D\eta_\circ = \mathcal{J}_\circ.$$

Therefore the multivariate CLT gives

$$\Delta_{n,\circ} \Rightarrow \mathcal{N}(0, \mathcal{J}_\circ).$$

□

H.4 Local facet observability beyond Gaussianity

We now derive the non-Gaussian analogue of facet observability. Fix a facet index $j \in \{1, \dots, r\}$, and suppose

$$d_\circ \in \operatorname{relint}(F_j) \cap \operatorname{relint}(\Delta_\Pi), \qquad F_j := \{v \in C_V : w_j^\top v = 0\}.$$

Theorem H.5 (Local facet observability through pullback Fisher information)**.** *Assume the regular exponential-family channel model* (H.1) *in a neighborhood of* $d_\circ \in \operatorname{relint}(F_j) \cap \operatorname{relint}(\Delta_\Pi)$.

1. *If*

$$w_j \notin \operatorname{range}(\mathcal{J}_\circ),$$

then there exists $h \in V$ with $w_j^\top h > 0$ such that, for every fixed $\gamma > 0$, the local alternatives

$$d_{+,n} := d_\circ + \frac{\gamma h}{\sqrt{n}}, \qquad d_{-,n} := d_\circ - \frac{\gamma h}{\sqrt{n}}$$

belong to Δ_Π for all sufficiently large n, lie on opposite sides of the facet F_j, and satisfy

$$\mathrm{TV}\big(P_{d_{+,n}}^{\otimes n}, P_{d_{-,n}}^{\otimes n}\big) \to 0.$$

Consequently,

$$\liminf_{n\to\infty} \inf_{\phi_n} \max\Big\{P_{d_{+,n}}^{\otimes n}(\phi_n = 0), P_{d_{-,n}}^{\otimes n}(\phi_n = 1)\Big\} \geq \frac{1}{2}.$$

2. *If*

$$w_j \in \operatorname{range}(\mathcal{J}_\circ),$$

define

$$\sigma_{j,\circ}^2 := w_j^\top \mathcal{J}_\circ^\dagger w_j, \qquad h_{j,\circ}^\star := \frac{\mathcal{J}_\circ^\dagger w_j}{\sigma_{j,\circ}^2}.$$

Then $w_j^\top h_{j,\circ}^\star = 1$, and for every fixed $\gamma > 0$, the local alternatives

$$d_{j,\pm,n} := d_\circ \pm \frac{\gamma h_{j,\circ}^\star}{\sqrt{n}}$$

lie on opposite sides of F_j for all sufficiently large n. Moreover, the local sequence of binary experiments generated by $(d_{j,+,n}, d_{j,-,n})$ converges to the one-dimensional Gaussian shift experiment

$$Z \sim \mathcal{N}(\pm\gamma, \sigma_{j,\circ}^2).$$

In particular,

$$\liminf_{n\to\infty} \inf_{\phi_n} \max\Big\{P_{d_{j,+,n}}^{\otimes n}(\phi_n = 0), P_{d_{j,-,n}}^{\otimes n}(\phi_n = 1)\Big\} \geq \Phi\Big(-\frac{\gamma}{\sigma_{j,\circ}}\Big). \tag{H.5}$$

Proof. We prove the two parts separately.

Part (1): local impossibility when $w_j \notin \operatorname{range}(\mathcal{J}_\circ)$. By Theorem H.3,

$$\operatorname{range}(\mathcal{J}_\circ) = \ker(\mathcal{J}_\circ)^\perp.$$

Hence $w_j \notin \operatorname{range}(\mathcal{J}_\circ)$ implies the existence of $h \in \ker(\mathcal{J}_\circ)$ with $w_j^\top h \neq 0$. Replacing h by $-h$ if necessary, assume $w_j^\top h > 0$. Again by Theorem H.3,

$$\ker(\mathcal{J}_\circ) = \ker(D\eta_\circ),$$

so $D\eta_\circ h = 0$.

Since $d_\circ \in \operatorname{relint}(F_j) \cap \operatorname{relint}(\Delta_\Pi)$, the facet-crossing lemma (Theorem A.13) and the relative interior of Δ_Π imply that for every fixed $\gamma > 0$, the points

$$d_{+,n} = d_\circ + \frac{\gamma h}{\sqrt{n}}, \qquad d_{-,n} = d_\circ - \frac{\gamma h}{\sqrt{n}}$$

belong to Δ_Π for all sufficiently large n, with

$$d_{+,n} \in C_V, \qquad d_{-,n} \notin C_V.$$

It remains to show asymptotic indistinguishability. Since η is C^2 and $D\eta_\circ h = 0$,

$$\eta(d_{\pm,n}) = \eta_\circ + O\Big(\frac{1}{n}\Big),$$

and therefore

$$\eta(d_{+,n}) - \eta(d_{-,n}) = o\Big(\frac{1}{n}\Big).$$

For exponential families, the one-sample Kullback–Leibler divergence has the exact form

$$\mathrm{KL}(P_\eta \,\|\, P_\xi) = A(\xi) - A(\eta) - \nabla A(\eta)^\top(\xi - \eta),$$

hence, by a second-order Taylor expansion of A,

$$\mathrm{KL}(P_\eta \,\|\, P_\xi) = O(\|\eta - \xi\|^2) \qquad \text{as } \eta - \xi \to 0.$$

Applying this with $\eta = \eta(d_{+,n})$ and $\xi = \eta(d_{-,n})$ yields

$$\mathrm{KL}(P_{d_{+,n}} \,\|\, P_{d_{-,n}}) = o\Big(\frac{1}{n^2}\Big).$$

Therefore

$$\mathrm{KL}(P_{d_{+,n}}^{\otimes n} \,\|\, P_{d_{-,n}}^{\otimes n}) = n\ \mathrm{KL}(P_{d_{+,n}} \,\|\, P_{d_{-,n}}) \to 0.$$

Pinsker's inequality then gives

$$\mathrm{TV}\big(P_{d_{+,n}}^{\otimes n}, P_{d_{-,n}}^{\otimes n}\big) \to 0.$$

Finally, for any test ϕ_n,

$$\max\Big\{P_{d_{+,n}}^{\otimes n}(\phi_n = 0), P_{d_{-,n}}^{\otimes n}(\phi_n = 1)\Big\} \geq \frac{1}{2}\Big(1 - \mathrm{TV}(P_{d_{+,n}}^{\otimes n}, P_{d_{-,n}}^{\otimes n})\Big),$$

so the limit inferior is at least $1/2$.

Part (2): local Gaussian limit when $w_j \in \mathrm{range}(\mathcal{J}_\circ)$. Since $w_j \in \mathrm{range}(\mathcal{J}_\circ)$, the pseudoinverse is well-defined on w_j, and

$$h_{j,\circ}^\star := \frac{\mathcal{J}_\circ^\dagger w_j}{\sigma_{j,\circ}^2}$$

satisfies

$$w_j^\top h_{j,\circ}^\star = \frac{w_j^\top \mathcal{J}_\circ^\dagger w_j}{\sigma_{j,\circ}^2} = 1.$$

As in Part (1), the facet-crossing lemma implies that, for every fixed $\gamma > 0$, the points

$$d_{j,\pm,n} = d_\circ \pm \frac{\gamma h_{j,\circ}^\star}{\sqrt{n}}$$

lie on opposite sides of F_j for all sufficiently large n.

Now apply Theorem H.4 with

$$h_\pm := \pm\gamma h_{j,\circ}^\star.$$

The central sequence $\Delta_{n,\circ}$ satisfies

$$\Delta_{n,\circ} \Rightarrow \mathcal{N}(0, \mathcal{J}_\circ) \qquad \text{under } P_\circ^{\otimes n},$$

and the log-likelihood ratio has the LAN form

$$\log \frac{dP_{h_\pm}^{(n)}}{dP_\circ^{(n)}} = h_\pm^\top \Delta_{n,\circ} - \frac{1}{2} h_\pm^\top \mathcal{J}_\circ h_\pm + o_{P_\circ^{\otimes n}}(1).$$

This identifies the local limit experiment as the Gaussian shift experiment on V with covariance $\mathcal{J}_\circ$. Projecting this limit experiment along the functional $w_j^\top \mathcal{J}_\circ^\dagger$ yields a scalar limit experiment

$$Z \sim \mathcal{N}(\pm\gamma, \sigma_{j,\circ}^2),$$

because

$$w_j^\top \mathcal{J}_\circ^\dagger(\mathcal{J}_\circ h_{j,\circ}^\star) = w_j^\top h_{j,\circ}^\star = 1$$

and

$$w_j^\top \mathcal{J}_\circ^\dagger \mathcal{J}_\circ \mathcal{J}_\circ^\dagger w_j = w_j^\top \mathcal{J}_\circ^\dagger w_j = \sigma_{j,\circ}^2.$$

The local asymptotic minimax bound (H.5) then follows from the exact minimax risk of the one-dimensional Gaussian shift experiment. □

Remark H.6 (Local analogue of the Gaussian main theorem). Theorem H.5 is the non-Gaussian counterpart of the main-text statistical theory. In the Gaussian model,

$$\eta(d) = A_V d, \qquad H_\circ = \Sigma^{-1}, \qquad \mathcal{J}_\circ = A_V^* \Sigma^{-1} A_V = G,$$

so the local Fisher geometry reduces exactly to the observable Gram geometry of Section 4.

H.5 Locally asymptotically optimal projected-score tests

The previous theorem gives the local lower bound. We now record a matching local test.

For a fixed facet j with $w_j \in \mathrm{range}(\mathcal{J}_\circ)$, define

$$Z_{j,n} := w_j^\top \mathcal{J}_\circ^\dagger \Delta_{n,\circ} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_j^\top \mathcal{J}_\circ^\dagger D\eta_\circ^* \big(S(Y_i) - \mu_\circ\big). \tag{H.6}$$

Consider the local sign test

$$\phi_{j,n}^{\mathrm{loc}} := \mathbb{I}\{Z_{j,n} \geq 0\}.$$

Proposition H.7 (Asymptotic optimality of the projected-score test)**.** *Under the hypotheses of Part (2) of Theorem H.5, for every fixed $\gamma > 0$,*

$$\lim_{n\to\infty} P_{d_{j,+,n}}^{\otimes n}\big(\phi_{j,n}^{\mathrm{loc}} = 0\big) = \lim_{n\to\infty} P_{d_{j,-,n}}^{\otimes n}\big(\phi_{j,n}^{\mathrm{loc}} = 1\big) = \Phi\Big(-\frac{\gamma}{\sigma_{j,\circ}}\Big).$$

Hence $\phi_{j,n}^{\mathrm{loc}}$ attains the local asymptotic lower bound in (H.5)*.*

Proof. By the LAN expansion in Theorem H.4 and Le Cam's third lemma, under the local alternatives $d_{j,\pm,n}$, the statistic $Z_{j,n}$ converges in distribution to

$$\mathcal{N}(\pm\gamma, \sigma_{j,\circ}^2).$$

Therefore

$$P_{d_{j,+,n}}^{\otimes n}(Z_{j,n} < 0) \to \Phi\Big(-\frac{\gamma}{\sigma_{j,\circ}}\Big), \qquad P_{d_{j,-,n}}^{\otimes n}(Z_{j,n} \geq 0) \to \Phi\Big(-\frac{\gamma}{\sigma_{j,\circ}}\Big),$$

which proves the claim. □

H.6 Examples

We close with three concrete channel classes.

Proposition H.8 (Bernoulli comparison channel)**.** *Suppose*

$$Y \in \{0,1\}, \qquad P_d(Y=1) = \sigma(\beta^\top d),$$

where $\sigma(t) = 1/(1+e^{-t})$ is the logistic link and $\beta \in V$. Then the model is a regular one-parameter exponential family with

$$\mathcal{J}_\circ = p_\circ(1-p_\circ)\,\beta\beta^\top, \qquad p_\circ := \sigma(\beta^\top d_\circ).$$

Consequently, a facet normal w_j is locally observable at $d_\circ$ if and only if $w_j \in \mathrm{span}(\beta)$.

Proof. This is the canonical Bernoulli exponential family with natural parameter

$$\eta(d) = \beta^\top d, \qquad A(\eta) = \log(1+e^\eta).$$

Hence

$$D\eta_\circ h = \beta^\top h, \qquad H_\circ = A''(\eta_\circ) = p_\circ(1-p_\circ),$$

so

$$\mathcal{J}_\circ = D\eta_\circ^* H_\circ D\eta_\circ = p_\circ(1-p_\circ)\,\beta\beta^\top.$$

Its range is $\mathrm{span}(\beta)$, proving the criterion. □

Proposition H.9 (Poisson count channel)**.** *Suppose*

$$Y \in \mathbb{N} \cup \{0\}, \qquad Y \sim \mathrm{Poisson}\big(e^{\beta^\top d}\big),$$

where $\beta \in V$. Then the model is a regular exponential family with

$$\mathcal{J}_\circ = \lambda_\circ\,\beta\beta^\top, \qquad \lambda_\circ := e^{\beta^\top d_\circ}.$$

Hence local observability again holds if and only if $w_j \in \mathrm{span}(\beta)$.

Proof. This is the canonical Poisson family with natural parameter

$$\eta(d) = \beta^\top d, \qquad A(\eta) = e^\eta.$$

Therefore

$$D\eta_\circ h = \beta^\top h, \qquad H_\circ = A''(\eta_\circ) = e^{\eta_\circ} = \lambda_\circ,$$

and the formula for $\mathcal{J}_\circ$ follows exactly as in Theorem H.8. □

Proposition H.10 (Independent multi-channel exponential families)**.** *Suppose $Y = (Y^{(1)}, \dots, Y^{(M)})$, where conditional on d, the coordinates are independent and each $Y^{(m)}$ follows a regular exponential family with natural parameter $\eta_m(d)$, sufficient statistic S_m, and log-partition A_m. Then the joint model is again a regular exponential family, and its pullback Fisher operator is*

$$\mathcal{J}_\circ = \sum_{m=1}^{M} D\eta_{m,\circ}^* H_{m,\circ} D\eta_{m,\circ}, \qquad H_{m,\circ} := \nabla^2 A_m(\eta_m(d_\circ)).$$

Thus heterogeneous channel fusion adds Fisher information across channels in the same way that it adds observable directions in the exact linear theory.

Proof. Independence implies that the joint log-density is the sum of the marginal log-densities, so the joint log-partition is

$$A(\eta_1, \dots, \eta_M) = \sum_{m=1}^{M} A_m(\eta_m).$$

Its Hessian is block diagonal with blocks $H_{m,\circ}$. Differentiating the stacked parameter map

$$d \mapsto (\eta_1(d), \dots, \eta_M(d))$$

and applying (H.2) yields the stated formula. □

Remark H.11 (Threshold alarms and probit channels). Binary threshold alarms and probit-type channels fit the same framework as Theorem H.8: one replaces the logistic link by another smooth strictly monotone link, and the resulting pullback Fisher operator remains of the form $c_\circ\, \beta\beta^\top$ for an appropriate local curvature constant $c_\circ > 0$.

I Hilbert-space extension

This appendix lifts the finite-dimensional theory to a separable Hilbert consequence space. The main text deliberately works in a finite-dimensional linear core, because that is the right level for exact finite-sample minimax theory. The Hilbert extension clarifies that the underlying order-observability phenomenon is fundamentally functional-analytic: the correct infinite-dimensional objects are the quotient induced by the observation operator, the closed dual cone, and the observable dual subspace

$$\overline{\mathrm{range}(A^*)} = \ker(A)^\perp.$$

The extension is formulated at the operator level. In particular, we do not develop an infinite-dimensional Gaussian-process testing theory here. Instead, we show that the exact identifiability results of Section 3 have clean Hilbert-space analogues, and that the finite-dimensional facet theorem is a special case of a more general dual-cone observability principle.

I.1 Hilbert consequence spaces and fused observation operators

Let $\mathcal{H}$ be a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and norm $\|\cdot\|_{\mathcal{H}}$. Let Π be any policy class, and assume that each $\pi \in \Pi$ induces a consequence vector

$$c^\pi \in \mathcal{H}.$$

Define the attainable difference set

$$\Delta_\Pi := \{c^\pi - c^{\pi'} : \pi, \pi' \in \Pi\} \subset \mathcal{H}.$$

Throughout this appendix, we assume that Δ_Π is compact in the norm topology of $\mathcal{H}$. Let

$$V_\Delta := \mathrm{cl}\left(\mathrm{span}(\Delta_\Pi)\right) \subseteq \mathcal{H}$$

denote the closed linear span of Δ_Π.

Let $\mathcal{G}_1, \ldots, \mathcal{G}_M$ be separable Hilbert spaces, and define the Hilbert direct sum

$$\mathcal{G} := \bigoplus_{m=1}^{M} \mathcal{G}_m.$$

For each m, let

$$A_m \in \mathcal{B}(\mathcal{H}, \mathcal{G}_m)$$

be a bounded linear observation operator. The fused observation operator is

$$A := \begin{bmatrix} A_1 \\ \vdots \\ A_M \end{bmatrix} \in \mathcal{B}(\mathcal{H}, \mathcal{G}).$$

We write

$$A_\Delta := A\restriction_{V_\Delta} \in \mathcal{B}(V_\Delta, \mathcal{G})$$

for its restriction to the effective span.

Finally, let $C \subseteq \mathcal{H}$ be a closed convex cone, and consider the binary welfare label

$$\psi_C(d) := \mathbb{I}\{d \in C\}, \qquad d \in \Delta_\Pi.$$

Since $\Delta_\Pi \subseteq V_\Delta$, only the restricted cone

$$C_\Delta := C \cap V_\Delta$$

matters for attainable comparisons.

Definition I.1 (Operator-identifiability on a subset)**.** Let $S \subseteq \mathcal{H}$, and let $\psi : S \to \{0, 1\}$. We say that ψ is *operator-identifiable through A on S* if

$$Ax = Ay \quad \Longrightarrow \quad \psi(x) = \psi(y), \qquad x, y \in S.$$

Equivalently, the fibers of A on S are ψ-monochromatic.

Remark I.2 (Why this formulation is the right infinite-dimensional one). In finite dimensions, exact identifiability from a Gaussian law can be phrased equivalently either at the level of means or at the level of observation laws, because the Gaussian location family is injective in the mean. In Hilbert spaces, the same issue depends on the chosen statistical channel. The operator formulation in Theorem I.1 isolates the exact order-theoretic core and remains valid for any statistical model whose law map is injective on $A(\Delta_\Pi)$.

I.2 Factorization on compact attainable sets

The first theorem is the Hilbert-space analogue of the finite-dimensional feedback-operator factorization theorem.

Theorem I.3 (Hilbert-space factorization theorem)**.** *Let $\Delta_\Pi \subset \mathcal{H}$ be compact, let $A \in \mathcal{B}(\mathcal{H}, \mathcal{G})$, and let $C \subseteq \mathcal{H}$ be a closed cone. Then the following are equivalent.*

1. *The welfare label $\psi_C(d) = \mathbb{I}\{d \in C\}$ is operator-identifiable through A on Δ_Π.*
2. *The fibers of A on Δ_Π are welfare-monochromatic:*
$$Ad = Ad' \quad \Longrightarrow \quad \big(d \in C \iff d' \in C\big), \qquad d, d' \in \Delta_\Pi.$$
3. *There exists a Borel map*
$$\bar{\psi} : A(\Delta_\Pi) \to \{0, 1\}$$
such that
$$\psi_C = \bar{\psi} \circ A \qquad \text{on } \Delta_\Pi.$$

4. *There exists a compact set*

$$\overline{C} \subseteq A(\Delta_\Pi)$$

such that, for every $d \in \Delta_\Pi$,

$$d \in C \quad \Longleftrightarrow \quad Ad \in \overline{C}.$$

Proof. The equivalence of (1) and (2) is exactly the definition of operator-identifiability.

We prove $(2) \Rightarrow (4) \Rightarrow (3) \Rightarrow (2)$.

Implication $(2) \Rightarrow (4)$**.** Define

$$\overline{C} := A(C \cap \Delta_\Pi).$$

Because $C \cap \Delta_\Pi$ is compact—being the intersection of the closed set C with the compact set Δ_Π—and A is continuous, $\overline{C}$ is compact. Also $\overline{C} \subseteq A(\Delta_\Pi)$.

Now fix $d \in \Delta_\Pi$. If $d \in C$, then $Ad \in \overline{C}$ by construction. Conversely, if $Ad \in \overline{C}$, then there exists $\tilde{d} \in C \cap \Delta_\Pi$ such that

$$A\tilde{d} = Ad.$$

By the monochromatic-fiber property in (2),

$$d \in C \iff \tilde{d} \in C.$$

Since $\tilde{d} \in C$, it follows that $d \in C$. Hence

$$d \in C \iff Ad \in \overline{C}, \qquad d \in \Delta_\Pi.$$

Implication $(4) \Rightarrow (3)$**.** Define

$$\bar{\psi}(y) := \mathbb{I}\{y \in \overline{C}\}, \qquad y \in A(\Delta_\Pi).$$

Since $\overline{C}$ is compact and $A(\Delta_\Pi)$ is compact, $\bar{\psi}$ is Borel measurable. The equivalence in (4) gives

$$\psi_C(d) = \bar{\psi}(Ad), \qquad d \in \Delta_\Pi.$$

Implication $(3) \Rightarrow (2)$**.** If $Ad = Ad'$, then

$$\psi_C(d) = \bar{\psi}(Ad) = \bar{\psi}(Ad') = \psi_C(d'),$$

so the fiber is monochromatic. □

Remark I.4 (Compactness is only used for measurability). The compactness of Δ_Π is used only to ensure that $\overline{C}$ is compact, hence Borel. The monochromatic-fiber formulation itself makes sense on arbitrary subsets of $\mathcal{H}$.

I.3 Observable dual subspace

We now identify the infinite-dimensional analogue of the row space from the main text.

Definition I.5 (Observable dual subspace)**.** Define

$$\mathcal{H}_{\text{obs}} := \ker(A_\Delta)^\perp = \overline{\text{range}(A_\Delta^*)} \subseteq V_\Delta.$$

We call $\mathcal{H}_{\text{obs}}$ the *observable dual subspace*.

Lemma I.6 (Hilbert-space range/orthogonality identity)**.** *For every bounded linear operator* $T \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ *between Hilbert spaces,*

$$\overline{\text{range}(T^*)} = \ker(T)^\perp.$$

In particular,

$$\mathcal{H}_{\text{obs}} = \overline{\text{range}(A_\Delta^*)}.$$

Proof. For every $u \in \mathrm{range}(T^*)$ and every $h \in \ker(T)$,

$$\langle u,\, h\rangle_{\mathcal{H}} = \langle T^* g,\, h\rangle_{\mathcal{H}} = \langle g,\, Th\rangle_{\mathcal{G}} = 0,$$

so $\mathrm{range}(T^*) \subseteq \ker(T)^\perp$, hence

$$\overline{\mathrm{range}(T^*)} \subseteq \ker(T)^\perp.$$

Conversely, if $h \in \ker(T)^\perp$, then h is orthogonal to $\ker(T)$. By the standard Hilbert-space decomposition theorem,

$$\mathcal{H} = \ker(T) \oplus \overline{\mathrm{range}(T^*)},$$

so $h \in \overline{\mathrm{range}(T^*)}$. Therefore

$$\overline{\mathrm{range}(T^*)} = \ker(T)^\perp.$$

Applying this with $T = A_\Delta$ proves the final claim. □

The next theorem is the infinite-dimensional analogue of facet observability for a *single* supporting functional.

Theorem I.7 (Linear welfare halfspaces and the observable dual subspace)**.** *Let $w \in V_\Delta$, and define the closed halfspace*

$$H_w := \{h \in V_\Delta : \langle w,\, h\rangle_{\mathcal{H}} \geq 0\}.$$

Then the following are equivalent.

1. *The label*
$$\psi_w(h) := \mathbb{I}\{h \in H_w\}$$
is operator-identifiable through A_Δ on V_Δ.

2. *w annihilates the unobservable directions:*
$$\langle w,\, k\rangle_{\mathcal{H}} = 0, \qquad \forall\, k \in \ker(A_\Delta).$$

3.
$$w \in \mathcal{H}_{\mathrm{obs}} = \overline{\mathrm{range}(A_\Delta^*)}.$$

If, in addition,

$$w \in \mathrm{range}(A_\Delta^*),$$

then there exists an exact observable certificate $u \in \mathcal{G}$ such that

$$\langle w,\, h\rangle_{\mathcal{H}} = \langle u,\, A_\Delta h\rangle_{\mathcal{G}}, \qquad \forall\, h \in V_\Delta.$$

Proof. We prove $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1)$.

Implication $(1) \Rightarrow (2)$**.** Assume ψ_w is operator-identifiable on V_Δ, and let $k \in \ker(A_\Delta)$. Suppose for contradiction that

$$\langle w,\, k\rangle_{\mathcal{H}} \neq 0.$$

Replacing k by $-k$ if necessary, assume

$$\langle w,\, k\rangle_{\mathcal{H}} < 0.$$

Then

$$A_\Delta k = 0 = A_\Delta 0,$$

but

$$0 \in H_w \qquad \text{while} \qquad k \notin H_w,$$

since $\langle w,\, 0\rangle = 0$ and $\langle w,\, k\rangle < 0$. This contradicts operator-identifiability. Hence w must annihilate $\ker(A_\Delta)$.

Implication $(2) \Rightarrow (3)$**.** If w annihilates $\ker(A_\Delta)$, then by definition

$$w \in \ker(A_\Delta)^\perp = \mathcal{H}_{\mathrm{obs}}.$$

Implication (3) $\Rightarrow$ (1)**.** Assume $w \in \mathcal{H}_{\mathrm{obs}} = \ker(A_\Delta)^\perp$. Let $h, h' \in V_\Delta$ satisfy

$$A_\Delta h = A_\Delta h'.$$

Then

$$h - h' \in \ker(A_\Delta),$$

so

$$\langle w,\, h - h' \rangle_{\mathcal{H}} = 0.$$

Hence

$$\langle w,\, h \rangle_{\mathcal{H}} = \langle w,\, h' \rangle_{\mathcal{H}},$$

which implies

$$h \in H_w \iff h' \in H_w.$$

Thus ψ_w is operator-identifiable on V_Δ.

Finally, if $w \in \mathrm{range}(A_\Delta^*)$, there exists $u \in \mathcal{G}$ with $A_\Delta^* u = w$, and then

$$\langle w,\, h \rangle_{\mathcal{H}} = \langle A_\Delta^* u,\, h \rangle_{\mathcal{H}} = \langle u,\, A_\Delta h \rangle_{\mathcal{G}} \qquad \forall\, h \in V_\Delta.$$

□

Remark I.8 (Closure versus exact certificates). In infinite dimensions, $\mathrm{range}(A_\Delta^*)$ need not be closed. The exact operator-identifiability of a linear halfspace depends only on $\overline{\mathrm{range}(A_\Delta^*)}$, not on $\mathrm{range}(A_\Delta^*)$ itself. The price of being only in the closure is that one may lose a *single bounded exact certificate* u, even though the label remains exactly identifiable.

Proposition I.9 (Approximate certificates from the observable dual subspace)**.** *Let* $w \in \mathcal{H}_{\mathrm{obs}} = \overline{\mathrm{range}(A_\Delta^*)}$. *Then there exists a sequence* $(u_n)_{n\geq 1} \subseteq \mathcal{G}$ *such that*

$$\sup_{d \in \Delta_\Pi} |\langle w,\, d \rangle_{\mathcal{H}} - \langle u_n,\, Ad \rangle_{\mathcal{G}}| \to 0.$$

More precisely, if

$$R_\Delta := \sup_{d \in \Delta_\Pi} \|d\|_{\mathcal{H}} < \infty,$$

then for every $\varepsilon > 0$ *one can choose* $u_\varepsilon \in \mathcal{G}$ *satisfying*

$$\sup_{d \in \Delta_\Pi} |\langle w,\, d \rangle_{\mathcal{H}} - \langle u_\varepsilon,\, Ad \rangle_{\mathcal{G}}| \leq R_\Delta\, \varepsilon.$$

Proof. Since $w \in \overline{\mathrm{range}(A_\Delta^*)}$, for every $\varepsilon > 0$ there exists $u_\varepsilon \in \mathcal{G}$ such that

$$\|A_\Delta^* u_\varepsilon - w\|_{\mathcal{H}} \leq \varepsilon.$$

For every $d \in \Delta_\Pi \subseteq V_\Delta$,

$$|\langle w,\, d \rangle_{\mathcal{H}} - \langle u_\varepsilon,\, Ad \rangle_{\mathcal{G}}| = |\langle w - A_\Delta^* u_\varepsilon,\, d \rangle_{\mathcal{H}}| \leq \|w - A_\Delta^* u_\varepsilon\|_{\mathcal{H}} \|d\|_{\mathcal{H}} \leq R_\Delta\, \varepsilon.$$

This proves the quantitative bound and hence the sequential claim. □

I.4 Closed convex cones and the dual-cone criterion

We now pass from single halfspaces to arbitrary closed convex cones.

Let $C_\Delta \subseteq V_\Delta$ be a closed convex cone. Its dual cone in V_Δ is

$$C_\Delta^* := \{w \in V_\Delta : \langle w,\, h \rangle_{\mathcal{H}} \geq 0,\ \forall\, h \in C_\Delta\}.$$

By the bipolar theorem for closed convex cones in Hilbert spaces,

$$C_\Delta = \{h \in V_\Delta : \langle w,\, h \rangle_{\mathcal{H}} \geq 0,\ \forall\, w \in C_\Delta^*\}. \tag{I.1}$$

Theorem I.10 (Dual-cone observability criterion)**.** *Let* $C_\Delta \subseteq V_\Delta$ *be a closed convex cone. Then the following are equivalent.*

1. *Membership in* C_Δ *is operator-identifiable through* A_Δ *on* V_Δ.

2. *Every dual supporting functional is observable:*
$$C_\Delta^* \subseteq \mathcal{H}_{\text{obs}} = \ker(A_\Delta)^\perp.$$

3. *For every $h, h' \in V_\Delta$,*
$$A_\Delta h = A_\Delta h' \quad \Longrightarrow \quad \big(h \in C_\Delta \iff h' \in C_\Delta\big).$$

Proof. The equivalence of (1) and (3) is exactly the definition of operator-identifiability.

We prove $(2) \Rightarrow (3) \Rightarrow (2)$.

Implication $(2) \Rightarrow (3)$**.** Assume $C_\Delta^* \subseteq \mathcal{H}_{\text{obs}} = \ker(A_\Delta)^\perp$. Let $h, h' \in V_\Delta$ satisfy
$$A_\Delta h = A_\Delta h'.$$
Then
$$k := h - h' \in \ker(A_\Delta).$$
Fix any $w \in C_\Delta^*$. Since $w \in \ker(A_\Delta)^\perp$,
$$\langle w,\, k\rangle_{\mathcal{H}} = 0,$$
hence
$$\langle w,\, h\rangle_{\mathcal{H}} = \langle w,\, h'\rangle_{\mathcal{H}}.$$
Therefore the two points have exactly the same pairing with every $w \in C_\Delta^*$. By the bipolar representation (I.1),
$$h \in C_\Delta \iff h' \in C_\Delta.$$
This proves (3).

Implication $(3) \Rightarrow (2)$**.** Assume (3), and let $w \in C_\Delta^*$. Suppose for contradiction that $w \notin \mathcal{H}_{\text{obs}} = \ker(A_\Delta)^\perp$. Then there exists $k \in \ker(A_\Delta)$ such that
$$\langle w,\, k\rangle_{\mathcal{H}} \neq 0.$$
Replacing k by $-k$ if necessary, assume
$$\langle w,\, k\rangle_{\mathcal{H}} < 0.$$
Now
$$A_\Delta k = 0 = A_\Delta 0.$$
Since $0 \in C_\Delta$ and $w \in C_\Delta^*$, the bipolar theorem implies that any $h \in C_\Delta$ must satisfy $\langle w,\, h\rangle \geq 0$. But
$$\langle w,\, k\rangle_{\mathcal{H}} < 0,$$
so $k \notin C_\Delta$. Thus
$$A_\Delta k = A_\Delta 0, \qquad 0 \in C_\Delta, \qquad k \notin C_\Delta,$$
contradicting (3). Therefore $w \in \mathcal{H}_{\text{obs}}$, and since $w \in C_\Delta^*$ was arbitrary, we obtain
$$C_\Delta^* \subseteq \mathcal{H}_{\text{obs}}.$$
□

Corollary I.11 (Generated-cone form)**.** *Suppose there exists a family $W \subseteq \mathcal{H}_{\text{obs}}$ such that*
$$C_\Delta = \{h \in V_\Delta : \langle w,\, h\rangle_{\mathcal{H}} \geq 0,\ \forall\, w \in W\}.$$
Then membership in C_Δ is operator-identifiable through A_Δ on V_Δ.

Proof. By definition of C_Δ, every $w \in W$ belongs to the dual cone C_Δ^*. Hence the closed convex cone generated by W inside V_Δ is contained in C_Δ^*. Conversely, the representation of C_Δ as an intersection of halfspaces induced by W implies that W is dual-generating for C_Δ, so C_Δ^* is contained in the closed cone generated by W. In particular, every element of C_Δ^* lies in the closed subspace $\mathcal{H}_{\text{obs}}$, because $\mathcal{H}_{\text{obs}}$ is closed and contains W. Thus $C_\Delta^* \subseteq \mathcal{H}_{\text{obs}}$, and Theorem I.10 applies. □

Remark I.12 (Recovery of the finite-dimensional facet theorem). When V_Δ is finite-dimensional and C_Δ is polyhedral with facet normals $w_1, \ldots, w_r$, one has
$$C_\Delta^* = \text{cone}\{w_1, \ldots, w_r\}.$$
Therefore
$$C_\Delta^* \subseteq \mathcal{H}_{\text{obs}} \quad \iff \quad w_j \in \mathcal{H}_{\text{obs}} = \text{range}(A_\Delta^*), \qquad j = 1, \ldots, r,$$
so Theorem I.10 reduces exactly to the facet-observability criterion of the main text.

I.5 Takeaway

The infinite-dimensional extension reveals the underlying structure of the whole paper:

1. Exact order identifiability is fundamentally a *quotient phenomenon*: welfare labels must descend through the observation operator.
2. The correct infinite-dimensional replacement for the row space is the *observable dual subspace*
$$\mathcal{H}_{\mathrm{obs}} = \overline{\mathrm{range}(A_\Delta^*)} = \ker(A_\Delta)^\perp.$$
3. For arbitrary closed convex cones, exact identifiability is governed by the dual cone:
$$\text{exact identifiability} \quad \Longleftrightarrow \quad C_\Delta^* \subseteq \mathcal{H}_{\mathrm{obs}}.$$

Thus the finite-dimensional theorem package of the main paper is not merely a matrix calculation. It is the polyhedral shadow of a more general functional analytic principle.