Counterfactual Sufficiency

State Representation and Response Geometry

Chae-Yeon Xon

Yonsei University

September 2026

Abstract

Which additional state coordinates preserve a specified policy calculation? I study this question relative to a supplied representation, target loss, and admissible feature class. Restricted response derivatives give a local benchmark: discarded eigenvalues determine leading decoder risk under shrinking perturbations. Global coordinate choice also depends on the variation left within feature level sets. A conditional-generator certificate bounds this risk, and its spectral slack enters the comparison between estimated certificate minimization and a risk-optimal coordinate map. The bound separates that slack from score estimation, sampling, and optimization error. A finite-graph counterpart evaluates fitted-decoder residuals. Occupational-choice and heterogeneous-agent exercises illustrate the consequences for model-generated policy responses. The analysis keeps representation loss, fitted prediction error, and identification of the underlying response map distinct.

JEL Codes: C14, C52, C61, C63, D81

Keywords: counterfactual sufficiency, state representation, response geometry, dimension reduction, active subspaces, structural models

Yonsei University. Email: eunixon@yonsei.ac.kr. This draft is circulated for discussion; comments and suggestions are welcome.

1. Introduction

A state record is useful for the policy questions it can support. Workers with the same schooling and total civilian experience may differ in their mix of experience and latent type. A wage or schooling intervention can make those differences relevant. A model that starts from the compressed record must then choose which additional coordinates to retain, subject to its measurement or computational budget.

I study this choice for a supplied state map q,[1] a family of policy responses, a target loss, a state distribution, and an admissible coordinate class. An augmentation b gives the decoder the record (q, b). The best decoder is the conditional mean of the weighted target, so the record's population risk is the target variation left among states with the same coordinates. The target specifies which errors matter. The coordinate class specifies which records are feasible. The state-design comparison is conditional on the supplied response surface.

The exact result comes first. Theorem 1 constructs the canonical quotient that retains the supplied record and every declared target response: every exact representation factors through this quotient. At a smooth state, any first-order augmentation needs at least $\operatorname{rank} A_{\mathfrak{T}}(s)$ coordinates, and an augmentation of that dimension attains closure at the state. Exact sufficiency and minimum local repair are therefore two parts of one target-relative statement. Compatibility across a region and membership in a feasible coordinate class remain additional requirements.

Restricted response derivatives give a tractable local benchmark. They measure target movements along state directions that leave q fixed. Their weighted Gram operator, $H_{\mathfrak{T}}(s)$, is an active-subspace operator on those omitted directions.[2] The spectral frontier gives the minimum linear loss, its attaining coordinates, and the exact zero-loss dimension. Over nonlinear continuous records, the leading worst-case term remains $\rho\sigma_{d+1}$ up to the stated quadratic remainder. Theorem 3 gives the eigenvalue tail a prediction interpretation under the shrinking perturbation law. A positive eigengap additionally controls the orientation of a minimizing projection.

At a finite scale, the distribution of target values within each retained record matters as well. A low-amplitude oscillation can have large derivatives. Retaining that direction may preserve most derivative energy while leaving most target variation unresolved. This raises the question studied here: when does minimizing a derivative-based upper bound give a useful comparison with a risk-optimal record?

[1] The supplied map may be an observed record, an existing model state, or a previous-stage approximation. The admissible augmentation class records which additional coordinates the analyst can construct.

[2] The restriction to the tangent space of a q-fiber removes directions already recorded by the supplied state map.

Theorems 4 and 5 answer this question under conditional spectral-gap and score-complexity conditions. A generator on each connected feature fiber yields a risk certificate. The first theorem gives an exact identity for its gap from decoder risk, characterizes equality, and gives a sharp upper bound. The second transfers that gap into an oracle inequality over the declared coordinate class, alongside score-estimation, sampling, and optimization errors. Accurate estimation of the certificate and a small certificate gap for the risk-optimal record have separate roles.

The implementation adds a fitted decoder and an evaluation design. Its prediction error contains both representation risk and the decoder's discrepancy from the conditional mean. A graph construction supplies a finite-state certificate for those fitted residuals under a specified resolution and edge rule.[3] Direct out-of-fold risk evaluates the prediction procedure more directly. The two scores can therefore select different members of the same coordinate menu.

The occupational-choice exercise keeps the supplied record, 55-target vector, decoder, and three-coordinate budget fixed. Across 20 core-state splits, two parameterizations, and two derivative-estimation runs, direct out-of-fold risk selects the held-out best candidate in 80 of 80 comparisons; the graph certificate does so in 63. Their median RMSE gains over the local counterfactual spectrum are 4.29 and 4.16 percent.[4] A heterogeneous-agent transition exercise then shows how changing the target loss changes the distributional directions worth retaining.

These calculations condition on the structural response map. Empirical use also requires evidence that identifies that map and supports its first-stage inference. A closed reduced Bellman calculation has further reward and transition requirements.[5] The paper states these conditions separately so that each representation claim can be read with its target, coordinate class, and source of evidence.

Related literature. Gradient-based ridge approximation provides the local spectral and conditional Poincaré tools (Constantine, Dow, and Wang 2014; Zahm et al. 2020). Constantine et al. (2017) show how oscillation separates derivative concentration from approximation accuracy. Nonlinear feature methods align a feature Jacobian with target gradients and fit a decoder on the learned coordinates (Bigoni, Marzouk, Prieur, and Zahm 2022). Verdière, Prieur, and Zahm (2025, Propositions 2 and 12) give Poincaré bounds for diffeomorphism-based scalar and vector-valued feature learning. These results organize dimension reduction around a feature class and its error bound.

[3] Changing the node sample, neighborhood rule, or graph resolution changes the finite diagnostic while leaving the population certificate defined below unchanged.

[4] The comparisons reuse one model-generated target surface across overlapping state splits. The 80 comparisons describe design repetitions on that surface.

[5] Target recovery alone concerns the declared response family. Planning closure additionally requires rewards and reduced transitions to factor through the retained record.

Here the supplied record is held fixed, derivatives are restricted to its omitted directions, and certificate slack is carried into a risk comparison over admissible augmentations. The conditional spectral decomposition of that slack uses standard symmetric-generator calculus (Bakry, Gentil, and Ledoux 2014). Its role is to expose the approximation term incurred when the certificate is used to choose a record. The resulting bound requires a positive conditional gap, the stated generator-domain conditions, and uniform control of estimated scores. It gives a class-relative comparison under these conditions.

Dynamic structural methods compute counterfactual responses from behavior and state transitions (Rust 1987; Bajari, Benkard, and Levin 2007); identification results characterize the objects recovered from such observations (Magnac and Thesmar 2002). Counterfactual-distribution inference (Chernozhukov, Fernández-Val, and Melly 2013) and sensitivity to latent-distribution assumptions (Christensen and Connault 2023) concern uncertainty about the response map. The state-design problem takes that map as its target. The response-field extension records where additional identification and sampling conditions enter when its derivatives are estimated.

Within dynamic discrete choice, Kalouptsidi, Scott, and Souza-Rodrigues (2021) give necessary and sufficient matrix conditions for counterfactual choice probabilities to be invariant to observationally unidentified payoffs. Their result characterizes identification inside a specified DDC structure. The canonical quotient and repair rank here are defined for a supplied record and an arbitrary declared response family.

Nonlinear realization and balanced model reduction study state descriptions that support a system's behavior (Sussmann 1976; Kotta, Moog, and Tõnso 2018; Moore 1981). The target-recovery problem here preserves a declared response family; planning closure additionally preserves rewards and reduced transitions. Rank tests and factor-number estimation address related dimension questions with different population objects (Chen et al. 2014; Bai and Ng 2002). Policy learning supplies another setting for class-relative risk comparisons (Athey and Wager 2021). Sequence-space and semistructural methods offer complementary ways to obtain the policy responses used in such comparisons (Auclert et al. 2021; Beraja 2023; Hebden and Winkler 2026).

For heterogeneous-agent estimation, Liu and Plagborg-Moller (2023) show that cross-sectional moments are sufficient under an exponential-family sampling density and give conditions that transport this sufficiency from latent micro states to observed variables. That model-specific result motivates careful separation between an available sufficient statistic and the minimum target-relative coordinate count studied here.

The paper moves from the state-design objective to local risk, global selection, implementation, and model-based comparisons. Proofs appear beside the three main theorems; supporting geometry, inference, and quantitative design details appear in the Online Appendix.

2. State representations and policy targets

The state-design analysis is conditional on a specified structural response map. The dynamic model below supplies one setting in which such a map can be computed. The target family, loss, and admissible coordinates then determine what its representation must preserve.

2.1. Structural environment and supplied state

Let $S_t \in \mathcal{S} \subseteq \mathbb{R}^K$ be a structural state, $A_t \in \mathcal{A}$ an action, $p \in \mathcal{P} \subseteq \mathbb{R}^{d_{\mathcal{P}}}$ a policy environment, and $\theta \in \Theta \subseteq \mathbb{R}^{p_\theta}$ a structural parameter. A stationary infinite-horizon discrete-choice model can be written as

$$V_\theta(s, p) = \log \sum_{a \in \mathcal{A}} \exp \left\{ \pi_\theta(a, s, p) + \beta \mathbb{E}_\theta [V_\theta(S', p) \mid s, a, p] \right\}, \tag{1}$$

where the log-sum form corresponds to independent type-I extreme-value choice shocks. The definitions apply to structural models that generate a differentiable counterfactual map. The parameter θ_0 is held fixed in the representation results; response data enter later through estimation of the response field. A supplied representation is a C^1 map

$$q : \mathcal{S} \to \mathcal{X}. \tag{2}$$

The map can collect observed aggregates, administrative records, state variables retained by a computational approximation, or a previous-stage representation. Fix a state region on which q has constant rank. The vertical space is

$$\mathcal{V}_s = \ker Dq(s). \tag{3}$$

A Riemannian metric on $\mathcal{S}$ supplies units for state perturbations. Within one chart, let $B(s) : \mathbb{R}^k \to \mathcal{V}_s$ be an isometry, where $k = \dim \mathcal{V}_s$.

2.2. Counterfactual family

A target view is $t = (u, \delta, c, h)$, where u is a policy direction, δ is a policy displacement, c indexes an outcome selector C_c, and h indexes a horizon. Let p_0 be the baseline policy. The finite target response is

$$R_t(s; \theta) = C_c \left[T_h(s, p_0 + \delta u; \theta) - T_h(s, p_0; \theta) \right]. \tag{4}$$

Let $\mathcal{T}$ be a finite family of views. Stack R_t into $\boldsymbol{R}_{\mathcal{T}}$. A continuous family can be treated by replacing sums with integrals.

DEFINITION 1 (State-design contract). *A state-design contract is*

$$\mathcal{E} = \big(q, \mathcal{T}, \{\omega_t, W_t\}_{t\in\mathcal{T}}, g, \mu, \varepsilon, \{\mathfrak{B}_d\}_{d\geq 0}\big), \tag{5}$$

where $W_t \geq 0$ is a declared target loss, g is the state metric, μ is the state distribution used for integrated risk, ε is a target-loss tolerance, and $\mathfrak{B}_d$ is the admissible class of d-coordinate augmentations. The contract is fixed before coordinate estimation and risk evaluation.

The exact quotient depends on $(q, \mathcal{T})$. Local spectra additionally depend on the state and target metrics. Integrated risk and selected dimension additionally depend on $(\mu, \varepsilon, \mathfrak{B}_d)$. Reporting results under economically relevant alternative contracts gives a sensitivity frontier for the state conclusion.

2.3. Population prediction criterion

Define the target map in its declared metric by

$$\boldsymbol{F}_{\mathcal{T}}(s) = \left\{\sqrt{\omega_t} W_t^{1/2} R_t(s)\right\}_{t\in\mathcal{T}} \tag{6}$$

Suppose $\boldsymbol{F}_{\mathcal{T}} \in L^2(\mu)$. For a measurable feature z, define

$$\mathfrak{R}(z) = \inf_h \mathbb{E}_\mu\left[\|\boldsymbol{F}_{\mathcal{T}}(S) - h(z(S))\|_2^2\right]. \tag{7}$$

The minimizing decoder is the conditional mean of $\boldsymbol{F}_{\mathcal{T}}(S)$ given $z(S)$. Thus $\mathfrak{R}(z)$ is the target variation left within the representation. Zero risk gives weighted-target factorization μ-almost surely; the exact equivalence relation is defined on every state in its stated domain.

2.4. Exact and loss-relevant distinctions

DEFINITION 2 (Counterfactual equivalence). *For $s, s' \in \mathcal{S}$, write $s \sim_{\mathcal{T},q} s'$ when $q(s) = q(s')$ and $\boldsymbol{R}_{\mathcal{T}}(s;\theta_0) = \boldsymbol{R}_{\mathcal{T}}(s';\theta_0)$.*

The quotient $\mathcal{S}/\sim_{\mathcal{T},q}$ records the distinctions in the structural state that are retained by q or used by the declared target family.

DEFINITION 3 (Loss-relevant counterfactual equivalence). *For $s, s' \in \mathcal{S}$, write $s \sim_{\mathcal{T},W,q} s'$ when $q(s) = q(s')$ and $W_t^{1/2} R_t(s;\theta_0) = W_t^{1/2} R_t(s';\theta_0)$ for every $t \in \mathcal{T}$.*

The loss-relevant quotient $\mathcal{Z}_W^\star = \mathcal{S}/\sim_{\mathcal{T},W,q}$ is a coarsening of the exact quotient. The two coincide when every W_t is positive definite on the span of attainable target differences. A singular decision loss retains the target combinations that affect the declared local decision problem.

THEOREM 1 (Canonical target quotient and minimum local repair). *Let* $\pi^\star : \mathcal{S} \to \mathcal{Z}^\star = \mathcal{S}/\!\sim_{\mathcal{T},q}$ *be the quotient map. There are unique maps* $\bar{q}$ *and* $\bar{R}_{\mathcal{T}}$ *on* $\mathcal{Z}^\star$ *such that*

$$q = \bar{q} \circ \pi^\star, \qquad \boldsymbol{R}_{\mathcal{T}} = \bar{R}_{\mathcal{T}} \circ \pi^\star. \tag{8}$$

Let $z : \mathcal{S} \to \mathcal{Z}$ *be any representation for which* $q = q_z \circ z$ *and* $\boldsymbol{R}_{\mathcal{T}} = R_z \circ z$*. Then there is a unique map* $h : z(\mathcal{S}) \to \mathcal{Z}^\star$ *satisfying*

$$\pi^\star = h \circ z. \tag{9}$$

At a state s*, define the stacked restricted derivative*

$$A_{\mathcal{T}}(s) = D_s \boldsymbol{R}_{\mathcal{T}}(s; \theta_0) B(s). \tag{10}$$

Every C^1 *augmentation* $b : \mathcal{S} \to \mathbb{R}^d$ *satisfying first-order target closure,*

$$\ker D(q, b)(s) \subseteq \ker D\boldsymbol{R}_{\mathcal{T}}(s; \theta_0), \tag{11}$$

satisfies $d \geq \operatorname{rank} A_{\mathcal{T}}(s)$*. An augmentation of dimension* $\operatorname{rank} A_{\mathcal{T}}(s)$ *attains first-order closure at* s*.*

The first part of the proposition factors the targets through equivalence classes. The second turns target variation along the vertical space into a coordinate count at a given state. Its attainment statement concerns first-order closure there. A feature used over a region must also satisfy compatibility and within-fiber conditions. The structural map that supplies these derivatives remains an input to this representation result.

PROPOSITION 1 (Finite-state coordinate count). *Let* $\mathcal{S}$ *be finite. If* q *fails to carry the exact target quotient, there is a scalar map* $b : \mathcal{S} \to \mathbb{R}$ *such that* (q, b) *carries it exactly.*

One unrestricted real coordinate can label every class of a finite quotient. Complexity in a discrete representation must therefore be defined through a declared cost, such as quotient cardinality, code length, a measurement dictionary, or regularity on a supplied graph.

2.5. Policy-information operator

Give target view t a weight $\omega_t > 0$ and a target loss $W_t \geq 0$. Define

$$A_t(s) = D_s R_t(s; \theta_0) B(s) \tag{12}$$

and

$$H_{\mathcal{T}}(s) = \sum_{t \in \mathcal{T}} \omega_t A_t(s)' W_t A_t(s). \tag{13}$$

The weights specify the declared target loss. Sample size and design precision enter a separate information operator.

PROPOSITION 2 (Kernel of the target operator). *The operator in* (13) *satisfies*

$$\ker H_{\mathcal{T}}(s) = \bigcap_{t \in \mathcal{T}} \ker\{W_t^{1/2} A_t(s)\}, \tag{14}$$

and

$$\operatorname{rank} H_{\mathcal{T}}(s) = \operatorname{rank} A_{\mathcal{T},W}(s), \tag{15}$$

where $A_{\mathcal{T},W}$ *stacks* $\sqrt{\omega_t} W_t^{1/2} A_t$. *If every* W_t *is positive definite on* $\operatorname{im} A_t(s)$*, then* $\ker H_{\mathcal{T}}(s) = \cap_t \ker A_t(s)$ *and* $\operatorname{rank} H_{\mathcal{T}}(s) = \operatorname{rank} A_{\mathcal{T}}(s)$. *Under a change of vertical orthonormal frame* $\widetilde{B}(s) = B(s)O(s)$, $\widetilde{H}_{\mathcal{T}}(s) = O(s)' H_{\mathcal{T}}(s) O(s)$.

The quotient tangent space can therefore be represented as

$$\mathcal{Q}_{\mathcal{T},W}(s) = \mathcal{V}_s / \{B(s) \ker H_{\mathcal{T}}(s)\}. \tag{16}$$

The map $B(s)$ carries the kernel from $\mathbb{R}^k$ into the vertical tangent space. The metric fixes the units of hidden-state perturbations.[6] Coordinate covariance, observability-Gramian, and Fisher-metric formulations cover geometric and distributional states. The finite-dimensional results use the stated frame.

[6] A coordinate rescaling carries the metric; fixing it changes the perturbation experiment and spectrum.

3. Local response geometry

The local criterion asks how much first-order target variation remains after a fixed number of vertical directions is retained. The resulting spectrum has a shrinking-neighborhood risk interpretation. Realizing those directions as a feature map introduces a separate geometric condition.

3.1. Dimension budgets

Let the ordered eigenvalues of $H_{\mathcal{T}}(s)$ be $\lambda_1(s) \geq \cdots \geq \lambda_k(s) \geq 0$. A rank-$d$ orthogonal projector P retains d vertical coordinates. Define the quadratic first-order target loss

$$\mathcal{L}_s(P) = \operatorname{tr}[(I - P)H_{\mathcal{T}}(s)]. \tag{17}$$

THEOREM 2 (Spectral compression frontier). *Among rank-d orthogonal projectors,*

$$\min_{\operatorname{rank} P = d} \mathcal{L}_s(P) = \sum_{j>d} \lambda_j(s), \tag{18}$$

and a minimizer is the projector $P_d(s)$ onto the leading d eigenvectors of $H_{\mathcal{T}}(s)$. If vertical perturbations satisfy $\|v\|_2 \leq \rho$, the smallest worst-case first-order target norm over rank-d linear encoders equals

$$\rho\sqrt{\lambda_{d+1}(s)}. \tag{19}$$

For a tolerance $\varepsilon \geq 0$, the quadratic effective dimension is

$$r_\varepsilon(s) = \min\left\{ d : \sum_{j>d} \lambda_j(s) \leq \varepsilon \right\}. \tag{20}$$

Zero first-order target loss is attainable with d coordinates if and only if $d \geq \operatorname{rank} H_{\mathcal{T}}(s)$. Hence $\operatorname{rank} H_{\mathcal{T}}(s)$ is the minimum loss-relevant coordinate budget at s.

The lower bound extends beyond linear encoding to every continuous d-coordinate encoder and decoder.[7] Its bound equals $\rho\sigma_{d+1}$ in the affine case and differs from that value by at most the declared quadratic remainder in the smooth case.

[7]Continuity rules out set-theoretic encodings that pack an entire high-dimensional state into one real number. The dimension bound therefore concerns records stable under the local perturbations used in the risk experiment.

3.2. Local prediction risk

Fix a supplied-fiber chart $\varphi_s : B^k_\rho \to q^{-1}(q(s))$ with $\varphi_s(0) = s$ and derivative $D\varphi_s(0) = B(s)$ in the chosen frame. Define the stacked weighted target increment

$$F_s(v) = \left\{\sqrt{\omega_t}W_t^{1/2}[R_t(\varphi_s(v)) - R_t(s)]\right\}_{t\in\mathcal{T}}, \qquad \|v\| \le \rho. \tag{21}$$

Let $\mathcal{A}^{\mathcal{T}}_s = DF_s(0)$ and assume the uniform expansion

$$F_s(v) = F_s(0) + \mathcal{A}^{\mathcal{T}}_s v + r_s(v), \qquad \sup_{\|v\|\le\rho} \|r_s(v)\| \le K_s\rho^2/2. \tag{22}$$

Let U be supported on B^k_1 and, for some $c > 0$, satisfy

$$\mathbb{E}U = 0, \qquad \mathbb{E}UU' = cI_k, \qquad \mathbb{E}[(I - P)U \mid PU] = 0 \tag{23}$$

for every rank-d orthogonal projector P under comparison. Spherical perturbation laws satisfy (23). For $\rho > 0$, define the optimal local decoding risk

$$\mathfrak{R}_{s,\rho}(P) = \inf_h \mathbb{E}\left[\|F_s(\rho U) - h(PU)\|_2^2\right]. \tag{24}$$

THEOREM 3 (Local spectral-risk expansion). *Under* (22) *and* (23)*, every rank-d orthogonal projector under comparison satisfies*

$$\left|\mathfrak{R}_{s,\rho}(P) - c\rho^2 \operatorname{tr}\{(I - P)H_{\mathcal{T}}(s)\}\right| \le K_s\rho^3\sqrt{c \operatorname{tr}\{(I - P)H_{\mathcal{T}}(s)\}} + \frac{K_s^2}{4}\rho^4. \tag{25}$$

Consequently,

$$\inf_{\operatorname{rank} P=d} \mathfrak{R}_{s,\rho}(P) = c\rho^2 \sum_{j>d} \lambda_j(s) + O(\rho^3). \tag{26}$$

If $\lambda_d(s) - \lambda_{d+1}(s) = \gamma_d(s) > 0$ *and* $P_{s,\rho}$ *minimizes* (24)*, then*

$$\|P_{s,\rho} - P_d(s)\|_{\mathrm{F}}^2 = O\left(\frac{\rho}{\gamma_d(s)}\right). \tag{27}$$

PROOF OF THEOREM 3. The conditional mean is the L^2-optimal decoder. Put $U_\perp = (I - P)U$. Equations (23) and (22) give

$$F_s(\rho U) - \mathbb{E}[F_s(\rho U) \mid PU] = \rho\mathcal{A}^{\mathcal{T}}_s U_\perp + \delta_\rho, \tag{28}$$

where $\delta_\rho = r_s(\rho U) - \mathbb{E}[r_s(\rho U) \mid PU]$. The conditional-variance identity gives

$$\mathbb{E}\|\delta_\rho\|_2^2 \le \mathbb{E}\|r_s(\rho U)\|_2^2 \le K_s^2 \rho^4/4. \tag{29}$$

Isotropy yields

$$\mathbb{E}\|\mathcal{A}_s^{\mathcal{T}} U_\perp\|_2^2 = c\,\mathrm{tr}\{(I-P)(\mathcal{A}_s^{\mathcal{T}})'\mathcal{A}_s^{\mathcal{T}}\} = c\,\mathrm{tr}\{(I-P)H_{\mathcal{T}}(s)\}. \tag{30}$$

Expansion of the squared norm and Cauchy–Schwarz give (25). The remainder is uniform over rank-d projectors because the residual trace is bounded by $\mathrm{tr}\, H_{\mathcal{T}}(s)$. Ky Fan's principle gives (26). If $P_{s,\rho}$ minimizes local risk, uniformity implies

$$\mathrm{tr}(P_d H_{\mathcal{T}}) - \mathrm{tr}(P_{s,\rho} H_{\mathcal{T}}) = O(\rho). \tag{31}$$

The eigengap inequality

$$\mathrm{tr}(P_d H_{\mathcal{T}}) - \mathrm{tr}(P H_{\mathcal{T}}) \ge \frac{\gamma_d(s)}{2}\|P - P_d\|_{\mathrm{F}}^2 \tag{32}$$

gives (27). □

The conditional-mean restriction in (23) makes the omitted linear response orthogonal to every decoder using PU. Isotropic covariance then turns its variance into a trace. The quadratic remainder controls the finite-radius error uniformly over the projectors under comparison. A positive eigengap is used for projector recovery, with existence of a minimizer maintained in that conclusion. The eigenvalue-tail risk expansion has its own conclusion even when an orientation is tied. The continuous-encoder frontier has the same leading term.

3.3. From directions to features

A leading projector specifies directions at each state. A feature map b specifies level sets. To connect the two, restrict attention to maps for which $Db|_{\mathcal{V}_s}$ has rank d, and let $P_b(s)$ project onto its retained vertical directions. The nonnegative orientation loss $\Delta_b(s)$ satisfies

$$\mathrm{tr}\{(I - P_b(s))H_{\mathcal{T}}(s)\} = \sum_{j>d} \lambda_j(s) + \Delta_b(s).$$

The spectral tail is a pointwise lower bound, and Δ_b is the additional loss associated with the chosen feature orientation. Realizing a smooth spectral field as coordinate level sets requires the Frobenius involutivity condition. The realization result separates an infimum bound from equality attained by an admissible coordinate map under that condition. These geometric requirements accompany the population risk criterion developed next.

4. Global decoder risk and coordinate selection

The population risk (7) averages target variation left within the level sets of a feature. Derivatives and the conditional state distribution bound that risk and identify the slack relevant for selecting a feature by an estimated upper bound. All comparisons use the stated state-design contract.

4.1. Why retained derivative variation can mislead

A coordinate rule selected from derivatives is ultimately used to predict target values. The following example holds the coordinate budget fixed and shows why the population distribution must enter that comparison.

PROPOSITION 3 (Local spectral concentration and global decoder risk). *For every integer $d \geq 1$, retained derivative share $\rho \in (0, 1)$, and risk ratio $K > 0$, there is a smooth scalar target on $[-1, 1]^{d+1}$ under the uniform distribution for which the leading d-dimensional eigenspace of the integrated derivative operator retains at least share ρ, while its optimal-decoder risk is at least K times the risk from another d-dimensional coordinate subspace.*

The oscillatory mechanism follows Constantine et al. (2017). Proposition 3 varies the amplitudes and frequencies to attain the stated derivative share and risk ratio at any retained dimension.

For a concrete case, let X, Y be independent uniform variables on $[-1, 1]$, let q be constant, and put $F(X, Y) = 0.05\phi_{40}(X) + \phi_1(Y)$, where $\phi_k(u) = \cos\{k\pi(u + 1)/2\}$. Retaining X captures 80 percent of integrated derivative energy. Its best-decoder risk is 0.5, compared with 0.00125 when Y is retained. The smaller-amplitude component has the larger derivatives. This same example will show how certificate slack affects the comparison of the two records.

4.2. Global decoding-risk certificate

Suppose a smooth submersion z admits conditional measures μ_ζ on its level sets $L_\zeta = z^{-1}(\zeta)$. Assume the scalar conditional Poincaré inequality

$$\int_{L_\zeta} |f - \mathbb{E}_{\mu_\zeta} f|^2 d\mu_\zeta \leq C(\zeta) \int_{L_\zeta} \|\nabla_{L_\zeta} f\|_g^2 d\mu_\zeta \tag{33}$$

for almost every ζ.

PROPOSITION 4 (Fiberwise target-risk certificate). *Suppose* (33) *holds, the coordinate restrictions of* $\boldsymbol{F}_{\mathcal{T}}$ *belong to* $H^1(L_\zeta, \mu_\zeta)$ *almost everywhere, and the expression below is integrable. Then*

$$\mathfrak{R}(z) \le \mathbb{E}_\mu \left[C(z(S)) \| D\boldsymbol{F}_{\mathcal{T}}(S) \Pi_{\ker Dz(S)} \|_{\mathrm{HS}}^2 \right]. \tag{34}$$

For $z = (q, b)$,

$$\mathfrak{R}(q, b) \le \mathbb{E}_\mu \left[C((q, b)(S)) \left\{ \sum_{j>d} \lambda_j(S) + \Delta_b(S) \right\} \right]. \tag{35}$$

If $\ker D(q, b) = \mathcal{K}_d$ *and* $C \le \overline{C}$ *almost everywhere, then*

$$\mathfrak{R}(q, b) \le \overline{C}\, \mathbb{E}_\mu \sum_{j>d} \lambda_j(S). \tag{36}$$

The Poincaré constant depends on the geometry and conditional distribution of each feature fiber. Equation (35) separates spectral truncation from the directional loss created by an admissible coordinate map. The argument follows gradient-based ridge approximation and nonlinear level-set feature learning (Zahm et al. 2020; Bigoni, Marzouk, Prieur, and Zahm 2022; Verdière, Prieur, and Zahm 2025). This is an upper bound under the conditional Poincaré inequality. Candidate selection also depends on the distance from that bound to the candidate's risk.

4.3. Disconnected and mixed state fibers

Suppose each regular level set of z is a countable union of connected components and let $C_z(S)$ be a measurable component label. Put

$$m(\zeta, c) = \mathbb{E}[\boldsymbol{F}_{\mathcal{T}}(S) \mid z(S) = \zeta, C_z(S) = c], \qquad \bar{m}(\zeta) = \mathbb{E}[\boldsymbol{F}_{\mathcal{T}}(S) \mid z(S) = \zeta]. \tag{37}$$

Assume a conditional Poincaré inequality on each component, with constant $C(\zeta, c)$.

PROPOSITION 5 (Mixed-fiber risk decomposition). *The optimal decoding risk satisfies*

$$\mathfrak{R}(z) \le \mathbb{E} \left[C(z(S), C_z(S)) \| D\boldsymbol{F}_{\mathcal{T}}(S) \Pi_{\ker Dz(S)} \|_{\mathrm{HS}}^2 \right] \tag{38}$$

$$+ \mathbb{E} \left[\| m(z(S), C_z(S)) - \bar{m}(z(S)) \|_2^2 \right]. \tag{39}$$

Before the Poincaré step, the two terms form the exact conditional-variance decomposition into variation within connected components and variation between their conditional target means. The weighted target factors through z, μ*-almost surely, when both right-hand terms vanish.*

The second term in (39) detects target distinctions across latent types, regimes, and disconnected strata. Within-stratum derivatives supply the first term.

4.4. Admissible-coordinate selection

Fix before estimation a nonempty class $\mathfrak{B}_d$ of smooth maps $b : \mathcal{S} \to \mathbb{R}^d$ for which $z_b = (q, b)$ has constant rank, $Db(s)|_{\mathcal{V}_s}$ has rank d, and the conditions of Proposition 5 hold. The class encodes the measurement and implementation restrictions placed on a state representation. Define

$$\mathfrak{R}_d^\star = \inf_{b \in \mathfrak{B}_d} \mathfrak{R}(q, b). \tag{40}$$

The class and its integrated risk complete this comparison. A diagnostic for the gap between the fiberwise Poincaré bound and decoder risk uses the nonnegative self-adjoint weighted Neumann generator $\mathcal{L}_{b,\zeta,c}$ associated with the conditional Dirichlet form on each connected component of a regular z_b-fiber. Assume compact resolvent, kernel consisting of constants on each connected component, and first nonzero eigenvalue $\nu_1(b, \zeta, c) > 0$. Put

$$r_b(S) = \boldsymbol{F}_{\mathfrak{T}}(S) - m(z_b(S), C_{z_b}(S)), \qquad \Gamma(b) = \mathbb{E}\|m(z_b(S), C_{z_b}(S)) - \bar{m}(z_b(S))\|_2^2. \tag{41}$$

The generator acts componentwise on r_b. Define the fiber certificate and its spectral-slack diagnostic by

$$\mathfrak{C}^\sharp(b) = \mathbb{E}\left[\nu_1(b, z_b(S), C_{z_b}(S))^{-1} \left\|D\boldsymbol{F}_{\mathfrak{T}}(S)\Pi_{\ker Dz_b(S)}\right\|_{\mathrm{HS}}^2\right] + \Gamma(b), \tag{42}$$

$$\mathfrak{D}(b) = \mathbb{E}\left[\left\|\left\{\nu_1(b, z_b(S), C_{z_b}(S))^{-1}\mathcal{L}_{b,z_b(S),C_{z_b}(S)} - I\right\} r_b\right\|_2^2\right]. \tag{43}$$

Write $\mathcal{A}_b r_b$ for the operator residual inside the norm in (43). For $z_b = (q, b)$, the derivative term in (42) equals the expectation of $\nu_1^{-1}\{\sum_{j>d}\lambda_j + \Delta_b\}$. The conditional generator yields a target-specific expression for the certificate gap.

THEOREM 4 (Exact fiber-certificate gap and sharp bound). *Suppose the conditional generators above have compact resolvent, r_b lies in their domains, and the displayed quantities, including the squared norms of $\mathcal{L}r_b$, are integrable. Then every $b \in \mathfrak{B}_d$ satisfies*

$$\mathfrak{C}^\sharp(b) - \mathfrak{R}(q, b) = \mathbb{E}\langle r_b, \mathcal{A}_b r_b\rangle_2 \geq 0, \tag{44}$$

and

$$0 \leq \mathfrak{C}^\sharp(b) - \mathfrak{R}(q, b) \leq \sqrt{\{\mathfrak{R}(q, b) - \Gamma(b)\}\mathfrak{D}(b)} \leq \sqrt{\mathfrak{R}(q, b)\mathfrak{D}(b)}. \tag{45}$$

$\mathfrak{C}^\sharp(b) = \mathfrak{R}(q,b)$ precisely when every nonzero conditional target residual lies in the first nonconstant eigenspace of its connected fiber, up to mixtures across eigenfunctions sharing the first eigenvalue. If $\mathfrak{R}(q,b) > \Gamma(b)$, the first upper bound in (45) *is attained precisely when $\mathcal{A}_b r_b = \kappa r_b$ in the direct-integral conditional L^2 space for one constant $\kappa \geq 0$. Residuals concentrated on one common normalized fiber frequency attain the bound.*

PROOF. Fix a connected component of a regular fiber and suppress its indices. Let $\{\phi_k\}_{k\geq 0}$ be an orthonormal eigenbasis of $\mathcal{L}$, with $\mathcal{L}\phi_k = \nu_k \phi_k$ and $0 = \nu_0 < \nu_1 \leq \nu_2 \leq \cdots$. Each scalar coordinate r_ℓ of the conditional target residual has mean zero and an expansion $r_\ell = \sum_{k\geq 1} a_{\ell k}\phi_k$. The conditional within-component risk, derivative certificate, and slack diagnostic are

$$V = \sum_{\ell,k\geq 1} a_{\ell k}^2, \tag{46}$$

$$C = \sum_{\ell,k\geq 1} \frac{\nu_k}{\nu_1} a_{\ell k}^2, \tag{47}$$

$$D = \sum_{\ell,k\geq 1} \left(\frac{\nu_k}{\nu_1} - 1\right)^2 a_{\ell k}^2. \tag{48}$$

The Dirichlet-form identity gives the middle expression. Therefore $C - V \geq 0$, and Cauchy–Schwarz gives

$$C - V = \sum_{\ell,k\geq 1} \left(\frac{\nu_k}{\nu_1} - 1\right) a_{\ell k}^2 \leq \sqrt{VD}. \tag{49}$$

Integrate over fiber values and component labels. The expansion for $C - V$ gives (44). Applying Cauchy–Schwarz in the direct-integral conditional L^2 space gives (45). The integrated within-component variance is $\mathfrak{R}(q,b) - \Gamma(b)$, while the between-component term enters both $\mathfrak{C}^\sharp(b)$ and $\mathfrak{R}(q,b)$ without approximation. The difference is zero exactly when every coefficient with $\nu_k > \nu_1$ vanishes. For a nonzero residual, equality in Cauchy–Schwarz holds exactly when $\mathcal{A}_b r_b = \kappa r_b$ for one nonnegative constant κ in that direct-integral space. □

Equation (44) is the conditional spectral decomposition of the Poincaré deficit, using the symmetric-generator calculus of Bakry, Gentil, and Ledoux (2014). The first nonconstant eigenvalue normalizes the within-component energy. Residual frequencies above it create slack. The between-component term $\Gamma(b)$ enters risk and certificate equally, so the diagnostic addresses the remaining, within-component approximation. Compact resolvent supplies the eigenbasis; the domain and integrability conditions justify the generator norms and their averaging.

Table 1 completes the oscillatory example. Uniform interval fibers use the Neumann generator. The certificate equals risk for the record X. Retaining Y leaves the high-frequency residual $0.05\phi_{40}(X)$; its certificate is 2 although its risk is 0.00125. The diagnostic in (43) is zero for X and $0.05^2(40^2-1)^2/2$ for Y. It records a loose certificate at the risk-improving coordinate.

TABLE 1. Derivative concentration, decoder risk, and certificate slack

Retained record	Derivative share	Decoder risk	Certificate $\mathfrak{C}^\sharp$	Slack $\mathfrak{C}^\sharp - \mathfrak{R}$
X	0.80	0.50000	0.50	0.00000
Y	0.20	0.00125	2.00	1.99875

Exact calculation for $F(X,Y) = 0.05\phi_{40}(X) + \phi_1(Y)$, independent uniform X, Y on $[-1,1]$, and a constant supplied record. Both candidates retain one coordinate. The certificate uses the conditional Neumann spectral gap. Values specialize the existing oscillatory construction; they are separate from the model experiments.

An affine target residual on a standard Gaussian fiber in orthonormal coordinates gives the complementary case. For its Ornstein–Uhlenbeck generator, $\mathcal{L}f = -\Delta f + x \cdot \nabla f$, a centered affine residual lies in the first nonconstant eigenspace, so $\mathfrak{D}(b) = 0$. On a Euclidean fiber with density p, the expression $\mathcal{L}f = -p^{-1}\operatorname{div}(p\nabla f)$ shows the inputs required to evaluate the diagnostic: target derivatives, conditional density scores, second derivatives, and the conditional gap. Theorems 4 and 5 use the stated positive-gap setting.

4.5. Selection over a prescribed class

The statistical bound uses a declared finite-complexity coordinate class. Let $\mathfrak{B}_d = \{b_\eta : \eta \in \Xi_d\}$, where Ξ_d lies in a Euclidean ball of radius D_Ξ in $\mathbb{R}^{p_d}$. Let $\psi_\eta(S)$ denote the nonnegative observation-level score in (42), including the between-component term, so that $\mathfrak{C}^\sharp(b_\eta) = \mathbb{E}\psi_\eta(S)$. For a positive score envelope M, assume

$$0 \le \psi_\eta(S) \le M, \qquad |\psi_\eta(S) - \psi_{\eta'}(S)| \le L\|\eta - \eta'\|_2 \tag{50}$$

almost surely. Cross-fitted first stages produce scores $\widehat{\psi}_{\eta,i}$ and

$$\widehat{\mathfrak{C}}_n(\eta) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\psi}_{\eta,i}, \qquad \rho_n = \sup_{\eta\in\Xi_d}\left|\frac{1}{n}\sum_{i=1}^{n}\{\widehat{\psi}_{\eta,i} - \psi_\eta(S_i)\}\right|. \tag{51}$$

For $\alpha \in (0,1)$, define

$$a_n(\alpha) = \frac{M}{\sqrt{n}} + M\sqrt{\frac{p_d \log\{1 + 4LD_\Xi\sqrt{n}/M\} + \log(2/\alpha)}{2n}}. \tag{52}$$

THEOREM 5 (Coordinate-selection oracle inequality). *For an optimization tolerance $\eta_n \ge 0$, let $\widehat{\eta}_d$ satisfy*

$$\widehat{\mathfrak{C}}_n(\widehat{\eta}_d) \le \inf_{\eta \in \Xi_d} \widehat{\mathfrak{C}}_n(\eta) + \eta_n. \tag{53}$$

Suppose $\eta_d^\star \in \Xi_d$ exists such that $b_d^\star = b_{\eta_d^\star}$ attains (40). *Suppose $S_1, \dots, S_n$ are i.i.d. from μ and the coordinate class and ideal score functions are fixed before this sample is drawn. Under* (50), *with probability at least $1 - \alpha$,*

$$0 \le \mathfrak{R}(q, b_{\widehat{\eta}_d}) - \mathfrak{R}_d^\star \le \sqrt{\{\mathfrak{R}_d^\star - \Gamma(b_d^\star)\}\mathfrak{D}(b_d^\star)} + 2\rho_n + 2a_n(\alpha) + \eta_n. \tag{54}$$

If $\mathfrak{D}(b_d^\star) \le \delta_d^2\{\mathfrak{R}_d^\star - \Gamma(b_d^\star)\}$, the first term is at most $\delta_d \mathfrak{R}_d^\star$. Writing $\overline{\mathfrak{D}}_d = \sup_{\eta \in \Xi_d} \mathfrak{D}(b_\eta)$, the same term is bounded by $\sqrt{M\overline{\mathfrak{D}}_d}$ because $\mathfrak{R}_d^\star \le M$.

PROOF. Put $P_n f = n^{-1}\sum_{i=1}^n f(S_i)$ and $Pf = \mathbb{E} f(S)$. When $L = 0$, all ψ_η agree almost surely and a single application of Hoeffding's inequality is bounded by (52). Suppose $L > 0$. Take an r-net of Ξ_d. Its cardinality is at most $(1 + 2D_\Xi/r)^{p_d}$. Hoeffding's inequality and a union bound give, with probability at least $1 - \alpha$,

$$\max_{\eta \text{ in the net}} |(P_n - P)\psi_\eta| \le M\sqrt{\frac{p_d \log(1 + 2D_\Xi/r) + \log(2/\alpha)}{2n}}. \tag{55}$$

The Lipschitz condition extends this bound from the net to Ξ_d at cost $2Lr$. Setting $r = M/(2L\sqrt{n})$ gives

$$\sup_{\eta \in \Xi_d} |(P_n - P)\psi_\eta| \le a_n(\alpha). \tag{56}$$

Consequently,

$$e_n := \sup_{\eta \in \Xi_d} |\widehat{\mathfrak{C}}_n(\eta) - \mathfrak{C}^\sharp(b_\eta)| \le \rho_n + a_n(\alpha). \tag{57}$$

The certificate sandwich and approximate minimization imply

$$\begin{aligned}
\mathfrak{R}(q, b_{\widehat{\eta}_d}) &\le \mathfrak{C}^\sharp(b_{\widehat{\eta}_d}) && (58)\\
&\le \widehat{\mathfrak{C}}_n(\widehat{\eta}_d) + e_n && (59)\\
&\le \widehat{\mathfrak{C}}_n(\eta_d^\star) + \eta_n + e_n && (60)\\
&\le \mathfrak{C}^\sharp(b_d^\star) + \eta_n + 2e_n && (61)\\
&\le \mathfrak{R}_d^\star + \sqrt{\{\mathfrak{R}_d^\star - \Gamma(b_d^\star)\}\mathfrak{D}(b_d^\star)} + \eta_n + 2\rho_n + 2a_n(\alpha). && (62)
\end{aligned}$$

The lower bound follows from (40). The score envelope and the certificate sandwich give $\mathfrak{R}_d^\star \le M$, so the uniform bound on $\mathfrak{D}$ gives the final assertion. □

The comparison uses a risk-optimal member of the declared class, whose existence is an assumption of the theorem. Its certificate slack is the approximation term. The other terms have different sources: ρ_n measures the uniform discrepancy introduced by estimated scores, $a_n(\alpha)$ controls sampling variation in the fixed ideal-score class, and η_n records numerical optimization tolerance. The proof first bounds those score errors uniformly and then transfers approximate certificate minimization to risk.

Cross-fitting enters the construction of $\widehat{\psi}$; a rate for ρ_n requires the relevant first-stage conditions. Uniform score and slack-estimation events, together with their failure probabilities, give the corresponding rate statement. The reportable bound there adds those events to the ideal-score sampling event. In the oscillatory example, certificate slack at the risk-optimal record remains present even if all three estimation and optimization terms vanish.

A constructive class can be defined on a declared q-adapted atlas. Within a chart (x, u) satisfying $q(x, u) = x$, let $\Psi_\eta(x, \cdot) : \mathbb{R}^k \to \mathbb{R}^k$ be a smoothly indexed diffeomorphism and set $b_\eta(x, u) = \pi_d \Psi_\eta(x, u)$. Compatibility restrictions on chart overlaps give a global feature. Affine maps, prespecified moment dictionaries, and fixed-architecture coupling flows provide finite-dimensional choices for Ξ_d. Equation (54) separates fiber-frequency slack, first-stage error, class complexity, and numerical optimization.

5. Implementation and response-field recovery

A representation comparison specifies the coordinate construction, decoder, selection score, and evaluation sample. The population results concern the best decoder for each representation. Computation and sampling determine how closely a fitted comparison approaches that criterion. In an implementation, first fix the policy experiments, target weights, supplied record, and coordinate budget. Construct candidate fields and train their decoders using the permitted training information. Then choose a candidate by the declared selection score and evaluate it on the designated evaluation states.[8]

5.1. Fitted decoders and independent evaluation

Let $z_b = (q, b)$ and $m_b(z) = \mathbb{E}[\boldsymbol{F}_{\mathcal{T}}(S) \mid z_b(S) = z]$. Conditional on a training sample that fixes b and a square-integrable decoder $\widehat{h}_b$, an independent evaluation state satisfies

$$\mathbb{E}\|\boldsymbol{F}_{\mathcal{T}}(S) - \widehat{h}_b(z_b(S))\|_2^2 = \mathfrak{R}(q, b) + \mathbb{E}\|m_b(z_b(S)) - \widehat{h}_b(z_b(S))\|_2^2. \tag{63}$$

[8] Using the evaluation loss to revise the candidate menu defines a new selection procedure and requires a fresh evaluation sample.

Conditional centering removes the cross term. Representation risk and decoder approximation therefore enter as separate, nonnegative quantities. Using the same decoder architecture across records holds the fitting rule fixed; the second term can still differ across records. Out-of-fold scores evaluate that fitted procedure under the specified fold construction. The squared-loss identity applies before the square root and normalization used to report RMSE. A held-out winner is the best evaluated member of the given menu, while $\mathfrak{R}_d^\star$ in Theorem 5 compares population best-decoder risks over the declared class.

5.2. Finite-state graph scores

The conditional spectral calculation has a finite-state analogue for fitted residuals. Fix a declared resolution for the supplied coordinates, which partitions the evaluation states into cells $g \in \mathcal{G}$. For each candidate b, choose a connected reversible weighted graph within every cell; its edge rule can depend on distances in b, and its nonnegative self-adjoint Laplacian $L_{b,g}$ has first nonzero eigenvalue $\nu_{1,b,g}$. The cell resolution, edge rule, and bandwidth are part of the state-design contract. Let $r_{b,g}$ be a target residual from an externally fitted or cross-fitted decoder, let $\bar{r}_{b,g}$ be its cell mean, and put $r^\circ_{b,g} = r_{b,g} - \bar{r}_{b,g}$. Write $\|\cdot\|_g$ for the cell empirical norm and let π_g be the cell frequency. Singleton cells retain only the residual-mean term; their centered energy and diagnostic contributions are defined to be zero.

PROPOSITION 6 (Finite-graph generator certificate). *Define*

$$\widehat{\mathfrak{R}}_G(b) = \sum_{g\in\mathcal{G}} \pi_g \left\{ \|r^\circ_{b,g}\|_g^2 + \|\bar{r}_{b,g}\|_2^2 \right\}, \tag{64}$$

$$\widehat{\mathfrak{E}}_G(b) = \sum_{g\in\mathcal{G}} \pi_g \left\{ \nu_{1,b,g}^{-1} \langle r^\circ_{b,g}, L_{b,g} r^\circ_{b,g} \rangle_g + \|\bar{r}_{b,g}\|_2^2 \right\}, \tag{65}$$

$$\widehat{\mathfrak{D}}_G(b) = \sum_{g\in\mathcal{G}} \pi_g \left\| \{\nu_{1,b,g}^{-1} L_{b,g} - I\} r^\circ_{b,g} \right\|_g^2. \tag{66}$$

Then

$$0 \le \widehat{\mathfrak{E}}_G(b) - \widehat{\mathfrak{R}}_G(b) = \sum_{g\in\mathcal{G}} \pi_g \langle r^\circ_{b,g}, \{\nu_{1,b,g}^{-1} L_{b,g} - I\} r^\circ_{b,g} \rangle_g \tag{67}$$

$$\le \sqrt{\left\{\widehat{\mathfrak{R}}_G(b) - \widehat{\Gamma}_G(b)\right\} \widehat{\mathfrak{D}}_G(b)}, \tag{68}$$

where $\widehat{\Gamma}_G(b) = \sum_g \pi_g \|\bar{r}_{b,g}\|_2^2$. Conditional on an auxiliary sample used to construct a finite candidate menu, graphs, and decoders, suppose n_G independent evaluation clusters contribute *scores in* $[0, M_G]$ *to* (65). *For a menu* $\mathfrak{B}_{d,G}$, *with probability at least* $1 - \alpha$,

$$\sup_{b \in \mathfrak{B}_{d,G}} \left|\widehat{\mathfrak{C}}_G(b) - \mathbb{E}\widehat{\mathfrak{C}}_G(b)\right| \leq M_G \sqrt{\frac{\log\{2|\mathfrak{B}_{d,G}|/\alpha\}}{2n_G}}. \tag{69}$$

Proposition 6 applies graph-Laplacian spectral calculus (Chung 1997) and replaces conditional density-score and boundary estimation by a declared graph on a discrete structural state grid. The graph certificate remains an upper bound for the cross-fitted residual risk on that grid. Its empirical role is candidate selection within the declared graph, coordinate resolution, decoder, and target metric. The finite-grid inequality applies to those fitted residuals. Relating it to the population conditional generator requires a separate graph-approximation argument. The concentration statement further assumes the auxiliary-sample construction and independent bounded evaluation clusters stated in the proposition. A three-way split can separate construction, scoring, and final evaluation when the resulting cluster scores satisfy those conditions. The deterministic graph identity can be checked on a fixed grid independently of this sampling claim. The reported model exercise uses its own training/test design.

The numerical comparisons therefore report the coordinate rule, its decoder, the construction and evaluation samples, and the score used to choose among candidates. Agreement with a held-out menu winner measures that comparison's selection performance. Repeated partitions of a fixed structural target surface evaluate sensitivity to the split; the independent experimental unit for inference remains the one specified by the sampling design.

5.3. Recovering the response field

A solved structural model supplies the target derivatives directly. A data-based implementation requires a first stage for the same target derivatives under the relevant identification and sampling conditions. The reconstruction combines local derivative estimates over a declared state cover. Its uniform error separates derivative estimation error, its quadratic product, and interpolation error (Proposition A10).

Cross-fitted quadratic products use the conditional centering and cross-product restrictions of Lemma A1; sample splitting is part of that construction. Cover conditions link the scattered evaluations to the recovered field. These requirements are distinct from the spectral calculation applied to a known target map.

Scope of the recovered field. Locally varying frames, eigenspace recovery, and response information enter through the structural criterion. The field in (A57) is a population object recovered from response data. A model-generated field can instead be calculated after behavioral estimation. The two implementations share a target and a metric while placing different requirements on the evidence.

Spectral-tail inference carries the trace and the operator jointly. Its assumptions include a joint first-stage limit and a negligible sieve bias. A numerical directional bootstrap accommodates a positive repeated boundary eigenvalue. Simultaneous upper bounds then support a confidence-protected coordinate count for a declared loss tolerance. These are conditional inferential results; the numerical examples evaluate specified model-generated fields.

5.4. Target recovery and Bellman closure

Counterfactual evaluation uses the exact quotient. A reduced Bellman calculation additionally uses reward and transition closure. Let $z = z(s)$ be a candidate reduced state. Maintain $0 < \beta < 1$, a finite action set, and uniformly bounded flow rewards on the stated state and policy region. These conditions give the unique bounded fixed point used by the discounted Bellman argument.

PROPOSITION 7 (Planning closure). *Suppose that for every action a and policy p in the computational family,*

$$\pi_\theta(a, s, p) = \bar{\pi}_\theta(a, z(s), p) \tag{70}$$

and the pushforward transition satisfies

$$\Pr_\theta\{z(S') \in B \mid s, a, p\} = \bar{P}_\theta(B \mid z(s), a, p) \tag{71}$$

for each measurable B. Then the Bellman value in (1) *factors as $V_\theta(s, p) = \bar{V}_\theta(z(s), p)$ and the associated logit conditional choice probabilities factor through $z(s)$.*

The proposition connects state compression to the state-equivalence conditions used in MDP model minimization (Givan, Dean, and Greig 2003). A target quotient can also be used after solving the full behavioral model to reduce counterfactual response storage and evaluation.

Multilevel and multi-index corrections specify the accuracy requirements for the computational implementation.

6. Equal-budget policy-response comparisons

The exercises ask whether representations preserve the responses generated by a specified model. The occupational-choice comparison evaluates equal-budget coordinates on held-out target vectors and distinguishes direct risk selection from the graph certificate. The transition-economy calculation studies local distributional directions and a decision-induced loss. Both use the declared state-design contracts.

6.1. Occupational choice and equal-budget prediction

The local spectrum supplies a relaxation whose global performance depends on the feature class, perturbation distribution, and decoder. The comparison therefore holds the supplied state, target vector, coordinate budget, and decoder fixed. The KW97 fixed-model exercise applies this distinction in a dynamic occupational-choice model. The exercise solves five policies on aligned state spaces under two distributed parameterizations. The policies change the schooling reward, white-collar wage, blue-collar wage, military wage, and both civilian wages. The parameterizations are held fixed, and all policy responses are generated by the model. Each parameterization contains 6,520 core states and four latent types.[9] The target vector contains 55 state-level policy responses: EMAX, five choice probabilities, and five alternative-specific mean values under each policy. Core states enter a 70/30 training-test split. The primary supplied representation records period, schooling, total civilian experience, and lagged-choice indicators. Its fiber contains the civilian experience mix, military experience, and three latent-type contrasts. Every method carries the same supplied representation. Counterfactual-fiber, baseline-value-fiber, and unrestricted-output methods receive three added coordinates and use the same quadratic ridge decoder. The full-state row uses all five omitted coordinates. Table 2 averages held-out target RMSE across two derivative-estimation runs.

TABLE 2. Held-out prediction under equal-budget KW97 representations

Representation	Distributed respy	Distributed original
Counterfactual fiber, $d = 3$	0.7356	0.7011
Baseline-value fiber, $d = 3$	0.6758	0.6844
Unrestricted output, $d = 3$	0.7121	0.7065
Full state, $d = 5$	0.6488	0.6435

Entries are mean held-out RMSE for the common 55-component policy-response vector. Representations use the same supplied map, state split, training-standardized metric, decoder, and simulation design.

Across the supplied-representation specifications, the counterfactual-fiber coordinates remain above the best equal-budget competitor under the stricter cases. In the primary specification, the first three counterfactual directions retain about 78–83 percent of positive derivative trace, yet their global held-out RMSE exceeds the baseline-value comparator by 8.9 percent under the respy parameterization and 2.4 percent under the original parameterization. All methods preserve the ranking of mean EMAX policy re-

[9] All four types belonging to one core state are assigned to the same training or evaluation fold.

sponses. The comparison illustrates the distinction in Proposition 3. Retained derivative share describes the local response geometry; held-out RMSE evaluates the fitted decoder. The latter also reflects the second term in (63).

The finite-graph exercise implements Proposition 6 on the same split. Two training folds produce out-of-fold residuals from the common quadratic ridge decoder. Exact cells of the primary supplied representation fix period, schooling, total civilian experience, and lagged choice. Within each cell, a complete reversible graph assigns inverse-distance weights in the candidate's added coordinates; the bandwidth is the median positive within-cell distance. The graph certificate, candidate menu, target normalization, and decoder use training states. Held-out targets enter only the final comparison. The menu contains the counterfactual-fiber spectrum, baseline-value spectrum, and conventional coordinates at budgets $d = 1, 2, 3$.

The main comparison repeats the construction across twenty core-state partitions; the single-split diagnostics provide a reference case.

The repeated-split comparison fixes the 55-target scale and active mask at the original design, then reruns the core-state partition, state standardization, local operators, ridge selection, graph construction, candidate selection, and held-out evaluation under 20 split seeds. Table 3 separates the direct risk score from the theorem-linked graph certificate. Minimum two-fold out-of-fold decoder risk matches the held-out oracle in all 80 model–run–split comparisons. It selects the baseline-value family in 75 comparisons and the counterfactual-fiber family in five. The median RMSE gain relative to the counterfactual-fiber family is 4.29 percent. The four-comparison mean gain is positive in every split and ranges from 2.76 to 7.67 percent. Paired 1,999-draw intervals exclude zero in all 75 comparisons where the direct selector chooses the baseline-value family. The graph certificate matches the held-out menu minimum in 63 of 80 comparisons. Its grid-level residual bound is given by Proposition 6. The 80 comparisons reuse model states under different splits and runs, so their selection frequencies describe stability within this design.

TABLE 3. Training-score selection across KW97 splits, $d = 3$

Training score	Oracle	BV / CF	Median gain	Split means	Positive CIs
Out-of-fold decoder risk	80 / 80	75 / 5	4.29	[2.76, 7.67]	75 / 80
Graph certificate	63 / 80	62 / 18	4.16	[0.61, 7.67]	60 / 80

The 80 comparisons comprise 20 core-state splits, two fixed-model parameterizations, and two derivative-estimation runs. Gains and ranges are percentage reductions in held-out target RMSE relative to the local counterfactual-fiber spectrum. Positive intervals count comparisons whose paired percentile interval excludes zero.

The 20-split comparison distinguishes the two training scores: direct risk reproduces the three-coordinate oracle throughout these comparisons, while the graph certificate provides the theorem-linked upper-bound diagnostic. The policies and targets are generated within the structural model. The exercise evaluates the coordinate-selection mechanism under a fixed model; the policy interpretation of its responses rests on the maintained structural model and separate identifying evidence. The paired percentile intervals condition on the fitted candidates, parameterization, graph rule, and realized split.

6.2. A distributional state in general equilibrium

The quantitative illustration is an incomplete-markets transition economy with one asset and three idiosyncratic income states. Households have CRRA utility with discount factor 0.94 and risk aversion 2. Next-period asset choices lie on 15 points over $[0, 20]$ and receive type-I extreme-value smoothing with scale 0.8. Income levels are $(0.5, 1, 1.5)$ with transition matrix

$$P_y = \begin{pmatrix} 0.90 & 0.09 & 0.01 \\ 0.05 & 0.90 & 0.05 \\ 0.01 & 0.09 & 0.90 \end{pmatrix}. \tag{72}$$

Production is $Y = K^{0.36}$ with depreciation 0.08. Competitive factor prices are

$$r(K) = 0.36K^{-0.64} - 0.08, \qquad w(K) = 0.64K^{0.36}. \tag{73}$$

The stationary equilibrium has $K = 4.4804$, $r = 0.0579$, and $w = 1.0981$. Each transition path is solved by backward household recursion and forward distribution propagation with market clearing for aggregate capital. The household distribution has 45 mass points. The supplied representation records total mass, aggregate assets, and the masses in the first two income states. The third income mass follows from total mass. In Fisher score coordinates the residual distribution space has dimension 41. The declared counterfactual family contains two balanced fiscal packages. A capital-income tax begins at 0.25 and decays geometrically at rate 0.5 with contemporaneous lump-sum rebates. A labor-income tax begins at 0.30 and has the same decay and rebate rule. The target vector contains capital, consumption, output, and the interest rate at horizons 0, 1, 2, 4, and 8 for both packages, giving 40 target components. The supplied map, fiscal packages, horizons, outcome normalization, state metric, and target metric are common across the scattered-state exercise. The response field is evaluated along 46 scattered Fisher-score directions that preserve the supplied representation. Central finite differences produce directional derivatives of the nonlinear general-equilibrium transition paths. A least-squares reconstruction of

TABLE 4. Distributional representation and fiscal-counterfactual loss

Coordinates d	Counterfactual loss share	Moment-dictionary loss share	Cumulative target share
1	0.274	0.526	0.726
2	0.031	0.439	0.969
3	0.012	0.376	0.988
4	0.005	0.338	0.995
5	0.001	0.304	0.999
6	0.0007	0.293	0.9993

Losses are divided by the zero-coordinate target loss. The moment dictionary contains powers of assets, asset-income interactions, and wealth-income cells. Each moment step selects the remaining dictionary element that gives the largest target-loss reduction.

the derivative map has RMSE $1.23e-08$, and the standard deviation of the directional response derivatives is $4.45e-03$. The resulting target spectrum has cumulative shares 0.726, 0.969, and 0.988 after one, two, and three coordinates. The comparison has a measurement interpretation. Under the unrestricted local linear class in Proposition A7, the target eigen-directions minimize residual derivative loss. The moment dictionary restricts added measurements to its listed economic moments. The two columns therefore answer different admissible-class questions. Table 4 reports the resulting class-conditional frontier for these two admissible classes. The distributional state can also be selected for a downstream policy decision. Let $G(s)$ be the 20×2 matrix whose columns are the two package-response paths. The planner chooses package weights $a \in \mathbb{R}^2$ to minimize

$$\frac{1}{2}\|d_0 + G(s)a\|^2 + \frac{\eta}{2}\|a\|^2, \qquad \eta = 2 \times 10^{-7}, \tag{74}$$

where d_0 is fixed at one half of the negative sum of the two baseline package responses. The Hessian calculation in Proposition A3 gives a target metric on $\operatorname{vec} G$. Three representations are estimated on a training set of scattered distributions: the equal-target metric, the decision metric, and the greedy moment dictionary. The response surface is predicted on held-out distributions and the planner resolves (74) using the predicted response. At $d = 2$, the decision-metric representation has policy-weight RMSE 1.08e-04, the equal-target representation has 1.41e-03, and the moment dictionary has 5.13e-03. Their mean regret relative to the mean full-state decision loss is 1.49e-05, 0.003, and 0.142. Figure 1 reports the target-loss and decision comparisons. The distributional derivative calculation is the finite-state counterpart of the Hilbert-space operator. Sequence-space methods compute equilibrium derivatives in heterogeneous-agent models (Auclert et al. 2021), and projection and parametric-distribution methods provide state representations (Reiter 2009; Winberry 2018). The calculation here starts from the supplied aggregate representation and measures the distributional directions required by the declared fiscal

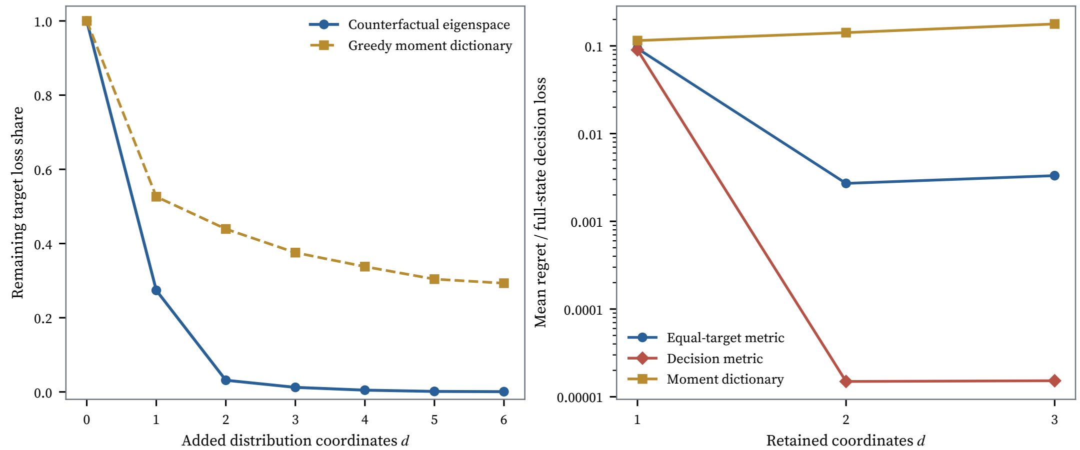

FIGURE 1. Distributional coordinates in the general-equilibrium transition economy

Left: remaining fiscal-counterfactual target loss at each dimension. Right: mean held-out policy regret relative to the mean full-state decision loss for the equal-target metric, the decision metric, and the moment dictionary. The decision-metric coordinates use the local regret metric in Proposition A3.

transition paths. Every response in this exercise is generated by the calibrated model. The reported losses describe internal approximation for the stated policies, metric, perturbation law, and coordinate classes. Empirical use requires an independently recovered intervention-response field evaluated under the same contract.

7. Conclusion

Which coordinates to retain depends on the policy calculation they will support. A supplied record, target loss, state distribution, and admissible augmentation class make that choice a decoding problem. The restricted derivative spectrum gives its local benchmark under the stated shrinking experiment.

At a finite scale, the target values left within a record determine decoder risk. A conditional-generator certificate bounds that risk, and its frequency diagnostic measures the slack relevant for coordinate selection. The oracle comparison carries this approximation term alongside score-estimation, sampling, and optimization error. The oscillatory example shows why each part belongs in the comparison.

The model exercises compare coordinate selection at fixed budgets and show how target weights change useful distributional directions. They evaluate fitted prediction procedures: the representation supplies information, and the decoder uses it. Empirical use additionally requires identification and estimation of the policy-response map.

References

Athey, Susan, and Stefan Wager. 2021. "Policy Learning With Observational Data." *Econometrica* 89(1):133–161.

Attia, Ahmed, Alen Alexanderian, and Arvind K. Saibaba. 2018. "Goal-Oriented Optimal Design of Experiments for Large-Scale Bayesian Linear Inverse Problems." *Inverse Problems* 34(9):095009.

Auclert, Adrien, Bence Bardoczy, Matthew Rognlie, and Ludwig Straub. 2021. "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models." *Econometrica* 89(5):2375–2408.

Bai, Jushan, and Serena Ng. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica* 70(1):191–221.

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin. 2007. "Estimating Dynamic Models of Imperfect Competition." *Econometrica* 75(5):1331–1370.

Bakry, Dominique, Ivan Gentil, and Michel Ledoux. 2014. *Analysis and Geometry of Markov Diffusion Operators*. Cham: Springer.

Beraja, Martin. 2023. "A Semistructural Methodology for Policy Counterfactuals." *Journal of Political Economy* 131(1):190–201.

Bigoni, Daniele, Youssef Marzouk, Clémentine Prieur, and Olivier Zahm. 2022. "Nonlinear Dimension Reduction for Surrogate Modeling Using Gradient Information." *Information and Inference: A Journal of the IMA* 11(4):1597–1639.

Borsuk, Karol. 1933. "Drei Sätze über die n-dimensionale euklidische Sphäre." *Fundamenta Mathematicae* 20:177–190.

Chen, Xiaohong, and Demian Pouzo. 2015. "Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models." *Econometrica* 83(3):1013–1079.

Chen, Xiaohong, Victor Chernozhukov, Sokbae Lee, and Whitney K. Newey. 2014. "Local Identification of Nonparametric and Semiparametric Models." *Econometrica* 82(2):785–809.

Cheng, Bing, and Howell Tong. 2026. "An Approach to Fisher-Rao Metric for Infinite Dimensional Non-Parametric Information Geometry." *Entropy* 28(4):374.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica* 81(6):2205–2268.

Christensen, Timothy, and Benjamin Connault. 2023. "Counterfactual Sensitivity and Robustness." *Econometrica* 91(1):263–298.

Chung, Fan R. K. 1997. *Spectral Graph Theory*. Providence: American Mathematical Society.

Constantine, Paul G., Armin Eftekhari, Jeffrey Hokanson, and Rachel A. Ward. 2017. "A Near-Stationary Subspace for Ridge Approximation." *Computer Methods in Applied Mechanics and Engineering* 326:402–421.

Constantine, Paul G., Eric Dow, and Qiqi Wang. 2014. "Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces." *SIAM Journal on Scientific Computing* 36(4):A1500–A1524.

Elmachtoub, Adam N., and Paul Grigas. 2022. "Smart Predict, then Optimize." *Management Science* 68(1):9–26.

Fang, Zheng, and Andres Santos. 2019. "Inference on Directionally Differentiable Functions." *Review of Economic Studies* 86(1):377–412.

Fujiwara, Akio. 2024. "Hommage to Chentsov's Theorem." *Information Geometry* 7:79–98.

Givan, Robert, Thomas Dean, and Matthew Greig. 2003. "Equivalence Notions and Model Minimization in Markov Decision Processes." *Artificial Intelligence* 147(1–2):163–223.

Gonçalves, Sílvia, Ana María Herrera, Lutz Kilian, and Elena Pesavento. 2024. "State-Dependent Local Projections." *Journal of Econometrics* 244(2):105702.

Grey, Zachary. 2026. "A Riemannian View on Active Subspaces." arXiv:2607.25163.

Guerron-Quintana, Pablo A., Atsushi Inoue, and Lutz Kilian. 2017. "Impulse Response Matching Estimators for DSGE Models." *Journal of Econometrics* 196(1):144–155.

Hahn, Juergen, and Thomas F. Edgar. 2002. "An Improved Method for Nonlinear Model Reduction Using Balancing of Empirical Gramians." *Computers & Chemical Engineering* 26(10):1379–1397.

Hahn, Juergen, Thomas F. Edgar, and Wolfgang Marquardt. 2003. "Controllability and Observability Covariance Matrices for the Analysis and Order Reduction of Stable Nonlinear Systems." *Journal of Process Control* 13(2):115–127.

Haji-Ali, Abdul-Lateef, Fabio Nobile, and Raul Tempone. 2016. "Multi-Index Monte Carlo: When Sparsity Meets Sampling." *Numerische Mathematik* 132(4):767–806.

Hebden, James, and Fabian Winkler. 2026. "Computation of Policy Counterfactuals in Sequence Space." *Journal of Economic Dynamics and Control* 182:105228.

Hong, Han, and Jessie Li. 2018. "The Numerical Delta Method." *Journal of Econometrics* 206(2):379–394.

Kotta, Ülle, Claude H. Moog, and Maris Tõnso. 2018. "Minimal Realizations of Nonlinear Systems." *Automatica* 95:207–212.

Kundu, Poorbita, and Nathan Wycoff. 2026. "Active Subspaces in Infinite Dimension." *Proceedings of Machine Learning Research* 300:2089–2097.

Lall, Sanjay, Jerrold E. Marsden, and Sonja Glavaski. 2002. "A Subspace Approach to Balanced Truncation for Model Reduction of Nonlinear Control Systems." *International Journal of Robust and Nonlinear Control* 12(6):519–535.

Lee, Junhyeong, Sangjin Jin, and Yongjae Lee. 2026. "Decision-Focused Learning via Tangent-Space Projection of Prediction Error." arXiv:2605.01361.

Lewis, A. S. 1996. "Derivatives of Spectral Functions." *Mathematics of Operations Research* 21(3):576–588.

Kalouptsidi, Myrto, Paul T. Scott, and Eduardo Souza-Rodrigues. 2021. "Identification of Counterfactuals in Dynamic Discrete Choice Models." *Quantitative Economics* 12(2):351–403.

Liu, Laura, and Mikkel Plagborg-Moller. 2023. "Full-Information Estimation of Heterogeneous Agent Models Using Macro and Micro Data." *Quantitative Economics* 14(1):1–35.

Liu, Laura, Mikkel Plagborg-Moller, and Nelson Matthew P. Tan. 2026. "Limited-Information Estimation of Heterogeneous Agent Models." arXiv:2608.13953.

Magnac, Thierry, and David Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes." *Econometrica* 70(2):801–816.

Moore, B. C. 1981. "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction." *IEEE Transactions on Automatic Control* 26(1):17–32.

Pistone, Giovanni, and Carlo Sempi. 1995. "An Infinite-Dimensional Geometric Structure on the Space of All the Probability Measures Equivalent to a Given One." *Annals of Statistics* 23(5):1543–1561.

Reiter, Michael. 2009. "Solving Heterogeneous-Agent Models by Projection and Perturbation." *Journal of Economic Dynamics and Control* 33(3):649–665.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55(5):999–1033.

Singer, Amit, and Hau-Tieng Wu. 2012. "Vector Diffusion Maps and the Connection Laplacian." *Communications on Pure and Applied Mathematics* 65(8):1067–1144.

Sussmann, Hector J. 1976. "Existence and Uniqueness of Minimal Realizations of Nonlinear Systems." *Mathematical Systems Theory* 10(1):263–284.

Verdière, Romain, Clémentine Prieur, and Olivier Zahm. 2025. "Diffeomorphism-Based Feature Learning Using Poincaré Inequalities on Augmented Input Space." *Journal of Machine Learning Research* 26(139):1–31.

Winberry, Thomas. 2018. "A Method for Solving and Estimating Heterogeneous Agent Macro Models." *Quantitative Economics* 9(3):1123–1151.

Zahm, Olivier, Paul G. Constantine, Clementine Prieur, and Youssef M. Marzouk. 2020. "Gradient-Based Dimension Reduction of Multivariate Vector-Valued Functions." *SIAM Journal on Scientific Computing* 42(1):A534–A558.