When Does Auxiliary Calibration Improve Target Recovery? Nuisance Geometry, Sample Complexity, and Timing

Anonymous authors
Paper under double-blind review

Abstract

Many learning problems combine target-bearing observations with auxiliary measurements that do not observe the target directly, but instead reveal a shared nuisance variable. When, and by how much, can such auxiliary measurements improve target recovery? We study this question in a Gaussian inverse experiment with an acquisition channel that mixes target and nuisance effects and a calibration channel that observes only the nuisance. We show that calibration value is determined not by marginal channel quality, but by the paired geometry relating the nuisance scores in the two channels. The closed relation pairing the two nuisance scores yields an exact resolvent formula for efficient target information, including nonclosed score ranges, and separates calibration-irrecoverable structural loss from finite-budget calibration loss. For unrestricted nuisance, we prove an exact minimax-risk reduction to the efficient Gaussian shift over compact convex target classes; for restricted nuisance classes, we give a nuisance-translation condition under which the reduction remains valid. Under a Hilbert-scale comparison, we derive matching two-resource minimax rates and the critical auxiliary sample size $m_{\mathrm{crit}}(n)$ needed to recover the nuisance-free target rate. In parabolic evolution models, calibration timing can change the effective ill-posedness and turn a polynomial auxiliary requirement into an exponential one, while nuisance constraints can change the minimax rate itself. The same geometry also yields target-oriented calibration allocation, which can differ from nuisance-optimal allocation. Experiments support these predictions and illustrate the additional errors-in-variables difficulty of learning the nuisance response.

1 Introduction

Auxiliary measurements are widely used in learning problems to reveal latent or nuisance structure that is not directly observed in target-bearing data. We study a particularly stark setting in which the auxiliary measurements never observe the target itself. Instead, they measure a shared nuisance variable that also affects the target-bearing observations. The central question is therefore not how accurately the auxiliary channel estimates the nuisance in isolation, but whether it resolves the nuisance directions that matter for recovering the target.

We formalize this setting through a Gaussian inverse experiment with two channels:

$$Y = \mathsf{A}x + \mathsf{B}\theta + n^{-1/2}\varepsilon, \qquad Z = \mathsf{C}\theta + m^{-1/2}\xi, \tag{1.1}$$

where x is the target, θ is a shared nuisance parameter, $\mathsf{A}, \mathsf{B}, \mathsf{C}$ are known bounded linear operators, and ε, ξ are independent standard Gaussian white noises. The acquisition channel Y contains both target and nuisance information; the calibration channel Z contains only nuisance information. Hereafter, *calibration* means auxiliary measurement of the shared nuisance, not predictive-probability calibration. The information levels n and m play the roles of sample sizes for the target-bearing and calibration observations, respectively. This leads to a two-resource sample-complexity problem: how does the minimax risk for x depend jointly on n, m, the coupling between the two nuisance channels, and the admissible nuisance class?

Related questions arise in transfer learning, semi-supervised learning, unknown-operator inverse problems, self-calibration, and auxiliary-sample deconvolution (Hanneke & Kpotufe, 2019; Kpotufe

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

& Martinet, 2021; Cavalier & Hengartner, 2005; Johannes & Schwarz, 2013; Ling & Strohmer, 2018; Johannes, 2009). These frameworks do not in general characterize how the nuisance directions observed in calibration align with those contaminating the target-bearing channel, which is the central issue here.

Marginal calibration accuracy is insufficient; target recovery depends on the *pairing* between nuisance effects in acquisition and calibration. Two models can have identical marginal Gram operators for their nuisance channels and yet leave completely different information about the target. In infinite dimensions this pairing is especially important because calibration may control the nuisance only in a weaker topology, the calibration range may be nonclosed, and limiting nuisance directions may be indistinguishable from target directions. We encode this pairing by a closed linear relation between the two nuisance scores. The resulting geometry determines efficient target information, separates structural from finite-budget loss, and governs minimax recovery, auxiliary sample complexity, timing, and calibration design.

Contributions. (i) We identify the paired nuisance-score geometry governing target recovery. Its exact resolvent remains valid for nonclosed calibration ranges and separates structural from finite-budget loss, while identical marginal nuisance-channel Grams can yield different target information (Theorem 2.1, Proposition 2.2). We then prove an exact minimax-risk reduction for unrestricted nuisance and give a translation condition extending it to restricted classes (Theorems 3.1 and D.9).

(ii) Under a two-sided Hilbert-scale link condition, we derive matching two-resource minimax rates without a common eigenbasis and identify the auxiliary sample complexity $m_{\mathrm{crit}}(n)$ (Theorems 3.3 and 3.4). In a parabolic initial-state model, calibration timing changes the effective ill-posedness, while nuisance constraints can change the late-calibration rate from logarithmic to polynomial; a growing-radius threshold connects the two regimes (Theorems 4.1 and 4.2, Proposition 4.3).

(iii) Efficient information induces a target-oriented design criterion and a closed-form allocation for independently exposed coordinates. The experiments separate this allocation gain from the cost of learning the nuisance response, including errors-in-variables attenuation under passive calibration data (Sections 5 and 6).

Related work. Our analysis connects semiparametric nuisance elimination (Begun et al., 1983; Trabs, 2015) with shorted operators and closed linear relations (Anderson & Trapp, 1975; Mitra & Puntanen, 1990; Hassi et al., 2007), and uses Hilbert-scale methods for the rate analysis (Mair & Ruymgaart, 1996; Mathé, 2019). Unlike operator-learning analyses based on a known diagonalizing basis (de Hoop et al., 2023), our comparison allows noncommuting transfer operators. The two-resource rates are related to auxiliary-sample deconvolution (Neumann, 1997; Johannes, 2009; Comte & Lacour, 2011), while the allocation problem connects to goal-oriented and multifidelity design (Attia et al., 2018; Peherstorfer et al., 2018). Our distinction is that the same paired nuisance geometry governs structural information loss, minimax recovery, auxiliary sample complexity, timing, and target-oriented allocation.

2 Paired nuisance geometry and efficient information

All Hilbert spaces are real and separable. We use the standard conventions for linear relations, quadratic-form order, generalized inverses, and spectral multipliers summarized in Appendix B.1.

2.1 Finite-dimensional formulation and infinite-dimensional obstructions

Take $\mathcal{X}, \mathbf{\Theta}, \mathcal{Y}, \mathcal{Z}$ finite-dimensional for a moment, so that $\mathsf{A}, \mathsf{B}, \mathsf{C}$ are matrices. After the harmless reparameterization $\theta \mapsto -\theta$, the joint experiment (1.1) has Fisher information

$$J_{n,m} = \begin{pmatrix} n\mathsf{A}^*\mathsf{A} & -n\mathsf{A}^*\mathsf{B} \\ -n\mathsf{B}^*\mathsf{A} & n\mathsf{B}^*\mathsf{B} + m\mathsf{C}^*\mathsf{C} \end{pmatrix} \tag{2.1}$$

on $\mathcal{X} \oplus \mathbf{\Theta}$, and the information about x after eliminating θ is its Schur complement onto the target block,

$$\begin{aligned} \mathsf{I}_{\mathrm{eff},n,m} &= n\mathsf{A}^*\mathsf{A} - n^2\mathsf{A}^*\mathsf{B}\big(n\mathsf{B}^*\mathsf{B} + m\mathsf{C}^*\mathsf{C}\big)^{-1}\mathsf{B}^*\mathsf{A}, \\ \langle \mathsf{I}_{\mathrm{eff},n,m} v, v\rangle &= \inf_{\theta\in\mathbf{\Theta}} \big[n\|\mathsf{A}v + \mathsf{B}\theta\|^2 + m\|\mathsf{C}\theta\|^2\big]. \end{aligned} \tag{2.2}$$

This is the semiparametric efficient information of the model (Begun et al., 1983; Trabs, 2015). Two features of calibration break it in infinite dimensions: $\mathsf{C}^*\mathsf{C}$ need not be boundedly invertible, so the inverse need not exist and the infimum need not be attained when calibration controls the nuisance only in a weaker topology than acquisition feels it; and some acquisition nuisance directions are limits of $\mathsf{B}\theta_j$ along sequences with $\mathsf{C}\theta_j \to 0$, which no budget separates from the target score and no formula in $\mathsf{C}^*\mathsf{C}$ alone sees. This motivates the paired nuisance-score relation introduced next.

2.2 Paired nuisance-score relation

Let $\mathbf{\Theta}$ be the nuisance parameter space and $\mathcal{Y}, \mathcal{Z}$ the acquisition and calibration score Hilbert spaces; $\mathcal{L}(H_1, H_2)$ denotes the bounded linear operators between Hilbert spaces. For $\mathsf{B} \in \mathcal{L}(\mathbf{\Theta}, \mathcal{Y})$ and $\mathsf{C} \in \mathcal{L}(\mathbf{\Theta}, \mathcal{Z})$ define the *paired nuisance-score relation*

$$\mathcal{G}_{BC} = \overline{\{(\mathsf{C}\theta, \mathsf{B}\theta) : \theta \in \mathbf{\Theta}\}}^{\mathcal{Z}\oplus\mathcal{Y}}, \tag{2.3}$$

a closed subspace of $\mathcal{Z} \oplus \mathcal{Y}$ recording which pairs of calibration and acquisition nuisance effects occur together. With $\operatorname{dom}\mathcal{G}_{BC} = \{z : \exists y,\ (z, y) \in \mathcal{G}_{BC}\}$ and $\operatorname{mul}\mathcal{G}_{BC} = \{y : (0, y) \in \mathcal{G}_{BC}\}$, set

$$\mathcal{U}_0 = \operatorname{mul}\mathcal{G}_{BC}, \quad \mathcal{Y}_0 = \mathcal{U}_0^{\perp}, \quad \mathcal{Z}_0 = \overline{\operatorname{dom}\mathcal{G}_{BC}}^{\mathcal{Z}}, \quad \mathsf{A}_0 = P_0\mathsf{A},$$

with P_0 the orthogonal projection of $\mathcal{Y}$ onto $\mathcal{Y}_0$ and $\mathsf{A} \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ the target score. Lemma D.2 shows $\mathcal{U}_0$ closed and $\mathcal{G}_{BC}$ orthogonally split as

$$\mathcal{G}_{BC} = \{(z, \mathsf{R}z + u) : z \in \operatorname{dom}\mathsf{R},\ u \in \mathcal{U}_0\}, \tag{2.4}$$

where $\mathsf{R} : \operatorname{dom}\mathsf{R} = \operatorname{dom}\mathcal{G}_{BC} \subseteq \mathcal{Z}_0 \to \mathcal{Y}_0$ is a closed densely defined operator. These two components play distinct roles in target recovery: $\mathcal{U}_0$ collects the directions no budget distinguishes from the target score, the *structural loss* that P_0 removes, and R is the recoverable transfer, unbounded exactly when calibration acts in the weaker topology. The closure in (2.3) is what admits limiting directions that no single nuisance tangent generates.

For information levels $n, m > 0$ define the target and nuisance score operators

$$\mathsf{S}_x v = (\sqrt{n}\mathsf{A}v, 0), \qquad \mathsf{S}_h\theta = (\sqrt{n}\mathsf{B}\theta, \sqrt{m}\mathsf{C}\theta), \tag{2.5}$$

let $\mathcal{T}_h = \overline{\operatorname{ran}\mathsf{S}_h}$ be the closed nuisance-score subspace, and let $\mathsf{S}_{\mathrm{eff}} = (I - \Pi_{\mathcal{T}_h})\mathsf{S}_x$ and $\mathsf{I}_{\mathrm{eff},n,m} = \mathsf{S}_{\mathrm{eff}}^*\mathsf{S}_{\mathrm{eff}}$ be the efficient score and efficient information. The relation $\mathcal{G}_{BC}$ determines $\mathcal{T}_h$ for every $n, m > 0$ and does not itself depend on the budgets. Appendices D and A establish its reparameterization invariance and derive it for evolving and locally nonlinear models.

2.3 Efficient projection and exact resolvent

Under the reparameterization $\theta \mapsto -\theta$ the joint score operator on $\mathcal{X} \oplus \mathbf{\Theta}$ is $\mathsf{S}_{\mathrm{joint}}(v, \theta) = \mathsf{S}_x v - \mathsf{S}_h\theta$, with Fisher block $\mathsf{J}_{n,m}$ of (2.1). Without any invertibility assumption, Proposition D.3 in Appendix D identifies $\mathsf{I}_{\mathrm{eff},n,m}$ with the *short* of $\mathsf{J}_{n,m}$ to the target subspace, so the variational form in (2.2) survives verbatim. The paired relation splits it into a structural-loss component and a finite-calibration component.

Theorem 2.1 (Efficient information from the paired relation)**.** *For every $n, m > 0$,*

$$\mathsf{I}_{\mathrm{eff},n,m} = n\mathsf{A}_0^*\left[I_{\mathcal{Y}_0} + (n/m)\mathsf{R}\mathsf{R}^*\right]^{-1}\mathsf{A}_0 \tag{2.6}$$

and, equivalently,

$$\langle \mathsf{I}_{\mathrm{eff},n,m}v, v\rangle = \inf_{z\in\operatorname{dom}\mathsf{R}}\left[n\|\mathsf{A}_0 v - \mathsf{R}z\|_{\mathcal{Y}_0}^2 + m\|z\|_{\mathcal{Z}_0}^2\right]. \tag{2.7}$$

The operator in brackets in (2.6) *is the bounded spectral resolvent of the possibly unbounded nonnegative self-adjoint operator $\mathsf{R}\mathsf{R}^*$. No closed-range assumption on C is required.*

The formula makes the two losses explicit. The kernel is $\ker\mathsf{I}_{\mathrm{eff},n,m} = \ker\mathsf{A}_0$, so $\mathcal{U}_0$ determines the structural loss even under arbitrarily informative calibration. As $m/n \to \infty$, $n^{-1}\mathsf{I}_{\mathrm{eff},n,m}$ tends to $\mathsf{A}_0^*\mathsf{A}_0$; as $m/n \to 0$, it tends to $\mathsf{A}_0^*\Pi_{\ker\mathsf{R}^*}\mathsf{A}_0$, retaining only reduced target scores orthogonal to $\overline{\operatorname{ran}\mathsf{R}}$. Information is independent of the calibration budget exactly when $\operatorname{ran}\mathsf{A}_0 \perp \operatorname{ran}\mathsf{R}$. Appendices D and E collect the linear-functional, no-calibration, and finite-resolution consequences.

2.4 MARGINAL CHANNEL GEOMETRY IS INSUFFICIENT

Proposition 2.2 (Marginal Gram operators do not determine efficient information)**.** *Let* $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$, $\mathbf{\Theta} = \mathbb{R}^2$, *and* $\mathsf{A} = I_{\mathbb{R}}$. *For* $\theta = (\theta_1, \theta_2)$, *define*

$$\mathsf{B}_{\parallel}\theta = \theta_1,\ \mathsf{C}_{\parallel}\theta = \theta_1,\ \mathsf{B}_{\perp}\theta = \theta_1,\ \mathsf{C}_{\perp}\theta = \theta_2. \tag{2.8}$$

Both models satisfy $\mathsf{B}\mathsf{B}^* = \mathsf{C}\mathsf{C}^* = I_{\mathbb{R}}$, *but* $\mathsf{I}^{\parallel}_{\mathrm{eff},n,m} = \frac{nm}{n+m} I_{\mathbb{R}}, \mathsf{I}^{\perp}_{\mathrm{eff},n,m} = 0$. *No universal functional of* $(\mathsf{B}\mathsf{B}^*, \mathsf{C}\mathsf{C}^*, n, m)$ *can recover the paired efficient information for all models of the form* (2.5).

This rules out a broad class of marginal reductions. In the parallel model calibration observes the nuisance direction that enters acquisition; in the transverse model it observes an orthogonal coordinate and leaves that nuisance unrestricted. Any summary discarding the pairing assigns the same input to two statistically different experiments.

3 MINIMAX REDUCTION AND AUXILIARY SAMPLE COMPLEXITY

The efficient-information characterization of Section 2 describes how auxiliary calibration changes the information available for the target. We now turn this characterization into a global minimax theory. We first show that, for unrestricted nuisance, the original joint experiment is exactly equivalent in minimax risk to the efficient Gaussian shift. We then derive two-resource minimax rates under a relative Hilbert-scale condition, and finally identify the critical auxiliary sample size $m_{\mathrm{crit}}(n)$ required to recover the nuisance-free target rate.

3.1 EXACT MINIMAX REDUCTION

The joint Gaussian experiment has mean $\mathsf{S}_x x + \mathsf{S}_h \theta$ in $\mathcal{Y} \oplus \mathcal{Z}$ and identity noise covariance in the isonormal sense. Its projection onto $\mathcal{T}_h^{\perp}$ is the efficient Gaussian shift

$$\mathbb{X}_{\mathrm{eff}}(g) = \langle \mathsf{S}_{\mathrm{eff}} x, g \rangle + W_{\mathrm{eff}}(g),\ g \in \mathcal{T}_h^{\perp}. \tag{3.1}$$

For a nonempty compact convex set $K \subseteq \mathcal{X}$, define $\mathcal{R}^{\mathrm{joint}}_{n,m}(K) = \inf_{\widehat{x}} \sup_{x \in K,\, \theta \in \mathbf{\Theta}} \mathbb{E}_{x,\theta} \|\widehat{x} - x\|^2$, *and* $\mathcal{R}^{\mathrm{eff}}_{n,m}(K) = \inf_{\widehat{x}} \sup_{x \in K} \mathbb{E}^{\mathrm{eff}}_x \|\widehat{x} - x\|^2$. The infima range over Borel estimators. Metric projection restricts their values to K without increasing squared loss.

Theorem 3.1 (Exact minimax reduction for unrestricted nuisance)**.** *For every nonempty compact convex* $K \subseteq \mathcal{X}$, $\mathcal{R}^{\mathrm{joint}}_{n,m}(K) = \mathcal{R}^{\mathrm{eff}}_{n,m}(K)$. *The equality remains valid when the nuisance score range* $\operatorname{ran} \mathsf{S}_h$ *is nonclosed.*

The unrestricted nuisance class makes the reduction exact without further conditions. For a restricted class $\mathbf{\Theta}_0$, equality still holds when $-\Pi_{\mathcal{T}_h} \mathsf{S}_x x \in \overline{\mathsf{S}_h(\mathbf{\Theta}_0)}$ for every $x \in K$ (Theorem D.9); if this translation condition fails, the constraint itself can carry target information discarded by projection. Appendices D.4 and H give the counterexample, radius conditions, and approximate-reduction bounds.

3.2 HILBERT-SCALE RATES

For the rate analysis that follows, restrict the target space to $(\ker \mathsf{A}_0)^{\perp}$ and retain the symbol $\mathcal{X}$ for this identifiable subspace. Let L be positive self-adjoint with compact inverse and eigensystem (e_k, λ_k) satisfying $\lambda_k \asymp k$. This specialization follows the Hilbert-scale formulation of statistical inverse problems (Mair & Ruymgaart, 1996; Cavalier, 2008).

Assumption 3.2 (Relative Hilbert-scale comparison)**.** *The reduced target score has full support in* $\mathcal{Y}_0$. *There are a unitary* $U : \mathcal{X} \to \mathcal{Y}_0$, *positive sequences* (a_k) *and* (g_k), *constants* $0 < c_R \le C_R < \infty$, *and* $\alpha \ge 0$ *such that* $\mathsf{A}_0 = U a(L), \operatorname{dom} \mathsf{R}^* = U \operatorname{dom} g(L), a_k \asymp \lambda_k^{-\alpha}$, *and, for every* $y \in \operatorname{dom} \mathsf{R}^*$,

$$c_R \|g(L) U^* y\|^2 \le \|\mathsf{R}^* y\|^2 \le C_R \|g(L) U^* y\|^2. \tag{3.2}$$

The two-sided bound (3.2) is a link condition in the sense of Hilbert-scale regularization (Natterer, 1984; Mathé & Pereverzev, 2003): the relative calibration operator must be of the same order as a spectral multiplier of L on a common form domain. Appendix G gives noncommuting examples satisfying this condition. Importantly, it does not require $\mathsf{R}\mathsf{R}^*$ and $\mathsf{A}_0\mathsf{A}_0^*$ to share an eigenbasis. Appendix F gives the rate comparison on a proper target-score support.

Theorem 3.3 (Relative form comparison)**.** *Under Assumption 3.2, uniformly in $n, m > 0$,*

$$\mathsf{I}_{\mathrm{eff},n,m} \asymp n\, a(L) \left[I + (n/m) g(L)^2\right]^{-1} a(L) \tag{3.3}$$

in two-sided quadratic-form order. For finitely supported $v = \sum_k v_k e_k$,

$$\langle \mathsf{I}_{\mathrm{eff},n,m} v, v\rangle \asymp \sum_{k\geq 1} \iota_k(n,m) v_k^2,\ \iota_k(n,m) = \frac{n a_k^2}{1 + (n/m) g_k^2}. \tag{3.4}$$

For $\beta, R > 0$, define

$$\mathcal{X}_\beta(R) = \left\{ x = \sum_{k\geq 1} x_k e_k : \sum_{k\geq 1} \lambda_k^{2\beta} x_k^2 \leq R^2 \right\}. \tag{3.5}$$

Since $\mathcal{X}_\beta(R)$ is compact and convex, write $\mathcal{R}_{n,m}(\beta, R)$ for the common risk in Theorem 3.1. Set $r_n = n^{-\frac{2\beta}{2\beta+2\alpha+1}}$.

Theorem 3.4 (Global minimax phase diagram)**.** *Under Assumption 3.2, as $n, m \to \infty$ along arbitrary joint sequences:*

(i) If $g_k \asymp k^\kappa$ and $p = 2\alpha + 2\kappa$, then

$$\mathcal{R}_{n,m}(\beta, R) \asymp r_n \vee \begin{cases} m^{-\frac{2\beta}{2\beta+p+1}}, & p > -1, \\ m^{-1}\log(e+m), & p = -1, \\ m^{-1}, & p < -1. \end{cases} \tag{3.6}$$

(ii) If $g_k \asymp k^\kappa e^{\Delta k^\nu}$ with $\Delta, \nu > 0$, then

$$\mathcal{R}_{n,m}(\beta, R) \asymp r_n \vee [\log(e+m)]^{-2\beta/\nu}. \tag{3.7}$$

(iii) If $g_k \asymp k^\kappa e^{-\delta k^\nu}$ with $\delta, \nu > 0$, then $\mathcal{R}_{n,m}(\beta, R) \asymp r_n \vee m^{-1}$.

The comparison constants depend only on the fixed structural parameters.

All three regimes follow from $\iota_k(n,m)^{-1} = (n a_k^2)^{-1} + g_k^2 (m a_k^2)^{-1}$. The upper bound uses unbiased efficient coordinates and a spectral cutoff in the target basis, even when $\mathsf{I}_{\mathrm{eff},n,m}$ is off diagonal; the lower bound uses only the upper quadratic-form comparison of Theorem 3.3, so the Assouad and Le Cam constructions need no common eigenbasis. The three branches have the same shape as the rates known for deconvolution with an auxiliary sample; Appendix F compares the two exponent structures.

Appendix G gives an exact diagonal product reduction that separates experiment equivalence from the rate comparison.

3.3 Auxiliary sample complexity

We next quantify the required auxiliary budget. Let $J_n \asymp n^{1/(2\beta+2\alpha+1)}$ be the acquisition oracle cutoff, the resolution supported by the acquisition channel when the nuisance is known. For $J \in \mathbb{N}$ and $m > 0$, define the accumulated calibration variance and the *auxiliary sample complexity*, the smallest auxiliary budget at which the target rate returns to its nuisance-free value r_n, by

$$\begin{aligned} V_{\mathrm{cal}}(J, m) &= \frac{1}{m} \sum_{k\leq J} \frac{g_k^2}{a_k^2}, \\ m_{\mathrm{crit}}(n) &= \inf\left\{ m \geq 1 : V_{\mathrm{cal}}(J_n, m) \leq c_* J_n^{-2\beta} \right\}. \end{aligned} \tag{3.8}$$

for a fixed constant $c_* > 0$; changing c_* changes only the comparison constants. Below $m_{\mathrm{crit}}(n)$ the auxiliary channel is the binding constraint and the rate is the calibration branch of (3.6); above it the

target is estimated at the rate it would have if the nuisance were known. In the algebraic regime,

$$m_{\mathrm{crit}}(n) \asymp \begin{cases} n^{\frac{2\beta+p+1}{2\beta+2\alpha+1}}, & p > -1, \\ \text{solution of } m/\log m \asymp n^{\frac{2\beta}{2\beta+2\alpha+1}}, & p = -1, \\ n^{\frac{2\beta}{2\beta+2\alpha+1}}, & p < -1. \end{cases}$$

Thus $m_{\mathrm{crit}}(n)$ directly quantifies the auxiliary sample size required to make nuisance uncertainty asymptotically nonbinding at the target resolution. The $p < -1$ branch is parametric because the transferred calibration variance is summable; Appendix F records the three cases. Section 4 shows that the polynomial auxiliary requirement above can become exponential when calibration is taken too late.

Remark F.4 in Appendix F balances the two branches when $n + m = N$ is fixed: the calibration budget of the balancing allocation grows with N while its share tends to zero whenever $p < 2\alpha$.

4 Calibration timing and nuisance constraints

Section 3 quantified how much auxiliary information is needed and showed that the efficient Gaussian shift exactly determines minimax risk when the nuisance is unrestricted. We now ask how that conclusion changes when auxiliary calibration is collected at different times and when the admissible nuisance class is restricted. Timing changes the ill-posedness of the relative transfer operator and hence the auxiliary sample complexity; nuisance constraints can remove the translations behind the exact reduction and change the minimax rate itself.

4.1 Timing under unrestricted nuisance

Consider the initial-state equation $\dot{w}_t + L_h^\nu w_t = 0, w_0 = b$, with $L_h e_k = k e_k$, and let

$$\mathsf{B}_{t_a} = L_h^{-s_Y} e^{-t_a L_h^\nu}, \ \mathsf{C}_\tau = L_h^{-s_Z} e^{-\tau L_h^\nu}. \tag{4.1}$$

Assume $t_a > 0$ or $s_Y \geq 0$, and $\tau > 0$ or $s_Z \geq 0$. With $\kappa = s_Z - s_Y$, the paired relation is the graph of

$$\mathsf{R}_\tau = L_h^\kappa e^{(\tau - t_a) L_h^\nu} \tag{4.2}$$

on its maximal spectral domain.

Theorem 4.1 (Calibration-timing phase transition)**.** *Let $\mathsf{A} = a(L_h)$ with $a_k \asymp k^{-\alpha}$, so that Assumption 3.2 holds with $U = I$. Then*

$$\mathcal{R}_{n,m}(\beta, R) \asymp \begin{cases} r_n \vee m^{-1}, & \tau < t_a, \\ r_n \vee r_m^{\mathrm{alg}}, & \tau = t_a, \\ r_n \vee [\log(e+m)]^{-2\beta/\nu}, & \tau > t_a, \end{cases} \tag{4.3}$$

where r_m^{alg} is the calibration term in (3.6)*. For each fixed τ, the constants may depend on $|\tau - t_a|$ and are uniform on compact timing intervals that stay strictly on one side of t_a. For fixed n, m, the efficient information is nonincreasing in τ. If $\tau > t_a$, then, for the cutoff sequence J_n fixed in* (3.8)*, the calibration budget preserving the oracle cutoff satisfies*

$$\log m_{\mathrm{crit}}(n) = 2(\tau - t_a) J_n^\nu + O(\log J_n). \tag{4.4}$$

Thus late calibration can turn a polynomial auxiliary requirement into an exponential one: R_τ is smoothing, algebraic, or severely ill-posed according as $\tau < t_a$, $\tau = t_a$, or $\tau > t_a$. Appendix G treats near-critical timing, multiple calibration channels, and a noncommuting variable-coefficient extension.

4.2 Nuisance constraints can change the rate

The preceding transition assumes unrestricted nuisance translations can absorb the target score. A bounded nuisance class may exclude them, so efficient projection need not remain equivalent in minimax risk and the constraint itself can provide target information. The next result analyzes this constrained experiment directly.

Theorem 4.2 (Timing with a bounded initial state)**.** *Let $\alpha \geq 0$ and $\beta, R, \rho, t_a, \tau > 0$ be fixed. Observe, independently across channels and coordinates,*

$$Y_k = k^{-\alpha} x_k + e^{-t_a k}\eta_k + n^{-1/2}\varepsilon_k, \qquad Z_k = e^{-\tau k}\eta_k + m^{-1/2}\xi_k, \tag{4.5}$$

with standard Gaussian noises, $\sum k^{2\beta} x_k^2 \leq R^2$, and $\|\eta\|_{\ell^2} \leq \rho$. The joint minimax squared ℓ^2 risk satisfies, as $n, m \to \infty$,

$$\mathcal{R}^{\mathrm{bd}}_{n,m}(\beta, R; \rho) \asymp r_n \vee \begin{cases} m^{-1}, & \tau < t_a, \\ m^{-1}(\log m)^{2\alpha+1}, & \tau = t_a, \\ m^{-t_a/\tau}(\log m)^{2\alpha}, & \tau > t_a, \end{cases} \tag{4.6}$$

where $r_n = n^{-2\beta/(2\beta+2\alpha+1)}$. Constants may depend on the fixed parameters, including ρ and the times, but not on the budgets.

By removing hard nuisance alternatives, the constraint changes the rate rather than only its constant. In this $\nu = 1$ equal-sensor-order model, it replaces the unrestricted late-calibration rate $(\log m)^{-2\beta}$ by the polynomial rate $m^{-t_a/\tau}(\log m)^{2\alpha}$ and makes the on-time branch parametric up to a logarithmic factor; the acquisition branch r_n is unchanged.

The proof uses a second cutoff: calibration correction is applied only where its variance is worth paying, while the nuisance-energy bound controls the remaining tail; Appendix H gives the estimator, a finite-budget form, and matching bounds.

Auxiliary sample complexity under the constraint. Let $q = 2\beta/(2\beta + 2\alpha + 1)$. Balancing the calibration branch in (4.6) with $r_n = n^{-q}$ gives

$$m_{\mathrm{bd}}(n) \asymp \begin{cases} n^q, & \tau < t_a, \\ n^q(\log n)^{2\alpha+1}, & \tau = t_a, \\ n^{q\tau/t_a}(\log n)^{2\alpha\tau/t_a}, & \tau > t_a. \end{cases} \tag{4.7}$$

At or above this scale, calibration is no longer rate-limiting and the joint risk has order r_n. In particular, late calibration requires only a polynomial auxiliary budget under a fixed energy bound, whereas the unrestricted nuisance class requires the exponential scale in (4.4).

Where the bounded and unrestricted regimes meet. A fixed radius and an unrestricted nuisance class are endpoints of a continuum. Let $\rho = \rho_m = m^\gamma$ for late calibration $\tau > t_a$, and define

$$\gamma_\star = \frac{t_a}{2(\tau - t_a)}, \qquad q(\gamma) = \frac{t_a - 2\gamma(\tau - t_a)}{\tau}. \tag{4.8}$$

Proposition 4.3 (Where the bounded and unrestricted rates part)**.** *In the model* (4.5) *with $\tau > t_a$ and $\rho_m = m^\gamma$ for a fixed $\gamma \geq 0$, as $n, m \to \infty$,*

$$\mathcal{R}^{\mathrm{bd}}_{n,m}(\beta, R; \rho_m) \asymp r_n \vee \begin{cases} m^{-q(\gamma)}(\log m)^{2\alpha}, & \gamma < \gamma_\star, \\ (\log m)^{-2\beta}, & \gamma > \gamma_\star. \end{cases} \tag{4.9}$$

For $\gamma < \gamma_\star$ the rate is attained by the rule (H.1) *with $K = J \wedge \left\lfloor \frac{\log(m\rho_m^2)}{2\tau} \right\rfloor$.*

Below $\gamma_\star$ the problem remains in the constrained regime; above it the hard translations return and recover the unrestricted logarithmic rate. Their least-favourable nuisance norm is $m^{\gamma_\star + o(1)}$, so the threshold identifies when the hard translations responsible for the unrestricted rate re-enter the admissible nuisance class; Appendix H gives a complementary approximate-reduction bound.

The value of auxiliary calibration is therefore determined not only by its sample size, but also by when it is collected and by which nuisance alternatives remain admissible. We next ask how a fixed calibration budget should be allocated when the experimenter can choose where auxiliary information is collected.

5 TARGET-ORIENTED CALIBRATION DESIGN

Section 2 identified the efficient target information induced by a calibration channel. We now treat that channel as a design variable and ask: given a fixed auxiliary-calibration budget, where should it be spent to improve target recovery? Section 4 changed *when* calibration is collected; here we change *where* information is supplied along the paired nuisance geometry. The objective is not to estimate the nuisance as accurately as possible in isolation, but to preserve the information most relevant for the target.

This distinction gives two design objectives. A *nuisance-optimal* design allocates effort to minimize uncertainty in the nuisance itself. A *target-oriented* design instead minimizes target uncertainty after nuisance elimination. The latter follows directly from efficient information: an allocation $\boldsymbol{m} = (m_j)_j$ changes the calibration information, which changes I_{eff}, and a target-risk functional of I_{eff} determines the design.

In finite dimensions, let $m_j \geq 0$ denote the information assigned to channel C_j, let $H = n\mathsf{B}^*\mathsf{B} + \sum_j m_j \mathsf{C}_j^*\mathsf{C}_j \succ 0$, and write $I(\boldsymbol{m}) = \mathsf{I}_{\text{eff}}(\boldsymbol{m})$ for the efficient target information. The Schur formula and its directional sensitivity are $I(\boldsymbol{m}) = n\mathsf{A}^*\mathsf{A} - n^2\mathsf{A}^*\mathsf{B}H^{-1}\mathsf{B}^*\mathsf{A}, \partial_{m_j}\langle I(\boldsymbol{m})v, v\rangle = \|\mathsf{C}_j\theta_v^*\|^2,\ and\ \theta_v^* = -nH^{-1}\mathsf{B}^*\mathsf{A}v$. Thus the marginal value of channel j is how strongly it observes the nuisance direction θ_v^* that masks the target direction v. When $I \succ 0$, criteria such as $\mathrm{tr}(WI^{-1})$ for a positive target weight W, or $\psi^* I^{-1}\psi$ for a scalar target functional, translate efficient information directly into target risk. Appendix I derives their sensitivities and weighted variants.

The independently exposed-coordinate model gives a closed form. The result is an oracle design statement: A, B, and the calibration gains d_j are known, coordinate exposures are independent, and each information unit has the same cost.

Proposition 5.1 (Target-oriented optimal allocation)**.** *Suppose A is invertible, B is known, $Z_j = d_j\eta_j + m_j^{-1/2}\xi_j$ with $d_j > 0$, and $\sum_j m_j = M$. If $\|A^{-1}Be_j\| > 0$ for every j, then the total target variance and its minimizing allocation are*

$$F_B(\boldsymbol{m}) = \frac{1}{n}\|A^{-1}\|_{\mathrm{HS}}^2 + \sum_j \frac{\|A^{-1}Be_j\|^2}{d_j^2 m_j}, \qquad m_j^{\mathrm{target}} = M\frac{\|A^{-1}Be_j\|/d_j}{\sum_\ell \|A^{-1}Be_\ell\|/d_\ell}. \tag{5.1}$$

A coordinate receives more calibration effort not merely when the nuisance is difficult to estimate there, but when uncertainty in that nuisance direction couples strongly into target-relevant directions. Calibration effort therefore follows target sensitivity to nuisance uncertainty rather than nuisance uncertainty alone.

By contrast, minimizing the calibration-only nuisance variance $\sum_j (d_j^2 m_j)^{-1}$ gives $m_j^{\mathrm{nuisance}} \propto d_j^{-1}$. On the active coordinates, the two allocations coincide only in the special geometry where $\|A^{-1}Be_j\|$ is constant in j; otherwise target recovery and nuisance estimation induce different weights. For correlated or non-coordinate calibration channels, the same efficient-information criterion defines the target-oriented design, but the closed form above is not claimed.

Known and learned nuisance responses. Proposition 5.1 assumes the acquisition-side response B is known. Pilot estimation adds response-estimation error distinct from allocation error. Passive noisy covariates additionally create errors-in-variables attenuation; Section 6 studies this implementation issue and Appendix I gives the details.

The same paired geometry that determines whether calibration is informative also determines where calibration effort should be spent. Auxiliary measurements should therefore be designed for the downstream target rather than for nuisance estimation in isolation. The next section tests this consequence and separates the gain from target-oriented allocation from the additional cost of learning the nuisance response.

6 NUMERICAL EVIDENCE

The experiments test the constrained timing rates and target-oriented design rule, with additional operator checks reported in the appendices.

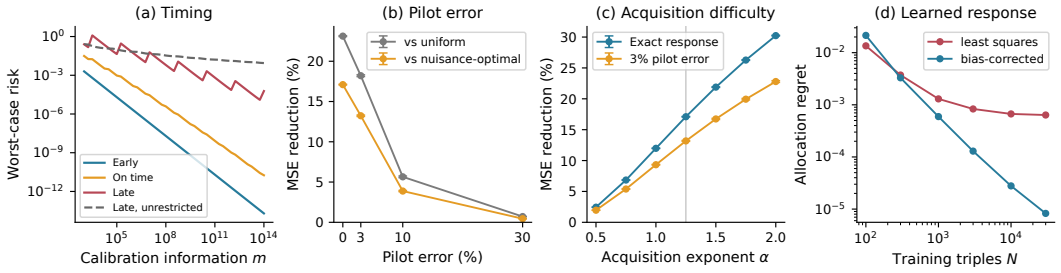

Figure 1: Constraints and target-oriented calibration. (a) Bounded-state timing risks versus unrestricted late calibration. (b,c) Held-out MSE gains of pilot-based target allocation across pilot quality and acquisition difficulty. (d) Allocation regret for least-squares and bias-corrected response learning. Intervals are obtained by bootstrapping 80 independent pilots or training sets, each with 1000 test problems; full protocols are in Appendices H.4 and I.

Figure 1(a) exhibits the three scaling laws predicted by Theorem 4.2 for its constructive rule; normalized risks remain stable over the grid. The constrained and unrestricted late rules separate only at large budgets—their risk ratio reaches 10 near $m = 10^{8.3}$—as detailed in Appendix H.4.

In the 24-dimensional equal-cost experiment, target-oriented allocation at 3% pilot error reduces held-out MSE by 13.2% against nuisance-optimal allocation and 18.2% against uniform (Figure 1(b)). Its true-response allocation regret is only 0.009%, whereas the remaining 35% gap to the known-response oracle is response-estimation error; Appendix I.3 gives intervals and the pilot-quality sweep. As α rises from 0.5 to 2, the gain over nuisance-optimal allocation increases from 2.0% to 22.8% under 3% pilot error (Figure 1(c)), following the widening gap between target weights $\|\mathsf{A}^{-1}\mathsf{B}e_j\|^2/d_j^2$ and nuisance weights d_j^{-2}.

Passive triples yield an errors-in-variables regression: least squares converges to the prediction-optimal but allocation-biased response $\mathsf{B}\Sigma_\eta(\Sigma_\eta + \Sigma_u)^{-1}$. Subtracting the design-known covariate-noise Gram is consistent and recovers the allocation performance as N increases. At $N = 3 \times 10^4$, the corrected allocation reaches regret 8.3×10^{-6} and a 16.50% gain against 17.10% with a known response, while least squares plateaus at 1.76% (Figure 1(d)); Appendix I.4 gives the fidelity and nonlinear checks. As an operator-level geometry check, a multi-coil linearization confirms that acquisition and nuisance scores couple through the reference image, so marginal channel summaries are insufficient. The shorted, resolvent, and variational information agree to relative error 10^{-15} and give the predicted structural-loss dimension; Appendix J.4 reports the construction. Appendix J verifies the phase exponents, growing-radius threshold, noncommuting comparison, and multiple-channel claims. A fastMRI SENSE study finds interior calibration-budget optima, but its overlapping data uses make it an empirical allocation tradeoff rather than a test of the rates above (Appendix J.5).

7 Conclusion

Auxiliary measurements that do not observe the target directly are useful only when they resolve nuisance directions that interfere with target recovery. We showed that this value is governed by the paired nuisance-score geometry rather than by the marginal quality of either channel. The resulting efficient-information characterization separates structural loss from finite-budget calibration loss and leads to a two-resource minimax theory in which the required auxiliary sample size depends on how calibration resolves target-relevant nuisance directions.

This perspective also clarifies the roles of timing, nuisance constraints, and calibration design. In parabolic models, calibration timing can change the effective ill-posedness and substantially alter the auxiliary sample complexity, while nuisance constraints can change the minimax rate itself. The same information geometry yields target-oriented allocation rules that can differ from allocations optimized for nuisance estimation. Our analysis is limited to the stated linear, spectral, and calibration-design settings; extending these results to growing-dimensional nonlinear models, correlated channel noise, and joint optimization of timing and allocation remains open. Overall, the value, timing, and allocation of auxiliary calibration are determined by the target-relevant nuisance geometry it resolves.

AI USE STATEMENT

Generative AI tools were used to aid and polish the writing of this paper, to support the retrieval and discovery of related work, and to assist in checking and revising portions of the mathematical arguments and proofs. All AI-assisted mathematical content was independently verified by the authors against the stated assumptions and derivations, and all references identified with AI assistance were checked against the original publications. The authors take full responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

The appendices contain all proofs and numerical protocols. The supplementary code has two parts. `target_calibration/` contains a self-contained script that reproduces the bounded-state timing study, the growing-radius and spectral-exponent checks, the pilot-design study, the sweep over acquisition ill-posedness, and the operator-learning study with its nonlinear robustness run, together with a local-check script for the scalar nuisance-radius and near-critical studies. Both need NumPy, SciPy and Matplotlib only and run in about a minute, and the accompanying tables contain pilot-level errors, allocations, and bootstrap summaries. `numerics/` contains the remaining studies: the phase diagram, the noncommuting and variable-coefficient timing experiments, the minimax reduction, the parameter sweeps, the multi-coil local geometry, and the fastMRI SENSE scan, with the tables and figures they produced. Appendix J.5 specifies the public fastMRI data protocol and cluster bootstrap.

REFERENCES

W. N. Anderson, Jr. and G. E. Trapp. Shorted operators. II. *SIAM Journal on Applied Mathematics*, 28(1):60–71, 1975. doi: 10.1137/0128007.

Ahmed Attia, Alen Alexanderian, and Arvind K. Saibaba. Goal-oriented optimal design of experiments for large-scale Bayesian linear inverse problems. *Inverse Problems*, 34(9), 2018. doi: 10.1088/1361-6420/aad210. URL `https://arxiv.org/abs/1802.06517`.

J. M. Begun, W. J. Hall, W.-M. Huang, and J. A. Wellner. Information and asymptotic efficiency in parametric-nonparametric models. *The Annals of Statistics*, 11(2):432–452, 1983. doi: 10.1214/aos/1176346151.

R. Bhatia. *Matrix Analysis*. Springer, New York, 1997. doi: 10.1007/978-1-4612-0653-8.

V. I. Bogachev. *Gaussian Measures*. American Mathematical Society, Providence, 1998.

Adrian Bürger, Dimitris Kouzoupis, Angelika Altmann-Dieses, and Moritz Diehl. A Schur complement method for optimum experimental design in the presence of process noise. *IFAC-PapersOnLine*, 50(1):14118–14124, 2017. doi: 10.1016/j.ifacol.2017.08.1853.

L. Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 2008. doi: 10.1088/0266-5611/24/3/034004.

L. Cavalier and N. W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21(4):1345–1361, 2005. doi: 10.1088/0266-5611/21/4/010.

F. Comte and C. Lacour. Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):601–627, 2011. doi: 10.1111/j.1467-9868.2011.00775.x.

Maarten V. de Hoop, Nikola B. Kovachki, Nicholas H. Nelsen, and Andrew M. Stuart. Convergence rates for learning linear operators from noisy data. *SIAM/ASA Journal on Uncertainty Quantification*, 11(2):480–513, 2023. doi: 10.1137/21M1442942.

Litao Ding and Peter Mathé. Minimax rates for statistical inverse problems under general source conditions. arXiv:1707.01706, 2017. URL `https://arxiv.org/abs/1707.01706`.

R. G. Douglas. On majorization, factorization, and range inclusion of operators on Hilbert space. *Proceedings of the American Mathematical Society*, 17(2):413–415, 1966. doi: 10.2307/2035178.

Steve Hanneke and Samory Kpotufe. On the value of target data in transfer learning. In *Advances in Neural Information Processing Systems 32*, pp. 9867–9877, 2019.

S. Hassi, Z. Sebestyén, H. S. V. de Snoo, and F. H. Szafraniec. A canonical decomposition for linear operators and linear relations. *Acta Mathematica Hungarica*, 115:281–307, 2007. doi: 10.1007/s10474-007-5247-y.

D. Henry. *Geometric Theory of Semilinear Parabolic Equations*, volume 840 of *Lecture Notes in Mathematics*. Springer, Berlin, 1981.

R. O. Hryniv and Y. V. Mykytyuk. Eigenvalue asymptotics for Sturm–Liouville operators with singular potentials. *Journal of Functional Analysis*, 238(1):27–57, 2006. doi: 10.1016/j.jfa.2006.04.015.

J. Johannes. Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5A): 2301–2323, 2009. doi: 10.1214/08-AOS652.

J. Johannes and M. Schwarz. Adaptive Gaussian inverse regression with partially unknown operator. *Communications in Statistics—Theory and Methods*, 42(7):1343–1362, 2013. doi: 10.1080/03610926.2012.731548.

T. Kato. *Perturbation Theory for Linear Operators*. Springer, Berlin, 2 edition, 1995. doi: 10.1007/978-3-642-66282-9.

Florian Knoll, Jure Zbontar, Anuroop Sriram, Matthew J Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J Geras, Joe Katsnelson, Hersh Chandarana, et al. fastmri: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning. *Radiology: Artificial Intelligence*, 2(1):e190007, 2020.

Samory Kpotufe and Guillaume Martinet. Marginal singularity and the benefits of labels in covariate-shift. *The Annals of Statistics*, 49(6):3299–3323, 2021. doi: 10.1214/21-AOS2084.

S. Ling and T. Strohmer. Self-calibration and bilinear inverse problems via linear least squares. *SIAM Journal on Imaging Sciences*, 11(1):252–292, 2018. doi: 10.1137/16M1103634.

B. A. Mair and F. H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics*, 56(5):1424–1444, 1996. doi: 10.1137/S0036139994264476.

M. Marcus and V. J. Mizel. Nemitsky operators on Sobolev spaces. *Archive for Rational Mechanics and Analysis*, 51:347–370, 1973. doi: 10.1007/BF00263040.

P. Mathé. Bayesian inverse problems with non-commuting operators. *Mathematics of Computation*, 88(320):2897–2912, 2019. doi: 10.1090/mcom/3439.

P. Mathé and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003. doi: 10.1088/0266-5611/19/3/319.

S. K. Mitra and S. Puntanen. The shorted operator statistically interpreted. *Calcutta Statistical Association Bulletin*, 40(1–4):97–102, 1990. doi: 10.1177/0008068319900508.

F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Analysis*, 18(1–2): 29–37, 1984. doi: 10.1080/00036818408839508.

M. H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330, 1997. doi: 10.1080/10485259708832708.

M. Oliver and C. Wulff. A-stable Runge–Kutta methods for semilinear evolution equations. *Journal of Functional Analysis*, 263:1981–2023, 2012. doi: 10.1016/j.jfa.2012.06.022.

A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer, New York, 1983. doi: 10.1007/978-1-4612-5561-1.

Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 60(3):550–591, 2018. doi: 10.1137/16M1082469.

Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, and Peter Boesiger. Sense: sensitivity encoding for fast mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

B. Simon. A canonical decomposition for quadratic forms with applications to monotone convergence theorems. *Journal of Functional Analysis*, 28(3):377–385, 1978. doi: 10.1016/0022-1236(78)90094-0.

M. Trabs. Information bounds for inverse problems with application to deconvolution and Lévy models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 51(4):1620–1650, 2015. doi: 10.1214/14-AIHP627.

A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009. doi: 10.1007/b13794.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998. doi: 10.1017/CBO9780511802256.

A. Zettl. *Sturm–Liouville Theory*. American Mathematical Society, Providence, 2005.

A Dynamical origins and local nonlinear limits

Let u be a fixed acquisition input and let q be a fixed calibration design. Let a denote the initial-state parameter and let r denote the dynamic nuisance parameter. The hidden state satisfies $\dot{h}_t = F(h_t, u_t, r_t)$ with $h_0 = H(a)$. At a reference parameter value (a_0, r_0), suppose the solution map with u fixed has a bounded derivative $\mathsf{K}_u : \mathbf{\Theta} \to \mathcal{V}$. A nuisance tangent is written as $\theta = (b, \rho)$, where b is the tangent perturbation of a and ρ is the tangent perturbation of r. The corresponding tangent trajectory $w = \mathsf{K}_u\theta$ solves

$$\dot{w}_t = D_h F_t w_t + D_r F_t \rho_t,\ w_0 = DH(a_0)b. \tag{A.1}$$

Thus b enters through the perturbed initial condition, whereas ρ enters continuously through the linearized dynamics; no $D_u F_t$ term appears because u is held fixed. If $\mathsf{B}_{0,u} : \mathcal{V} \to \mathcal{Y}$ and $\mathsf{C}_{0,q} : \mathcal{V} \to \mathcal{Z}$ are the acquisition and calibration observation derivatives with respect to the tangent trajectory, respectively, then $\mathsf{B} = \mathsf{B}_{0,u}\mathsf{K}_u$ and $\mathsf{C} = \mathsf{C}_{0,q}\mathsf{K}_u$. Hence the same nuisance tangent θ first generates a common tangent trajectory through K_u, which is then observed through the two different observation derivatives. This shared tangent trajectory induces the pairing between B and C.

Theorem A.1 (Evolution-generated relative operator)**.** *Let $\mathsf{E}(t, s)$ be a strongly continuous evolution family on a Hilbert space $\mathcal{H}_{\mathrm{st}}$, with bounded injective operators of dense range, and let $J_Y : \mathcal{H}_{\mathrm{st}} \to \mathcal{Y}$ and $J_Z : \mathcal{H}_{\mathrm{st}} \to \mathcal{Z}$ be bounded isomorphisms. For an initial-state tangent, set $\mathsf{B}_{t_a} = J_Y\mathsf{E}(t_a, 0)$ and $\mathsf{C}_\tau = J_Z\mathsf{E}(\tau, 0)$. Then $\mathcal{G}_{BC}$ has zero multivalued part and is the graph of*

$$\mathsf{R}_{t_a,\tau} = \begin{cases} J_Y\mathsf{E}(t_a, \tau)J_Z^{-1}, & \tau \le t_a, \\ J_Y\mathsf{E}(\tau, t_a)^{-1}J_Z^{-1}, & \tau > t_a, \end{cases} \tag{A.2}$$

where the late-calibration operator has domain $J_Z \operatorname{ran}\mathsf{E}(\tau, t_a)$. If $\mathsf{E}(t, s) = e^{-(t-s)G}$ for a nonnegative self-adjoint G and J_Y, J_Z are unitary, then $\mathsf{R}_{t_a,\tau} = J_Y e^{(\tau-t_a)G}J_Z^*$ on its maximal spectral domain, and in that case $\mathsf{R}_{t_a,\tau_1}\mathsf{R}_{t_a,\tau_1}^* \preceq \mathsf{R}_{t_a,\tau_2}\mathsf{R}_{t_a,\tau_2}^*$ for $\tau_1 \le \tau_2$, so the efficient information is nonincreasing in the calibration time.

Initial-state propagation is single valued, but forcing paths can generate a multivalued relation because later calibration may smooth away a component that acquisition still detects.

Proposition A.2 (Forced diffusion and the graph/non-graph boundary)**.** *Let L be positive self-adjoint with compact inverse, set $E_t = e^{-tL^\nu}$, fix $0 < t_a < \tau$, and write $\Delta = \tau - t_a$. For $\mathbf{\Theta} = L^2(0, t_a; \mathcal{H}_{\mathrm{st}})$, define*

$$K_{t_a}\rho = \int_0^{t_a} E_{t_a-s}\rho_s\, ds,\ \mathsf{C}\rho = E_\Delta K_{t_a}\rho, \tag{A.3}$$

and, for $\varphi \in \mathcal{H}_{\mathrm{st}}$ and $J \in \mathcal{L}(\mathcal{H}_{\mathrm{st}})$, $\mathsf{B}\rho = (\langle\varphi, K_{t_a}\rho\rangle, JE_\Delta K_{t_a}\rho)$ in $\mathcal{Y} = \mathbb{R} \oplus \mathcal{H}_{\mathrm{st}}$. If $\varphi = E_\Delta q$, then $\mathcal{G}_{BC} = \operatorname{gra}[z \mapsto (\langle q, z\rangle, Jz)]$ and $\mathcal{U}_0 = \{0\}$. If $\varphi \notin \operatorname{ran}E_\Delta$, then

$$\mathcal{G}_{BC} = \{(z, (u, Jz)) : z \in \mathcal{H}_{\mathrm{st}},\ u \in \mathbb{R}\},\ \mathcal{U}_0 = \mathbb{R} \oplus \{0\}. \tag{A.4}$$

A concrete nonlinear realization is given by a semilinear reaction-diffusion equation on $D = (0, 1)$. Let G_0 be a positive uniformly elliptic Neumann operator on $L^2(D)$ and set $\mathcal{H}_{\mathrm{st}} = \operatorname{dom} G_0^{1/2} = H^1(D)$. For $\Phi, \sigma_Y, \sigma_Z \in C_b^3(\mathbb{R})$, interpreted as Nemytskii maps, consider the mild solution

$$\dot{h}_t + G_1 h_t = \Phi(h_t) + f_t,\ h_0 = a, \tag{A.5}$$

and the observation means

$$\mu_Y(x, a) = Kx + J_Y \sigma_Y(h_{t_a}(a)),\ \mu_Z(a) = J_Z \sigma_Z(h_\tau(a)). \tag{A.6}$$

At a reference trajectory h^0, the Fréchet derivative in the initial-state direction b is the solution of $\dot{w}_t + G_1 w_t = \Phi'(h_t^0) w_t$ with $w_0 = b$. The resulting scores are

$$\mathsf{A}v = Kv,\ \mathsf{B}b = J_Y[\sigma_Y'(h_{t_a}^0) w_{t_a}(b)],\ \mathsf{C}b = J_Z[\sigma_Z'(h_\tau^0) w_\tau(b)]. \tag{A.7}$$

The $H^1(D)$ algebra property and twice continuous Fréchet differentiability of the semiflow give an $O(N^{-1})$ mean remainder, uniformly on compact subsets of every fixed finite-dimensional local parameter space. Propositions B.1 and B.2 in Appendix B state the full assumptions and prove that the corresponding experiment converges in total variation on compact sets to the Gaussian shift

$$q_\gamma(v, \theta) = (\sqrt{\gamma_Y}(\mathsf{A}v + \mathsf{B}\theta), \sqrt{\gamma_Z}\mathsf{C}\theta)\,. \tag{A.8}$$

The following corollary makes the connection with the paired relation explicit.

Corollary A.3 (Paired efficient information in the local asymptotic normality (LAN) limit)**.** *Let $E_X \subseteq \mathcal{X}$ and $E_\Theta \subseteq \boldsymbol{\Theta}$ be finite dimensional, and restrict the limiting experiment in* (A.8) *to $E_X \oplus E_\Theta$. Set $\mathsf{A}_E = \mathsf{A}|_{E_X}$, $\mathsf{C}_E = \mathsf{C}|_{E_\Theta}$, and $\mathsf{B}_E = \mathsf{B}|_{E_\Theta}$. Let $\mathcal{G}_{BC,E}$ be the paired relation generated by $(\mathsf{C}_E, \mathsf{B}_E)$, and let $\mathcal{U}_{0,E}$, $P_{0,E}$, R_E, and $\mathsf{A}_{0,E} = P_{0,E}\mathsf{A}_E$ be its decomposition. If $\gamma_Z > 0$, the efficient information for the object coordinate is*

$$\mathsf{I}_{\mathrm{eff}}^{\gamma,E} = \gamma_Y \mathsf{A}_{0,E}^* \left[I + (\gamma_Y/\gamma_Z)\mathsf{R}_E \mathsf{R}_E^*\right]^{-1} \mathsf{A}_{0,E}. \tag{A.9}$$

If $\gamma_Z = 0$, then

$$\mathsf{I}_{\mathrm{eff}}^{(\gamma_Y,0),E} = \gamma_Y \mathsf{A}_E^* \Pi_{\overline{(\operatorname{ran} \mathsf{B}_E)}^\perp} \mathsf{A}_E = \gamma_Y \mathsf{A}_{0,E}^* \Pi_{\ker \mathsf{R}_E^*} \mathsf{A}_{0,E}. \tag{A.10}$$

Corollary A.3 identifies the efficient information in the finite-dimensional local limit and shows that the same multivalued/operator-part decomposition is preserved under linearization of the nonlinear model. The result applies to each fixed finite-dimensional local experiment. Transferring nonparametric rates to the semilinear model requires control of the remainder as the local dimension grows.

B Fixed-dimensional local Gaussian linearization

B.1 Analytic conventions

All Hilbert spaces are real and separable. Every Gaussian shift is realized, through an orthonormal basis chosen for the experiment and adapted to the relevant closed subspaces when needed, on a finite-dimensional Euclidean space or on $\mathbb{R}^{\mathbb{N}}$ with its product Borel σ-field. All experiments are carried by standard Borel spaces, orthogonal restrictions are measurable statistics, and estimators and Markov kernels are Borel measurable. For Hilbert spaces H_1, H_2, let $\mathcal{L}(H_1, H_2)$ denote the bounded linear operators. For a linear relation $G \subseteq H_1 \oplus H_2$, set $\operatorname{dom} G = \{x : \exists y,\ (x, y) \in G\}$, $\operatorname{ran} G = \{y : \exists x,\ (x, y) \in G\}$, $\ker G = \{x : (x, 0) \in G\}$, and $\operatorname{mul} G = \{y : (0, y) \in G\}$. The graph of an operator T is denoted by $\operatorname{gra} T$; adjoints, orthogonality, and quadratic forms refer to the stated Hilbert structures.

For bounded self-adjoint operators A and B on the same Hilbert space, write $A \preceq B$ when $\langle Ax, x\rangle \le \langle Bx, x\rangle$ for every x. For possibly unbounded nonnegative self-adjoint operators, the same notation denotes closed quadratic-form order: $A \preceq B$ means $\operatorname{dom} B^{1/2} \subseteq \operatorname{dom} A^{1/2}$ and $\left\|A^{1/2}x\right\|^2 \le \left\|B^{1/2}x\right\|^2$ for every $x \in \operatorname{dom} B^{1/2}$. Write $A \asymp B$ when $cB \preceq A \preceq CB$ for constants $0 < c \le C < \infty$. For positive scalar quantities, $a \lesssim b$ means $a \le Cb$, $a \gtrsim b$ means $b \lesssim a$, and $a \asymp b$ means both comparisons hold. Unless a theorem states otherwise, comparison constants may depend on fixed structural parameters but not on the asymptotic indices under consideration. Set $u \vee v = \max\{u, v\}$.

Whenever $R : \operatorname{dom} R \subseteq H_1 \to H_2$ is closed and densely defined, RR^* denotes the nonnegative self-adjoint operator associated with the closed form $q_R(y) = \|R^* y\|^2$ on $\operatorname{dom} R^*$. For $t > 0$, $[I_{H_2} + tRR^*]^{-1}$ denotes its bounded spectral resolvent. All comparisons involving such products use closed quadratic-form order. These conventions follow the standard closed-form and linear-relation frameworks (Kato, 1995; Simon, 1978; Hassi et al., 2007).

If I is a bounded nonnegative self-adjoint operator with spectral resolution E_I, define its Moore–Penrose inverse and inverse square root by

$$I^\dagger = \int_{(0,\infty)} \lambda^{-1}\, dE_I(\lambda), \qquad \operatorname{dom} I^\dagger = \left\{ h : \int_{(0,\infty)} \lambda^{-2}\, d\langle E_I(\lambda)h, h\rangle < \infty \right\},$$

$$I^{\dagger/2} = (I^\dagger)^{1/2} = \int_{(0,\infty)} \lambda^{-1/2}\, dE_I(\lambda), \quad \operatorname{dom} I^{\dagger/2} = \left\{ h : \int_{(0,\infty)} \lambda^{-1}\, d\langle E_I(\lambda)h, h\rangle < \infty \right\}.$$

Both operators act as zero on $\ker I$; an estimability statement separately requires the relevant vector to belong to $\operatorname{ran} I^{1/2}$. In a Gaussian shift $\mathbb{X}(g) = \langle Sx, g\rangle + W(g)$, a square-integrable linear unbiased estimator of $x \mapsto \langle \psi, x\rangle$ is a statistic $\mathbb{X}(g)$ with $S^* g = \psi$.

Whenever L is positive self-adjoint with compact inverse and eigensystem (e_k, λ_k), and $\zeta = (\zeta_k)$ is a positive sequence, $\zeta(L)$ denotes the closed spectral multiplier

$$\zeta(L)x = \sum_{k\geq 1} \zeta_k x_k e_k,\ \operatorname{dom}\zeta(L) = \left\{ x = \sum_{k\geq 1} x_k e_k : \sum_{k\geq 1} \zeta_k^2 |x_k|^2 < \infty \right\}.$$

Proposition B.1 (Fixed-dimensional local linearization and LAN for Gaussian observations)**.** *Let $E \subseteq \mathcal{X} \oplus \boldsymbol{\Theta}$ be finite dimensional, write $h = (v, \theta) \in E$, and let $n_N, m_N > 0$ satisfy $n_N/N \to \gamma_Y > 0$ and $m_N/N \to \gamma_Z \geq 0$. For each N, let $\mu_{Y,N} : E \to \mathcal{Y}$ and $\mu_{Z,N} : E \to \mathcal{Z}$. Let W_Y and W_Z be independent isonormal Gaussian processes on $\mathcal{Y}$ and $\mathcal{Z}$, defined on a fixed canonical product space. After centering at $h = 0$ and whitening, let $P_{N,h}$ denote the law of the two independent processes*

$$\mathbb{Y}_N(y) = \sqrt{n_N}\ \langle \mu_{Y,N}(h) - \mu_{Y,N}(0), y\rangle_{\mathcal{Y}} + W_Y(y), \quad y \in \mathcal{Y},$$
$$\mathbb{Z}_N(z) = \sqrt{m_N}\ \langle \mu_{Z,N}(h) - \mu_{Z,N}(0), z\rangle_{\mathcal{Z}} + W_Z(z), \quad z \in \mathcal{Z}.$$

Assume that the mean increments satisfy

$$\begin{aligned} \mu_{Y,N}(h) - \mu_{Y,N}(0) &= N^{-1/2}(\mathsf{A}v + \mathsf{B}\theta) + r_{Y,N}(h), \\ \mu_{Z,N}(h) - \mu_{Z,N}(0) &= N^{-1/2}\mathsf{C}\theta + r_{Z,N}(h), \end{aligned} \tag{B.1}$$

and that, for every compact $K \subset E$,

$$\sup_{h\in K} \left[\sqrt{n_N}\, \|r_{Y,N}(h)\|_{\mathcal{Y}} + \sqrt{m_N}\, \|r_{Z,N}(h)\|_{\mathcal{Z}} \right] \longrightarrow 0. \tag{B.2}$$

Let P_h^γ denote the law of the isonormal Gaussian shift on $\mathcal{Y} \oplus \mathcal{Z}$ with score

$$q_\gamma(h) = (\sqrt{\gamma_Y}(\mathsf{A}v + \mathsf{B}\theta), \sqrt{\gamma_Z}\mathsf{C}\theta)\,. \tag{B.3}$$

Then $\sup_{h\in K} \|P_{N,h} - P_h^\gamma\|_{\mathrm{TV}} \to 0$ for every compact $K \subset E$. For every fixed $h \in E$, the log-likelihood ratio $\log(dP_{N,h}/dP_{N,0})$ satisfies the LAN expansion with information operator $q_\gamma^* q_\gamma$.

Proposition B.2 (Semilinear reaction-diffusion realization)**.** *Let $D = (0,1)$, and let G_0 be the positive self-adjoint Neumann realization on $L^2(D)$ of $-\partial_\xi(\mathfrak{a}(\xi)\partial_\xi) + q(\xi) + 1$, where $\mathfrak{a} \in C^1(\overline{D})$ is uniformly positive and $q \in L^\infty(D)$ is nonnegative. Its domain consists of the functions $u \in H^2(D)$ satisfying $(\mathfrak{a}u')(0) = (\mathfrak{a}u')(1) = 0$. Set $\mathcal{H}_{\mathrm{st}} = \operatorname{dom} G_0^{1/2} = H^1(D)$ with its graph norm. Let G_1 be the part of G_0 in $\mathcal{H}_{\mathrm{st}}$; its domain consists of those $u \in \operatorname{dom} G_0$ for which $G_0 u \in \mathcal{H}_{\mathrm{st}}$. Set $\boldsymbol{\Theta} = \mathcal{H}_{\mathrm{st}}$. The semigroup generated by $-G_1$ is the restriction of e^{-tG_0} to $\mathcal{H}_{\mathrm{st}}$. Interpret $\Phi, \sigma_Y, \sigma_Z \in C_b^3(\mathbb{R})$ as Nemytskii maps on $\mathcal{H}_{\mathrm{st}}$. Fix $a_0 \in \mathcal{H}_{\mathrm{st}}$, $f \in C([0,T]; \mathcal{H}_{\mathrm{st}})$, and $t_a, \tau \in [0,T]$. For a in a neighborhood of a_0, let $h(a) \in C([0,T]; \mathcal{H}_{\mathrm{st}})$ be the unique mild solution of*

$$\dot{h}_t + G_1 h_t = \Phi(h_t) + f_t,\ h_0 = a. \tag{B.4}$$

Let $K \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, $J_Y \in \mathcal{L}(\mathcal{H}_{\mathrm{st}}, \mathcal{Y})$, and $J_Z \in \mathcal{L}(\mathcal{H}_{\mathrm{st}}, \mathcal{Z})$, and define

$$\mu_Y(x, a) = Kx + J_Y \sigma_Y(h_{t_a}(a)),\ \mu_Z(a) = J_Z \sigma_Z(h_\tau(a)). \tag{B.5}$$

Set $h^0 = h(a_0)$. For $x_0 \in \mathcal{X}$ and local directions (v, b) in a fixed finite-dimensional subspace of $\mathcal{X} \oplus \mathcal{H}_{\mathrm{st}}$, let $x_N = x_0 + N^{-1/2} v$ and $a_N = a_0 + N^{-1/2} b$, and define

$$\mu_{Y,N}(v, b) = \mu_Y(x_N, a_N),\ \mu_{Z,N}(v, b) = \mu_Z(a_N).$$

Then the expansion (B.1) *holds with*

$$\mathsf{A}v = Kv,\ \mathsf{B}b = J_Y \left[\sigma_Y'(h^0_{t_a}) w_{t_a}(b)\right],\ \mathsf{C}b = J_Z \left[\sigma_Z'(h^0_\tau) w_\tau(b)\right], \tag{B.6}$$

where $w(b) \in C([0, T]; \mathcal{H}_{\mathrm{st}})$ is the mild solution of $\dot{w}_t + G_1 w_t = \Phi'(h^0_t) w_t$ with $w_0 = b$. On every compact subset of the chosen finite-dimensional parameter space, both remainders are $O(N^{-1})$, with a constant determined by a bounded H^1-neighborhood of the reference trajectory, T, the coefficient bounds, and the observation-operator norms. The remainder bound (B.2) *holds whenever $n_N, m_N = O(N)$.*

Proof of Proposition B.1. Let $\mathcal{G} = \mathcal{Y} \oplus \mathcal{Z}$. For $h = (v, \theta) \in E$, define

$$q_N(h) = \left(\sqrt{n_N}\left[N^{-1/2}(\mathsf{A}v + \mathsf{B}\theta) + r_{Y,N}(h)\right], \sqrt{m_N}\left[N^{-1/2}\mathsf{C}\theta + r_{Z,N}(h)\right]\right). \tag{B.7}$$

By the definition of $P_{N,h}$, the local law is the Cameron–Martin shift with mean $q_N(h)$, while the limiting law has mean $q_\gamma(h)$ from (B.3). Equation (B.2), compactness of K, and convergence of n_N/N and m_N/N give

$$\sup_{h \in K} \|q_N(h) - q_\gamma(h)\|_{\mathcal{G}} \longrightarrow 0. \tag{B.8}$$

For Gaussian shifts with common covariance, the Hellinger formula and the standard comparison between Hellinger and total-variation distance imply

$$\|P_q - P_{\widetilde{q}}\|_{\mathrm{TV}} \le C \|q - \widetilde{q}\|_{\mathcal{G}} \tag{B.9}$$

for a universal constant C; see Bogachev (1998). This proves compact-uniform total-variation convergence.

Let $\mathbb{W}_N$ be the centered isonormal process under the baseline law. The Cameron–Martin likelihood formula gives

$$\log \frac{dP_{N,h}}{dP_{N,0}} = \mathbb{W}_N(q_N(h)) - \frac{1}{2} \|q_N(h)\|^2_{\mathcal{G}}. \tag{B.10}$$

For fixed h, define $\Delta_N(h) = \mathbb{W}_N(q_\gamma(h))$. By (B.8), $\mathbb{W}_N(q_N(h) - q_\gamma(h)) = o_{P_{N,0}}(1)$ and $\|q_N(h)\|^2 = \|q_\gamma(h)\|^2 + o(1)$. Hence

$$\log \frac{dP_{N,h}}{dP_{N,0}} = \Delta_N(h) - \frac{1}{2} \left\langle q_\gamma^* q_\gamma h, h \right\rangle + o_{P_{N,0}}(1). \tag{B.11}$$

The process Δ_N is centered Gaussian and linear on E, with covariance $\mathrm{Cov}(\Delta_N(h_1), \Delta_N(h_2)) = \left\langle q_\gamma^* q_\gamma h_1, h_2 \right\rangle$. This is the usual finite-dimensional LAN expansion (van der Vaart, 1998). The argument is not uniform over subspaces whose dimensions increase with N. □

Lemma B.3 (Second-order Nemytskii regularity on $H^1(0,1)$)**.** *Let $D = (0, 1)$ and let $\psi \in C_b^3(\mathbb{R})$. Then the Nemytskii map $N_\psi : H^1(D) \to H^1(D)$ is twice continuously Fréchet differentiable on bounded sets, with*

$$DN_\psi(h)v = \psi'(h)v,\ D^2 N_\psi(h)[v_1, v_2] = \psi''(h) v_1 v_2.$$

The operator norms of $DN_\psi(h)$ and $D^2 N_\psi(h)$ are uniformly bounded when h ranges over a bounded subset of $H^1(D)$.

Proof. The embedding $H^1(D) \hookrightarrow L^\infty(D)$ and the Banach-algebra estimate

$$\|uv\|_{H^1} \le C \|u\|_{H^1} \|v\|_{H^1}$$

imply that the displayed first and second derivatives define bounded multilinear maps. Pointwise Taylor formulas for ψ and ψ' combined with the algebra estimate give the Fréchet remainders in $H^1(D)$. On an H^1-bounded set, the embedding confines all function values to one compact interval, on which ψ', ψ'', and $\psi^{(3)}$ are uniformly continuous. This proves continuity of both derivatives and the uniform operator bounds. General Sobolev superposition results are given by Marcus & Mizel (1973). □

Proof of Proposition B.2. Let $S_1(t) = e^{-tG_0}|_{\mathcal{H}_{\mathrm{st}}}$. Since $\mathcal{H}_{\mathrm{st}} = \operatorname{dom} G_0^{1/2}$ with the graph norm, the spectral calculus for G_0 shows that $(S_1(t))_{t\geq 0}$ is a bounded analytic C_0-semigroup on $\mathcal{H}_{\mathrm{st}}$, with generator $-G_1$. The mild equation is

$$h_t(a) = S_1(t)a + \int_0^t S_1(t-s)\left[\Phi(h_s(a)) + f_s\right] ds. \tag{B.12}$$

Lemma B.3 applies to Φ, σ_Y, and σ_Z. In particular, multiplication by $\Phi'(h_t^0)$ is bounded on $H^1(D)$ uniformly in $t \in [0,T]$. Standard differentiability theory for semilinear semiflows applied to (B.12) gives a neighborhood U of a_0 for which $a \mapsto h(a)$ is twice continuously Fréchet differentiable from U into $C([0,T];\mathcal{H}_{\mathrm{st}})$ (Henry, 1981; Pazy, 1983; Oliver & Wulff, 2012). We record direct estimates because they determine the compact-uniform second-order remainder. For u with $a_0 + u \in U$, set $d_t = h_t(a_0+u) - h_t(a_0)$. On a bounded trajectory neighborhood containing all such solutions, the first derivative of N_Φ is uniformly bounded. The difference of the two mild equations and Gronwall's inequality therefore give

$$\sup_{0\leq t\leq T} \|d_t\|_{H^1} \leq C_1 \|u\|_{H^1}. \tag{B.13}$$

Let $w(u)$ solve the linearized mild equation

$$w_t(u) = S_1(t)u + \int_0^t S_1(t-s)\left[\Phi'(h_s^0)w_s(u)\right] ds, \tag{B.14}$$

and put $z_t = d_t - w_t(u)$. The second-order Taylor formula for N_Φ gives

$$z_t = \int_0^t S_1(t-s)\left[\Phi'(h_s^0)z_s + \mathcal{Q}_s(d_s)\right] ds,\ \|\mathcal{Q}_s(d_s)\|_{H^1} \leq C_2 \|d_s\|_{H^1}^2. \tag{B.15}$$

Combining (B.13) with Gronwall's inequality yields

$$\sup_{0\leq t\leq T} \|z_t\|_{H^1} \leq C_3 \|u\|_{H^1}^2. \tag{B.16}$$

The constants C_1, C_2, C_3 depend only on T, a semigroup bound on $[0,T]$, the C_b^3 bounds of Φ, the H^1 algebra and embedding constants, and the chosen bounded trajectory neighborhood.

Applying the same second-order Nemytskii estimate to σ_Y and σ_Z, and then the bounded maps J_Y and J_Z, gives

$$\mu_Y(x_0+v, a_0+u) - \mu_Y(x_0,a_0) = Kv + \mathsf{B}u + R_Y(u),$$
$$\mu_Z(a_0+u) - \mu_Z(a_0) = \mathsf{C}u + R_Z(u),$$

with $\|R_Y(u)\| + \|R_Z(u)\| \leq C_K \|u\|_{H^1}^2$. Here C_K is uniform when (v,u) ranges over a compact subset of the fixed finite-dimensional local parameter space, and depends additionally on the C_b^3 bounds of σ_Y, σ_Z and the norms of J_Y, J_Z. Taking $v = N^{-1/2}v_0$ and $u = N^{-1/2}b$ gives $O(N^{-1})$ remainders uniformly on compact sets. Multiplication by $\sqrt{n_N}$ and $\sqrt{m_N}$ proves (B.2) when $n_N, m_N = O(N)$. □

Proof of Corollary A.3. For $\gamma_Z > 0$, the paired relation associated with the scaled nuisance scores $\sqrt{\gamma_Z}\mathsf{C}_E$ and $\sqrt{\gamma_Y}\mathsf{B}_E$ has multivalued part $\sqrt{\gamma_Y}\mathcal{U}_{0,E}$ and operator part $\sqrt{\gamma_Y/\gamma_Z}\mathsf{R}_E$. The reduced object score is $\sqrt{\gamma_Y}\mathsf{A}_{0,E}$. Applying Theorem 2.1 to the limiting Gaussian shift gives

$$\mathsf{I}_{\mathrm{eff}}^{\gamma,E} = \gamma_Y \mathsf{A}_{0,E}^* \left[I + (\gamma_Y/\gamma_Z)\mathsf{R}_E\mathsf{R}_E^*\right]^{-1} \mathsf{A}_{0,E}. \tag{B.17}$$

If $\gamma_Z = 0$, the nuisance score range is $\overline{\operatorname{ran}\mathsf{B}_E}$. Orthogonal projection therefore gives $\mathsf{I}_{\mathrm{eff}}^{(\gamma_Y,0),E} = \gamma_Y \mathsf{A}_E^* \Pi_{(\overline{\operatorname{ran}\mathsf{B}_E})^\perp} \mathsf{A}_E$. The relation decomposition implies $\overline{\operatorname{ran}\mathsf{B}_E} = \mathcal{U}_{0,E} \oplus \overline{\operatorname{ran}\mathsf{R}_E}$, which yields the second expression in (A.10). □

C EVOLUTION AND FORCING GENERATED RELATIONS

Proof of Theorem A.1. The evolution identity is

$$\mathsf{E}(t,r)\mathsf{E}(r,s) = \mathsf{E}(t,s),\ 0 \le s \le r \le t \le T. \tag{C.1}$$

Assume first that $\tau \le t_a$. For every $\theta \in \mathcal{H}_{\mathrm{st}}$,

$$\mathsf{B}_{t_a}\theta = J_Y\mathsf{E}(t_a,\tau)\mathsf{E}(\tau,0)\theta = J_Y\mathsf{E}(t_a,\tau)J_Z^{-1}\mathsf{C}_\tau\theta. \tag{C.2}$$

The range of C_τ is dense in $\mathcal{Z}$, and the transfer operator is bounded. The closure of its graph restricted to $\operatorname{ran}\mathsf{C}_\tau$ is therefore its full graph.

Assume next that $\tau > t_a$. The bounded injective operator $\mathsf{E}(\tau,t_a)$ has dense range, so its inverse is closed and densely defined on $\operatorname{ran}\mathsf{E}(\tau,t_a)$. Set

$$\mathsf{R}_{t_a,\tau} = J_Y\mathsf{E}(\tau,t_a)^{-1}J_Z^{-1},\ \operatorname{dom}\mathsf{R}_{t_a,\tau} = J_Z\operatorname{ran}\mathsf{E}(\tau,t_a). \tag{C.3}$$

For $x = \mathsf{E}(t_a,0)\theta$,

$$\mathsf{C}_\tau\theta = J_Z\mathsf{E}(\tau,t_a)x,\ \mathsf{B}_{t_a}\theta = J_Y x. \tag{C.4}$$

The unclosed paired relation is the graph of $\mathsf{R}_{t_a,\tau}$ restricted to $J_Z\mathsf{E}(\tau,t_a)\operatorname{ran}\mathsf{E}(t_a,0)$. Let $x \in \mathcal{H}_{\mathrm{st}}$ be arbitrary and set $z = J_Z\mathsf{E}(\tau,t_a)x \in \operatorname{dom}\mathsf{R}_{t_a,\tau}$. Since $\operatorname{ran}\mathsf{E}(t_a,0)$ is dense in $\mathcal{H}_{\mathrm{st}}$, choose $x_j \in \operatorname{ran}\mathsf{E}(t_a,0)$ such that $x_j \to x$ in $\mathcal{H}_{\mathrm{st}}$, and set $z_j = J_Z\mathsf{E}(\tau,t_a)x_j$. Then $z_j \to z$ in $\mathcal{Z}$ and $\mathsf{R}_{t_a,\tau}z_j = J_Y x_j \to J_Y x = \mathsf{R}_{t_a,\tau}z$ in $\mathcal{Y}$. The graph of the restriction is dense in $\operatorname{gra}\mathsf{R}_{t_a,\tau}$, so the restriction is a graph core. This proves (A.2). In both timing cases the closed relation is an operator graph, so its multivalued part is zero.

For the self-adjoint semigroup,

$$\mathsf{R}_{t_a,\tau} = J_Y\exp[(\tau-t_a)G]J_Z^*,\ \mathsf{R}_{t_a,\tau}\mathsf{R}_{t_a,\tau}^* = J_Y\exp[2(\tau-t_a)G]J_Y^*. \tag{C.5}$$

Scalar monotonicity of $\lambda \mapsto \exp[2(\tau-t_a)\lambda]$ gives the form order asserted in Theorem A.1. The resolvent in Theorem 2.1 is operator monotone decreasing, which proves the efficient-information order without imposing commutation with $\mathsf{A}\mathsf{A}^*$. □

Proof of Proposition A.2. Write $K = K_{t_a}$ and $E = E_\Delta$. The controllability map K is bounded from $L^2(0,t_a;\mathcal{H}_{\mathrm{st}})$ into $\mathcal{H}_{\mathrm{st}}$. Its adjoint is

$$(K^*x)(s) = E_{t_a-s}x. \tag{C.6}$$

If $K^*x = 0$ almost everywhere, strong continuity at $s = t_a$ gives $x = 0$. Hence $\ker K^* = \{0\}$ and $D = \operatorname{ran}K$ is dense in $\mathcal{H}_{\mathrm{st}}$.

The paired relation is the closure of

$$\{(Ex, (\langle\varphi,x\rangle, JEx)) : x \in D\}. \tag{C.7}$$

Since E is injective with dense range, $E(D)$ is dense in $\mathcal{H}_{\mathrm{st}}$. Consider first the scalar relation

$$T_0 = \{(Ex, \langle\varphi,x\rangle) : x \in D\} \subseteq \mathcal{H}_{\mathrm{st}} \oplus \mathbb{R}. \tag{C.8}$$

A scalar $c \in \mathbb{R}$ belongs to $\operatorname{dom}T_0^*$ exactly when there is $q_c \in \mathcal{H}_{\mathrm{st}}$ such that

$$c\langle\varphi,x\rangle = \langle q_c, Ex\rangle \tag{C.9}$$

for every $x \in D$. Density of D and self-adjointness of E show that this is equivalent to $c\varphi = Eq_c$.

If $\varphi = Eq$, then T_0 is the restriction to $E(D)$ of the bounded functional $z \mapsto \langle q,z\rangle$. Its closure is the full graph of that functional. Adding the continuous component Jz proves the graph conclusion in Proposition A.2.

If $\varphi \notin \operatorname{ran}E$, then $\operatorname{dom}T_0^* = \{0\}$. For a densely defined relation, the multivalued part of its closure is $(\operatorname{dom}T_0^*)^\perp$; hence $\operatorname{mul}\overline{T_0} = \mathbb{R}$. Since $\overline{\operatorname{dom}T_0} = \mathcal{H}_{\mathrm{st}}$, the operator part of $\overline{T_0}$ maps into the zero orthogonal complement of its multivalued part and has closed dense domain. It is therefore the zero operator on all of $\mathcal{H}_{\mathrm{st}}$, so $\overline{T_0} = \mathcal{H}_{\mathrm{st}} \times \mathbb{R}$. The second acquisition component in (C.7) is the continuous map $z \mapsto Jz$. Consequently,

$$\mathcal{G}_{BC} = \{(z,(u,Jz)) : z \in \mathcal{H}_{\mathrm{st}},\ u \in \mathbb{R}\}, \tag{C.10}$$

which gives all claims in (A.4). □

D SHORTING, CLOSED RELATIONS, AND THE RESOLVENT

D.1 PAIRED RELATION AND INTRINSICITY

Lemma D.1 (Intrinsicity of the paired nuisance-score relation)**.** *For every $n, m > 0$,*

$$\mathcal{T}_h = \left\{ \left(\sqrt{n}\, y, \sqrt{m}\, z \right) : (z, y) \in \mathcal{G}_{BC} \right\}. \tag{D.1}$$

The relation $\mathcal{G}_{BC}$ determines the closed nuisance-score subspace for every pair of information levels, whereas $\mathcal{G}_{BC}$ itself does not depend on n or m.

The relation is also invariant under dense reparameterization of the nuisance tangent. Let $\widetilde{\mathbf{\Theta}}$ be a Hilbert space and let $Q \in \mathcal{L}(\widetilde{\mathbf{\Theta}}, \mathbf{\Theta})$ have dense range. Set $\mathsf{B}_Q = \mathsf{B}Q$ and $\mathsf{C}_Q = \mathsf{C}Q$. Then

$$\overline{\{(\mathsf{C}_Q \vartheta, \mathsf{B}_Q \vartheta) : \vartheta \in \widetilde{\mathbf{\Theta}}\}}^{\mathcal{Z} \oplus \mathcal{Y}} = \mathcal{G}_{BC}. \tag{D.2}$$

The paired relation is invariant under bounded reparameterizations of the nuisance tangent with dense range.

Lemma D.2 (Orthogonal operator-part decomposition)**.** *The space $\mathcal{U}_0 = \operatorname{mul} \mathcal{G}_{BC}$ is closed. For $z \in \operatorname{dom} \mathcal{G}_{BC}$, choose $y \in \mathcal{Y}$ with $(z, y) \in \mathcal{G}_{BC}$ and set $\mathsf{R}z = P_0 y$. This rule is independent of the choice of y and defines a closed densely defined operator $\mathsf{R} : \operatorname{dom} \mathsf{R} \subseteq \mathcal{Z}_0 \to \mathcal{Y}_0$ with $\operatorname{dom} \mathsf{R} = \operatorname{dom} \mathcal{G}_{BC}$, and* (2.4) *holds. Consequently every acquisition score paired with z has the form $\mathsf{R}z + u$ with $u \in \mathcal{U}_0$: the first summand is determined by the calibration score and the second is not.*

Proof of Lemma D.2. The set $\mathcal{G}_{BC}$ is a closed linear subspace of $\mathcal{Z} \oplus \mathcal{Y}$. Hence $\{0\} \oplus \mathcal{U}_0 = \mathcal{G}_{BC} \cap (\{0\} \oplus \mathcal{Y})$ is closed, and so is $\mathcal{U}_0$.

If $(z, y_1), (z, y_2) \in \mathcal{G}_{BC}$, then $(0, y_1 - y_2) \in \mathcal{G}_{BC}$, so $y_1 - y_2 \in \mathcal{U}_0$ and $P_0 y_1 = P_0 y_2$. The operator R is well defined. Its graph is

$$\operatorname{gra} \mathsf{R} = \mathcal{G}_{BC} \cap (\mathcal{Z} \oplus \mathcal{Y}_0), \tag{D.3}$$

which is closed. Since $\operatorname{dom} \mathsf{R} = \operatorname{dom} \mathcal{G}_{BC}$ and $\mathcal{Z}_0 = \overline{\operatorname{dom} \mathcal{G}_{BC}}^{\mathcal{Z}}$, the domain of R is dense in $\mathcal{Z}_0$. If $(z, y) \in \mathcal{G}_{BC}$, then $y = P_0 y + (I - P_0) y = \mathsf{R}z + u$ with $u \in \mathcal{U}_0$. Conversely, sums of $(z, \mathsf{R}z)$ and $(0, u)$ belong to $\mathcal{G}_{BC}$. This proves (2.4). A closed densely defined operator has a nonnegative self-adjoint product $\mathsf{R}\mathsf{R}^*$ on its codomain. □

Proof of Lemma D.1. Define $\Lambda_{n,m} : \mathcal{Z} \oplus \mathcal{Y} \to \mathcal{Y} \oplus \mathcal{Z}$ by $\Lambda_{n,m}(z, y) = (\sqrt{n}\, y, \sqrt{m}\, z)$. This map is bounded and boundedly invertible, and $\Lambda_{n,m}(\mathsf{C}\theta, \mathsf{B}\theta) = \mathsf{S}_h \theta$ for every $\theta \in \mathbf{\Theta}$. Since $\Lambda_{n,m}$ is a homeomorphism,

$$\Lambda_{n,m} \mathcal{G}_{BC} = \overline{\{\Lambda_{n,m}(\mathsf{C}\theta, \mathsf{B}\theta) : \theta \in \mathbf{\Theta}\}} = \overline{\operatorname{ran} \mathsf{S}_h} = \mathcal{T}_h. \tag{D.4}$$

For the reparameterized score pair, the inclusion

$$\overline{\{(\mathsf{C}Q\vartheta, \mathsf{B}Q\vartheta) : \vartheta \in \widetilde{\mathbf{\Theta}}\}} \subseteq \mathcal{G}_{BC} \tag{D.5}$$

is immediate. Conversely, let $\theta \in \mathbf{\Theta}$. Since $\operatorname{ran} Q$ is dense, there is a sequence (ϑ_j) in $\widetilde{\mathbf{\Theta}}$ such that $Q\vartheta_j \to \theta$. The boundedness of B and C gives $(\mathsf{C}Q\vartheta_j, \mathsf{B}Q\vartheta_j) \to (\mathsf{C}\theta, \mathsf{B}\theta)$. Taking closures proves the reverse inclusion and hence the asserted equality. □

Proof of Proposition 2.2. The paired relations are

$$\{(\mathsf{C}_\parallel \theta, \mathsf{B}_\parallel \theta) : \theta \in \mathbb{R}^2\} = \{(z, z) : z \in \mathbb{R}\},\ \{(\mathsf{C}_\perp \theta, \mathsf{B}_\perp \theta) : \theta \in \mathbb{R}^2\} = \mathbb{R} \oplus \mathbb{R}. \tag{D.6}$$

For the parallel model,

$$\inf_{\theta \in \mathbb{R}^2} \left[n|v + \theta_1|^2 + m|\theta_1|^2 \right] = \frac{nm}{n+m} |v|^2. \tag{D.7}$$

For the transverse model,

$$\inf_{\theta \in \mathbb{R}^2} \left[n|v + \theta_1|^2 + m|\theta_2|^2 \right] = 0, \tag{D.8}$$

attained at $\theta_1 = -v$ and $\theta_2 = 0$. Direct calculation gives the identical marginal Gram operators. The final non-reducibility statement follows because the proposed functional would receive the same arguments in the two models but would have to return different efficient information operators. □

D.2 Efficient information as a shorted operator

Proposition D.3 (Efficient information as a shorted operator)**.** *The short of $\mathsf{J}_{n,m}$ to the object subspace is*

$$\mathrm{Short}_{\mathcal{X}\oplus\{0\}}(\mathsf{J}_{n,m}) = \begin{bmatrix} \mathsf{I}_{\mathrm{eff},n,m} & 0 \\ 0 & 0 \end{bmatrix}, \tag{D.9}$$

and, for every $v \in \mathcal{X}$,

$$\langle \mathsf{I}_{\mathrm{eff},n,m} v, v\rangle = \inf_{\theta\in\mathbf{\Theta}} \left[n\|\mathsf{A}v + \mathsf{B}\theta\|_{\mathcal{Y}}^2 + m\|\mathsf{C}\theta\|_{\mathcal{Z}}^2 \right]. \tag{D.10}$$

For a bounded nonnegative operator M on a Hilbert space H and a closed subspace $S \subseteq H$, the shorted operator $\mathrm{Short}_S(M)$ is characterized by

$$\langle \mathrm{Short}_S(M)s, s\rangle = \inf_{u\in S^\perp} \langle M(s+u), s+u\rangle\,,\ s \in S. \tag{D.11}$$

Proof of Proposition D.3. Apply (D.11) with $H = \mathcal{X}\oplus\mathbf{\Theta}$, $S = \mathcal{X}\oplus\{0\}$, and $M = \mathsf{J}_{n,m}$. For $s = (v, 0)$ and $u = (0, \theta)$,

$$\langle \mathsf{J}_{n,m}(v,\theta), (v,\theta)\rangle = n\,\|\mathsf{A}v - \mathsf{B}\theta\|_{\mathcal{Y}}^2 + m\,\|\mathsf{C}\theta\|_{\mathcal{Z}}^2\,. \tag{D.12}$$

The infimum is the squared distance from the object score $(\sqrt{n}\mathsf{A}v, 0)$ to the closure of the nuisance score range. Replacing θ by $-\theta$ gives the form in (D.10). The infimum equals $\langle \mathsf{I}_{\mathrm{eff},n,m} v, v\rangle$, proving (D.9). □

D.3 Exact resolvent and its consequences

Proof of Theorem 2.1. Let $a = \mathsf{A}v$. Since the image of $\theta \mapsto (\mathsf{C}\theta, \mathsf{B}\theta)$ is dense in $\mathcal{G}_{BC}$ and the objective is continuous, Proposition D.3 gives

$$\frac{1}{n}\,\langle \mathsf{I}_{\mathrm{eff},n,m} v, v\rangle = \inf_{(z,y)\in\mathcal{G}_{BC}} \left[\|a-y\|_{\mathcal{Y}}^2 + (m/n)\,\|z\|_{\mathcal{Z}}^2 \right]. \tag{D.13}$$

By (2.4), for fixed z, orthogonal projection onto $\mathcal{U}_0$ gives

$$\inf_{u\in\mathcal{U}_0} \|a - \mathsf{R}z - u\|_{\mathcal{Y}}^2 = \|P_0 a - \mathsf{R}z\|_{\mathcal{Y}_0}^2\,, \tag{D.14}$$

which proves (2.7).

Set $t = n/m$, $T = \mathsf{R}\mathsf{R}^*$, $a_0 = P_0 a$, and $Q_t = (I_{\mathcal{Y}_0} + tT)^{-1}$. The range of Q_t lies in $\mathrm{dom}\,T$, so $\mathsf{R}^* Q_t a_0 \in \mathrm{dom}\,\mathsf{R}$. Define $z_t = t\mathsf{R}^* Q_t a_0$. Then

$$\mathsf{R}z_t = tTQ_t a_0 = (I_{\mathcal{Y}_0} - Q_t)a_0. \tag{D.15}$$

For $z = z_t + h$, the cross term in $\|a_0 - \mathsf{R}z\|^2 + t^{-1}\|z\|^2$ vanishes because $\mathsf{R}^* Q_t a_0 = t^{-1} z_t$. The vector z_t is the unique minimizer and the minimum is

$$\|Q_t a_0\|^2 + t^{-1}\|z_t\|^2 = \langle a_0, Q_t a_0\rangle\,. \tag{D.16}$$

Multiplication by n proves (2.6). □

Corollary D.4 (Linear functionals and calibration limits)**.** *In the efficient Gaussian shift, the functional $x \mapsto \langle \psi, x\rangle$ admits a square-integrable linear unbiased estimator if and only if $\psi \in \mathrm{ran}\,\mathsf{I}_{\mathrm{eff},n,m}^{1/2}$. Its minimum variance is*

$$\|\mathsf{I}_{\mathrm{eff},n,m}^{\dagger/2}\psi\|^2. \tag{D.17}$$

In the strong operator topology,

$$\begin{aligned} \lim_{m/n\to\infty} n^{-1}\mathsf{I}_{\mathrm{eff},n,m} &= \mathsf{A}_0^*\mathsf{A}_0, \\ \lim_{m/n\to 0} n^{-1}\mathsf{I}_{\mathrm{eff},n,m} &= \mathsf{A}_0^*\Pi_{\ker\mathsf{R}^*}\mathsf{A}_0. \end{aligned} \tag{D.18}$$

Proof of Corollary D.4. Let $\mathsf{S}_{\text{eff}} = U_{\text{eff}}\mathsf{I}_{\text{eff},n,m}^{1/2}$ be the polar decomposition of the efficient score. A statistic $\mathbb{X}_{\text{eff}}(g)$ is a square-integrable linear unbiased estimator of $x \mapsto \langle \psi, x\rangle$ exactly when $\mathsf{S}_{\text{eff}}^* g = \psi$. Since $\mathsf{S}_{\text{eff}}^* = \mathsf{I}_{\text{eff},n,m}^{1/2} U_{\text{eff}}^*$, such a finite-norm solution exists if and only if $\psi \in \operatorname{ran} \mathsf{I}_{\text{eff},n,m}^{1/2}$. The minimum-norm solution is

$$g_{\text{eff}} = U_{\text{eff}}\mathsf{I}_{\text{eff},n,m}^{\dagger/2}\psi, \|g_{\text{eff}}\|^2 = \left\|\mathsf{I}_{\text{eff},n,m}^{\dagger/2}\psi\right\|^2. \tag{D.19}$$

This proves (D.17).

Set $t = n/m$. By (2.6), the relevant spectral multiplier is $(1+t\lambda)^{-1}$. As $t \downarrow 0$, it converges pointwise to 1, whereas as $t \uparrow \infty$, it converges pointwise to the indicator of $\{0\}$. Spectral calculus therefore gives

$$(I_{\mathcal{Y}_0} + t\mathsf{R}\mathsf{R}^*)^{-1} \longrightarrow I_{\mathcal{Y}_0} \tag{D.20}$$

strongly as $t \downarrow 0$, and

$$(I_{\mathcal{Y}_0} + t\mathsf{R}\mathsf{R}^*)^{-1} \longrightarrow \Pi_{\ker \mathsf{R}^*} \tag{D.21}$$

strongly as $t \uparrow \infty$. Substitution into (2.6) proves (D.18). □

Corollary D.5 (Calibration gain decomposition)**.** *In the strong operator topology,*

$$n^{-1}\mathsf{I}_{\text{or},n} - \lim_{m/n\to 0} n^{-1}\mathsf{I}_{\text{eff},n,m} = \underbrace{\mathsf{A}^*\Pi_{\mathcal{U}_0}\mathsf{A}}_{\textit{structural loss}} + \underbrace{\mathsf{A}_0^*\Pi_{\overline{\operatorname{ran}\mathsf{R}}}\mathsf{A}_0}_{\textit{what calibration buys}}, \tag{D.22}$$

and the second term is the whole gain: $\lim_{m/n\to\infty} n^{-1}\mathsf{I}_{\text{eff},n,m} - \lim_{m/n\to 0} n^{-1}\mathsf{I}_{\text{eff},n,m} = \mathsf{A}_0^*\Pi_{\overline{\operatorname{ran}\mathsf{R}}}\mathsf{A}_0$. *In particular calibration changes nothing, at any budget, if and only if* $\operatorname{ran}\mathsf{A}_0 \perp \operatorname{ran}\mathsf{R}$*, and no calibration budget recovers the first term.*

Proof of Corollary D.5. By Corollary D.4, in the strong operator topology $n^{-1}\mathsf{I}_{\text{eff},n,m}$ tends to $\mathsf{A}_0^*\mathsf{A}_0$ as $m/n \to \infty$ and to $\mathsf{A}_0^*\Pi_{\ker \mathsf{R}^*}\mathsf{A}_0$ as $m/n \to 0$. Since $\ker \mathsf{R}^* = (\overline{\operatorname{ran}\mathsf{R}})^\perp$ inside $\mathcal{Y}_0$,

$$\mathsf{A}_0^*\mathsf{A}_0 - \mathsf{A}_0^*\Pi_{\ker \mathsf{R}^*}\mathsf{A}_0 = \mathsf{A}_0^*\Pi_{\overline{\operatorname{ran}\mathsf{R}}}\mathsf{A}_0,$$

which is the stated gain. For (D.22), write $\mathsf{A} = P_0\mathsf{A} + (I - P_0)\mathsf{A} = \mathsf{A}_0 + \Pi_{\mathcal{U}_0}\mathsf{A}$; the two summands have orthogonal ranges, so $\mathsf{A}^*\mathsf{A} = \mathsf{A}_0^*\mathsf{A}_0 + \mathsf{A}^*\Pi_{\mathcal{U}_0}\mathsf{A}$. Subtracting $\mathsf{A}_0^*\Pi_{\ker \mathsf{R}^*}\mathsf{A}_0$ from $n^{-1}\mathsf{I}_{\text{or},n} = \mathsf{A}^*\mathsf{A}$ gives (D.22). The gain vanishes exactly when $\Pi_{\overline{\operatorname{ran}\mathsf{R}}}\mathsf{A}_0 = 0$, that is when $\operatorname{ran}\mathsf{A}_0 \perp \operatorname{ran}\mathsf{R}$, and in that case (2.6) gives $\mathsf{I}_{\text{eff},n,m} = n\mathsf{A}_0^*\mathsf{A}_0$ for every m, so no budget changes anything. The first term of (D.22) does not depend on m at all. □

D.4 Gaussian nuisance translations and global minimax reduction

Lemma D.6 (Closure stability of bounded Gaussian decision risks)**.** *Let H be a real separable Hilbert space, let $M \subseteq H$ be any nonempty subset, and let $\overline{M}$ be its closure. Let Q_h denote the canonical Gaussian shift on H with Cameron–Martin mean h. Let $\mathcal{P}_0$ be a standard Borel parameter space, let $\mathcal{A}$ be a standard Borel action space, let $\ell : \mathcal{P}_0 \times \mathcal{A} \to \mathbb{R}$ be bounded and Borel measurable, and let δ be a Borel Markov kernel from the sample space to $\mathcal{A}$. For an arbitrary map $\mu : \mathcal{P}_0 \to H$, define*

$$R_\delta(x,t) = \int\int_{\mathcal{A}} \ell(x,a)\,\delta(da \mid \omega) Q_{\mu(x)+t}(d\omega).$$

Then

$$\sup_{x\in\mathcal{P}_0,\, t\in M} R_\delta(x,t) = \sup_{x\in\mathcal{P}_0,\, t\in\overline{M}} R_\delta(x,t).$$

Proof. If $t_j \to t$ in H, then (B.9) gives $\left\|Q_{\mu(x)+t_j} - Q_{\mu(x)+t}\right\|_{\text{TV}} \to 0$ for every fixed x. Since the loss is bounded,

$$|R_\delta(x,t_j) - R_\delta(x,t)| \le 2\|\ell\|_\infty \left\|Q_{\mu(x)+t_j} - Q_{\mu(x)+t}\right\|_{\text{TV}} \to 0.$$

The map $t \mapsto R_\delta(x,t)$ is continuous for every x, and its supremum over M equals its supremum over $\overline{M}$. Taking the supremum over x proves the claim. □

Lemma D.7 (Gaussian nuisance-translation reduction)**.** *Let H and Θ_0 be real separable Hilbert spaces, let $N \in \mathcal{L}(\Theta_0, H)$, set $T = \overline{\operatorname{ran} N}^H$, and let $\Theta_1 \subseteq \Theta_0$ be nonempty. Let $\mathcal{P}_0$ be a compact metric parameter space, let $\mathcal{A}$ be a compact metric action space, let $\mu : \mathcal{P}_0 \to H$ be continuous, and let $\ell : \mathcal{P}_0 \times \mathcal{A} \to \mathbb{R}$ be bounded and continuous. Let Q_h denote the canonical Gaussian shift law on H with mean h, and let $Q_h^\perp$ denote the corresponding law on $T^\perp$. Define*

$$\begin{aligned} \mathcal{R}_{\Theta_1}^{\mu,N} &= \inf_\delta \sup_{x \in \mathcal{P}_0,\, \theta \in \Theta_1} \int\int_{\mathcal{A}} \ell(x,a)\, \delta(da \mid \omega) Q_{\mu(x)+N\theta}(d\omega), \\ \mathcal{R}^\perp &= \inf_\kappa \sup_{x \in \mathcal{P}_0} \int\int_{\mathcal{A}} \ell(x,a)\, \kappa(da \mid \omega_\perp) Q^\perp_{\Pi_{T^\perp}\mu(x)}(d\omega_\perp), \end{aligned} \tag{D.23}$$

where the infima range over Borel Markov kernels on the respective sample spaces. Then

$$\mathcal{R}_{\Theta_1}^{\mu,N} \le \mathcal{R}^\perp, \tag{D.24}$$

with equality whenever $-\Pi_T \mu(x) \in \overline{N(\Theta_1)}$ for every $x \in \mathcal{P}_0$; in particular for $\Theta_1 = \Theta_0$. If $\mathcal{P}_0 = \mathcal{A} = K$ is a compact convex subset of a separable Hilbert space and $\ell(x,a) = \|a - x\|^2$, both infima are unchanged when the rules are restricted to deterministic Borel estimators.

Proof. Choose the canonical Gaussian realization in an orthonormal basis adapted to $H = T \oplus T^\perp$. The observation decomposes into independent coordinates with means $\Pi_{T^\perp}\mu(x)$ and $\Pi_T \mu(x) + N\theta$. A rule for the projected experiment can ignore the T-coordinate in the joint experiment, so $\mathcal{R}_{\Theta_1}^{\mu,N} \le \mathcal{R}_{\Theta_0}^{\mu,N} \le \mathcal{R}^\perp$.

For the reverse inequality under the stated condition, Lemma D.6 shows that the nuisance supremum is unchanged when $N(\Theta_1)$ is replaced by its closure. Fix a joint rule δ. Choose $t_x = -\Pi_T \mu(x)$, which the condition places in that closure and which is therefore an admissible nuisance translation. The T-coordinate then has the centered Gaussian law Q_0^T and is independent of the projected observation. Integrating δ against Q_0^T defines the projected Markov kernel

$$\kappa_\delta(da \mid \omega_\perp) = \int \delta(da \mid \omega_\perp, \omega_T) Q_0^T(d\omega_T).$$

The worst-case risk of δ in the joint experiment is at least the worst-case risk of κ_δ in the projected experiment. Taking the infimum over δ gives $\mathcal{R}_{\Theta_1}^{\mu,N} \ge \mathcal{R}^\perp$. For $\Theta_1 = \Theta_0$ the condition holds because $\Pi_T \mu(x) \in T = \overline{\operatorname{ran} N}$.

In the squared-error case, metric projection restricts all actions to K without increasing loss. For a randomized rule supported on K, its conditional Bochner barycenter is a deterministic Borel estimator with no larger squared error by Jensen's inequality. Deterministic estimators suffice. □

Proposition D.8 (The nuisance radius of Corollary D.10 is sharp)**.** *Let $H = \mathcal{X} = \Theta_0 = \mathbb{R}$, let $S = N = I$, let $K = [-1,1]$, and observe $Y = x + \theta + \varepsilon$ with $\varepsilon \sim N(0,1)$. Here $T = \mathbb{R}$, the projected experiment has no informative coordinate, and its minimax risk is 1. Restrict the nuisance to $[-c_{\mathrm{nuis}}, c_{\mathrm{nuis}}]$ and write $\mathcal{R}(c_{\mathrm{nuis}})$ for the constrained joint minimax squared-error risk. Then*

$$\mathcal{R}(c_{\mathrm{nuis}}) \le \frac{1}{1 + (1 - c_{\mathrm{nuis}})^2} < 1 \quad \text{for } 0 \le c_{\mathrm{nuis}} < 1, \qquad \mathcal{R}(c_{\mathrm{nuis}}) = 1 \quad \text{for } c_{\mathrm{nuis}} \ge 1. \tag{D.25}$$

The radius that Corollary D.10 prescribes here is $\sup_{x \in K} |x| = 1$, so the reduction fails at every smaller radius and holds from that radius on.

Proof. For $c_{\mathrm{nuis}} \ge 1$ the parameter points $(x,\theta) = (1,-1)$ and $(-1,1)$ are both admissible and induce the same law $N(0,1)$. For every estimator $\widehat{x}$ the sum of the squared errors at these two points is $(\widehat{x}-1)^2 + (\widehat{x}+1)^2 = 2\widehat{x}^2 + 2 \ge 2$, so the maximum of the two risks is at least 1; the constant estimator $\widehat{x} = 0$ has maximal risk 1. Hence $\mathcal{R}(c_{\mathrm{nuis}}) = 1$, the unrestricted value.

For $0 \le c_{\mathrm{nuis}} < 1$ put $u = 1 - c_{\mathrm{nuis}}$ and $\lambda = u/(1+u^2)$, and take $\widehat{x} = \mathrm{proj}_{[-1,1]}(\lambda Y)$. Metric projection does not increase squared error for $x \in [-1,1]$, so

$$\mathbb{E}_{x,\theta}(\widehat{x} - x)^2 \le \mathbb{E}_{x,\theta}(\lambda Y - x)^2 = [(1-\lambda)x - \lambda\theta]^2 + \lambda^2 \le [(1-\lambda) + \lambda c_{\mathrm{nuis}}]^2 + \lambda^2,$$

uniformly over $|x| \le 1$ and $|\theta| \le c_{\mathrm{nuis}}$. The right-hand side is $(1-\lambda u)^2 + \lambda^2$, minimized at the stated λ with value $1/(1+u^2)$, which proves (D.25). □

Theorem D.9 (Minimax reduction for restricted nuisance classes)**.** *For every nonempty $\mathbf{\Theta}_0 \subseteq \mathbf{\Theta}$ and every nonempty compact convex $K \subseteq \mathcal{X}$,*

$$\mathcal{R}^{\mathrm{joint}}_{n,m}(K; \mathbf{\Theta}_0) \le \mathcal{R}^{\mathrm{eff}}_{n,m}(K), \tag{D.26}$$

with equality whenever

$$-\Pi_{\mathcal{T}_h} \mathsf{S}_x x \in \overline{\mathsf{S}_h(\mathbf{\Theta}_0)}^{\mathcal{Y}\oplus\mathcal{Z}} \qquad \text{for every } x \in K. \tag{D.27}$$

Condition (D.27) *holds for $\mathbf{\Theta}_0 = \mathbf{\Theta}$, which recovers Theorem 3.1.*

Proof of Theorems 3.1 and D.9. Apply Lemma D.7 with $H = \mathcal{Y} \oplus \mathcal{Z}$, $\mathcal{P}_0 = K$, $\mathcal{A} = K$, $\mu(x) = \mathsf{S}_x x$, $N = \mathsf{S}_h$, $\ell(x,a) = \|a - x\|^2$, and $\Theta_1 = \mathbf{\Theta}_0$. The inequality (D.26) is the first assertion of the lemma and (D.27) is its equality condition; taking $\mathbf{\Theta}_0 = \mathbf{\Theta}$ gives Theorem 3.1. □

Corollary D.10 (A sufficient nuisance radius)**.** *Suppose* $\mathrm{ran}(\mathsf{B} \oplus \mathsf{C})$ *is closed, and let $\kappa_0 < \infty$ be such that every t in that range is $(\mathsf{B}\theta, \mathsf{C}\theta)$ for some θ with $\|\theta\| \le \kappa_0 \|t\|$. If $\mathbf{\Theta}_0$ contains the ball of radius*

$$\rho \ge \kappa_0 \sqrt{\frac{n}{n \wedge m}} \sup_{x \in K} \|\mathsf{A}x\|, \tag{D.28}$$

then (D.27) *holds and the reduction is exact. In particular, when the calibration channel is at least as informative as the acquisition channel, $m \ge n$, the radius $\kappa_0 \sup_{x\in K} \|\mathsf{A}x\|$ suffices for every pair of budgets.*

Proof of Corollary D.10. Fix $x \in K$ and put $t = -\Pi_{\mathcal{T}_h} \mathsf{S}_x x$. Since $\mathsf{S}_x x = (\sqrt{n}\mathsf{A}x, 0)$ and orthogonal projections do not increase norms, $\|t\| \le \sqrt{n}\,\|\mathsf{A}x\|$. By Lemma D.1, $\mathcal{T}_h$ is the image of $\mathcal{G}_{BC}$ under the bounded bijection $(z,y) \mapsto (\sqrt{n}y, \sqrt{m}z)$, so $t \in \overline{\mathrm{ran}\,\mathsf{S}_h}$ and, $\mathrm{ran}(\mathsf{B} \oplus \mathsf{C})$ being closed, $t = \mathsf{S}_h\theta$ for some θ. Writing $t = (t_1, t_2)$, that θ satisfies $(\mathsf{B}\theta, \mathsf{C}\theta) = (t_1/\sqrt{n}, t_2/\sqrt{m})$, a vector of norm at most $\|t\|/\sqrt{n \wedge m}$. Hence θ may be chosen with

$$\|\theta\| \le \kappa_0 \frac{\|t\|}{\sqrt{n \wedge m}} \le \kappa_0 \sqrt{\frac{n}{n \wedge m}}\, \|\mathsf{A}x\|,$$

which is at most ρ by (D.28). So $t \in \mathsf{S}_h(\mathbf{\Theta}_0)$ and (D.27) holds. □

E FINITE-RESOLUTION CONSEQUENCES

Define $\mathsf{I}_{\mathrm{or},n} = n\mathsf{A}^*\mathsf{A}$ and $\Delta_{n,m} = \mathsf{I}_{\mathrm{or},n} - \mathsf{I}_{\mathrm{eff},n,m}$.

Proposition E.1 (Finite-rank information loss)**.** *If $\dim \overline{\mathrm{ran}\,\mathsf{S}_h} = r < \infty$ and $\mathsf{I}_{\mathrm{or},n}$ is compact, then $\mathrm{rank}\,\Delta_{n,m} \le r$. If (μ_k) and (λ_k) are the eigenvalues of $\mathsf{I}_{\mathrm{or},n}$ and $\mathsf{I}_{\mathrm{eff},n,m}$ in nonincreasing order, then*

$$\mu_{k+r} \le \lambda_k \le \mu_k. \tag{E.1}$$

Proof. Let $\mathcal{T}_h$ be the closure of the nuisance score range and $\mathsf{S}_x v = (\sqrt{n}\mathsf{A}v, 0)$. Then $\Delta_{n,m} = \mathsf{S}_x^* \Pi_{\mathcal{T}_h} \mathsf{S}_x$. If $\dim \mathcal{T}_h = r$, the loss has rank at most r. The min–max principle for a nonnegative perturbation of rank at most r gives (E.1); see Bhatia (1997). □

Proposition E.2 (Restricted oracle equivalence)**.** *Let $E_J \subset \mathcal{X}$ be finite dimensional, let P_J be its orthogonal projection, and restrict the object parameter to E_J. Set*

$$\mathsf{I}_{\mathrm{or},J} = P_J \mathsf{I}_{\mathrm{or},n} P_J|_{E_J} \succ 0, \;\; \rho_J = \left\| \mathsf{I}_{\mathrm{or},J}^{-1/2} P_J \Delta_{n,m} P_J \mathsf{I}_{\mathrm{or},J}^{-1/2} \right\|. \tag{E.2}$$

If $\rho_J < 1$, then

$$(1-\rho_J)\mathsf{I}_{\mathrm{or},J} \preceq P_J \mathsf{I}_{\mathrm{eff},n,m} P_J|_{E_J} \preceq \mathsf{I}_{\mathrm{or},J}. \tag{E.3}$$

The condition $\rho_{J_{n,m}} \to 0$ is equivalent to uniform convergence of restricted-experiment information bounds to their oracle values over linear functionals normalized by oracle variance.

Proof. Set $K_J = \mathsf{I}_{\mathrm{or},J}^{-1/2} P_J \Delta_{n,m} P_J \mathsf{I}_{\mathrm{or},J}^{-1/2}$. Then

$$P_J \mathsf{I}_{\mathrm{eff},n,m} P_J|_{E_J} = \mathsf{I}_{\mathrm{or},J}^{1/2}(I-K_J)\mathsf{I}_{\mathrm{or},J}^{1/2}.$$

Since $0 \preceq K_J \preceq \rho_J I$, (E.3) follows. For a functional represented by $\psi \in E_J$, put $q = \mathsf{I}_{\mathrm{or},J}^{-1/2}\psi$. Under the normalization $\|q\| = 1$, the oracle variance is one and the effective variance is $\langle (I-K_J)^{-1}q, q\rangle$. Its maximal excess over the unit sphere is $\rho_J/(1-\rho_J)$, proving the equivalence. □

The restriction in Proposition E.2 is essential. Let $\mathcal{X} = \mathbb{R}^2$, $E = \operatorname{span}\{e_1\}$, $\mathsf{I}_{\mathrm{or},n} = I_2$, and, for $0 < \varepsilon < 1/2$, set

$$D_\varepsilon = \begin{bmatrix} \varepsilon & \sqrt{\varepsilon(1-2\varepsilon)} \\ \sqrt{\varepsilon(1-2\varepsilon)} & 1-2\varepsilon \end{bmatrix}, \; I_{\mathrm{eff},\varepsilon} = I_2 - D_\varepsilon. \tag{E.4}$$

Then $D_\varepsilon \succeq 0$ has rank one and the compressed relative loss on E tends to zero, but $\left\langle I_{\mathrm{eff},\varepsilon}^{-1} e_1, e_1 \right\rangle = 2$ while $\langle \mathsf{I}_{\mathrm{or},n}^{-1} e_1, e_1\rangle = 1$. Full-model functional bounds therefore require shorting to E_J, direct inverse-form comparison, or a reducing-subspace condition.

F Relative form comparison and minimax reconstruction

F.1 Restricted object-score support

Remark F.1 (Relation to deconvolution with an auxiliary sample). The three branches of (3.6) have the same shape as the rates known for deconvolution with an unknown error distribution estimated from an auxiliary sample. For a target in a Sobolev space and an ordinary smooth error of index a, Johannes (2009) obtains a risk of order $n^{-2(p-s)/(2(p+a)+1)} + m^{-(1\wedge\beta)}$ with $\beta = (p-s)/(s+a)$, and proves that the second term cannot be removed when the two samples are independent. In both cases the calibration contribution saturates at the parametric rate once the transfer is favourable enough, and below that threshold it is algebraic with an exponent set by the interplay of object smoothness and the transfer. The mechanisms differ in where the threshold comes from: there it is the source-condition index of the target relative to the smoothness of the error, here it is the exponent $p = 2\alpha + 2\kappa$ of the transfer profile relative to the object score, and (3.6) additionally exhibits the logarithmic law at the boundary $p = -1$. Our framework also varies the coupling between the two nuisance scores. Proposition 2.2 shows how this coupling can create structural loss even when the marginal channel Grams are unchanged.

Let $\mathcal{H}_A = \overline{\operatorname{ran}\mathsf{A}_0}^{\mathcal{Y}_0}$ and let P_A be the orthogonal projection onto $\mathcal{H}_A$. With $t = n/m$ and $\mathsf{T}_R = \mathsf{R}\mathsf{R}^*$, the exact identity is $\mathsf{I}_{\mathrm{eff},n,m} = n\mathsf{A}_0^* P_A (I_{\mathcal{Y}_0} + t\mathsf{T}_R)^{-1} P_A \mathsf{A}_0$; it follows from $P_A\mathsf{A}_0 = \mathsf{A}_0$ and $\mathsf{A}_0^* P_A = \mathsf{A}_0^*$. If $\mathcal{H}_A$ reduces T_R, then $\mathsf{T}_A = \mathsf{T}_R|_{\mathcal{H}_A}$ is nonnegative self-adjoint, spectral calculus gives $P_A(I_{\mathcal{Y}_0} + t\mathsf{T}_R)^{-1}|_{\mathcal{H}_A} = (I_{\mathcal{H}_A} + t\mathsf{T}_A)^{-1}$, and therefore $\mathsf{I}_{\mathrm{eff},n,m} = n\mathsf{A}_0^*(I_{\mathcal{H}_A} + t\mathsf{T}_A)^{-1}\mathsf{A}_0$. Theorems 3.3 and 3.4 remain valid with $\mathcal{Y}_0$ replaced by $\mathcal{H}_A$ whenever there are a unitary $U : \mathcal{X} \to \mathcal{H}_A$ and spectral multipliers $a(L), g(L)$ such that $\mathsf{A}_0 = Ua(L)$ and T_A is two-sided form-comparable to $Ug(L)^2U^*$ on the common form domain. If $\mathcal{H}_A$ does not reduce T_R, the compressed family $P_A(I_{\mathcal{Y}_0} + t\mathsf{T}_R)^{-1}|_{\mathcal{H}_A}$ need not be the resolvent of the restriction of one nonnegative self-adjoint operator to $\mathcal{H}_A$; a direct compressed-resolvent comparison is then an additional assumption.

Proof of Theorem 3.3. Set $t = n/m$ and let $\mathsf{T}_g = Ug(L)^2U^*$, interpreted as the nonnegative self-adjoint operator associated with the closed form $q_{\mathsf{T}_g}(y) = \|g(L)U^*y\|^2$ on $U \operatorname{dom} g(L)$. The form of $\mathsf{R}\mathsf{R}^*$ is $q_R(y) = \|\mathsf{R}^*y\|^2$ on the same domain. Assumption 3.2 gives

$$c_R\mathsf{T}_g \preceq \mathsf{R}\mathsf{R}^* \preceq C_R\mathsf{T}_g \tag{F.1}$$

in closed-form order. Because the two closed forms have the same domain, the order-reversing property of the resolvent under closed quadratic-form order applies; see (Kato, 1995, Chapter VI) and Simon (1978). Hence

$$(I + tC_R\mathsf{T}_g)^{-1} \preceq (I + t\mathsf{R}\mathsf{R}^*)^{-1} \preceq (I + tc_R\mathsf{T}_g)^{-1}. \tag{F.2}$$

For $q \geq 0$ and $c > 0$, $(1 + tcq)^{-1}$ is uniformly comparable to $(1 + tq)^{-1}$, with constants depending only on c. Functional calculus and (2.6) therefore prove (3.3). Evaluation on finitely supported vectors gives (3.4). □

Lemma F.2 (Gaussian comparison tools)**.** *Let P_x^{eff} denote the law of the efficient Gaussian shift in* (3.1)*.*

(i) For $x, y \in \mathcal{X}$,

$$\mathrm{KL}(P_x^{\mathrm{eff}}, P_y^{\mathrm{eff}}) = \frac{1}{2}\left\|\mathsf{S}_{\mathrm{eff}}(x-y)\right\|^2 = \frac{1}{2}\left\langle \mathsf{I}_{\mathrm{eff},n,m}(x-y), x-y\right\rangle. \tag{F.3}$$

(ii) Let D be a positive diagonal operator with $De_k = d_k e_k$, $d_k > 0$, and suppose $cD \preceq I \preceq CD$ for bounded nonnegative I. Then $\operatorname{ran} D^{1/2} \subseteq \operatorname{ran} I^{1/2}$ and, for every $h \in \operatorname{ran} D^{1/2}$,

$$\left\|I^{\dagger/2}h\right\|^2 \leq c^{-1}\left\|D^{-1/2}h\right\|^2. \tag{F.4}$$

In particular, $\left\|I^{\dagger/2}e_k\right\|^2 \leq c^{-1}d_k^{-1}$.

(iii) Let $K \subset \mathbb{N}$ be finite and let $x_\omega = \sum_{k\in K}\omega_k\delta_k e_k$, $\omega_k \in \{-1, 1\}$. If every vertex belongs to the parameter set, $I \preceq CD$, and $2Cd_k\delta_k^2 \leq \kappa_0 < 2$ for all $k \in K$, then

$$\inf_{\widehat{x}}\sup_{\omega}\mathbb{E}_\omega\left\|\widehat{x} - x_\omega\right\|^2 \geq \frac{1}{2}\left(1 - \sqrt{\kappa_0/2}\right)\sum_{k\in K}\delta_k^2. \tag{F.5}$$

An upper quadratic-form bound on the information is sufficient for the hypercube lower bound; no common eigenbasis for I and D is required.

(iv) A hypercube $x_\omega = \sum_{k\in K}\omega_k\delta_k e_k$ lies in $\mathcal{X}_\beta(R)$ whenever $\sum_{k\in K}\lambda_k^{2\beta}\delta_k^2 \leq R^2$. In particular, the band construction $K = \{J, \dots, 2J-1\}$ with $\delta_k = c_0 J^{-\beta-1/2}$ is admissible for all J if $c_0 \leq R(2C_\lambda)^{-\beta}$. At the critical algebraic index, the choice $\delta_k = c_0(mk)^{-1/2}$ for $k_0 \leq k \leq J$ and $J \leq C_J m^{1/(2\beta)}$ is admissible when c_0 is small enough depending only on R, β, C_λ, C_J.

Proof. Part (i) is the Cameron–Martin formula for Gaussian shifts with common covariance. For part (ii), $cD \preceq I$ and Douglas's theorem give $D^{1/2} = c^{-1/2}I^{1/2}C_0$ for a contraction C_0. If $h = D^{1/2}u$, the equation $I^{1/2}v = h$ therefore has a solution with $\|v\| \leq c^{-1/2}\|u\|$. The minimum-norm solution is $I^{\dagger/2}h$, which proves (F.4); see Douglas (1966).

For part (iii), neighboring vertices that differ only in coordinate k have separation $2\delta_k e_k$. Parts (i) and the bound $I \preceq CD$ give Kullback–Leibler divergence at most $2Cd_k\delta_k^2 \leq \kappa_0$. Pinsker's inequality bounds the corresponding total-variation distance by $\sqrt{\kappa_0/2}$. Averaging the coordinatewise testing errors over neighboring pairs gives (F.5); this is Assouad's lemma in the form of Tsybakov (2009, Chapter 2). Finally, part (iv) follows directly from the definition of $\mathcal{X}_\beta(R)$. For the band construction,

$$\sum_{k=J}^{2J-1}\lambda_k^{2\beta}\delta_k^2 \leq c_0^2(2C_\lambda)^{2\beta}.$$

For the critical construction, $\sum_{k\leq J}k^{2\beta-1} \leq C_\beta J^{2\beta}$, so the weighted norm is at most $c_0^2C_\lambda^{2\beta}C_\beta C_J^{2\beta}$. □

Let $D_{n,m}$ be the positive diagonal operator defined by $D_{n,m}e_k = \iota_k(n,m)e_k$. Theorem 3.3 gives $c_I D_{n,m} \preceq \mathsf{I}_{\mathrm{eff},n,m} \preceq C_I D_{n,m}$. If $\mathsf{S}_{\mathrm{eff}} = U_{\mathrm{eff}}\mathsf{I}_{\mathrm{eff},n,m}^{1/2}$ is the polar decomposition, define

$$\widehat{x}_k = \mathbb{X}_{\mathrm{eff}}\left(U_{\mathrm{eff}}\mathsf{I}_{\mathrm{eff},n,m}^{\dagger/2}e_k\right),\ \widehat{x}_J = \sum_{k\leq J}\widehat{x}_k e_k. \tag{F.6}$$

Douglas's range-inclusion theorem makes every displayed coordinate well defined. Each $\widehat{x}_k$ is unbiased for x_k, although $\mathsf{I}_{\mathrm{eff},n,m}$ need not be diagonal in the basis (e_k).

Proposition F.3 (Off-diagonal cutoff risk)**.** *Under Assumption 3.2,*

$$\sup_{x\in\mathcal{X}_\beta(R)} \mathbb{E}_x^{\mathrm{eff}} \left\|\widehat{x}_J - x\right\|^2 \le C\left[R^2\lambda_{J+1}^{-2\beta} + \sum_{k\le J}\iota_k(n,m)^{-1}\right]. \tag{F.7}$$

Proof of Proposition F.3. Let $D_{n,m}$ be diagonal with entries $\iota_k(n,m)$. Theorem 3.3 gives $c_I D_{n,m} \preceq \mathsf{I}_{\mathrm{eff},n,m} \preceq C_I D_{n,m}$. The estimator in (F.6) is unbiased for $P_J x$, and Lemma F.2(ii) gives

$$\sum_{k\le J}\mathrm{Var}(\widehat{x}_k) = \sum_{k\le J}\left\|\mathsf{I}_{\mathrm{eff},n,m}^{\dagger/2}e_k\right\|^2 \le c_I^{-1}\sum_{k\le J}\iota_k(n,m)^{-1}.$$

The ellipsoid tail satisfies $\left\|(I-P_J)x\right\|^2 \le R^2\lambda_{J+1}^{-2\beta}$, proving (F.7). □

Remark F.4 (Fixed total budget). Suppose $n+m=N$, and write ρ_m for the calibration term in (3.6). Theorem 3.4 gives risk of order $r_{N-m}\vee\rho_m$. Balancing these terms yields an asymptotically optimal allocation, $\rho_{m^\star}\asymp r_{N-m^\star}$, or $m^\star \asymp m_{\mathrm{crit}}(N-m^\star)$. Fix monotone representatives and define the balancing allocation by equality. Its calibration budget grows with N, sublinearly when $p<2\alpha$ (the boundary $p=-1$ included), so the absolute budget grows while its share tends to zero; at $p=2\alpha$ the two have the same order, and for $p>2\alpha$, $m^\star/N\to 1$. These are allocation scales; the exact finite-budget optimum also depends on the risk constants. Appendix J.5 explores the corresponding acquisition and calibration tradeoff in MRI.

Proof of Theorem 3.4. All constants in this proof may depend on the fixed structural parameters listed in Theorem 3.4, but they are independent of n, m, the cutoff J, and the finite lower-bound blocks.

Algebraic relative calibration.

Assume $g_k \asymp k^\kappa$ and put $p=2\alpha+2\kappa$. By (3.4),

$$\sum_{k\le J}\iota_k(n,m)^{-1} \asymp n^{-1}J^{2\alpha+1} + m^{-1}\sum_{k\le J}k^p. \tag{F.8}$$

The last sum has order J^{p+1} for $p>-1$, $\log(e+J)$ for $p=-1$, and one for $p<-1$. Balancing the bias $J^{-2\beta}$ with the acquisition variance gives $J_n \asymp n^{1/(2\beta+2\alpha+1)}$. If $p>-1$, balancing the calibration variance gives $J_m \asymp m^{1/(2\beta+p+1)}$. Taking $J\asymp J_n\wedge J_m$ in Proposition F.3 proves the upper bound in the first line of (3.6). At $p=-1$, take $J_m \asymp [m/\log(e+m)]^{1/(2\beta)}$; for $p<-1$, set $J_m\asymp m^{1/(2\beta)}$ and take $J\asymp J_n\wedge J_m$.

For the acquisition lower bound, use the hypercube

$$x_\omega = \delta_J\sum_{k=J}^{2J-1}\omega_k e_k,\ \delta_J = c_0 J^{-\beta-1/2},\ \omega_k\in\{-1,1\}. \tag{F.9}$$

Choose $c_0\le R(2C_\lambda)^{-\beta}$, so Lemma F.2(iv) places every vertex in $\mathcal{X}_\beta(R)$. For neighboring vertices, Theorem 3.3 and Lemma F.2(i) give divergence at most $CnJ^{-2\alpha}\delta_J^2$. With J a sufficiently large constant multiple of J_n, this bound is below the fixed κ_0 in Lemma F.2(iii), while $J\delta_J^2\asymp J^{-2\beta}\asymp r_n$ for any fixed multiple.

If $p>-1$, the information upper bound also gives $\langle \mathsf{I}_{\mathrm{eff},n,m}e_k,e_k\rangle \le Cmk^{-p}$. The same construction, with J a sufficiently large constant multiple of $m^{1/(2\beta+p+1)}$, satisfies the neighboring-divergence condition and yields the calibration term. At $p=-1$, let J be a fixed small multiple of $m^{1/(2\beta)}$ and take $\delta_k = c_0(mk)^{-1/2}$ for $k_0\le k\le J$. Lemma F.2(iv) verifies the ellipsoid constraint for a constant c_0 depending only on R,β,C_λ. Since $\langle \mathsf{I}_{\mathrm{eff},n,m}e_k,e_k\rangle\le Cmk$, every neighboring divergence is bounded by a constant multiple of c_0^2, and part (iii) gives $\sum_{k=k_0}^J\delta_k^2\asymp m^{-1}\log m$. If $p<-1$, choose a fixed k_0 and compare 0 with $c_0m^{-1/2}e_{k_0}$. Part (i) and $\langle \mathsf{I}_{\mathrm{eff},n,m}e_{k_0},e_{k_0}\rangle\le Cmk_0^{-p}$ give bounded

divergence for small c_0, while Le Cam's two-point lemma (Tsybakov, 2009, Chapter 2) gives a risk lower bound of order m^{-1}. This proves part (i) of Theorem 3.4.

Severe relative calibration.

Assume $g_k \asymp k^\kappa \exp(\Delta k^\nu)$ with $\Delta, \nu > 0$, and put $p = 2\alpha + 2\kappa$. Let

$$V_\Delta(J) = \sum_{k \le J} k^p \exp(2\Delta k^\nu). \tag{F.10}$$

The largest term gives a lower bound and a polynomial multiple of that term gives an upper bound, so

$$\log V_\Delta(J) = 2\Delta J^\nu + O(\log J). \tag{F.11}$$

Proposition F.3 gives risk at most $C[J^{-2\beta} + n^{-1}J^{2\alpha+1} + m^{-1}V_\Delta(J)]$. Let J_m be the largest integer satisfying $m^{-1}V_\Delta(J_m) \le R^2 J_m^{-2\beta}$. Then $J_m \asymp [\log(e+m)]^{1/\nu}$, and $J \asymp J_n \wedge J_m$ gives the upper bound.

For the calibration lower bound, use $x_0 = 0$ and $x_1 = c_1 J^{-\beta} e_J$, with $c_1 \le R C_\lambda^{-\beta}$ so that both points lie in $\mathcal{X}_\beta(R)$. Lemma F.2(i), together with the information upper bound, gives

$$\mathrm{KL}(x_0, x_1) \le C m J^{-p-2\beta} \exp(-2\Delta J^\nu). \tag{F.12}$$

Choose the smallest J for which the right-hand side is below a fixed constant. Then $J \asymp [\log(e + m)]^{1/\nu}$, and Le Cam's lemma yields the second term in (3.7). The acquisition lower bound is the band construction above, which completes part (ii).

By (3.8) and the profile comparisons for a_k and g_k, the defining inequality for $m_{\mathrm{crit}}(n)$ is equivalent, up to fixed constants, to $m^{-1}V_\Delta(J_n) \lesssim J_n^{-2\beta}$. Therefore

$$\log m_{\mathrm{crit}}(n) = 2\Delta J_n^\nu + O(\log J_n). \tag{F.13}$$

Smoothing relative calibration.

If $g_k \asymp k^\kappa \exp(-\delta k^\nu)$ with $\delta, \nu > 0$, then $\sum_{k \ge 1} k^{2\alpha} g_k^2 < \infty$. Proposition F.3, with $J = J_n$, gives the upper bound $r_n + m^{-1}$. For a fixed k_0, compare 0 with $c_0 m^{-1/2} e_{k_0}$. The point lies in $\mathcal{X}_\beta(R)$ for all sufficiently large m, and Lemma F.2(i) together with $\langle \mathsf{I}_{\mathrm{eff},n,m} e_{k_0}, e_{k_0} \rangle \le C m a_{k_0}^2 g_{k_0}^{-2}$ gives bounded divergence. Le Cam's lemma yields cm^{-1}, while the acquisition band construction yields r_n. This proves part (iii).

□

F.2 Calibration budgets

Let $J_n \asymp n^{1/(2\beta+2\alpha+1)}$ be the acquisition oracle cutoff. With V_{cal} and m_{crit} defined in (3.8), the algebraic estimate (F.8) gives

$$m_{\mathrm{crit}}(n) \asymp \begin{cases} n^{\frac{2\beta+p+1}{2\beta+2\alpha+1}}, & p > -1, \\ \text{the solution of } m/\log m \asymp n^{\frac{2\beta}{2\beta+2\alpha+1}}, & p = -1, \\ n^{\frac{2\beta}{2\beta+2\alpha+1}}, & p < -1. \end{cases} \tag{F.14}$$

The last line is parametric because the transferred calibration variance is summable. In the severe regime, (F.13) instead gives a budget exponential in the oracle resolution.

G Exact sequence reduction and parabolic timing

G.1 Product experiment and exact diagonal reduction

Let (a_k) be a bounded real sequence, let (c_k) and (d_k) be real sequences, and assume $a_k d_k \ne 0$ for every k. Define the nuisance Hilbert space

$$\boldsymbol{\Theta}_{c,d} = \left\{ \eta : \sum_{k \ge 1} (c_k^2 + d_k^2)\eta_k^2 < \infty \right\}, \ \|\eta\|^2_{\boldsymbol{\Theta}_{c,d}} = \sum_{k \ge 1} (c_k^2 + d_k^2)\eta_k^2. \tag{G.1}$$

The score maps $\eta \mapsto (c_k\eta_k)$ and $\eta \mapsto (d_k\eta_k)$ are bounded from $\mathbf{\Theta}_{c,d}$ to ℓ^2. On $\Omega = \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$ with the product Borel σ-field, let

$$P_{x,\eta}^{n,m} = \bigotimes_{k\geq 1} N(a_k x_k + c_k\eta_k, n^{-1}) \otimes N(d_k\eta_k, m^{-1}). \tag{G.2}$$

Equivalently,

$$Y_k = a_k x_k + c_k\eta_k + n^{-1/2}\varepsilon_k,\ Z_k = d_k\eta_k + m^{-1/2}\xi_k, \tag{G.3}$$

with independent standard Gaussian noises. Set $g_k = |c_k/d_k|$ and $\sigma_k^2 = n^{-1} + m^{-1}g_k^2$. The efficient diagonal experiment is

$$\widetilde{Y}_k = a_k x_k + \sigma_k\widetilde{\varepsilon}_k,\ k \geq 1, \tag{G.4}$$

where $(\widetilde{\varepsilon}_k)$ are independent standard Gaussian variables. For a nonempty compact convex $K \subseteq \ell^2$, let $\mathcal{R}_{n,m}^{\mathrm{seq}}(K)$ and $\mathcal{R}_{n,m}^{\mathrm{diag}}(K)$ denote the minimax squared-error risks in (G.2) and (G.4), respectively.

Proposition G.1 (Exact diagonal reduction and rate realization)**.** *For every nonempty compact convex $K \subseteq \ell^2$,*

$$\mathcal{R}_{n,m}^{\mathrm{seq}}(K) = \mathcal{R}_{n,m}^{\mathrm{diag}}(K). \tag{G.5}$$

If $|a_k| \asymp k^{-\alpha}$ and $|c_k/d_k| \asymp g_k$, where (g_k) satisfies one of the three profiles in Theorem 3.4, then

$$\mathcal{R}_{n,m}^{\mathrm{seq}}(\mathcal{X}_\beta(R)) \asymp \mathcal{R}_{n,m}(\beta, R). \tag{G.6}$$

For two object alternatives x and x' such that $\Delta x = x - x'$ has finite support, minimize the joint Kullback–Leibler divergence over nuisance differences $\Delta\eta = \eta - \eta'$ supported on the same finite set. The minimizing nuisance difference is given coordinatewise by

$$\Delta\eta_k^\star = -\frac{na_kc_k}{nc_k^2 + md_k^2}\Delta x_k. \tag{G.7}$$

With this choice, the contribution of coordinate k to the minimized divergence is $\frac{1}{2}\iota_k(n,m)(\Delta x_k)^2$. Fix a reference pair (x^0, η^0). For a family whose differences are supported on a common finite set, the affine assignment

$$\eta_k^\star(x) = \eta_k^0 - \frac{na_kc_k}{nc_k^2 + md_k^2}(x_k - x_k^0)$$

belongs to $\mathbf{\Theta}_{c,d}$ and realizes the minimizing nuisance difference for every pair in the family.

Proof. Whiten each coordinate pair and rotate it into the direction orthogonal to the nuisance vector $(\sqrt{n}c_k, \sqrt{m}d_k)$. The orthogonal coordinate is a deterministic rescaling of

$$Y_k - (c_k/d_k)Z_k = a_k x_k + n^{-1/2}\varepsilon_k - (c_k/d_k)m^{-1/2}\xi_k, \tag{G.8}$$

and hence has the law in (G.4). Finite-support nuisance sequences are dense in the nuisance score subspace. The closure-stability and Gaussian translation reduction in Lemmas D.6 and D.7 therefore give the exact equality (G.5).

The cutoff estimator $\widehat{x}_J = \sum_{k\leq J} a_k^{-1}\widetilde{Y}_k e_k$ satisfies

$$\mathbb{E}_x\|\widehat{x}_J - x\|^2 = \sum_{k>J} x_k^2 + \sum_{k\leq J}\left[\frac{1}{na_k^2} + \frac{|c_k/d_k|^2}{ma_k^2}\right]. \tag{G.9}$$

The calculations in Appendix F give the upper bounds. For two alternatives, the contribution of coordinate k to their joint Kullback–Leibler divergence is

$$\frac{1}{2}\left[n(a_k\Delta x_k + c_k\Delta\eta_k)^2 + md_k^2(\Delta\eta_k)^2\right].$$

Minimizing this quadratic function over $\Delta\eta_k$ gives (G.7), and substitution gives $\frac{1}{2}\iota_k(n,m)(\Delta x_k)^2$. For a family supported on a common finite set, the affine assignment stated in the proposition belongs to $\mathbf{\Theta}_{c,d}$ and realizes these minimizing differences simultaneously. The Assouad and Le Cam constructions from Appendix F then prove (G.6). □

G.2 INITIAL-STATE PARABOLIC OPERATOR

The score maps in (4.1) are bounded under the conditions stated before that display. Their coordinate multipliers are

$$(\mathsf{B}_{t_a} b)_k = k^{-s_Y} e^{-t_a k^\nu} b_k,\ (\mathsf{C}_\tau b)_k = k^{-s_Z} e^{-\tau k^\nu} b_k.$$

Every calibration multiplier is nonzero, so C_τ is injective with dense range. The quotient on $\operatorname{ran} \mathsf{C}_\tau$ has multiplier $r_k = k^\kappa e^{(\tau - t_a)k^\nu}$. Its closure is the maximal multiplication operator in (4.2), with domain

$$\operatorname{dom} \mathsf{R}_\tau = \left\{ z = \sum_{k \ge 1} z_k e_k : \sum_{k \ge 1} k^{2\kappa} e^{2(\tau - t_a)k^\nu} |z_k|^2 < \infty \right\}. \tag{G.10}$$

Finite-support vectors form a graph core, so the paired relation is a graph and has zero multivalued part.

Proof of Theorem 4.1. The preceding calculation gives $g_k = k^\kappa e^{(\tau - t_a)k^\nu}$. The three lines of (4.3) follow from Theorem 3.4. On compact intervals strictly before or after t_a, the smoothing or severe profile constants are uniform because $|\tau - t_a|$ is bounded away from zero.

If $\tau_1 \le \tau_2$, then $k^{2\kappa} e^{2(\tau_1 - t_a)k^\nu} \le k^{2\kappa} e^{2(\tau_2 - t_a)k^\nu}$. Hence $\mathsf{R}_{\tau_1} \mathsf{R}_{\tau_1}^* \preceq \mathsf{R}_{\tau_2} \mathsf{R}_{\tau_2}^*$ in form order, and Theorem 2.1 gives information monotonicity. The critical budget follows from (F.13) with $\Delta = \tau - t_a$. □

Proposition G.2 (Uniformity across the near-critical window)**.** *Let $J_0 = J_0(n, m)$ be an integer cutoff with $J_0 \asymp J_n \wedge J_m$, where J_n and J_m are the on-time acquisition and calibration cutoffs in the proof of Theorem 3.4, and write $\mathcal{R}_{n,m}(\beta, R; \Delta)$ for the minimax risk of the model* (4.1) *at calibration offset $\Delta = \tau - t_a$. For every $c > 0$ there are constants $0 < c_\star \le C_\star < \infty$, depending only on c and on the fixed structural parameters, such that*

$$c_\star\, \mathcal{R}_{n,m}(\beta, R; 0) \le \mathcal{R}_{n,m}(\beta, R; \Delta) \le C_\star\, \mathcal{R}_{n,m}(\beta, R; 0) \tag{G.11}$$

for all sufficiently large n, m and all admissible Δ with $|\Delta|\, J_0^\nu \le c$. The constants can be chosen as fixed structural multiples of $e^{-2^{\nu+1}c}$ and $e^{2^{\nu+1}c}$, respectively.

Proof of Proposition G.2. Write $g_{\Delta,k} = k^\kappa e^{\Delta k^\nu}$ and, by (3.4),

$$\iota_k(n, m; \Delta)^{-1} = \frac{1}{n a_k^2} + e^{2\Delta k^\nu}\, \frac{g_{0,k}^2}{m a_k^2}. \tag{G.12}$$

Only the second term depends on Δ, and it does so through a positive factor. Set $\theta = e^{2^{\nu+1}c}$. For every mode $k \le 2J_0$ the hypothesis $|\Delta| J_0^\nu \le c$ gives $2|\Delta| k^\nu \le 2^{\nu+1} |\Delta| J_0^\nu \le 2^{\nu+1} c$, hence

$$\theta^{-1}\, \iota_k(n, m; 0)^{-1} \le \iota_k(n, m; \Delta)^{-1} \le \theta\, \iota_k(n, m; 0)^{-1}, \qquad k \le 2J_0. \tag{G.13}$$

Consider first $\Delta > 0$. Information monotonicity in τ, already proved above, gives $\mathcal{R}_{n,m}(\beta, R; \Delta) \ge \mathcal{R}_{n,m}(\beta, R; 0)$, which is the lower bound in (G.11) with $c_\star = 1$. For the upper bound, apply Proposition F.3 at the cutoff J_0: the bias term $R^2 \lambda_{J_0+1}^{-2\beta}$ does not involve Δ, and by the right inequality in (G.13) the variance term satisfies $\sum_{k \le J_0} \iota_k(n, m; \Delta)^{-1} \le \theta \sum_{k \le J_0} \iota_k(n, m; 0)^{-1}$. Hence the cutoff risk at Δ is at most θ times the cutoff risk at $\Delta = 0$, which by the choice of J_0 is of the order of $\mathcal{R}_{n,m}(\beta, R; 0)$.

Now let $\Delta < 0$. Monotonicity gives the upper bound with $C_\star = 1$. The acquisition lower bound r_n holds at every offset because $\iota_k(n, m; \Delta) \le n a_k^2$. This proves the required lower bound when acquisition dominates the on-time rate. When calibration dominates, $J_0 \asymp J_m$. For $p > -1$, use a band of length proportional to J_0 below $2J_0$, with separations $c_0 J_0^{-\beta - 1/2}$. For $p = -1$, use $\delta_k = c_0 (mk)^{-1/2}$ for $k \le J_0$; the ellipsoid constraint holds and $\log J_0 \asymp \log m$. For $p < -1$, use a fixed coordinate with separation $c_0 m^{-1/2}$. The information comparison (G.13) allows the same constructions after multiplying each separation by $\theta^{-1/2}$. The Assouad and Le Cam bounds then yield the on-time calibration rate multiplied by θ^{-1}.

In both cases the constants are $\theta^{\pm 1}$ times the fixed constants of Theorem 3.4, and θ depends only on c and ν. □

G.3 MULTIPLE CALIBRATION CHANNELS

Multiple calibration channels fit the same framework. With channels $\mathsf{C}_1, \dots, \mathsf{C}_k$ at information levels $m_1, \dots, m_k$, Proposition D.3 gives

$$\langle \mathsf{I}_{\mathrm{eff},n,m} v, v\rangle = \inf_{\theta\in\mathbf{\Theta}} \Big[n\|\mathsf{A}v + \mathsf{B}\theta\|_{\mathcal{Y}}^2 + \sum_{j=1}^{k} m_j \|\mathsf{C}_j\theta\|^2 \Big], \tag{G.14}$$

For fixed A and B, the calibration channels enter through the weighted Gram $\sum_j m_j \mathsf{C}_j^*\mathsf{C}_j$. Increasing this Gram in quadratic-form order increases the efficient information. Gram order therefore gives a sufficient design comparison; directions irrelevant to the target can make the converse fail, as illustrated below.

For parabolic channels the induced sandwich is explicit.

Proposition G.3 (Multiple calibration channels and the earliest-time rate)**.** *In the setting of Theorem 4.1, let the calibration channels be C_{τ_j} of* (4.1) *at times $\tau_1 \le \cdots \le \tau_k$ with information levels $m_1, \dots, m_k$, and put $m_{\mathrm{tot}} = \sum_j m_j$. Write $\mathcal{R}_{n,m}(\beta, R; \tau)$ for the single-channel risk of Theorem 4.1. Then*

$$\mathcal{R}_{n,m_{\mathrm{tot}}}(\beta, R; \tau_1) \;\lesssim\; \mathcal{R}_n^{\mathrm{multi}}(\beta, R) \;\lesssim\; \mathcal{R}_{n,m_1}(\beta, R; \tau_1), \tag{G.15}$$

with the comparison constants of Theorem 3.4. If the times are fixed and $m_1 \ge c m_{\mathrm{tot}}$ for a fixed $c > 0$, the two ends of (G.15) *agree up to constants. The multi-channel risk then has the single-channel rate at $(\tau_1, m_{\mathrm{tot}})$. For arbitrary budget allocations, the sandwich holds with the distinct budgets displayed.*

At a fixed total calibration budget, assigning all information to the earliest available time maximizes the weighted Gram in this model. A fixed finite collection of late channels remains in the severe calibration regime; an early channel receiving a fixed positive budget share gives the smoothing regime. The distinction governs total recovery risk when calibration is the limiting source of error.

Proof of Proposition G.3. The variational form (G.14) is Proposition D.3 applied to the nuisance score $\theta \mapsto (\sqrt{n}\mathsf{B}\theta, \sqrt{m_1}\mathsf{C}_{\tau_1}\theta, \dots, \sqrt{m_k}\mathsf{C}_{\tau_k}\theta)$, whose nuisance-nuisance Fisher block is $n\mathsf{B}^*\mathsf{B} + \sum_j m_j \mathsf{C}_{\tau_j}^*\mathsf{C}_{\tau_j}$. Monotonicity in the weighted Gram is immediate from that form, since the objective is increasing in $\sum_j m_j \langle \mathsf{C}_j^*\mathsf{C}_j\theta, \theta\rangle$ pointwise in θ.

The channels of (4.1) are spectral multipliers, $(\mathsf{C}_{\tau_j} b)_k = k^{-s_Z} e^{-\tau_j k^\nu} b_k$, so

$$\sum_j m_j \mathsf{C}_{\tau_j}^*\mathsf{C}_{\tau_j} \quad \text{has multiplier} \quad k^{-2s_Z} \sum_j m_j e^{-2\tau_j k^\nu}.$$

Since $\tau_1 = \min_j \tau_j$, every exponential obeys $e^{-2\tau_j k^\nu} \le e^{-2\tau_1 k^\nu}$, so

$$m_1 e^{-2\tau_1 k^\nu} \le \sum_j m_j e^{-2\tau_j k^\nu} \le m_{\mathrm{tot}} e^{-2\tau_1 k^\nu} \qquad \text{for every } k.$$

Substituting into (3.4) shows that the efficient per-mode variance of the multi-channel experiment lies, mode by mode, between the one of the single channel at $(\tau_1, m_{\mathrm{tot}})$ and the one at (τ_1, m_1). The bias term is common to all three. Both the cutoff upper bound of Proposition F.3 and the hypercube and two-point lower bounds of the proof of Theorem 3.4 depend on the experiment only through those per-mode variances and the corresponding information upper bound, so both transfer with the same constants, which gives (G.15). If $m_1 \ge c m_{\mathrm{tot}}$ the two ends differ by the fixed factor c^{-1} in the calibration term, hence by a constant in the risk. □

Scope of the design comparisons. Gram order is sufficient for target-information order, but it is not necessary. For $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, $\mathbf{\Theta} = \mathbb{R}^2$, $\mathsf{A} = 1$ and $\mathsf{B} = (1, 0)$, the calibration Grams $H_1 = \mathrm{diag}(2, 1)$ and $H_2 = \mathrm{diag}(1, 2)$ are unordered, while their efficient information values satisfy $2n/(n+2) > n/(n+1)$. The second nuisance coordinate does not affect acquisition.

The earliest-budget condition also has statistical content. Take $\alpha = 0$, $\beta = 1$, $\kappa = 0$, $\nu = 1$, $\tau_1 = t_a$ and $\tau_2 = t_a + 1$, with $n = M$, $m_1 = \log M$, $m_2 = M$. The effective calibration precision on

mode k is $h_k = \log M + Me^{-2k}$, so $\iota_k = Mh_k/(M + h_k) \le h_k$. At $k = \lceil \log M \rceil$, the two alternatives 0 and $k^{-1}e_k$ lie in $\mathcal{X}_1(1)$ and have Kullback–Leibler divergence tending to zero. Thus the multi-channel risk is at least of order $(\log M)^{-2}$, whereas putting all calibration information at τ_1 gives order $M^{-2/3}$. A vanishing earliest-budget share can therefore change the rate in the total budget.

G.4 Noncommuting variable-coefficient evolution, $\nu \in (0, 2]$

Proposition G.4 (Variable-coefficient evolution with frequency-localized sensor mixing)**.** *Fix $T > 0$ and $\nu \in (0, 2]$, let $D = (0, 1)$, and let*

$$Gu = -\partial_\xi(\mathfrak{a}(\xi)\partial_\xi u) + q(\xi)u + u,\ u|_{\partial D} = 0, \tag{G.16}$$

where $\mathfrak{a} \in C^2(\overline{D})$ is uniformly positive and $q \in L^\infty(D)$ is nonnegative. Set $\ell_\mathfrak{a} = \int_0^1 \mathfrak{a}(\xi)^{-1/2}\, d\xi$ and $c_\mathfrak{a} = \pi/\ell_\mathfrak{a}$. Let $L = G^{1/2}$, with eigenpairs (ϕ_k, λ_k) ordered increasingly. Then

$$\lambda_k = c_\mathfrak{a} k + O(k^{-1}),\ \lambda_k^2 = c_\mathfrak{a}^2 k^2 + O(1). \tag{G.17}$$

Let $\{(p_j, q_j)\}_{j\ge1}$ partition the mode indices into disjoint pairs and let Q rotate each (ϕ_{p_j}, ϕ_{q_j}) through an angle ω_j satisfying

$$0 < |\omega_j| < \frac{\pi}{4},\ \sup_{j\ge1} e^{T\delta_j^{(\nu)}} |\sin\omega_j| < \infty, \qquad \delta_j^{(\nu)} = \left|\lambda_{q_j}^\nu - \lambda_{p_j}^\nu\right|. \tag{G.18}$$

For the pairing $q_j = p_j + d$ at a fixed offset $d \ge 1$, $\delta_j^{(\nu)} \asymp p_j^{\nu-1}$, bounded for $\nu \le 1$ and divergent for $\nu > 1$. Fix $\alpha \ge 0$, $s_Y, s_Z \in \mathbb{R}$, $t_a \in [0, T]$, and set $\kappa = s_Z - s_Y$. Let $I \subseteq [0, T]$ be compact; assume $t_a > 0$ or $s_Y \ge 0$, and if $s_Z < 0$, assume $\inf I > 0$. Define

$$\mathsf{A} = L^{-\alpha},\ \mathsf{B}_{t_a} = QL^{-s_Y}e^{-t_aL^\nu},\ \mathsf{C}_\tau = L^{-s_Z}e^{-\tau L^\nu}. \tag{G.19}$$

For every $\tau \in I$, the paired relation is the graph of $\mathsf{R}_\tau = QL^\kappa e^{(\tau-t_a)L^\nu}$. Assumption 3.2 holds with $g_{\tau,k} = \lambda_k^\kappa e^{(\tau-t_a)\lambda_k^\nu}$ and constants uniform for $\tau \in I$. Also,

$$g_{\tau,k} \asymp k^\kappa \exp\left[(\tau - t_a)c_\mathfrak{a}^\nu k^\nu\right] \tag{G.20}$$

uniformly for $\tau \in I$, the restriction $\nu \le 2$ entering here through $\lambda_k^\nu - c_\mathfrak{a}^\nu k^\nu = O(k^{\nu-2})$. If, for some j,

$$(\lambda_{p_j}^{-2\alpha} - \lambda_{q_j}^{-2\alpha})(g_{\tau,p_j}^2 - g_{\tau,q_j}^2)\sin\omega_j\cos\omega_j \ne 0, \tag{G.21}$$

then $[\mathsf{R}_\tau\mathsf{R}_\tau^*, \mathsf{A}\mathsf{A}^*] \ne 0$. The minimax risk has the corresponding order in (4.3)*, with exponent ν and effective timing constant $c_\mathfrak{a}^\nu|\tau - t_a|$, and the efficient information remains nonincreasing in τ.*

The strength of mixing the angle condition permits is governed by ν alone, because $\delta_j^{(\nu)} \asymp j^{\nu-1}$:

- *(i) For $\nu \le 1$, and in particular for the semigroup e^{-tL} of Theorem 4.2, the gaps of any fixed-offset pairing are bounded, so* (G.18) *holds for* every *sequence of angles in $(0, \pi/4)$. A fixed rotation of order one, applied to every pair and at any fixed frequency offset, is admissible, and $Q - I$ is then not compact.*
- *(ii) For $\nu > 1$, and in particular for the heat semigroup $\nu = 2$, the gaps grow and* (G.18) *forces $|\sin\omega_j| \lesssim e^{-Tcj^{\nu-1}}$. Taking $\sin\omega_j = \varepsilon e^{-T\delta_j^{(\nu)}}$ with $\varepsilon > 0$ small makes $Q - I$ Hilbert–Schmidt and keeps the commutator nonzero whenever* (G.21) *holds.*

Proof of Proposition G.4. The operator G in (G.16) is positive self-adjoint with compact inverse. Define

$$s(\xi) = \int_0^\xi \mathfrak{a}(r)^{-1/2}\, dr,\ 0 \le s \le \ell_\mathfrak{a}, \tag{G.22}$$

and let $\xi = \xi(s)$ denote the inverse change of variables. The Liouville transformation

$$(Uu)(s) = \mathfrak{a}(\xi(s))^{1/4}u(\xi(s)) \tag{G.23}$$

is unitary from $L^2(0,1;d\xi)$ onto $L^2(0,\ell_{\mathfrak{a}};ds)$ because $ds=\mathfrak{a}(\xi)^{-1/2}d\xi$. A direct calculation gives

$$UGU^{-1}=-\frac{d^2}{ds^2}+V(s), \tag{G.24}$$

where

$$V(s)=q(\xi(s))+1+\frac{1}{4}\mathfrak{a}''(\xi(s))-\frac{[\mathfrak{a}'(\xi(s))]^2}{16\mathfrak{a}(\xi(s))}. \tag{G.25}$$

The assumptions $\mathfrak{a}\in C^2(\overline{D})$, $\inf\mathfrak{a}>0$, and $q\in L^\infty(D)$ imply $V\in L^\infty(0,\ell_{\mathfrak{a}})$. Because $\mathfrak{a}^{1/4}$ is continuous and strictly positive on $\overline{D}$, the transformation preserves zero boundary traces. The boundary condition $u(0)=u(1)=0$ becomes the Dirichlet condition at $s=0$ and $s=\ell_{\mathfrak{a}}$; see Zettl (2005).

Let μ_k denote the eigenvalues of G. Unitary equivalence identifies them with the Dirichlet eigenvalues of $-d^2/ds^2+V$ on $(0,\ell_{\mathfrak{a}})$. After the rescaling $s=\ell_{\mathfrak{a}}t$, the normalized Dirichlet problem on $(0,1)$ has potential $\ell_{\mathfrak{a}}^2V(\ell_{\mathfrak{a}}t)\in L^\infty(0,1)$ and eigenvalues $\ell_{\mathfrak{a}}^2\mu_k$. The standard Dirichlet Sturm–Liouville asymptotics therefore give

$$\ell_{\mathfrak{a}}\sqrt{\mu_k}=\pi k+O(k^{-1}); \tag{G.26}$$

see Hryniv & Mykytyuk (2006, Theorem 1.1). In the notation of Proposition G.4, $L=G^{1/2}$ and $L\phi_k=\lambda_k\phi_k$, so $\mu_k=\lambda_k^2$. Therefore $\lambda_k=c_{\mathfrak{a}}k+O(k^{-1})$, and squaring gives the second relation in (G.17). Since $\lambda_k=c_{\mathfrak{a}}k+O(k^{-1})$, raising to the power ν gives $\lambda_k^\nu=c_{\mathfrak{a}}^\nu k^\nu+O(k^{\nu-2})$, which is bounded exactly when $\nu\le 2$. Consequently,

$$\delta_j^{(\nu)}=\left|\lambda_{q_j}^\nu-\lambda_{p_j}^\nu\right|=c_{\mathfrak{a}}^\nu\left|q_j^\nu-p_j^\nu\right|+O(p_j^{\nu-2}), \tag{G.27}$$

so for the pairing $q_j=p_j+d$ the gaps are of order $p_j^{\nu-1}$: they tend to $dc_{\mathfrak{a}}$ when $\nu=1$ and grow like $2dc_{\mathfrak{a}}^2p_j$ when $\nu=2$. The same expansion gives, uniformly for $\tau,t_a\in[0,T]$,

$$\lambda_k^\kappa e^{(\tau-t_a)\lambda_k^\nu}\asymp k^\kappa e^{(\tau-t_a)c_{\mathfrak{a}}^\nu k^\nu}, \tag{G.28}$$

because λ_k/k stays bounded above and below and $\lambda_k^\nu-c_{\mathfrak{a}}^\nu k^\nu=O(1)$. This proves (G.20) and places the model in the smoothing, algebraic, or severe profile of Theorem 3.4 according as $\tau<t_a$, $\tau=t_a$, or $\tau>t_a$.

For every $\theta\in L^2(D)$, the identity $\kappa=s_Z-s_Y$ and the spectral calculus give

$$\mathsf{B}_{t_a}\theta=QL^{-s_Y}e^{-t_aL^\nu}\theta=QL^\kappa e^{(\tau-t_a)L^\nu}L^{-s_Z}e^{-\tau L^\nu}\theta=Qg_\tau(L)\mathsf{C}_\tau\theta. \tag{G.29}$$

Hence the unclosed paired score relation is

$$\{(z,Qg_\tau(L)z):z\in\operatorname{ran}\mathsf{C}_\tau\}. \tag{G.30}$$

Every spectral multiplier $\lambda_k^{-s_Z}e^{-\tau\lambda_k^\nu}$ of C_τ is nonzero. Therefore the algebraic spectral span $\mathcal{D}_{\mathrm{fin}}=\operatorname{span}\{\phi_k:k\ge1\}$ is contained in $\operatorname{ran}\mathsf{C}_\tau$. For $z=\sum_{k\ge1}z_k\phi_k\in\operatorname{dom}g_\tau(L)$, define $z^{(N)}=\sum_{k\le N}z_k\phi_k$. Then

$$\left\|z^{(N)}-z\right\|^2+\left\|g_\tau(L)(z^{(N)}-z)\right\|^2=\sum_{k>N}\left(1+g_{\tau,k}^2\right)|z_k|^2\longrightarrow0. \tag{G.31}$$

The space $\mathcal{D}_{\mathrm{fin}}$ is a graph core for the maximal spectral multiplier $g_\tau(L)$. Since Q is unitary, $\|Qg_\tau(L)z\|=\|g_\tau(L)z\|$, so the same space is a graph core for $Qg_\tau(L)$. Consequently,

$$\overline{\{(\mathsf{C}_\tau\theta,\mathsf{B}_{t_a}\theta):\theta\in L^2(D)\}}^{L^2(D)\oplus L^2(D)}=\operatorname{gra}(Qg_\tau(L)). \tag{G.32}$$

Hence the paired relation is the graph of $\mathsf{R}_\tau=Qg_\tau(L)=QL^\kappa e^{(\tau-t_a)L^\nu}$ and has zero multivalued part.

For $g_{\tau,k}=\lambda_k^\kappa e^{(\tau-t_a)\lambda_k^\nu}$, the bounded ratio of adjacent eigenvalues and (G.27) imply

$$\max\left\{\frac{g_{\tau,q_j}}{g_{\tau,p_j}},\frac{g_{\tau,p_j}}{g_{\tau,q_j}}\right\}\le C_\kappa e^{T\delta_j^{(\nu)}}. \tag{G.33}$$

On the jth block, let $W_{\tau,j} = \mathrm{diag}(g_{\tau,p_j}, g_{\tau,q_j})$; the argument is blockwise and never uses that the paired modes are adjacent. The matrix $W_{\tau,j} Q_j^* W_{\tau,j}^{-1}$ has diagonal entries $\cos\omega_j$ and off-diagonal entries bounded by a constant multiple of $|\sin\omega_j| e^{T\delta_j^{(\nu)}}$. The angle condition in Proposition G.4 therefore gives uniform upper and lower bounds for this matrix and its inverse. Summing over blocks yields

$$c\|g_\tau(L)y\|^2 \le \|g_\tau(L)Q^*y\|^2 \le C\|g_\tau(L)y\|^2, \tag{G.34}$$

uniformly for $\tau \in I$, and Q^* preserves $\mathrm{dom}\, g_\tau(L)$. Since $\mathsf{R}_\tau^* = g_\tau(L)Q^*$, Assumption 3.2 holds.

On the block spanned by ϕ_{p_j}, ϕ_{q_j}, the off-diagonal entry of $\mathsf{R}_\tau \mathsf{R}_\tau^* = Q\,\mathrm{diag}(g_{\tau,p_j}^2, g_{\tau,q_j}^2)Q^*$ is $\sin\omega_j \cos\omega_j (g_{\tau,p_j}^2 - g_{\tau,q_j}^2)$. The condition in (G.21) therefore implies a nonzero commutator with $\mathsf{A}\mathsf{A}^* = L^{-2\alpha}$.

The minimax rates follow from Theorems 3.4 and 4.1. Pointwise order of $g_\tau(L)^2$ in τ, followed by conjugation with Q, gives information monotonicity. Finally, for $\nu \le 1$ the gaps $\delta_j^{(\nu)}$ are bounded, so (G.18) holds for every angle sequence and no decay is imposed; a constant angle then gives $\sum_j |\sin\omega_j|^2 = \infty$, so $Q - I$ is not compact. For $\nu > 1$, if $\sin\omega_j = \varepsilon e^{-T\delta_j^{(\nu)}}$, then (G.27) gives $\sum_j |\sin\omega_j|^2 < \infty$. On each rotation block, $\|Q_j - I\|_{\mathrm{HS}}^2 = 4(1-\cos\omega_j) \lesssim |\sin\omega_j|^2$. Summation over the blocks therefore gives $Q - I \in \mathcal{S}_2$, so $Q - I$ is Hilbert–Schmidt. □

H Bounded initial states and approximate reduction

H.1 Proof of Theorem 4.2

Write $d = 2\beta + 2\alpha + 1$. All comparison constants in this proof may depend on the fixed parameters in the theorem. Budgets are sufficiently large; no relation between their rates of growth is imposed.

Upper bound. For integers $0 \le K \le J$, define

$$\widehat{x}_k = \begin{cases} k^\alpha (Y_k - e^{(\tau - t_a)k} Z_k), & k \le K, \\ k^\alpha Y_k, & K < k \le J, \\ 0, & k > J. \end{cases} \tag{H.1}$$

The target truncation bias and the remaining nuisance bias have disjoint coordinate supports. Independence of the noises gives the exact worst-case risk of this rule:

$$\begin{aligned} \sup_{x,\eta} \mathbb{E}\|\widehat{x} - x\|^2 = R^2 (J+1)^{-2\beta} + \frac{1}{n}\sum_{k=1}^{J} k^{2\alpha} + \frac{1}{m}\sum_{k=1}^{K} k^{2\alpha} e^{2(\tau - t_a)k} \\ + \rho^2 \max_{K < k \le J} k^{2\alpha} e^{-2t_a k}, \end{aligned} \tag{H.2}$$

where a maximum over an empty set is zero. Equality follows by concentrating the target on mode $J+1$ and the nuisance on a maximizing mode of the disjoint middle block. Take $J = \lfloor n^{1/d} \rfloor$; the first two terms are $O(r_n)$.

If $\tau < t_a$, choose $K = J$. The last term vanishes and $\sum_{k \ge 1} k^{2\alpha} e^{-2(t_a - \tau)k} < \infty$, giving the calibration bound $O(m^{-1})$.

If $\tau = t_a$, put $K = J \wedge \lceil (\log m)/(2t_a) \rceil$. The calibration variance is at most $C m^{-1} (\log m)^{2\alpha+1}$. When $K < J$, the profile $k^{2\alpha} e^{-2t_a k}$ is decreasing for sufficiently large K and its tail maximum is at most $C m^{-1} (\log m)^{2\alpha}$; when $K = J$, the tail is absent. This proves the on-time upper bound.

If $\tau > t_a$, put $\Delta = \tau - t_a > 0$ and $K = J \wedge \lceil (\log m)/(2\tau) \rceil$. The geometric-series bound

$$\sum_{k=1}^{K} k^{2\alpha} e^{2\Delta k} \le C K^{2\alpha} e^{2\Delta K}$$

shows that the calibration variance is at most $C m^{-t_a/\tau} (\log m)^{2\alpha}$. If $K < J$, the nuisance tail has the same upper bound; otherwise it is zero. Integer rounding multiplies the exponential terms by fixed constants because the exponent is linear in k.

Acquisition lower bound. Set $\eta = 0$. On the block $J \le k < 2J$, use $x_k = \omega_k c n^{-1/2} k^\alpha$, with independent signs $\omega_k \in \{-1, 1\}$ and $J = \lfloor c_1 n^{1/d} \rfloor$. Fixed sufficiently small positive c, c_1 ensure $\sum k^{2\beta} x_k^2 \le R^2$. Neighboring sign vectors have joint Kullback–Leibler divergence $2c^2$, since calibration carries no target mean. Coordinatewise binary testing, or Assouad's lemma, yields a lower bound of order $n^{-1} J^{2\alpha+1} \asymp r_n$.

Calibration lower bounds. All following families satisfy

$$x_k = -k^\alpha e^{-t_a k} \eta_k. \tag{H.3}$$

Their acquisition means are identically zero, so their testing distances depend only on calibration, independently of n.

For $\tau < t_a$, use only mode 1 and let $\eta_1 = \pm c m^{-1/2}$. For any fixed sufficiently small $c > 0$, these two points are in both parameter classes for large m, their calibration divergence is $2c^2 e^{-2\tau}$, and their squared target separation is $4c^2 e^{-2t_a}/m$. The two-point bound gives $c' m^{-1}$.

For $\tau = t_a$, set $K = \lfloor (\log m)/(2t_a) \rfloor$ and, for $1 \le k \le K$, let

$$\eta_k = \omega_k c_0 m^{-1/2} e^{t_a k}, \qquad x_k = -\omega_k c_0 m^{-1/2} k^\alpha.$$

Since $m^{-1} \sum_{k \le K} e^{2t_a k} \le e^{2t_a}/(e^{2t_a} - 1)$, a fixed positive c_0 can ensure $\|\eta\| \le \rho$ and a neighboring divergence $2c_0^2 \le 1/8$. Also

$$\sum_{k \le K} k^{2\beta} x_k^2 \le C c_0^2 m^{-1} K^{2\beta+2\alpha+1} \longrightarrow 0,$$

so the whole hypercube eventually belongs to the target ellipsoid. The same coordinatewise testing bound gives $c' m^{-1} \sum_{k \le K} k^{2\alpha} \asymp m^{-1} (\log m)^{2\alpha+1}$.

For $\tau > t_a$, choose $k_m = \lfloor (\log m)/(2\tau) \rfloor$ and a fixed $0 < h \le \min\{\rho, e^{-\tau}/4\}$. Use $\eta_{k_m} = \pm h$ and zero elsewhere, with x defined by (H.3). The calibration divergence is $2mh^2 e^{-2\tau k_m} \le 2h^2 e^{2\tau} \le 1/8$. The target constraint holds eventually because $h^2 k_m^{2\beta+2\alpha} e^{-2t_a k_m} \to 0$. The squared target separation is

$$4h^2 k_m^{2\alpha} e^{-2t_a k_m} \asymp m^{-t_a/\tau} (\log m)^{2\alpha}.$$

The two-point bound proves the last branch. Combining each calibration lower bound with the acquisition lower bound completes the proof. □

H.2 A GROWING NUISANCE RADIUS

Proof of Proposition 4.3. Throughout $\Delta = \tau - t_a > 0$, $\rho_m = m^\gamma$, and comparison constants depend only on $\alpha, \beta, R, t_a, \tau, \gamma$.

Upper bound, $\gamma < \gamma_\star$. Apply (H.5) with $J = \lfloor n^{1/d} \rfloor$ and $K = J \wedge \lfloor \log(m\rho_m^2)/(2\tau) \rfloor$, so that $e^{2\tau K} \asymp m\rho_m^2$. The first two terms are $O(r_n)$. The calibration variance obeys

$$\frac{1}{m} \sum_{k \le K} k^{2\alpha} e^{2\Delta k} \le \frac{K^{2\alpha} e^{2\Delta K}}{m(1 - e^{-2\Delta})} \asymp K^{2\alpha} \frac{(m\rho_m^2)^{\Delta/\tau}}{m},$$

and, since $K \ge \alpha/t_a$ for large m, the uncorrected tail obeys $\rho_m^2 K^{2\alpha} e^{-2t_a K} \asymp \rho_m^2 K^{2\alpha} (m\rho_m^2)^{-t_a/\tau}$. Using $\Delta/\tau = 1 - t_a/\tau$, both are of order $\rho_m^{2-2t_a/\tau} m^{-t_a/\tau} K^{2\alpha}$. With $\rho_m = m^\gamma$ the power of m is $2\gamma - 2\gamma t_a/\tau - t_a/\tau = -q(\gamma)$, and $K \asymp \log m$, giving $m^{-q(\gamma)} (\log m)^{2\alpha}$. The two cutoffs balance the two terms up to a fixed factor: the calibration variance increases and the uncorrected tail decreases in K, so no other choice of K lowers the larger of them by more than a constant.

Lower bound, $\gamma < \gamma_\star$. Use the cancellation family (H.3) at the single mode $k_m = \lfloor \log(m\rho_m^2)/(2\tau) \rfloor$ with $\eta_{k_m} = \pm h$, $h = c_1 \rho_m$ for a fixed small $c_1 > 0$, and zero elsewhere. The acquisition means agree. The calibration divergence is $2mh^2 e^{-2\tau k_m} \asymp c_1^2 m\rho_m^2 (m\rho_m^2)^{-1} = c_1^2$, at most $1/8$ for c_1 small. The target constraint holds because

$$k_m^{2\beta} x_{k_m}^2 = h^2 k_m^{2\beta+2\alpha} e^{-2t_a k_m} \asymp (\log m)^{2\beta+2\alpha} m^{-q(\gamma)},$$

which tends to 0 for $\gamma < \gamma_\star$, since then $q(\gamma) > 0$ and a power beats every logarithm. The squared target separation is $4h^2 k_m^{2\alpha} e^{-2t_a k_m} \asymp m^{-q(\gamma)} (\log m)^{2\alpha}$, and the two-point bound applies. The acquisition lower bound of the proof above holds at every radius and gives r_n.

The case $\gamma > \gamma_\star$. The upper bound is Theorem 4.1, because $\{\|\eta\| \le \rho_m\} \subseteq \ell^2$. For the lower bound, inspect the unrestricted construction. Perturbing mode k by $\delta\eta_k = ce^{\tau k}/\sqrt{m}$ keeps the calibration divergence bounded and, through (H.3), moves the target by $\delta x_k = -ck^\alpha e^{\Delta k}/\sqrt{m}$. The ellipsoid constraint $k^{2\beta}\delta x_k^2 \le R^2$ permits k up to $k^\dagger = (1+o(1))\log m/(2\Delta)$, and at that mode the squared separation is of order $k^{-2\beta} \asymp (\log m)^{-2\beta}$, the rate of Theorem 4.1. The norm this construction requires is

$$\|\eta\| = ce^{\tau k^\dagger}/\sqrt{m} = m^{\tau/(2\Delta)-1/2+o(1)} = m^{t_a/(2\Delta)+o(1)} = m^{\gamma_\star+o(1)}. \tag{H.4}$$

Hence for $\gamma > \gamma_\star$ the construction lies in $\{\|\eta\| \le \rho_m\}$ for large m, and the unrestricted lower bound applies unchanged.

The boundary. At $\gamma = \gamma_\star$ the two expressions in (4.9) agree up to logarithmic factors, both being $(\log m)^{O(1)}$ times a constant; the upper bound of the first case then ceases to tend to zero and the operative bound is the second. We state the result away from this point. □

H.3 The finite-budget form

Corollary H.1 (Explicit finite-budget upper bound)**.** *For every $n, m > 0$ and all integers $0 \le K \le J$, the rule* (H.1) *satisfies*

$$\sup_{x\in\mathcal{X}_\beta(R),\,\|\eta\|\le\rho} \mathbb{E}\|\widehat{x}-x\|^2 = \frac{R^2}{(J+1)^{2\beta}} + \frac{S_\alpha(J)}{n} + \frac{1}{m}\sum_{k\le K} k^{2\alpha}e^{2\Delta k} + \rho^2 \max_{K<k\le J} k^{2\alpha}e^{-2t_a k}, \tag{H.5}$$

an empty maximum being zero.

Proof of Corollary H.1. The identity (H.5) is (H.2), which holds for every n, m and every pair $0 \le K \le J$; no asymptotics enter its derivation. For the displayed bounds, the acquisition sum satisfies $S_\alpha(J) \le J^{2\alpha} + \int_0^J t^{2\alpha}\,dt$. For $\Delta < 0$ and $K = J$ the calibration sum is bounded by the convergent series $c_\Delta(\alpha) = \sum_{k\ge1} k^{2\alpha}e^{-2|\Delta|k}$, and the tail maximum is empty. For $\Delta = 0$ the calibration sum is $S_\alpha(K)$. For $\Delta > 0$, monotonicity of $k \mapsto k^{2\alpha}$ gives

$$\sum_{k\le K} k^{2\alpha}e^{2\Delta k} \le K^{2\alpha}\sum_{k\le K} e^{2\Delta k} \le \frac{K^{2\alpha}e^{2\Delta K}}{1-e^{-2\Delta}}.$$

The profile $k \mapsto k^{2\alpha}e^{-2t_a k}$ decreases for $k \ge \alpha/t_a$, so once $K \ge \alpha/t_a$, which for $K = \lceil(\log m)/(2\tau)\rceil$ holds as soon as $m \ge e^{2\tau\alpha/t_a}$, the tail maximum is attained at $k = K+1$ and is at most $\rho^2 K^{2\alpha}e^{-2t_a K}$. Collecting these gives the following explicit bound,

$$\sup \mathbb{E}\|\widehat{x}-x\|^2 \le \frac{R^2}{(J+1)^{2\beta}} + \frac{S_\alpha(J)}{n} + \begin{cases} m^{-1}\,c_\Delta(\alpha), & \Delta < 0, \\ m^{-1}S_\alpha(K) + \rho^2 K^{2\alpha}e^{-2t_a K}, & \Delta = 0, \\ K^{2\alpha}\big(\theta_\Delta e^{2\Delta K}m^{-1} + \rho^2 e^{-2t_a K}\big), & \Delta > 0, \end{cases} \tag{H.6}$$

where $\theta_\Delta = (1-e^{-2\Delta})^{-1}$, $c_\Delta(\alpha) = \sum_{k\ge1} k^{2\alpha}e^{-2|\Delta|k} < \infty$ and $S_\alpha(J) \le J^{2\alpha} + J^{2\alpha+1}/(2\alpha+1)$. All four quantities are computable from $\alpha, \beta, R, \rho, t_a, \tau, J, K$ alone. □

Scope of the spectral specialization. The theorem uses the semigroup profile e^{-tk}, with fixed $\alpha \ge 0$ and fixed positive radii. Its comparison constants depend on ρ and the observation times. For a classical one-dimensional heat spectrum proportional to k^2, neighboring exponential weights have unbounded ratios, so the rounding step in the late lower bound requires a separate analysis. The noncommuting heat result of Proposition G.4 continues to concern the unrestricted nuisance model. Source-condition minimax theory (Ding & Mathé, 2017) explains why restricting exponentially damped contamination can improve rates; the result here identifies its joint acquisition–calibration risk and the separate cutoffs that attain it.

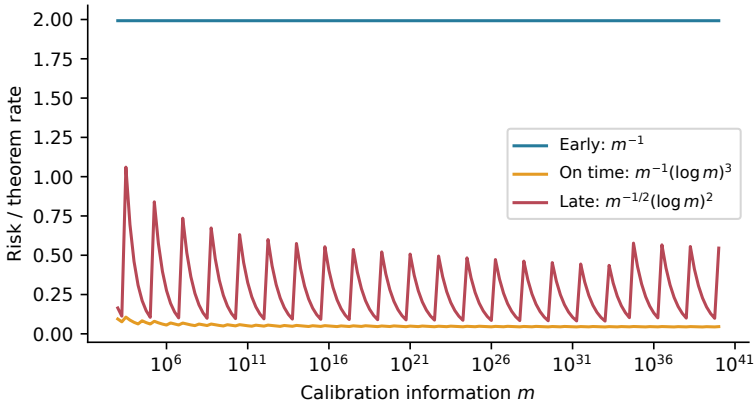

Figure 2: Worst-case two-cutoff risk divided by its calibration rate in Theorem 4.2, with $\alpha = 1$. The bounded oscillation at late times reflects integer cutoff changes.

H.4 Bounded-timing numerical check

For Theorem 4.2, take $\alpha = \beta = R = \rho = t_a = 1$, $\tau \in \{0.5, 1, 2\}$, $n = m^6$, and 149 logarithmically spaced budgets $10^3 \le m \le 10^{40}$. Evaluate the exact worst-case risk (H.2) at the theorem's J, K. The acquisition polynomial sum is analytic, so arbitrarily large J introduces no finite spectral ceiling. The early summable calibration series is evaluated through $\min\{J, 1500\}$; the omitted positive tail is below floating-point precision. The unrestricted late comparator minimizes the single-cutoff risk over $1 \le J \le 250$; every optimum is interior to this search range. The normalized bounded risks stay between fixed positive values at large budgets. Over the large-budget range reported in the tables, the normalized late risk oscillates within a factor of 7.1. The integer cutoff accounts for this pattern: $K = \lceil (\log m)/(2\tau) \rceil$ increases in unit steps, and each step multiplies $e^{2(\tau - t_a)K}$ by $e^{2(\tau - t_a)} = e^2 \approx 7.39$ at these parameters (Figure 2). The ratio of the unrestricted late single-cutoff risk to this rule's risk is 1.037 at $m = 10^3$ and crosses 2, 10 and 100 near $m = 10^{4.9}$, $10^{8.3}$ and $10^{11.7}$; the difference between the two rates becomes substantial over several decades of budget. The curves report the risks of the two specified estimators.

The growing-radius threshold. Proposition 4.3 is checked on the same exact risk with $\alpha = 1$, $t_a = 1$ and $\tau = 2$, so $\gamma_\star = 1/2$. For $\gamma \in \{0, 0.1, 0.2, 0.3, 0.4, 0.45\}$ we minimize the calibration branch of (H.5) over K at eleven budgets from 10^{20} to 10^{60}, working in logarithms so the grid can run far enough to separate the power from the logarithmic factor, and regress log risk on $\log m$ after removing $2\alpha \log\log(m\rho_m^2)$. The fitted exponent matches $q(\gamma)$ to within 8.1×10^{-4} at every γ, and the minimizing cutoff stays within one integer of $\log(m\rho_m^2)/(2\tau)$ throughout. Results are in `growing_radius_check.csv`.

H.5 An approximate translation certificate

Proposition H.2 (Approximate minimax reduction)**.** *Let K be nonempty, compact and convex, $D = \mathrm{diam}(K)$, and $\mathbf{\Theta}_0$ a nonempty nuisance class. In the notation of Theorem D.9, define*

$$\epsilon = \sup_{x \in K} \mathrm{dist}\big(-\Pi_{\mathcal{T}_h} \mathsf{S}_x x, \overline{\mathsf{S}_h(\mathbf{\Theta}_0)}\big).$$

Then

$$0 \le \mathcal{R}^{\mathrm{eff}}_{n,m}(K) - \mathcal{R}^{\mathrm{joint}}_{n,m}(K; \mathbf{\Theta}_0) \le D^2 \min\{1, \epsilon/\sqrt{2\pi}\}. \tag{H.7}$$

Proof. Projection to $\mathcal{T}_h^\perp$ gives the first inequality. If $D = 0$ the claim is immediate; if $\epsilon = \infty$, bounded loss gives the stated bound. Otherwise, restrict an arbitrary joint rule to values in K. Its loss lies in $[0, D^2]$. For each x and $\delta > 0$, choose a nuisance score in $\mathsf{S}_h(\mathbf{\Theta}_0)$ whose distance from $-\Pi_{\mathcal{T}_h} \mathsf{S}_x x$ is at most $\epsilon + \delta$. Compare its joint law to the Gaussian law Q_x having mean $\mathsf{S}_{\mathrm{eff}} x$ in $\mathcal{T}_h^\perp$ and mean zero in $\mathcal{T}_h$. Two identity-covariance Gaussian shifts with Cameron–Martin mean difference h

have total variation $2\Phi(\|h\|/2)-1 \le \min\{1, \|h\|/\sqrt{2\pi}\}$, using total variation $\sup_A |P(A)-Q(A)|$. Pointwise in x, the loss under Q_x is at most the worst joint loss plus $D^2 \min\{1, (\epsilon+\delta)/\sqrt{2\pi}\}$. Under Q_x, the $\mathcal{T}_h$-coordinate is independent parameter-free noise. Integrating the rule over that coordinate and taking its conditional barycenter in K gives an efficient-experiment rule with no larger risk. Take the supremum over x, then the infimum over joint rules, and finally $\delta \downarrow 0$. The nuisance choices are pointwise comparisons; no measurable selector is used to define an experiment or an estimator. □

The bound preserves a relative minimax rate when $D^2\epsilon$ is controlled at the scale of that risk.

I TARGET-ORIENTED CALIBRATION: IDENTITIES AND EXPERIMENT

I.1 INFORMATION SENSITIVITY

Let the spaces be finite-dimensional and $H = n\mathsf{B}^*\mathsf{B} + \sum_j m_j \mathsf{C}_j^*\mathsf{C}_j \succ 0$. Minimizing the quadratic objective in (G.14) gives $\theta_v^* = -nH^{-1}\mathsf{B}^*\mathsf{A}v$ and the Schur formula in Section 5. Since $\partial_{m_j} H^{-1} = -H^{-1}\mathsf{C}_j^*\mathsf{C}_j H^{-1}$,

$$\partial_{m_j} I = n^2 \mathsf{A}^*\mathsf{B}H^{-1}\mathsf{C}_j^*\mathsf{C}_j H^{-1}\mathsf{B}^*\mathsf{A}.$$

Taking its quadratic form gives the sensitivity identity in Section 5. Applying $\partial I^{-1} = -I^{-1}(\partial I)I^{-1}$ gives the scalar-variance derivative. For a positive semidefinite target weight W, the criterion $\mathrm{tr}(WI^{-1})$ has derivative

$$-\partial_{m_j} \mathrm{tr}(WI^{-1}) = \|n\mathsf{C}_j H^{-1}\mathsf{B}^*\mathsf{A}I^{-1}W^{1/2}\|_{\mathrm{HS}}^2.$$

With channel cost c_j per information unit, dividing this marginal decrease by c_j compares local gains per cost. These identities specialize classical optimal-design differentiation after nuisance elimination; Schur-complement methods also appear in dynamical experimental design (Bürger et al., 2017).

I.2 A SOLVABLE DESIGN PROBLEM

For invertible square A, consider the independent channels

$$Y = Ax + B\eta + n^{-1/2}\varepsilon, \qquad Z_j = d_j\eta_j + m_j^{-1/2}\xi_j, \qquad d_j > 0, \quad \sum_{j=1}^{D} m_j = M. \tag{I.1}$$

Information levels $m_j > 0$ are continuous exposure times, with the same cost per unit. For known B, the unbiased estimator $A^{-1}(Y - BD_d^{-1}Z)$ has covariance equal to I^{-1}, where $D_d = \mathrm{diag}(d_j)$. Its total variance is

$$F_B(\boldsymbol{m}) = \frac{1}{n}\|A^{-1}\|_{\mathrm{HS}}^2 + \sum_{j=1}^{D} \frac{w_j(B)}{m_j}, \qquad w_j(B) = \frac{\|A^{-1}Be_j\|^2}{d_j^2}. \tag{I.2}$$

For positive weights, Cauchy–Schwarz gives the global optimum

$$m_j^{\mathrm{target}} = M\frac{\sqrt{w_j(B)}}{\sum_\ell \sqrt{w_\ell(B)}}. \tag{I.3}$$

The calibration-only nuisance variance is $\sum_j (m_j d_j^2)^{-1}$, whose minimizing allocation is $m_j^{\mathrm{nuisance}} = Md_j^{-1}/\sum_\ell d_\ell^{-1}$. The two criteria differ through the effect of each nuisance coordinate on target recovery. The closed-form solution fixes the allocation before test observations are generated.

The equal-quality example. With $n = m = 1$, $A = \mathrm{diag}(0.1, 1)$, $B = I_2$, $C_1 = \mathrm{diag}(\sqrt{0.1}, 10)$, and $C_2 = C_1P$ for the coordinate swap P, the two designs have the same calibration-only nuisance variance 10.01. The acquisition variance is 101; the calibration matrices add 1000.01 and 11, giving target variances 1101.01 and 112, a 9.83-fold gap caused solely by which nuisance coordinate overlaps the weak target direction. For unrestricted x, η this is a Gaussian location problem and the equivariant estimator attains the minimax variance, as also follows from increasingly diffuse Gaussian priors.

Pilot error	Uniform	Nuisance	Target	Oracle	Target gain vs nuisance
0%	1.521	1.411	1.169	1.169	17.10% [17.05, 17.15]
3%	1.941	1.829	1.587	1.175	13.23% [13.14, 13.32]
10%	6.042	5.931	5.701	1.174	3.88% [3.81, 3.95]
30%	42.001	41.890	41.702	1.171	0.45% [0.43, 0.47]

Table 1: Mean held-out target MSE at equal acquisition, calibration and pilot costs. Brackets are 95% pilot-cluster bootstrap intervals for the percentage reduction relative to nuisance-optimal exposure. The oracle knows the true nuisance response for allocation and reconstruction.

I.3 Independent-pilot protocol

The reproducible experiment uses $D = 24$, $A = \mathrm{diag}(j^{-1.25})$, $B = Q\,\mathrm{diag}(e^{-0.03j})$, and $d_j = e^{-0.015j}$. In each adjacent coordinate pair starting at odd index j, Q rotates by $0.45 + 0.2\sin j$ radians. The matrices AA^* and BB^* do not commute. Acquisition information and total calibration exposure are both 200000.

Pilot and costs. With the target absent, apply each known unit nuisance state e_j eight times and observe $Be_j + \sigma_p g$, with independent standard Gaussian vector g. The sample means give $\widehat{B}$ from 192 vector probe observations. For each nominal quality $q \in \{0, 0.03, 0.1, 0.3\}$, set $\sigma_p = q\|B\|_{\mathrm{HS}}\sqrt{8}/24$; hence $\mathbb{E}\|\widehat{B} - B\|_{\mathrm{HS}}^2 = q^2\|B\|_{\mathrm{HS}}^2$. The simulator uses B to specify these noise scenarios; the design rule receives only $\widehat{B}$, A, and the known calibration gains. Acquisition, calibration, and pilot budgets are fixed across policies. A vector acquisition or probe is charged one unit and a scalar calibration exposure one unit, giving total cost $200000 + 200000 + 192$ under this convention. Pilot quality varies through the probe-noise variance at a fixed observation count.

Policies and reconstruction. Compare uniform exposure, nuisance-optimal exposure, and target-optimal exposure from (I.3) with B replaced by $\widehat{B}$. All three reconstruct with the same $\widehat{B}$ in $A^{-1}(Y - \widehat{B}D_d^{-1}Z)$. The oracle uses true B for both allocation and reconstruction; it provides a known-model benchmark at the same total budget. Each policy uses the same candidate family of positive coordinate exposures. The implementable policies use the pilot and final observations.

Held-out evaluation. For each quality, draw 80 independent pilots and 1000 independent test problems per pilot. Draw $x_j = 100u_j/j$ with u uniform on the unit sphere, and draw η independently on the unit sphere, so every test point satisfies the target Sobolev and nuisance norm constraints with target radius 100 and nuisance radius 1, respectively. The target amplitude fixes signal units and does not affect these linear rules' error distributions; zero-output and acquisition-only error controls are also recorded in the tables. Fresh independent channel noise generates Y, Z; common random numbers pair policies within each test problem. The reported squared Euclidean errors estimate risk averaged over this test distribution. Seed 20260908 fixes pilot and test draws. We bootstrap 4000 times over independent pilots, retaining each pilot's test problems and policy pairing together, to obtain 95% intervals.

Prediction and model error. For fixed $\widehat{B}$, the nominal variance is $F_{\widehat{B}}(\boldsymbol{m})$. Under the stated test distribution, the population MSE of each implementable policy is exactly

$$F_{\widehat{B}}(\boldsymbol{m}) + \frac{1}{D}\|A^{-1}(B - \widehat{B})\|_{\mathrm{HS}}^2. \tag{I.4}$$

The second term is the response-estimation error, which is independent of the exposure allocation. The nominal criterion ranks the population MSE of these three policies exactly in this linear model. The tables report the two terms, empirical MSE, and the allocation diagnostic $F_B(\boldsymbol{m})/\min_{\boldsymbol{u}} F_B(\boldsymbol{u}) - 1$. This diagnostic evaluates the chosen exposure under the true response. It is computed after allocation and uses B only for evaluation. It differs from the full MSE gap to the known-B oracle, which also includes response-estimation error.

Results. Table 1 reports target MSE and paired reductions. The target policy beats both implementable baselines in all 80 pilot-level empirical comparisons at each quality. At 3% pilot error, its

true-score allocation regret averages 0.0085%, versus 20.63% for nuisance-optimal exposure and 30.08% for uniform exposure. At 10% and 30% pilot error the target allocation regret is 0.235% and 2.44%. Total reconstruction MSE rises much more rapidly because the separate score-error term in (I.4) is amplified by A^{-1}.

Varying the ill-posedness. We repeat the protocol across the acquisition exponent. Only α in $A = \mathrm{diag}(j^{-\alpha})$ changes; the dimension, decay profiles, Q, budgets, probe count, evaluation distribution, and replication scheme stay fixed. The sweep uses a separate random stream. Against nuisance-optimal allocation at 3% pilot error the reduction is 1.96, 5.39, 9.33, 13.18, 16.74, 19.94 and 22.80 percent at $\alpha = 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2$, with bootstrap intervals narrower than 0.3 percentage points throughout; with an exact response it runs from 2.45 to 30.23 percent. For this sensor family, increasing α increases the disparity between the target weights $w_j = \|A^{-1}Be_j\|^2/d_j^2$ in (I.2) and the nuisance weights d_j^{-2}. The empirical gains increase across the tested exponents. The value at $\alpha = 1.25$ agrees with the main study to 0.05 percentage points, the two differing only in random stream. Allocation regret under the true response also grows with α, from 0.0046% to 0.117%, and remains smaller than the full reconstruction gap to the known-response oracle. Results are in `target_design_alpha_sweep.csv`.

I.4 Learning the response from passively observed data

The protocol above sets known nuisance states. Neither multi-coil imaging nor a parabolic initial state permits that, so we repeat the study with the response *learned* from triples (x_i, Y_i, Z_i) in which the nuisance is whatever nature supplied and is seen only through the calibration channel. Writing $V_i = Y_i - \mathsf{A}x_i$ and $W_i = D_d^{-1}Z_i$,

$$V_i = \mathsf{B}\eta_i + n^{-1/2}\varepsilon_i, \qquad W_i = \eta_i + u_i, \qquad \mathrm{Cov}(u_i) = \Sigma_u = \mathrm{diag}\big(D/(m_{\mathrm{tr}}d_j^2)\big), \tag{I.5}$$

an errors-in-variables regression: the covariate is the nuisance, and it is observed with noise set by the calibration exposure m_{tr} spent per training measurement. Training measurements are taken at lower calibration fidelity than the deployment budget m, as in multifidelity practice (Peherstorfer et al., 2018); this is what makes Σ_u non-negligible.

Attenuation, and why the usual validation criterion selects it. Ordinary least squares $\widehat{\mathsf{B}}_{\mathrm{ls}} = (\sum_i V_i W_i^\top)(\sum_i W_i W_i^\top)^{-1}$ converges as $N \to \infty$ to

$$\mathsf{B}\,\Sigma_\eta(\Sigma_\eta + \Sigma_u)^{-1} \neq \mathsf{B}, \tag{I.6}$$

an attenuation that no amount of training data removes; because d_j decays, the coordinates measured least accurately are shrunk most. The limit in (I.6) is exactly the best linear predictor of V from W, so selecting a model by held-out predictive error on the training pairs selects the attenuated operator. The operator that best predicts the acquisition channel from the calibration channel is not the operator the allocation and the reconstruction need. Subtracting the known covariate noise,

$$\widehat{\mathsf{B}}_{\mathrm{c}} = \Big(\sum_i V_i W_i^\top\Big)\Big(\sum_i W_i W_i^\top - \lambda N \Sigma_u\Big)^{-1}, \tag{I.7}$$

is consistent at $\lambda = 1$. It needs no tuning, because the design fixes Σ_u: λ is set to the largest value in $[0, 1]$ for which the corrected Gram keeps least eigenvalue at least $0.05\,\mathrm{tr}(\sum_i W_i W_i^\top)/D$, a function of the training Gram alone. The guard binds only at small N and low fidelity; the mean share retained is 0.15, 0.51 and 0.99 at $m_{\mathrm{tr}} = 10^3$ for $N = 50, 100, 300$, is 1 from $N = 100$ on at $m_{\mathrm{tr}} = 5 \times 10^3$, and at $m_{\mathrm{tr}} = 2 \times 10^4$ falls below 1 only at $N = 50$, to 0.97 (`operator_learning_guard.csv`).

Protocol. Everything not named here follows Appendix I.3: the same model, budgets, evaluation distribution and pairing. The replication unit is an independent training set, 80 of them per configuration, each followed by 1000 fresh test problems; the bootstrap resamples training sets. Both estimators are formed from the same training set, all three policies then receive the same $\widehat{\mathsf{B}}$ and reconstruct with it, and the training set costs N acquisitions and Nm_{tr} calibration units shared by every policy. Training calibration is allocated uniformly, since the target allocation is not known when the training data are collected; the iterated version is the sequential design we do not study. No test data enter training or allocation.

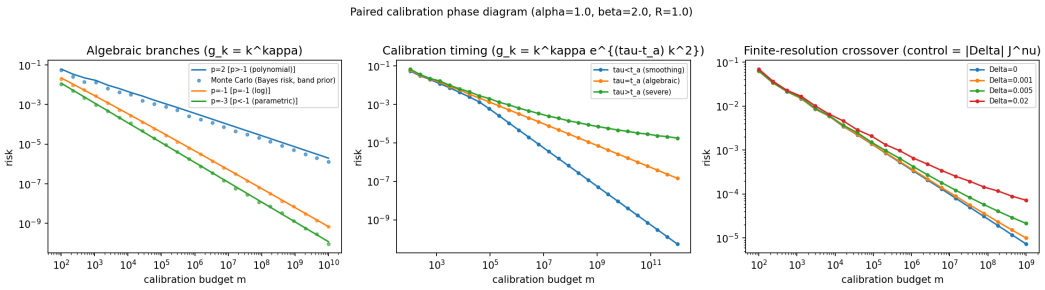

Figure 3: Diagonal sequence model. Optimized cutoff-estimator risk versus calibration budget m across the three algebraic regimes (left), and the calibration-timing transition (middle, right) between smoothing, algebraic, and severely ill-posed behaviour. Fitted log–log slopes match the exponents of Theorems 3.4 and 4.1; the severe branch follows a logarithmic law with slow convergence of the fitted logarithmic exponent. Dots are the Monte Carlo prior-averaged risk of the cutoff estimator under the band prior.

Results. Along N at $m_{\mathrm{tr}} = 5 \times 10^3$ the two estimators separate completely. Least squares stalls: from $N = 3 \times 10^3$ to $N = 3 \times 10^4$ its response error moves only from 2.0×10^{-2} to 1.9×10^{-2}, its allocation regret from 8.3×10^{-4} to 6.4×10^{-4}, and its reduction against nuisance-optimal allocation from 1.68% to 1.76%. The corrected estimator improves as $1/N$, reaching response error 1.5×10^{-4}, regret 8.3×10^{-6} and a reduction of 16.50% with 95% interval $[16.45, 16.55]\%$, against the ceiling of 17.10% available with a known response. At $N = 100$ the correction is worse than least squares, 0.56% against 0.63%: removing the bias costs variance, and with this little data the trade is not yet worth making. Along the training fidelity at $N = 10^4$, the reduction attained by least squares is 0.07%, 1.73% and 10.92% at $m_{\mathrm{tr}} = 10^3, 5 \times 10^3, 2 \times 10^4$, while the corrected estimator gives 8.69%, 15.33% and 16.66%. At the lowest fidelity the uncorrected pipeline recovers essentially none of the available gain. Results are in `operator_learning_rates.csv`.

A learned nonlinear response. As a robustness run, not a main result, we replace the linear fit by a two-layer $24 \to 48 \to 24$ network with $\tanh$ activation, trained by Adam on at most 4000 triples with early stopping on a 20% split, and use its average Jacobian for the allocation and the network itself for reconstruction. It remains close to least squares, consistent with attenuation: at $N = 10^4$ and $m_{\mathrm{tr}} = 5 \times 10^3$ the Jacobian sits at squared distance 2.3×10^{-2} from B but 1.0×10^{-3} from the attenuated operator of (I.6), a factor of 22 closer, rising to a factor of 101 at $m_{\mathrm{tr}} = 10^3$; the corresponding reductions are 1.16% and -0.05%. Flexibility does not help, because the attenuation is a property of the training criterion rather than of the model class. Results are in `operator_learning_nonlinear.csv`.

Using the known nuisance radius in reconstruction. As a second reconstruction rule, project the calibration estimate $D_d^{-1} Z$ onto the unit nuisance ball and use $A^{-1}(Y - \widehat{B}\Pi_{\|\eta\| \le 1}(D_d^{-1} Z))$. Keep every allocation, pilot, and test draw unchanged. At 3% pilot error, target allocation reduces its held-out MSE by 11.88% relative to nuisance-optimal allocation, with 95% interval $[11.78, 11.98]\%$. The gains at 0%, 10% and 30% pilot error are 15.15%, 3.78%, and 0.58%. Here the linear variance criterion is a design surrogate for a nonlinear constrained estimator; its effect is measured on independent test problems. The pilot-level and summary tables include these constrained-estimator errors alongside the linear ones.

J Numerical constructions and protocols

This appendix records the constructions and protocols behind Section 6. The calculations use the conventions of Appendix B.1.

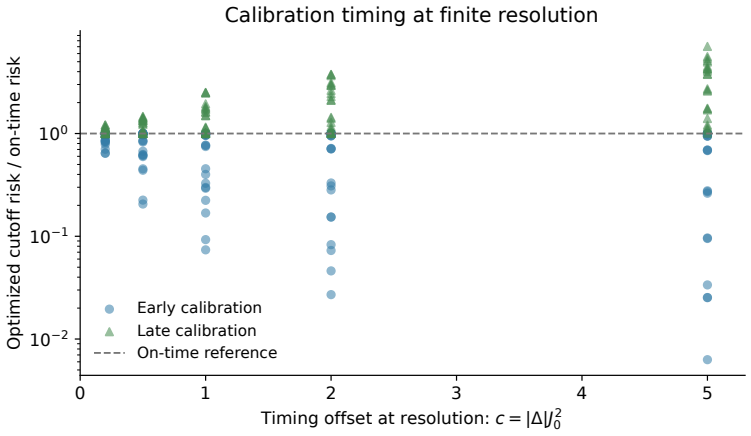

Figure 4: Timing offsets at the on-time resolution. Each point represents one acquisition and calibration budget pair, with $c = |\Delta| J_0^2$ and J_0 the optimal on-time cutoff. The vertical axis compares the optimized cutoff risk at offset Δ with its on-time value. The dashed line is the on-time reference.

J.1 Diagonal sequence model

The diagonal model (G.3) is instantiated coordinatewise. On mode k we set the object score $a_k = k^{-\alpha}$ and the calibration/acquisition transfer through a profile g_k taken from the parabolic construction of Section 4, so that $\tau - t_a$ selects algebraic ($\tau = t_a$), exponentially decaying ($\tau < t_a$), or exponentially growing ($\tau > t_a$) transfer profiles. The efficient per-mode variance is σ_k^2/a_k^2 with $\sigma_k^2 = n^{-1} + m^{-1} g_k^2$, as in Appendix G. The object set is the Sobolev ellipsoid (3.5) of smoothness β; for a spectral-cutoff estimator that keeps the first K modes we evaluate the worst-case risk over the ellipsoid analytically (bias from the truncated tail plus efficient variance of the kept modes) and minimize over K, and we cross-check with a Monte Carlo risk under Gaussian noise. The acquisition budget n is set far above the calibration budget m so the acquisition-oracle floor is negligible and the calibration branch is isolated; risk is fit as a log–log slope against m. For the severe branch ($\tau > t_a$) the predicted law is logarithmic, so we additionally fit the slope of $\log(\text{risk})$ against $\log \log m$ and track its slow convergence to the predicted exponent over budgets up to $m = 10^{240}$. Numerical guards clip the exponentially growing g_k to avoid overflow without affecting the retained modes. The Monte Carlo curve averages the same cutoff estimator over a prior whose energy is concentrated on the ellipsoid boundary near the cutoff scale, so it is a prior-averaged risk of that estimator and lies below the worst-case curve.

The near-critical window. We evaluate a diagonal model with $\alpha = 1$, $\beta = 2$, $R = 1$, $\kappa = 0$, and $\nu = 2$. The cutoff risk is

$$F_\Delta(J) = (J+1)^{-4} + n^{-1} \sum_{k \le J} k^2 + m^{-1} \sum_{k \le J} k^2 e^{2\Delta k^2}, \qquad F_\Delta(0) = 1.$$

For $n \in \{10^4, 10^6, 10^8\}$ and eight logarithmically spaced m values from 10^3 to 10^{10}, let J_0 minimize F_0. For each $c \in \{0.2, 0.5, 1, 2, 5\}$, set $\Delta = \pm c/J_0^2$ and minimize F_Δ again. We use $t_a = 6$ so all resulting calibration times are nonnegative. The ratio $\min_J F_\Delta(J)/\min_J F_0(J)$ measures the effect of timing on this estimator class. The finite search is certified by the increasing acquisition variance: once that term alone exceeds the best risk found, no larger cutoff can improve it. Figure 4 displays the ratios, with one as the on-time reference. Proposition G.2 concerns minimax rates up to structural constants; the plotted quantities are exact cutoff risks.

J.2 Beyond the diagonal model: constructions and figures

The four experiments of Appendix J.2 use one finite-dimensional construction: given score operators $(\mathsf{A}, \mathsf{B}, \mathsf{C})$ we form $\mathcal{U}_0 = \overline{\mathsf{B}(\ker \mathsf{C})}$, the projection P_0 onto $\mathcal{Y}_0 = \mathcal{U}_0^\perp$, the reduced score $\mathsf{A}_0 = P_0\mathsf{A}$ and the transfer operator $\mathsf{R} = P_0\mathsf{B}\mathsf{C}^+$, and evaluate the efficient information either as the Schur

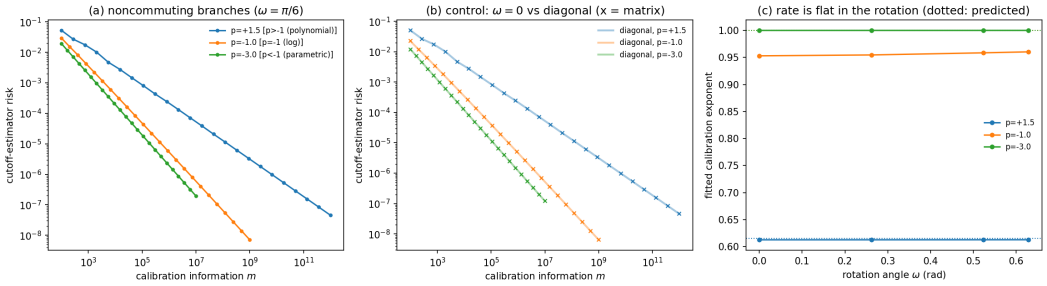

Figure 5: Rates without a common eigenbasis. Left: the three algebraic branches with a pair rotation inserted (cutoff-estimator risk). Middle: at zero rotation the computation reproduces the diagonal reference of Figure 3 to machine precision. Right: fitted calibration exponents are flat in the rotation angle and match the predictions of Theorem 3.4, although the risk itself moves by up to 84% across the same range.

complement of the joint Fisher block (2.1) or through the resolvent (2.6). In the rate experiments the acquisition budget n is large, so the resolvent is evaluated in the equivalent form $\mathsf{I}_{\mathrm{eff},n,m} = \mathsf{A}_0^*[n^{-1}I + m^{-1}\mathsf{R}\mathsf{R}^*]^{-1}\mathsf{A}_0$, whose conditioning is that of $\mathsf{R}\mathsf{R}^*$ rather than of n/m.

Marginal separation. With $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$, $\mathbf{\Theta} = \mathbb{R}^2$, $\mathsf{A} = I$ and the two score pairs of Proposition 2.2, we observe $Y = x + \mathsf{B}\theta + n^{-1/2}\varepsilon$ and $Z = \mathsf{C}\theta + m^{-1/2}\zeta$. The efficient estimator subtracts the calibration channel when it is informative about the acquisition nuisance and ignores it otherwise; risks are averaged over 2×10^5 replicates at each of 26 calibration budgets and agree with the closed forms $n^{-1} + m^{-1}$ and $\rho^2 + n^{-1}$ to under 1%. For the transverse model, the rule $\widehat{x} = Y$ has worst-case risk $\rho^2 + n^{-1}$ over $|\theta_1| \le \rho$; the simulation includes both endpoints. This is the risk of the stated rule under a bounded nuisance, whereas Proposition 2.2 characterizes information with unrestricted nuisance directions.

Noncommuting rates. The ellipsoid, the smoothness β and the acquisition budget follow Appendix J.1. The diagonal cutoff risk generalizes to $\min_K [R^2\lambda_{K+1}^{-2\beta} + \mathrm{tr}(P_K \mathsf{I}_{\mathrm{eff},n,m} P_K)^{-1}]$, the trace replacing the sum of per-mode efficient variances; the two coincide when $\mathsf{I}_{\mathrm{eff},n,m}$ is diagonal. Cutoffs are restricted to complete rotation blocks (even J): a cutoff that splits a coupled pair requires a leakage-bias term, which the compressed inverse alone omits; with complete blocks the quantity is the exact worst-case risk of the block-wise unbiased cutoff estimator. Each regime is swept over a range of m on which the mode truncation is verified, against a finely truncated diagonal reference, to leave the risk unchanged. Because the pair rotation mixes only (ϕ_{2j-1}, ϕ_{2j}) and the remaining factors are diagonal, $\mathsf{I}_{\mathrm{eff},n,m}$ is exactly block diagonal with 2×2 blocks, so all cutoffs are evaluated in closed form; this is checked against the dense computation.

Variable-coefficient timing. We use $\mathfrak{a}(\xi) = 1 + \frac{1}{2}\sin^2(2\pi\xi)$ and $q(\xi) = 0.3(1 + \cos^2(4\pi\xi))$, second-order finite differences on 8000 interior points, and retain the first 2000 eigenmodes; the Weyl check is reported over the first 300, where the discretization error is below 10^{-3}. The rotation angles are chosen as $\sin\omega_j = s_0 \exp(-T(\lambda_{2j}^2 - \lambda_{2j-1}^2))$ with $s_0 = 0.6 < \sin(\pi/4)$ and $T = 0.02$, which covers every observation time used ($t_a = 0.01$ and $|\tau - t_a| \le 0.01$), so that (G.18) holds by construction; the weighted off-diagonal entries are bounded by 0.51 over all pairs and calibration times, and only the first 4 pairs carry a rotation above 0.01 radians. The transfer operator is formed directly as $\mathsf{R} = Q\,\mathrm{diag}(\lambda_k^\kappa e^{\Delta\lambda_k^2})$ rather than as $\mathsf{B}\mathsf{C}^+$, because $e^{+\tau G}$ overflows long before the product does; the block-wise and the dense evaluation agree to 6×10^{-17}. The exponentially growing profile of the late branch is clipped exactly as in Appendix J.1. The severe branch approaches its $[\log m]^{-2\beta/\nu}$ asymptotic form at very large budgets. Over the tested range, the experiment resolves the three branches and agrees with the diagonal model at the same budgets.

Order-one sensor mixing and the spectral exponent. The study above uses $\nu = 2$, where (G.18) forces the angles to decay. To test part (i) of Proposition G.4 we hold the angle at $\omega = 0.6$ radians

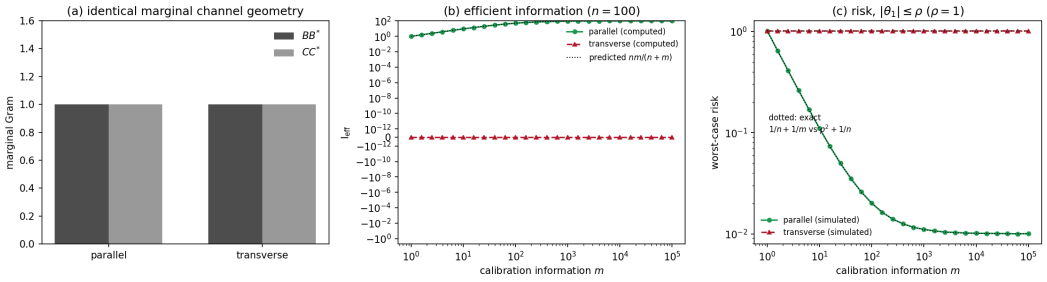

Figure 6: Marginal channel geometry is insufficient (Proposition 2.2). Left: the two models have identical marginal Grams. Middle: the computed efficient information is $nm/(n+m)$ in the parallel model and exactly zero in the transverse one. Right: simulated worst-case risk; the parallel model improves as $n^{-1} + m^{-1}$ while the transverse model remains at a calibration-independent floor.

on *every* pair, so that $Q - I$ is not compact, and pair mode k with mode $k + d$ for fixed offsets $d \in \{1, 4, 16, 32\}$. The spectrum is the same Sturm–Liouville operator on 8000 interior points with 1200 modes retained; $c_{\mathfrak{a}} = 3.485677$, and $|\lambda_k^2 - c_{\mathfrak{a}}^2 k^2|$ stays below 1.3×10^3 over the first 300 modes, a relative error of 1.2×10^{-3} at mode 300.

The angle condition controls the block matrix $W_{\tau,j} Q_j^* W_{\tau,j}^{-1}$, whose singular values determine the two-sided constants of Assumption 3.2; its determinant is one, so its condition number is $\sigma_{\max}^2$. At $|\tau - t_a| = 0.02$ and $\nu = 1$ the largest condition number over all blocks is 1.08, 1.38, 4.15 and 29.2 for the four offsets, so the link condition holds with fixed constants at a fixed order-one angle. At $\nu = 2$ and the same angles the weighted gaps reach 5.6×10^2 to 1.7×10^4 and the condition numbers overflow, showing the contrasting $\nu = 2$ behavior.

At $\nu = 1$ the rates survive the mixing. On the on-time branch, where the cutoffs stay interior to the retained spectrum, the fitted calibration exponent is 0.5322 against the predicted $2\beta/(2\beta + 2\alpha + 2\kappa + 1) = 0.5333$ for $\alpha = 1$, $\beta = 2$, $\kappa = 1/4$. The efficient information is genuinely off diagonal: on the late branch, replacing $\mathsf{I}_{\mathrm{eff},n,m}$ by its diagonal part changes the computed cutoff risk by 14% at offset 16 and by 28% at offset 32. The summable early branch is not resolved here, because its optimal cutoff runs into the 1200-mode budget; the table records the realized cutoffs and flags that fit as non-interior. Results are in `spectral_exponent_dichotomy.csv` and `spectral_exponent_rates.csv`.

Bounded-nuisance counterexample. For the scalar model of Proposition D.8, use

$$\widehat{x}_c = \mathrm{proj}_{[-1,1]}(\lambda(c) Y), \qquad \lambda(c) = \begin{cases} (1-c)/(1+(1-c)^2), & 0 \le c < 1, \\ 0, & c \ge 1. \end{cases}$$

For fixed x and $\lambda > 0$, the risk derivative in the Gaussian mean has the sign of $\lambda \mathbb{E}[Y \mid |\lambda Y| < 1] - x$. The truncated Gaussian mean increases with the untruncated mean, so any stationary point is a minimum. The maximum over $\theta \in [-c, c]$ is therefore attained at an endpoint. We evaluate the corresponding Gaussian integrals analytically and maximize over $x \in [-1, 1]$ numerically, using a dense grid followed by refinement of candidate maxima. The curve in Figure 8 is this rule's worst-case risk. The bound $1/(1 + (1-c)^2) < 1$ proves strict improvement for every $c < 1$; for $c \ge 1$, the zero rule has risk 1 and the indistinguishable endpoint pair proves minimax equality.

Several calibration channels. With channels at times τ_j and budgets m_j the efficient per-mode variance uses the weighted Gram $\sum_j m_j e^{-2\tau_j k^\nu}$ in place of a single $m e^{-2\tau k^\nu}$, and everything else follows Appendix J.1. We evaluate the cutoff risk for six configurations and compare it with the single-channel risks at the earliest time with the total and with the earliest budget, and separately sweep the total budget for a late-only design against an early one.

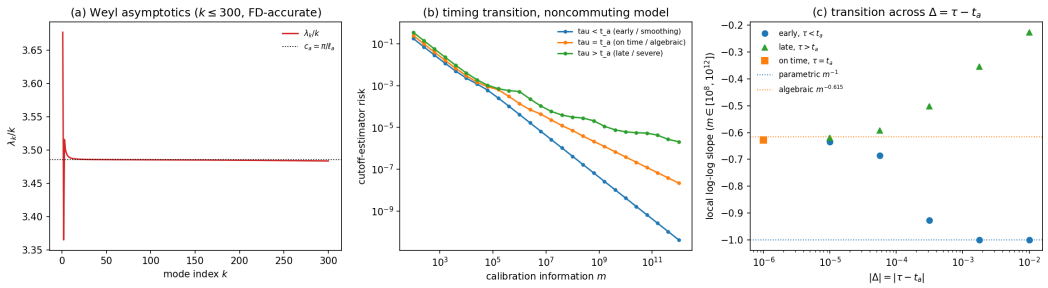

Figure 7: Calibration timing for a variable-coefficient heat equation with frequency-localized sensor mixing (Proposition G.4). Left: the discrete spectrum follows the predicted Weyl law. Middle: the three timing branches separate as in the diagonal model. Right: sweeping $\Delta = \tau - t_a$ through zero traces the transition from the parametric early branch to the severely ill-posed late branch.

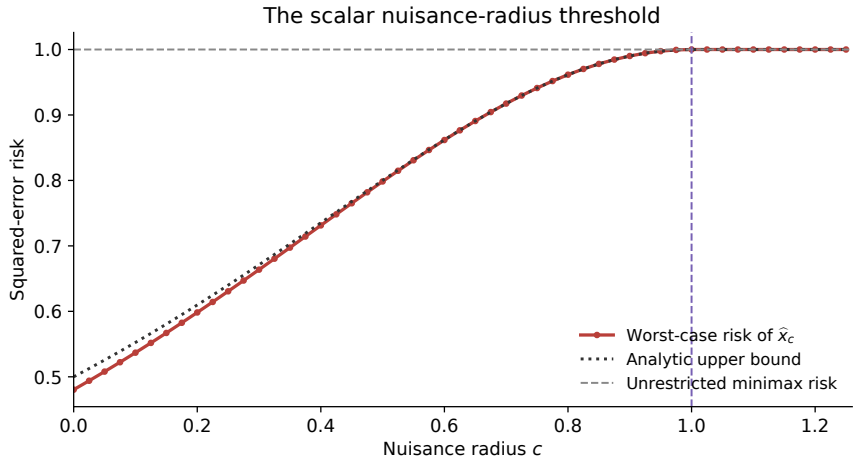

Figure 8: The nuisance-radius threshold in the scalar model (Proposition D.8). The solid curve is the numerically maximized risk of $\widehat{x}_c = \mathrm{proj}_{[-1,1]}(\lambda(c)Y)$, and the dotted curve is its analytic upper bound for $c < 1$. At $c \ge 1$, $\lambda(c) = 0$ and the rule attains the minimax risk 1.

J.3 Systematic parameter sweeps

Choice of fit window. Every fitted exponent in Appendix J.2 is read off a finite range of calibration budgets, so we select a fit window before estimating the slope. For each configuration we select it automatically on the exactly computable diagonal risk, requiring two conditions simultaneously. First, the cutoff risk truncated at the working number of modes must agree with a reference truncated at 20000 modes to within 10^{-3}; this excludes ranges in which the retained spectrum is exhausted. Second, the acquisition-only oracle risk, defined by the $m \to \infty$ limit, must contribute at most a tenth of the total. This second condition excludes the acquisition floor, where risk becomes insensitive to m and a power fit would no longer measure the calibration exponent. Configurations admitting no window of at least four decades are reported as skipped rather than fitted; eight of the 96 grid points are skipped on this ground.

Rate grid. We sweep $\alpha \in \{0.5, 1, 1.5, 2\}$, $\beta \in \{1, 1.5, 2, 3\}$ and $p = 2\alpha + 2\kappa \in \{2, 1.5, 0.5, -1, -2, -3\}$, so that each of the three regimes of Theorem 3.4 is represented at every smoothness pair, with the pair rotation active throughout. On the boundary branch $p = -1$ the predicted law is $m^{-1} \log m$, so no power law is correct on a finite window; we therefore compare the fitted slope with that of $m^{-1} \log m$ over the same window. The residual median discrepancy of 1.4×10^{-2} on that branch reflects the difference between the true cutoff risk and the idealized law, and does not shrink with the window.

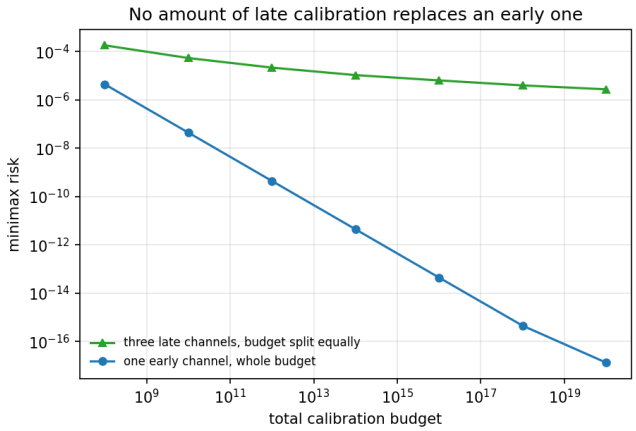

Figure 9: Early versus late calibration (Proposition G.3). Three channels after the acquisition time sharing a budget, against one channel before it holding the whole budget. The late design follows the logarithmic law of (3.7), the early one is parametric, and the gap grows without bound.

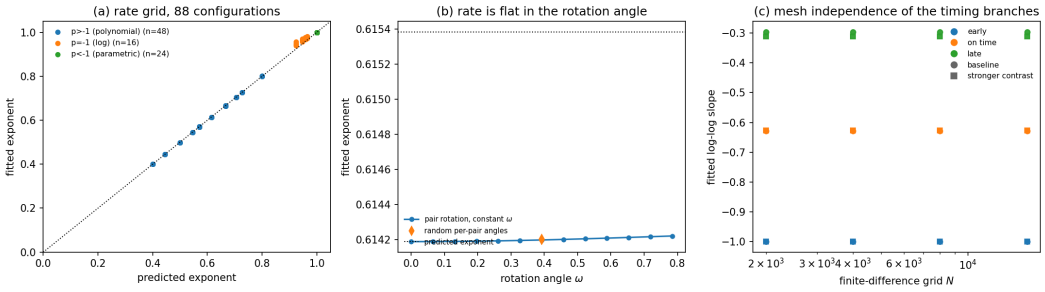

Figure 10: Systematic sweeps behind Appendix J.2. (a) Fitted against predicted calibration exponents over 88 admissible (α, β, p) configurations. (b) The exponent is flat in the rotation angle over the whole admissible range and under randomized per-pair angles. (c) The three timing branches are unchanged across finite-difference grids, mode counts and coefficient pairs.

Mesh refinement. The timing study is repeated on finite-difference grids of 2000, 4000, 8000 and 16000 points, retaining 1000 or 2000 modes, for the coefficient pair of Appendix J.2 and for a second pair with stronger contrast, $\mathfrak{a}(\xi) = 0.4 + 1.6\xi^2$ and $q(\xi) = 1 + 2\sin^2(\pi\xi)$, with the angle horizon of the timing experiment throughout.

J.4 MRI COIL-SENSITIVITY LOCAL MODEL

The multi-coil forward map is $T(\theta, x) = P_\Omega \mathcal{F}(S(\theta) \odot x)$ on a finite Cartesian grid, with C coils, sampling pattern Ω, and coil sensitivities $S(\theta)$ expanded in a smooth (low-frequency discrete-cosine) basis with coefficients θ. We linearize about a reference (θ_0, x_0): the object score is $\mathsf{A}\,\delta x = P_\Omega \mathcal{F}(S(\theta_0) \odot \delta x)$ and the acquisition nuisance score is $\mathsf{B}\,\delta\theta = P_\Omega \mathcal{F}((\partial_\theta S\,\delta\theta) \odot x_0)$, so B depends on x_0 while A does not. The calibration channel observes the sensitivities through a separate low-resolution readout $\mathsf{C}\,\delta\theta$. Complex operators are realified into real block form (Appendix B.1) so the real Hilbert-space theory applies verbatim. We verify A and B against finite-difference Jacobians of T (relative error $\approx 5 \times 10^{-10}$).

Given $(\mathsf{A}, \mathsf{B}, \mathsf{C})$ we form the paired relation (2.3) and extract $\mathcal{U}_0 = \operatorname{mul} \mathcal{G}_{BC}$ by the null space of C mapped through B and projected off $\overline{\operatorname{ran}}\,\mathsf{C}$-consistent directions, $\mathcal{Y}_0 = \mathcal{U}_0^\perp$, P_0, $\mathsf{A}_0 = P_0\mathsf{A}$, and the transfer operator R of Lemma D.2. Efficient information is computed three ways and compared: the shorted operator (D.9) via the Schur complement of the joint Fisher block (2.1), the exact resolvent (2.6), and the variational form (D.10) solved as a least-squares problem in θ. The three agree

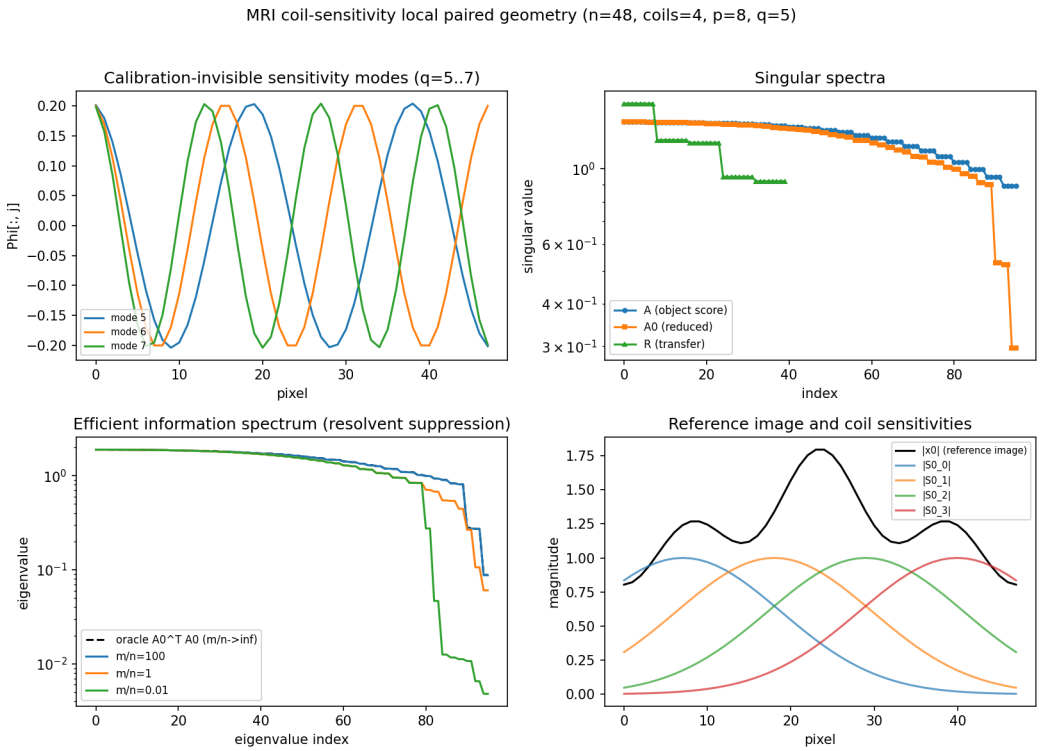

Figure 11: Finite-dimensional MRI paired geometry. Structural-loss modes $\mathcal{U}_0$ (calibration-invisible sensitivity directions), singular spectra of the reduced score A_0 and transfer operator R, and the efficient-information spectrum, computed by three independent formulas that agree to machine precision.

to relative error 10^{-15}; $\ker \mathsf{I}_{\mathrm{eff},n,m} = \ker \mathsf{A}_0$ holds to numerical zero; and the calibration limits (D.18) are recovered as $m/n \to \infty$ and $m/n \to 0$ to relative error 10^{-9}. In the reported instance $\dim \mathcal{Y} = 384$, $\dim \mathcal{U}_0 = 24$, $\operatorname{rank} \mathsf{A}_0 = 96$, and $\operatorname{rank} \mathsf{R} = 40$. The calibration derivative C observes the first $q = 5$ of the $p = 8$ smooth modes of each coil; the structural loss of the instance is therefore fixed by this choice, $\dim \mathcal{U}_0 = 2 \times 4 \times (p - q) = 24$, as specified by the observation model.

J.5 Real-data illustration: calibration budget at a fixed total budget

Data and reconstruction. We use fully sampled multi-coil fastMRI volumes from two collections (Knoll et al., 2020). The knee set contributes 200 volumes, uniformly 15 coils, phase-encode widths 320 to 388. The brain set contributes 200 volumes sampled evenly across the collection, which differs from the knee set in anatomy, in contrast, in matrix size (widths 213 to 396) and in receive array, with channel counts from 2 to 28; since the channel count governs how well an array can unfold an aliased image, the reported brain numbers restrict to the 411 of 600 slices with at least twelve channels. Quality comparisons are made within each collection. Root-sum-of-squares images are the ground truth. Reconstruction is classical Cartesian SENSE (Pruessmann et al., 1999): the encoding operator is $Ex = \{\Omega \odot \mathcal{F}(S_c \odot x)\}_{c=1}^{C}$ and we solve $(E^*E + \lambda' I)x = E^*y$ by 40 conjugate-gradient iterations with $\lambda' = 10^{-3} \max |E^*y|$, a relative regularization of about 10^{-7}, so that early stopping is the operative regularizer. Sensitivities are estimated from the auto-calibration signal (ACS): a central fully sampled block of width w yields low-resolution coil images normalized by their root-sum-of-squares. The total number of sampled phase-encode lines is fixed at $\mathrm{round}(\mathrm{width}/R)$; the w central lines are fully sampled and the remaining budget is placed equispaced over the outer lines. Quality is measured after centre-cropping to 320×320 and resolving the global scale by least squares to the ground-truth magnitude. We report NMSE, a normalized magnitude-error criterion for this illustration; its normalization and scale alignment differ from the Hilbert-space loss in the theory.

Conditions and noise. Three conditions are compared: *oracle* (sensitivities from the fully sampled k-space of the same slice, a privileged baseline that is exact at full sampling because the ground truth is the root-sum-of-squares of the same coil images), *finite* (sensitivities from the width-w ACS of the noisy sampled data, which are reused for reconstruction) and *calibration-blind* (unit sensitivities). Complex Gaussian noise is added to the sampled k-space with standard deviation 0.02 times the root-mean-square of the fully sampled k-space of the slice, fixed once per slice and independent of the sampling pattern, with three realizations per slice seeded from a process-independent hash of the slice identifier. Three central slices from each of the 200 volumes of a collection are used, at accelerations $R \in \{2, 3, 4, 6, 8\}$ and ACS widths $w \in \{8, 16, 24, 32, 48, 64, 96\}$; a point is skipped when w exceeds the total budget, and (R, w) cells reachable only in the widest volumes are excluded, because they would be averaged over a much smaller population than their neighbours. Unreadable volumes are skipped with a warning; one of the 204 local knee volumes is truncated.

Connection to the allocation analysis. ACS lines both estimate sensitivities and contribute to imaging. Their signal depends on the object, and their noise is reused in reconstruction. The scan therefore studies an empirical allocation tradeoff with overlapping data uses. Remark F.4 supplies the theoretical motivation: balancing calibration and acquisition costs gives an allocation scale that changes with the total budget. Here we measure the finite-grid optimum and its uncertainty separately on each collection.

Uncertainty. Slices of one volume and noise realizations of one slice are not independent. All uncertainty statements use a cluster bootstrap over volumes (2000 resamples; the slices and noise realizations of a resampled volume are kept together). For each R we report the mean NMSE of every condition with 95% bootstrap intervals, the paired differences between adjacent ACS widths for the finite condition, and the bootstrap distribution of the best width w^*. We call w^* an interior optimum only if it is not at an end of the admissible grid and the paired differences to both grid neighbours exclude zero; otherwise we report a plateau.

Outcome. On both collections the finite-calibration curve lies between the oracle and the calibration-blind baseline at every acceleration and every width (Figure 12); each reported cell averages 1791 slice-by-seed reconstructions from 199 knee volumes, or between 1179 and 1233 from 137 brain volumes. An interior optimum appears at two of the five accelerations, and at the same two on both collections. On the knee set the best width is $w^* = 24$ at $R = 3$ (NMSE 0.0456, 95% interval $[0.0406, 0.0510]$), with paired differences to its grid neighbours of -1.05×10^{-3}, interval $[-1.35, -0.75] \times 10^{-3}$, and $+0.79 \times 10^{-3}$, interval $[+0.47, +1.14] \times 10^{-3}$, both excluding zero; and $w^* = 16$ at $R = 4$ (0.1122, $[0.1045, 0.1204]$), with paired differences -2.49×10^{-2} and $+2.04 \times 10^{-3}$. On the brain set it is $w^* = 32$ at $R = 3$ (0.0210, $[0.0188, 0.0238]$; paired differences -1.66×10^{-3} and $+3.98 \times 10^{-3}$) and $w^* = 24$ at $R = 4$ (0.0596, $[0.0547, 0.0648]$; -5.13×10^{-3} and $+5.39 \times 10^{-3}$), again all excluding zero. At $R = 2$, $R = 6$ and $R = 8$ the error decreases over the whole admissible range on both collections, so the best admissible width is the widest one the total budget allows and no interior optimum is visible: at $R = 6$ and $R = 8$ because w cannot exceed the total budget, and at $R = 2$ because error still decreases at the widest tested width.

On both collections, the best ACS width decreases between the two accelerations with resolved interior optima: from 24 to 16 on the knee set and from 32 to 24 on the brain set. The brain optima are one grid step wider at both accelerations (Figure 13). This is a finite-grid observation of the allocation tradeoff; the other accelerations yield boundary optima on the tested grid. Scored by PSNR rather than NMSE the knee optima move one grid step wider, to 32 and 24, at which point the two collections agree; the location of the optimum therefore depends on the evaluation metric, and NMSE is our primary reconstruction criterion. For the oracle condition, PSNR reaches its numerical floor: root-sum-of-squares images peak near 10^{-4} and the mean squared error is near 10^{-14}. Its reported PSNR is then determined by the image peak, so we evaluate the oracle using NMSE.

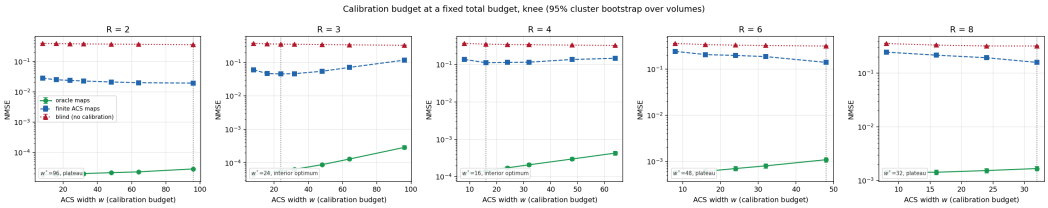

Figure 12: Classical SENSE on fastMRI knee slices at a fixed total number of phase-encode lines (the brain panels are in the released figures): NMSE against the ACS width w at five accelerations for oracle, finite-ACS and calibration-blind sensitivities. Error bars are 95% cluster-bootstrap intervals over volumes; the dotted line marks the best finite width w^* with its classification.

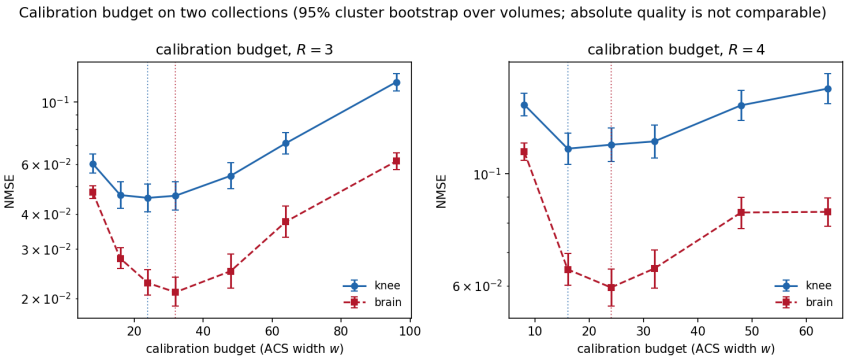

Figure 13: The two collections at the accelerations whose admissible range contains an interior optimum. Between these two accelerations, both optima move to a smaller ACS width; the brain optimum is one grid step wider than the knee optimum at each acceleration (dotted lines). Error bars are 95% cluster-bootstrap intervals over volumes; absolute quality is not comparable between the collections.