OUTPUT-WIDTH RATES FOR SPECTRAL PROJECTOR APPROXIMATION UNDER WASSERSTEIN SHIFT

Anonymous authors
Paper under double-blind review

ABSTRACT

A spectral projector may vary continuously even when every narrow continuous factor incurs order-one error somewhere. To the best of our knowledge, we establish the first sharp joint characterization of output-factor width, and Wasserstein robustness for spectral projector prediction on real Grassmann manifolds. For predictors $WW^\top$ on $\mathrm{Gr}(k, N)$, let W have m output columns, $\|W\|_F \le \sqrt{k}$, and $\mathrm{Lip}_F(W) \le L$. For $k \le m < N$, set $r = m - k + 1$. At fixed dimensions, in the large-L, small-ε regime, the optimal in-distribution squared operator-norm risk under Haar sampling is $\Theta(L^{-r})$, while the optimal risk over a W_p ball of radius ε is $\Theta(L^{-r} + \varepsilon^{pr/(p+r)})$ for $1 \le p < \infty$. Imposing a density-ratio cap $M \ge 1$ yields $\Theta(\min\{ML^{-r}, L^{-r} + \varepsilon^{pr/(p+r)}\})$. Matching lower bounds and explicit continuous constructions turn a topological existence obstruction into quantitative approximation and distribution-shift exponents. The lower bounds apply to every neural architecture satisfying the stated constraints; matching neural upper rates remain architecture-dependent. The exponents persist under input densities bounded above and away from zero, bi-Lipschitz coordinate changes, and uniformly gapped spectral deformations. Consequently, at fixed subthreshold output width, vanishing in-distribution error can coexist with an unavoidable robust-error floor at a fixed positive shift radius, while the density cap determines whether this floor persists. Frozen-backbone Fourier neural operator experiments on elliptic, Darcy, and plate problems support this distinction: widening only the output factor head substantially reduces targeted shifted error, whereas a sixteenfold increase in the size of a one-column plate model leaves its shifted-error floor essentially unchanged. Analytic rate tests and structural-mode experiments provide additional checks across ranks and discretizations.

1 INTRODUCTION

Neural operators can amortize repeated eigenspace computations in reduced models, subsurface flow, and structural dynamics (Kovachki et al., 2023; Azizzadenesheli et al., 2024; Trifonov et al., 2026; Rudikov et al., 2026). Let $A(\xi)$ be a parameter-dependent operator, and let the columns of $U_k(\xi)$ span an isolated k-dimensional eigenspace. Its spectral projector is

$$P(\xi) = U_k(\xi)U_k(\xi)^\top.$$

We approximate P with $S_W = WW^\top$, where $W(\xi) \in \mathbb{R}^{N\times m}$. The integer m is the number of factor columns. It is independent of the depth and hidden width of the network that produces W.

An eigengap makes P locally stable (Yu et al., 2014; Li et al., 2025), but it does not guarantee a continuous narrow factor. A factor may fit most of the input distribution and still incur order-one error on a small set. The size of that set is controlled by the Lipschitz budget, while its existence is controlled by the output representation. We quantify both effects and then ask how Wasserstein shifts can concentrate mass near the defect.

Setting and rates. The canonical domain is $X = \mathrm{Gr}(k, N)$ with Haar probability μ and the principal-angle metric. For $\ell_W(x) = \|W(x)W(x)^\top - P(x)\|_{\mathrm{op}}^2$, consider factors satisfying $\mathrm{Lip}(W) \le L$ and either $\|W(x)\|_F \le \sqrt{k}$ or $\|W(x)\|_F = \sqrt{k}$. If $k \le m < N$ and $r = m - k + 1$,

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.

we prove

$$\inf_{W} \int_X \ell_W \, d\mu \asymp L^{-r}, \tag{1}$$

$$\inf_{W} \sup_{\mathsf{W}_p(\mu,\nu)\le\varepsilon} \int_X \ell_W \, d\nu \asymp L^{-r} + \varepsilon^{pr/(p+r)}. \tag{2}$$

For a density cap $d\nu/d\mu \le M$, the second quantity becomes

$$\Theta\Big(\min\{ML^{-r}, L^{-r} + \varepsilon^{pr/(p+r)}\}\Big). \tag{3}$$

The constants are uniform for fixed dimensions and p, large L, small ε, and $M \ge 1$. The corresponding exponent for $p = \infty$ is r. At $m = N$, the choice $W = P$ is exact.

The lower bound restricts the target to a projective r-dimensional slice. On every such slice, a subthreshold factor loses a target direction at some point. Lipschitz continuity then spreads unit-scale error over mass $\Omega(L^{-r})$. A clipped polar factor confines the error to a set of the same order and yields the upper bounds. Moving mass from a ball of radius t to the defect costs order t^{p+r}, which gives the transport exponent $pr/(p+r)$. A density cap limits the amount of mass that can be moved and produces (3).

Relation to earlier works. Manifold interpolation and subspace regression model parameter-dependent spaces (Zimmermann, 2021; Bhattacharya et al., 2021; Trifonov et al., 2026; Lanthaler, 2023; Boullé & Townsend, 2024); local spectral learning and Rayleigh–Ritz methods compare or refine eigenspaces (Rudikov et al., 2026; Yang et al., 2026; Davis & Kahan, 1970; Kato, 1995; Stewart & Sun, 1990). Sign- and basis-invariant networks and neural eigensolvers address how spectral objects are encoded (Lim et al., 2023; Tahmasebi & Jegelka, 2025; Chen et al., 2022; Ryu et al., 2024; Li et al., 2024). Topological obstructions and frame constructions determine whether continuous representatives exist (Zhou et al., 2019; Esmaeili et al., 2023; Dym et al., 2024; Ballas et al., 2025; Heightman et al., 2026). Our rates add the approximation scale L, the number of output columns, and the transport constraint to this existence question. Neural-operator approximation theory establishes universality and architecture-dependent error bounds (Kovachki et al., 2021; 2023), while geometric and convolutional variants address irregular meshes and discretization effects (Lu et al., 2023; Helwig et al., 2023; Raonic et al., 2023). PDEBench provides standardized scientific-ML tasks for broader evaluation (Takamoto et al., 2022). Our rates complement perturbation results for gapped eigenspaces and Wasserstein risk bounds (Yu et al., 2014; Li et al., 2025; Mohajerin Esfahani & Kuhn, 2018; Gao & Kleywegt, 2023; Sinha et al., 2018; Bai et al., 2023; Lara Benitez et al., 2024; Mouli et al., 2024).

Consequences for neural operators. A in-distribution target error η requires $L = \Theta(\eta^{-1/r})$ in the continuous class. Adding one column below the exact threshold changes this to $\Theta(\eta^{-1/(r+1)})$. Increasing hidden capacity at fixed m may localize the defect but cannot remove it. At a fixed positive shift radius, the unrestricted shifted risk therefore retains a positive floor. With a density cap, it vanishes precisely when $M = o(L^r)$. Moreover, $\Theta(L^r)$ independent validation points may be needed to encounter the defect (Proposition 15).

Corollary 4 states the resulting lower bound for a frozen feature extractor followed by a pointwise factor head. We test this setting by changing only the last head of FNOs trained on a variable-coefficient elliptic problem, the standard random-permeability Darcy problem, and Kirchhoff–Love plate modes. In the elliptic and Darcy tests, $m = 3$ is subthreshold and $m = 4$ is exact; in the plate test the corresponding widths are one and two. We also enlarge a one-column plate FNO from 19,117 to 304,561 parameters. Held-out error decreases, but its targeted shifted error does not.

The theory also covers bounded-positive input densities, bi-Lipschitz coordinate changes, and uniformly gapped deformations. A separate result treats sign-reversing real eigenlines on closed parameter paths. Analytic experiments test the exponents across ranks, transport orders, norm constraints, and density caps. Plate and TU1402 experiments evaluate the closed-path result with numerical modes and certified Lipschitz budgets. These experiments concern continuous factor prediction; their feasible attacks are lower bounds on robust risk, and the neural upper rates remain architecture dependent.

2 APPROXIMATION AND DISTRIBUTIONAL ROBUSTNESS

Let $1 \le k < N$ and let $X = \mathrm{Gr}(k, N)$ carry Haar probability measure and the principal-angle geodesic metric. In this canonical model, $x \in X$ is itself a k-dimensional subspace and the target $P(x)$ is its orthogonal projector. For a physical parameter ξ and operator $A(\xi)$, the corresponding application is obtained through the map $\xi \mapsto P(\xi)$ onto a selected gapped eigenspace. We also use the computable projective subfamily, with $d = N - k$. Here $\mathbb{RP}^d = (\mathbb{R}^{d+1} \setminus \{0\})/\!\sim$, where $x \sim \lambda x$ for every nonzero scalar λ, and $[x]$ denotes the resulting line through x (equivalently, for unit vectors, $[x] = [-x]$). Thus

$$P_k([x]) = \mathrm{diag}(I_{k-1}, xx^\top), \qquad [x] \in \mathbb{RP}^d, \tag{4}$$

whose image bundle is $\underline{\mathbb{R}}^{k-1} \oplus \gamma$.

A neural predictor outputs a continuous factor $W : X \to \mathbb{R}^{N\times m}$ and predicts $S_W = WW^\top$. Define

$$\begin{aligned}
\mathcal{W}_{\le}(L) &= \left\{ W : \sup_x \|W(x)\|_F \le \sqrt{k},\ \mathrm{Lip}_F(W) \le L \right\}, \\
\mathcal{W}_{=}(L) &= \{ W \in \mathcal{W}_{\le}(L) : \|W(x)\|_F = \sqrt{k}\ \forall x \}, \\
\mathcal{R}(W) &= \int_X \|S_W(x) - P(x)\|_{\mathrm{op}}^2 \, d\mu(x).
\end{aligned} \tag{5}$$

The second class fixes $\mathrm{tr}\, S_W = k$. No block structure or pointwise rank condition is imposed. These are function classes, not a particular network architecture. Any factor network satisfying the same output, norm, and Lipschitz constraints is a member, so a lower bound for either class is also a neural approximation lower bound. Its continuous upper bound becomes a neural upper bound only after an architecture-specific approximation argument.

Theorem 1 (Best nominal approximation rate)**.** *Suppose $k \le m < N$ and set $r = m - k + 1$. Every continuous factor has $\sup_x \|S_W(x) - P(x)\|_{\mathrm{op}} \ge 1$, while*

$$\inf_{W \in \mathcal{W}_{\le}(L)} \mathcal{R}(W) \asymp L^{-r}, \qquad \inf_{W \in \mathcal{W}_{=}(L)} \mathcal{R}(W) \asymp L^{-r} \tag{6}$$

for sufficiently large L, with constants depending only on N, k, m. The same exponent holds for squared Frobenius risk. Across all widths, an exact continuous factor exists if and only if $m \ge N$; for $m \ge N$ one can choose an exact factor with Lipschitz constant at most $\sqrt{2}$.

The lower bound restricts the target projector family to a projective r-slice. On every rotated slice, the columns of PW must lose a target direction somewhere; otherwise their kernel would contradict the standard obstruction to a global frame. Lipschitz continuity spreads unit error over a region of probability order L^{-r}. More explicitly, if $h_W(x) = \|W(x)W(x)^\top - P(x)\|_{\mathrm{op}}$, then $h_W(z) \ge 1$ at an obstructed point and $\mathrm{Lip}(h_W) \le 1 + 2\sqrt{k}L$. Radial integration on the slice gives

$$\mathcal{R}(W) \ge \frac{1}{D_r} \int_0^{(1+2\sqrt{k}L)^{-1}} (1 - (1 + 2\sqrt{k}L)t)^2 \sin^{r-1} t \, dt \gtrsim L^{-r}. \tag{7}$$

For the matching upper bound, fix an orthonormal $N \times m$ matrix U and let $s(P) = \sigma_k(PU)$. Clipping the singular values of PU/δ at one produces an exact output whenever $s(P) \ge \delta$ and has Lipschitz constant $O(\delta^{-1})$. The fixed-dimensional small-ball bound, consistent with rectangular small-singular-value estimates (Rudelson & Vershynin, 2009),

$$\mu\{P : s(P) \le \delta\} \le D_{N,k,m}\delta^r \tag{8}$$

is the matching geometric fact: error is confined to a region of codimension r. Taking $\delta \asymp L^{-1}$ proves the upper rate. A globally nonvanishing normalization provides the same rate at fixed trace. The proof and constants are in Appendices A.3 and A.4.

2.1 WIDTH, TRANSPORT, AND DENSITY CONCENTRATION

Let $\mathcal{D}_{p,\varepsilon}(W)$ be the supremum of $\int \ell_W \, d\nu$ over the population Wasserstein ball $\mathsf{W}_p(\mu, \nu) \le \varepsilon$. For finite p, W_p^p minimizes $\int d(x, y)^p \, d\pi$ over couplings, while $p = \infty$ bounds their essential maximum

displacement. The label at a moved input y is its own projector $P(y)$. To control concentration, define

$$\mathcal{D}_{p,\varepsilon}^{(M)}(W) = \sup_{\substack{\mathsf{W}_p(\mu,\nu)\le\varepsilon \\ \nu\ll\mu,\ d\nu/d\mu\le M}} \int_X \ell_W \, d\nu, \qquad M \ge 1. \tag{9}$$

The cap is relative to the population input law, not an empirical sample.

Theorem 2 (Best approximation under Wasserstein shifts)**.** *Fix $k \le m < N$, set $r = m - k + 1$, and let $\mathcal{C}(L)$ be either factor class in (5). Define $a_p = pr/(p+r)$ for $1 \le p < \infty$ and $a_\infty = r$. Then*

$$\inf_{W\in\mathcal{C}(L)} \mathcal{D}_{p,\varepsilon}(W) \asymp L^{-r} + \varepsilon^{a_p}, \tag{10}$$

$$\inf_{W\in\mathcal{C}(L)} \mathcal{D}_{p,\varepsilon}^{(M)}(W) \asymp \min\{ML^{-r},\ L^{-r} + \varepsilon^{a_p}\}. \tag{11}$$

The constants are uniform for $L \ge L_0$, $0 \le \varepsilon \le \varepsilon_0$, and $M \ge 1$, at fixed dimensions and transport order. Both statements remain valid for smooth strictly positive shifted densities, with the same cap in (11). For $m \ge N$ and $L \ge \sqrt{2}$, both best errors are zero for every radius and cap.

Nonuniform input distributions. Exact Haar sampling is not necessary. If $\mu_f = f\mu$ with fixed $0 < c \le f \le C < \infty$, both rates hold with constants depending on c, C and with the cap measured relative to μ_f. Indeed, for every fixed factor, capped or uncapped risk under μ_f lies between c and C times its risk under μ at the same radius and cap. The optional smooth-density statement additionally assumes that f is smooth. Global bi-Lipschitz coordinate changes also preserve the exponents. Appendix B.6 proves these comparisons.

Corollary 3 (Nominal approximation and shifted error)**.** *At fixed sufficiently small $\varepsilon > 0$ and subthreshold width, nominal error tends to zero as $L \to \infty$, while the best unrestricted shifted error is bounded below by $c\varepsilon^{a_p} > 0$. In contrast, for a cap $M(L) \ge 1$, the best density-capped shifted error tends to zero if and only if $M(L) = o(L^r)$. In the shift-dominated regime $\varepsilon^{a_p} \gg L^{-r}$, the crossover to the unrestricted order is given by the balance*

$$M - 1 \asymp L^r \varepsilon^{a_p}, \qquad \min\{ML^{-r}, L^{-r} + \varepsilon^{a_p}\} = L^{-r} + \min\{(M-1)L^{-r}, \varepsilon^{a_p}\}. \tag{12}$$

These are best-approximation rates over the continuous factor class. For a neural subclass under the same constraints, their lower bounds still hold. Matching neural upper rates require the network to approximate the explicit factor with controlled Lipschitz constant; they do not follow from training a particular architecture. For finite p and desired shift exponent $0 < \beta < p$, solving $pr/(p+r) \ge \beta$ gives the smallest eligible subthreshold width

$$m_\beta = k - 1 + \left\lceil \frac{p\beta}{p-\beta} \right\rceil, \tag{13}$$

provided $m_\beta < N$; otherwise width N admits exact representation. The rule applies after verifying that the target family has the stated projective geometry.

Corollary 4 (Frozen neural-operator output heads)**.** *Let $\Xi = Y \times \mathbb{RP}^{d-1}$ carry a product law whose second factor is Haar, and let*

$$P(y,[x]) = Q(y)xx^\top Q(y)^\top, \qquad Q(y)^\top Q(y) = I_d,$$

with Q Lipschitz. Let an arbitrary continuous neural-operator feature extractor Z_ψ be frozen, and let a pointwise head H_m produce a factor $W_{\psi,m} = \mathrm{reshape}(H_m Z_\psi)$ with $\|W_{\psi,m}\|_F \le 1$ and $\mathrm{Lip}(W_{\psi,m}) \le L$. If $m < d$, every such head satisfies the nominal and finite-p shifted lower bounds

$$cL^{-m}, \qquad c\big(L^{-m} + \varepsilon^{pm/(p+m)}\big),$$

respectively, with c independent of the hidden width, Fourier modes and depth of Z_ψ. Width $m = d$ removes this representation obstruction; realization by a fixed feature extractor is architecture-specific.

Appendix B.1 gives the fiberwise proof. Experiment 2 isolates the conclusion by holding Z_ψ fixed and changing only H_m.

Proof idea. On each projective r-slice, the obstruction forces a unit-error point. Moving mass $\asymp t^r$ to it costs $O(t^{p+r})$, which gives the unrestricted exponent. A density cap limits the moved mass to $\min\{(M-1)L^{-r}, t^r\}$, while the same clipped factor attains the two upper bounds. Appendices B–B.5 give the measurable couplings and uniform constants.

Proposition 5 (Lipschitz orthogonal changes of basis)**.** *Let $U : X \to O(N)$ be T-Lipschitz and $\widetilde{P}(x) = U(x)P(x)U(x)^\top$. For either factor norm constraint, let $\mathcal{B}_P(L)$ denote the best nominal, unrestricted shifted, or density-capped shifted error at fixed (p, ε, M). Then*

$$\mathcal{B}_{\widetilde{P}}(L + \sqrt{k}T) \le \mathcal{B}_P(L), \qquad \mathcal{B}_P(L + \sqrt{k}T) \le \mathcal{B}_{\widetilde{P}}(L). \tag{14}$$

Thus the exponents and exact-width threshold are preserved.

Corollary 6 (Uniformly gapped spectral deformations)**.** *Let $A_t(x)$ be smooth and real symmetric on compact $X \times [0,1]$. Suppose a rank-k cluster has a uniform positive gap. If its initial projector P_0 satisfies the stated rates, its terminal projector P_1 has the same exponents and exact-width threshold. Constants may depend on the homotopy and gap.*

Closed parameter paths of real modes. On $S^1 \times Y$, where Y is a compact rectangle, a C^1 real rank-one projector with a sign-reversing fiber and a bounded-positive source density obeys (11) with $r = 1$ at width one; two columns factor it exactly (Theorem 18). The single-circle case needs no prescribed reference homotopy (Corollary 17). Interval eigengap and perturbation checks can verify the sign reversal for numerical modes (Proposition 16); sampled gaps alone cannot.

3 Experiments

The experiments comprise an analytic rate calculation, paired FNO head interventions, a learned elliptic operator, a Darcy solution operator, and two structural-mode operators. The analytic calculation tests the exponents. The paired models hold their learned features fixed while changing the number of factor columns. Dataset sizes and evaluation details are given in Appendices C and D.

Experimental setup and evaluation. Each task maps ξ to a rank-k projector $P_k(\xi)$ through a factor $W_\phi(\xi) \in \mathbb{R}^{N \times m}$, with the target fixed as m varies. Checkpoints use independent validation projector loss. We report held-out Error, Error under the specified shift, and fitted log–log orders. Except for the analytic W_∞ calculation, the reported shifts are feasible; random geodesic steps are law-preserving controls. Plate and TU1402 use certified Lipschitz budgets and a predeclared joint budget–radius schedule. Tables 1–2 give the main results. The appendix records the protocols and seed checks.

The pointwise Error and its worst-case expectation are

$$\begin{aligned} \widehat{P}_\phi(\xi) &= W_\phi(\xi)W_\phi(\xi)^\top, \qquad \ell_\phi(\xi) = \|\widehat{P}_\phi(\xi) - P_k(\xi)\|_{\mathrm{op}}^2, \\ \mathcal{D}_{p,\varepsilon}(W_\phi) &= \sup_{\mathsf{W}_p(\mu,\nu) \le \varepsilon} \int \ell_\phi \, d\nu. \end{aligned} \tag{15}$$

Held-out Error uses $\nu = \mu$; the numerical shifts use feasible couplings described below and in the appendices.

Experiment 1: Approximation rate and distribution shift. On $\mathbb{RP}^7$ with $k = 1$, deterministic quadrature evaluates the nominal squared-operator Error of the explicit clipped factors against their certified Lipschitz budgets $L_\delta = \sqrt{2}/\delta$ (Figure 3). The orders fitted directly from the quadrature values are $1.000, 2.000, 3.000, 4.000$ in $O(L^{-k})$, matching $k = r = m$ for widths $m = 1, 2, 3, 4$. These calculations test the constructive upper-bound family; the theorem supplies the lower bound for every admissible continuous factor. The separate Wasserstein experiment reports the fitted shift orders in Figure 4; the fixed-radius plateau in Figure 4 appears when the shift term dominates while nominal Error continues to decrease. Matrix-loss, transport-moment, geometry, and finite linear-program checks are independent.

The density-cap sweep in Appendix C.1 and Figure 5 isolates the effect of an input-density constraint. A rank-two calculation gives $\delta^2/3$ at width three, verifying $r = m - k + 1$ (Appendix C.2); a

Table 1: Neural-operator results for Experiments 2–4 at $\varepsilon = 0.05$. Errors are medians over five seeds. The fitted order is obtained from the recorded radius sweep and compared with the theoretical order $6/5$. The $m = 3$ and $m = 4$ heads share the same frozen FNO backbone.

problem	Error (m=3)	shifted Error (m=3)	shifted Error (m=4)	fitted order	theory
elliptic	1.49×10^{-5}	2.63×10^{-2}	1.34×10^{-5}	1.186	1.200
FNO N=1024	1.02×10^{-3}	2.67×10^{-2}	6.58×10^{-4}	1.271	1.200
FNO N=4096	8.14×10^{-4}	2.70×10^{-2}	3.95×10^{-4}	1.240	1.200
Darcy q=24	1.96×10^{-3}	2.82×10^{-2}	1.15×10^{-3}	1.334	1.200
Darcy q=32	1.92×10^{-3}	2.78×10^{-2}	1.29×10^{-3}	1.264	1.200

validation-power control is reported in Appendix B.7. Additional parameter-grid results are retained with the source package rather than repeated in the manuscript.

The explicit factor and its loss are

$$\begin{aligned} P([x]) = xx^\top, \qquad s = \|U^\top x\|_2, \qquad W_\delta([x]) = \frac{xx^\top U}{\max\{s,\delta\}}, \\ \ell_{W_\delta}([x]) = (1 - s^2/\delta^2)_+^2, \qquad \mathrm{Lip}(W_\delta) \le L_\delta = \frac{\sqrt{2}}{\delta}, \\ \mathcal{R}_\mu(W_\delta) \sim \frac{B(m/2,3)}{B(m/2,(8-m)/2)}\,\delta^m = C_m L_\delta^{-m}. \end{aligned} \tag{16}$$

Here $U = [e_1, \dots, e_m] \in \mathbb{R}^{8\times m}$ and x is a unit representative of $[x]$; the last line is the $\delta \downarrow 0$ asymptotic on the Haar law, with $C_m = 2^{m/2}B(m/2,3)/B(m/2,(8-m)/2)$.

Experiment 2: Output-head intervention in an FNO. For $[x] \in \mathbb{RP}^3$ and $\theta \in [-1,1]^{12}$, the variable-coefficient four-field Dirichlet problem is

$$\begin{aligned} \mathcal{A}_{\theta,[x]}u = &-\partial_s(a_\theta(s)\partial_s u) - \partial_t(b_\theta(t)\partial_t u) + (\alpha I_4 - xx^\top)u, \\ a_\theta(s) = 1 + 0.2\tanh&\left(\frac{1}{\sqrt{6}}\sum_{j=1}^{6}\theta_j \sin(j\pi s)\right), \\ b_\theta(t) = 1 + 0.2\tanh&\left(\frac{1}{\sqrt{6}}\sum_{j=1}^{6}\theta_{6+j} \sin(j\pi t)\right), \qquad P_1(\theta,[x]) = v_1 v_1^\top, \quad \mathcal{A}_{\theta,[x]}v_1 = \lambda_1 v_1. \end{aligned} \tag{17}$$

The finite-difference grids use $N = 4q^2$ unknowns, with $q = 16, 32$. A standard two-dimensional FNO takes the coefficient fields, projective parameter and coordinates as 14 input channels. Four Fourier layers produce a feature field Z_ψ; a pointwise linear map is the only width-dependent component:

$$\begin{aligned} C_{\theta,[x]}(s,t) = \big(a_\theta(s), b_\theta(t), \mathrm{vech}(xx^\top), s, t\big), \quad Z_\psi = \mathrm{FNO}_\psi(C_{\theta,[x]}), \\ W_{\psi,m} = \mathrm{reshape}(H_m Z_\psi)\,, \qquad H_m : \mathbb{R}^{32} \to \mathbb{R}^{4m}, \qquad \widehat{P}_{\psi,m} = W_{\psi,m}W_{\psi,m}^\top. \end{aligned} \tag{18}$$

For each seed, we train a width-four FNO, freeze its backbone, delete the fourth factor column to initialize the width-three head, and fine-tune only the two heads on identical minibatches. Their backbone tensors remain bitwise identical. The wider head adds 132 parameters (0.025% and 0.016% of the two models). Each resolution uses 10,000/2000/2000 samples, five seeds and three robust-evaluation coefficient fields per seed.

Held-out Error is below 1.31×10^{-3} in every run. At $\varepsilon = 0.05$, the wider head reduces median Shifted Error by $40.6\times$ and $68.5\times$ (Figure 1 and Table 1); all 30 paired comparisons favor it, by 18.8–$103.7\times$. The narrow-head exponents are 1.271 and 1.240, versus $6/5$, while median random-shift changes are 0.18% and 0.40%. Changing 132 output parameters removes the observed narrow-head floor on the shared features. Appendix D gives the all-seed ranges, search protocol, and residual checks; the attacks are feasible lower bounds.

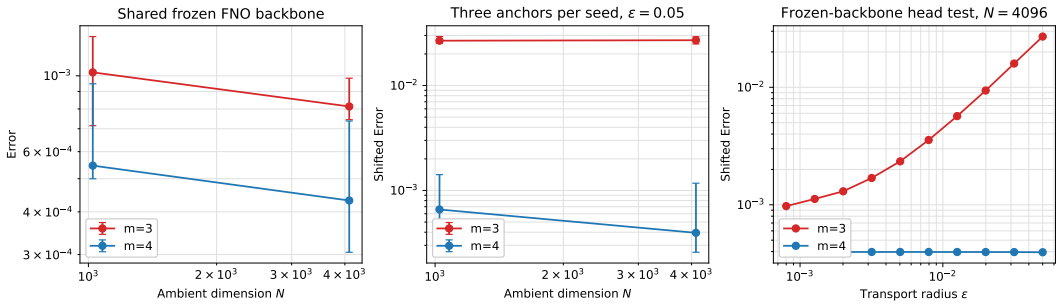

Figure 1: Experiment 2: frozen-backbone FNO output-head intervention over five seeds. Left: held-out Error. Center: targeted Shifted Error at $\varepsilon = 0.05$. Right: the radius sweep at $N = 4096$. Points are medians and bars span all seeds and coefficient anchors. Both heads use exactly the same learned and frozen FNO backbone; the only trainable difference is the final $4m$-channel pointwise map.

Table 2: Structural-mode results for Experiments 5–6. Plate uses $\varepsilon = 0.05$ and TU1402 uses 0.049. The fitted order is obtained from the recorded radius sweep and compared with the theoretical order $2/3$.

problem	Error (m=1)	shifted Error (m=1)	shifted Error (m=2)	fitted order	theory
plate $q = 8$	4.72×10^{-3}	1.47×10^{-1}	2.70×10^{-6}	0.666	0.667
plate $q = 12$	4.62×10^{-3}	1.42×10^{-1}	2.60×10^{-6}	0.662	0.667
TU1402	1.15×10^{-2}	8.99×10^{-2}	1.19×10^{-5}	0.664	0.667

Experiment 3: Elliptic spectral projectors. The input is $[x] \in \mathbb{RP}^3$, and the target is the lowest-eigenvalue projector of the four-field Dirichlet operator

$$\begin{aligned} \mathcal{A}_{[x]}u &= -u'' + [2I_4 - xx^\top + 0.2H(s)]u, \qquad [x] \in \mathbb{RP}^3, \\ u(0) &= u(1) = 0, \quad \mathcal{A}_{[x]}v_1 = \lambda_1 v_1, \qquad P_1([x]) = v_1 v_1^\top. \end{aligned} \tag{19}$$

where $H(s) = H(s)^\top$ and $\|H(s)\|_{\mathrm{op}} \le 1$. All three frozen narrow factors on (23) have refined unit-loss destinations. The common shift amplifies only narrow-factor Error, with a fitted Error-difference exponent near the best-approximation reference; Haar-preserving controls remain near nominal Error (Table 1).

Experiment 4: Darcy-flow solution projectors. Following the standard FNO Darcy benchmark (Li et al., 2021), draw $a \in \{3, 12\}$ by thresholding $g \sim \mathcal{N}(0, (-\Delta + 9I)^{-2})$ and solve

$$-\nabla\cdot(a\nabla s_a) = 1 \quad \text{in } (0,1)^2, \qquad s_a = 0 \quad \text{on } \partial(0,1)^2, \tag{20}$$

and learn the solution-subspace projector $vv^\top$, where $v = (s_a/\|s_a\|_2) \otimes x$ for $[x] \in \mathbb{RP}^3$. Each fixed-a fiber is an isometric projective target with exact factor width four. On $q = 24, 32$ grids and five seeds, the shared-backbone $m = 3, 4$ intervention adds only 132 parameters (0.011%). At $\varepsilon = 0.05$, narrow-head Shifted Error is 24.4 and 21.6 times the paired wide-head value; the $q = 32$ fitted exponent is 1.264 versus $6/5$ (Table 1). Appendix D gives the protocol, controls, and solver audit.

Experiment 5: Parametric plate modes. A simply supported Kirchhoff–Love plate on a varying Winkler foundation (Jiang et al., 2023) tests the transfer result with a different exact-width threshold. Along one closed parameter loop, a continuous unit representative of the second mode changes sign while its rank-one projector closes; the gap and doublet calculation are given in Appendix D. Hence a continuous width-one factor has a defect, whereas width two permits an exact factor. The labels are eigenvectors of the full Galerkin matrix rather than prescribed analytic modes, so this is not a manufactured-solution test.

The dimensionless plate operator and target second-mode projector are

$$\begin{aligned} A_\theta &= \pi^{-4}\Delta^2 + 10 + \tfrac{1}{2}V_\theta, && \theta \in \mathbb{R}/\pi\mathbb{Z}, \\ V_\theta(x,y) &= 2\cos(2\theta)[\cos(2\pi x) - \cos(2\pi y)] + 4\sin(2\theta)\cos(\pi x)\cos(\pi y), \\ A_\theta\varphi_2(\theta) &= \lambda_2(\theta)\varphi_2(\theta), \\ P_1(\theta) &= \varphi_2(\theta)\varphi_2(\theta)^\top, \end{aligned} \tag{21}$$

with $w = \Delta w = 0$ on the boundary; A_θ is assembled in the sine-Galerkin basis at $q = 8, 12$.

At $q = 12$, a two-dimensional FNO takes the foundation field and coordinates and predicts the mode factor. We train $m = 2$ to learn a feature map, freeze the backbone, discard its trained head, and train fresh $m = 1$ and $m = 2$ heads with an identical first-column initialization and identical minibatches. The extra column adds 25 parameters (0.033%). Across five seeds, median held-out Error falls from 2.49×10^{-2} to 8.18×10^{-3}; at $\varepsilon = 0.05$, median Shifted Error falls from 1.52×10^{-1} to 8.54×10^{-3}, and all five pairs favor $m = 2$. As a separate control, end-to-end narrow FNOs with 19,117–304,561 parameters reduce median held-out Error from 2.47×10^{-3} to 4.17×10^{-4}, while median Shifted Error remains in $[0.145, 0.147]$ (Figure 2). Separate certified-budget ReLU tests recover the L^{-1} and $2/3$ rates over 20 seeds (Table 3); full controls are in the appendix.

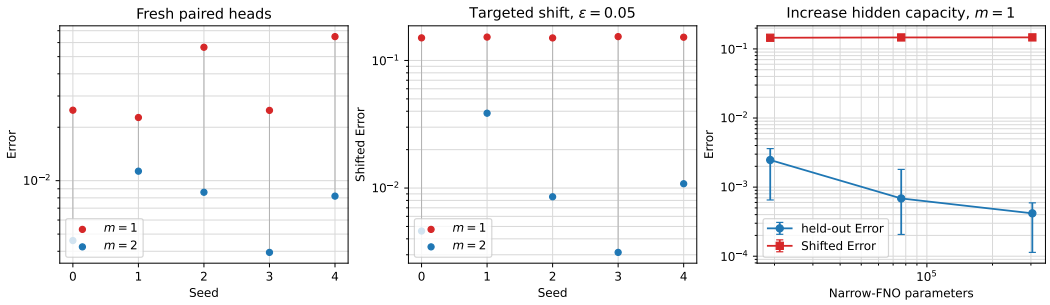

Figure 2: Experiment 5: controlled FNO interventions on numerical plate modes. Left and center: fresh paired heads on the same frozen features; vertical lines join the same seed. Right: increasing hidden capacity at fixed $m = 1$ lowers held-out Error but leaves the targeted Shifted Error unchanged. Error bars give the full five-seed range.

Experiment 6: Open structural-dynamics benchmark. The public COST TU1402 virtual bridge (Tatsis & Chatzi, 2019) supplies a finite-element modal task. At each parameter value it solves a generalized eigenproblem and predicts the first-mode projector on 120 sensors:

$$\begin{gathered} K(\theta,\rho)\phi_1 = \omega_1^2 M\phi_1, \qquad z_1 = \frac{S\phi_1}{\|S\phi_1\|_2}, \qquad P_1(\theta,\rho) = z_1 z_1^\top, \\ k_{\mathrm{mid}} = k_\star e^{0.04\rho\cos\theta}, \\ k_{\mathrm{left}} = 10^{10} e^{0.50\rho\sin\theta}, \qquad k_{\mathrm{right}} = 10^{10} e^{-0.50\rho\sin\theta}, \\ (\theta,\rho) \in S^1 \times [0.8, 1.2]. \end{gathered} \tag{22}$$

Here K and M are the assembled stiffness and consistent mass matrices, and S restricts displacement to the sensor degrees of freedom. The loop fixes $\rho = 1$; varying ρ gives the two-parameter family covered by Theorem 18. Interval bounds at 306 full-mode anchors support a continuous eigengap and nonzero sensor projection. All five initial and all 25 fresh narrow factors have refined unit-loss roots; feasible continuous-source Error is at least 0.09125 and 0.09066, respectively. Table 3 tests the L^{-1} nominal law and the joint budget–radius law over 20 seed groups. Appendix D gives the population coupling, two-parameter, cut-chart, and seed controls. These results concern the fixed loop, not a certified L rate over the full annulus.

Table 3: Certified-budget tests in Experiments 5–6 for randomly initialized width-one factors. Values are 20-seed median Error exponents and full ranges, fitted over all four budgets. The joint schedule is $\varepsilon = 0.25L^{-3/2}$.

physical task	Error exponent k vs. L	Shifted-Error exponent k vs. ε
Plate	$-0.983\,[-0.990, -0.975]$	$0.663\,[0.661, 0.665]$
TU1402	$-0.978\,[-0.982, -0.954]$	$0.664\,[0.662, 0.664]$
reference	-1	$2/3$

4 Discussion and limitations

The rates separate two model choices. Hidden capacity controls how sharply a factor changes; output width controls whether a continuous factor exists. A loop or chart audit can therefore be used before training: if m columns do not admit a continuous factor, one can increase m, predict the projector directly, or divide the parameter domain into charts. The paired FNO results show the effect of the first option while keeping the learned features fixed.

The conclusions require a verified target geometry and fixed ambient dimensions. The continuous upper construction does not by itself give a rate for a specified neural architecture. Likewise, the numerical attacks give feasible lower bounds rather than exact robust risks. Output width, the parameter domain, and targeted shift diagnostics should therefore be reported together with average error.

REPRODUCIBILITY STATEMENT

The accompanying implementation records model configurations, seeds, checkpoints, independent-test metrics, active-search coordinates, and double-precision refinement results. All reported numerical values come from those saved outputs. Figure export code and the underlying result tables accompany the source package. The proofs and numerical protocol are stated explicitly, including the matched-width, random-shift, and coefficient-extrapolation controls used by the representation diagnostics. No additional training results are inferred from the analytic constructions.

AI USE STATEMENT

Generative AI tools, accessed through OpenAI Codex and ChatGPT, were used for checking of proof arguments, coding and literature retrieval. The human authors take full responsibility for the correctness of all claims, proofs, references, implementations, figures, and reported results in the submitted manuscript.

REFERENCES

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, 6(5):320–328, 2024. doi: 10.1038/s42254-024-00712-5. URL https://doi.org/10.1038/s42254-024-00712-5.

Xingjian Bai, Guangyi He, Yifan Jiang, and Jan Obłój. Wasserstein distributional robustness of neural networks. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/53be3798fcc46e68ca0819c29a004652-Abstract-Conference.html.

Samuel A. Ballas, Tom Needham, and Clayton Shonkwiler. On the existence of Parseval frames for vector bundles. *Transactions of the American Mathematical Society, Series B*, 12:395–416, 2025. doi: 10.1090/btran/223. URL https://www.math.colostate.edu/~clayton/research/papers/framebundles.pdf.

Michael V. Berry. Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London. Series A*, 392(1802):45–57, 1984. doi: 10.1098/rspa.1984.0023.

Kaushik Bhattacharya, Benedikt Hosseini, Nikola B. Kovachki, and Andrew M. Stuart. Model reduction and neural networks for parametric PDEs. *SMAI Journal of Computational Mathematics*, 7:121–157, 2021. doi: 10.5802/smai-jcm.74. URL https://doi.org/10.5802/smai-jcm.74.

Nicolas Boullé and Alex Townsend. A mathematical guide to operator learning. In *Handbook of Numerical Analysis*, volume 25, pp. 83–125. Elsevier, 2024. doi: 10.1016/bs.hna.2024.05.003. URL https://doi.org/10.1016/bs.hna.2024.05.003.

Ziyu Chen, Yingzhou Li, and Xiuyuan Cheng. SpecNet2: Orthogonalization-free spectral embedding by neural networks. In *Proceedings of Mathematical and Scientific Machine Learning*, volume 190 of *Proceedings of Machine Learning Research*, pp. 33–48, 2022. URL https://proceedings.mlr.press/v190/chen22a.html.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi: 10.1137/0707001.

Nadav Dym, Hannah Lawrence, and Jonathan W. Siegel. Equivariant frames and the impossibility of continuous canonicalization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12228–12267. PMLR, 2024. URL https://proceedings.mlr.press/v235/dym24a.html.

Babak Esmaeili, Robin Walters, Heiko Zimmermann, and Jan-Willem van de Meent. Topological obstructions and how to avoid them. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/1c12ccfc7720f6b680edea17300bfc2b-Abstract.html.

Rui Gao and Anton J. Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *Mathematics of Operations Research*, 48(2):603–655, 2023. doi: 10.1287/moor.2022.1275. URL https://arxiv.org/abs/1604.02199.

Allen Hatcher. Vector bundles and K-theory, 2017. URL https://pi.math.cornell.edu/~hatcher/VBKT/VB.pdf. Version 2.2, November 2017.

Timothy Heightman, Elena Orlova, Philip Mantrov, and Aleksei Ustimenko. Topology obstructs pure foundation neural quantum states, 2026. URL https://arxiv.org/abs/2609.07591.

Jacob Helwig, Xuan Zhang, Cong Fu, Jerry Kurtin, Stephan Wojtowytsch, and Shuiwang Ji. Group equivariant fourier neural operators for partial differential equations. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12907–12930. PMLR, 2023. URL https://proceedings.mlr.press/v202/helwig23a.html.

Hao-Jie Jiang, Tong-Bo Chen, Yu-Xiang Ren, and Ning-Hua Gao. Analytical solutions of free vibration for rectangular thin plate and right-angle triangle plate on the Winkler elastic foundation based on the symplectic superposition method. *Journal of Mechanics*, 39:395–415, 2023. doi: 10.1093/jom/ufad032. URL https://academic.oup.com/jom/article/doi/10.1093/jom/ufad032/7395018.

Tosio Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer, 2 edition, 1995. doi: 10.1007/978-3-642-66282-9. URL https://doi.org/10.1007/978-3-642-66282-9.

Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On universal approximation and error bounds for fourier neural operators. *Journal of Machine Learning Research*, 22(290):1–76, 2021. URL https://jmlr.org/papers/v22/21-0806.html.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://www.jmlr.org/papers/v24/21-1524.html.

Samuel Lanthaler. Operator learning with PCA-Net: Upper and lower complexity bounds. *Journal of Machine Learning Research*, 24(318):1–67, 2023. URL https://jmlr.org/papers/v24/23-0478.html.

Jose Antonio Lara Benitez, Takashi Furuya, Florian Faucher, Anastasis Kratsios, Xavier Tricoche, and Maarten V. de Hoop. Out-of-distributional risk bounds for neural operators with applications to the helmholtz equation. *Journal of Computational Physics*, 513:113168, 2024. doi: 10.1016/j.jcp.2024.113168. URL https://doi.org/10.1016/j.jcp.2024.113168.

H. Li, J. Sun, and Z. Zhang. Deep eigenspace network for parametric non-self-adjoint eigenvalue problems, 2025. URL https://arxiv.org/abs/2512.20058.

Tianbo Li, Zekun Shi, Jiaxi Zhao, and Min Lin. Amortized eigendecomposition for neural networks. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-4404. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/fa9c7d0f08785e607b7eacac49805d6e-Abstract-Conference.html.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Derek Lim, Joshua David Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Q-UHqMorzil.

Lu Lu, Pengzhan Jin, and George Em Karniadakis. Geometric fourier neural operator. *Journal of Machine Learning Research*, 24:1–26, 2023. URL https://jmlr.org/papers/v24/23-0064.html.

Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1–2):115–166, 2018. doi: 10.1007/s10107-017-1172-1.

S. Chandra Mouli, Danielle C. Maddix, Shima Alizadeh, Gaurav Gupta, Andrew Stuart, Michael W. Mahoney, and Bernie Wang. Using uncertainty quantification to characterize and improve out-of-domain learning for PDEs. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 36372–36418. PMLR, 2024. URL https://proceedings.mlr.press/v235/mouli24a.html.

Bogdan Raonic, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of PDEs. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-3376. URL https://papers.neurips.cc/paper_files/paper/2023/hash/f3c1951b34f7f55ffaecada7fde6bd5a-Abstract-Conference.html.

Mark Rudelson and Roman Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, 62(12):1707–1739, 2009. doi: 10.1002/cpa.20294.

Alexander Rudikov, Vladimir Fanaskov, Sergei Stepanov, Buzheng Shan, Ekaterina Muravleva, Yalchin Efendiev, and Ivan Oseledets. Locally subspace-informed neural operators for efficient multiscale pde solving. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=RqXxkiCYip.

Jongha Jon Ryu, Xiangxiang Xu, Hasan Sabri Melihcan Erol, Yuheng Bu, Lizhong Zheng, and Gregory W. Wornell. Operator SVD with neural networks via nested low-rank approximation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 42870–42905. PMLR, 2024. URL https://proceedings.mlr.press/v235/ryu24b.html.

Aman Sinha, Hongseok Namkoong, and John C. Duchi. Certifying some distributional robustness with principled adversarial training. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=Hk6kPgZA-.

G. W. Stewart and Ji-guang Sun. *Matrix Perturbation Theory*. Computer Science and Scientific Computing. Academic Press, Boston, 1990. ISBN 9780126702309.

Behrooz Tahmasebi and Stefanie Jegelka. Generalization bounds for canonicalization: A comparative study with group averaging. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/b36dc39b319ba6ba2a0fd7601951efb4-Abstract-Conference.html.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. PDEBench: An extensive benchmark for scientific machine learning. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0117. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/0a9747136d411fb83f0cf81820d44afb-Abstract-Datasets_and_Benchmarks.html.

Konstantinos Tatsis and Eleni Chatzi. A numerical benchmark for system identification under operational and environmental variability. In *Proceedings of the 8th International Operational Modal Analysis Conference*, pp. 101–106, Copenhagen, Denmark, 2019. URL https://www.research-collection.ethz.ch/handle/20.500.11850/383295.

Vladislav Trifonov, Alexander Rudikov, Vladimir Fanaskov, Ekaterina Muravleva, and Ivan Oseledets. Deep learning for subspace regression. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=HF60Lu1Maj.

Zherui Yang, Tao Du, and Ligang Liu. Learning Laplacian eigenspace with mass-aware neural operators on point clouds. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference*, pp. 6:1–6:11, 2026. doi: 10.1145/3799902.3811185. URL https://arxiv.org/abs/2605.24390.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians, 2014. URL https://arxiv.org/abs/1405.0680.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5745–5753, 2019.

Ralf Zimmermann. Manifold interpolation. In *Model Order Reduction, Volume 1: System- and Data-Driven Methods and Algorithms*, pp. 229–274. De Gruyter, Berlin and Boston, 2021. doi: 10.1515/9783110498967-007. URL https://arxiv.org/abs/1902.06502.

A PROOFS AND COMPLEMENTARY RESULTS

A.1 TWO EIGENSOLVER-DEFINED REALIZATIONS

The first learned experiment uses a four-field Dirichlet eigenproblem

$$\mathcal{A}_{[x]}u = -u'' + [2I - xx^\top + 0.2H(s)]u, \qquad [x] \in \mathbb{RP}^3, \tag{23}$$

where $H(s)$ is symmetric and has operator norm at most one. Its 24×24 Galerkin discretization is coercive, its lowest eigenvalue is simple with gap at least 0.6, and its eigenline bundle is isomorphic to the tautological line. Hence exact representation requires at least four output columns. Proposition 7 and Appendix A.5 give the verification. This realizes the same obstruction in a well-conditioned numerical spectral problem; it does not claim an advantage over a classical eigensolver.

The second uses a nonlinear weighted graph family. Let $J \in \mathbb{R}^{5\times 4}$ have orthonormal columns spanning $\mathbf{1}^\perp$ and set $Q_0([x]) = Jxx^\top J^\top$. On the complete graph, define

$$w_{ij}([x]) = 1 + (Q_0([x]))_{ij} + \eta \sin(3(Q_0([x]))_{ij} + \phi_{ij}), \qquad \eta = 0.03, \tag{24}$$

with fixed phases, and let the target be the projector onto the smallest nonconstant Laplacian eigenvector. At $\eta = 0$ the Laplacian is $5I - \mathbf{1}\mathbf{1}^\top - Q_0$: the target eigenvalue is 4, the remaining nonconstant eigenvalues are 5, and the target line is tautological. The perturbation has operator norm at most $2(5-1)\eta$, so the target gap along the homotopy is at least $1 - 4(5-1)\eta = 0.52$; all edge weights are at least $0.5 - \eta = 0.47$. Proposition 8 proves that this target requires four output columns for exact representation and inherits the same nominal and robust exponents through Proposition 5.

A.2 ELLIPTIC REALIZATION STATEMENT

Proposition 7 (Coupled elliptic realization)**.** *Consider the spectral-Galerkin discretization of*

$$\mathcal{A}_{[x]}u = -u'' + [\alpha I - \beta xx^\top + \varepsilon H(s)]u, \qquad [x] \in \mathbb{RP}^3,$$

with Dirichlet boundary conditions, where $H(s) = H(s)^\top$ and $\operatorname{ess\,sup}_s \|H(s)\|_{\mathrm{op}} \le 1$. Let $g_0 = \min\{\beta, 3\pi^2\}$. If $2\varepsilon < g_0$ and $\pi^2 + \alpha - \beta - \varepsilon > 0$, then the lowest Galerkin eigenvalue is simple with gap at least $g_0 - 2\varepsilon$, and its eigenline bundle is isomorphic to the tautological line. Consequently its exact continuous output width for exact representation is four. For the experiment's parameters $(\alpha, \beta, \varepsilon) = (2, 1, 0.2)$, the operator is coercive by more than 10.66 and the target gap is at least 0.6.

Proposition 8 (Nonlinear graph realization)**.** *For the five-node graph family in (24), more generally with $0 \le \eta < 1/16$, every edge weight is positive and the smallest nonconstant Laplacian eigenvalue is simple with gap at least $1 - 16\eta$. Its eigenline bundle is smoothly orthogonally transported from the tautological line over $\mathbb{RP}^3$. Consequently its exact continuous output width for exact representation is four, and the sharp nominal and Wasserstein-robust exponents for widths $m = 1, 2, 3$ are those of the projective target. At the experimental value $\eta = 0.03$, the certified edge-weight and gap lower bounds are 0.47 and 0.52.*

A.3 MATCHING RATES ON THE FULL REAL GRASSMANN MANIFOLD

Let $X = \mathrm{Gr}(k, N)$ with its invariant probability measure. The distance is the Euclidean norm of the vector of principal angles. Consequently,

$$\|P(x) - P(y)\|_{\mathrm{op}} \le \operatorname{dist}(x, y), \qquad \|P(x) - P(y)\|_F \le \sqrt{2}\operatorname{dist}(x, y).$$

All dimensions are fixed as $L \to \infty$.

Lower bound and exact representation. Fix a $(k-1)$-plane F and an $(r+1)$-plane $V \subset F^\perp$, where $r = m - k + 1 \le N - k$. The map $[x] \mapsto F \oplus \operatorname{span}\{x\}$ embeds $\mathbb{RP}(V)$ isometrically into X. The tautological bundle restricts to $\underline{\mathbb{R}}^{k-1} \oplus \gamma$, so its inverse class in degree r is $\alpha^r \neq 0$. Every rotated slice has the same property. Haar-rotate the slice by $O(N)$. A uniform point on the random slice is uniform on X, since its law is invariant and $O(N)$ acts transitively. The missing-direction argument and radial integral in Proposition 11 therefore give exactly the lower constants in (6) on X. In particular no continuous exact factor has $k \le m < N$ columns. For $m < k$ this follows pointwise from rank. For $m \ge N$, $W = [P, 0]$ is exact, has Frobenius norm $\sqrt{k}$, and is $\sqrt{2}$-Lipschitz.

An elementary small-singular-value bound. Fix $U \in \mathbb{R}^{N\times m}$ with $U^\top U = I_m$, and set $s(P) = \sigma_k(PU)$, with singular values in decreasing order. For $k \le m < N$, let

$$r = m - k + 1, \qquad a = r/2, \qquad b = (Nk - r)/2 > 0.$$

For $0 < \varepsilon \le 1/\sqrt{k}$ we claim

$$\mathbb{P}\{s(P) \le \varepsilon\} \le k I_{k\varepsilon^2}(a, b) \le D_{N,k,m}\varepsilon^r, \quad D_{N,k,m} = k^{1+r/2}\frac{B(a, \min\{b, 1\})}{B(a, b)}. \tag{25}$$

The last bound also holds for $1/\sqrt{k} < \varepsilon \le 1$, because $D_{N,k,m}\varepsilon^r \ge k \ge 1$ there.

By rotational invariance take U to be the first m coordinate vectors. Let $G \in \mathbb{R}^{N\times k}$ have independent standard normal entries, and write $Z = G(G^\top G)^{-1/2}$. Then $P = ZZ^\top$ has the required Haar law. If A comprises the first m rows of G, the nonzero singular values of PU are those of $A(G^\top G)^{-1/2}$. Thus

$$s(P) \le \varepsilon \implies \sigma_k(A) \le \varepsilon\|G\|_{\mathrm{op}} \le \varepsilon\|G\|_F.$$

Indeed, apply A to the normalized vector proportional to $(G^\top G)^{-1/2}v$, where v minimizes the transformed singular value. Choose a unit minimizing vector for A, and a coordinate j of magnitude at least $1/\sqrt{k}$. Projecting Av orthogonally to the span of the other $k-1$ columns shows that

$$d_j := \mathrm{dist}(a_j, \mathrm{span}\{a_i : i \ne j\}) \le \sqrt{k}\,\varepsilon\|G\|_F.$$

Conditional on the other columns, the orthogonal component of a_j has $r = m - k + 1$ independent normal coordinates. Its squared norm is χ^2_r and is independent of both its parallel component and all the other entries. The conditional distributions of these component norms do not depend on the orientation of the span. Therefore d_j^2 and $\|G\|_F^2 - d_j^2$ are independent chi-squared variables with r and $Nk - r$ degrees of freedom. It follows that $d_j^2/\|G\|_F^2 \sim \mathrm{Beta}(a, b)$. A union bound over the k columns proves the first inequality. For the second, substitute $t = zv$ in the beta integral and use $(1 - zv)^{b-1} \le 1$ if $b \ge 1$, or $(1 - zv)^{b-1} \le (1 - v)^{b-1}$ otherwise. The column-distance method and rectangular singular-value exponents are classical (Rudelson & Vershynin, 2009). The argument above provides the fixed-dimensional bound needed here without an asymptotic random-matrix limit or an additive exceptional-probability term.

Clipped polar factors. Let $\mathcal{B}_{\mathrm{op}}$ be the operator-norm unit ball in $\mathbb{R}^{N\times m}$. Define

$$F_\varepsilon(P) = \mathrm{Proj}_{\mathcal{B}_{\mathrm{op}}}(PU/\varepsilon), \tag{26}$$

where the projection uses the Frobenius metric. It clips each singular value of PU/ε at one. This follows by aligning singular vectors and minimizing the sum of squared singular-value differences. As a Hilbert-space projection onto a closed convex set, it is nonexpansive: the two projection inequalities, added together, give $\|\mathrm{Proj}(A) - \mathrm{Proj}(B)\|_F^2 \le \langle A - B, \mathrm{Proj}(A) - \mathrm{Proj}(B)\rangle_F$. Consequently,

$$\|F_\varepsilon\|_F \le \sqrt{k}, \quad \mathrm{Lip}_F(F_\varepsilon) \le \sqrt{2}/\varepsilon, \quad \|F_\varepsilon F_\varepsilon^\top - P\|_{\mathrm{op}} = (1 - s(P)^2/\varepsilon^2)_+.$$

These identities remain valid at a rank-deficient PU. The risk is at most $\mathbb{P}\{s(P) < \varepsilon\}$, so (25) gives the upper exponent r.

A globally nonvanishing normalization. For $0 < \delta \le 1$, set

$$\begin{aligned} t(P) &= \min\{1, s(P)/\delta\}, \\ A_\delta(P) &= F_\delta(P) + (1 - t(P))\sqrt{k/m}\,U, \\ \overline{F}_\delta(P) &= \sqrt{k}\,A_\delta(P)/\|A_\delta(P)\|_F. \end{aligned} \tag{27}$$

The inner product $\langle F_\delta, U\rangle_F$ equals $\langle F_\delta, PU\rangle_F \ge 0$, by the aligned singular vectors. Each of the first k clipped singular values is at least t, including the case $s = 0$. Hence

$$\|A_\delta\|_F^2 \ge k\{t^2 + (1 - t)^2\} \ge k/2.$$

The normalization has no singularity anywhere on X. Singular-value perturbation gives $\mathrm{Lip}(t) \le 1/\delta$. Combining this with (39) yields

$$\|\overline{F}_\delta\|_F = \sqrt{k}, \qquad \mathrm{Lip}_F(\overline{F}_\delta) \le B_k/\delta, \qquad B_k = 2 + \sqrt{2k}. \tag{28}$$

For $s \geq \delta$, the added term vanishes and the factor is exact. For $s < \delta$, the difference of positive semidefinite matrices with trace k has operator norm at most k. Thus

$$\mathcal{R}(\overline{F}_\delta) \leq k^2 D_{N,k,m} \delta^r.$$

This construction is allowed to increase the output rank above k in the exceptional region, up to the permitted output width m.

Proof of Theorem 1 on the full Grassmann manifold. Combine the slice lower bound with the constructions above, taking $\varepsilon = \sqrt{2}/L$ and $\delta = B_k/L$. One may take $C_{N,k,m} = 2^{r/2} D_{N,k,m}$ and $\overline{C}_{N,k,m} = k^2 B_k^r D_{N,k,m}$. For squared Frobenius risk the lower bound is unchanged. The clipped construction has squared Frobenius error at most k on its exceptional set, and the normalized construction has error at most $2k^2$ there. The same exponent follows. The earlier exact-representation argument completes the proof. Appendix A.4 gives the projective-subfamily proof and sharper constants for that special case. □

A.4 Arbitrary coupled factors and the rank-k rate

We use the canonical metric and probability measure specified in the main text. In particular, a projective line has circumference π. The constants below depend on this metric convention. Write $P = P_k$, $S = WW^\top$, and $h_W(p) = \|S(p) - P(p)\|_{\mathrm{op}}$.

Lemma 9 (A missing target direction on every projective slice)**.** *Let $k \leq m < k + d$, and set $r = m - k + 1$. For any continuous $W : \mathbb{RP}^d \to \mathbb{R}^{(k+d)\times m}$ and any $(r+1)$-dimensional subspace $V \subset \mathbb{R}^{d+1}$, there are $p \in \mathbb{RP}(V)$ and a unit $v \in \operatorname{im} P(p)$ such that $W(p)^\top v = 0$. In particular $h_W(p) \geq 1$.*

Proof. Let $Y = \mathbb{RP}(V) \cong \mathbb{RP}^r$. The image bundle of $P|_Y$ is $E = \underline{\mathbb{R}}^{k-1} \oplus \gamma$, where γ denotes the tautological line bundle. Suppose for contradiction that $T(p) = P(p)W(p)$ has rank k at every point of Y. It then defines a surjection of vector bundles

$$T : \underline{\mathbb{R}}^m \longrightarrow E.$$

Its kernel is a vector bundle K of rank $m - k = r - 1$, and an orthogonal splitting gives $\underline{\mathbb{R}}^m \cong E \oplus K$. The kernel bundle is used only under this constant-rank assumption. For an explicit continuity check, the matrix

$$B(p) = T(p)T(p)^\top + I - P(p)$$

is invertible: it is positive definite on $\operatorname{im} P(p)$ and is the identity on its orthogonal complement. Thus the kernel projector

$$I_m - T(p)^\top B(p)^{-1} T(p)$$

is continuous.

The classical cohomology and characteristic-class identities (Hatcher, 2017) are

$$H^*(\mathbb{RP}^r; \mathbb{F}_2) = \mathbb{F}_2[\alpha]/(\alpha^{r+1}), \qquad w(\gamma) = 1 + \alpha, \qquad w(E \oplus K) = w(E)w(K).$$

The splitting would therefore require

$$w(K) = (1 + \alpha)^{-1} = 1 + \alpha + \cdots + \alpha^r. \tag{29}$$

This implies $w_r(K) = \alpha^r \neq 0$, whereas a rank-$(r-1)$ bundle has $w_r(K) = 0$. The contradiction proves that T loses rank somewhere. At that point choose a unit target vector v orthogonal to $\operatorname{im} T$. Then $W^\top v = W^\top P v = T^\top v = 0$, and

$$h_W(p) \geq |v^\top (S - P) v| = 1.$$

□

Lemma 10 (Lipschitz bounds for the observable error)**.** *For $W \in \mathcal{W}_{\leq}(L)$, $\operatorname{Lip}(h_W) \leq 1 + 2\sqrt{k}L$. For $W \in \mathcal{W}_{=}(L)$,*

$$\operatorname{Lip}_{\mathrm{op}}(WW^\top) \leq \sqrt{k}L, \qquad \operatorname{Lip}(h_W) \leq 1 + \sqrt{k}L.$$

Proof. The projector identity

$$\|P(p) - P(q)\|_{\mathrm{op}} = \sin \operatorname{dist}(p, q) \le \operatorname{dist}(p, q)$$

holds because only the last rank-one block changes. The product difference bound gives

$$\|W(p)W(p)^\top - W(q)W(q)^\top\|_{\mathrm{op}} \le (\|W(p)\|_{\mathrm{op}} + \|W(q)\|_{\mathrm{op}})\|W(p) - W(q)\|_{\mathrm{op}},$$

which proves the first assertion.

For the improved bound, restrict W to a unit-speed geodesic. This restriction is absolutely continuous, with derivative Z almost everywhere satisfying $\|Z\|_F \le L$ and $\langle W, Z\rangle_F = 0$. For a unit vector v, set $a = \|W^\top v\|^2 \in [0, k]$. Since $\|W\|_F^2 = k$,

$$\left\| vv^\top W - \frac{a}{k} W \right\|_F^2 = a - \frac{a^2}{k}.$$

Consequently,

$$\begin{aligned} |v^\top (WW^\top)' v| &= 2 \left| \left\langle vv^\top W - \frac{a}{k} W, Z \right\rangle_F \right| \\ &\le 2\sqrt{a - a^2/k}\, \|Z\|_F \le \sqrt{k}\, \|Z\|_F. \end{aligned}$$

Taking the supremum over unit v, integrating along a shortest geodesic, and using the reverse triangle inequality for the error norm proves the claim. This improvement requires constant Frobenius norm. □

Proposition 11 (Explicit lower bound and error tails)**.** *Define*

$$D_r = \int_0^{\pi/2} \sin^{r-1} t \, dt, \qquad c_r = \frac{2(2/\pi)^{r-1}}{r(r+1)(r+2)D_r}. \tag{30}$$

In Lemma 9, suppose h_W is K-Lipschitz with $K \ge 1$. Then

$$\mathcal{R}(W) \ge \frac{1}{D_r} \int_0^{1/K} (1 - Kt)^2 \sin^{r-1} t \, dt \ge c_r K^{-r}, \tag{31}$$

$$\mu_d\{h_W \ge \tau\} \ge \frac{1}{D_r} \int_0^{(1-\tau)/K} \sin^{r-1} t \, dt, \qquad 0 \le \tau < 1. \tag{32}$$

Proof. On each projective r-slice, Lemma 9 supplies a point p_0 with $h_W(p_0) \ge 1$. Thus

$$h_W(p) \ge (1 - K \operatorname{dist}(p, p_0))_+.$$

The radial probability density around a point of $\mathbb{RP}^r$ is $\sin^{r-1} t / D_r$ for $0 \le t \le \pi/2$. Since $1/K \le 1 < \pi/2$, integration gives the first bound on each slice. Using $\sin t \ge 2t/\pi$ on that interval and substituting $s = Kt$ gives

$$\frac{(2/\pi)^{r-1}}{D_r} K^{-r} \int_0^1 (1 - s)^2 s^{r-1} \, ds = c_r K^{-r}.$$

The level-set bound follows from the ball of radius $(1 - \tau)/K$.

To obtain the ambient result, fix a standard projective r-slice and rotate it by a Haar-uniform element of $O(d+1)$. A uniform point on the rotated slice has distribution μ_d. Fubini's theorem transfers each slice's bound to the full risk and tail probability. Only the moving $\mathbb{R}^{d+1}$ parameter coordinates are rotated. No measurable selection of a bad point or direction across slices is needed. □

A.4.1 A MATCHING CONSTRUCTION WITH A NORM UPPER BOUND

Let $U = [e_1, \dots, e_r] \in \mathbb{R}^{(d+1)\times r}$, write $P_1 = xx^\top$, and set $\rho = \|P_1 U\|_F = \|U^\top x\|$. For $0 < \varepsilon \le 1$, define the rectangular block-diagonal factor

$$V_\varepsilon(P_1) = \frac{P_1 U}{\max\{\varepsilon, \rho\}}, \qquad W_\varepsilon = \operatorname{diag}(I_{k-1}, V_\varepsilon) \in \mathbb{R}^{(k+d)\times m}. \tag{33}$$

The map $A \mapsto A/\max\{\varepsilon, \|A\|_F\}$ is the Euclidean projection of A/ε onto the Frobenius unit ball. It is $1/\varepsilon$-Lipschitz. Also $\|P_1(p)U - P_1(q)U\|_F \le \sqrt{2}\,\mathrm{dist}(p,q)$. Hence

$$\|W_\varepsilon\|_F \le \sqrt{k}, \qquad \mathrm{Lip}_F(W_\varepsilon) \le \sqrt{2}/\varepsilon. \tag{34}$$

Since $(P_1U)(P_1U)^\top = \rho^2 P_1$, the error is exactly

$$h_{W_\varepsilon} = (1 - \rho^2/\varepsilon^2)_+. \tag{35}$$

Let

$$a = \frac{r}{2}, \qquad b = \frac{d+1-r}{2} > 0,$$

and write B for the beta function. A uniform $x \in S^d$ can be generated by normalizing a vector of independent standard Gaussian variables. The ratio of the sum of its first r squared coordinates to the total squared norm therefore gives $\rho^2 \sim \mathrm{Beta}(a,b)$. Integrating (35) yields the exact formula

$$\mathcal{R}(W_\varepsilon) = \frac{\varepsilon^r}{B(a,b)} \int_0^1 (1-v)^2 v^{a-1}(1-\varepsilon^2 v)^{b-1}\,dv. \tag{36}$$

In particular,

$$\mathcal{R}(W_\varepsilon) \sim \varepsilon^r \frac{B(a,3)}{B(a,b)} \quad \text{as } \varepsilon \downarrow 0, \qquad \mathcal{R}(W_\varepsilon) \le \frac{B(a, \min\{b,1\}+2)}{B(a,b)}\varepsilon^r. \tag{37}$$

For the last bound, use $(1-\varepsilon^2 v)^{b-1} \le 1$ when $b \ge 1$, and $(1-\varepsilon^2 v)^{b-1} \le (1-v)^{b-1}$ when $b < 1$. The asymptotic follows by dominated convergence, for example by restricting to $\varepsilon \le 1/2$.

A.4.2 A MATCHING CONSTRUCTION WITH EXACTLY CONSTANT TRACE

For $0 < \delta \le 1$, set

$$\begin{aligned} t &= \min\{1, \rho/\delta\}, \\ A_\delta(P_1) &= \frac{P_1 U}{\max\{\delta, \rho\}} + (1-t)\frac{U}{\sqrt{r}}, \\ \overline{V}_\delta(P_1) &= \frac{A_\delta(P_1)}{\|A_\delta(P_1)\|_F}, \qquad \overline{W}_\delta = \mathrm{diag}(I_{k-1}, \overline{V}_\delta). \end{aligned} \tag{38}$$

There is no division by a vanishing matrix in this construction. For $\rho \le \delta$, the two summands in A_δ have norms t and $1-t$, and their inner product is nonnegative because $\langle P_1U, U\rangle_F = \rho^2$. It follows that

$$\|A_\delta\|_F^2 \ge t^2 + (1-t)^2 \ge \frac{1}{2}.$$

For $\rho \ge \delta$, $A_\delta = P_1U/\rho$ has norm one. The triangle inequality also gives $\|A_\delta\|_F \le 1$ everywhere.

The scalar ρ is 1-Lipschitz in the projective metric. Indeed, choose representatives of two inputs with nonnegative inner product, and bound their projected norm difference by their chordal distance, which is at most their geodesic distance. The first term of A_δ is $\sqrt{2}/\delta$-Lipschitz and the second is $1/\delta$-Lipschitz. For nonzero Euclidean vectors A, B,

$$\|A - B\|^2 = (\|A\| - \|B\|)^2 + \|A\|\|B\| \left\| \frac{A}{\|A\|} - \frac{B}{\|B\|} \right\|^2. \tag{39}$$

Applying this identity to matrix vectorizations shows that normalization on the above set has Lipschitz constant at most $\sqrt{2}$. Consequently,

$$\|\overline{W}_\delta\|_F = \sqrt{k}, \qquad \mathrm{Lip}_F(\overline{W}_\delta) \le (2+\sqrt{2})/\delta. \tag{40}$$

When $\rho \ge \delta$, $\overline{W}_\delta \overline{W}_\delta^\top = P_k$. Inside the exceptional region, the error is supported on the moving block and is the difference of two positive semidefinite trace-one matrices. Its operator norm is at most one. Using the regularized incomplete beta function I_z, we obtain

$$\mathcal{R}(\overline{W}_\delta) \le \mathbb{P}\{\rho < \delta\} = I_{\delta^2}(a,b) \le \frac{B(a, \min\{b,1\})}{B(a,b)}\delta^r. \tag{41}$$

The last inequality follows by the same substitution and comparison as in (37). At $\rho = 0$, the moving output block is $UU^\top/r$. For $r \ge 2$ the output then has rank m, which is allowed by (5). The construction must not be inserted into a smaller class that requires output rank exactly k everywhere.

Projective-subfamily case of Theorem 1. Lemmas 9 and 10, followed by Proposition 11, give both lower bounds. Take $\varepsilon = \sqrt{2}/L$ in (33) and $\delta = (2+\sqrt{2})/L$ in (38). The main-text constants can be chosen as

$$C_{d,r} = 2^{r/2}\frac{B(a, \min\{b,1\}+2)}{B(a,b)},$$
$$\overline{C}_{d,r} = (2+\sqrt{2})^r \frac{B(a,\min\{b,1\})}{B(a,b)}. \tag{42}$$

These upper bounds satisfy the same column, norm, and Lipschitz budgets as their corresponding lower bounds. All parameters other than L are fixed in the claimed asymptotic rates.

Squared Frobenius error is at least squared operator error. For the clipped construction the error is a scalar multiple of a rank-one projector, so these two errors agree. For the normalized construction the nonzero error block is the difference of two positive semidefinite trace-one matrices. Its squared Frobenius norm is at most two. Thus the Frobenius-risk upper bound also has exponent r.

For $m < k$, every point has a missing target direction simply by dimension, so exact representation is impossible. For $k \le m < k+d$, Lemma 9 rules it out. For $m \ge N = k+d$, take $W_* = [P_k, 0]$. Then $W_* W_*^\top = P_k$, $\|W_*\|_F = \sqrt{k}$, and

$$\|W_*(p) - W_*(q)\|_F = \sqrt{2}\sin \operatorname{dist}(p,q) \le \sqrt{2}\,\operatorname{dist}(p,q).$$

This proves the width threshold and a sufficient Lipschitz budget. It does not establish the least possible Lipschitz constant of an exact rank-k factor. □

A.5 Coupled elliptic PDE realization

We prove Proposition 7 and record the experimental constants. Let $q \ge 1$ and $\psi_\ell(s) = \sqrt{2}\sin(\ell\pi s)$, $1 \le \ell \le q$, be the $L^2(0,1)$-orthonormal Galerkin basis. At $\varepsilon = 0$, the matrix of (23) is

$$A_0([x]) = \operatorname{diag}_{1\le\ell\le q}(\ell^2\pi^2 + \alpha) \otimes I_4 - \beta I_q \otimes xx^\top. \tag{43}$$

Its lowest eigenvector is $\psi_1 \otimes x$, with eigenvalue $\pi^2 + \alpha - \beta$. The nearest competitors are $\psi_1 \otimes x^\perp$, at separation β, and $\psi_2 \otimes x$, at separation $3\pi^2$ when $q \ge 2$. Thus the base gap is at least $g_0 = \min\{\beta, 3\pi^2\}$, with equality when $q \ge 2$.

Let H_q be the Galerkin compression of multiplication by $H(s)$. For every Galerkin vector v,

$$|\langle v, H_q v\rangle| = \left|\int_0^1 v(s)^\top H(s) v(s)\, ds\right| \le \int_0^1 \|v(s)\|_2^2\, ds,$$

so $\|H_q\|_{\mathrm{op}} \le 1$. Weyl's inequality along $A_t = A_0 + t\varepsilon H_q$ gives coercivity at least $\pi^2 + \alpha - \beta - \varepsilon$ and target gap at least $g_0 - 2\varepsilon$. Under the proposition's assumptions the target projector $P_t([x])$ is therefore smooth in $([x], t)$ and its image bundles at $t = 0$ and $t = 1$ are isomorphic.

At $t = 0$ the image line is the tautological line over $\mathbb{RP}^3$ embedded by the first spatial mode. If $B : \mathbb{R}^4 \to \mathbb{R}^{4q}$ denotes this isometric embedding, then

$$W_0([x]) = Bxx^\top \in \mathbb{R}^{4q\times 4}, \qquad W_0 W_0^\top = P_0([x]). \tag{44}$$

This is projectively invariant and supplies four exact columns. A smooth orthogonal transport $U([x])$ from P_0 to P_1 gives the exact factor UW_0 at $t = 1$. Conversely, any exact factor with fewer than four columns would pull back through $U^\top$ to such a factor of the tautological line, contradicting the width threshold on $\mathbb{RP}^3$. Multiplying factors by this transport, or its transpose, preserves pointwise error and changes the Lipschitz budget by only a fixed additive constant. Hence the nominal and robust exponents transfer to this family.

For the learned experiment, $q = 6$, $\alpha = 2$, $\beta = 1$, and $\varepsilon = 0.2$. Hence

$$\lambda_{\min}(A_1) \ge \pi^2 + 0.8 > 10.66, \qquad \operatorname{gap}(A_1) \ge 0.6. \tag{45}$$

Projecting the perturbed ground-state equation onto the complement of P_0 also yields

$$\sup_{[x]} \|P_1([x]) - P_0([x])\|_{\mathrm{op}} \le \frac{\varepsilon}{g_0 - \varepsilon} = \frac{1}{4}. \tag{46}$$

Thus the close-projector transport is valid globally. The ten supplied coefficient features are the upper-triangular entries of $xx^\top$ and hence determine the projective input without loss. The reported transported-factor oracle is an algebraic implementation check of (44), not a substitute for these bounds.

A.6 Higher-rank elliptic spectral clusters

The rank-one realization extends to an entire rank-k cluster without requiring gaps between eigenvalues inside that cluster. Let $P \in \mathrm{Gr}(k,4)$, $1 \le k < 4$, and consider the q-mode Galerkin matrix of

$$\mathcal{A}_{t,P}u = -u'' + [2I_4 - P + 0.2tH(s)]u, \qquad t \in [0,1], \quad u(0) = u(1) = 0, \tag{47}$$

where H is the bounded symmetric field used above. Write $J : \mathbb{R}^4 \to \mathbb{R}^{4q}$ for the isometry into the first sine mode, and $Q_t(P)$ for the projector onto the lowest k eigenvalues.

At $t = 0$ the target eigenvalue is $\pi^2 + 1$ with multiplicity k, $Q_0(P) = JPJ^\top$, and its gap from the complementary spectrum is one. The Galerkin perturbation has norm at most $0.2t$, since compression into an orthonormal trial space does not increase the multiplication-operator norm. Weyl's inequalities give, uniformly in P, t, and $q \ge 1$,

$$\lambda_{k+1}(A_{t,P}) - \lambda_k(A_{t,P}) \ge 0.6, \qquad \lambda_1(A_{t,P}) \ge \pi^2 + 0.8. \tag{48}$$

Thus the full rank-k spectral projector is smooth, including at any internal eigenvalue coincidence. The orthogonal transport from Corollary 6 supplies a smooth $O(P) \in O(4q)$ with $Q_1(P) = O(P)JPJ^\top O(P)^\top$.

The ambient discretization dimension $4q$ does not set the exact factor width for this family. An exact width-four factor is $O(P)JP$, whereas an exact narrower factor would, after applying $J^\top O(P)^\top$, factor the tautological projector on $\mathrm{Gr}(k,4)$, contradicting Theorem 1. For approximation, compression gives

$$\|BB^\top - P\|_{\mathrm{op}} \le \|WW^\top - Q_1(P)\|_{\mathrm{op}}, \qquad B = J^\top O(P)^\top W. \tag{49}$$

It preserves the norm inequality and changes the Lipschitz budget by at most a constant depending on the fixed smooth transport. It therefore transfers the lower rates to arbitrary ambient factors, including factors of fixed trace. Conversely, transporting and embedding the reference factors preserves both norm constraints and the loss exactly, and transfers the upper rates. The nominal, unrestricted, and capped laws consequently hold with $r = m - k + 1$ for $k \le m < 4$ and exact width four. They also hold under the comparable nonuniform input laws considered in Appendix B.6.

The higher-rank experiment uses $k = 2$, $q = 4$, and widths $2, 3, 4$ on the full $\mathrm{Gr}(2,4)$. Unlike the projective subfamily in (4), its parameterization fixes no common target direction. Its predictions concern a two-dimensional subspace, and the loss is unchanged by rotating the two eigenvectors used to store its label. This remains a generated elliptic diagnostic, not a claim that arbitrary physical parameter domains carry this bundle.

A domain where the obstruction is absent. The same operator admits an exact width-two factor when its inputs are restricted to a graph chart. Let $B \in [-0.4, 0.4]^{2\times 2}$ and set

$$V(B) = \begin{bmatrix} I_2 \\ B \end{bmatrix} (I_2 + B^\top B)^{-1/2}, \qquad P(B) = V(B)V(B)^\top. \tag{50}$$

The columns of $V(B)$ form a smooth orthonormal basis across this whole parameter box. Consequently $O(P(B))JV(B)$ is a smooth exact factor of $Q_1(P(B))$ with two columns and Frobenius norm $\sqrt{2}$. Compactness supplies a finite Lipschitz constant. This control retains the spectral problem and target rank while removing the nontrivial global bundle. The comparison does not assert that equal neural training budgets imply equal optimization difficulty on the two domains.

A.7 Nonlinear graph realization

We prove Proposition 8. Since $Q_0\mathbf{1} = 0$, the complete-graph Laplacian formed from weights $1 + (Q_0)_{ij}$ is

$$L_0 = 5I - \mathbf{1}\mathbf{1}^\top - Q_0.$$

It has eigenvalue zero on $\operatorname{span}\{\mathbf{1}\}$, eigenvalue four on $\operatorname{im} Q_0$, and eigenvalue five on the orthogonal three-dimensional complement. Thus the smallest nonconstant eigenspace is the tautological line and its spectral gap is one.

Each sinusoidal edge perturbation has magnitude at most η. The induced Laplacian perturbation E has absolute row sum at most $2(5-1)\eta$, hence $\|E\|_{\mathrm{op}} \le 8\eta$. Weyl's inequality applied to both neighboring eigenvalues leaves a gap of at least $1 - 2\|E\|_{\mathrm{op}} \ge 1 - 16\eta$. Also, every off-diagonal entry of an orthogonal projector is at least $-1/2$, so $w_{ij} \ge 1/2 - \eta > 0$.

Applying these bounds along the homotopy from zero to η gives a smooth isolated eigenprojector $Q_t([x])$. Kato parallel transport, defined by $\partial_t U_t = [\partial_t Q_t, Q_t]U_t$ and $U_0 = I$, is orthogonal and satisfies $Q_t = U_t Q_0 U_t^\top$. Smoothness on the compact parameter space gives a finite Lipschitz constant in $[x]$. The exact-representation statement follows by transporting factors in both directions, and Proposition 5 transfers the nominal and robust rates. Substituting $\eta = 0.03$ gives the stated numerical certificates.

B Distributional robustness: constructions and proofs

All Wasserstein balls in this section are centered at the population Haar measure and use the principal-angle geodesic metric. Their elements are arbitrary Borel probability measures on the same compact domain. The adversarial law can depend on the candidate factor. The loss always compares the prediction at the transported input with that input's own target projector. No labels are corrupted or held fixed across inputs.

B.1 Proof of the neural-operator output-head corollary

Fix y. Compression by the isometry $Q(y)$ cannot increase operator norm, so

$$\|W_{\psi,m}W_{\psi,m}^\top - P\|_{\mathrm{op}} \ge \|(Q(y)^\top W_{\psi,m})(Q(y)^\top W_{\psi,m})^\top - xx^\top\|_{\mathrm{op}}.$$

The compressed factor has Frobenius norm at most one and Lipschitz constant at most L on the projective fiber. The projective cases of Theorems 1 and 2, with $k = 1$ and $r = m$, therefore give the stated lower bounds uniformly in y. Integrating gives the nominal result. For the shifted result, apply the measurable finite-net selection in Lemma 12 jointly to $(y, [x])$ and move only the projective coordinate; the first marginal remains the product law and the same transport-cost estimate applies. Finally, $W(y, [x]) = Q(y)xx^\top \in \mathbb{R}^{N\times d}$ is a well-defined continuous exact factor at width d, because $(xx^\top)^2 = xx^\top$.

B.2 A transport lower bound from obstructed slices

For an integer $q \ge 1$, let $v_q(t)$ be the normalized volume of a radius-t geodesic ball in $\mathbb{RP}^q$. For $0 \le t \le \pi/2$, homogeneity makes it independent of its center, and

$$v_q(t) = \frac{2|S^{q-1}|}{|S^q|}\int_0^t \sin^{q-1} u\, du. \tag{51}$$

Fix $t_0 \in (0, \pi/2)$. There are constants $a_q, b_q > 0$ such that $a_q t^q \le v_q(t) \le b_q t^q$ for $0 \le t \le t_0$.

Lemma 12 (Transport from a family of obstructed slices)**.** *Let a compact group G act continuously and isometrically on a compact geodesic space X. Let $i : \mathbb{RP}^q \to X$ be an isometric embedding whose Haar-rotated slice measures average to μ. Suppose a nonnegative Lipschitz function h has, on every rotated slice, at least one point at which $h \ge a > 0$. For $1 \le p < \infty$, there are $c, \varepsilon_0 > 0$, independent of the Lipschitz constant of h, such that*

$$\sup_{\mathsf{W}_p(\mu,\nu)\le\varepsilon} \int h^2\, d\nu \ge c\varepsilon^{pq/(p+q)} \quad (0 \le \varepsilon \le \varepsilon_0). \tag{52}$$

For $p = \infty$, the lower bound is $c\varepsilon^q$.

Proof. Write $K = \mathrm{Lip}(h)$. If $K > 0$, choose a finite $a/(2K)$-net $z_1, \dots, z_J$ in $\mathbb{RP}^q$. If $K = 0$, one point suffices. On every slice there is an index with $h(gi(z_j)) \ge a/2$. Let $j(g)$ be the first such index

and $z(g) = z_{j(g)}$. Each selection set is Borel, since its definition uses finitely many inequalities involving the continuous functions $g \mapsto h(gi(z_j))$. This resolves measurability without assuming a continuous choice of topological defect.

Independently draw g from Haar measure on G and z uniformly from $\mathbb{RP}^q$. Set $x = gi(z)$, so the marginal of x is μ. For a radius $0 < t \le t_0$, define

$$y = \begin{cases} gi(z(g)), & d_{\mathbb{RP}^q}(z, z(g)) \le t, \\ x, & \text{otherwise.} \end{cases}$$

The resulting pair defines a Borel coupling. Its transported mass is exactly $v_q(t)$, and all moved mass lands where $h \ge a/2$. Thus, for its second marginal ν_t,

$$\int h^2 \, d\nu_t \ge \frac{a^2}{4} v_q(t), \qquad \mathsf{W}_p(\mu, \nu_t)^p \le t^p v_q(t) \le b_q t^{p+q}.$$

Taking $t = (\varepsilon^p / b_q)^{1/(p+q)}$ proves (52), provided this radius is at most t_0. For $p = \infty$ take $t = \varepsilon \le t_0$. The finite net can become arbitrarily fine as K increases. Its size does not enter the transported mass or the constants. □

This coupling need not be a deterministic transport map on X, because the same input can belong to multiple slices. That is allowed in the definition of Wasserstein distance. It is also why counting overlapping slice neighborhoods is unnecessary.

B.3 Matching upper bound from a small sublevel set

Lemma 13 (Transport to a small sublevel set)**.** *Suppose h is Borel measurable, $0 \le h \le H$, $s : X \to [0,1]$ is 1-Lipschitz, $h = 0$ on $\{s \ge \delta\}$, and $\mu\{s \le u\} \le D u^q$ for $0 < u \le 1$. If $\delta + t \le 1$, then*

$$\sup_{\mathsf{W}_p(\mu,\nu) \le \varepsilon} \int h^2 \, d\nu \le H^2 \left[D(\delta + t)^q + \frac{\varepsilon^p}{t^p} \right] \quad (1 \le p < \infty). \tag{53}$$

For $p = \infty$ and $\delta + \varepsilon \le 1$, the bound is $H^2 D(\delta + \varepsilon)^q$.

Proof. Compactness gives an optimal coupling for the continuous finite-p transport cost. If $h(y) > 0$ and $d(x, y) \le t$, then $s(x) \le s(y) + t < \delta + t$. The complementary event has probability at most ε^p / t^p by Markov's inequality. This proves (53). For $p = \infty$, compactness also gives a coupling supported on $d(x, y) \le \varepsilon$. One can obtain it as a weak limit of couplings supported on $d(x, y) \le \varepsilon + 1/n$. The same argument then has no complementary event. □

B.4 Proof of the sharp laws

Proof of Proposition 5. For an admissible factor W, set $\widetilde{W}(x) = U(x)W(x)$. Orthogonality preserves both Frobenius norm constraints, while

$$\|\widetilde{W}(x) - \widetilde{W}(y)\|_F \le \|W(x) - W(y)\|_F + \|U(x) - U(y)\|_{\mathrm{op}} \|W(y)\|_F,$$

so $\mathrm{Lip}_F(\widetilde{W}) \le L + \sqrt{k}T$. Moreover,

$$\|\widetilde{W}\widetilde{W}^\top - \widetilde{P}\|_{\mathrm{op}} = \|U(WW^\top - P)U^\top\|_{\mathrm{op}} = \|WW^\top - P\|_{\mathrm{op}}$$

pointwise. Hence the nominal integrals agree, and so do the integrals under every shifted law in the same Wasserstein ball. This proves the first inequality in (14). Apply the same argument to $U^\top$ for the reverse inequality. Additive changes in L do not alter the large-L exponents, and exact factors correspond in both directions. □

Proof of Corollary 6. Uniform spectral separation and the Riesz projector formula make $P_t(x)$ smooth in (x, t). Solve Kato's equation $\partial_t U_t = [\partial_t P_t, P_t] U_t$ with $U_0 = I$. Its generator is skew-symmetric, so U_t is orthogonal, and differentiating $U_t P_0 U_t^\top$ shows that it equals P_t. Compactness gives a finite Lipschitz constant for U_1 on X. Proposition 5 completes the proof. □

Proof of Theorem 2. For $k \le m < N$, set $q = r = m - k + 1$. On every projective r-slice used in Appendix A.3, the missing-direction argument gives $h = \|WW^\top - P\|_{\mathrm{op}} \ge 1$ somewhere. These are isometric slices, and their rotations average to Haar measure. The loss is Lipschitz for every admissible factor. Hence Lemma 12, with $a = 1$, gives the shift term for every factor, uniformly in L. The nominal lower bound of Theorem 1 also applies because $\nu = \mu$ is admissible. The maximum of these two lower bounds is at least half their sum, with an adjustment of constants.

For the upper bound, take the trace-normalized clipped polar factor $\overline{F}_\delta$ from (27), with $\delta = B_k/L$ and $B_k = 2 + \sqrt{2k}$. It belongs to both factor classes, has $h \le k$, and is exact when $s(P) = \sigma_k(PU) \ge \delta$. Singular-value perturbation and $\|P(x) - P(y)\|_{\mathrm{op}} \le d(x, y)$ show that s is 1-Lipschitz. The small-ball bound (25) supplies exponent r. Apply Lemma 13 and choose $t = \varepsilon^{p/(p+r)}$ for $\varepsilon > 0$. Since $(\delta + t)^r \le 2^{r-1}(\delta^r + t^r)$, this gives $C(L^{-r} + \varepsilon^{pr/(p+r)})$. At $\varepsilon = 0$ use the nominal upper bound directly. Choose L_0 and ε_0 so that $\delta + t \le 1$ throughout the stated regime. The infinite-p case follows by the corresponding part of that lemma.

Finally, when $m \ge N$, the factor $[P, 0]$ is exact at every input. Its loss remains zero under every probability distribution.

For the smooth-density restriction and $\varepsilon > 0$, first construct the lower-bound law ν using radius $\varepsilon/2$, and mix it equally with μ. This preserves at least half of its risk and remains within distance $\varepsilon/2$ of μ, for finite p and for $p = \infty$. Average this mixture under a smooth probability density on $O(N)$ supported sufficiently near the identity. Convolution on the compact homogeneous space gives a smooth density, bounded below by $1/2$ because Haar measure is invariant. The rotations can be chosen to move every input by at most η, giving Wasserstein distance at most η from the mixture for every p. Uniform continuity of h^2 makes its risk converge to the mixture's risk as $\eta \to 0$. Choose $0 < \eta < \varepsilon/2$ small enough to retain at least half that risk. The result is an admissible smooth positive-density law with the same lower-bound order. Its density upper bound can depend on W and ε. At radius zero, μ already has a smooth positive density. Upper bounds hold because the restricted ambiguity set is smaller. This completes the smooth-density assertion. □

B.5 Wasserstein shifts with a bounded density

Proof of the density-constrained part of Theorem 2. Write $h = h_W = \|WW^\top - P\|_{\mathrm{op}}$ and $r = m - k + 1$. For every admissible factor, $\mathrm{Lip}(h) \le K = 1 + 2\sqrt{k}L$. Let $Y = \mathbb{RP}^r$ have uniform probability measure σ, and fix an isometric slice $i : Y \to X$ as in Lemma 12. Its rotations by $G = O(N)$ average to μ. Every rotated slice contains a point at which $h \ge 1$.

A high-loss ball and a capped coupling. Set $\rho = (4K)^{-1}$. Choose a finite ρ-net $z_1, \dots, z_J$ in Y. On each rotated slice at least one index satisfies $h(gi(z_j)) \ge 3/4$. Select the first such index and denote its center by $z(g)$. This is a Borel selection, since each $g \mapsto h(gi(z_j))$ is continuous. Consequently

$$h(gi(z)) \ge \tfrac{1}{2} \qquad \text{for } z \in B_g := B_\rho(z(g)).$$

Fix a small $t_0 > 0$. By (51), there are $a, b > 0$ depending only on r, t_0 such that $au^r \le v_r(u) \le bu^r$ for $0 \le u \le t_0$. Increase L_0 if necessary so that $\rho \le t_0$. Already, averaging the high-loss balls over G gives

$$\int h^2 \, d\mu \ge \tfrac{1}{4} v_r(\rho) \ge \tfrac{a}{4}\rho^r. \tag{54}$$

Choose $\lambda \ge 2$ with $b\lambda^{-r} \le a/2$. For $\lambda\rho \le t \le t_0$, define

$$A_g = B_t(z(g)) \setminus B_g, \qquad v = v_r(\rho), \qquad w = v_r(t) - v, \qquad \eta = \min\{(M-1)v, w\}.$$

Here $w \ge (a/2)t^r > 0$. The density

$$f_g = 1 + \frac{\eta}{v}\mathbf{1}_{B_g} - \frac{\eta}{w}\mathbf{1}_{A_g} \tag{55}$$

integrates to one and satisfies $0 \le f_g \le M$, including when $M = 1$. Let $\sigma_{A_g}, \sigma_{B_g}$ be normalized restrictions of σ, and let $\Delta(z) = (z, z)$. A coupling from σ to $f_g\sigma$ is

$$\pi_g = \Delta_\# \left[\left(1 - \frac{\eta}{w}\mathbf{1}_{A_g}\right)\sigma\right] + \eta\, \sigma_{A_g} \otimes \sigma_{B_g}. \tag{56}$$

It moves only mass η, from A_g into B_g, through distances at most $t + \rho \le 3t/2$. The finite-valued measurable center selection makes these measure-valued constructions Borel measurable in g.

Push (56) forward by $(z, z') \mapsto (gi(z), gi(z'))$ and average over Haar measure dg. The first marginal is μ. If ν is the second marginal, then for every Borel set $E \subset X$,

$$\nu(E) = \int_G \int_Y \mathbf{1}_E(gi(z)) f_g(z)\, d\sigma(z)\, dg \le M \int_G \int_Y \mathbf{1}_E(gi(z))\, d\sigma(z)\, dg = M\mu(E). \tag{57}$$

Thus the ambient law has density at most M, even though each slice law is singular relative to ambient Haar measure. Overlap between slices does not affect this argument. The retained and added mass in B_g give

$$\int h^2\, d\nu \ge \tfrac{1}{4}(v + \eta) \ge \tfrac{a}{8}\left[\rho^r + \min\{(M-1)\rho^r, t^r\}\right], \tag{58}$$

$$\mathsf{W}_p(\mu, \nu)^p \le \eta(t + \rho)^p \le b(3/2)^p t^{p+r}, \qquad 1 \le p < \infty. \tag{59}$$

For $p = \infty$, the same coupling has maximum displacement at most $3t/2$.

Uniform lower bounds. For finite p, take $t = (\varepsilon^p/[b(3/2)^p])^{1/(p+r)}$ and reduce ε_0 so that $t \le t_0$. If $t \ge \lambda\rho$, (58)–(59) give an admissible law with risk at least a constant times

$$\rho^r + \min\{(M-1)\rho^r, \varepsilon^{pr/(p+r)}\}. \tag{60}$$

If $t < \lambda\rho$, including $\varepsilon = 0$, then $\varepsilon^{pr/(p+r)} \lesssim \rho^r$. The whole expression in (60) is therefore $O(\rho^r)$, uniformly in M, and the admissible nominal law in (54) suffices. Since $\rho \asymp L^{-1}$ for $L \ge L_0$, this proves the finite-p lower bound for every factor in either class, with constants independent of $M \ge 1$.

For $p = \infty$, set $t = 2\varepsilon/3$. If $t \ge \lambda\rho$, the coupling is feasible and (58) gives the same bound with ε^r in place of $\varepsilon^{pr/(p+r)}$. If $t < \lambda\rho$, then $\varepsilon^r \lesssim \rho^r$, and again the nominal law supplies the entire lower bound. This proves uniformity also at arbitrarily small radii and as $M \downarrow 1$.

A common factor for both upper bounds. Use the same trace-normalized clipped polar factor $\overline{F}_\delta$, with $\delta = B_k/L$, as in the unrestricted proof. Its nominal risk is at most $C_1 L^{-r}$, so every capped law satisfies

$$\int h^2_{\overline{F}_\delta}\, d\nu \le M \int h^2_{\overline{F}_\delta}\, d\mu \le C_1 M L^{-r}.$$

The unrestricted bound for this very factor is $C_2(L^{-r} + \varepsilon^{pr/(p+r)})$. Taking the minimum gives the claimed upper bound with constant $\max\{C_1, C_2\}$, independent of M. The argument for $p = \infty$ uses its unrestricted bound $C_2(L^{-r} + \varepsilon^r)$.

Smooth positive densities under the same cap. For $\varepsilon > 0$, first construct the lower-bound law at radius $\varepsilon/2$ and mix it equally with μ. Its density lies between $1/2$ and $(M+1)/2 \le M$, its risk is at least half the constructed risk, and its distance from μ remains at most $\varepsilon/2$. Smooth it by averaging under a smooth probability density on $O(N)$ supported sufficiently near the identity, as in the unrestricted proof. Haar invariance preserves both density bounds. The rotations can move every point by less than $\varepsilon/2$; continuity of h^2 allows retention of at least half of the mixed risk. The result is smooth, strictly positive, satisfies the same cap, and is within the original radius. Replacing ε by $\varepsilon/2$ changes the rate only by a fixed factor. At radius zero, or when $M = 1$, the Haar law itself suffices. □

The exact identity

$$\min\{MA, A + B\} = A + \min\{(M-1)A, B\}, \qquad A, B \ge 0,\ M \ge 1,$$

with $A = L^{-r}$ and $B = \varepsilon^{pr/(p+r)}$, identifies the concentration threshold $(M-1)L^{-r} \asymp B$; use $B = \varepsilon^r$ for $p = \infty$. A fixed finite cap restores a vanishing optimal robust risk. More generally, at fixed sufficiently small $\varepsilon > 0$, $M(L) = o(L^r)$ forces that risk to zero, whereas $\liminf_{L\to\infty} M(L)/L^r > 0$ leaves a positive lower bound on the best shifted error.

B.6 COMPARABLE NONUNIFORM INPUT LAWS AND CHANGES OF COORDINATES

The assumption of an exactly Haar input law can be weakened to a density bounded above and below by fixed positive constants. This does not follow by identifying Wasserstein balls with different centers. The following couplings give the required comparison and preserve the density cap even as $M \downarrow 1$.

Write $\mathcal{D}^{(M)}_{\lambda,p,\varepsilon}(W)$ for (9) with reference probability measure λ in both the Wasserstein constraint and the density cap. When the cap is omitted, write $\mathcal{D}_{\lambda,p,\varepsilon}(W)$.

Lemma 14 (Comparison of reference probability measures)**.** *Let $\widetilde{\mu} = f\mu$ on the same compact metric space, with $0 < c \le f \le C < \infty$ almost everywhere. For any fixed factor W, any $M \ge 1$, any $\varepsilon \ge 0$, and $1 \le p \le \infty$,*

$$c\,\mathcal{D}^{(M)}_{\mu,p,\varepsilon}(W) \le \mathcal{D}^{(M)}_{\widetilde{\mu},p,\varepsilon}(W) \le C\,\mathcal{D}^{(M)}_{\mu,p,\varepsilon}(W). \tag{61}$$

For finite p, the sharper comparison is

$$c\,\mathcal{D}^{(M)}_{\mu,p,\varepsilon c^{-1/p}}(W) \le \mathcal{D}^{(M)}_{\widetilde{\mu},p,\varepsilon}(W) \le C\,\mathcal{D}^{(M)}_{\mu,p,\varepsilon C^{-1/p}}(W). \tag{62}$$

The same statements hold with the density caps omitted. Only nonnegativity of the loss is needed for these comparisons.

Proof. Since both measures are probabilities, $c \le 1 \le C$. Let π be a coupling from μ to an admissible law ν and let $\Delta(x) = (x, x)$. The measure

$$\widetilde{\pi} = c\pi + \Delta_{\#}(\widetilde{\mu} - c\mu)$$

is a probability coupling with first marginal $\widetilde{\mu}$ and second marginal $\widetilde{\nu} = c\nu + \widetilde{\mu} - c\mu$. All component measures are nonnegative. If $\nu \le M\mu$, then

$$\widetilde{\nu} \le \widetilde{\mu} + c(M-1)\mu \le \widetilde{\mu} + (M-1)\widetilde{\mu} = M\widetilde{\mu}.$$

The pth transport moment is multiplied by c, while the integral of the nonnegative loss is at least $c\int \ell_W\, d\nu$. For $p = \infty$, the added diagonal mass does not increase the essential maximum displacement. This proves the left inequalities.

Conversely, given a coupling $\widetilde{\pi}$ from $\widetilde{\mu}$ to $\widetilde{\nu} \le M\widetilde{\mu}$, set

$$\pi = \frac{1}{C}\widetilde{\pi} + \Delta_{\#}\left(\mu - \frac{1}{C}\widetilde{\mu}\right).$$

Its first marginal is μ and its second marginal satisfies

$$\nu = \frac{1}{C}\widetilde{\nu} + \mu - \frac{1}{C}\widetilde{\mu} \le \mu + \frac{M-1}{C}\widetilde{\mu} \le M\mu.$$

Its loss is at least $C^{-1}\int \ell_W\, d\widetilde{\nu}$, and its finite-$p$ cost is divided by C. Its maximum displacement is no larger than that of $\widetilde{\pi}$. Taking suprema proves the right inequalities. Compactness supplies optimal transport couplings; equivalently, one may use arbitrarily near-optimal couplings and take limits. Without caps the same constructions require no density inequalities. Finally, monotonicity in the radius and $c \le 1 \le C$ imply (61) from (62) for finite p. □

Taking infima over the same factor class in (61) proves all the nominal, unrestricted, and density-capped rates for $\widetilde{\mu}$. In particular, the density cap is relative to $\widetilde{\mu}$, and

$$\inf_{W \in \mathcal{C}(L)} \mathcal{D}^{(M)}_{\widetilde{\mu},p,\varepsilon}(W) \asymp \min\{ML^{-r}, L^{-r} + \varepsilon^{a_p}\}, \tag{63}$$

uniformly for $M \ge 1$ and in the same (L, ε) regime as Theorem 2. If its Haar lower and upper constants are c_*, C_*, valid choices here are cc_* and CC_*. For fixed sufficiently small positive radius, optimal capped risk tends to zero if and only if $M(L) = o(L^r)$. In the shift-dominated regime, the concentration crossover remains $(M-1)L^{-r} \asymp \varepsilon^{a_p}$. These statements are uniform over all measurable densities satisfying the same fixed bounds c, C; no derivatives of f are required.

Scope of the smooth-density assertion. If f is smooth and strictly positive, the preceding coupling transformations also preserve smooth strictly positive shifted densities relative to Haar measure. Indeed their transformed densities are $f - c + cg$ and $1 - f/C + \widetilde{g}/C$, where $g, \widetilde{g}$ are the original shifted densities. Both expressions are smooth and positive, and the cap inequalities proved above are unchanged. Thus the smooth-density clause of Theorem 2 also extends under this extra assumption on f. For a merely measurable f, the rates above refer to all admissible measures, with measurable densities when a cap is imposed. A uniform smooth-density claim would be false: at $M = 1$ the only admissible law is $\widetilde{\mu}$ itself, which need not be smooth. The same issue arises at radius zero without a cap.

Bi-Lipschitz coordinates. Let $F : X \to Y$ be a bijection satisfying

$$a\, d_X(x, x') \le d_Y(Fx, Fx') \le b\, d_X(x, x'), \qquad 0 < a \le b < \infty.$$

Use target $P_Y = P \circ F^{-1}$ and input law $\lambda = F_{\#}\widetilde{\mu}$ on Y. Density caps relative to the reference law are preserved exactly under pushforward. A coupling of p-Wasserstein radius ε/b on X is feasible at radius ε on Y; pulling a feasible coupling back gives radius at most ε/a on X. Both statements also hold for $p = \infty$. Meanwhile, composition with F multiplies a factor's Lipschitz constant by at most b, and composition with F^{-1} multiplies it by at most $1/a$. If $\mathcal{B}_{X,\mu}(L, \varepsilon, M)$ denotes the optimal capped risk, and $\mathcal{B}_{Y,\lambda}$ uses the corresponding factor class and target on Y, these facts and (61) give

$$c\, \mathcal{B}_{X,\mu}(bL, \varepsilon/b, M) \le \mathcal{B}_{Y,\lambda}(L, \varepsilon, M) \le C\, \mathcal{B}_{X,\mu}(aL, \varepsilon/a, M). \tag{64}$$

The same comparison holds without a cap and for nominal risk. Consequently all exponents and the exact factor-width threshold are preserved. This argument uses only metric and measure pushforwards, so it requires neither smooth coordinates nor a smooth input density. Preservation of the optional smooth-density restriction would additionally require suitable smoothness, for example a smooth diffeomorphism and a smooth positive f. The comparison is global; it does not extend the obstructed-family lower bound to a support contained in one trivializing chart.

B.7 How many nominal validation points reveal the defect?

The preceding rates have a direct implication for evaluating a frozen factor. Let $A_W = \{x : \|W(x)W(x)^\top - P(x)\|_{\mathrm{op}}^2 \ge 1/4\}$. For an independent nominal validation set of size n, write $H_n(W) = \mathbb{P}\{\text{at least one validation point lies in } A_W\}$.

Proposition 15 (Nominal validation power)**.** *For $k \le m < N$, $r = m - k + 1$, and sufficiently large L, there are positive constants c, C, depending only on N, k, m, such that*

$$cL^{-r} \le \mu(A_W) \quad \textit{for every } W \in \mathcal{W}_{\le}(L), \qquad \inf_{W \in \mathcal{W}_{\le}(L)} \mu(A_W) \le CL^{-r}. \tag{65}$$

The same assertion holds for $\mathcal{W}_{=}(L)$. For a factor fixed independently of the validation set, $H_n(W) = 1 - (1 - \mu(A_W))^n$. Consequently, at any fixed desired confidence $1 - \alpha \in (1/2, 1)$, the worst-case nominal sample size needed to witness the high-loss event has order $L^r \log(1/\alpha)$. This is a detection statement, not a uniform generalization bound for a factor selected using the validation samples.

Proof. Write $h_W = \|WW^\top - P\|_{\mathrm{op}}$ and $K = 1 + 2\sqrt{k}L$. On every rotated projective r-slice there is a point where $h_W \ge 1$. Since h_W is K-Lipschitz, the slice ball of radius $1/(2K)$ around that point lies in A_W. The radial slice density used in (32) gives mass at least cL^{-r}. Averaging the rotated slice laws gives the first inequality under Haar measure. For the reverse inequality, the clipped factor (26) with clipping scale $\sqrt{2}/L$ is exact whenever $s(P) \ge \sqrt{2}/L$. Hence A_W lies in that singular-value tube, whose mass is at most CL^{-r} by (25). The trace-normalized construction (27), with scale B_k/L, is also exact outside such a tube and proves the assertion for $\mathcal{W}_{=}(L)$. The binomial identity follows from independence. Solving $1 - (1 - q)^n \ge 1 - \alpha$ with $q \asymp L^{-r}$ yields the stated order. □

For the explicit rank-one factor of Appendix C on $\mathbb{RP}^7$, take $\delta = \sqrt{2}/L$. Its event A_W is exactly $s^2 \le L^{-2}$, so the hit probability is

$$q_m(L) = I_{L^{-2}}\left(\frac{m}{2}, \frac{8-m}{2}\right), \qquad n_{.95}(m, L) = \left\lceil \frac{\log(0.05)}{\log(1 - q_m(L))} \right\rceil. \tag{66}$$

Table 4 shows the exact 95% detection sizes. Although wider factors have lower nominal risk, their unavoidable bad region is correspondingly harder to observe by uniform random testing. This is the regime in which a targeted search can be more informative than increasing a modest independent test set.

Table 4: Experiment 1 extension: exact number of independent Haar validation points needed for 95% probability of observing squared operator loss at least $1/4$ for the explicit clipped factor on $\mathbb{RP}^7$. The factor has Lipschitz constant at most L. Values are exact beta-CDF computations, not fitted sample-complexity exponents.

L	$m=1$	$m=2$	$m=3$	$m=4$
8	11	64	457	4,132
16	23	256	3,626	65,613
32	46	1,023	28,936	1,047,765

As an independent geometric check, we drew two million normalized eight-dimensional Gaussian directions at each of $L = 8, 16$ and counted $s^2 \le L^{-2}$ for all four widths. The largest absolute standardized deviation from the beta-CDF expectation was 2.17 over the eight predeclared comparisons. The script, random seed, hit counts, and exact probabilities are included in the source package.

C Analytic transport protocol

Table 5: Scale of the six experiments. Data/evaluation gives either train/validation/test sizes or fresh attack/random samples. Repeats include controls.

No.	experiment	domain/size	data or evaluation	repeats
1	analytic	18 (k, N, m) triples	quadrature/transport	522 fits/checks
2	frozen FNO heads	$N = 1024, 4096$	10,000/2,000/2,000	20 heads/30 anchors
3	elliptic, rank one	24×24, $m = 3, 4$	65,536/8,192	three pairs
	elliptic, rank two	16×16, $m =$ 2:4	16,384 per source law	nine + three chart
4	Darcy FNO	$q = 24, 32$, $m = 3, 4$	5,000/1,000/1,000	10 pairs/30 anchors
5	plate	$q = 8, 12$, $m = 1, 2$	Galerkin/self-attacks	20 frozen models
	plate L sweep	$q = 8$, $K = 512, 1024$	four certified budgets	$120 + 72$ stages
6	TU1402 loop/chart	1200 elements	832 two-input labels	$10 + 25$ models
	TU1402 L sweep	2814 DOFs	$K = 512, 1024$; FE labels	$100 + 40$ stages

The experiment evaluates an explicit norm-bounded rank-one factor on $\mathrm{Gr}(1, 8) = \mathbb{RP}^7$. It does not train a network. For $U = [e_1, \dots, e_m]$ and a unit representative x, let $s = \|U^\top x\|$ and

$$W_\delta([x]) = \frac{xx^\top U}{\max\{s, \delta\}}, \qquad h_\delta([x])^2 = (1 - s^2/\delta^2)_+^2.$$

This is the clipped polar factor of Appendix A.3. It is sign-invariant, has Frobenius norm at most one, and has Lipschitz constant at most $\sqrt{2}/\delta$. Writing $\theta = \arcsin s$ for distance to the defect set $Z = \{[x] : U^\top x = 0\}$, its Haar density is

$$f_m(\theta) = \frac{2 \sin^{m-1}\theta \cos^{7-m}\theta}{B(m/2, (8-m)/2)}, \qquad 0 < \theta < \pi/2.$$

Finite transport moments. The attack sends every input with $\theta \le a$ to its nearest point of Z and leaves the rest unchanged. It evaluates the target at the moved input. Its transport cost is exactly $\int_0^a \theta^p f_m(\theta)\, d\theta$, which is set to ε^p. Its Error is

$$F_m(a) + \int_a^{\arcsin \delta} h_\delta(\theta)^2 f_m(\theta)\, d\theta,$$

Table 6: Measured convergence of the clipped factors in Figure 3. The Error is the quadrature value at $L = 1.414 \times 10^4$. The fitted order is obtained from the eight largest recorded values of L and compared with the theoretical order in $O(L^{-k})$.

m	Error	fitted order k	theoretical order k
1	1.0865×10^{-4}	1.0000	1
2	1.0000×10^{-8}	2.0000	2
3	7.7607×10^{-13}	3.0000	3
4	5.0000×10^{-17}	4.0000	4

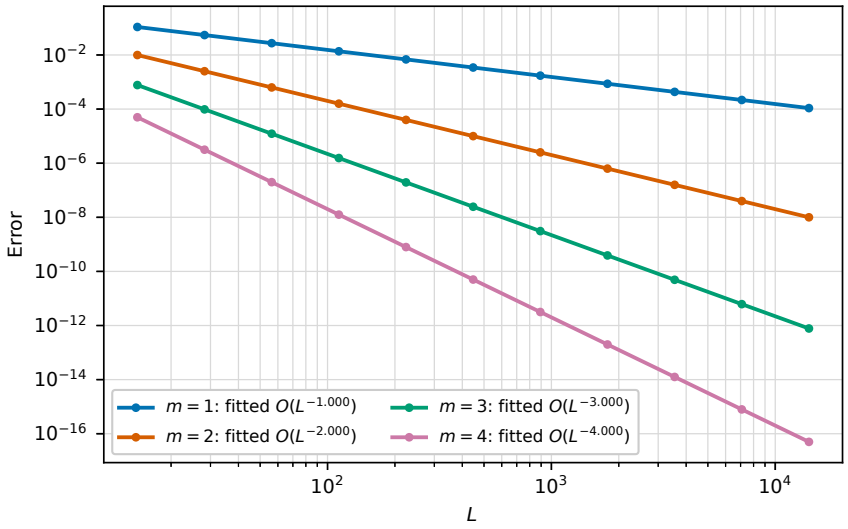

Figure 3: Experiment 1: Error versus L for the explicit clipped factors on $\mathbb{RP}^7$. Every point is a deterministic quadrature result. The legend reports the order fitted from the eight largest values of L; the theoretical prediction is $O(L^{-m})$.

with the integral interpreted as zero when $a \geq \arcsin \delta$. This is a feasible shifted Error for the particular factor, not an exact solution of the finite-p inner maximization. An analytic upper bound is $\min\{1, \inf_{t>0}[F_m(\arcsin \delta + t) + \varepsilon^p / t^p]\}$, with F_m clipped at one.

We use $m = 1, 2, 3, 4$, $p = 1, 2$, $\delta = 10^{-5}$, and 17 log-spaced radii from 10^{-6} to 10^{-2}. Fits use the first seven radii, fixed before execution. The observed $p = 1$ exponents are $0.50000, 0.66662, 0.74981, 0.79954$. The $p = 2$ exponents are $0.66667, 1.00000, 1.19998, 1.33327$. Their largest discrepancy from $pm/(p + m)$ is below 4.6×10^{-4}. A second sweep fixes $p = 2, \varepsilon = 10^{-3}$ and varies δ over 17 log-spaced values from 10^{-4} to 0.3, exposing the nonzero distribution-shift term.

Bounded displacement. For this monotone radial loss, a radius-ε perturbation can reduce θ to $\max\{0, \theta - \varepsilon\}$. The radial distance is 1-Lipschitz and this reduction is attained along a minimizing geodesic, so integrating that loss gives the exact W_∞ shifted Error of this factor. At $\delta = 10^{-7}$, 13 radii from 10^{-4} to 10^{-2} give first-seven-point exponents $0.99980, 1.99960, 2.99940, 3.99920$.

Checks and reproducibility. There are 256 deterministic curve points. Relative quadrature tolerance is 2×10^{-10} with absolute tolerance 10^{-25} for rare events. All feasible-attack values lie below their transport upper bounds and all 136 finite-p costs match the prescribed moments to relative error below 10^{-8}. At $m = 1$ and $m = 3$, two independently sampled sets of 60,000 unit vectors verify the transport geometry and Error within 0.91 estimated standard errors. Matrix eigendecomposition agrees with the scalar loss formula within 2.3×10^{-15}. An independent finite transport linear program and scalar dual agree to 1.0×10^{-12}. This last check concerns the discretized problem, not an exact

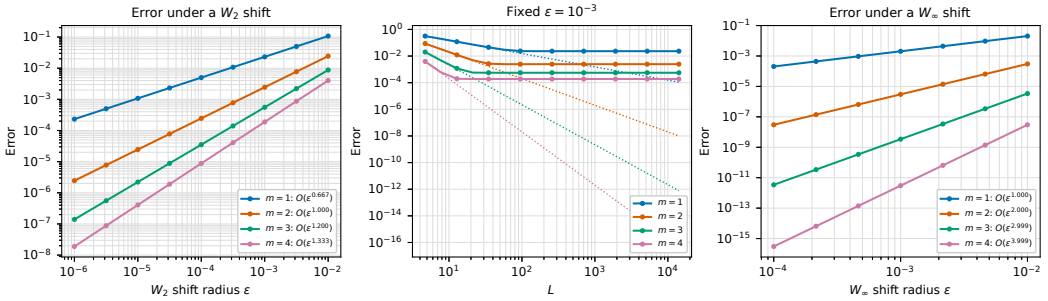

Figure 4: Experiment 1: Error under Wasserstein shifts. Left: measured Error under the feasible W_2 shift; the legend gives orders fitted from the first seven radii. Center: shifted Error (solid) and held-out Error (dotted) as L increases at fixed $\varepsilon = 10^{-3}$. Right: exact W_∞-shifted Error and fitted orders. All curves show computed Error values; no theoretical guide or upper-bound curve is drawn.

solution of the continuous problem. Six unit tests cover geometry, transport budgets, duality, limiting cases, and the unit-gap upper construction. The script, CSV curves, source hash, and diagnostic outputs accompany the paper. These checks were implemented within the same AI-assisted workflow and do not constitute an external mathematical review.

C.1 Density-constrained analytic transport

This extension uses the same norm-bounded rank-one factor and radial Haar law as the preceding protocol. Let F_m be the radial CDF and set

$$T_{a,M}(\theta) = \begin{cases} F_m^{-1}(F_m(\theta)/M), & 0 \le \theta \le a, \\ \theta, & a < \theta \le \pi/2, \end{cases} \qquad b = F_m^{-1}(F_m(a)/M). \tag{67}$$

Keep the angular coordinates in the m-dimensional sensor space and its orthogonal complement fixed. This defines a sign-invariant projective map outside measure-zero coordinate degeneracies, with geodesic displacement $\theta - T_{a,M}(\theta)$. Its target density relative to full Haar measure is M on $[0, b]$, zero on $(b, a]$, and one above a. Thus the cap holds on the original space, not merely on a radial marginal. Choose the largest $a \le \pi/2$ satisfying

$$\int_0^a [\theta - T_{a,M}(\theta)]^2 f_m(\theta)\, d\theta \le \varepsilon^2. \tag{68}$$

The Error under this feasible shifted law is

$$M \int_0^b h_\delta(\theta)^2 f_m(\theta)\, d\theta + \int_a^{\pi/2} h_\delta(\theta)^2 f_m(\theta)\, d\theta. \tag{69}$$

If $b \ge \arcsin \delta$, all nonzero loss lies where the density is M, and (69) equals $M\mathcal{R}(W_\delta)$. The universal density bound therefore gives the exact capped Error for this particular factor in that regime. Other evaluated values are feasible, with analytic upper envelope $\min\{M\mathcal{R}(W_\delta), U_{2,\varepsilon}(W_\delta)\}$, where U is the preceding tube/transport bound. The densities in this experiment have jumps and a zero annulus. The theorem's smooth-density extension is established by its proof rather than by these numerical laws.

Protocol and results. For $m = 1, 2, 3, 4$, $p = 2$, and $\varepsilon = 10^{-3}$, the clipping sweep uses 25 logarithmically spaced $\delta \in [10^{-4}, 0.2]$ with $M = 1, 4, 16$ and an unrestricted comparison. At $\delta = \varepsilon = 10^{-3}$, the cap sweep uses 49 logarithmically spaced $M \in [1, 10^{12}]$. All 596 points are retained. The predeclared first-seven-point clipping fits have slopes $1.000, 2.000, 3.000, 4.000$ for widths one through four; the full-precision values are saved in the source package. This window lies in the proved identity $\mathcal{D}^{(M)} = M\mathcal{R}(W_\delta)$, explaining the agreement with the nominal power. The unrestricted comparison has a constant plateau on the same window. Figure 5 displays only the computed capped Errors used to test the transition between these regimes.

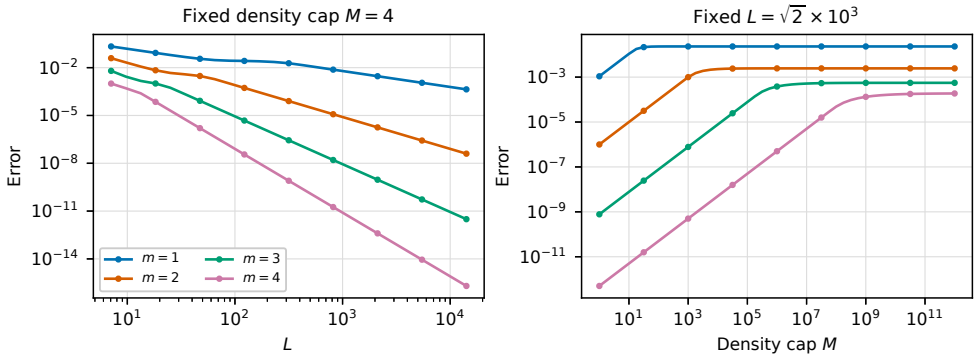

Figure 5: Experiment 1: Error under density-constrained shifts for an explicit rank-one factor on $\mathbb{RP}^7$, with $p = 2$ and $\varepsilon = 10^{-3}$. Left: Error versus L at $M = 4$. Right: Error versus M at $L = \sqrt{2} \times 10^3$. Every curve is a computed Error; colors denote output width.

Numerical verification. The inverse CDF uses the inverse regularized beta function. Transport moments use 256-point Gauss–Legendre quadrature and are rechecked with 512 points. Across the reported grid the largest relative discrepancy is 6.8×10^{-15} and the largest mass residual is 1.2×10^{-16}. Loss integrals use relative tolerance 2×10^{-11}. Every feasible curve point respects the transport budget and density bound. Twelve separate finite transport LPs at 80 and 160 radial bins enforce source masses, destination capacities, and the moment constraint directly. Their primal–dual gaps are below 2.8×10^{-17}, and the cases $M = 1$ and $\varepsilon = 0$ recover held-out Error. These are finite-problem checks, not certified continuum error estimates. The saved script, curve tables, configuration, and source hash accompany the manuscript.

C.2 An exact rank-two nominal-risk calculation

A second analytic check distinguishes output width from the codimension of the high-error region on the full $\mathrm{Gr}(2, 4)$. Take $U = [e_1, e_2, e_3]$ and the norm-bounded clipped factor $F_\delta(P)$ in (26), with $m = 3$ and $0 < \delta \le 1$. For any orthonormal basis V of P, the two largest squared singular values of PU are the eigenvalues of $V^\top U U^\top V = I_2 - (V^\top e_4)(V^\top e_4)^\top$. They equal 1 and $Y = 1 - P_{44}$. Haar invariance gives $P_{44} \sim \mathrm{Beta}(1, 1)$, so Y is uniform on $[0, 1]$. Consequently,

$$\mathcal{R}_\mu(F_\delta) = \int_0^{\delta^2} (1 - y/\delta^2)^2 \, dy = \frac{\delta^2}{3}. \tag{70}$$

The exponent is $r = m - k + 1 = 2$, rather than $m = 3$. This is an exact fixed-factor formula, not a neural training result or a numerical estimate of the best approximation error.

Consider also the smooth nonuniform density

$$f_\alpha(P) = 1 + \alpha(P_{11} + P_{22} - 1), \qquad 0 \le \alpha < 1. \tag{71}$$

Its Haar mean is one and $1 - \alpha \le f_\alpha \le 1 + \alpha$. Conditional orthogonal symmetry in the first three coordinates and $\mathrm{tr}\, P = 2$ imply $\mathbb{E}[P_{11} + P_{22} \mid Y] = 2(1 + Y)/3$. Hence

$$\mathcal{R}_{f_\alpha \mu}(F_\delta) = \left(1 - \frac{\alpha}{3}\right) \frac{\delta^2}{3} + \frac{\alpha \delta^4}{18}. \tag{72}$$

The leading exponent remains two under this change of input law.

The source-package table checks both identities with a fixed sample of 500,000 independent Gaussian-QR planes and unbiased importance weights for $\alpha = 0.6$. All six clipping scales are retained. The largest absolute discrepancy is 1.65 estimated Monte Carlo standard errors. Separately, dense SVD reconstruction at 2048 planes and five clipping scales checks the full matrix loss against the scalar expression to 1.1×10^{-13}. Rotating the two basis columns changes their projector by at most 1.2×10^{-15}. These checks use actual rank-two matrices, in addition to the one-dimensional exact integration.

The six-scale Monte Carlo table is in the source package.

D BENCHMARK PROTOCOLS AND ADDITIONAL CHECKS

This appendix records the information needed to interpret the experiments in Section 3. The source package contains the checkpoints, machine-readable configurations, per-seed results, attack coordinates, and figure scripts. All robust values below are obtained from explicit feasible couplings and are therefore lower bounds on the corresponding Wasserstein suprema.

D.1 PAIRED OUTPUT-HEAD INTERVENTIONS

The elliptic and Darcy experiments use the same intervention. We first train an $m = 4$ FNO, copy and freeze its lifting, Fourier, and pointwise backbone layers, and delete the fourth factor column to initialize the $m = 3$ head. Only the two final heads are then optimized, on identical minibatches. The paired backbone tensors agree bitwise after training. Each head maps 32 features to $4m$ channels; the additional factor column therefore costs 132 parameters.

For the variable-coefficient elliptic problem, the $q = 16, 32$ grids have $N = 4q^2 = 1024, 4096$ degrees of freedom. The input consists of the two coefficient fields, the ten entries of $xx^\top$, and two coordinates. Four width-32 Fourier layers retain 8 and 10 modes per direction. Each resolution uses $10{,}000/2000/2000$ train/validation/test examples, five seeds, 1600 end-to-end Adam steps, and 1000 paired-head steps. The coefficient bounds $a_\theta, b_\theta \in [0.8, 1.2]$ imply a discrete eigengap of one for the four-field ground state.

The Darcy experiment follows the random-permeability distribution of Li et al. (2021). A Gaussian field with covariance $(-\Delta + 9I)^{-2}$ is thresholded to $a \in \{3, 12\}$; a conservative five-point scheme solves (20). FP64 conjugate-gradient residuals are at most 2.00×10^{-7}, and three sparse direct solves agree within 3.25×10^{-8}. Four width-32, 12-mode Fourier layers are trained on $5000/1000/1000$ examples at $q = 24, 32$. We use five seeds, 1600 backbone steps, 900 paired-head steps, and three coefficient anchors per seed.

For each fixed coefficient anchor, a multistart search finds a projective input with nearly unit narrow-head loss. A common empirical W_2 coupling moves the nearest source atoms to that point, splitting one boundary atom to match ε^2. Both heads are evaluated on these same moved inputs. The source and random-control sample sizes are 8192 and 2048. Isotropic projective steps of the same radius form the control. All 30 elliptic seed–anchor comparisons and all 30 Darcy comparisons favor $m = 4$ at $\varepsilon = 0.05$; the medians and fitted finite-range exponents appear in Table 1.

The small elliptic model in Experiment 3 is evaluated in the same way after training is complete. Its coupling uses 65,536 fresh projective inputs, and the random control uses 8192 inputs. The three $m = 3$ destinations have unit loss to numerical precision, whereas the paired $m = 4$ losses lie between 4.78×10^{-6} and 1.45×10^{-5}. Direct radial quadrature under the Haar law gives the continuous-source exponent in Table 1; doubling the quadrature resolution changes the excess Errors by at most 1.5×10^{-7} relatively.

D.2 PLATE MODES

For the simply supported plate, the sine-Galerkin basis $\phi_{ij} = 2\sin(i\pi x)\sin(j\pi y)$ diagonalizes the bending term. On $E = \mathrm{span}\{\phi_{12}, \phi_{21}\}$, the perturbed doublet is

$$A_\theta|_E = 35I + \frac{1}{2}\begin{pmatrix} -\cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{pmatrix}. \tag{73}$$

The lower eigenvector changes sign when θ traverses $\mathbb{R}/\pi\mathbb{Z}$. The complementary spectrum is at least 17 away from the doublet, while the Schur correction is at most $4/17$; hence the second mode has a uniform gap of at least $9/17$. This verifies the hypotheses used for the width-one obstruction.

At $q = 12$, a four-layer width-24 FNO with four modes per direction maps the foundation field and coordinates to the factor. Five $m = 2$ backbones are trained for 1400 steps and frozen. Fresh $m = 1, 2$ heads, with the same first-column initialization, are trained for 700 identical minibatches on $8192/2048/2048$ labels. The second column adds 25 parameters to the 76,249-parameter narrow model. All five pairs favor $m = 2$ under the width-one-targeted coupling; the median shifted errors are 0.152 and 8.54×10^{-3}. Searches performed independently for both widths give the same qualitative separation.

The hidden-capacity control trains width-one FNOs with 12, 24, and 48 hidden channels, corresponding to 19,117, 76,249, and 304,561 parameters. Their median held-out errors decrease from 2.47×10^{-3} to 4.17×10^{-4}, whereas their median shifted errors remain in $[0.145, 0.147]$. Checkpoint reload, periodicity, the factor-norm cap, and dense versus reduced operator loss were checked independently; the last two loss evaluations agree within 2.63×10^{-15}.

The certified-budget ReLU experiment uses periodic inputs $(\cos 2\theta, \sin 2\theta)$ and four predeclared budgets. Spectral normalization and a fixed-knot construction provide certified Lipschitz bounds. Twenty independent seed groups are evaluated on fresh labels and on the joint schedule $\varepsilon = 0.25L^{-3/2}$. Table 3 reports medians and complete seed ranges. These finite-budget fits test compatibility with the exponents; they are not estimates of the minimax constants.

D.3 TU1402 STRUCTURAL MODES

The COST TU1402 model supplies stiffness and consistent mass matrices. The input loop varies three support stiffnesses as in (22); labels are freshly solved generalized eigenvectors restricted to 120 displacement sensors. Interval checks at 306 anchors verify a nonzero sensor restriction, a continuous first-mode gap, and a negative full-mode overlap around the loop. These checks certify the sign reversal used by Proposition 16.

The factor network has two tanh layers with 128 units, Frobenius projection onto the unit ball, 3000 Adam steps, and five seeds. Validation projector loss selects the checkpoints before any defect search. A 256-point scan is refined with a fresh finite-element solve at every scalar root evaluation. All initial and all 25 fresh width-one models have a unit-loss destination. The empirical coupling again moves the nearest source mass and splits one boundary atom. At $\varepsilon = 0.0491$, the median width-one error rises from 1.15×10^{-2} to 8.99×10^{-2}; the paired width-two value is 1.19×10^{-5}. A two-parameter annulus $S^1 \times [0.8, 1.2]$, cut-chart controls, fixed-load projection, and the 20-seed certified-budget study use independent samples and preserve this separation. The reported attacks remain feasible lower bounds, and the rate claim applies to the certified loop rather than the entire annulus.

D.4 NUMERICAL SAFEGUARDS

All reported checkpoints are selected without using test or attack data. Targets are recomputed at moved inputs, operator losses are evaluated in double precision, and empirical transport costs include the fractional boundary atom. Random shifts use the same radius as the targeted shift. Saved audits compare reduced and dense losses, verify projector symmetry and idempotence, check factor-norm constraints, and record solver residuals. The finite searches cannot certify global adversarial optima; they provide reproducible witnesses for the width-dependent errors reported in the main text.

E APPLYING AND EVALUATING THE LAWS ON PARAMETER LOOPS

E.1 A VERIFIABLE EIGENLINE OBSTRUCTION ALONG A PARAMETER LOOP

The sign change of a continuously followed real mode can be certified from discrete eigenpairs if an interval perturbation bound rules out unresolved behavior between samples. The following sufficient condition is conservative but does not require a prescribed algebraic form of the target projector. The underlying sign holonomy is classical (Berry, 1984).

Proposition 16 (Mode continuation and sign holonomy). *Let $A : \mathbb{R}/T\mathbb{Z} \to \mathrm{Sym}(n)$ be C^1, and fix an ordered eigenvalue index k. Choose a partition $0 = t_0 < \cdots < t_J = T$. At each t_j, suppose the kth eigenvalue is simple, with certified gap lower bound $g_j > 0$ to every other eigenvalue. Suppose also that*

$$\sup_{t \in [t_j, t_{j+1}]} \|A(t) - A(t_j)\|_{\mathrm{op}} \le \delta_j, \qquad 2\delta_j < g_j. \tag{74}$$

Let u_j be any unit target eigenvector at t_j, setting $u_J = u_0$. Then all adjacent inner products are nonzero, the target eigenprojector $P(t)$ is C^1 on the entire circle, and its sign holonomy is

$$\sigma = \prod_{j=0}^{J-1} \mathrm{sign}(u_j^\top u_{j+1}) \in \{-1, 1\}. \tag{75}$$

If $\sigma = -1$, every continuous periodic single-column factor $w(t)$ has a point z satisfying

$$\|w(z)w(z)^\top - P(z)\|_{\mathrm{op}}^2 \geq 1. \tag{76}$$

Proof. Weyl's inequalities imply that the target gap on interval j is at least $g_j - 2\delta_j > 0$. Thus the ordered target eigenvalue remains simple and its projector is C^1. For a unit eigenvector $u(t)$ on this interval, project the identity $(A(t_j) - \lambda_k(t))u(t) = -(A(t) - A(t_j))u(t)$ onto $u_j^\perp$. The restricted operator on the left has inverse norm at most $(g_j - \delta_j)^{-1}$, since $|\lambda_k(t) - \lambda_k(t_j)| \leq \delta_j$. Consequently

$$\|(I - u_j u_j^\top)u(t)\| \leq s_j := \frac{\delta_j}{g_j - \delta_j} < 1. \tag{77}$$

The target line therefore stays in the neighborhood where projection of u_j onto that line is nonzero. Normalizing this projection gives a continuous unit lift with positive inner product with u_j throughout the interval. At the right endpoint its sign relative to u_{j+1} is exactly the corresponding factor in (75). Concatenating these lifts shows that a continuous lift on $[0, T]$ satisfies $u(T) = \sigma u(0)$. Changing the sign of any stored u_j changes two factors in the product, so the result is independent of those choices. When $\sigma = -1$, the continuous scalar $u(t)^\top w(t)$ has opposite endpoint values and hence a zero z. At that point $u(z)^\top(ww^\top - P)u(z) = -1$, proving (76). □

Computed-eigenpair caveat. For a validated candidate residual η_j and target separation $d_j > \eta_j$, spectral expansion bounds the line-angle sine by η_j/d_j. If the computed adjacent overlap has rigorous arithmetic error χ_j, its sign agrees with the exact overlap when

$$|b_j| > \chi_j + e_j + e_{j+1}, \qquad e_j = \sqrt{2 - 2\sqrt{1 - (\eta_j/d_j)^2}}. \tag{78}$$

The interval perturbations also need certified bounds; sampled gaps and residuals alone cannot certify the sign product. The TU1402 numerical application below does not claim validated interval eigenpairs.

Consequences for average and transported loss. Use the circular geodesic distance and uniform probability dt/T. Suppose $\sigma = -1$, $\sup_t \|w(t)\| \leq B$, and $\mathrm{Lip}(w) \leq L$. If $D_j \geq \sup_{[t_j, t_{j+1}]} \|A'(t)\|_{\mathrm{op}}$, then

$$\mathrm{Lip}(P) \leq \Gamma := \max_j \frac{D_j}{g_j - 2\delta_j}, \qquad \mathrm{Lip}\big(\|ww^\top - P\|_{\mathrm{op}}\big) \leq K := 2BL + \Gamma.$$

Indeed, a locally parallel unit eigenvector satisfies $\|u'(t)\| \leq \|A'(t)\|_{\mathrm{op}}/\operatorname{gap}(t)$ and $\|P'(t)\|_{\mathrm{op}} = \|u'(t)\|$. The point z in (76) therefore gives

$$\mathcal{R}(w) \geq \frac{2}{T}\int_0^{T/2} (1 - Ks)_+^2 \, ds, \qquad \mathcal{R}(w) \geq \frac{2}{3TK} \quad \text{if } K \geq 2/T. \tag{79}$$

For any $0 \leq a \leq T/2$, move every source point within distance a of z to z and leave the remaining points fixed. The moved mass is $2a/T$, and its charged pth moment is exactly $2a^{p+1}/((p+1)T)$. Hence, with

$$a_\varepsilon = \min\left\{\frac{T}{2}, \left(\frac{(p+1)T\varepsilon^p}{2}\right)^{1/(p+1)}\right\},$$

the unrestricted population risk satisfies

$$\mathcal{D}_{p,\varepsilon}(w) \geq \frac{2a_\varepsilon}{T}, \qquad 1 \leq p < \infty. \tag{80}$$

For $p = \infty$, use $a_\varepsilon = \min\{\varepsilon, T/2\}$. The identity coupling also retains the lower bound in (79). Thus, for fixed operator family, a fixed norm bound and large L, these inequalities give lower orders L^{-1} and $L^{-1} + \varepsilon^{p/(p+1)}$ at small radii ($L^{-1} + \varepsilon$ for $p = \infty$). These inequalities alone are lower bounds. On the one-dimensional circle, an explicit construction also gives matching bounds, as follows; no such conclusion is asserted for arbitrary parameter spaces.

Corollary 17 (Sharp rates for a nontrivial eigenline loop). *Let $P : \mathbb{R}/T\mathbb{Z} \to \mathrm{Sym}(n)$, $n \geq 2$, be a C^1 rank-one orthogonal projector with negative sign holonomy, and let $\Gamma \geq \sup_t \|P'(t)\|_{\mathrm{op}}$. Use the uniform input law and circular geodesic metric. For either class of periodic factors with $\mathrm{Lip}(w) \leq L$ and $\sup_t \|w(t)\| \leq 1$ or $\|w(t)\| = 1$ everywhere, set $\alpha_p = p/(p+1)$ for finite p and $\alpha_\infty = 1$. Then*

$$\begin{aligned} \inf_w \mathcal{R}(w) &\asymp L^{-1}, \\ \inf_w \mathcal{D}_{p,\varepsilon}(w) &\asymp L^{-1} + \varepsilon^{\alpha_p}, \\ \inf_w \mathcal{D}^{(M)}_{p,\varepsilon}(w) &\asymp \min\{ML^{-1}, L^{-1} + \varepsilon^{\alpha_p}\}. \end{aligned} \tag{81}$$

Here the cap is relative to dt/T. Constants are uniform for large L, small $\varepsilon \geq 0$, and every $M \geq 1$, at fixed T, Γ, p. The robust statements also hold for smooth strictly positive shifted densities under the same cap. Two columns admit an exact periodic factor of Frobenius norm one and Lipschitz constant at most $\sqrt{\Gamma^2 + (\pi/T)^2}$.

Proof. A unit lift obtained by parallel continuation satisfies $u(t+T) = -u(t)$ and $\|u'(t)\| \leq \Gamma$. We construct one norm-equality factor, which supplies upper bounds for both classes. Use an interval $I = [T/2 - a, T/2 + a]$. On I, choose a unit $v \perp u$ by solving $v' = -u\langle u', v\rangle$ from any perpendicular initial unit vector. Both u and v then evolve under the skew-symmetric matrix $\Omega = u'u^\top - uu'^\top$, with $\|\Omega\|_{\mathrm{op}} \leq \Gamma$. Let θ increase linearly from zero to π on I, and define

$$w(t) = \begin{cases} u(t), & t \leq T/2 - a, \\ \cos\theta(t)u(t) + \sin\theta(t)v(t), & t \in I, \\ -u(t), & t \geq T/2 + a. \end{cases}$$

This is periodic and Lipschitz, has norm one, and has Lipschitz constant at most $\Gamma + \pi/(2a)$. Its squared loss is $\sin^2\theta$ on I and zero outside. Taking $a = \pi/L$ is admissible for $L \geq 2\Gamma$ and $L \geq 4\pi/T$. A smooth angle with zero endpoint derivatives can replace the linear angle if a C^1 factor is desired, changing only constants.

Thus either class contains one factor whose loss is bounded by one and supported on an interval of radius $O(L^{-1})$. For a coupling of finite-p radius ε, splitting at displacement t bounds its risk by $C_T(L^{-1} + t) + \varepsilon^p/t^p$. Choose $t = \varepsilon^{p/(p+1)}$; at radius zero use nominal risk. For $p = \infty$ the bound is $C_T(L^{-1} + \varepsilon)$. The same factor also has capped risk at most $M\mathcal{R}(w)$. Taking the smaller of these bounds proves all upper rates.

The preceding pointwise obstruction and (79)–(80) give the nominal and unrestricted lower bounds. For the capped lower bound, let $\rho = [2(2L + \Gamma)]^{-1}$. The interval B of radius ρ about an obstructed point has loss at least $1/4$ and mass $v = 2\rho/T$. For $2\rho \leq t \leq T/4$, the surrounding annulus A has mass $b = 2(t - \rho)/T$. Move $\eta = \min\{(M-1)v, b\}$ uniformly from A into B. The density $1 + (\eta/v)\mathbf{1}_B - (\eta/b)\mathbf{1}_A$ lies in $[0, M]$, the resulting risk is at least $(v + \eta)/4$, and the charged moment is at most $\eta(t + \rho)^p \leq C_{T,p}t^{p+1}$. Taking $t \asymp \varepsilon^{p/(p+1)}$ gives $c[\rho + \min\{(M-1)\rho, \varepsilon^{\alpha_p}\}]$. When this t is smaller than 2ρ, the nominal bound alone covers that entire expression. For $p = \infty$, take $t = 2\varepsilon/3$ and use the same small-radius split. Since $\rho \asymp L^{-1}$, this proves the capped lower bound uniformly even as $M \downarrow 1$. Mixing with the uniform law and smoothing by small circular rotations, with radius slack, proves the smooth-density version exactly as in Appendix B.5.

Finally, $W(t) = u(t)[\cos(\pi t/T), \sin(\pi t/T)]$ is periodic and satisfies $WW^\top = P$, $\|W\|_F = 1$, and $\|W'\|_F^2 = \|u'\|^2 + (\pi/T)^2$. □

This corollary uses the classical classification of real lines on a circle and the same localization and transport estimates as the main theorem. It extends their scope to any such loop, without a prescribed reference homotopy, rather than introducing a separate topological mechanism.

Generalized and measured modes. For fixed positive-definite mass M, whiten $Kv = \lambda Mv$ to $M^{-1/2}KM^{-1/2}$; mode overlaps use $v_i^\top Mv_j$. A sensor map S preserves the sign conclusion only where the measured mode $SM^{-1/2}u$ is nonzero. Appendix D checks this for TU1402 through a sensor-nullspace eigenvalue comparison, conditional on double-precision nodal eigenpairs.

E.2 Risk laws on a multi-parameter family of mode loops

A single obstructed curve need not carry positive probability under a higher-dimensional input law. The following result identifies a regime in which the one-dimensional risk law survives a transverse parameter change. It is stated for a product domain to make the source law and transport cost explicit; uniformly bi-Lipschitz coordinates give the same exponents.

Theorem 18 (Mode-loop risk on a product parameter domain)**.** *Let $Y \subset \mathbb{R}^q$ be a compact rectangle with nonempty interior, and let $X = (\mathbb{R}/T\mathbb{Z}) \times Y$ have the product metric of circular geodesic distance and Euclidean distance. Suppose $P : X \to \mathrm{Sym}(n)$ is a C^1 rank-one orthogonal projector, $n \geq 2$, whose real eigenline reverses sign around the circle for one $y \in Y$. Let μ have density between two fixed positive constants relative to uniform probability on X. For either norm constraint $\sup_X \|w\| \leq 1$ or $\|w\| = 1$ everywhere, let $\mathrm{Lip}(w) \leq L$ and $\ell_w = \|ww^\top - P\|_{\mathrm{op}}^2$. For each fixed $1 \leq p < \infty$, uniformly for $L \geq L_0$, $0 \leq \varepsilon \leq \varepsilon_0$, and $M \geq 1$,*

$$\inf_w \int_X \ell_w \, d\mu \asymp L^{-1},$$

$$\inf_w \sup_{\mathsf{W}_p(\mu,\nu)\leq\varepsilon} \int_X \ell_w \, d\nu \asymp L^{-1} + \varepsilon^{p/(p+1)},$$

$$\inf_w \sup_{\substack{\mathsf{W}_p(\mu,\nu)\leq\varepsilon \\ \nu\leq M\mu}} \int_X \ell_w \, d\nu \asymp \min\{ML^{-1}, L^{-1} + \varepsilon^{p/(p+1)}\}. \tag{82}$$

The constants may depend on P, Y, T, p and the density bounds, but not on L, ε, M. With two output columns there is an exact periodic factor W with $WW^\top = P$ and finite Lipschitz constant.

Proof. The sign around a fiber cannot change as y varies continuously through the connected set Y while P remains rank one. Hence every fiber has negative sign holonomy. For each fixed y, the argument of Proposition 16 gives an angle where $h_w := \|ww^\top - P\|_{\mathrm{op}} \geq 1$. Uniformly on X, $\mathrm{Lip}(h_w) \leq 2L + \mathrm{Lip}(P) =: K$. For measurability, select on each fiber the first maximizer of h_w on a fixed angular grid of mesh at most $(8K)^{-1}$. This defines a measurable $z(y)$ with $h_w(z(y), y) \geq 7/8$. On the angular interval B_y of radius $\rho = (8K)^{-1}$ about $z(y)$, the loss is at least $1/2$. The uniform lower density bound and integration over Y give $\int \ell_w \, d\mu \gtrsim \rho \asymp L^{-1}$.

For the unrestricted transport lower bound, take an angular radius $t \leq T/8$ and, for every y, send source mass within distance t of $z(y)$ to $(z(y), y)$, leaving the rest fixed. Its pth transport moment is at most Ct^{p+1} and its destination risk is at least ct. Taking t to be a sufficiently small multiple of $\varepsilon^{p/(p+1)}$ proves the shifted term; the nominal bound covers the case $t \lesssim \rho$.

For the density-capped lower bound, let A_y be the angular annulus between radii ρ and t, where $t \geq 2\rho$. On each fiber, move mass under the conditional law μ_y in the disintegration $\mu(d\theta, dy) = \mu_Y(dy)\mu_y(d\theta)$,

$$\eta_y = \min\{(M-1)\mu_y(B_y),\ \mu_y(A_y)/2\}$$

from the conditional source law on A_y to its conditional law on B_y. The resulting density relative to μ lies in $[0, M]$. Since $\mu_y(B_y) \asymp \rho$, $\mu_y(A_y) \asymp t$ uniformly, the added risk on B_y is bounded below by $c\min\{(M-1)\rho, t\}$. Every move has length at most $2t$, so its moment is at most Ct^{p+1}. For $t < 2\rho$ the nominal lower bound already covers the claimed expression. This proves the capped lower order, including $M = 1$.

For the matching upper bounds, cut the circle at one angle. Since $[0, T] \times Y$ is contractible, the pulled-back line admits a C^1 unit lift $u(\theta, y)$ with $u(T, y) = -u(0, y)$. Over a fixed angular strip of width $2a$, choose a C^1 unit vector v perpendicular to u and rotate from u to $-u$ as in the proof of Corollary 17. The resulting periodic norm-one factor has Lipschitz constant $O(1 + a^{-1})$ and loss supported on that strip. With $a \asymp L^{-1}$ its nominal risk is $O(L^{-1})$. For any transport coupling of radius ε, split it at displacement t: the source mass within distance t of the strip is $O(a + t)$, and Markov's inequality bounds the remaining mass by ε^p/t^p. Taking $t \asymp \varepsilon^{p/(p+1)}$ proves the unrestricted upper bound; $\nu \leq M\mu$ gives the additional capped bound $O(ML^{-1})$ for the same factor.

Finally, $W(\theta, y) = u(\theta, y)[\cos(\pi\theta/T), \sin(\pi\theta/T)]$ is periodic, has two columns and Frobenius norm one, and satisfies $WW^\top = P$ with finite Lipschitz constant. □