Geometry-Selective Distribution Comparison via Normal Log-Density Taylor Signatures

Anonymous authors
Paper under double-blind review

Abstract

MMD and KSD quantify distributional discrepancy but do not explicitly distinguish changes along local tangent and normal directions. We propose the Normal Log-Density Taylor Signature (NLTS), which uses a reference distribution P to define local normal directions and compares reference log-density responses across samples from P and Q. Under a local manifold model with small normal noise, we show that the negative-log-density Hessian identifies the local normal subspace, while, for spectrally isolated normal directions, projected scores provide the corresponding first-order Taylor coefficients. Synthetic experiments verify the intended normal–tangent selectivity. On CIFAR-10-C, NLTS remains effective with learned scores, while in single-cell perturbations its power is associated with geometric changes characterized separately by local PCA. Together, these results support the use of reference-defined local density geometry for selective high-dimensional distribution comparison. Code will be made publicly available.

1 Introduction

Discrepancy measures for quantifying differences between two probability distributions P and Q are fundamental tools in statistics and machine learning. Methods such as maximum mean discrepancy (MMD) (Gretton et al., 2012) and Stein discrepancies (Liu et al., 2016; Chwialkowski et al., 2016) are widely used as general-purpose tools for detecting distributional change. However, a detected change can have different geometric meanings depending on the directions in which it occurs. When high-dimensional data concentrate near a low-dimensional manifold (Fefferman et al., 2016; Brown et al., 2023), changes along tangent directions, i.e., variations already represented in the reference data, correspond to redistributions of probability mass along the manifold. By contrast, changes in normal, or off-manifold, directions are not captured by redistribution along the manifold. For applications concerned with departures from reference-supported variation, this distinction can be important, yet general-purpose discrepancies do not explicitly distinguish between tangent and normal changes.

To this end, we propose the *Normal Log-Density Taylor Signature* (NLTS), which uses the reference distribution P to define local normal directions and compares the distributions of the resulting reference log-density responses evaluated on samples from P and Q. In practice, however, the data manifold—and hence its local normal directions—is typically not known explicitly. The key challenge is therefore to recover this local normal structure from the reference distribution itself. We show that, under a small-noise local model, the local normal space at a point near the manifold is encoded by the dominant spectral subspace of the ambient negative-log-density Hessian evaluated at that point. NLTS leverages this spectral correspondence to form a compact signature of the first- and second-order behavior of the reference log density along the recovered normal directions. Our contributions are as follows:

- We show, under a small-noise local model, that the dominant spectral subspace of the negative-log-density Hessian recovers the local normal space, while, for spectrally isolated normal directions, the corresponding score projections encode the leading first-order normal-coordinate Taylor coefficients (Sec. 3.1; Props. 1 and 2).
- Building on this correspondence, we develop NLTS, which summarizes the first- and second-order behavior of the reference log density along recovered normal directions and compares the distribution of the resulting signature under P and Q (Sec. 3.2).

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 evaluate NLTS across controlled and real-data settings. On synthetic distributions with known tangent–normal structure, we validate its intended directional behavior under controlled normal and tangential perturbations (Sec. 4.1). On CIFAR-10-C (Hendrycks & Dietterich, 2019), we assess its behavior on high-dimensional natural images using a learned reference field (Sec. 4.2). Finally, on sci-Plex3 K562 drug-perturbation data (Srivatsan et al., 2020) from the scPerturb harmonized release (Peidli et al., 2024), we explore whether NLTS detection sensitivity correlates with drug-induced geometric changes characterized separately by local PCA (Sec. 4.3).

The code used in all experiments will be made publicly available.

2 Preliminaries

2.1 Distribution comparison and related work

Let P and Q denote reference and comparison distributions on $\mathbb{R}^D$. Given samples $\mathbf{x}_1, \ldots, \mathbf{x}_n \sim P$ and $\mathbf{y}_1, \ldots, \mathbf{y}_m \sim Q$, common discrepancy measures include MMD (Gretton et al., 2012), Energy Distance (Székely & Rizzo, 2013), and Wasserstein distance. When the reference score $\mathbf{s}_P(\mathbf{x}) = \nabla_{\mathbf{x}} \log p_P(\mathbf{x})$ is available, Stein identities yield score-based discrepancies such as KSD (Liu et al., 2016). NLTS also uses the reference score, but exploits the local differential structure of the reference log density for distribution comparison.

In high dimensions, finite-sample performance can degrade with dimension for kernel- and distance-based two-sample tests (Ramdas et al., 2015), while empirical Wasserstein distance has intrinsic-dimension-dependent sample complexity (Weed & Bach, 2019). A common strategy is to reduce the comparison to low-dimensional projections, as in Sliced Wasserstein (Bonneel et al., 2015), Sliced KSD / maxSKSD (Gong et al., 2021a), and Active Slices (Gong et al., 2021b). These methods use shared directions, chosen randomly or optimized for distributional comparison; NLTS instead derives sample-specific directions from the local differential structure of the reference score. Baseline details are in App. B.

Related work has applied statistical distribution comparison to high-dimensional dataset shifts, including dimensionality reduction with two-sample tests such as MMD (Rabanser et al., 2019) and learned kernels for increasing two-sample test power (Liu et al., 2020). CIFAR-10-C has also been used for online shift detection (Luo et al., 2024) and for MMD-based shift detection that triggers adaptation (Bhope et al., 2025).

Scores and log-density Hessians have also been used to recover data-manifold geometry, including intrinsic dimension from diffusion models (Stanczuk et al., 2024) and local Hessian spectra (Osada, 2026). Whereas these works use scores or Hessian spectra to estimate manifold geometry, NLTS uses reference-defined local geometry as a representation for distribution comparison.

2.2 Problem setup: Tangent and normal geometry

Under the manifold hypothesis (Fefferman et al., 2016), assume that P concentrates near a smooth d-dimensional embedded manifold $\mathcal{M} \subset \mathbb{R}^D$. For $\mathbf{m} \in \mathcal{M}$, let $T_{\mathbf{m}}\mathcal{M}$ and $N_{\mathbf{m}}\mathcal{M}$ denote the tangent and normal spaces with $\mathbb{R}^D = T_{\mathbf{m}}\mathcal{M} \oplus N_{\mathbf{m}}\mathcal{M}$. By the tubular neighborhood theorem (Lee, 2012), within a sufficiently small neighborhood we can locally choose a smooth orthonormal normal frame $\mathbf{U}_N(\mathbf{m}) \in \mathbb{R}^{D\times r}$, where $r = D - d$, and write

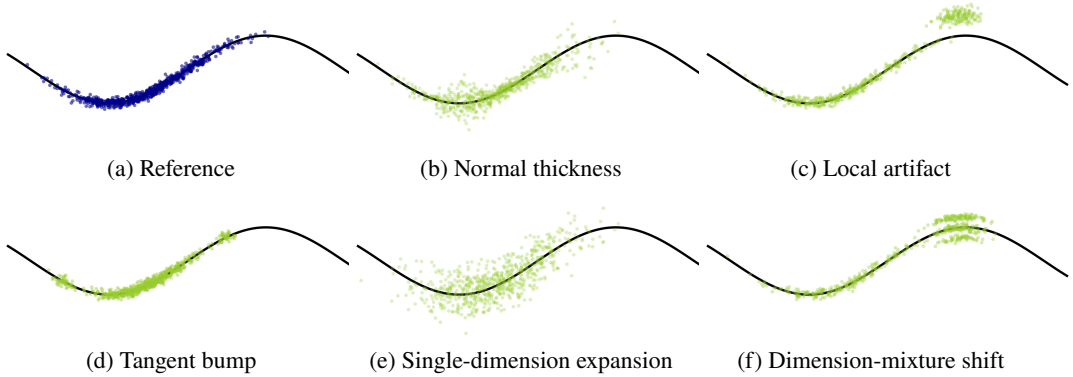
$$\mathbf{x} = F(\mathbf{m}, \boldsymbol{\xi}) = \mathbf{m} + \mathbf{U}_N(\mathbf{m})\boldsymbol{\xi}, \qquad \boldsymbol{\xi} \in \mathbb{R}^r. \tag{1}$$

In these coordinates, let the induced joint density be

$$f_P(\mathbf{m}, \boldsymbol{\xi}) = q_P(\mathbf{m})\rho_P(\boldsymbol{\xi} \mid \mathbf{m}),$$

where q_P is the manifold marginal density with respect to volume measure and ρ_P is the normal-coordinate conditional density with respect to Lebesgue measure. Thus, q_P describes how probability mass is distributed along the manifold, whereas ρ_P describes the conditional variation in the normal coordinates. Fig. 1 illustrates the corresponding synthetic shifts.

Throughout, *local density geometry* refers to the local differential structure of the ambient log density. We use ''curvature'' for second-order structure represented by the negative log-density Hessian, not for the intrinsic or extrinsic curvature of $\mathcal{M}$.



(a) Reference (b) Normal thickness (c) Local artifact

(d) Tangent bump (e) Single-dimension expansion (f) Dimension-mixture shift

Figure 1: Synthetic shift schematics. Relative to (a), panels (b), (c), and (e) alter local normal conditional structure; (d) illustrates a representative tangential specificity control; and (f) reweights components with different local-dimensional regimes. Normal-geometry shift constructions are given in App. D.1; tangential controls are detailed in App. D.4.

3 Method

NLTS is a fixed-reference construction: the reference score and Hessian of P define local normal directions and the associated first- and second-order log-density responses. Applying this P-defined map to samples from both P and Q yields the signature distributions to be compared. Sec. 3.1 develops the underlying Taylor structure, from its representation in ideal normal coordinates to its recovery from the ambient score and Hessian (Props. 1 and 2), and characterizes its response to departures of Q from P. Sec. 3.2 then constructs the practical NLTS signature, defines how it is used to compare Q against a fixed reference P, and states the population-level condition under which the induced signature distributions are distinguishable.

3.1 Local Taylor structure and normal directions

We begin with the idealized setting where the manifold and normal coordinates are known, and characterize the first- and second-order Taylor coefficients of the reference negative log density along normal directions. Let P_σ be a family of reference distributions with fixed manifold marginal and a normal-coordinate conditional density that concentrates as $\sigma \to 0$. Specifically, we assume $q_{P,\sigma}(\boldsymbol{m}) \equiv q_P(\boldsymbol{m})$. Write the negative log density of the normal conditional distribution as

$$\ell^N_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m}) := -\log \rho_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m}). \tag{2}$$

Near $\boldsymbol{\xi} = \mathbf{0}$, it admits the local expansion

$$\ell^N_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m}) = \ell^N_{P,\sigma}(\mathbf{0} \mid \boldsymbol{m}) + \frac{1}{2\sigma^2}\boldsymbol{\xi}^\top \boldsymbol{A}(\boldsymbol{m})\boldsymbol{\xi} + \widetilde{r}_\sigma(\boldsymbol{m}, \boldsymbol{\xi}), \qquad \boldsymbol{A}(\boldsymbol{m}) \succ 0. \tag{3}$$

Here, $\boldsymbol{A}(\boldsymbol{m}) \in \mathbb{R}^{r\times r}$ characterizes the anisotropy of local concentration in the normal coordinates. The regularity assumptions are given in App. A.2.1. For $\boldsymbol{\xi}$ in the typical tube, $\|\boldsymbol{\xi}\| = O(\sigma)$, consider a small displacement $\boldsymbol{\eta} \in \mathbb{R}^r$. A local Taylor expansion gives

$$\ell^N_{P,\sigma}(\boldsymbol{\xi}+\boldsymbol{\eta} \mid \boldsymbol{m}) - \ell^N_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m}) = -\boldsymbol{g}_{P,\sigma}(\boldsymbol{m}, \boldsymbol{\xi})^\top \boldsymbol{\eta} + \frac{1}{2}\boldsymbol{\eta}^\top \boldsymbol{K}_{P,\sigma}(\boldsymbol{m}, \boldsymbol{\xi})\boldsymbol{\eta} + o(\|\boldsymbol{\eta}\|_2^2). \tag{4}$$

By Eq. (3) and the stated regularity assumptions,

$$\begin{aligned} \boldsymbol{g}_{P,\sigma}(\boldsymbol{m}, \boldsymbol{\xi}) &:= \nabla_{\boldsymbol{\xi}} \log \rho_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m}) = -\frac{1}{\sigma^2}\boldsymbol{A}(\boldsymbol{m})\boldsymbol{\xi} + O(1), \\ \boldsymbol{K}_{P,\sigma}(\boldsymbol{m}, \boldsymbol{\xi}) &:= \nabla^2_{\boldsymbol{\xi}} \ell^N_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m}) = \frac{1}{\sigma^2}\boldsymbol{A}(\boldsymbol{m}) + O(1), \end{aligned} \tag{5}$$

where the $O(1)$ terms remain uniformly bounded over the typical tube as $\sigma \to 0$; see App. A.2.1 for the derivation. Thus, the normal-coordinate score $\boldsymbol{g}_{P,\sigma}$ and curvature $\boldsymbol{K}_{P,\sigma}$ have leading terms $-\boldsymbol{A}(\boldsymbol{m})\boldsymbol{\xi}/\sigma^2$ and $\boldsymbol{A}(\boldsymbol{m})/\sigma^2$, respectively.

Recovery from ambient score and Hessian. If the manifold and its normal coordinates were known, the normal-coordinate score and curvature could be evaluated directly. In real data, however, the normal coordinate system is unknown, so $\boldsymbol{g}_{P,\sigma}$ and $\boldsymbol{K}_{P,\sigma}$ are not directly available. We next show that their leading-order information can be recovered from the reference score and Hessian in the ambient space. The ambient reference score and negative-log-density Hessian are

$$\boldsymbol{s}_{P,\sigma}(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \log p_{P,\sigma}(\boldsymbol{x}), \qquad \boldsymbol{H}_{P,\sigma}(\boldsymbol{x}) = \nabla^2_{\boldsymbol{x}}[-\log p_{P,\sigma}(\boldsymbol{x})] = -\nabla_{\boldsymbol{x}} \boldsymbol{s}_{P,\sigma}(\boldsymbol{x}). \tag{6}$$

Let $\boldsymbol{A}(\boldsymbol{m})\boldsymbol{v}_j = \alpha_j \boldsymbol{v}_j$, with $\alpha_1 \geq \cdots \geq \alpha_r > 0$. Denote the dominant eigenpairs of $\boldsymbol{H}_{P,\sigma}(\boldsymbol{x})$ by $(\lambda_j(\boldsymbol{x}), \boldsymbol{u}_j(\boldsymbol{x}))$, ordered by descending eigenvalue. For notational simplicity, we suppress the dependence of $\boldsymbol{u}_j$ and λ_j on σ. We first establish the second-order correspondence.

Proposition 1 (Second-order correspondence and normal-space recovery)**.** *Under the regularity assumptions of App. A.2.1, for $\boldsymbol{m} \in \mathcal{M}$,*

$$\begin{aligned} \boldsymbol{H}_{P,\sigma}(\boldsymbol{m}) &= \frac{1}{\sigma^2} \boldsymbol{U}_N(\boldsymbol{m}) \boldsymbol{A}(\boldsymbol{m}) \boldsymbol{U}_N(\boldsymbol{m})^\top + \boldsymbol{R}_\sigma(\boldsymbol{m}), \qquad \|\boldsymbol{R}_\sigma(\boldsymbol{m})\|_{\text{op}} = O(1), \\ \lambda_j(\boldsymbol{m}) &= \frac{\alpha_j}{\sigma^2} + O(1), \qquad j = 1, \ldots, r. \end{aligned} \tag{7}$$

For any $\boldsymbol{x}$ at which the eigensystem is defined, let $\widehat{N}_\sigma(\boldsymbol{x}) := \operatorname{span}\{\boldsymbol{u}_1(\boldsymbol{x}), \ldots, \boldsymbol{u}_r(\boldsymbol{x})\}$. Then, on the manifold,

$$\left\| \sin\Theta\Big(\widehat{N}_\sigma(\boldsymbol{m}), N_{\boldsymbol{m}}\mathcal{M}\Big) \right\|_{\text{op}} = O(\sigma^2). \tag{8}$$

For a typical-tube sample $\boldsymbol{x} = F(\boldsymbol{m}, \boldsymbol{\xi})$ with $\|\boldsymbol{\xi}\| = O(\sigma)$,

$$\left\| \sin\Theta\Big(\widehat{N}_\sigma(\boldsymbol{x}), N_{\boldsymbol{m}}\mathcal{M}\Big) \right\|_{\text{op}} = O(\sigma). \tag{9}$$

Details and the proof are given in App. A.2.2. We next establish the first-order correspondence.

Proposition 2 (First-order correspondence)**.** *Assume that the corresponding eigenvalue α_j is separated from the other normal eigenvalues by a positive eigengap. For a typical-tube sample $\boldsymbol{x} = F(\boldsymbol{m}, \boldsymbol{\xi})$ with $\|\boldsymbol{\xi}\| = O(\sigma)$, choose the sign of $\boldsymbol{u}_j(\boldsymbol{x})$ to align with $\boldsymbol{U}_N(\boldsymbol{m})\boldsymbol{v}_j$. Then*

$$\begin{aligned} \boldsymbol{U}_N(\boldsymbol{m})^\top \boldsymbol{s}_{P,\sigma}(\boldsymbol{x}) &= \boldsymbol{g}_{P,\sigma}(\boldsymbol{m}, \boldsymbol{\xi}) + O(1), \\ a_{j,\sigma}(\boldsymbol{x}) := \boldsymbol{u}_j(\boldsymbol{x})^\top \boldsymbol{s}_{P,\sigma}(\boldsymbol{x}) &= \boldsymbol{v}_j^\top \boldsymbol{g}_{P,\sigma}(\boldsymbol{m}, \boldsymbol{\xi}) + O(1) \\ &= -\frac{\alpha_j}{\sigma^2}\xi_j + O(1), \qquad \xi_j = \boldsymbol{v}_j^\top \boldsymbol{\xi}. \end{aligned} \tag{10}$$

The proof is given in App. A.2.3. Prop. 1 is a subspace statement and requires only separation of the dominant normal block from the tangent spectrum, whereas the coordinate-wise interpretation above additionally requires the internal eigengap stated in Prop. 2. Repeated eigenvalues are discussed in App. A.4.3. Together, the two propositions recover the leading-order normal-space structure and, under the additional internal eigengap, the corresponding coordinate-wise projected-score structure.

This correspondence also has a direct ambient-space Taylor interpretation: along a Hessian eigendirection, $-a_{j,\sigma}$ and $\lambda_j/2$ are the linear and quadratic coefficients of the reference negative log density. Let $\ell_{P,\sigma}(\boldsymbol{x}) := -\log p_{P,\sigma}(\boldsymbol{x})$. Then

$$\ell_{P,\sigma}(\boldsymbol{x} + \boldsymbol{\delta}) - \ell_{P,\sigma}(\boldsymbol{x}) = -\boldsymbol{s}_{P,\sigma}(\boldsymbol{x})^\top \boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^\top \boldsymbol{H}_{P,\sigma}(\boldsymbol{x})\boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|_2^2). \tag{11}$$

In particular, since $\boldsymbol{H}_{P,\sigma}(\boldsymbol{x})\boldsymbol{u}_j(\boldsymbol{x}) = \lambda_j(\boldsymbol{x})\boldsymbol{u}_j(\boldsymbol{x})$,

$$\ell_{P,\sigma}(\boldsymbol{x} + t\boldsymbol{u}_j) - \ell_{P,\sigma}(\boldsymbol{x}) = -a_{j,\sigma}(\boldsymbol{x})t + \frac{1}{2}\lambda_j(\boldsymbol{x})t^2 + o(t^2). \tag{12}$$

See App. A.2.4 for the derivation.

Taylor response to normal shifts. Having identified the P-defined Taylor quantities, we next ask what kinds of differences between P and Q they can capture. For the Gaussian normal-coordinate model $\boldsymbol{\Xi}_P \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}^{-1})$ with $\boldsymbol{K}\boldsymbol{v}_j = \kappa_j \boldsymbol{v}_j$, let $b_{P,j}(\boldsymbol{\xi}) = \boldsymbol{v}_j^\top \boldsymbol{g}_P(\boldsymbol{\xi})$. If $\boldsymbol{\Xi}_Q \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_Q)$, then

$$\begin{aligned} b_{P,j}(\boldsymbol{\Xi}_P) &\sim \mathcal{N}(0, \kappa_j), \\ b_{P,j}(\boldsymbol{\Xi}_Q) &\sim \mathcal{N}(-\kappa_j \mu_j, \kappa_j^2 \tau_j^2), \end{aligned} \tag{13}$$

where $\mu_j = \boldsymbol{v}_j^\top \boldsymbol{\mu}$ and $\tau_j^2 = \boldsymbol{v}_j^\top \boldsymbol{\Sigma}_Q \boldsymbol{v}_j$. Hence the squared coefficient used by NLTS changes in distribution whenever $\mu_j \neq 0$ or $\tau_j^2 \neq \kappa_j^{-1}$. By Prop. 2, the ambient feature a_j^2 has the same leading-order interpretation; the exact statement and proof are given in App. A.2.6. By contrast, Prop. 3 shows that manifold-marginal reweighting with the normal conditional fixed leaves the local normal Taylor structure unchanged at each manifold point.

Proposition 3 (Normal-conditional invariance under manifold-marginal reweighting)**.** *For the tubular joint density $f_P(\boldsymbol{m}, \boldsymbol{\xi}) = q_P(\boldsymbol{m})\rho_P(\boldsymbol{\xi} \mid \boldsymbol{m})$, let $w(\boldsymbol{m}) > 0$ depend only on $\boldsymbol{m}$, and set $Z := \int_{\mathcal{M}} w(\boldsymbol{m}) q_P(\boldsymbol{m})\, d\mathrm{vol}_{\mathcal{M}}(\boldsymbol{m})$, with $0 < Z < \infty$. Define the reweighted joint density by $f_Q(\boldsymbol{m}, \boldsymbol{\xi}) := Z^{-1} w(\boldsymbol{m}) f_P(\boldsymbol{m}, \boldsymbol{\xi})$. Then, for every fixed $\boldsymbol{m}$,*

$$\rho_Q(\boldsymbol{\xi} \mid \boldsymbol{m}) = \rho_P(\boldsymbol{\xi} \mid \boldsymbol{m}), \tag{14}$$

and, wherever the required derivatives exist, the two distributions have identical normal-coordinate scores and normal-coordinate negative-log-density Hessians.

The proof is given in App. A.2.5. This pointwise invariance does not, however, imply population-level invariance: reweighting manifold locations can still change the distribution of the signature introduced below and yield a nonzero population discrepancy. The precise detectability condition is given in Eq. (18).

3.2 NLTS signature and fixed-reference comparison

In principle, retaining (a_j, λ_j) for all Hessian eigendirections would retain the first- and second-order Taylor coefficients in the Hessian eigenbasis at each point. In practice, however, a full Hessian eigendecomposition at every sample is computationally expensive, so we retain only the top $k \ll D$ dominant directions. Here, k is a computational truncation parameter rather than an estimate of the codimension. In the fixed-reference construction that follows, we suppress the σ subscript and write $\boldsymbol{s}_P$, $\boldsymbol{H}_P$, and a_j for $\boldsymbol{s}_{P,\sigma}$, $\boldsymbol{H}_{P,\sigma}$, and $a_{j,\sigma}$, respectively.

To obtain a feature invariant to the arbitrary sign of each eigenvector, we use a_j^2. This removes the eigenvector-sign ambiguity at the cost of discarding the sign of the linear coefficient. We further define

$$r_k(\boldsymbol{x}) := \|(\boldsymbol{I} - \boldsymbol{U}_k(\boldsymbol{x})\boldsymbol{U}_k(\boldsymbol{x})^\top)\boldsymbol{s}_P(\boldsymbol{x})\|_2^2 = \|\boldsymbol{s}_P(\boldsymbol{x})\|_2^2 - \sum_{j=1}^{k} a_j(\boldsymbol{x})^2, \quad \boldsymbol{U}_k(\boldsymbol{x}) := [\boldsymbol{u}_1(\boldsymbol{x}), \ldots, \boldsymbol{u}_k(\boldsymbol{x})]. \tag{15}$$

The quantity r_k is not restricted to the normal component; it represents the total score energy remaining outside the top-k eigenspace. It therefore retains the remaining score energy outside the selected eigenspace without computing additional eigenvectors.

Combining these components, we define the sample-level signature as

$$\mathbf{S}_{P,k}(\boldsymbol{x}) = [\lambda_1(\boldsymbol{x}), \ldots, \lambda_k(\boldsymbol{x}), a_1(\boldsymbol{x})^2, \ldots, a_k(\boldsymbol{x})^2, r_k(\boldsymbol{x})]. \tag{16}$$

For repeated eigenvalues, the individual a_j^2 coordinates are basis dependent and therefore do not have the coordinate-wise interpretation of Prop. 2. The corresponding block energy is basis invariant; this case and a blockwise alternative are discussed in App. A.4.3. If a degenerate eigenvalue block straddles the top-k boundary, the selected top-k projector can itself be basis dependent. The Hessian enters the signature in two distinct ways: its eigendirections define the projected-score coefficients a_j, while its eigenvalues provide the curvature coefficients λ_j. Their complementary roles are examined empirically in Sec. 4.1 and App. E.2. Further properties of the representation are discussed in App. A.4.

We apply the same P-defined map to both reference and comparison samples and compare the induced distributions after reference-only standardization T_P using the maximum mean discrepancy (MMD) (Gretton et al., 2012):

$$\mathcal{D}_{P,k}(P, Q) := \mathrm{MMD}_\kappa^2 \left((T_P \circ \mathbf{S}_{P,k})_\# P, (T_P \circ \mathbf{S}_{P,k})_\# Q \right). \tag{17}$$

Thus, the comparison is fixed-reference: only the score and Hessian of P are required, and no corresponding quantities are estimated for Q. Details of the standardization and multiscale RBF–MMD are given in App. A.5.1.

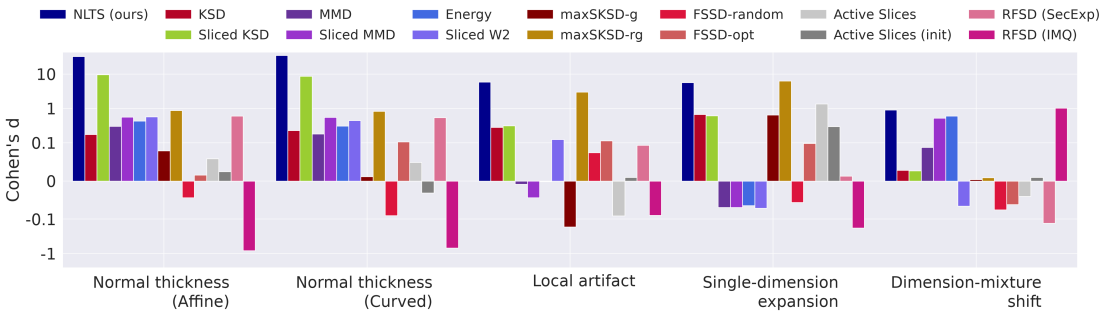


Figure 2: Cohen's d across the five target synthetic shifts. NLTS (**blue**) maintains substantial separation across the controlled changes in local density geometry.

Population-level detectability. The Gaussian analysis in App. A.2.6 characterizes when a canonical normal shift changes an individual fixed-reference Taylor feature. At the representation level, because κ is characteristic (Gretton et al., 2012),

$$\mathcal{D}_{P,k}(P,Q) > 0 \quad \Longleftrightarrow \quad (T_P \circ \mathbf{S}_{P,k})_{\#}P \neq (T_P \circ \mathbf{S}_{P,k})_{\#}Q. \tag{18}$$

This gives the representation-level detectability condition: NLTS has positive population discrepancy exactly when the two induced signature distributions differ. Equivalent characterizations, including an explicit invisible alternative, are given in App. A.3.

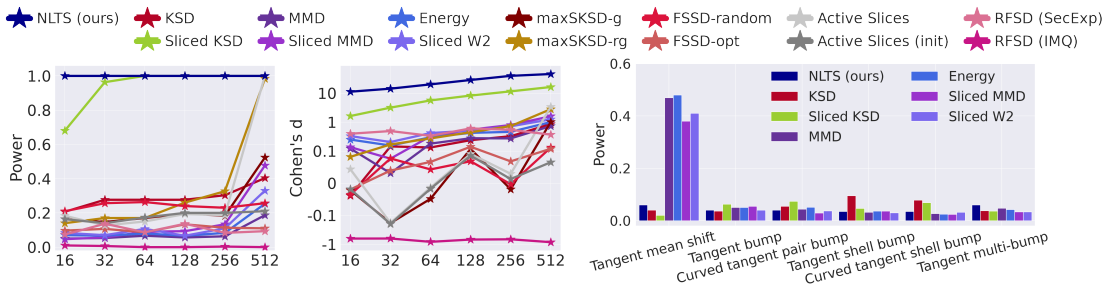
Practical score and Hessian computation. When the reference score is not analytically available, we use a learned score $\widehat{\boldsymbol{s}}_P$ in place of $\boldsymbol{s}_P$ (e.g., via denoising score matching (Vincent, 2011)). The required spectral directions can be estimated without forming a dense Hessian, using matrix-free directional products and iterative eigensolvers. In our experiments, we use $|\widehat{\lambda}_j|$ for the curvature features to improve numerical stability. The relationship between the theoretical symmetric Hessian and the benchmark-specific learned operators, together with their numerical stabilization, is described in App. A.5.3. The population-level characterization above applies analogously to the practical absolute-curvature variant after replacing $\mathbf{S}_{P,k}$ by the corresponding feature map.

4 Experiments

Across three experiments, we evaluate (i) directional sensitivity under controlled geometric shifts, (ii) performance on CIFAR-10-C image shifts (Hendrycks & Dietterich, 2019), and (iii) associations with local-PCA diagnostics in single-cell drug perturbations. Evaluation uses empirical power for synthetic shifts, alternative rejection rate for CIFAR-10-C, detection power for single-cell, and Cohen's d for synthetic/CIFAR-10-C; for target shifts, rejection/power values closer to 1 and larger Cohen's d indicate stronger detection, whereas low rejection is desired for tangential specificity controls. Baselines and common protocols are given in App. B and App. C. Within each experiment, all Stein-based baselines and NLTS use the same reference score. Unless stated otherwise, we use $k = 2$ and multiscale RBF–MMD in signature space. Sensitivity to these fixed design choices is examined on the controlled synthetic benchmark in App. E.1.

4.1 Synthetic geometric shifts: Controlled test of directional sensitivity

The synthetic benchmark controls the tangent/normal structure and perturbation direction, providing a direct test of the directional sensitivity of NLTS. We separate the evaluation into five target geometric shifts (Fig. 2) and six tangential specificity controls (Fig. 3b); the five target shifts modify normal-direction structure or local dimensionality and are referred to below collectively as normal-geometry shifts. Fig. 1 illustrates the target mechanisms together with a representative tangential control. For the five target shifts, we use ambient dimension $D = 256$ and tangent dimension $d = 8$ except for Dimension-mixture shift, which uses $D = 1024$ and mixes components with tangent dimensions $d \in \{128, 256, 512\}$. Data-generating mechanisms and evaluation settings are given in App. D.1–D.2. Reference scores and Hessian–vector products are evaluated analytically, removing score-model approximation error from this test.



(a) Curved normal thickness: power (left), Cohen's d (right)

(b) Tangential specificity controls

Figure 3: Synthetic specificity and sample-size behavior. (a) Sample-size dependence for Curved normal thickness ($D = 256$); left: empirical power, right: Cohen's d. (b) Empirical rejection rates across six tangential specificity controls ($D = 256$, $d = 8$, $n = 256$). NLTS (**blue**) maintains low response across the tangential controls while showing strong and sample-efficient response to the normal-thickness shift.

Normal-geometry shifts. Across the five target shifts, NLTS maintains substantial Cohen's d (Fig. 2). In the corresponding testing results, NLTS reaches power 1.00 for both normal-thickness shifts, Local artifact, and Single-dimension expansion, with no systematic Type-I error inflation in the validated null settings. On Curved normal thickness, its response is already strong at small sample sizes and remains high as n increases; Sliced KSD also reaches power 1.00 for $n \geq 64$ (Fig. 3a). Scenario-wise numerical results and the corresponding Affine sample-size sweep are given in App. D.3.

Tangential controls. Across six tangential perturbations, NLTS remains near the nominal rejection level, including mean-shift, pair-bump, shell, multi-bump, and curved variants (Fig. 3b). Full scenario definitions and complementary Cohen's d results are given in App. D.4. Taken together, NLTS is the only evaluated method that shows sensitivity across all five normal-geometry shifts while maintaining low response across all six tangential controls. By contrast, the baseline methods show scenario-dependent responses within both families, responding to some normal-geometry and tangential shifts but not to others. Thus, their detection patterns do not consistently distinguish normal-geometry from tangential changes.

First- and second-order Taylor components. Isolating the first-order projected-score and second-order curvature components reveals complementary behavior: on Curved normal thickness, their empirical powers are 1.00 and 0.05, respectively, whereas on Local artifact they are 0.11 and 0.98. Full component-wise results and interpretation are given in App. E.2.

4.2 CIFAR-10-C: High-dimensional real-image stress test

We evaluate whether NLTS can detect corruption-induced distributional changes in high-dimensional natural images ($D = 3072$) using CIFAR-10-C (Hendrycks & Dietterich, 2019). The tangent–normal decomposition of each corruption relative to the clean-image distribution is unknown. Nevertheless, CIFAR-10-C corruptions can introduce variations that are not well represented by variation along the clean-data manifold and may therefore contain nontrivial normal-direction components. We use this benchmark as a high-dimensional real-image stress test of NLTS, rather than as a setting with known tangent–normal ground truth. We derive the learned reference vector field from a publicly available pretrained DDPM (Ho et al., 2020) trained on clean CIFAR-10 (Krizhevsky et al., 2009)[1]. Dataset details, corruption types, baseline methods, and the CIFAR-10-C-specific score and directional-derivative approximations are described in Apps. F.1 and F.2.

Results. At severity 1 and $n = 128$, NLTS is the only method that maintains consistently positive and substantial Cohen's d across all 19 corruption types (Fig. 4); every baseline exhibits weak or negative separation for at least one corruption. This pattern persists in the full baseline comparison (Fig. 10 in App. F.3).

[1] `huggingface.co/google/ddpm-cifar10-32`

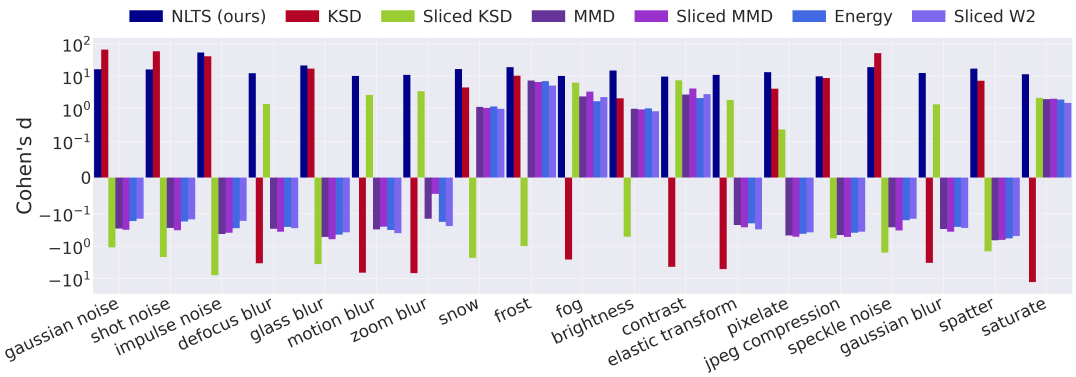
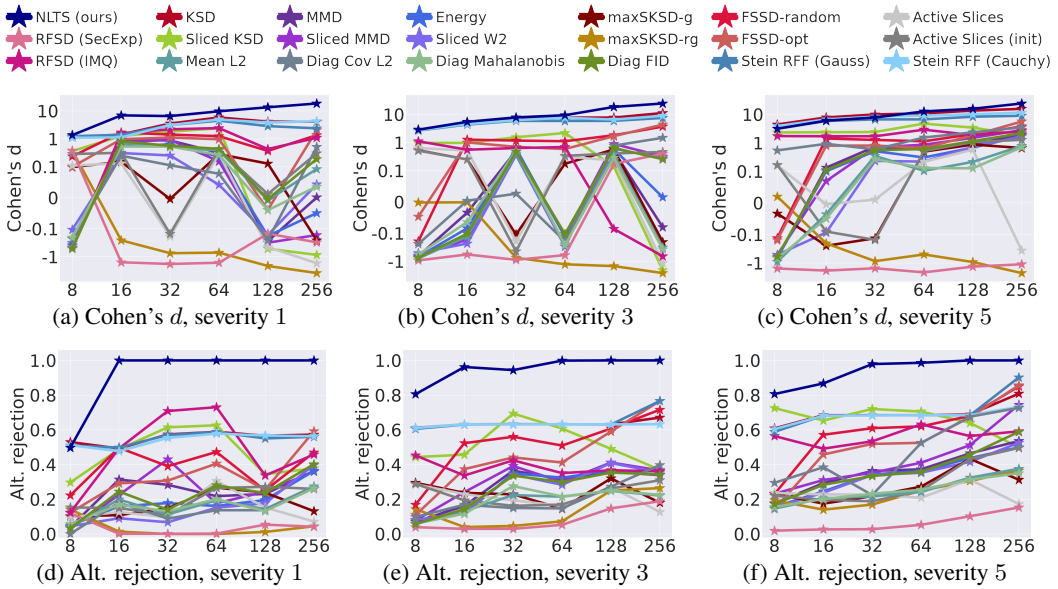


Figure 4: CIFAR-10-C corruption-wise Cohen's d at severity 1 and $n = 128$. NLTS (**blue**) maintains positive separation across all 19 corruptions.



(a) Cohen's d, severity 1 (b) Cohen's d, severity 3 (c) Cohen's d, severity 5

(d) Alt. rejection, severity 1 (e) Alt. rejection, severity 3 (f) Alt. rejection, severity 5

Figure 5: CIFAR-10-C sample-size dependence at severities 1, 3, and 5. Across the 19 corruption types, the top row shows median Cohen's d and the bottom row mean alternative rejection rate, both as functions of the number of samples per distribution. NLTS (**blue**) shows consistently large Cohen's d values and high rejection rates across severities and sample sizes.

Across severities 1, 3, and 5, NLTS also shows strong sample efficiency (Fig. 5). At severity 1, its mean alternative rejection rate reaches 1.00 from $n = 16$ onward; at severities 3 and 5, it is already high at the smallest sample size and approaches saturation as n increases. KSD attains similar or larger Cohen's d values in some of the smallest-sample settings; the $n = 8$, severity-1 numerical comparison and corresponding Type-I errors are reported in Table 11.

Together, these results show that NLTS remains effective for diverse shifts in high-dimensional image data, even when the underlying tangent–normal structure is not explicitly available.

4.3 Single-cell drug perturbations: Association with local-PCA diagnostics

We examine whether NLTS detection sensitivity is associated with drug-induced local geometric changes. We use sci-Plex3 K562 drug-perturbation data (Srivatsan et al., 2020) from the scPerturb harmonized release (Peidli et al., 2024) and analyze 10 drugs across four doses. Separately from NLTS, reference-only local-PCA bases define three geometry descriptors: Effective-rank contraction, Normal thickness change, and Tangent mean energy, which we compare with NLTS and KSD

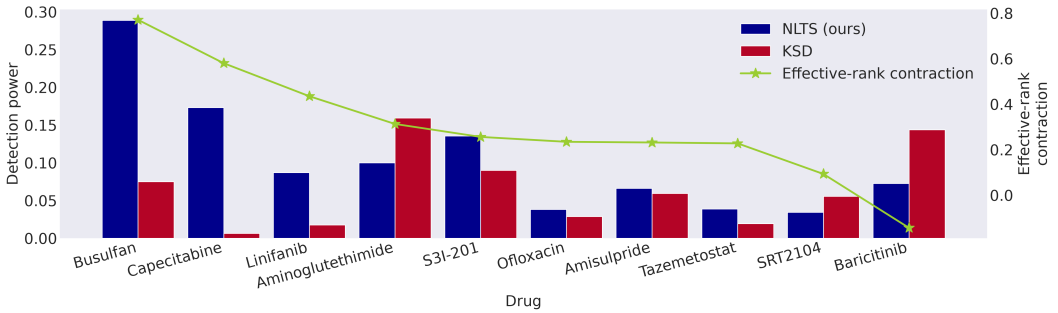


Figure 6: Drug-level detection power (bars, left axis) and Effective-rank contraction (line, right axis). NLTS (**blue**) shows a clearer positive association with contraction than KSD.

Table 1: Drug-level associations between local-PCA descriptors and detection power across 10 drugs. Partial Spearman correlation is used when an adjustment is listed; otherwise ordinary Spearman correlation is reported. p-values are nominal and uncorrected.

		NLTS		KSD	
Descriptor	Adjustment	ρ	p	ρ	p
Effective-rank contraction	Tangent mean energy	0.692	0.0388	0.046	0.907
Normal thickness change	—	-0.600	0.0667	0.067	0.855
Tangent mean energy	Effective-rank contraction	0.197	0.611	-0.371	0.325

detection power using the same reference score model. Dataset preprocessing, local-PCA descriptor definitions, score-model training, and sampling protocols are described in Apps. G.1, G.2, and G.3.

Results. Effective-rank contraction is positively associated with NLTS power after adjusting for Tangent mean energy ($\rho = 0.692$, $p = 0.0388$), whereas the corresponding KSD correlation is near zero ($\rho = 0.046$, $p = 0.907$) (Table 1). Given the 10-drug sample and uncorrected nominal p-values, we interpret this as exploratory evidence that NLTS sensitivity covaries with reference-defined local geometry, not as direct validation of tangent–normal selectivity. Additional analyses are in App. G.4.

5 Discussion, Limitations, and Conclusion

NLTS is designed for geometric selectivity rather than universal two-sample sensitivity. Because both distributions are mapped through the same reference-derived signature, changes not expressed in the retained Taylor components may be attenuated or lost. This deliberate compression is also what allows the comparison to emphasize changes in local normal density geometry rather than treating all distributional differences equally. This provides a way to incorporate reference-defined local geometry directly into distribution comparison.

Limitations. (i) Our theory is local and assumes a smooth small-noise manifold with a normal–tangent spectral gap; the coordinate-wise first-order interpretation additionally requires each retained normal eigenvalue to be separated from the rest of the normal spectrum. It does not cover alternatives placing substantial mass far from the reference distribution. (ii) Practice uses learned scores and approximate Hessian directions, but we do not quantify how these errors propagate to the NLTS signature or test power. (iii) Tangent–normal ground truth is available only synthetically; CIFAR-10-C and single-cell therefore provide indirect, rather than direct, evidence of geometric selectivity.

Conclusion. NLTS illustrates how a reference-derived representation can steer two-sample sensitivity toward differences expressed in local normal density geometry.

AI Disclosure

We used OpenAI's ChatGPT in several parts of the research and writing workflow. For the synthetic experiments, it assisted with the mathematical formalization of author-designed perturbation scenarios and with generating the corresponding implementation code; the experimental scenarios themselves were designed by the authors. For the theoretical development, ChatGPT was used to critically examine mathematical claims and proofs, including checking assumptions, searching for possible errors, and considering counterexamples. The final mathematical statements and proofs were independently reviewed and verified by the authors.

For the single-cell experiments, ChatGPT was used to obtain background orientation in the relevant biology, to assist in evaluating candidate datasets, and to critically review the experimental design. It was also used for code generation, modification, debugging, and review throughout the experimental workflow. AI-assisted code was tested by the authors through execution checks and inspection of the resulting outputs.

ChatGPT was additionally used to improve grammar and readability in the writing of the paper. The scientific content and claims were determined by the authors. It was also used to summarize and compare existing literature and to search for potentially relevant papers and other information sources; citations and factual claims used in the paper were checked by the authors against the original papers or primary sources.

We did not use generative AI to develop the theoretical model or conceptual framework, provide critical ingredients for proofs, propose or refine research hypotheses, clean or reformat datasets, perform qualitative or thematic data analysis, or interpret experimental results. We reviewed all AI-assisted work and take responsibility for the final content of this paper, including its text, mathematical claims, code, and experimental artifacts.

Reproducibility Statement

Implementation details of the NLTS representation, including signature standardization, finite-sample discrepancy computation, and practical score and Hessian approximations, are given in App. A.5.1–A.5.3. Baseline definitions, implementation settings, and locked revisions of public reference implementations are documented in App. B, and the common resampling, calibration, and evaluation procedures are specified in App. C. Dataset construction or preprocessing and experiment-specific protocols are provided for the synthetic experiments in Apps. D.1–D.2, for CIFAR-10-C in Apps. F.1 and F.2, and for the single-cell experiments in Apps. G.1–G.3. The code used in all experiments will be made publicly available.

References

Rahul Atul Bhope, K. R. Jayaram, Praveen Venkateswaran, and Nalini Venkatasubramanian. Shift happens: Mixture of experts based continual adaptation in federated learning. In *Proceedings of the 26th International Middleware Conference*, Middleware '25, pp. 455 – 468, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715549. doi: 10.1145/3721462.3770784. URL https://doi.org/10.1145/3721462.3770784.

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein Barycenters of Measures. *Journal of Mathematical Imaging and Vision*, 1(51):22–45, 2015. doi: 10.1007/s10851-014-0506-3.

Bradley CA Brown, Anthony L. Caterini, Brendan Leigh Ross, Jesse C Cresswell, and Gabriel Loaiza-Ganem. Verifying the union of manifolds hypothesis for image data. In *The Eleventh International Conference on Learning Representations*, 2023.

Kacper Chwialkowski, Heiko Strathmann, and Arthur Gretton. A kernel test of goodness of fit. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2606–2615, New York, New York, USA, 20–22 Jun 2016. PMLR.

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.

Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

Wenbo Gong, Yingzhen Li, and José Miguel Hernández-Lobato. Sliced kernelized stein discrepancy. In *International Conference on Learning Representations*, 2021a.

Wenbo Gong, Kaibo Zhang, Yingzhen Li, and Jose Miguel Hernandez-Lobato. Active slices for sliced stein discrepancy. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3766–3776. PMLR, 18–24 Jul 2021b.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=HJz6tiCqYm`.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Jonathan Huggins and Lester Mackey. Random feature stein discrepancies. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL `https://proceedings.neurips.cc/paper_files/paper/2018/file/0f840be9b8db4d3fbd5ba2ce59211f55-Paper.pdf`.

Wittawat Jitkrittum, Wenkai Xu, Zoltan Szabo, Kenji Fukumizu, and Arthur Gretton. A linear-time kernel goodness-of-fit test. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, 2nd edition, 2012. ISBN 978-1-4419-9981-8. doi: 10.1007/978-1-4419-9982-5.

Feng Liu, Wenkai Xu, Jie Lu, Guangquan Zhang, Arthur Gretton, and Danica J. Sutherland. Learning deep kernels for non-parametric two-sample tests. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6316–6326. PMLR, 13–18 Jul 2020.

Qiang Liu, Jason Lee, and Michael Jordan. A kernelized stein discrepancy for goodness-of-fit tests. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 276–284, New York, New York, USA, 20–22 Jun 2016. PMLR.

Rachel Luo, Rohan Sinha, Yixiao Sun, Ali Hindy, Shengjia Zhao, Silvio Savarese, Edward Schmerling, and Marco Pavone. Online distribution shift detection via recency prediction. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 16251–16263, 2024. doi: 10.1109/ICRA57147.2024.10611114.

Genki Osada. Local hessian spectral filtering for robust intrinsic dimension estimation. In *Forty-third International Conference on Machine Learning*, 2026.

Stefan Peidli, Tessa D. Green, Ciyue Shen, Torsten Gross, Joseph Min, Samuele Garda, Bo Yuan, Linus J. Schumacher, Jake P. Taylor-King, Debora S. Marks, Augustin Luna, Nils Blüthgen, and Chris Sander. scperturb: harmonized single-cell perturbation data. *Nature Methods*, 21(3): 531–540, Mar 2024. ISSN 1548-7105. doi: 10.1038/s41592-023-02144-y.

Stephan Rabanser, Stephan Günnemann, and Zachary Lipton. Failing loudly: An empirical study of methods for detecting dataset shift. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/846c260d715e5b854ffad5f70a516c88-Paper.pdf.

Aaditya Ramdas, Sashank J. Reddi, Barnabás Póczos, Aarti Singh, and Larry Wasserman. On the decreasing power of kernel and distance based nonparametric hypothesis tests in high dimensions. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pp. 3571 – 3577. AAAI Press, 2015. ISBN 0262511290.

Sanjay R Srivatsan, José L McFaline-Figueroa, Vijay Ramani, Lauren Saunders, Junyue Cao, Jonathan Packer, Hannah A Pliner, Dana L Jackson, Riza M Daza, Lena Christiansen, Fan Zhang, Frank Steemers, Jay Shendure, and Cole Trapnell. Massively multiplex chemical transcriptomics at single-cell resolution. *Science (New York, N.Y.)*, 367(6473):45–51, 2020. ISSN 1095-9203. doi: 10.1126/science.aax6234.

Jan Pawel Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. Diffusion models encode the intrinsic dimension of data manifolds. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 46412–46440. PMLR, 21–27 Jul 2024.

Gábor J. Székely and Maria L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013. ISSN 0378-3758. doi: https://doi.org/10.1016/j.jspi.2013.03.018.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance. *Bernoulli*, 25(4A):pp. 2620–2648, 2019. ISSN 13507265, 15739759.

A Supplementary theory, method, and experimental details

This appendix collects technical definitions, proofs, baseline details, and experiment-specific protocols omitted from the main text for readability.

A.1 Geometric and spectral preliminaries

A.1.1 Tubular coordinates and change of variables

Let $\mathcal{M} \subset \mathbb{R}^D$ be a smooth d-dimensional embedded manifold. Because the normal bundle need not admit a global smooth frame, we fix a local coordinate patch $\mathcal{U} \subset \mathcal{M}$ and a smooth orthonormal normal frame $\boldsymbol{U}_N(\boldsymbol{m}) \in \mathbb{R}^{D\times r}$ on $\mathcal{U}$, where $r = D - d$. We follow the standard geometric background, including the tubular-neighborhood theorem, in (Lee, 2012). On a sufficiently small local tubular neighborhood,

$$F(\boldsymbol{m}, \boldsymbol{\xi}) = \boldsymbol{m} + \boldsymbol{U}_N(\boldsymbol{m})\boldsymbol{\xi}, \qquad \boldsymbol{\xi} \in \mathbb{R}^r, \tag{19}$$

is a local diffeomorphism.

With respect to manifold volume and Lebesgue measure, write the induced joint density as

$$f_P(\boldsymbol{m}, \boldsymbol{\xi}) = q_P(\boldsymbol{m})\rho_P(\boldsymbol{\xi} \mid \boldsymbol{m}). \tag{20}$$

If $J_F(\boldsymbol{m}, \boldsymbol{\xi})$ denotes the volume Jacobian of F, then

$$p_P(F(\boldsymbol{m}, \boldsymbol{\xi})) = \frac{q_P(\boldsymbol{m})\rho_P(\boldsymbol{\xi} \mid \boldsymbol{m})}{J_F(\boldsymbol{m}, \boldsymbol{\xi})}, \tag{21}$$

and hence

$$-\log p_P(F(\boldsymbol{m}, \boldsymbol{\xi})) = -\log q_P(\boldsymbol{m}) - \log \rho_P(\boldsymbol{\xi} \mid \boldsymbol{m}) + \log J_F(\boldsymbol{m}, \boldsymbol{\xi}). \tag{22}$$

A.1.2 Principal angles and eigenspace perturbation

Let $E, F \subset \mathbb{R}^D$ be two k-dimensional subspaces with orthonormal basis matrices $\boldsymbol{U}_E, \boldsymbol{U}_F \in \mathbb{R}^{D\times k}$. If the singular values of $\boldsymbol{U}_E^\top \boldsymbol{U}_F$ are $\cos\theta_1, \ldots, \cos\theta_k$, then $\theta_1, \ldots, \theta_k$ are the principal angles. We measure subspace error by

$$\|\sin\Theta(E, F)\|_{\text{op}} = \max_i \sin\theta_i. \tag{23}$$

Consider symmetric matrices $\boldsymbol{H}$ and $\widehat{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{E}_H$, and let $\gamma > 0$ be the eigengap separating the target eigenspace from the remainder of the spectrum. When the perturbation is sufficiently small relative to γ, a Davis–Kahan-type bound (Davis & Kahan, 1970) gives

$$\|\sin\Theta(\widehat{E}, E)\|_{\text{op}} \le C\frac{\|\boldsymbol{E}_H\|_{\text{op}}}{\gamma} \tag{24}$$

for a constant $C > 0$.

A.2 Supplementary method derivations and proofs

A.2.1 Local quadratic model for the normal conditional density

For the reference family P_σ used in the method, the manifold marginal is fixed across σ:

$$q_{P,\sigma}(\boldsymbol{m}) \equiv q_P(\boldsymbol{m}). \tag{25}$$

The normal conditional density is assumed to concentrate as $\sigma \to 0$, while the marginal distribution on the manifold remains fixed.

Consider Eq. (3),

$$\ell^N_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m}) = \ell^N_{P,\sigma}(\boldsymbol{0} \mid \boldsymbol{m}) + \frac{1}{2\sigma^2}\boldsymbol{\xi}^\top \boldsymbol{A}(\boldsymbol{m})\boldsymbol{\xi} + \widetilde{r}_\sigma(\boldsymbol{m}, \boldsymbol{\xi}). \tag{26}$$

On the local patch, assume that $\boldsymbol{A}(\boldsymbol{m})$ is uniformly positive definite:

$$0 < a_* \le \lambda_{\min}(\boldsymbol{A}(\boldsymbol{m})) \le \lambda_{\max}(\boldsymbol{A}(\boldsymbol{m})) \le a^* < \infty. \tag{27}$$

We also assume that q_P is positive on the local patch and that the required tangential derivatives of $\ell^N_{P,\sigma}(\mathbf{0} \mid \boldsymbol{m})$ are uniformly $O(1)$ as $\sigma \to 0$, while the quantities entering the expansions, including $\boldsymbol{A}$, $\boldsymbol{U}_N$, $\log q_P$, $\log J_F$, and the remainder terms, have the required tangential, normal, and mixed derivatives uniformly bounded on the fixed patch and the typical tube. By construction, $\widetilde{r}_\sigma(\boldsymbol{m}, \mathbf{0}) = 0$. We further assume $\nabla_{\boldsymbol{\xi}} \widetilde{r}_\sigma(\boldsymbol{m}, \mathbf{0}) = \mathbf{0}$, and

$$\sup_{\boldsymbol{m}} \left\| \nabla^2_{\boldsymbol{\xi}} \widetilde{r}_\sigma(\boldsymbol{m}, \mathbf{0}) \right\|_{\mathrm{op}} \le C. \tag{28}$$

Hence

$$\boldsymbol{K}_{P,\sigma}(\boldsymbol{m}, \mathbf{0}) = \nabla^2_{\boldsymbol{\xi}} \ell^N_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m})\big|_{\boldsymbol{\xi}=\mathbf{0}} = \frac{\boldsymbol{A}(\boldsymbol{m})}{\sigma^2} + O(1). \tag{29}$$

To extend the same leading structure throughout the typical tube $\|\boldsymbol{\xi}\|_2 \le C_0\sigma$, assume

$$\sup_{\|\boldsymbol{\xi}\| \le C_0\sigma} \|\nabla_{\boldsymbol{\xi}} \widetilde{r}_\sigma(\boldsymbol{m}, \boldsymbol{\xi})\|_2 = O(1), \qquad \sup_{\|\boldsymbol{\xi}\| \le C_0\sigma} \left\| \nabla^2_{\boldsymbol{\xi}} \widetilde{r}_\sigma(\boldsymbol{m}, \boldsymbol{\xi}) \right\|_{\mathrm{op}} = O(1), \tag{30}$$

uniformly over the patch. Then, uniformly in the tube,

$$\boldsymbol{g}_{P,\sigma}(\boldsymbol{m}, \boldsymbol{\xi}) = -\frac{1}{\sigma^2} \boldsymbol{A}(\boldsymbol{m})\boldsymbol{\xi} + O(1), \qquad \boldsymbol{K}_{P,\sigma}(\boldsymbol{m}, \boldsymbol{\xi}) = \frac{1}{\sigma^2} \boldsymbol{A}(\boldsymbol{m}) + O(1). \tag{31}$$

A.2.2 Proof of Prop. 1

By the tubular change of variables,

$$-\log p_{P,\sigma}(F(\boldsymbol{m}, \boldsymbol{\xi})) = -\log q_P(\boldsymbol{m}) + \ell^N_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m}) + \log J_F(\boldsymbol{m}, \boldsymbol{\xi}). \tag{32}$$

At $\boldsymbol{\xi} = \mathbf{0}$, i.e., on the manifold, the normal–normal block of the ambient Hessian has leading term $\sigma^{-2}\boldsymbol{A}(\boldsymbol{m})$ by Eq. (29). Under the stated regularity assumptions, the tangent–tangent and tangent–normal blocks, as well as the remaining normal corrections arising from q_P, the tubular map, the volume Jacobian, and the manifold geometry, are $O(1)$. Therefore

$$\boldsymbol{H}_{P,\sigma}(\boldsymbol{m}) = \frac{1}{\sigma^2} \boldsymbol{U}_N(\boldsymbol{m}) \boldsymbol{A}(\boldsymbol{m}) \boldsymbol{U}_N(\boldsymbol{m})^\top + \boldsymbol{R}_\sigma(\boldsymbol{m}), \qquad \|\boldsymbol{R}_\sigma(\boldsymbol{m})\|_{\mathrm{op}} = O(1). \tag{33}$$

The leading operator

$$\boldsymbol{H}_{0,\sigma}(\boldsymbol{m}) = \frac{1}{\sigma^2} \boldsymbol{U}_N(\boldsymbol{m}) \boldsymbol{A}(\boldsymbol{m}) \boldsymbol{U}_N(\boldsymbol{m})^\top \tag{34}$$

has eigenvalues $\alpha_j(\boldsymbol{m})/\sigma^2$ on $N_{\boldsymbol{m}}\mathcal{M}$ and zero on its orthogonal complement. Eq. (27) therefore gives a normal-to-tangent spectral gap of order $\Theta(\sigma^{-2})$. Weyl's inequality yields $\lambda_j(\boldsymbol{H}_{P,\sigma}(\boldsymbol{m})) = \alpha_j/\sigma^2 + O(1)$. Applying the eigenspace perturbation bound above to $\boldsymbol{R}_\sigma(\boldsymbol{m}) = O(1)$ gives

$$\left\| \sin\Theta\left(\widehat{N}_\sigma(\boldsymbol{m}), N_{\boldsymbol{m}}\mathcal{M} \right) \right\|_{\mathrm{op}} = O(\sigma^2). \tag{35}$$

For a typical-tube point $\boldsymbol{x} = F(\boldsymbol{m}, \boldsymbol{\xi})$ with $\|\boldsymbol{\xi}\| = O(\sigma)$, the uniform tube regularity assumptions give the ambient decomposition

$$\boldsymbol{H}_{P,\sigma}(\boldsymbol{x}) = \frac{1}{\sigma^2} \boldsymbol{U}_N(\boldsymbol{m}) \boldsymbol{A}(\boldsymbol{m}) \boldsymbol{U}_N(\boldsymbol{m})^\top + \boldsymbol{E}_\sigma(\boldsymbol{x}), \qquad \|\boldsymbol{E}_\sigma(\boldsymbol{x})\|_{\mathrm{op}} = O(\sigma^{-1}). \tag{36}$$

The $O(\sigma^{-1})$ term accounts for the variation of the tubular frame, Jacobian, and mixed Hessian blocks away from the manifold center. Because the normal spectral gap remains $\Theta(\sigma^{-2})$, the same perturbation argument gives

$$\left\| \sin\Theta\left(\widehat{N}_\sigma(\boldsymbol{x}), N_{\boldsymbol{m}}\mathcal{M} \right) \right\|_{\mathrm{op}} = O(\sigma). \tag{37}$$

A.2.3 Derivation of Prop. 2

In tubular coordinates,

$$\log p_{P,\sigma}(F(\boldsymbol{m}, \boldsymbol{\xi})) = \log q_P(\boldsymbol{m}) + \log \rho_{P,\sigma}(\boldsymbol{\xi} \mid \boldsymbol{m}) - \log J_F(\boldsymbol{m}, \boldsymbol{\xi}). \tag{38}$$

Differentiating with respect to $\boldsymbol{\xi}$ at fixed $\boldsymbol{m}$ gives

$$\boldsymbol{U}_N(\boldsymbol{m})^\top \boldsymbol{s}_{P,\sigma}(F(\boldsymbol{m},\boldsymbol{\xi})) = \boldsymbol{g}_{P,\sigma}(\boldsymbol{m},\boldsymbol{\xi}) - \nabla_{\boldsymbol{\xi}} \log J_F(\boldsymbol{m},\boldsymbol{\xi}) = \boldsymbol{g}_{P,\sigma}(\boldsymbol{m},\boldsymbol{\xi}) + O(1). \tag{39}$$

In the typical tube, $\|\boldsymbol{s}_{P,\sigma}(\boldsymbol{x})\|_2 = O(\sigma^{-1})$. Let $\boldsymbol{A}(\boldsymbol{m})\boldsymbol{v}_j = \alpha_j \boldsymbol{v}_j$. To interpret an individual eigendirection, assume the internal eigengap condition

$$\min_{\ell \neq j} |\alpha_j - \alpha_\ell| \geq c_j > 0. \tag{40}$$

Under this internal eigengap, the corresponding leading eigenvalue is separated from the remaining eigenvalues by $\Theta(\sigma^{-2})$. The typical-tube perturbation of the leading operator is $O(\sigma^{-1})$, as in the proof of Prop. 1. Applying the same eigenspace perturbation argument to the one-dimensional eigenspace and choosing the sign consistently gives

$$\boldsymbol{u}_j(\boldsymbol{x}) = \boldsymbol{U}_N(\boldsymbol{m})\boldsymbol{v}_j + O(\sigma). \tag{41}$$

Consequently,

$$\boldsymbol{u}_j(\boldsymbol{x})^\top \boldsymbol{s}_{P,\sigma}(\boldsymbol{x}) = \boldsymbol{v}_j^\top \boldsymbol{g}_{P,\sigma}(\boldsymbol{m},\boldsymbol{\xi}) + O(1). \tag{42}$$

Eq. (31) further gives

$$\boldsymbol{v}_j^\top \boldsymbol{g}_{P,\sigma}(\boldsymbol{m},\boldsymbol{\xi}) = -\frac{\alpha_j}{\sigma^2}\xi_j + O(1), \qquad \xi_j = \boldsymbol{v}_j^\top \boldsymbol{\xi}, \tag{43}$$

and therefore

$$\boldsymbol{u}_j(\boldsymbol{x})^\top \boldsymbol{s}_{P,\sigma}(\boldsymbol{x}) = -\frac{\alpha_j}{\sigma^2}\xi_j + O(1). \tag{44}$$

A.2.4 Taylor expansion of the ambient negative log density

Let $\ell_{P,\sigma}(\boldsymbol{x}) := -\log p_{P,\sigma}(\boldsymbol{x})$. If $\ell_{P,\sigma}$ is twice continuously differentiable in a neighborhood of $\boldsymbol{x}$, the second-order Taylor expansion gives

$$\ell_{P,\sigma}(\boldsymbol{x}+\boldsymbol{\delta}) - \ell_{P,\sigma}(\boldsymbol{x}) = -\boldsymbol{s}_{P,\sigma}(\boldsymbol{x})^\top \boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^\top \boldsymbol{H}_{P,\sigma}(\boldsymbol{x})\boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|_2^2). \tag{45}$$

Writing $\boldsymbol{H}_{P,\sigma}(\boldsymbol{x}) = \boldsymbol{U}(\boldsymbol{x})\boldsymbol{\Lambda}(\boldsymbol{x})\boldsymbol{U}(\boldsymbol{x})^\top$, $\boldsymbol{z} = \boldsymbol{U}(\boldsymbol{x})^\top\boldsymbol{\delta}$, and $\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{U}(\boldsymbol{x})^\top \boldsymbol{s}_{P,\sigma}(\boldsymbol{x})$ gives

$$\ell_{P,\sigma}(\boldsymbol{x}+\boldsymbol{\delta}) - \ell_{P,\sigma}(\boldsymbol{x}) = -\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{z} + \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{\Lambda}(\boldsymbol{x})\boldsymbol{z} + o(\|\boldsymbol{\delta}\|_2^2). \tag{46}$$

Thus, score projections and Hessian eigenvalues are the first- and second-order coefficients of the same ambient negative-log-density increment. Under C^3 regularity with locally bounded third derivatives, the remainder can be strengthened to $O(\|\boldsymbol{\delta}\|_2^3)$; the weaker $o(\|\boldsymbol{\delta}\|_2^2)$ form suffices for the main text.

A.2.5 Proof of Prop. 3

Reweight the tubular joint density $f_P(\boldsymbol{m},\boldsymbol{\xi}) = q_P(\boldsymbol{m})\rho_P(\boldsymbol{\xi} \mid \boldsymbol{m})$ by a positive weight $w(\boldsymbol{m})$ depending only on $\boldsymbol{m}$, with the normalizing constant Z defined in Prop. 3:

$$f_Q(\boldsymbol{m},\boldsymbol{\xi}) = \frac{w(\boldsymbol{m})}{Z} f_P(\boldsymbol{m},\boldsymbol{\xi}). \tag{47}$$

For fixed $\boldsymbol{m}$,

$$\rho_Q(\boldsymbol{\xi} \mid \boldsymbol{m}) = \frac{f_Q(\boldsymbol{m},\boldsymbol{\xi})}{\int f_Q(\boldsymbol{m},\boldsymbol{\xi}')\, d\boldsymbol{\xi}'} = \rho_P(\boldsymbol{\xi} \mid \boldsymbol{m}). \tag{48}$$

Hence the normal-coordinate conditional density and its $\boldsymbol{\xi}$-derivatives are preserved pointwise. Because the reweighting may change the manifold marginal, however, this pointwise invariance does not imply $(\mathbf{S}_{P,k})_\# P = (\mathbf{S}_{P,k})_\# Q$ or a vanishing population discrepancy.

A.2.6 When normal shifts change fixed-reference Taylor coefficients

Fix $\boldsymbol{m}$ and consider the Gaussian normal-coordinate reference

$$\boldsymbol{\Xi}_P \sim \mathcal{N}(\mathbf{0}, \boldsymbol{K}^{-1}), \qquad \boldsymbol{K} \succ 0, \qquad \boldsymbol{K}\boldsymbol{v}_j = \kappa_j \boldsymbol{v}_j, \qquad \|\boldsymbol{v}_j\|_2 = 1. \tag{49}$$

Up to an additive constant, its normal-coordinate negative log density and score are

$$\ell_P^N(\boldsymbol{\xi}) = \frac{1}{2}\boldsymbol{\xi}^\top \boldsymbol{K}\boldsymbol{\xi}, \qquad \boldsymbol{g}_P(\boldsymbol{\xi}) = -\boldsymbol{K}\boldsymbol{\xi}. \tag{50}$$

Define the projected reference score

$$b_{P,j}(\boldsymbol{\xi}) := \boldsymbol{v}_j^\top \boldsymbol{g}_P(\boldsymbol{\xi}) = -\kappa_j \xi_j, \qquad \xi_j = \boldsymbol{v}_j^\top \boldsymbol{\xi}. \tag{51}$$

Along the same reference eigendirection,

$$\ell_P^N(\boldsymbol{\xi} + t\boldsymbol{v}_j) - \ell_P^N(\boldsymbol{\xi}) = -b_{P,j}(\boldsymbol{\xi})t + \frac{1}{2}\kappa_j t^2. \tag{52}$$

Thus $-b_{P,j}(\boldsymbol{\xi})$ is the linear Taylor coefficient and $\kappa_j/2$ is the quadratic coefficient along $\boldsymbol{v}_j$. By Prop. 2, under its eigengap condition, the corresponding ambient projected score a_j has the same leading-order behavior in the small-noise regime.

Now consider a Gaussian alternative

$$\boldsymbol{\Xi}_Q \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_Q), \qquad \boldsymbol{\Sigma}_Q \succ 0, \tag{53}$$

and write

$$\mu_j := \boldsymbol{v}_j^\top \boldsymbol{\mu}, \qquad \tau_j^2 := \boldsymbol{v}_j^\top \boldsymbol{\Sigma}_Q \boldsymbol{v}_j. \tag{54}$$

Proposition 4 (When Gaussian normal shifts change a fixed-reference Taylor coefficient)**.** *Under Eqs. (49)–(54),*

$$\begin{aligned} b_{P,j}(\boldsymbol{\Xi}_P) &\sim \mathcal{N}(0, \kappa_j), \\ b_{P,j}(\boldsymbol{\Xi}_Q) &\sim \mathcal{N}(-\kappa_j \mu_j, \kappa_j^2 \tau_j^2). \end{aligned} \tag{55}$$

Consequently,

$$b_{P,j}(\boldsymbol{\Xi}_P) \overset{d}{=} b_{P,j}(\boldsymbol{\Xi}_Q) \quad \Longleftrightarrow \quad \mu_j = 0 \text{ and } \tau_j^2 = \kappa_j^{-1}. \tag{56}$$

Moreover, the same condition is necessary and sufficient after the sign-invariant squaring used by the practical first-order NLTS feature:

$$b_{P,j}(\boldsymbol{\Xi}_P)^2 \overset{d}{=} b_{P,j}(\boldsymbol{\Xi}_Q)^2 \quad \Longleftrightarrow \quad \mu_j = 0 \text{ and } \tau_j^2 = \kappa_j^{-1}. \tag{57}$$

Proof. Because $b_{P,j}$ is linear, the two laws in Eq. (55) follow from the Gaussian projection formula. Equality of these Gaussian laws holds if and only if their means and variances agree, which gives Eq. (56).

For the squared variables, the reverse implication in Eq. (57) is immediate. For the forward implication, write

$$m = -\kappa_j \mu_j, \qquad s^2 = \kappa_j^2 \tau_j^2. \tag{58}$$

If the squared laws are equal, equality of their first moments gives

$$m^2 + s^2 = \kappa_j, \tag{59}$$

while equality of their second moments gives

$$m^4 + 6m^2 s^2 + 3s^4 = 3\kappa_j^2. \tag{60}$$

Substituting $s^2 = \kappa_j - m^2$ into the latter identity yields $-2m^4 = 0$. Hence $m = 0$ and $s^2 = \kappa_j$, which is equivalent to $\mu_j = 0$ and $\tau_j^2 = \kappa_j^{-1}$. □

Corollary 1 (Normal location shift)**.** *If $\boldsymbol{\Sigma}_Q = \boldsymbol{K}^{-1}$ and $\mu_j \neq 0$ for some j, then*

$$b_{P,j}(\boldsymbol{\Xi}_P)^2 \overset{d}{\neq} b_{P,j}(\boldsymbol{\Xi}_Q)^2, \tag{61}$$

and, in particular,

$$\mathbb{E}_Q[b_{P,j}(\boldsymbol{\Xi})^2] = \kappa_j + \kappa_j^2 \mu_j^2 > \kappa_j = \mathbb{E}_P[b_{P,j}(\boldsymbol{\Xi})^2]. \tag{62}$$

Thus, even a normal location shift that leaves the reference Hessian unchanged alters the fixed-reference squared first-order Taylor response.

Corollary 2 (Normal scale (thickness) shift)**.** *If $\boldsymbol{\mu} = \mathbf{0}$ and, for some j,*

$$\tau_j^2 = \boldsymbol{v}_j^\top \boldsymbol{\Sigma}_Q \boldsymbol{v}_j \neq \frac{1}{\kappa_j}, \tag{63}$$

then

$$b_{P,j}(\boldsymbol{\Xi}_P)^2 \overset{d}{\neq} b_{P,j}(\boldsymbol{\Xi}_Q)^2, \tag{64}$$

and, in particular,

$$\mathbb{E}_Q[b_{P,j}(\boldsymbol{\Xi})^2] = \kappa_j^2 \tau_j^2 \neq \kappa_j = \mathbb{E}_P[b_{P,j}(\boldsymbol{\Xi})^2]. \tag{65}$$

Thus, a normal scale / thickness shift also changes the fixed-reference squared first-order Taylor response.

In particular, if either condition holds for at least one retained direction $j \leq k$, then the joint law of $(b_{P,1}^2, \ldots, b_{P,k}^2)$ differs under P and Q, because equality of the joint laws would imply equality of every marginal law.

Distribution-specific Hessians must be distinguished from fixed-reference responses. For a quadratic Gaussian reference, the reference Hessian $\boldsymbol{K}$ is independent of $\boldsymbol{\xi}$, so changing the location or thickness of Q need not change the Hessian eigenvalues of the reference field. The change instead appears in the distribution of the reference first-order Taylor response evaluated at Q samples. These statements are exact for the Gaussian normal-coordinate model above; through Prop. 2, they describe the leading-order mechanism of the ambient projected-score feature in the small-noise regime. They are not, by themselves, a population-level detectability statement for the practical NLTS signature as a whole.

A.3 Population-level detectability of NLTS

The preceding Gaussian analysis characterizes when a canonical normal shift changes an individual fixed-reference Taylor feature. We now address the population-level question: whether the joint distribution of the standardized NLTS signature differs under P and Q. For compactness, write

$$\Phi_{P,k} := \widetilde{\mathbf{S}}_{P,k} = T_P \circ \mathbf{S}_{P,k}, \tag{66}$$

where T_P is the deterministic reference-only population preprocessing used in the signature comparison. We assume throughout that T_P is well defined and measurable. For the coordinate-wise population standardization used below, zero reference scales are assigned unit scale, so T_P is an invertible affine map on the signature space and does not itself discard signature information.

We call an alternative Q *population-detectable* by the fixed-reference signature if

$$\mathcal{D}_{P,k}(P, Q) > 0, \tag{67}$$

and *population-invisible* if

$$\mathcal{D}_{P,k}(P, Q) = 0. \tag{68}$$

These notions concern separation by the population representation and do not, by themselves, imply a finite-sample power guarantee.

Proposition 5 (Population-level characterization of fixed-reference detectability)**.** *Let P and Q be Borel probability measures on $\mathbb{R}^D$, let $\Phi_{P,k}$ be measurable, and let κ be a bounded characteristic kernel on the signature space. Then*

$$\mathcal{D}_{P,k}(P, Q) = 0 \quad \Longleftrightarrow \quad (\Phi_{P,k})_\# P = (\Phi_{P,k})_\# Q \tag{69}$$

and equivalently

$$\mathcal{D}_{P,k}(P, Q) = 0 \quad \Longleftrightarrow \quad P(A) = Q(A) \quad \textit{for every } A \in \sigma(\Phi_{P,k}). \tag{70}$$

Consequently,

$$\mathcal{D}_{P,k}(P, Q) > 0 \quad \Longleftrightarrow \quad \exists A \in \sigma(\Phi_{P,k}) \textit{ such that } P(A) \neq Q(A). \tag{71}$$

If, in addition, $Q \ll P$ and

$$r(\boldsymbol{x}) := \frac{dQ}{dP}(\boldsymbol{x}), \tag{72}$$

then the zero-discrepancy condition is equivalent to

$$\mathbb{E}_P\left[r(X) \,|\, \Phi_{P,k}(X)\right] = 1 \qquad P\textit{-a.s.} \tag{73}$$

Hence Q is population-detectable if and only if

$$P(\mathbb{E}_P\left[r(X) \,|\, \Phi_{P,k}(X)\right] \neq 1) > 0. \tag{74}$$

Proof. Because κ is characteristic (Gretton et al., 2012),

$$\mathrm{MMD}^2_\kappa(\mu,\nu) = 0 \quad \Longleftrightarrow \quad \mu = \nu \tag{75}$$

for probability measures μ and ν on the signature space. Applying this property to $(\Phi_{P,k})_\# P$ and $(\Phi_{P,k})_\# Q$ gives Eq. (69).

By the definition of a pushforward measure,

$$(\Phi_{P,k})_\# P = (\Phi_{P,k})_\# Q \tag{76}$$

if and only if

$$P\Big(\Phi^{-1}_{P,k}(B)\Big) = Q\Big(\Phi^{-1}_{P,k}(B)\Big) \tag{77}$$

for every Borel set B in the signature space. The preimages $\Phi^{-1}_{P,k}(B)$ form $\sigma(\Phi_{P,k})$, which proves Eq. (70). Eq. (71) is its negation.

Now suppose $Q \ll P$. For every $A \in \sigma(\Phi_{P,k})$,

$$Q(A) = \int_A r \, dP. \tag{78}$$

Therefore

$$P(A) = Q(A) \quad \text{for every } A \in \sigma(\Phi_{P,k}) \tag{79}$$

if and only if

$$\int_A (r-1)\, dP = 0 \quad \text{for every } A \in \sigma(\Phi_{P,k}). \tag{80}$$

By the defining property of conditional expectation, this is equivalent to

$$\mathbb{E}_P\left[r - 1 \,|\, \sigma(\Phi_{P,k})\right] = 0 \qquad P\text{-a.s.}, \tag{81}$$

or equivalently to Eq. (73). □

Interpretation. Prop. 5 characterizes the success and failure conditions of NLTS at the population level. The method distinguishes P from Q exactly when their difference is visible through the σ-algebra generated by the fixed-reference representation $\Phi_{P,k}$.

When $Q \ll P$, this statement can be expressed directly in terms of the density-ratio perturbation. In particular,

$$r(X) - 1 = \mathbb{E}_P\left[r(X) - 1 \,|\, \Phi_{P,k}(X)\right] + \left\{r(X) - 1 - \mathbb{E}_P\left[r(X) - 1 \,|\, \Phi_{P,k}(X)\right]\right\}. \tag{82}$$

The first term is the component of the distributional perturbation visible through the fixed-reference signature. NLTS has positive population discrepancy exactly when this conditional-mean component is nonzero on a set of positive P-probability. Differences contained entirely in the second term are invisible to the population signature.

Corollary 3 (Identifiability under an injective signature)**.** *Suppose $Q \ll P$ and assume that $\Phi_{P,k}$ is a Borel isomorphism on a Borel set of P-measure one. Then*

$$\mathcal{D}_{P,k}(P,Q) = 0 \quad \Longleftrightarrow \quad P = Q. \tag{83}$$

In particular,

$$P \neq Q \quad \Longrightarrow \quad \mathcal{D}_{P,k}(P,Q) > 0. \tag{84}$$

Proof. Let A be a P-full-measure Borel set on which $\Phi_{P,k}$ is a Borel isomorphism onto its image. Since $Q \ll P$, we also have $Q(A) = 1$. If

$$(\Phi_{P,k})_\# P = (\Phi_{P,k})_\# Q, \tag{85}$$

applying the measurable inverse of $\Phi_{P,k}|_A$ to the common pushforward measure recovers the same probability measure on A. Hence $P = Q$. The converse is immediate. □

The injectivity condition above is only a sufficient condition for universal population-level separation; it is not assumed for the practical NLTS signature. The practical representation deliberately compresses the original sample space through top-k spectral truncation, squared projected-score coordinates, and aggregation of the remaining score energy. It can therefore discard distributional information.

Proposition 6 (An alternative invisible to the practical signature)**.** *Let*

$$P = \mathcal{N}(\mathbf{0}, \boldsymbol{K}^{-1}), \qquad \boldsymbol{K} = \mathrm{diag}(\kappa_1, \ldots, \kappa_D), \qquad \kappa_1 > \cdots > \kappa_D > 0. \tag{86}$$

For any $0 < |\varepsilon| < 1$*, define* $Q \ll P$ *by*

$$\frac{dQ}{dP}(\boldsymbol{x}) = 1 + \varepsilon \tanh(x_1). \tag{87}$$

Then $Q \neq P$*, but the practical fixed-reference signature has the same law under* P *and* Q*:*

$$(\mathbf{S}_{P,k})_{\#}P = (\mathbf{S}_{P,k})_{\#}Q. \tag{88}$$

Consequently, for any deterministic measurable reference-only preprocessing T_P*,*

$$\mathcal{D}_{P,k}(P, Q) = 0. \tag{89}$$

Proof. For the reference in Eq. (86),

$$\boldsymbol{s}_P(\boldsymbol{x}) = -\boldsymbol{K}\boldsymbol{x}, \qquad \boldsymbol{H}_P(\boldsymbol{x}) = \boldsymbol{K}. \tag{90}$$

For the coordinate eigenvectors of $\boldsymbol{K}$, the practical top-k signature is

$$\mathbf{S}_{P,k}(\boldsymbol{x}) = \left[\kappa_1, \ldots, \kappa_k,\ \kappa_1^2 x_1^2, \ldots, \kappa_k^2 x_k^2,\ \sum_{j=k+1}^{D} \kappa_j^2 x_j^2\right]. \tag{91}$$

Thus $\mathbf{S}_{P,k}$ is invariant under the sign change $x_1 \mapsto -x_1$.

The density ratio in Eq. (87) is strictly positive. Moreover, by symmetry of the Gaussian marginal,

$$\mathbb{E}_P[\tanh(X_1)] = 0, \tag{92}$$

so it integrates to one and defines a valid probability distribution. For $\varepsilon \neq 0$, $Q \neq P$.

Because P and $\mathbf{S}_{P,k}$ are invariant under $x_1 \mapsto -x_1$, whereas $\tanh(x_1)$ is odd,

$$\mathbb{E}_P\left[\tanh(X_1) \,|\, \mathbf{S}_{P,k}(X)\right] = 0 \qquad P\text{-a.s.} \tag{93}$$

and hence

$$\mathbb{E}_P\left[\frac{dQ}{dP}(X) \,\middle|\, \mathbf{S}_{P,k}(X)\right] = 1 \qquad P\text{-a.s.} \tag{94}$$

By Prop. 5, Eq. (88) follows. Applying any deterministic measurable T_P preserves equality of the two pushforward laws, which yields Eq. (89). □

This construction does not imply that tangential changes are generally invisible, nor that normal changes are always detectable. Corollaries 1 and 2 show that canonical normal location and scale changes alter the fixed-reference squared first-order Taylor response in the Gaussian normal-coordinate model. Prop. 5 addresses the separate representation-level question: NLTS has positive population discrepancy exactly when the standardized fixed-reference signature has a different law under P and Q. Prop. 6 shows explicitly that some alternatives are lost by the practical representation.

A.4 Properties of the NLTS representation

A.4.1 Nonredundancy of first- and second-order features

The features a^2 and λ are not generally redundant under fixed-reference evaluation. For the one-dimensional Gaussian reference $P = \mathcal{N}(0, \kappa^{-1})$,

$$\lambda_P(\xi) = \kappa, \qquad a_P(\xi)^2 = \kappa^2 \xi^2. \tag{95}$$

If $Q = \mathcal{N}(\mu, \kappa^{-1})$ with $\mu \neq 0$, then

$$(\lambda_P)_{\#}P = (\lambda_P)_{\#}Q, \qquad (a_P^2)_{\#}P \neq (a_P^2)_{\#}Q. \tag{96}$$

Conversely, first-order quantities do not generally determine second-order ones. For example, for the smooth confining potential $V(x) = x^4/4 + x^2/2 + x$, the squared score $[-V'(x)]^2$ equals 1 at both $x = 0$ and $x = -1$, while $V''(0) = 1$ and $V''(-1) = 4$. Thus, first- and second-order features do not determine one another in general.

A.4.2 Sample-dependent eigendirections versus globally shared projections

A globally shared projection applies a single $\boldsymbol{V} \in \mathbb{R}^{D\times k}$ to all samples. NLTS instead solves, at each $\boldsymbol{x}$,

$$\boldsymbol{H}_P(\boldsymbol{x})\boldsymbol{u}_j(\boldsymbol{x}) = \lambda_j(\boldsymbol{x})\boldsymbol{u}_j(\boldsymbol{x}) \tag{97}$$

and defines the sample-dependent subspace

$$E_{P,k}(\boldsymbol{x}) = \mathrm{span}\{\boldsymbol{u}_1(\boldsymbol{x}), \ldots, \boldsymbol{u}_k(\boldsymbol{x})\}. \tag{98}$$

Prop. 1 gives this dominant subspace a local normal-space interpretation in the ideal small-noise regime.

A.4.3 Spectral multiplicity and basis dependence

For an eigenspace E of multiplicity greater than one with orthonormal basis matrix $\boldsymbol{U}_E$, the block energy

$$\|\boldsymbol{U}_E^\top \boldsymbol{s}_P(\boldsymbol{x})\|_2^2 \tag{99}$$

is invariant to the choice of orthonormal basis within E. For a simple eigenvalue, squaring removes the remaining sign ambiguity of the associated eigenvector. For a degenerate eigenspace, individual a_j^2 values are basis dependent, whereas the corresponding blockwise projected-score energy provides a basis-invariant alternative.

More generally, if a repeated eigenvalue block intersects the top-k truncation boundary, the selected projector $\boldsymbol{U}_k\boldsymbol{U}_k^\top$, and hence r_k, can also depend on the chosen basis. A fully basis-invariant construction would retain or aggregate the entire degenerate block. For an isotropic Gaussian normal block, however, the projected-score law is rotationally invariant for any fixed orthonormal basis, so an isotropic scale-shift response does not depend on that basis.

A.4.4 Top-k truncation and residual score energy

If the full Hessian eigensystem were available, $\{(\lambda_j, a_j)\}_{j=1}^D$ would retain more local differential information than the top-k representation. We use $k \ll D$ primarily to avoid a full sample-wise Hessian eigendecomposition.

With $\boldsymbol{U}_k = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]$,

$$r_k(\boldsymbol{x}) = \|(\boldsymbol{I} - \boldsymbol{U}_k\boldsymbol{U}_k^\top)\boldsymbol{s}_P(\boldsymbol{x})\|_2^2 = \|\boldsymbol{s}_P(\boldsymbol{x})\|_2^2 - \sum_{j=1}^{k} a_j(\boldsymbol{x})^2. \tag{100}$$

Thus, r_k retains the first-order score energy in the unselected subspace as a single scalar without computing additional eigenvectors. It is not a purely normal quantity: it may contain both unselected normal components and tangent/background components, and it does not recover the second-order tail of the unselected Hessian eigenvalues.

A.5 Implementation details

A.5.1 Signature comparison

At the population level, let $\boldsymbol{\mu}_P^S$ denote the coordinate-wise mean of the reference signature and let $\boldsymbol{D}_P^S$ be the diagonal matrix of its coordinate-wise standard deviations. Each diagonal scale is taken to be strictly positive; if a reference coordinate has zero scale, we assign unit scale. Thus T_P is an invertible affine transformation of the signature space. We define

$$T_P(\boldsymbol{z}) = (\boldsymbol{D}_P^S)^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_P^S), \qquad \widetilde{\mathbf{S}}_{P,k} = T_P \circ \mathbf{S}_{P,k}. \tag{101}$$

Empirical standardization, treatment of near-zero-variance coordinates, and the benchmark-specific finite-sample MMD estimators are described in App. A.5.2.

Following the kernel two-sample framework of Gretton et al. (2012), using bandwidths $h_1, \ldots, h_L$, we define the multiscale RBF kernel

$$\kappa(\boldsymbol{z}, \boldsymbol{z}') = \frac{1}{L}\sum_{\ell=1}^{L} \exp\left(-\frac{\|\boldsymbol{z} - \boldsymbol{z}'\|_2^2}{2h_\ell^2}\right), \tag{102}$$

Because each RBF kernel is characteristic, their positive finite average κ is also characteristic. The population discrepancy is

$$\mathrm{MMD}^2_\kappa \left((\widetilde{\mathbf{S}}_{P,k})_\# P, (\widetilde{\mathbf{S}}_{P,k})_\# Q \right). \tag{103}$$

The empirical MMD statistic and calibration follow the experiment-specific protocol.

A.5.2 Finite-sample signature discrepancy across benchmarks

For empirical signature samples $\{\boldsymbol{z}_i^P\}_{i=1}^{n_P}$ and $\{\boldsymbol{z}_i^Q\}_{i=1}^{n_Q}$, standardization parameters are estimated from the reference signatures only. Let $\widehat{\boldsymbol{\mu}}_P^S$ be their coordinate-wise mean and let $\widehat{\boldsymbol{D}}_P^S$ be the diagonal matrix of coordinate-wise standard deviations. We apply the same reference-fitted map to both groups:

$$\widehat{T}_P(\boldsymbol{z}) = (\widehat{\boldsymbol{D}}_P^S)^{-1}(\boldsymbol{z} - \widehat{\boldsymbol{\mu}}_P^S). \tag{104}$$

For numerically degenerate coordinates, the corresponding diagonal scale is set to one.

We then pool the standardized reference and comparison signatures and set a base bandwidth h by the median heuristic on pairwise Euclidean distances. For the synthetic and single-cell experiments, the multiscale bandwidth set is

$$\mathcal{H}_{\text{synth/sc}} = \left\{ \frac{h}{2},\, h,\, 2h,\, 4h \right\}, \tag{105}$$

and we average the unbiased empirical RBF–MMD^2 estimates over these bandwidths. For the single-cell experiment, if more than 512 pooled samples are available for bandwidth estimation, at most 512 are selected using a fixed random seed. This follows the same multiscale-MMD convention as the synthetic experiment.

For CIFAR-10-C, we use the same reference-side standardization and pooled median heuristic, but use

$$\mathcal{H}_{\text{CIFAR}} = \left\{ \frac{h}{4},\, \frac{h}{2},\, h,\, 2h,\, 4h \right\}, \tag{106}$$

and average the biased empirical RBF–MMD^2 estimates over these bandwidths.

Thus, all three benchmarks share the same reference-defined signature, reference-side standardization, and median-heuristic multiscale RBF kernel. The bandwidth grid and finite-sample MMD estimator differ slightly between the synthetic/single-cell and CIFAR-10-C pipelines; these are implementation choices of the corresponding experimental pipelines rather than theoretical components of the method.

A.5.3 Practical score and Hessian computation

For an exact reference score,

$$\boldsymbol{H}_P(\boldsymbol{x}) = -\nabla_{\boldsymbol{x}} \boldsymbol{s}_P(\boldsymbol{x}) = \nabla^2_{\boldsymbol{x}}[-\log p_P(\boldsymbol{x})] \tag{107}$$

is symmetric. Learned score fields, including denoising-score-matching estimators (Vincent, 2011), need not have an exactly symmetric Jacobian under finite approximation error. When an explicitly symmetric approximation to the theoretical Hessian is desired, a natural choice is

$$\widehat{\boldsymbol{H}}_P^{\text{sym}}(\boldsymbol{x}) = -\frac{1}{2}\left[\nabla_{\boldsymbol{x}} \widehat{\boldsymbol{s}}_P(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} \widehat{\boldsymbol{s}}_P(\boldsymbol{x})^\top \right]. \tag{108}$$

If $\widehat{\boldsymbol{s}}_P = -\nabla_{\boldsymbol{x}} E_\theta$ is defined by a scalar energy model, Jacobian symmetry holds structurally.

The required k directions can be computed matrix-free with an iterative eigensolver, without forming the full matrix. For the symmetrized operator in Eq. (108), each matrix–vector product can be obtained from a Jacobian–vector product and a vector–Jacobian product; when the score is a gradient field, Hessian–vector products suffice. Thus, k is a truncation parameter controlling computation and signature dimension rather than an estimate of the normal codimension. The precise learned operator and its numerical stabilization are experiment-specific and are stated in the corresponding protocol.

In practical learned or noisy settings, estimated eigenvalue signs can be unstable. We therefore use $|\widehat{\lambda}_j|$ for the practical curvature features to suppress spurious sign flips. This numerical stabilization is distinct from the signed Taylor-coefficient interpretation of λ_j in the theoretical development and from the positive normal-curvature asymptotics of Prop. 1.

B BASELINE METHODS

We summarize the main comparison methods and implementation settings used in the experiments. Sample-based discrepancies use samples from both the reference and comparison distributions. Stein-based methods are evaluated as goodness-of-fit statistics using the reference score $\boldsymbol{s}_P$. Within each experiment, all Stein-based baselines and NLTS use the same reference score.

For adaptive Stein methods, samples used for direction or test-location selection are separated from those used to evaluate the final statistic. Test thresholds and Type-I error evaluation follow the common protocol in App. C.

MMD and Energy Distance. For a positive-definite kernel k, the population MMD (Gretton et al., 2012) is

$$\mathrm{MMD}_k^2(P,Q) = \mathbb{E}k(\boldsymbol{x},\boldsymbol{x}') + \mathbb{E}k(\boldsymbol{y},\boldsymbol{y}') - 2\mathbb{E}k(\boldsymbol{x},\boldsymbol{y}), \tag{109}$$

where $\boldsymbol{x},\boldsymbol{x}' \sim P$ and $\boldsymbol{y},\boldsymbol{y}' \sim Q$ are independent. We use the RBF kernel

$$k_h(\boldsymbol{x},\boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{2h^2}\right), \tag{110}$$

with bandwidth h set by the median heuristic on pairwise Euclidean distances unless stated otherwise.

Energy Distance (Székely & Rizzo, 2013) is

$$\mathcal{E}(P,Q) = 2\mathbb{E}\|\boldsymbol{x}-\boldsymbol{y}\|_2 - \mathbb{E}\|\boldsymbol{x}-\boldsymbol{x}'\|_2 - \mathbb{E}\|\boldsymbol{y}-\boldsymbol{y}'\|_2. \tag{111}$$

For raw RBF–MMD in the synthetic and single-cell experiments, we use the unbiased empirical estimator based on off-diagonal terms. For CIFAR-10-C, both groups are standardized using the coordinate-wise mean and standard deviation estimated from the reference side, after which we use the biased empirical RBF–MMD estimator.

Sliced discrepancies. For Sliced Wasserstein-2 (Sliced W_2), we follow the sliced Wasserstein construction (Bonneel et al., 2015). For Sliced MMD, we analogously average one-dimensional MMDs over random projections. In both cases, samples are projected onto random unit directions $\boldsymbol{\theta}_\ell$, obtained by normalizing independent standard Gaussian vectors, and the resulting one-dimensional discrepancies are averaged. For L directions,

$$\mathrm{SMMD}(P,Q) = \frac{1}{L}\sum_{\ell=1}^{L} \mathrm{MMD}^2\left(\boldsymbol{\theta}_\ell^\top \boldsymbol{x}, \boldsymbol{\theta}_\ell^\top \boldsymbol{y}\right). \tag{112}$$

Sliced W_2 is defined analogously by averaging W_2^2 over the one-dimensional projected distributions.

The final synthetic completion run uses $L = 64$ random directions, while CIFAR-10-C uses $L = 128$.

KSD and random Sliced KSD. For KSD (Liu et al., 2016; Chwialkowski et al., 2016), define the Stein operator for the reference score $\boldsymbol{s}_P$ by

$$\mathcal{T}_P \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{s}_P(\boldsymbol{x})^\top \boldsymbol{f}(\boldsymbol{x}) + \nabla\cdot \boldsymbol{f}(\boldsymbol{x}). \tag{113}$$

KSD combines this operator with a positive-definite kernel; we use an RBF kernel with median-heuristic bandwidth.

KSD is computed from comparison samples $\boldsymbol{y}_i \sim Q$ and the corresponding reference scores $\boldsymbol{s}_P(\boldsymbol{y}_i)$. The synthetic and single-cell experiments use the U-statistic estimator based on off-diagonal terms. CIFAR-10-C uses the biased estimator and at most 256 samples for KSD evaluation.

For random Sliced KSD (Gong et al., 2021a), we construct

$$y_i^{(\ell)} = \boldsymbol{\theta}_\ell^\top \boldsymbol{y}_i, \qquad s_i^{(\ell)} = \boldsymbol{\theta}_\ell^\top \boldsymbol{s}_P(\boldsymbol{y}_i), \tag{114}$$

for random unit directions $\boldsymbol{\theta}_\ell$, and average the resulting one-dimensional KSD values. We use 64 directions for the synthetic experiment, 128 for CIFAR-10-C, and 16 for the single-cell experiment.

Finite-set Stein discrepancy. For FSSD-random and FSSD-opt (Jitkrittum et al., 2017), we use the public reference implementation listed in Table 2. Both methods use $J = 5$ test locations.

FSSD-random fits a Gaussian distribution to held-out selector samples and draws the J test locations from that fit. The initial Gaussian-kernel width is set to the squared median pairwise distance among the selector samples.

FSSD-opt starts from the same initialization and jointly optimizes the test locations and Gaussian-kernel width using the reference implementation's optimization routine, with reg $= 10^{-2}$ and $\text{tol}_{\text{fun}} = 10^{-5}$. The final synthetic completion run uses at most 100 optimization iterations.

For the CIFAR-10-C full sweep, the main profile uses a 20-iteration optimization budget to control computational cost. Rather than running the reference implementation's internal null simulation within each repeat, we use the common empirical calibration in App. C.

Max-sliced KSD and Active Slices. For maxSKSD-g and maxSKSD-rg (Gong et al., 2021a), and Active Slices (Gong et al., 2021b), we use the public reference implementation listed in Table 2.

maxSKSD uses three projection directions. In maxSKSD-g, the projection parameter r is fixed while g is optimized; maxSKSD-rg jointly optimizes both r and g.

Active Slices also uses three active directions. Active Slices (init) uses the SVD-based initialization returned by the reference implementation without further optimization. Active Slices starts from the same initialization, keeps r fixed, and updates g by gradient-based optimization.

For the final synthetic completion run, we use 100 optimization iterations, learning rate 10^{-3}, and mini-batch size 50. For the CIFAR-10-C full sweep, we use 20 iterations, learning rate 10^{-2}, and mini-batch size 50. As with FSSD, the main full sweep uses the common empirical calibration rather than the reference implementation's per-repeat bootstrap calibration.

RFSD and Stein random-feature baselines. For RFSD and the Stein RFF baselines (Huggins & Mackey, 2018), we use the public reference implementation listed in Table 2. All variants use 100 random features.

For RFSD (IMQ), we use the reference implementation's L1-IMQ configuration with $\gamma = 0.25$, target_df $= 0.5$, and $c = 4m_2$, where m_2 is the median pairwise L_2 distance among evaluation samples.

For RFSD (SechExp), we use the L2-SechExp configuration with $\gamma = 0.25$. Let m_1 be the median pairwise L_1 distance and define

$$\alpha_{\text{aux}} = \frac{\gamma}{3}, \qquad \xi = \frac{4\alpha_{\text{aux}}}{2 + \alpha_{\text{aux}}}. \tag{115}$$

The initial scale is

$$\text{scale} = \frac{1}{m_1\sqrt{\pi/2}}, \tag{116}$$

and for $d > 1$ we further set

$$\text{scale} \leftarrow \frac{\text{scale}}{2\xi}. \tag{117}$$

For CIFAR-10-C, we additionally evaluate Stein RFF (Gauss) and Stein RFF (Cauchy). Stein RFF (Gauss) uses Gaussian random Fourier features with $\sigma^2 = m_2^2$, while Stein RFF (Cauchy) uses Cauchy random Fourier features with scale m_2. These are random-feature Stein discrepancies incorporating the reference score, not standard random-feature MMDs.

Moment baselines. For CIFAR-10-C, we additionally evaluate simple first- and second-moment baselines: Mean L2, Diag Cov L2, Diag Mahalanobis, and Diag FID.

Let $\boldsymbol{\mu}_X, \boldsymbol{\mu}_Y$ and $\boldsymbol{v}_X, \boldsymbol{v}_Y$ denote the sample-mean and coordinate-wise variance vectors. For numerical stabilization, we set $\boldsymbol{v}_X \leftarrow \boldsymbol{v}_X + 10^{-6}\mathbf{1}$ and $\boldsymbol{v}_Y \leftarrow \boldsymbol{v}_Y + 10^{-6}\mathbf{1}$.

Mean L2 is

$$\|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|_2, \tag{118}$$

Table 2: Public reference implementations used for the Stein baselines. The revision column reports the Git commit used in our experiments.

Method	Revision
FSSD[1]	`039a95ed`
maxSKSD / Active Slices[2]	`bac05c07`
RFSD / Stein RFF[3]	`d77218b4`

Diag Cov L2 is

$$\|\boldsymbol{v}_X - \boldsymbol{v}_Y\|_2, \tag{119}$$

Diag Mahalanobis uses the reference variance:

$$\sum_j \frac{(\mu_{X,j} - \mu_{Y,j})^2}{v_{X,j}}, \tag{120}$$

and Diag FID is the diagonal-Gaussian analogue of the Fréchet term used in FID (Heusel et al., 2017):

$$\|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|_2^2 + \sum_j \left(v_{X,j} + v_{Y,j} - 2\sqrt{v_{X,j} v_{Y,j}}\right). \tag{121}$$

Public implementations and locked revisions. For reproducibility, we fix not only the repository but also the Git revision for each external Stein-baseline implementation. Table 2 summarizes the versions used in our experiments.

C Common experimental evaluation protocol

Let T denote the scalar statistic returned by a method, with larger values indicating stronger evidence against the no-shift hypothesis. A *repeat* denotes one Monte Carlo realization in which the sample sets are regenerated or resampled according to the experiment-specific sampling procedure and the statistic or geometry diagnostic is recomputed from scratch.

C.1 Resampling and repeat counts

The number of repeats depends on the benchmark. For the synthetic experiment, each alternative setting uses 100 repeats, with an independent set of 300 null-calibration repeats and 300 null-validation repeats. For CIFAR-10-C, each corruption, severity, and sample-size setting uses 50 main resampling repeats, together with a separate 100-repeat clean-only null evaluation. For the single-cell experiment, null calibration and null validation each use 1000 repeats. Alternative-repeat settings and the detection protocol are given in App. G.3, and the resampling procedure for the local-PCA diagnostics is given in App. G.2.

C.2 Empirical calibration and rejection rates

Let $T_1^{(0)}, \ldots, T_{N_{\text{cal}}}^{(0)}$ denote the null statistics used for calibration, and set the nominal significance level to $\alpha = 0.05$. The rejection threshold is

$$\widehat{c}_{1-\alpha} = \widehat{Q}_{1-\alpha}\left(T_1^{(0)}, \ldots, T_{N_{\text{cal}}}^{(0)}\right), \tag{122}$$

where $\widehat{Q}_{1-\alpha}$ is the empirical $(1-\alpha)$ quantile. When independent null-validation repeats are available, the Type-I error is estimated as

$$\widehat{\text{TypeI}} = \frac{1}{N_{\text{val}}} \sum_{i=1}^{N_{\text{val}}} \mathbf{1}\left[T_i^{(0,\text{val})} > \widehat{c}_{1-\alpha}\right]. \tag{123}$$

[1] `github.com/wittawatj/kernel-gof`

[2] `github.com/WenboGong/Sliced_Kernelized_Stein_Discrepancy`

[3] `bitbucket.org/jhhuggins/random-feature-stein-discrepancies`

The alternative rejection rate is

$$\widehat{R}_{H_1} = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathbf{1}\left[T_i^{(1)} > \widehat{c}_{1-\alpha}\right]. \tag{124}$$

We refer to this quantity as empirical power when size control is verified. When method-dependent clean-null inflation remains, as in some CIFAR-10-C comparisons, we instead report it as an alternative rejection or detection rate. Accordingly, we use *empirical power* for the size-validated synthetic experiments, *alternative rejection rate* for CIFAR-10-C, where some methods exhibit clean-null inflation, and *detection power* for the single-cell analysis, where condition-specific rejection rates are aggregated across score-model seeds, sample sizes, and doses. The benchmark-specific construction of the calibration statistics is given in the corresponding experimental protocol.

C.3 Cohen's d reporting

For the synthetic and CIFAR-10-C experiments, we use Cohen's d as a standardized descriptive measure of the separation between the null and alternative statistic distributions. Let $\overline{T}_0, s_0$ and $\overline{T}_1, s_1$ denote the sample means and standard deviations of the null and alternative statistics, respectively. We define

$$d = \frac{\overline{T}_1 - \overline{T}_0}{\sqrt{(s_0^2 + s_1^2)/2}}. \tag{125}$$

Cohen's d is not test power; it is interpreted as a standardized descriptive measure of null–alternative separation for a given statistic. The benchmark-specific construction of the H_0 and H_1 statistic samples is given in the corresponding experimental section.

D Controlled synthetic benchmark details

D.1 Synthetic data generation

Fig. 1 schematically illustrates low-dimensional analogues of the data-generating mechanisms used in the synthetic experiments. For the main synthetic experiments, we use distributions for which both the score and the negative-log-density Hessian can be evaluated analytically. Unless otherwise stated, the affine and curved Gaussian-tube experiments use ambient dimension $D = 256$ and manifold (tangent) dimension $d = 8$. The primary Dimension-mixture condition instead uses $D = 1024$ with component tangent dimensions $(128, 256, 512)$.

D.1.1 Reference distribution families

The synthetic scenarios are built from the following reference families.

- **Affine Gaussian tube.** Let D be the ambient dimension and d the tangent dimension. We independently sample

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d), \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{D-d}), \qquad \sigma_n = 0.05, \tag{126}$$

  and set $\mathbf{Y} = [\mathbf{Z}^\top, \boldsymbol{\varepsilon}^\top]^\top$. For a geometry seed g, a fixed orthogonal transformation $\mathbf{R}_g$ defines

$$\mathbf{X} = \mathbf{R}_g \mathbf{Y} + \boldsymbol{\mu}. \tag{127}$$

- **Curved Gaussian tube.** Starting from the same latent variables, we apply a nonlinear shear to the first canonical normal coordinate,

$$Y_{d+1} = \varepsilon_1 + A \sin(\omega Z_1), \qquad A = 0.32, \qquad \omega = 1.15. \tag{128}$$

  Writing $c(Z_1) = A\omega \cos(\omega Z_1)$, the corresponding first tangent and normal directions in canonical coordinates are

$$\mathbf{t}_1(Z_1) = \frac{\mathbf{e}_1 + c(Z_1)\mathbf{e}_{d+1}}{\sqrt{1 + c(Z_1)^2}}, \qquad \mathbf{n}_1(Z_1) = \frac{-c(Z_1)\mathbf{e}_1 + \mathbf{e}_{d+1}}{\sqrt{1 + c(Z_1)^2}}, \tag{129}$$

  so that $\mathbf{t}_1(Z_1)^\top \mathbf{n}_1(Z_1) = 0$. Thus, the local tangent and normal directions vary with position along the curved reference geometry.

- **Heterogeneous local-geometry mixture.** For the Dimension-mixture family, we define three fixed components P_1, P_2, P_3 with different tangent dimensions and form the reference distribution

$$P = \sum_{j=1}^{3} w_{P,j} P_j. \tag{130}$$

The component geometries remain fixed; the corresponding alternative changes only their mixture weights, as specified below.

D.1.2 Perturbation scenarios

The primary target-shift comparison uses the following five scenarios.

- **Affine normal thickness.** Using the affine reference family, we leave the tangent latent distribution unchanged and modify only the normal scale:

$$\sigma_{n,Q} = (1+\delta)\sigma_{n,P}. \tag{131}$$

- **Curved normal thickness.** We apply the same normal-scale change in Eq. 131 to the curved reference family. The perturbation therefore changes normal thickness while the local normal direction itself varies with position.
- **Local artifact.** Let z_1 denote the first standardized tangent coordinate and let $\mathbf{v}_1(\mathbf{X})$ and $\mathbf{v}_2(\mathbf{X})$ denote the first two local normal directions of the reference distribution. The alternative adds a position-dependent normal displacement

$$\mathbf{X}_Q = \mathbf{X}_P + \delta \left[\sin(\nu z_1)\mathbf{v}_1(\mathbf{X}_P) + \frac{1}{2}\,\mathrm{sign}\{\sin(2\nu z_1)\}\,\mathbf{v}_2(\mathbf{X}_P) \right], \qquad \nu = 6. \tag{132}$$

For the affine family the normal directions are constant, whereas in the curved variant they are evaluated locally at each sample.

- **Single-dimension expansion.** For one local normal direction $\mathbf{v}_1(\mathbf{X}_P)$, we sample $\eta \sim \mathcal{N}(0,1)$ and set

$$\mathbf{X}_Q = \mathbf{X}_P + \delta\eta\,\mathbf{v}_1(\mathbf{X}_P). \tag{133}$$

This increases variability along a single normal direction without changing the tangent latent distribution.

- **Dimension-mixture shift.** We keep the three component distributions fixed and change only their mixture composition:

$$P = \sum_{j=1}^{3} w_{P,j} P_j, \qquad Q = \sum_{j=1}^{3} w_{Q,j} P_j. \tag{134}$$

For ambient dimension D, the component tangent dimensions are

$$(d_1, d_2, d_3) = \left(\frac{D}{8}, \frac{D}{4}, \frac{D}{2} \right), \tag{135}$$

giving $(32, 64, 128)$ for $D = 256$ and $(128, 256, 512)$ for $D = 1024$. The reference and alternative weights are

$$\mathbf{w}_P = (0.4, 0.2, 0.4), \qquad \mathbf{w}_Q = \mathbf{w}_P + \alpha(-0.2, 0.3, -0.1), \qquad \alpha \in \{0.5, 1.0, 1.5\}. \tag{136}$$

The weight perturbation preserves both total mass and the mixture-weighted average tangent dimension:

$$\sum_{j=1}^{3}(w_{Q,j} - w_{P,j}) = 0, \qquad \sum_{j=1}^{3}(w_{Q,j} - w_{P,j})d_j = 0. \tag{137}$$

The moment-matched construction additionally enforces equality of the population first and second moments,

$$\mathbb{E}_P[\mathbf{X}] = \mathbb{E}_Q[\mathbf{X}], \qquad \mathrm{Cov}_P(\mathbf{X}) = \mathrm{Cov}_Q(\mathbf{X}), \tag{138}$$

so that the comparison primarily reflects a change in the population composition of heterogeneous local-dimensional regimes rather than a simple global mean or covariance shift.

D.2 Synthetic evaluation protocol

For the five target shifts in the primary synthetic comparison, we fix representative evaluation conditions before aggregation. The numerical scale of the perturbation parameter is scenario-specific and should not be interpreted as representing equal distributional difficulty across different scenarios.

For each alternative setting, Cohen's d is computed from independently repeated H_0 and H_1 statistics using Eq. 125. The hypothesis-test threshold is estimated from a separate set of null-calibration samples, and Type-I error is evaluated on independent null-validation samples that are not used for calibration. Thus, the statistic samples used for effect-size estimation, threshold calibration, and Type-I evaluation serve distinct roles.

Table 3 summarizes the geometric role of each target shift and its role in the evaluation.

Table 3: Roles of the primary target-shift scenarios.

Scenario	Geometry role	Evaluation role
Affine normal thickness	Normal-scale change	Positive control
Curved normal thickness	Position-dependent normal-scale change	Positive control
Local artifact	Localized normal displacement	Positive control
Single-dimension expansion	Variability increase along one normal direction	Positive control
Dimension-mixture shift	Mixture-composition shift across local-dimensional regimes	Hard positive condition

D.3 Additional synthetic results

Tables 4–6 provide complementary numerical summaries of the synthetic experiments. Table 4 reports Cohen's d and Type-I error for the affine and curved normal-thickness shifts. Table 5 reports empirical power as a function of sample size for Affine normal thickness. Table 6 summarizes Local artifact, Single-dimension expansion, and Dimension-mixture shift.

Table 4: Affine and curved normal-thickness shifts. Cohen's d and Type-I error are medians across the representative settings that passed validation and across geometry seeds.

	Affine		Curved	
Method	Cohen's d	Type-I error	Cohen's d	Type-I error
NLTS	32.574	0.040	34.976	0.053
KSD	0.175	0.047	0.227	0.027
Sliced KSD	9.659	0.057	8.581	0.047
MMD	0.301	0.057	0.180	0.033
Sliced MMD	0.564	0.060	0.550	0.027
Energy	0.428	0.063	0.307	0.033
Sliced W_2	0.568	0.060	0.444	0.027

Table 5: Affine normal-thickness shift ($D = 256$, strength 0.5): empirical power across sample sizes.

Method	16	32	64	128	256	512
NLTS	1.00	1.00	1.00	1.00	1.00	1.00
Sliced KSD	0.70	0.96	1.00	1.00	1.00	1.00
Sliced MMD	0.01	0.07	0.05	0.09	0.19	0.44
Sliced W_2	0.05	0.09	0.07	0.08	0.14	0.29
KSD	0.24	0.29	0.24	0.27	0.28	0.42
MMD	0.04	0.09	0.03	0.05	0.06	0.13
Energy	0.03	0.09	0.03	0.06	0.09	0.18

Fig. 7 complements these aggregate summaries with representative curves. Panels (a,b) show empirical power and Cohen's d as functions of sample size for Affine normal thickness. Panels (c,d) show empirical power as a function of the mixture-weight parameter α for Dimension-mixture shift at $D = 1024$ and $D = 256$, respectively.

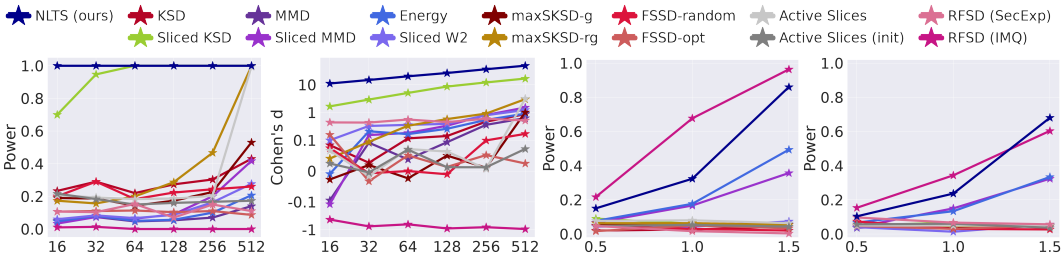
D.4 Tangential specificity controls

Scenario definitions. To assess tangential specificity separately from the five target shifts, we evaluate six tangential perturbation settings: a mean-shift control and five bump-based variants, including affine and curved constructions. Let

$$\mathbf{Z}_P \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d) \tag{139}$$

Table 6: Local artifact, Single-dimension expansion, and Dimension-mixture shift. Each pair of columns reports median Cohen's d and median Type-I error over the available validated settings.

	Local artifact		Single-dimension expansion		Dimension-mixture shift	
Method	Cohen's d	Type-I	Cohen's d	Type-I	Cohen's d	Type-I
NLTS	5.838	0.040	5.617	0.040	0.895	0.073
KSD	0.281	0.047	0.666	0.047	0.028	0.050
Sliced KSD	0.311	0.057	0.615	0.057	0.027	0.063
MMD	-0.009	0.057	-0.069	0.057	0.088	0.043
Sliced MMD	-0.044	0.060	-0.069	0.060	0.519	0.037
Energy	-0.002	0.063	-0.065	0.063	0.601	0.050
Sliced W_2	0.125	0.060	-0.071	0.060	-0.066	0.040



(a) Affine normal thickness: power (b) Affine normal thickness: Cohen's d (c) Dimension mixture ($D = 1024$): power (d) Dimension mixture ($D = 256$): power

Figure 7: Additional synthetic results. (a,b) Sample-size dependence of empirical power and Cohen's d for Affine normal thickness. (c,d) Empirical power across mixture-weight strengths for Dimension-mixture shift at $D = 1024$ and $D = 256$, respectively.

denote the standardized tangent coordinates of the reference distribution. In all six settings, the normal noise variables are generated exactly as under the reference distribution and are left unchanged. Except for Tangent mean shift, the affected tangent coordinate or tangent block is centered and rescaled so that its first two moments match those of the reference distribution.

Let $B \sim \text{Bernoulli}(w)$ indicate whether a sample receives the localized perturbation, let S be uniform on $\{-1, +1\}$, and let all scalar or vector Gaussian variables below be mutually independent standard normal variables in the relevant dimension.

- **Tangent mean shift.** Only the first standardized tangent coordinate is translated:

$$Z_{1,Q} = Z_{1,P} + \delta, \qquad Z_{j,Q} = Z_{j,P} \quad (j \geq 2). \tag{140}$$

- **Tangent pair bump.** With probability w, the first tangent coordinate is replaced by a draw from a symmetric two-component Gaussian mixture centered at $\pm c$:

$$\widetilde{Z}_1 = \frac{(1-B)G_0 + B\,(Sc + \tau G_1)}{\sqrt{(1-w) + w(c^2 + \tau^2)}}. \tag{141}$$

  The symmetry preserves zero mean, and the denominator restores unit variance.

- **Curved tangent pair bump.** The perturbation in Eq. 141 is applied to a tangent latent coordinate before mapping the sample through the curved embedding. Thus, the perturbation is tangential in local manifold coordinates even though its ambient-space direction varies with position.

- **Tangent shell perturbation.** For a b-dimensional tangent block, let $\mathbf{U}$ be uniformly distributed on the unit sphere in $\mathbb{R}^b$. With probability w, the Gaussian tangent block is replaced by a noisy spherical shell of radius r:

$$\widetilde{\mathbf{Z}}_{1:b} = \frac{(1-B)\mathbf{G}_0 + B\,(r\mathbf{U} + \tau \mathbf{G}_1)}{\sqrt{(1-w) + w(r^2/b + \tau^2)}}. \tag{142}$$

  Spherical symmetry preserves zero mean and, after rescaling, identity covariance within the affected tangent block.

- **Curved tangent shell perturbation.** The tangent-block perturbation in Eq. 142 is applied before the curved embedding. The perturbation is therefore defined in local tangent coordinates while the corresponding ambient-space directions vary along the manifold.
- **Tangent multi-bump perturbation.** For a b-dimensional tangent block, let J be uniform on $\{1, \dots, b\}$ and let $\mathbf{e}_J$ denote the corresponding coordinate basis vector. With probability w, probability mass is redistributed toward one of the symmetric locations $\pm c\mathbf{e}_J$:

$$\widetilde{\mathbf{Z}}_{1:b} = \frac{(1-B)\mathbf{G}_0 + B\,(Sc\mathbf{e}_J + \tau\mathbf{G}_1)}{\sqrt{(1-w) + w(\tau^2 + c^2/b)}}. \tag{143}$$

  Averaging over J and S preserves zero mean and, after rescaling, identity covariance within the affected tangent block.

The five settings other than Tangent mean shift modify higher-order structure in the tangent distribution while preserving the first two moments of the affected standardized tangent coordinates. Together, these settings extend the specificity analysis beyond a simple tangential mean shift. All tangential specificity controls use ambient dimension $D = 256$ and manifold (tangent) dimension $d = 8$, with $n = 256$ samples per distribution and three geometry seeds. For the bump-based controls, we set the bump weight to 0.1, the center to 2.0, the width to 0.25, and the bump dimension to 2.

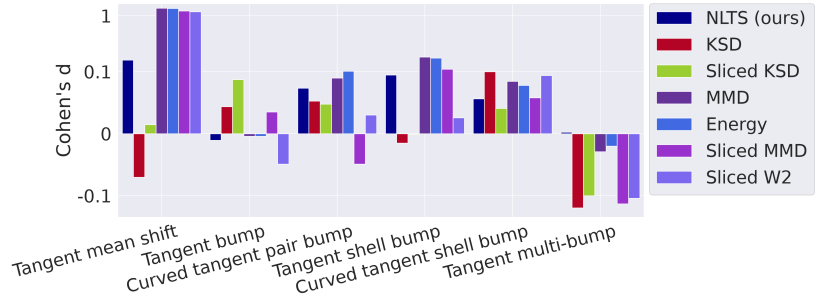


Figure 8: Cohen's d view of the tangential specificity controls. Cohen's d remains close to zero for NLTS across the six tangential perturbations, complementing the empirical rejection-rate results reported in Fig. 3b of the main text.

Results. The main text reports empirical rejection rates across the six controls (Fig. 3b), while Fig. 8 provides the complementary Cohen's d view. The baseline methods show heterogeneous behavior across these perturbations. For Tangent mean shift, Energy, MMD, Sliced MMD, and Sliced W_2 attain empirical rejection rates of 0.48, 0.47, 0.38, and 0.41, respectively, whereas several baselines remain close to the nominal level for the moment-matched pair, shell, and multi-bump perturbations. The same overall pattern is reflected in the Cohen's d view: baseline responses vary across tangential perturbations, rather than showing a consistent tendency. In contrast, NLTS maintains low response across all six controls on both metrics, indicating that its low sensitivity to tangential perturbations is not specific to a single perturbation scenario.

E Design ablations of NLTS

This section examines two aspects of the NLTS design. We first study the sensitivity of the main configuration to the number of retained Hessian eigendirections, k, and to the discrepancy metric used in signature space. We then compare the isolated projected-score and Hessian-eigenvalue components of the fixed-reference representation.

E.1 Parameter ablations

In each comparison, all design choices other than the ablated component are held fixed. Each table entry reports ''Cohen's d / empirical power.'' We use $n = 128$ for Affine and Curved normal

thickness and $n = 256$ for the remaining geometry-positive scenarios. When a scenario contains multiple evaluated settings, we report the median across those settings.

Ablation of retained eigendirections k. We fix the joint signature in Eq. 16 and multiscale RBF–MMD, and compare $k = 2$ with $k = 4$. The results are shown in Table 7.

Table 7: Ablation of the number of retained Hessian eigendirections k. Each entry reports Cohen's d / empirical power. The joint NLTS signature and multiscale RBF–MMD are held fixed.

Scenario	n	$k = 2$	$k = 4$
Affine normal thickness	128	**32.54 / 1.00**	27.67 / 1.00
Curved normal thickness	128	**35.70 / 1.00**	30.16 / 1.00
Dimension-mixture shift	256	**1.80 / 0.610**	0.92 / 0.215
Local artifact	256	**7.87 / 1.00**	6.68 / 1.00
Curved local artifact	256	9.38 / 1.00	**9.70 / 1.00**
Single-dimension expansion	256	7.77 / 1.00	**8.00 / 1.00**

Although $k = 4$ yields a slightly larger Cohen's d in some settings, both $k = 2$ and $k = 4$ already achieve power 1.00 for Curved local artifact and Single-dimension expansion. By contrast, in the more difficult Dimension-mixture shift, $k = 2$ achieves power 0.610, substantially above the 0.215 obtained with $k = 4$. Thus, increasing the number of retained eigendirections from two to four does not provide a consistent gain in detection performance. Moreover, computing additional Hessian eigendirections requires additional iterations of the eigensolver. We therefore use the fixed choice $k = 2$ throughout the main experiments as a practical trade-off between detection performance and computation.

Ablation of discrepancy metric. Next, we fix $k = 2$ and the joint signature in Eq. 16, and compare Energy Distance, median-bandwidth RBF–MMD, and multiscale RBF–MMD. The results are shown in Table 8.

Table 8: Ablation of the discrepancy metric in signature space. Each entry reports Cohen's d / empirical power. The joint NLTS signature with $k = 2$ is held fixed.

Scenario	n	Energy	Median RBF–MMD	Multiscale RBF–MMD
Affine normal thickness	128	19.26 / 1.00	25.97 / 1.00	**32.54 / 1.00**
Curved normal thickness	128	17.82 / 1.00	27.77 / 1.00	**35.70 / 1.00**
Dimension-mixture shift	256	1.33 / 0.375	1.10 / 0.290	**1.80 / 0.610**
Local artifact	256	7.95 / 1.00	**8.04 / 1.00**	7.87 / 1.00
Curved local artifact	256	9.60 / 1.00	**9.76 / 1.00**	9.38 / 1.00
Single-dimension expansion	256	**8.12 / 1.00**	7.82 / 1.00	7.77 / 1.00

For the easier shifts, all three metrics reach power 1.00, and the largest Cohen's d varies across metrics. Thus, multiscale RBF–MMD is not pointwise optimal in every scenario. For the more difficult Dimension-mixture shift, however, multiscale RBF–MMD achieves power 0.610, compared with 0.375 for Energy Distance and 0.290 for median-bandwidth RBF–MMD. In this benchmark, multiscale RBF–MMD gives the highest power among the three choices on the Dimension-mixture setting. We therefore use multiscale RBF–MMD as the fixed discrepancy metric in the main experiments.

These ablations are used to choose a single default configuration, not to tune the method separately for each test condition. After selecting $k = 2$ and multiscale RBF–MMD here, we keep the same configuration across scenarios, corruption types, severities, and sample sizes in the main experiments. Accordingly, the reported main results reflect fixed-configuration performance rather than per-setting oracle selection.

Tangential controls under the selected design. Finally, we check whether the low response to the tangential controls is preserved across these design choices. Table 9 reports the responses to Tangent bump and Curved tangent bump when varying k and the discrepancy metric.

Across all design choices, Cohen's d remains close to zero and the empirical rejection rates remain near the nominal Type-I level on the tangential controls. Thus, the low response on these controls is

Table 9: Responses on tangential controls under the design ablations. The left two columns vary k with multiscale RBF–MMD fixed, whereas the right three columns vary the discrepancy metric with $k = 2$ fixed. Each entry reports Cohen's d / empirical power.

Scenario	n	$k = 2$	$k = 4$	Energy	Median RBF–MMD	Multiscale RBF–MMD
Tangent bump	128	-0.013 / 0.06	0.002 / 0.06	0.000 / 0.07	0.030 / 0.09	-0.013 / 0.06
Curved tangent bump	128	0.047 / 0.04	0.002 / 0.03	0.080 / 0.05	0.070 / 0.05	0.047 / 0.04

consistently observed across the tested values of k and across the three signature-space discrepancies. In particular, it is preserved by the selected configuration of $k = 2$ with multiscale RBF–MMD.

E.2 First- vs. second-order component analysis

As indicated by Eq. 12, the NLTS representation contains a first-order coefficient $a_j = u_j^\top s_P$ and a second-order curvature coefficient λ_j. Prop. 2 further shows that a_j captures first-order normal-coordinate information along a dominant Hessian direction. To compare these components in isolation, we fix $k = 2$ and multiscale RBF–MMD and compare the projected-score component

$$S_a(x) = \left[a_1(x)^2, a_2(x)^2\right] \tag{144}$$

with the curvature-magnitude component used in the practical implementation,

$$S_\lambda(x) = \left[|\lambda_1(x)|, |\lambda_2(x)|\right]. \tag{145}$$

This is not a generic ablation over arbitrary subsets of the signature. Rather, it is a mechanistic comparison of the first- and second-order quantities that arise from the local expansion underlying the method. The results are shown in Table 10.

Table 10: First- vs. second-order component analysis. Each entry reports Cohen's d / empirical power. S_a uses projected-score energies along Hessian eigendirections, whereas S_λ uses only the corresponding curvature magnitudes.

Scenario	n	First-order S_a	Second-order S_λ
Affine normal thickness	128	**3.90 / 1.00**	-0.17 / 0.04
Curved normal thickness	128	**4.15 / 1.00**	0.12 / 0.05
Dimension-mixture shift	256	0.04 / 0.065	**1.02 / 0.48**
Local artifact	256	0.22 / 0.11	**2.36 / 0.98**
Curved local artifact	256	**5.03 / 1.00**	3.29 / 0.95
Single-dimension expansion	256	0.46 / 0.13	**1.28 / 0.78**

For Affine and Curved normal thickness, S_a reaches power 1.00, whereas the response of S_λ alone remains close to the null level. This difference is naturally explained by the local quadratic model. For a dominant Hessian eigenpair,

$$a_j(x) = u_j(x)^\top s_P(x) \approx -\frac{\alpha_j(m)}{\sigma^2}\xi_j, \qquad \lambda_j(x) \approx \frac{\alpha_j(m)}{\sigma^2}. \tag{146}$$

Hence, when the normal-coordinate distribution changes in thickness while the reference curvature $\alpha_j(m)$ is held fixed, the distribution of S_a changes directly, whereas S_λ is unchanged to leading order. The difference in power between the two components on the normal-thickness controls is therefore consistent with the fixed-reference local model.

Importantly, S_a is not a score-only feature. The eigendirections $u_j(x)$ that define S_a are obtained from the reference Hessian $H_P(x)$, and in the regime of Prop. 1 the dominant eigenspace approximates the local normal space. The Hessian therefore provides a geometry-adapted local coordinate system in which the high-dimensional score vector can be decomposed and its normal-direction variation isolated. Because S_a uses Hessian-derived eigendirections, this comparison shows that the Hessian enters the representation through the directional decomposition even when the eigenvalue magnitudes themselves carry little signal.

The role of S_λ is different. It retains the local curvature magnitude of the reference Hessian field. In general, $\alpha_j(m)$ need not be constant along the manifold, and for a nonquadratic normal density, $\lambda_j(F(m, \xi))$ may also depend on the normal coordinate ξ. Consequently, if samples from Q occupy

regions of the fixed reference curvature field with different frequencies, then even without estimating H_Q one may have

$$(\lambda_{P,j})_{\#}P \neq (\lambda_{P,j})_{\#}Q. \tag{147}$$

Thus, S_λ is not generally redundant with S_a. Indeed, a_j and λ_j are the first- and second-order coefficients of the same local negative-log-density expansion, and we retain them as complementary descriptors of local density geometry.

F Additional CIFAR-10-C experimental details and results

F.1 Dataset and evaluation protocol

Dataset. CIFAR-10-C (Hendrycks & Dietterich, 2019) is a distribution-shift benchmark constructed from CIFAR-10 (Krizhevsky et al., 2009), a dataset of 32×32 RGB natural images, by applying predefined image corruptions to the test images. We use the 10,000 images in the CIFAR-10 test split as the clean reference set. The version of CIFAR-10-C considered here contains 19 corruption types, with 10,000 images available for each corruption at each severity level from 1 to 5. We evaluate severities 1, 3, and 5.

Evaluation sampling. For each corruption, severity, and sample size

$$n \in \{8, 16, 32, 64, 128, 256\}, \tag{148}$$

we compute each discrepancy statistic using n images from each distribution.

For each repeat, we sample two disjoint subsets $\mathcal{X}^{(1)}_{\text{clean}}$ and $\mathcal{X}^{(2)}_{\text{clean}}$, each containing n images from the clean test set, and define the clean-vs-clean statistic as

$$T^{(0)} = T\Big(\mathcal{X}^{(1)}_{\text{clean}}, \mathcal{X}^{(2)}_{\text{clean}}\Big). \tag{149}$$

Separately, we sample n clean images and n corrupted images from the same corruption–severity condition and define

$$T^{(1)} = T(\mathcal{X}_{\text{clean}}, \mathcal{X}_{\text{corr}}) \tag{150}$$

as the clean-vs-corrupted statistic.

We repeat this sampling procedure 50 times for each setting and compute Cohen's d from the resulting $T^{(0)}$ and $T^{(1)}$ values using Eq. 125. The empirical 95th percentile of $T^{(0)}$ is used as the rejection threshold, and the fraction of $T^{(1)}$ values exceeding this threshold is reported as the alternative rejection rate.

To assess the null behavior of each method independently, we additionally perform a separate null-only evaluation consisting of 100 clean-vs-clean comparisons at the same sample size n. This evaluation is used only for the Type-I error audit and is not included in the computation of Cohen's d.

Image examples. Fig. 9 shows two clean images together with the corresponding examples under all 19 corruptions at severity 1.

F.2 Diffusion-score implementation

Reference diffusion-derived vector field. We use a DDPM (Ho et al., 2020) pretrained on CIFAR-10 as the reference model.[2] For an image $x \in [-1, 1]^D$ and diffusion step t, we use the cumulative DDPM scheduler coefficient $\bar{\alpha}_t$ to construct

$$\widetilde{x}_t(x) = \sqrt{\bar{\alpha}_t}\, x. \tag{151}$$

Thus, instead of sampling the Gaussian perturbation in the usual forward diffusion process,

$$x_t = \sqrt{\bar{\alpha}_t}\, x + \sqrt{1 - \bar{\alpha}_t}\, \epsilon, \tag{152}$$

we map every image to the diffusion-time input through the same deterministic transformation $x \mapsto \sqrt{\bar{\alpha}_t} x$.

[2] `https://huggingface.co/google/ddpm-cifar10-32`

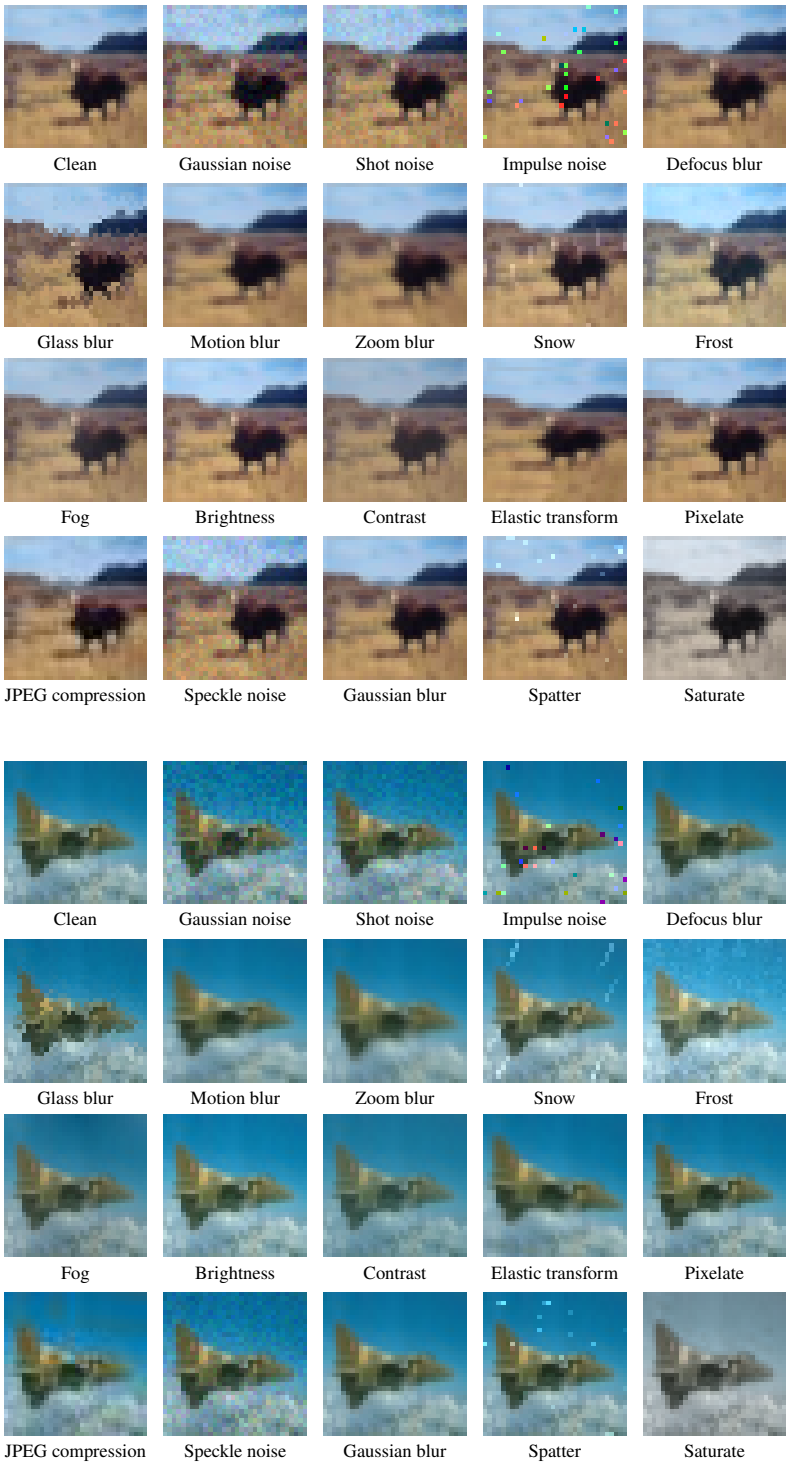


Figure 9: CIFAR-10-C severity 1 examples for two CIFAR-10 test images.

Let $\widehat{\epsilon}_\theta(\widetilde{x}_t(x), t)$ denote the UNet noise prediction. We define the reference vector field used in this experiment as

$$\widetilde{s}_t(x) = -\frac{\widehat{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\, x, t)}{\sqrt{1-\bar{\alpha}_t}}, \qquad t = 25. \tag{153}$$

Accordingly, we treat $\widetilde{s}_t$ as a deterministic vector field over clean-image coordinates induced by the pretrained diffusion model at the fixed diffusion step t. The same reference field is provided to Stein baselines that require a score.

Centered finite-difference local operator. To make spectral estimation tractable with a large diffusion model, we avoid explicitly forming the full Jacobian and instead evaluate its local action through centered finite differences. For a unit direction v,

$$\widetilde{H}_{t,\delta}(x)v := -\frac{\widetilde{s}_t(\Pi(x+\delta v)) - \widetilde{s}_t(\Pi(x-\delta v))}{2\delta}, \qquad \delta = 10^{-3}. \tag{154}$$

Here, $\Pi(x) = \mathrm{clip}(x, -1, 1)$ projects the perturbed image back onto the valid pixel range.

In regions where clipping is inactive, writing $\widetilde{s}_t(x) = \widehat{s}_t(\sqrt{\bar{\alpha}_t}x)$ gives, by the chain rule,

$$-D_x\widetilde{s}_t(x) = -\sqrt{\bar{\alpha}_t}\, D_{x_t}\widehat{s}_t(x_t)\big|_{x_t=\sqrt{\bar{\alpha}_t}x}. \tag{155}$$

Eq. 154 can therefore be viewed as a finite-difference approximation to the local Jacobian action of the learned reference field at a fixed diffusion step, expressed in clean-image coordinates. For computational tractability, we use this direct Jacobian action without explicit symmetrization. Explicit symmetrization would additionally require transpose-Jacobian products through the large diffusion model, substantially increasing the cost of each iterative spectral update. Accordingly, $\widetilde{H}_{t,\delta}$ is a benchmark-specific local operator induced by the learned reference field rather than an exactly symmetric Hessian.

The main local estimator starts from a random orthonormal initialization and performs eight block power-type iterations, applying $\widetilde{H}_{t,\delta}$ twice per iteration. The directions are orthonormalized after each iteration. For the resulting directions $u_j(x)$, we compute

$$\widetilde{\lambda}_j(x) = u_j(x)^\top \widetilde{H}_{t,\delta}(x)u_j(x), \qquad \widetilde{a}_j(x) = u_j(x)^\top \widetilde{s}_t(x). \tag{156}$$

F.3 Additional CIFAR-10-C results

Corruption-wise results. Fig. 10 shows Cohen's d separately for all 19 corruption types at severity 1 and $n = 128$. NLTS achieves an alternative rejection rate of 1.00 for all 19 corruptions under this condition. Although it does not attain the largest Cohen's d for every individual corruption, it is the only method in this comparison that achieves an alternative rejection rate of 1.00 across all 19 corruption types.

Low-sample regime at severity 1. Table 11 reports the numerical comparison at the smallest evaluated sample size, $n = 8$, for severity 1. At this most sample-limited condition, NLTS has the largest median Cohen's d among the methods shown in Table 11. KSD and Stein RFF attain similar alternative rejection rates, but these values should be interpreted jointly with their corresponding Type-I errors.

G Single-cell drug perturbation details

This section supplements Sec. 4.3 with details on the dataset and preprocessing, reference-only local-PCA diagnostics, the detection and evaluation protocol, and additional results omitted from the main text. The local-PCA diagnostics are constructed separately from the NLTS signature using only the local neighborhood structure of the training controls.

G.1 Dataset and preprocessing

Single-cell perturbation experiments measure cellular molecular states at single-cell resolution after interventions such as drug treatment. Here, we analyze changes in the transcriptomic distribution induced by drug treatment.

We use the sci-Plex3 dataset (Srivatsan et al., 2020) from the scPerturb harmonized release (Peidli et al., 2024), which provides public single-cell perturbation datasets in a common format. sci-Plex3

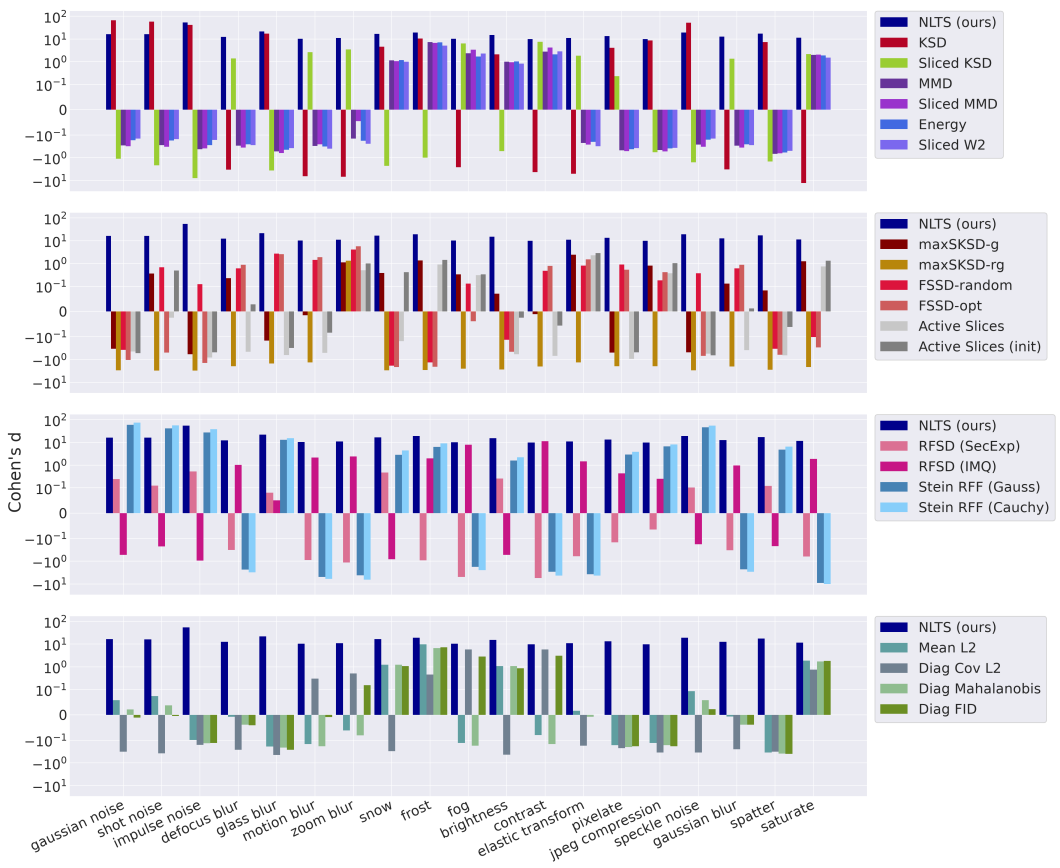


Figure 10: CIFAR-10-C corruption-wise comparison at severity 1 and $n = 128$ samples per distribution. Each bar shows the corruption-wise Cohen's d. Methods are separated across panels for readability.

is a large-scale chemical perturbation dataset in which many small-molecule compounds are administered at multiple concentrations and the resulting transcriptional responses are measured by single-cell RNA sequencing. We restrict the analysis to the K562 cell line, derived from human chronic myelogenous leukemia, and use vehicle-treated cells as the reference population.

As quality control, we remove cells with fewer than 200 detected genes and exclude genes detected in fewer than 3 cells. After filtering, 173,172 K562 cells remain, including 3,927 vehicle controls. We split the controls into

$$N_{\text{train}} = 2356, \qquad N_{\text{eval}} = 1571 \tag{157}$$

training controls and held-out evaluation controls, respectively.

Each drug is evaluated at four concentrations,

$$\mathcal{D} = \{10, 100, 1000, 10000\} \text{ nM}, \qquad |\mathcal{D}| = 4, \tag{158}$$

where the dose denotes the drug concentration applied to the corresponding cell population. After quality-control filtering, we retain only drugs with at least 256 available cells at each of the four doses. The resulting 10 drugs are Busulfan, SRT2104, Ofloxacin, Aminoglutethimide, Capecitabine, Amisulpride, S3I-201, Tazemetostat, Linifanib, and Baricitinib.

For each cell, total counts are normalized to 10^4 and then log-transformed. We select 2,000 highly variable genes (HVGs) according to the variance of log-normalized expression in the training controls. HVG selection and all subsequent fitted transformations use only the training controls and do not use drug-treated cells.

We then fit PCA to the 2,000-dimensional HVG expression vectors of the training controls and map them to 50-dimensional coordinates in the primary analysis. The PCA coordinates are centered using

Table 11: CIFAR-10-C at $n = 8$ samples per distribution and severity 1. Cohen's d is the median over 19 corruptions. Alt. rejection is the mean alternative rejection rate over 19 corruptions; it is interpreted as empirical power only when Type-I error is adequately controlled. Type-I error is measured by an independent clean-vs-clean evaluation at the same sample size.

Method	Median Cohen's d	Alt. rejection	Type-I error
NLTS	1.361	0.497	0.100
KSD	1.120	0.529	0.160
Sliced KSD	0.364	0.297	0.100
MMD	-0.393	0.039	0.100
Sliced MMD	-0.386	0.034	0.080
Energy	-0.324	0.048	0.070
Sliced W_2	-0.106	0.044	0.040
maxSKSD-g	0.099	0.097	0.140
maxSKSD-rg	0.380	0.138	0.020
FSSD-random	0.237	0.214	0.000
FSSD-opt	0.100	0.122	0.000
Active Slices	0.106	0.095	0.150
Active Slices (init)	0.268	0.153	0.110
RFSD (SecExp)	0.378	0.102	0.000
RFSD (IMQ)	-0.476	0.121	0.430
Stein RFF (Gauss)	1.249	0.505	0.030
Stein RFF (Cauchy)	1.095	0.507	0.060
Mean L2	-0.456	0.037	0.110
Diag Cov L2	-0.227	0.000	0.090
Diag Mahalanobis	-0.204	0.058	0.090
Diag FID	-0.511	0.033	0.140

the training controls and further rescaled by a single global scale factor shared across all dimensions. The resulting fitted transformation,

$$x \longmapsto T_P(x), \tag{159}$$

is fixed and applied identically to the held-out controls and drug-treated cells.

This PCA is a global preprocessing step shared by the detector and the geometry diagnostics. The local-PCA diagnostics below apply a separate PCA to the neighborhood of each reference anchor within this fixed coordinate system.

G.2 Reference-only local-PCA diagnostics

We next define the reference-derived tangent–normal decomposition and the geometry descriptors used in Sec. 4.3.

G.2.1 Local tangent-space construction from reference controls

Local tangent space. We select 64 reference anchors

$$a_1, \ldots, a_{64} \tag{160}$$

from the training controls and keep them fixed across all diagnostics. These anchors specify the locations at which local geometry is evaluated.

For each anchor a, we apply PCA to the centered local cloud formed by its 64 nearest neighbors among the training controls. The 64 reference anchors and the 64 nearest neighbors associated with each anchor play distinct roles: the former specify the evaluation locations, whereas the latter define the neighborhood used to estimate the local tangent space at each location.

Let d_a^{raw} be the smallest dimension whose cumulative explained variance reaches at least 90%. We define

$$d_a = \min\{d_{\max}, \max(2, d_a^{\text{raw}})\}, \tag{161}$$

using $d_{\max} = 40$ in the primary analysis. Sensitivity to the dimension cap is assessed with $d_{\max} \in \{20, 30, 40\}$.

Let $U_a \in \mathbb{R}^{D \times d_a}$ contain the leading d_a principal directions as columns. We define the tangent and normal projectors as

$$\Pi_{T,a} = U_a U_a^\top, \qquad \Pi_{N,a} = I - U_a U_a^\top. \tag{162}$$

Thus, each reference anchor a yields a d_a-dimensional local tangent subspace and its $(D - d_a)$-dimensional orthogonal complement. The geometry descriptors below measure drug-induced changes relative to this fixed tangent–normal decomposition.

G.2.2 Geometry descriptors

In the synthetic experiments, the data-generating mechanism can be controlled directly, so changes such as normal-thickness variation or tangential displacement can be specified in advance. In the single-cell setting, by contrast, a drug-induced distribution shift need not correspond to a single predefined geometric category. We therefore describe the observed local distributional changes relative to the tangent–normal structure estimated from vehicle-treated reference controls.

We use three descriptors: Normal thickness change, which measures changes in normal residual spread; Effective-rank contraction, which measures concentration of the local covariance spectrum; and Tangent mean energy, which measures local mean displacement along the reference tangent space.

For two cell samples P and Q, let $\mathcal{N}_P(a)$ and $\mathcal{N}_Q(a)$ denote the 32 cells in each sample nearest to anchor a. Each descriptor is computed at every anchor from these corresponding local neighborhoods.

Normal thickness change. For each point x, define its residual in the orthogonal complement of the local tangent space by

$$r_a(x) = \Pi_{N,a}(x - a). \tag{163}$$

For $S \in \{P, Q\}$, define the residual mean and dispersion within the 32-point local neighborhood as

$$\overline{r}_{S,a} = \frac{1}{32} \sum_{x \in \mathcal{N}_S(a)} r_a(x), \tag{164}$$

and

$$v_{S,a} = \frac{1}{32} \sum_{x \in \mathcal{N}_S(a)} \|r_a(x) - \overline{r}_{S,a}\|_2^2 . \tag{165}$$

We define Normal thickness change as

$$\Delta_{\text{thick}}(a) = \frac{v_{Q,a} - v_{P,a}}{v_{P,a} + 10^{-8}}. \tag{166}$$

Positive values indicate increased residual spread in the local normal space defined from the reference controls, whereas negative values indicate reduced spread.

Effective-rank contraction. Let $C_{P,a}$ and $C_{Q,a}$ denote the covariance matrices of the corresponding 32-point local neighborhoods. For a covariance matrix C with eigenvalues $\gamma_1, \ldots, \gamma_D$, define the participation-ratio effective rank as

$$r_{\text{eff}}(C) = \frac{\left(\sum_{j=1}^{D} \gamma_j\right)^2}{\sum_{j=1}^{D} \gamma_j^2} = \frac{\text{tr}(C)^2}{\text{tr}(C^2)}. \tag{167}$$

The anchor-level change in effective rank is

$$\Delta r_{\text{eff}}(a) = r_{\text{eff}}(C_{Q,a}) - r_{\text{eff}}(C_{P,a}), \tag{168}$$

and in the figures and association analyses we use

$$C_{\text{eff}}(a) = -\Delta r_{\text{eff}}(a) \tag{169}$$

as the Effective-rank contraction. Thus, $C_{\text{eff}} > 0$ indicates that the local covariance spectrum under Q is more concentrated in a smaller number of dominant directions than under P.

Effective-rank contraction is computed from the covariance spectrum of the local neighborhood cloud. Under a local Gaussian approximation with covariance Σ, the negative-log-density Hessian is Σ^{-1}. Under this approximation, the covariance and negative-log-density Hessian spectra are related through reciprocal eigenvalues. We therefore use Effective-rank contraction as a separate second-order geometry diagnostic, not as a direct estimate of the NLTS signature.

Tangent mean energy. At each anchor, define the local means

$$\mu_{P,a} = \frac{1}{32} \sum_{x \in \mathcal{N}_P(a)} x, \qquad \mu_{Q,a} = \frac{1}{32} \sum_{x \in \mathcal{N}_Q(a)} x, \tag{170}$$

and let $\Delta\mu_a = \mu_{Q,a} - \mu_{P,a}$. The Tangent mean energy is

$$E_T(a) = \|\Pi_{T,a} \Delta\mu_a\|_2^2 , \tag{171}$$

which measures the magnitude of local mean displacement along the reference tangent space.

G.3 Detection and evaluation protocol

G.3.1 Score model and detectors

The detectors use a scalar-potential score model trained by denoising score matching (DSM) (Vincent, 2011) using only the training controls; drug-treated cells are not used for score-model fitting. Rather than directly parameterizing a vector-valued score, we parameterize a scalar function $\ell_\theta : \mathbb{R}^D \to \mathbb{R}$ and define

$$\widehat{\boldsymbol{s}}_P(x) := \nabla_x \ell_\theta(x). \tag{172}$$

DSM trains this gradient field to approximate the reference score; ℓ_θ itself is therefore used as a scalar potential and is not treated as a normalized density estimate.

This parameterization also makes the local second-order operator symmetric by construction:

$$\widehat{\boldsymbol{H}}_P(x) := -\nabla_x \widehat{\boldsymbol{s}}_P(x) = -\nabla_x^2 \ell_\theta(x). \tag{173}$$

Hessian–vector products are evaluated by automatic differentiation. To target directions of largest absolute curvature, block power iteration is applied to $\widehat{\boldsymbol{H}}_P(x)^2$. After eight iterations, the resulting directions $\boldsymbol{u}_j$ are used to compute

$$\widehat{\lambda}_j = \boldsymbol{u}_j^\top \widehat{\boldsymbol{H}}_P \boldsymbol{u}_j, \qquad \widehat{a}_j = \boldsymbol{u}_j^\top \widehat{\boldsymbol{s}}_P. \tag{174}$$

Table 12 summarizes the score-model configuration used in the primary analysis. NLTS uses $k = 2$ and multiscale RBF–MMD. These detector settings are fixed across drugs, doses, and sample sizes. For each score-model seed, NLTS and KSD use the same learned reference score.

Table 12: Scalar-potential score-model configuration used in the primary single-cell analysis.

Setting	Value
Input dimension	50
Network depth	4
Hidden width	256
Residual quadratic base	Enabled
DSM noise scale σ	0.08
Optimization steps	50,000
Batch size	256
Learning rate	2×10^{-4}
Weight decay	10^{-6}
Score-model seeds	$0, 1, 2$

G.3.2 Detection protocol

We use sample sizes

$$n \in \{64, 128, 256\} \tag{175}$$

and perform 50 alternative repeats for each drug–dose–sample-size condition. The test level, null calibration, and independent null validation follow the common protocol in App. C.

To assess sensitivity to the global PCA dimension and DSM smoothing scale, we also evaluate

$$D_{\text{PCA}} \in \{50, 100, 200, 500\}, \qquad \sigma \in \{0.04, 0.08, 0.16\}. \tag{176}$$

For each (D_{PCA}, σ) combination, we train a separate single-σ DSM model for 50,000 optimization steps. The primary detection results do not average across PCA dimensions or smoothing scales. The geometry–power association reported in the main text uses $D_{\text{PCA}} = 50$ and $\sigma = 0.08$.

G.3.3 Drug-level aggregation for geometry--power association

To evaluate the drug-level association between geometry and detection performance, we aggregate each quantity to a single value per drug.

Geometry aggregation. For each drug–dose condition and repeat, we draw two disjoint samples P and P_0 from the held-out vehicle controls and a sample Q from the drug-treated cells. All three samples have size

$$|P| = |P_0| = |Q| = 256. \tag{177}$$

Thus, P and P_0 are both untreated control samples, whereas Q contains cells from the corresponding drug–dose condition.

For each anchor a, we select the 32 cells nearest to the anchor from each 256-cell sample, yielding $\mathcal{N}_P(a)$, $\mathcal{N}_{P_0}(a)$, and $\mathcal{N}_Q(a)$.

Let $G_a(P, Q)$ denote an anchor-level geometry descriptor. We aggregate across the 64 anchors using the median,

$$G(P, Q) = \text{median}_{a\in\{a_1,\dots,a_{64}\}}\, G_a(P, Q). \tag{178}$$

Within the same repeat, the treated comparison (P, Q) and the control–control comparison (P, P_0) share the same sample P. For drug g, dose d, and repeat r, we define

$$\widetilde{G}^{(r)}_{g,d} = G(P, Q) - G(P, P_0). \tag{179}$$

This subtraction centers the descriptor by the corresponding control–control variation observed in that repeat.

We repeat this sampling procedure 50 times for each drug g and dose d and define the drug–dose-level geometry descriptor as

$$\overline{G}_{g,d} = \frac{1}{50}\sum_{r=1}^{50} \widetilde{G}^{(r)}_{g,d}. \tag{180}$$

Finally, we average equally over the four doses,

$$\overline{G}_g = \frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}} \overline{G}_{g,d}, \tag{181}$$

to obtain the drug-level geometry descriptor for drug g. Each dose receives equal weight regardless of the number of available cells in that drug–dose condition.

Detection-power aggregation. For drug g, dose d, sample size n, score-model seed s, and method m, we define the empirical rejection rate over the 50 alternative repeats as

$$\widehat{\pi}^{(m)}_{g,d,n,s} = \frac{1}{50}\sum_{r=1}^{50} \mathbf{1}\Big\{T^{(m)}_{g,d,n,s,r} > \tau_{m,n,s}\Big\}. \tag{182}$$

For a fixed method, we first average equally over the three score-model seeds,

$$\widehat{\pi}_{g,d,n} = \frac{1}{3}\sum_{s=1}^{3} \widehat{\pi}_{g,d,n,s}, \tag{183}$$

and then over the three sample sizes,

$$\widehat{\pi}_{g,d} = \frac{1}{3}\sum_{n\in\{64,128,256\}} \widehat{\pi}_{g,d,n}, \tag{184}$$

to obtain dose-level detection power. Finally, we average equally over the four doses,

$$\widehat{\pi}_g = \frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}} \widehat{\pi}_{g,d}, \tag{185}$$

to obtain drug-level detection power.

Thus, the main-text analysis with one point per drug uses the 10 pairs

$$\left(\overline{G}_g, \widehat{\pi}_g\right), \qquad g = 1, \ldots, 10. \tag{186}$$

Individual doses, sample sizes, score-model seeds, and repeats are therefore not pooled as separate observations; instead, condition-specific rejection rates are aggregated with equal weights to obtain one drug-level quantity.

G.3.4 Geometry--power association protocol

Table 1 evaluates rank associations between the drug-level geometry descriptors and drug-level detection power defined in Sec. G.3.3 across the 10 drugs.

For Effective-rank contraction and Tangent mean energy, we use partial Spearman correlation to account for their covariation. Let X_g denote the target geometry descriptor, Y_g the drug-level detection power, and Z_g the other geometry descriptor used for adjustment. Let x_g, y_g, and z_g denote their respective rank-transformed values. We regress x_g and y_g separately on z_g and compute

$$e_{X,g} = x_g - (\widehat{\alpha}_X + \widehat{\beta}_X z_g), \qquad e_{Y,g} = y_g - (\widehat{\alpha}_Y + \widehat{\beta}_Y z_g). \tag{187}$$

The partial Spearman correlation is then

$$\rho_{XY \cdot Z} = \mathrm{Corr}\left(e_X, e_Y\right). \tag{188}$$

Accordingly, the association between Effective-rank contraction and detection power adjusts for Tangent mean energy, whereas the association between Tangent mean energy and detection power adjusts for Effective-rank contraction. For Normal thickness change, we use the ordinary Spearman rank correlation between the drug-level descriptor and detection power without adjustment.

For the partial correlations, nominal p-values are computed using the standard t approximation with one adjusted variable, giving $10 - 3 = 7$ degrees of freedom. Because the analysis contains only 10 drugs and evaluates multiple exploratory associations, we report nominal p-values without multiple-testing correction.

G.4 Additional single-cell results

This subsection reports the local-PCA geometry profile, unadjusted associations, and auxiliary visualizations omitted from the main text.

G.4.1 Drug-level local-PCA geometry profile

Fig. 11 shows the three geometry descriptors for the 10 drugs.

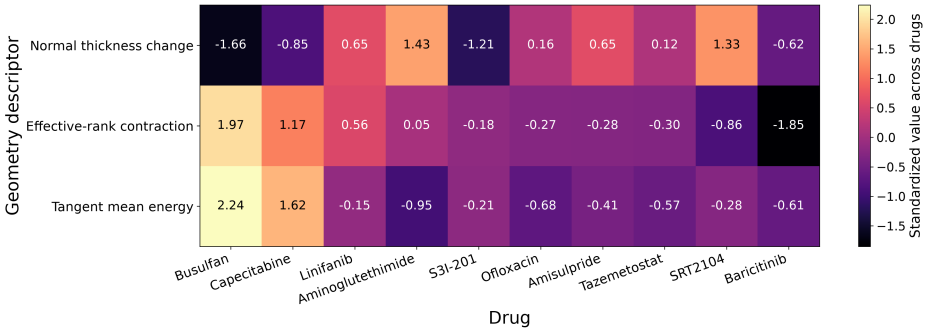


Figure 11: Drug-level local-PCA geometry profile for the single-cell drug perturbation experiment. The three columns correspond to Normal thickness change, Effective-rank contraction, and Tangent mean energy. Drug-level values are the equally weighted dose averages defined in Eq. 181. For visualization, each descriptor is standardized across drugs.

G.4.2 Marginal geometry--power relationships

Table 13 reports the unadjusted Spearman rank associations between drug-level geometry descriptors and drug-level detection power across the 10 drugs.

Table 13: Unadjusted Spearman rank associations between the geometry descriptors and drug-level detection power. Drug-level quantities are defined in Sec. G.3.3. Reported p-values are nominal and are not corrected for multiple testing.

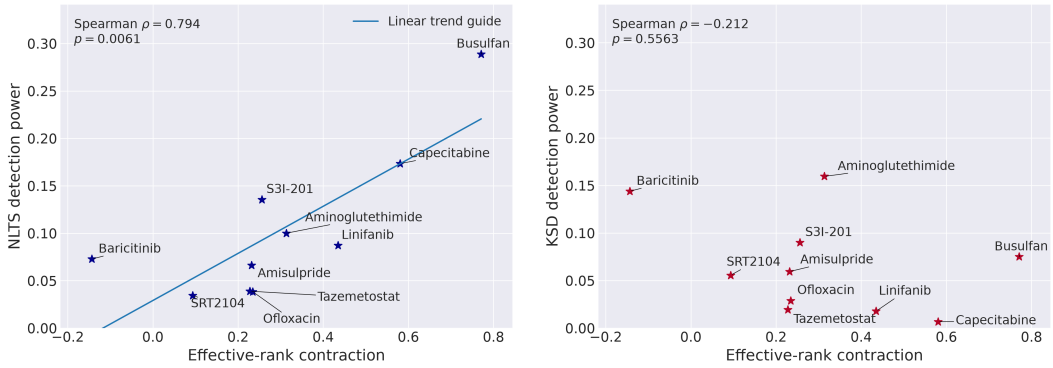
	NLTS		KSD	
Descriptor	ρ	p	ρ	p
Normal thickness change	-0.600	0.0667	0.067	0.8548
Effective-rank contraction	0.794	0.0061	-0.212	0.5563
Tangent mean energy	0.564	0.0897	-0.418	0.2291

Effective-rank contraction and Tangent mean energy are themselves associated across the 10 drugs,

$$\rho = 0.588, \qquad p = 0.0739. \tag{189}$$

The partial Spearman analysis in the main text therefore uses the protocol defined in Sec. G.3.4 to account for this covariation.

Fig. 12 visualizes the marginal association between Effective-rank contraction and detection power for NLTS and KSD. For NLTS, Effective-rank contraction is positively associated with detection power ($\rho = 0.794$, nominal $p = 0.0061$), whereas the corresponding association for KSD is weak ($\rho = -0.212$, nominal $p = 0.5563$).



(a) NLTS (ours) (b) KSD

Figure 12: Marginal association between Effective-rank contraction and drug-level detection power. Each point represents one drug, with drug-level quantities defined in Sec. G.3.3. (a) NLTS: $\rho = 0.794$, nominal $p = 0.0061$. (b) KSD: $\rho = -0.212$, nominal $p = 0.5563$. Linear trend lines are included only as visual guides and are not used to estimate the Spearman correlations.