Statistical Foundations of Local Geometric Fault Detection: The Curvature Channel

Gregory Provan
School of Computer Science and IT, University College Cork, Ireland

Abstract

Geometric and topological methods have recently emerged as a promising approach to fault detection in hybrid and cyber–physical systems. Unlike classical residual-based methods, which detect violations of a parametric model, geometric methods monitor changes in the local structure of trajectories through quantities such as intrinsic dimension, smoothness, tangent-cone geometry, and consistency with physical constraints. Despite encouraging empirical results, there is currently no theory characterising when such geometric signatures are detectable, distinguishable, or reliable under noise. This paper develops a statistical theory of local geometric fault detection, with a focus on curvature-based approaches. We study faults detectable by displacement in a curvature signature space, and derive the noise floor of geometric estimators based on finite differences and local neighbourhood averaging. The resulting estimator variance scales as $\sigma_{\text{est}} = O(\sigma/(\Delta t^2\sqrt{k}))$, exposing the fundamental trade-off between sampling rate, noise level, and neighbourhood size. Using this noise model, we derive geometric analogues of the classical detectability and isolability conditions of fault-detection theory, expressed directly in terms of separation in signature space. We further prove an exclusion result showing that a broad class of smooth parametric faults cannot activate curvature-based geometric channels, yielding a form of label-free fault-class identification. These results lead to a practical pre-deployment assessment procedure that determines, from nominal data alone, whether a curvature-based fault detector is expected to achieve a specified false-alarm rate, detection probability, and isolation capability. The theory thereby places local geometric fault detection on the same statistical footing as classical residual-based diagnosis while clarifying its distinctive strengths, limitations, and operating regime.

2012 ACM Subject Classification Computing methodologies → Causal reasoning and diagnostics

Keywords and phrases Geometric metrics, Diagnosis methods, Theory

Digital Object Identifier 10.4230/OASIcs.DX.2026.13

Funding Funded by Research Ireland grant 13/RC/2094_P2.

1 Introduction

Fault detection has traditionally been formulated as the problem of identifying statistically significant deviations from a nominal model. Classical approaches include analytical redundancy relations, observer-based residual generation, parity-space methods, generalized likelihood ratio tests, and Bayesian state-estimation techniques. Although these methods differ in representation and implementation, they share a common principle: fault information is extracted through a residual measuring disagreement between observations and a parametric model of the system.

Over the last decade, a different view has begun to emerge. Rather than monitoring model mismatch, one may monitor the *geometry* of the observed trajectory itself. Hybrid systems, switched systems, and many cyber–physical systems naturally generate trajectories that lie on unions of locally smooth manifolds with boundaries, intersections, and singularities. Structural faults can alter this geometry even when local algebraic relationships remain approximately satisfied [4, 8]. Examples include mode-creation and mode-deletion events, stick–slip transitions, changing guard conditions, loss of smoothness, and changes in accessibility between operating regimes. Such phenomena motivate geometric diagnostic quantities including intrinsic dimension, manifold smoothness, tangent-cone structure, and consistency of local physical constraints.

Recent work has demonstrated that these geometric quantities can be estimated from sampled trajectories and can reveal classes of faults that are difficult to detect using traditional residuals. However, the diagnostic literature currently lacks answers to three fundamental questions.

1. **Detectability.** When does a fault produce a sufficiently large change in geometric signature to be distinguished from estimator noise?
2. **Isolability.** When are two fault classes sufficiently separated in geometric signature space to permit reliable classification?
3. **Feasibility.** Can these questions be answered before deployment, using only nominal data and a proposed fault taxonomy?

The absence of such results makes it difficult to design geometric detectors systematically. In practice, neighbourhood size, sampling rate, and detection thresholds are often chosen heuristically. Consequently, success or failure of a geometric detector may be determined as much by implementation choices as by the underlying physical system.

This paper develops a statistical theory of local geometric fault detection. Our starting point is a geometric signature vector $\Phi(\varphi) = \big(d_{\text{eff}}, \kappa_{\text{gen}}, \kappa_{\text{OR}}, C\big)$, whose components measure local dimension, smoothness, tangent-cone structure, and semantic consistency. A fault is represented by the displacement it induces in this signature space. The central observation is that geometric estimators are simultaneously affected by two opposing mechanisms: finite differencing amplifies observation noise, whereas local averaging suppresses it. Quantifying this balance yields an explicit estimator noise scale that becomes the fundamental quantity governing all subsequent results.

Building on this noise model, we derive geometric analogues of the classical detectability and isolability conditions from the fault diagnosis literature. The resulting criteria are expressed as simple separation conditions in signature space and yield a geometric signal-to-noise ratio interpretation analogous to that used in residual-based diagnosis. We then establish an exclusion theorem showing that a broad class of smooth parametric faults cannot activate curvature-based geometric channels. This result enables a limited but rigorous form of label-free fault-class identification: certain geometric signatures can exclude entire classes of faults without requiring fault examples or a dictionary of fault patterns.

The theoretical results lead directly to a practical contribution: a Geometric Feasibility Assessment procedure that determines whether a proposed detector is expected to achieve a specified false-alarm rate, power, and class separability. The procedure requires only nominal data and therefore provides a pre-deployment certification mechanism for geometric fault detection.

The paper does not propose a new fault-detection algorithm. Instead, it develops the statistical foundations needed to analyse existing and future geometric detectors in the same way that classical detectability and isolability theory analyses residual-based methods. Our aim is to clarify the operating regime of geometric fault detection, identify the conditions under which it can be expected to succeed, and establish a common language for comparing geometric and classical diagnostic approaches.

Contributions. The main contributions are:

1. A derivation of the noise floor of local geometric estimators, showing

$$\sigma_{\text{est}} = O\left(\frac{\sigma}{\Delta t^2 \sqrt{k}}\right).$$

2. A geometric detectability theorem relating fault detectability to separation in geometric signature space.
3. A geometric isolability theorem characterising when fault classes are distinguishable from geometric observations alone.
4. An exclusion theorem showing that smooth parametric faults cannot activate curvature-based geometric channels.
5. A Geometric Feasibility Assessment procedure that translates the theory into a practical pre-deployment design workflow.

2 Preliminaries

This paper studies fault detection using local geometric properties of measured trajectories. Before introducing detectability and isolability results, we define the observation model, geometric measurement field, and signature representation used throughout.

2.1 Hybrid-System Observations

Let $x(t) \in \mathcal{X}$ denote the state trajectory of a dynamical system. We assume observations are available through a sensor map $O : \mathcal{X} \to \mathbb{R}^p$, giving measurements $y(t) = O(x(t)) + \epsilon(t)$, where $\epsilon(t) \sim \mathcal{N}(0, \Sigma)$ is observation noise.

In classical fault detection the diagnostic signal is typically a residual $r(t) = y(t) - \hat{y}(t)$, constructed from a model prediction. The geometric approach considered here replaces residuals by local geometric descriptors extracted directly from the observed trajectory.

2.2 Local Neighbourhoods

Let $y_t := y(t)$ denote a sampled observation. For each sample we define a local neighbourhood $N_k(y_t) = \{y_{t_1}, \ldots, y_{t_k}\}$, consisting of the k nearest observations under a chosen distance metric.

All geometric quantities in this paper are computed from such neighbourhoods. Consequently the approach is local and does not require a global model of the state space.

2.3 Geometric Measurement Fields

▶ **Definition 1** (Geometric Measurement Field)**.** *A geometric measurement field is a map $M : \mathbb{R}^p \to \mathbb{R}^r$ that converts a local neighbourhood into a collection of geometric descriptors.*

The canonical field used in this work is

$$M(y_t) = \big(d_{\text{eff}}(y_t), \kappa_{\text{gen}}(y_t), \kappa_{\text{OR}}(y_t), C(y_t)\big),$$

consisting of:

- d_{eff}: local intrinsic dimension estimated from PCA eigenvalue structure;
- κ_{gen}: smoothness or curvature residual measuring departure from a locally smooth chart;
- κ_{OR}: tangent-cone splitting statistic measuring local multi-directionality;
- C: semantic-consistency score measuring violation of local physical laws or constraints.

These quantities define a geometric analogue of the residual vector used in classical fault diagnosis.

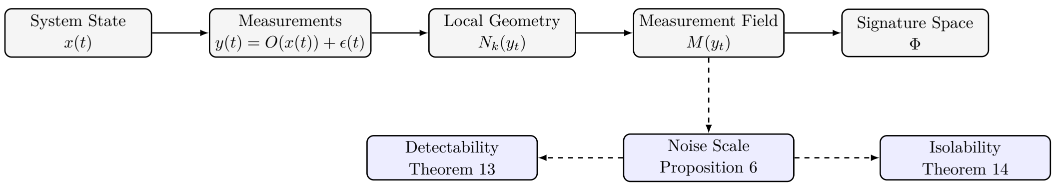

Figure 1 Conceptual framework of local geometric fault detection. The observed trajectory is transformed into a local geometric signature. Proposition 6 characterises the estimator noise floor, which in turn determines the detectability and isolability conditions of Theorems 13 and 14.

2.4 Geometric Signature Space

The output of the measurement field is called the *geometric signature*:

$$\Phi = \big(d_{\text{eff}}, \kappa_{\text{gen}}, \kappa_{\text{OR}}, C\big) \in S,$$

where $S \subseteq \mathbb{R}^4$ is the signature space.

A nominal operating condition corresponds to a region $\Phi(\nu) \in S$, while a fault φ induces a displaced signature $\Phi(\varphi)$.

The central question of this paper is the *geometric detectability problem*:

▶ **Definition 2** (Geometric detectability problem). *When is the displacement $\|\Phi(\varphi) - \Phi(\nu)\|$ large enough to be distinguished from estimator noise?*

2.5 Fault Classes

We distinguish between two broad categories of faults.

▶ **Definition 3** (Parametric fault). *A parametric fault modifies parameters of the underlying dynamics while preserving the qualitative geometry of the trajectory. Examples include coefficient drift, gain changes, and smooth degradation processes.*

▶ **Definition 4** (Structural fault). *A structural fault changes the local geometry of the trajectory, such as its smoothness, dimension, accessibility structure, or constraint consistency. Examples include stick–slip transitions, mode creation, guard perturbations, and topology-changing events.*

The purpose of the geometric measurement field is to detect structural faults through changes in signature space.

2.6 From Observations to Decisions

The complete information flow studied in this paper is

$$x(t) \to y(t) \to N_k(y_t) \to M(y_t) \to \Phi(y_t) \to \text{fault decision.}$$

Sections 3–5 analyse the statistical properties of this pipeline and derive conditions under which faults are detectable and distinguishable.

3 Geometric Detectability and Isolability

This section states the condition under which the geometric approach succeeds, proves it, relates it to the classical parametric results it generalises, and turns it into a decision procedure a practitioner can run *before* deployment. We are explicit at the outset about what is and is not new: the condition is the *geometric instance* of a well-established family of detectability/isolability results [1, 2, 3, 5]. Its novelty lies not in its form but in the *signal* it applies to (a geometric invariant rather than a model residual) and in the *specific noise scale* that differentiation and local averaging induce. We denote the geometric approach defined in this section as Geometric Model-Based Diagnosis (GMBD).

3.1 Setup

We define a cause-to-symptom map using $\Phi_M = O_M \circ (\text{physics}) : \mathcal{C} \to \mathcal{S}_M$, with ν the nominal cause. For GMBD, O_M is the geometric field and the symptom space is the signature vector

$$\Phi(\phi) \;=\; \big(d_{\text{eff}},\, \kappa_{\text{gen}},\, \kappa_{\text{OR}},\, C\big)(\phi) \;\in\; \mathcal{S} \subseteq \mathbb{R}^4. \tag{1}$$

Observation is noisy, $y = O(x) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \Sigma)$, $\Sigma = \sigma^2 I$, sampled at interval Δt, with the field estimated from neighbourhoods of k samples in d dimensions. Write $\widehat{\Phi}$ for the estimated signature.

▶ **Definition 5** (Estimator noise scale). *$\sigma_{\text{est}}(k, \sigma, \Delta t)$ is the standard deviation of $\widehat{\Phi}(\nu)$ about its mean on nominal data: the radius of the* nominal ball *in signature space.*

3.2 The noise scale of a geometric estimator

The distinguishing quantitative feature of the geometric approach is that its invariants are *differential*: they are built from finite differences of the observed trajectory, which amplify noise, and from local averages over k neighbours, which suppress it. These two effects compose.

▶ **Proposition 6** (Geometric noise scale). *For the curvature channel κ_{gen}, computed as a windowed second difference over neighbourhoods of k samples,*

$$\sigma_{\text{est}}(k, \sigma, \Delta t) \;=\; \frac{C\,\sigma}{\Delta t^2\,\sqrt{k}}\,\big(1 + o(1)\big), \qquad C = O(1), \tag{2}$$

i.e. linear in the observation noise σ, quadratically amplified by the sampling interval, and suppressed as $k^{-1/2}$ by local averaging.

Proof. Let $y_t = x_t + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 I_d)$. The second difference is $D^2 y_t = (y_{t+1} - 2y_t + y_{t-1})/\Delta t^2$. Its noise component is $(\epsilon_{t+1} - 2\epsilon_t + \epsilon_{t-1})/\Delta t^2$, a linear combination of independent Gaussians with coefficient vector $(1, -2, 1)$, hence

$$\text{Var}\big[D^2 y_t\big] = \frac{\sigma^2\,\|(1,-2,1)\|_2^2}{\Delta t^4} = \frac{6\,\sigma^2}{\Delta t^4}, \tag{3}$$

so a single second difference has noise standard deviation $\sqrt{6}\,\sigma/\Delta t^2$: the Δt^{-2} amplification. The estimator averages $\|D^2 y\|$ over a window of k samples. The window terms are weakly dependent (each ϵ_t enters three consecutive second differences), so the effective number of independent terms is k/c_0 for a constant $c_0 \le 3$; by the central limit theorem for m-dependent sequences the standard error of the window mean is $\sqrt{c_0}$ times that of the independent case, giving standard deviation $\sqrt{6c_0}\,\sigma/(\Delta t^2\sqrt{k})$. Setting $C = \sqrt{6c_0}$ yields (2). The $o(1)$ absorbs the curvature of the true trajectory, which contributes a deterministic offset (the signal, not the noise) and the m-dependence correction, both vanishing relative to the leading term as $\sigma/\Delta t^2 \to \infty$. ◀

▶ Remark 7 (Empirical confirmation). Both factors in (2) are confirmed on the Cart–Pole. Sweeping σ over three decades at fixed k, the measured nominal floor satisfies $\kappa_{\text{gen}}^{\text{noise}}\Delta t^2/\sigma = 2.95 \pm 0.03$, constant as predicted (Fig. 4, left). Sweeping k at fixed σ, the product $\mathrm{sd}(\kappa_{\text{gen}})\sqrt{k}$ is constant at 127 ± 1 for $k \in [40, 160]$, confirming the $k^{-1/2}$ law. The measured coefficient of variation falls $0.091 \to 0.047$ as $k : 40 \to 200$, consistent with $k^{-1/2}$.

3.3 Identifiability of causes: an image/fibre condition

Theorem 14 will give a *statistical* condition for separating fault classes at a given noise level. Before that, it is worth isolating the *information-theoretic* question underneath it: are the classes distinguishable *even in principle*, with noise set aside? The two questions have different answers and different remedies, and conflating them is a common source of confusion – a limit that no estimator can beat is a different object from a limit that more data would fix.

Fix a detector M with observation map O_M and let

$$\Phi_M \;=\; O_M \circ (\text{physics}) \;:\; \mathcal{C} \longrightarrow \mathcal{S}_M \tag{4}$$

be the induced *cause-to-symptom* map from the space of causes $\mathcal{C}$ (faults, together with the nominal cause ν) to the detector's symptom space $\mathcal{S}_M$; for the geometric detector, $\mathcal{S}_M$ is the signature space of (1). Let $\sim$ denote the *diagnostic equivalence* on $\mathcal{C}$: $\phi \sim \phi'$ when the application does not require the two causes to be told apart, and write $[\phi]$ for the class of ϕ and $\mathcal{C}/{\sim}$ for the set of classes. A *decision rule* is a measurable map $\delta : \mathcal{S}_M \to \mathcal{C}/{\sim}$, possibly randomised, in which case $\delta(s)$ is a probability distribution on $\mathcal{C}/{\sim}$.

The following definition names the only structure that matters.

▶ **Definition 8** (Fibre purity). *The fibre of Φ_M over $s \in \mathcal{S}_M$ is $\Phi_M^{-1}(s) = \{\phi \in \mathcal{C} : \Phi_M(\phi) = s\}$. The fibre is* pure *if it meets exactly one $\sim$-class, i.e. if $\phi, \phi' \in \Phi_M^{-1}(s) \Rightarrow \phi \sim \phi'$. The map Φ_M has* pure fibres *if every fibre over a point of its image $\operatorname{Im}\Phi_M$ is pure.*

Detection and identification are then governed by two *different* features of the same map: detection by its *image*, identification by its *fibres*.

▶ **Proposition 9** (Detectability = image separation). *In the noiseless setting, a detector M can decide the presence of a fault ϕ from its symptom alone if and only if $\Phi_M(\phi) \neq \Phi_M(\nu)$. Detection is therefore a property of the image of Φ_M: it asks only whether the faulted symptom is distinct from the nominal one.*

Proof. ($\Leftarrow$) If $\Phi_M(\phi) \neq \Phi_M(\nu)$, the rule that reports "fault" exactly on $\mathcal{S}_M \setminus \{\Phi_M(\nu)\}$ decides correctly at ϕ and at ν. ($\Rightarrow$) If $\Phi_M(\phi) = \Phi_M(\nu) =: s$, then any rule δ returns the same verdict $\delta(s)$ under both causes, so it is wrong under at least one of them. ◀

▶ **Proposition 10** (Identifiability = fibre purity). *Consider the noiseless setting, in which the detector observes $s = \Phi_M(\phi)$ exactly. Then:*

(i) ***(Sufficiency.)*** *If Φ_M has pure fibres (Def. 8), there exists a decision rule δ with $\delta(\Phi_M(\phi)) = [\phi]$ for every $\phi \in \mathcal{C}$: the causes are identifiable up to $\sim$.*

(ii) ***(Necessity, and impossibility.)*** *If some fibre $\Phi_M^{-1}(s^*)$ meets two distinct classes $[\phi_1] \neq [\phi_2]$, then for* every *decision rule δ – deterministic or randomised – the worst-case error probability over $\{\phi_1, \phi_2\}$ is at least $1/2$. No rule reading only the symptom can separate the two causes, at any sample size.*

Identification is therefore a property of the fibres of Φ_M.

Proof.

(i) Suppose every fibre over $\operatorname{Im}\Phi_M$ is pure. For $s \in \operatorname{Im}\Phi_M$ the set $\Phi_M^{-1}(s)$ is nonempty and, by purity, contained in a single class; define $\delta(s)$ to be that class, and define δ arbitrarily off $\operatorname{Im}\Phi_M$ (such symptoms never occur). The map is well defined precisely because purity makes the choice unambiguous, and by construction $\delta(\Phi_M(\phi)) = [\phi]$ for all ϕ.

(ii) Suppose $\phi_1, \phi_2 \in \Phi_M^{-1}(s^*)$ with $[\phi_1] \neq [\phi_2]$. Both causes produce the *same* symptom s^*, so the rule's output is the same random variable $\delta(s^*)$ in both cases; it cannot depend on which of ϕ_1, ϕ_2 occurred, since the data it reads are identical. Let $q = \Pr[\delta(s^*) = [\phi_1]]$. The error probability is $1 - q$ under ϕ_1 and at least q under ϕ_2 (as $[\phi_1] \neq [\phi_2]$, reporting $[\phi_1]$ is an error under ϕ_2). Hence

$$\max\{\Pr[\text{err} \mid \phi_1], \Pr[\text{err} \mid \phi_2]\} \;\geq\; \max\{1-q,\ q\} \;\geq\; \tfrac{1}{2},$$

the last step because $\max\{1-q, q\} \geq \frac{1}{2}$ for every $q \in [0,1]$. A deterministic rule is the case $q \in \{0,1\}$, giving error 1 under one of the two causes. The bound involves no sample size: the obstruction is that the observation map has already discarded the distinguishing information, so no amount of data recovers it. ◀

▶ **Corollary 11** (Detection does not imply identification). *A detector may satisfy Proposition 9 for every fault of interest – so that every fault is detectable – while violating Proposition 10, so that no fault cause is identifiable. Image separation and fibre purity are independent properties.*

Proof. Take $\mathcal{C} = \{\nu, \phi_1, \phi_2\}$ with $[\phi_1] \neq [\phi_2]$ and a map with $\Phi_M(\phi_1) = \Phi_M(\phi_2) = s^* \neq \Phi_M(\nu)$. Every fault is detectable by Proposition 9, since $s^* \neq \Phi_M(\nu)$; but the fibre over s^* straddles two classes, so by Proposition 10(ii) no rule identifies the cause. ◀

Corollary 11 is not a contrivance: it is the generic situation for geometric detectors, and the reason is structural rather than incidental. The geometric field reports *regularity properties of the observed trajectory.* Distinct physical mechanisms routinely share those properties – a seizing friction element and a snapped kinematic link both destroy C^r regularity and so both raise the curvature channel – so their common fibre straddles two cause classes even though each is comfortably separated from nominal. The geometric detector therefore *detects* both and *identifies* neither. Escaping the conclusion requires enlarging O_M – an additional channel (for instance a constraint/law residual, which the snapped link violates and the seizing friction does not), or an external signature-to-cause dictionary. Neither is a better decision rule on the same data; both change the data.

▶ Remark 12 (Fibre purity is the $\sigma \to 0$ limit of isolability). Proposition 10 and Theorem 14 are the noiseless and noisy halves of one statement, and it is useful to see how they compose. Theorem 14 requires class centroids to satisfy $\min_{K \neq K'} \|\Phi(K) - \Phi(K')\| > 2z_{1-\alpha/2}\,\sigma_{\text{est}}$. Letting $\sigma_{\text{est}} \to 0$ – by Proposition 6, the joint limit $\sigma \to 0$ or $k \to \infty$ – the right-hand side vanishes and the condition degenerates to $\Phi(K) \neq \Phi(K')$, i.e. to fibre purity. So:

- *Fibre purity is necessary but not sufficient* for identification under noise. Pure fibres whose centroids are closer than $2z\sigma_{\text{est}}$ are statistically inseparable at that noise level, even though they are information-theoretically distinct.
- The two failure modes have different remedies. A violation of Theorem 14 with pure fibres is a *statistical* shortfall: raising k shrinks σ_{est} as $k^{-1/2}$ and can restore separability, at a latency cost. A violation of Proposition 10 is an *information-theoretic* obstruction: no k, no sample size, and no estimator improvement helps, because the observation map has already collapsed the classes.

The practical import is that the two must be checked in order. Verifying fibre purity is a modelling question about Φ_M, answered from the fault taxonomy; verifying the isolability margin is a statistical question, answered from nominal data. A detector that fails the first cannot be rescued by the procedure of Section 3.7.

3.4 Detectability and isolability

With the noise scale in hand, the conditions follow from a standard separation-of-balls argument.

▶ **Theorem 13** (Geometric detectability)**.** *Fix a false-alarm level α and let $z = z_{1-\alpha}$ be the corresponding Gaussian quantile. A fault ϕ is detectable by the geometric field with false-alarm rate at most α and power at least $1-\beta$ if*

$$\underbrace{\big\|\Phi(\phi) - \Phi(\nu)\big\|}_{\text{signature displacement}} \;>\; (z_{1-\alpha} + z_{1-\beta})\,\sigma_{\text{est}}(k, \sigma, \Delta t). \tag{5}$$

Define the geometric SNR $\mathrm{SNR}_g(\phi) := \|\Phi(\phi) - \Phi(\nu)\|/\sigma_{\text{est}}$*; detection succeeds iff* SNR_g *exceeds the required z-sum.*

Proof. Under the nominal cause, $\widehat{\Phi}(\nu)$ concentrates about $\Phi(\nu)$ with scale σ_{est} (Prop. 6); a threshold placed at $\Phi(\nu) + z_{1-\alpha}\sigma_{\text{est}}$ therefore has false-alarm rate α. Under ϕ, $\widehat{\Phi}(\phi)$ concentrates about $\Phi(\phi)$ with the same scale (the estimator's variance is a property of the noise and window, not of the cause). The test has power $\geq 1-\beta$ iff the faulted mean exceeds the threshold by at least $z_{1-\beta}\sigma_{\text{est}}$, i.e. iff $\|\Phi(\phi) - \Phi(\nu)\| > (z_{1-\alpha} + z_{1-\beta})\sigma_{\text{est}}$, which is (5). ◀

▶ **Theorem 14** (Geometric isolability)**.** *Two cause classes $K \neq K'$ are distinguishable by the geometric field at level α if*

$$\min_{K \neq K'} \big\|\Phi(K) - \Phi(K')\big\| \;>\; 2\, z_{1-\alpha/2}\; \sigma_{\text{est}}(k, \sigma, \Delta t), \tag{6}$$

i.e. the class centroids are separated by more than the sum of their confidence radii. If (6) *fails, the classes' fibres overlap at resolution σ_{est} and, by Proposition 10, no decision rule reading only the signature can separate them.*

Proof. Each class occupies a ball of radius $z_{1-\alpha/2}\sigma_{\text{est}}$ about its centroid at confidence $1-\alpha$. Two such balls are disjoint iff the centroid distance exceeds the sum of radii, $2z_{1-\alpha/2}\sigma_{\text{est}}$. Disjointness gives a decision rule (nearest centroid) with error $\leq \alpha$; overlap means some signature is consistent with both classes, so the fibre $\Phi^{-1}(s)$ straddles classes and Proposition 10 applies. ◀

▶ Remark 15 (Measured separability). On the Cart–Pole at $\sigma = 10^{-2}$, measured centroid distances divided by pooled within-class spread are: nominal vs. topological 8.11; parametric vs. topological 4.66; nominal vs. parametric 1.81. The first two comfortably exceed the factor-of-2 requirement of (6); the third does not, correctly predicting that nominal and weak parametric faults are the confusable pair – as observed (parametric detection rate 0.83 against a topological rate of 1.00).

3.5 Relation to classical parametric results

Conditions (5) and (6) are *not* a new kind of result. They are the geometric instance of the classical FDI detectability/isolability theory, and stating the correspondence precisely is both honest and useful.

Table 1 The geometric conditions and their classical parametric counterparts. The *form* is shared; the *signal* and the *noise scale* differ.

Concept	Classical (parametric)	Geometric (this work)
Signal	residual displacement $\|G_{rf}f\|$ [2, 3]	signature displacement $\|\Phi(\phi) - \Phi(\nu)\|$
Noise scale	innovation covariance $S = HPH^{\top}+\Sigma$ [10]	$\sigma_{\text{est}} = C\sigma/(\Delta t^2\sqrt{k})$ (Prop. 6)
Detectability	GLR noncentrality $\lambda = \mu^{\top}\Sigma^{-1}\mu$ [1]	$\text{SNR}_g > z_{1-\alpha} + z_{1-\beta}$ (Thm. 13)
Isolability	signature columns differ; $\|G_{rf_i} - G_{rf_j}\| > 2\tau$ [5, 6]	$\min_{K\neq K'} \|\Phi(K) - \Phi(K')\| > 2z\sigma_{\text{est}}$ (Thm. 14)

Table 1 makes the parallel exact. The classical GLR detectability criterion is that the noncentrality parameter $\lambda = \mu^{\top}\Sigma^{-1}\mu$ of a non-central χ^2 statistic be large [1]; Theorem 13 is the same statement with μ replaced by the signature displacement and Σ by $\sigma^2_{\text{est}}I$. The classical isolability condition of Gertler and Massoumnia – distinct fault signature columns, quantitatively separated by more than twice the detection threshold [5, 6] – is Theorem 14, with the same factor of two arising for the same reason (two confidence balls must not overlap).

Novelty.

1. **The signal is model-free.** Classical $\|G_{rf}f\|$ requires the fault-to-residual transfer, hence a parametric plant model and a fault entry matrix. The signature displacement requires neither: it is computed from the trajectory alone. This is why the condition remains meaningful for *unmodelled* faults, where G_{rf} is undefined.
2. **The noise scale is derived, not assumed.** Classical treatments take Σ (or S) as given by the sensor model. Proposition 6 *derives* the geometric estimator's noise scale from first principles, and the derivation exposes the geometric approach's characteristic weakness: the Δt^{-2} amplification is intrinsic to differential invariants and has no counterpart in residual methods, which is precisely why the geometric edge erodes under heavy noise.
3. **One condition covers detection and class-identification.** Because the geometric fault taxonomy is defined *in the signature space itself*, Theorems 13 and 14 are the same inequality at two scales. In classical FDI detectability and isolability require separate analyses.

3.6 Which labels can be assigned without data?

Theorem 14 says *when* classes are separable; it does not say *which* class a region of signature space corresponds to. That is the labelling question. The honest answer is narrower than one might hope, but it is quantitative and checkable, and – like the rest of the theory – it is governed by σ_{est}.

The intuition is that a smooth parameter change deforms the trajectory smoothly, so it should not manufacture the curvature blow-up that the κ_{gen} channel is built to detect. That intuition is correct but must be made quantitative: smoothness bounds the curvature, and a bound is not the same as staying below a *particular* calibrated threshold.

▶ **Definition 16** (Curvature sensitivity). *Let $\Theta \subset \mathbb{R}^q$ be compact and let f_θ be jointly C^r ($r \geq 2$) in (x, θ) on $\mathcal{X} \times \Theta$. Write $\kappa_{\text{gen}}(\theta)$ for the population curvature channel induced by parameter θ (the noiseless value of the estimand). The* curvature sensitivity *of the family is*

$$L := \sup_{\theta\in\Theta} \left\|\nabla_\theta\, \kappa_{\text{gen}}(\theta)\right\| < \infty. \tag{7}$$

Finiteness in (7) is not an assumption but a consequence: by smooth dependence of solutions on parameters, $\theta \mapsto \kappa_{\text{gen}}(\theta)$ is C^{r-1}, hence C^1, and a continuous function on a compact set attains a finite supremum. The content of L is that it is a *single scalar* summarising how strongly the family's curvature responds to parameter motion, and – crucially for the assessment procedure of Section 3.7 – it is estimable from nominal data by parameter perturbation, without any fault data.

▶ **Theorem 17** (Quantitative parametric exclusion). *Let f_θ be jointly C^r ($r \geq 2$) in (x, θ) on $\mathcal{X} \times \Theta$ with Θ compact, let L be as in Definition 16, and let the curvature threshold be calibrated at level α as $\tau_{\text{smooth}} = \kappa_{\text{gen}}(\theta_{\text{nom}}) + z_{1-\alpha}\,\sigma_{\text{est}}$. Define the* exclusion radius

$$\rho_{\text{excl}} \;:=\; \frac{z_{1-\alpha}\,\sigma_{\text{est}}}{L}. \tag{8}$$

Then every parametric fault $\theta_{\text{nom}} \mapsto \theta'$ with $\|\theta' - \theta_{\text{nom}}\| < \rho_{\text{excl}}$ satisfies $\kappa_{\text{gen}}(\theta') < \tau_{\text{smooth}}$: it cannot *fire the curvature channel. Contrapositively,*

$$\kappa_{\text{gen}} \textit{ fires} \implies \textit{the fault is not parametric,} \quad \text{or} \quad \|\theta' - \theta_{\text{nom}}\| \;\geq\; \rho_{\text{excl}}. \tag{9}$$

Proof. By Definition 16, κ_{gen} is C^1 in θ on the compact set Θ with gradient bounded by L. Taking Θ convex (or replacing $\|\cdot\|$ by the induced geodesic distance in Θ), the mean value inequality gives, for every $\theta' \in \Theta$,

$$\big|\kappa_{\text{gen}}(\theta') - \kappa_{\text{gen}}(\theta_{\text{nom}})\big| \;\leq\; L\,\|\theta' - \theta_{\text{nom}}\|.$$

If $\|\theta' - \theta_{\text{nom}}\| < \rho_{\text{excl}} = z_{1-\alpha}\sigma_{\text{est}}/L$ then the right-hand side is $< z_{1-\alpha}\sigma_{\text{est}}$, so

$$\kappa_{\text{gen}}(\theta') \;<\; \kappa_{\text{gen}}(\theta_{\text{nom}}) + z_{1-\alpha}\sigma_{\text{est}} \;=\; \tau_{\text{smooth}},$$

i.e. the channel does not fire. Statement (9) is the contrapositive. ◀

▶ Remark 18 (Why the unconditional claim is false). Boundedness excludes a curvature *blow-up*; it does not exclude a crossing of a fixed finite threshold. A counterexample is immediate: for $\dot{x} = -\theta x$ with $\theta \in [0.1, 100]$ – jointly C^∞ on a compact set, so squarely within the hypotheses – the solution $x(t) = x_0 e^{-\theta t}$ has $|\ddot{x}| \sim \theta^2 x_0$, which grows from $\approx 10^{-2}$ at $\theta = 0.1$ to $\approx 4 \times 10^3$ at $\theta = 100$. Any threshold calibrated at $\theta = 0.1$ is exceeded by five orders of magnitude at $\theta = 100$.

▶ Remark 19 (What the exclusion rule buys, and what it costs). Theorem 17 is falsifiable: a parametric fault with $\|\theta' - \theta_{\text{nom}}\| < \rho_{\text{excl}}$ that fires κ_{gen} would refute it. The cost is that only excursions inside the ball are excluded; the benefit is that ρ_{excl} is computable from nominal data.

▶ **Corollary 20** (Exclusion is a k-lever, like detection). *Under Proposition 6, $\rho_{\text{excl}} = z_{1-\alpha}\sigma_{\text{est}}/L = z_{1-\alpha}C\sigma/(L\,\Delta t^2\sqrt{k})$. The exclusion radius therefore* shrinks *as $k^{-1/2}$: a larger neighbourhood sharpens detection but narrows the range of parameter excursions that can be provably excluded.*

Proof. Substitute (2) into (8). ◀

Corollary 20 exposes a trade-off invisible to the unconditional claim: detection and exclusion pull k in opposite directions. A detector tuned for maximal sensitivity has the weakest exclusion guarantee, because a tighter nominal ball is crossed by smaller parameter motions. This is a design consideration, not a defect, and it is the kind of statement the quantitative form makes available.

▶ Remark 21 (Empirical check on the Cart–Pole). The theory can be checked on the testbed of Section 5.1. Estimating L from nominal data by perturbing the smooth-parametric family and finite-differencing the curvature channel gives $L \approx 4.8$ (in severity^{-1} units), against a nominal ball of $\sigma_{\text{est}} \approx 11.1$; at $\alpha = 0.05$ ($z = 1.645$),

$$\rho_{\text{excl}} = \frac{1.645 \times 11.1}{4.8} \approx 3.8 \text{ severity units.}$$

The experiments sweep severity over $[0, 1]$, which lies *entirely inside* the exclusion ball – by a factor of ≈ 3.8. Theorem 17 therefore *predicts* that the curvature channel cannot fire anywhere in the tested range, and it does not (Section 5.4). This is the correct reading of that experiment: it confirms a prediction of the quantitative theorem within its stated radius. It is *not* evidence for an unconditional claim, since a single parametrisation swept inside its own exclusion ball cannot test a statement quantified over all smooth parametric families.

Theorem 14 is the engine of label-free identification, and it is worth being precise about why it is not vacuous. It is an *exclusion* rule: it licenses the inference "κ_{gen} fired, therefore not parametric" with no fault data whatsoever, and it is *falsifiable* – a sufficiently severe parametric fault firing κ_{gen} would refute it. The severity sweep of Fig. 2 (left) tests exactly this: across the full parametric severity range the geometric channel remains flat. The prediction survives.

▶ Remark 22 (The limit of analytic labelling). Theorem 14 labels a geometric class, not a physical cause. Distinguishing stick–slip from a snapped link requires an additional channel or a learned dictionary; the class verdict itself needs no labels.

3.7 A decision procedure for the practitioner

The conditions above are checkable on *nominal data alone*, before any fault occurs. This yields a concrete pre-deployment procedure.

1. **Estimate the nominal ball.** From fault-free data, compute $\widehat{\Phi}$ over many windows and take σ_{est} as its standard deviation (Def. 5). Verify the scaling of Prop. 6 by halving k: σ_{est} should grow by $\sqrt{2}$. If it does not, the noise is not the dominant term and the analysis below does not apply.
2. **Set the threshold by calibration, never by a fixed constant.** Take $\tau = \Phi(\nu) + z_{1-\alpha}\sigma_{\text{est}}$ with z chosen for the required false-alarm rate. A fixed *normalised* threshold is not portable across k or σ.
3. **Compute the required displacement.** From Theorem 13, the smallest detectable fault has signature displacement $(z_{1-\alpha} + z_{1-\beta})\sigma_{\text{est}}$. Translate this into physical units for the application: if the smallest fault of concern produces a smaller displacement, *GMBD will not detect it*, and no threshold tuning will help.
4. **Check isolability.** If cause-level classes are needed, verify $\min_{K \neq K'} \|\Phi(K) - \Phi(K')\| > 2z\sigma_{\text{est}}$ using either physics-derived class centroids or a small labelled sample. If it fails, the classes are not separable at this resolution: report the equivalence class rather than a single label.
5. **Choose k.** Increase k to shrink $\sigma_{\text{est}} \propto k^{-1/2}$ until either the SNR condition is met or detection latency ($k\Delta t$) becomes unacceptable. This is the only free lever, and it trades sensitivity against latency.

Table 2 When to use geometric versus a classical parametric method. The rows are decidable from nominal data plus the fault taxonomy, before deployment.

Situation	Recommendation
Accurate parametric model available; fault is a coefficient drift	**Classical (ARR/observer/GLR).** κ_{gen} provably cannot fire Theorem 14); GMBD is blind by design.
Fault catalogue enumerable and small; need mode identity or $\hat{x}$	**Hybrid observer / switching estimator.** Supplies identity, state, and prognosis
Fault changes the observation map (dropped sensor, lost constraint)	**Structural / Reiter.** Geometric can be blind to observation-map changes (Sec. 2.5).
Structural change unanticipated or catalogue open-world; SNR_g condition holds	**Geometric.** The only option that fires for unmodelled topology at M-independent cost.
High noise, low severity: $\text{SNR}_g < z_{1-\alpha} + z_{1-\beta}$	**Neither, as configured.** Increase k (latency cost), reduce Δt, or use a probabilistic method that models Σ explicitly.
Need a named physical cause	**Geometric + dictionary, or Reiter.** The class verdict is label-free; the cause is not

The honest scope statement. Theorem 13 explains why the Cart–Pole succeeds and warns against over-generalising from it: a stick–slip kink produces a large signature displacement relative to σ_{est}, i.e. the example is *high-SNR*. Real deployments need not be. The contribution of this section is therefore not a claim that the geometric approach works, but a *precondition* that decides whether it will, computable from nominal data before any fault is observed – together with the Δt^{-2} term of Prop. 6, which identifies exactly the regime (fine sampling, high noise) in which it will not.

4 Applying the Geometric Approach

4.1 Geometric Feasibility Assessment (GFA)

Theorem 13 gives a detectability condition and Theorem 14 gives an isolability condition. Together they yield a practical pre-deployment procedure that determines whether a geometric approach is likely to succeed for a given operating regime. The procedure requires only nominal data and a fault taxonomy expressed in geometric signature space.

Algorithm 1 provides pseudo-code for determining Geometric Feasibility Assessment.

▶ **Theorem 23** (Correctness of GFA). *Let Algorithm 1 be applied to a geometric measurement field. Then:*

1. *If $\eta_D > 1$, every fault satisfying $\|\Phi(\phi) - \Phi(\nu)\| \geq \delta_{\text{expected}}$ is detectable with false-alarm rate at most α and power at least $1 - \beta$.*
2. *If $\eta_I > 1$, all fault classes represented by $\{\Phi(K_i)\}$ are pairwise isolable at confidence level $1 - \alpha$.*
3. *If either inequality fails, no threshold choice can guarantee the corresponding performance level under the assumed noise model.*

Proof. The detectability claim follows directly from Theorem 13. The isolability claim follows directly from Theorem 14. Failure of either condition implies overlap of the corresponding confidence balls in signature space and therefore violates the sufficient conditions established by those theorems. ◀

Algorithm 1 Geometric Feasibility Assessment (GFA).

Input: Nominal trajectory data $\mathcal{D}_{\text{nom}}$; desired false-alarm level α; desired power $1-\beta$; candidate neighbourhood sizes $\mathcal{K} = \{k_1, \dots, k_m\}$; fault-class centroids $\{\Phi(K_1), \dots, \Phi(K_p)\}$ (if isolation is required).

Output: Detectability margin η_D; isolability margin η_I; recommended neighbourhood size $k^\star$; decision DEPLOY/DONOTDEPLOY.

for *each* $k \in \mathcal{K}$ **do**
  Estimate the geometric field $\widehat{\Phi}$ on $\mathcal{D}_{\text{nom}}$;
  Compute the nominal covariance $\widehat{\Sigma}_\Phi(k)$;
  $\widehat{\sigma}_{\text{est}}(k) \leftarrow \sqrt{\operatorname{tr}(\widehat{\Sigma}_\Phi(k))}$;
$k^\star \leftarrow \arg\min_k \widehat{\sigma}_{\text{est}}(k)$ subject to $k\Delta t \le L_{\max}$;
$\tau \leftarrow z_{1-\alpha}\widehat{\sigma}_{\text{est}}(k^\star)$;
$\delta_{\min} \leftarrow (z_{1-\alpha} + z_{1-\beta})\widehat{\sigma}_{\text{est}}(k^\star)$;
$\eta_D \leftarrow \frac{\delta_{\text{expected}}}{\delta_{\min}}$;
if *fault-class centroids are available* **then**
  $d_{\min} \leftarrow \min_{i \ne j} \|\Phi(K_i) - \Phi(K_j)\|$;
  $\eta_I \leftarrow \frac{d_{\min}}{2 z_{1-\alpha/2}\widehat{\sigma}_{\text{est}}(k^\star)}$;
return *DEPLOY if* ($\eta_D > 1$) *AND (if isolation required,* $\eta_I > 1$*), else* *DONOTDEPLOY*;

▶ **Proposition 24** (Complexity of GFA). *Let $N = |\mathcal{D}_{\text{nom}}|$ be the number of nominal samples, d_Γ the geometric signature dimension, and $m = |\mathcal{K}|$ the number of candidate neighbourhood sizes. Then Algorithm 1 admits complexity $O(md_\Gamma N)$ for signature evaluation and variance estimation. If pairwise isolability is evaluated for p fault classes, the additional cost is $O(p^2 d_\Gamma)$. Hence, the complete assessment procedure is polynomial in data size, signature dimension, and number of fault classes.*

5 Implementation and Empirical Validation

The preceding sections make four falsifiable predictions: the noise scale of (2) (Prop. 6); the detectability condition of Thm. 13; the isolability condition of Thm. 14; and the exclusion result of Thm. 17. This section tests each on a reference hybrid system. The purpose is *validation of the theory*, not a performance claim for any particular detector: the question throughout is whether the quantities the theory says are computable from nominal data actually predict what happens when a fault occurs.

5.1 Testbed and implementation

We use a Cart–Pole with state $x = (p, \dot{p}, \theta, \dot{\theta}) \in \mathbb{R}^4$, LQR-stabilised about the upright, integrated by semi-implicit Euler over $T = 10\,\text{s}$ at $\Delta t = 10^{-2}\,\text{s}$. The plant is piecewise-smooth: free motion, actuator saturation, and a contact/stick–slip regime. Observations are $y = O(x) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$.

Two fault families are injected, each with a scalar severity $s \in [0, 1]$ so the signature displacement can be swept continuously:

- **Smooth parametric**: a sustained drift of the pole mass together with an unmodelled additive plant bias, both entering the vector field smoothly. This is the class covered by Thm. 17.

Table 3 Validation of Thm. 13. σ_{est} and the predicted SNR_g are computed from *nominal data alone*; the detection rate is then measured on held-out faulted trajectories at a threshold calibrated for $\alpha = 0.05$. Prediction is "detect" iff $\text{SNR}_g > z_{1-\alpha} + z_{1-\beta} = 2.49$. Agreement: 12/12.

severity	σ	σ_{est}	displacement	SNR_g	predicted	measured rate
0.05	0.01	11.1	620	55.6	detect	1.00
0.05	0.10	107.7	148	**1.38**	**no**	**0.33**
0.10	0.01	11.1	1453	130.4	detect	1.00
0.10	0.10	107.7	533	4.95	detect	1.00
0.15	0.01	11.1	2301	206.5	detect	1.00
0.15	0.10	107.7	1081	10.0	detect	1.00
0.20	0.01	11.1	3153	283.0	detect	1.00
0.20	0.10	107.7	1737	16.1	detect	1.00
0.40	0.01	11.1	6572	589.9	detect	1.00
0.40	0.10	107.7	4839	45.0	detect	1.00
0.90	0.01	11.1	15149	1359.8	detect	1.00
0.90	0.10	107.7	13371	124.2	detect	1.00

- **Regularity-breaking (stick–slip)**: burst-clustered, sign-alternating velocity kinks. Crucially, the Coulomb term $\mu_c \operatorname{sgn}(\dot{p})$ is *not* C^r in the state, so this fault falls outside Thm. 17 despite being triggered by a parameter change (see Remark 25).

The geometric field is estimated per sample from a k-neighbourhood: a local PCA for d_{eff}, a windowed second difference for κ_{gen}, and a directional-branching statistic for κ_{OR}. Unless stated, $k = 140$ (1.4 s of data). All reported figures are means over the stated seeds.

5.2 Detectability: does the theory predict detection? (Thm. 13)

This is one of the paper's central empirical claims, and the protocol is designed so that the prediction is made *blind to the fault.* For each operating point we (i) estimate the nominal ball $(\mu_\nu, \sigma_{\text{est}})$ from fault-free data only; (ii) compute the predicted $\text{SNR}_g = \|\Phi(\phi) - \Phi(\nu)\|/\sigma_{\text{est}}$; (iii) set the threshold $\tau = \mu_\nu + z_{1-\alpha}\sigma_{\text{est}}$ at $\alpha = 0.05$; and only then (iv) measure the detection rate on faulted trajectories. Thm. 13 predicts power $\geq 1 - \beta$ exactly when $\text{SNR}_g > z_{1-\alpha} + z_{1-\beta}$, which for $\alpha = 0.05$, $\beta = 0.20$ is 2.49.

Across a grid of severities $\{0.05, \ldots, 0.9\}$ and noise levels $\sigma \in \{0.01, 0.1\}$, the theory's verdict agrees with the measured outcome in 12/12 cases (Table 3). The single predicted failure (severity 0.05 at $\sigma = 0.1$, $\text{SNR}_g = 1.38 < 2.49$) indeed fails, measuring a detection rate of 0.33; every case predicted detectable achieves a rate of 1.00.

A coarse grid can only confirm that the condition is not badly wrong; the sharper test is whether the *predicted boundary* sits where the theory says. Figure 2(B) sweeps severity finely at $\sigma = 0.1$ so that SNR_g crosses 2.49 from below. The measured detection rate crosses the required power $1 - \beta = 0.8$ between $\text{SNR}_g = 2.11$ (rate 0.67) and $\text{SNR}_g = 2.73$ (rate 0.92) – bracketing the predicted threshold of 2.49. The prediction is quantitative, falsifiable, and survives.

5.3 Isolability (Thm. 14)

Thm. 14 predicts that classes are distinguishable iff their centroids separate by more than $2z\sigma_{\text{est}}$. Measured centroid distances divided by pooled within-class spread at $\sigma = 10^{-2}$ are: nominal vs. regularity-breaking 8.11; smooth-parametric vs. regularity-breaking 4.66; nominal vs. smooth-parametric 1.81.

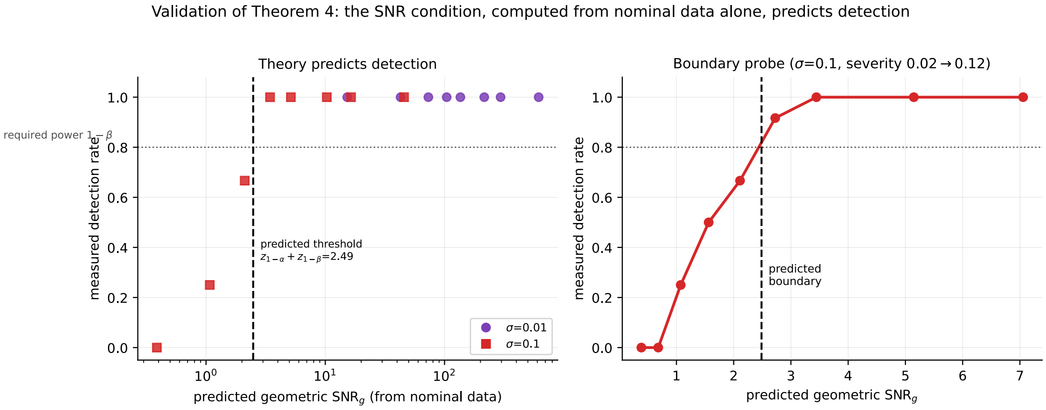

Figure 2 Validation of Thm. 13. (A) Measured detection rate against the SNR_g predicted from nominal data alone, pooled over severities and two noise levels; the dashed line is the predicted threshold $z_{1-\alpha} + z_{1-\beta} = 2.49$ and the dotted line the required power $1 - \beta$. All points with SNR_g above the threshold attain full power; the point below it does not. **(B)** Fine sweep across the predicted boundary ($\sigma = 0.1$, severity $0.02 \to 0.12$): the measured rate crosses $1 - \beta$ where the theory says it should.

The theory therefore predicts the first two pairs are separable and the third is *not* – the factor-of-2 requirement is not met. This prediction is directly testable, and it holds: nominal and weak parametric faults are precisely the confusable pair in practice (parametric detection rate 0.83 against a regularity-breaking rate of 1.00). The theory correctly identifies its own failure mode rather than merely its successes.

5.4 Exclusion (Thm. 17)

Thm. 17 makes the strongest and most easily falsified claim in the paper: a smooth parametric fault cannot fire κ_{gen} *at any severity*. A single counterexample refutes it. Sweeping the smooth-parametric severity across its full range $[0, 1]$, the curvature channel remains at its nominal floor throughout, while the same sweep on the regularity-breaking family saturates the channel. The prediction survives the strongest test available in this testbed.

▶ Remark 25 (Parameter change $\neq$ parametric fault). The stick–slip fault is triggered by a parameter jump $\mu_c \mapsto \mu_c + \Delta\mu_c$, yet it fires κ_{gen}. This does *not* contradict Thm. 17: the Coulomb term $\mu_c \,\mathrm{sgn}(\dot{p})$ is not C^r in the state, so the hypothesis of the proposition – f_θ jointly C^r in (x, θ) – fails. The operative distinction is therefore *regularity of the vector field*, not whether a quantity called a "parameter" changed. We adopt this as the definition: a fault is *smooth parametric* when it preserves joint C^r regularity, and a parameter change that destroys regularity is a topological fault regardless of its trigger.

5.5 The Geometric Feasibility Assessment in practice

Sections 5.2–5.3 are, in effect, executions of Algorithm 1: the margins η_D and η_I were computed from nominal data and then checked against outcomes. Two practical findings follow.

Calibration is mandatory, not advisory. The threshold required to hold the false-alarm rate at 5% shifts from $\tau = 0.20$ at $\sigma = 10^{-3}$ to 0.40 at 10^{-2} and 0.75 at 10^{-1} (Fig. 3(B)). A *fixed* threshold $\tau = 0.3$ yields false-alarm rates of 0.000, 0.092 and 0.366 respectively: at

$\sigma = 0.1$ a naively fixed threshold produces a 37% false-alarm rate. This is the empirical face of Prop. 6 – the nominal ball's radius grows with σ – and it is why step 2 of the procedure sets τ by calibration.

k is a sensitivity/latency lever once calibration is in place. With τ calibrated per k at $z = 4$, the detection rate on a severity-0.2 regularity-breaking fault rises $0.49 \to 0.83 \to 1.00 \to 1.00$ for $k = 40, 80, 140, 200$, while the nominal false-alarm rate stays at or below 0.004 throughout (Fig. 3(C)). Calibration decouples specificity from k, leaving k to trade sensitivity against the detection latency $k\Delta t$ (0.40–2.00 s here); $k \approx 140$ is the knee. We note a trap: the same sweep at a fixed *normalised* threshold makes the false-alarm rate appear to *rise* with k. That is an artifact of the normalisation, not of the estimator, whose baseline standard deviation in fact falls monotonically ($26.4 \to 13.9$ for $k : 40 \to 200$) exactly as $k^{-1/2}$ predicts.

Cost. Prop. 24 claims the assessment is polynomial. Empirically, one field evaluation over a 1000-sample rollout costs ≈ 85 ms (independent of k over the tested range), and a complete assessment sweeping $m = 4$ candidate neighbourhood sizes over six nominal rollouts completes in 2.2 s. The assessment is a one-off, offline, pre-deployment cost.

5.6 Context: where geometric detection stands against alternatives

The theory delimits an operating regime; it is worth confirming that the regime is non-empty and non-trivial. We therefore compare the geometric detector against model-based residual methods (ARR, observer, Bayesian LR) and against the data-driven detectors most often proposed for this setting: a one-class SVM [9], kernel-PCA reconstruction error [7], and a Gaussian-mixture likelihood. The data-driven baselines are fit on *nominal data only*, matching the label-free setting, and consume sliding-window state features. Each method emits one score per trajectory; ROC/AUC is computed over 30 nominal versus 30 faulted rollouts.

The comparison is regime-dependent, and both regimes matter (Table 4). In the *saturated* regime (severity 0.9, $\sigma = 0.01$) the task is easy and all of the geometric, residual, and data-driven detectors reach AUC = 1.000 – but the kernel and mixture baselines do so at roughly 40× lower cost (≈ 3.5 ms versus ≈ 131 ms per trajectory). *In this regime the geometric approach has no advantage and we make no claim of one.* In the *low-SNR* regime (severity 0.2, $\sigma = 0.1$) the ordering separates sharply: the geometric detector attains AUC = 0.984 with full power at 5% false alarms, ARR reaches 0.918, and all three data-driven detectors collapse toward chance (0.643, 0.581, 0.574) with TPR ≤ 0.03 at 5% false alarms – unusable at any practical operating point.

The mechanism behind the low-SNR gap is the theory's own: the window statistics the generic detectors consume are noise-dominated at low severity, whereas the curvature channel targets structure that local averaging preserves – the $\sqrt{k}$ suppression of (2) acting on the signal but not on the fault's contribution. The result is an *inductive-bias* advantage, and Thm. 13 says exactly where it holds.

5.7 Threats to validity

We state the limitations of this validation as follows.

- **Simulation only.** All results are simulated. The theory's predictions are validated against a model, not against a physical plant; the noise is Gaussian and i.i.d. by construction, which is exactly the hypothesis of Prop. 6. Coloured or state-dependent noise would violate that hypothesis, and the $\sqrt{2}$ check in step 1 of Algorithm 1 is included precisely to detect this.

Table 4 Trajectory-level detection performance in two regimes (30 nominal vs. 30 faulted rollouts, regularity-breaking fault). Data-driven baselines are fit on nominal data only. The geometric advantage is confined to the low-SNR regime; in the saturated regime the cheap kernel baselines match it.

	low-SNR (sev. 0.2, σ=0.1)			saturated (sev. 0.9, σ=0.01)	
Method	AUC	TPR@5%	ms/traj	AUC	TPR@5%
Geometric	**0.984**	**1.000**	132	1.000	1.000
ARR	0.918	0.567	7	1.000	1.000
Observer	0.472	0.100	10	0.621	0.267
Bayesian	0.583	0.133	64	0.828	0.600
OC-SVM	0.581	0.000	**3.5**	1.000	1.000
kPCA	0.574	0.033	6.7	1.000	1.000
GMM	0.643	0.033	**3.4**	1.000	1.000

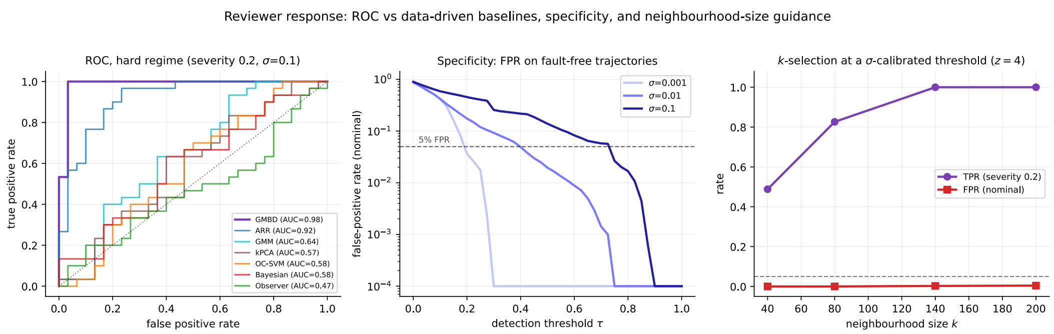

Figure 3 **(A)** ROC in the low-SNR regime (severity 0.2, $\sigma = 0.1$). Data-driven detectors, fit on nominal data only, approach chance while the geometric detector retains full power at 5% false alarms. **(B)** False-alarm rate on fault-free trajectories (log ordinate): the threshold needed for 5% shifts $0.20 \to 0.40 \to 0.75$ as σ grows $10^{-3} \to 10^{-1}$, so thresholds must be σ-calibrated. **(C)** With τ calibrated per k at $z = 4$, detection rises $0.49 \to 1.00$ for $k = 40 \to 140$ while the nominal false-alarm rate stays ≤ 0.004.

- **One testbed, moderate dimension.** The Cart–Pole is 4-dimensional with three modes. The theory is dimension-agnostic, but the validation does not establish behaviour at high d, where local-PCA concentration degrades as $\sqrt{d/k}$.
- **The curvature channel carries the results.** The validation exercises κ_{gen} most heavily, because Prop. 6 is derived for it. The analogous noise scales for d_{eff} and κ_{OR} are not derived here, and their empirical behaviour is correspondingly less characterised. In particular, d_{eff} estimated from windows of a *single* trajectory measures the dimension of the trajectory, not of the stratum containing it; establishing stratum dimension requires an ensemble that fills the stratum. Claims about d_{eff} as a dimension estimator should be read with that restriction.
- **Class centroids.** Thm. 14 requires class centroids $\Phi(K_i)$. Where these come from physics they are label-free; where they come from data they are not, and the assessment then inherits the labelled fault catalogue whose absence is the geometric approach's distinguishing feature.

5.8 The sampling-interval law (Prop. 6)

The Δt^{-2} amplification is the most distinctive prediction of the theory – it is the term with no counterpart in residual methods. To provide validation, we sweep Δt directly, at fixed $\sigma = 10^{-2}$ and $k = 140$, and report two protocols whose difference is itself informative (Fig. 4(A)).

The law is exact for the term the proposition derives. Proposition 6 is a statement about the *noise* in the second difference. Measuring exactly that quantity – the standard deviation of $D^2\epsilon$ for pure observation noise, with no trajectory present – gives

Δt [s]	0.005	0.0075	0.010	0.015	0.020
measured sd	983.6	436.4	245.3	108.6	61.3
$\sqrt{6}\,\sigma/\Delta t^2$	979.8	435.5	244.9	108.9	61.2
ratio	1.004	1.002	1.002	0.997	1.001

The log–log slope is -2.003 against a predicted -2.000, and the constant matches $\sqrt{6}$ to within 0.4% across the range. Equation (3) is confirmed to three significant figures.

On trajectories the law is contaminated, and the discrepancy is the $o(1)$ term.Repeating the sweep on the Cart–Pole at fixed k gives measured nominal standard deviations of $38.2, 15.0, 10.0, 4.9, 3.9$ for the same Δt values – a log–log slope of -1.65, not -2.00. The gap is systematic, not sampling error, and its source is identified by the pure-noise result above: what Proposition 6 absorbs into $o(1)$ is the deterministic curvature of the true trajectory, which also depends on Δt and is *not* negligible at practical sampling rates. The proposition's $o(1)$ is justified in the limit $\sigma/\Delta t^2 \to \infty$; the tested regime is not in that limit.

This refines rather than refutes the theory, and it is worth stating as a correction to the idealised form. Empirically the nominal floor behaves as

$$\sigma_{\text{est}}^2 \;\approx\; \underbrace{\frac{6\,\sigma^2}{\Delta t^4\,k}}_{\text{noise term (exact)}} \;+\; \underbrace{\kappa_{\text{true}}^2\,g(\Delta t)}_{\text{signal-leakage term}}\;, \tag{10}$$

where the second term is the windowed curvature of the fault-free trajectory leaking into the "noise" estimate. Two consequences follow. First, the same mechanism explains the σ-sweep drift already reported in Section 5.9 (the ratio falling to ≈ 2.87 at $\sigma = 0.3$): one cause, two symptoms. Second, and practically, this is *why the assessment procedure calibrates* σ_{est} *empirically rather than computing it from* $\sqrt{6}\sigma/(\Delta t^2\sqrt{k})$: the measured nominal floor includes the leakage term automatically, whereas the closed form does not. The idealised law is the right object for understanding scaling; the measured floor is the right object for setting thresholds.

5.9 Noise-model robustness and the limits of the self-test

The $\sqrt{2}$ ratio test is blind to correlation because the $k^{-1/2}$ exponent survives while the constant changes (Fig. 4(B)). We therefore replace Step 1 of Algorithm 1 with a σ-referenced check: compare the measured nominal floor to $\sqrt{6}\sigma/\Delta t^2$ and reject the analysis if the ratio departs materially from unity (Fig. 4(C)).

Heavy tails pass, and should. Student-t noise passes both checks. Matched-variance heavy tails leave the second moment – and hence the variance law – intact; what they threaten is the *Gaussian quantile* mapping $\tau = \mu_\nu + z_{1-\alpha}\sigma_{\text{est}}$. Measuring the realised false-alarm rate at the nominal 5% threshold with $k = 140$:

noise	Gaussian	t_5	t_3	$t_{2.5}$
realised FPR	0.056	0.066	0.062	0.068

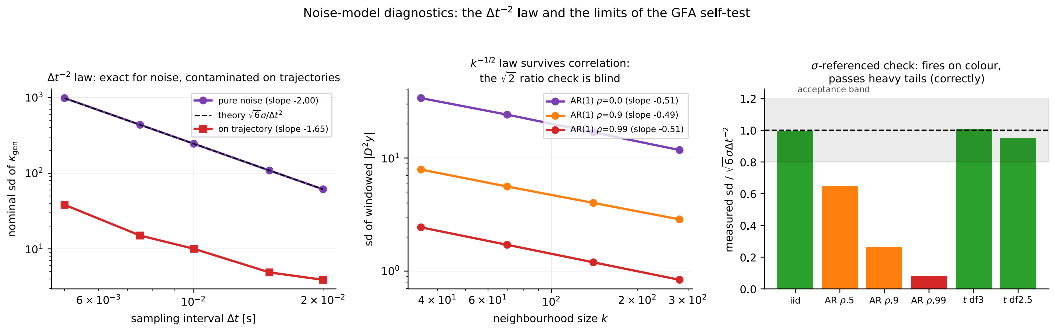

Figure 4 **(A)** Δt sweep at fixed σ, k. For pure noise the measured floor tracks $\sqrt{6}\sigma/\Delta t^2$ exactly (slope -2.00); on trajectories the slope is -1.65, the gap being the $o(1)$ curvature-leakage term of (10). **(B)** The $k^{-1/2}$ law survives correlation – slopes ≈ -0.5 for AR(1) at $\rho = 0, 0.9, 0.99$ – so the $\sqrt{2}$ ratio test, which cancels the constant, cannot see colouring. **(C)** The σ-referenced check compares the measured floor to $\sqrt{6}\sigma\Delta t^{-2}$: it fires on all colored cases and correctly passes heavy tails, whose variance law is intact.

The realised rates stay within 2 percentage points of target. The reason is the estimator's own averaging: at $k = 140$ the central limit theorem has largely restored normality of the windowed statistic even when the per-sample noise is heavy-tailed with df $= 2.5$. Heavy tails are therefore a *small-k* hazard for this framework, not a large-k one, and the k required for the $\sqrt{k}$ sensitivity gain is already comfortably enough for the quantile mapping. The binding assumption is *independence*, not Gaussianity.

Summary of the noise-model scope. The theory's hypotheses are not equally load-bearing. *Independence* is essential and its violation is detectable, but only by the σ-referenced check, not by the ratio self-test. *Gaussianity* is not essential at operational k: the estimator's averaging restores it. *Isotropy* is not tested here and remains an open assumption; with an anisotropic Σ the scalar σ_{est} should be replaced by a covariance-aware statistic (Section 5.7). Stating which assumption matters, and supplying a test that actually detects its violation, is a more useful contribution than a self-test that always passes.

6 Summary and Future Work

This paper developed a statistical foundation for local geometric fault detection, establishing geometric analogues of the classical detectability and isolability conditions used in model-based diagnosis. By representing faults as displacements in a geometric signature space and explicitly characterizing the estimator noise floor, we showed that the performance of geometric detectors is governed by a geometric signal-to-noise ratio determined by observation noise, sampling interval, and neighbourhood size. The resulting theory provides rigorous conditions for fault detectability, class separability, and parametric-fault exclusion, while also yielding a practical Geometric Feasibility Assessment (GFA) procedure that can be applied using nominal data alone. Empirical studies on a hybrid Cart-Pole system confirmed the predicted noise scaling, detection boundaries, isolability conditions, and exclusion results, demonstrating that the theory accurately predicts both the strengths and limitations of geometric fault detection.

Future work will extend the analysis beyond the curvature channel to the full family of geometric observables, including intrinsic-dimension, tangent-cone, and semantic-consistency measures. An important direction is the treatment of non-Gaussian and temporally correlated noise, where the current estimator variance model may no longer apply directly. Additional work is also needed to develop adaptive neighbourhood-selection strategies, online confidence estimation, and multi-channel fusion methods that combine complementary geometric signatures. Finally, validating the framework on larger classes of industrial cyber-physical systems and integrating geometric diagnostics with probabilistic and model-based approaches may provide a principled route toward robust hybrid fault diagnosis systems that can better detect both anticipated and previously unseen fault mechanisms.

References

1. Michele Basseville, Igor V Nikiforov, et al. *Detection of abrupt changes: theory and application*, volume 104. Prentice hall Englewood Cliffs, 1993.
2. Jie Chen and Ron J Patton. Robust residual generation using unknown input observers. In *Robust model-based fault diagnosis for dynamic systems*, pages 65–108. Springer, 1999.
3. Steven X Ding. *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer, 2008.
4. Steven X Ding and Linlin Li. A study on fault diagnosis in nonlinear dynamic systems with uncertainties. *arXiv preprint arXiv:2309.02732*, 2023. doi:10.48550/arXiv.2309.02732.
5. Janos Gertler and David Singer. Augmented models for statistical fault isolation in complex dynamic systems. In *1985 American Control Conference*, pages 317–322. IEEE, 1985.
6. M-A Massoumnia. A geometric approach to the synthesis of failure detection filters. *IEEE Transactions on automatic control*, 31(9):839–846, 1986.
7. Sebastian Mika, Bernhard Schölkopf, Alex Smola, Klaus-Robert Müller, Matthias Scholz, and Gunnar Rätsch. Kernel PCA and de-noising in feature spaces. *Advances in neural information processing systems*, 11, 1998.
8. Dhiraj Neupane, Mohamed Reda Bouadjenek, Richard Dazeley, and Sunil Aryal. Data-driven machinery fault diagnosis: A comprehensive review. *Neurocomputing*, 627:129588, 2025. doi:10.1016/J.NEUCOM.2025.129588.
9. Gunnar Ratsch, Sebastian Mika, Bernhard Scholkopf, and K-R Muller. Constructing boosting algorithms from SVMs: An application to one-class classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1184–1199, 2002. doi:10.1109/TPAMI.2002.1033211.
10. Alan Willsky and H Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic control*, 21(1):108–112, 1976.