Quotient Geometry of Statistical Manifolds Under Dimensional Collapse

Ian Todd

Sydney Medical School, University of Sydney

`itod2305@uni.sydney.edu.au`

Abstract

We develop the quotient geometry of statistical manifolds under collapse maps—smooth maps $\pi : \mathcal{M} \to \mathbb{R}^k$ with $k < \dim(\mathcal{M})$. The main results are: (1) a *fiber structure theorem* showing that the vertical distribution $\mathcal{V} = \ker(d\pi)$ foliates $\mathcal{M}$ into submanifolds along which the Fisher metric becomes degenerate, with points on the same fiber being statistically non-identifiable in the sense of Watanabe; (2) a *quotient metric theorem* establishing when the Fisher metric descends to a well-defined Riemannian metric on the quotient $\mathcal{M}/\sim_\pi$, and characterizing obstructions in terms of α-connection geometry; and (3) *covering number bounds* relating the metric geometry of the quotient to the number of distinguishable equivalence classes at finite resolution. We show that a recent result on minimal embedding dimension for recurrent processes is a special case of this framework. The theory provides a unified geometric treatment of dimensional reduction, sufficient statistics, and singularities in statistical models.

Keywords: information geometry, quotient manifolds, Fisher metric, dimensional reduction, singular statistical models

1 Introduction

When a statistical manifold is observed through a lower-dimensional map, what geometric structure survives? This paper develops the quotient geometry of statistical manifolds under *collapse maps*—smooth maps $\pi : \mathcal{M} \to \mathbb{R}^k$ with $k < \dim(\mathcal{M})$ that reduce dimension while preserving some statistical content.

The question connects to several classical themes in information geometry:

- **Sufficient statistics:** A sufficient statistic $T : \mathcal{X} \to \mathbb{R}^k$ induces a collapse map on the parameter space. The quotient structure encodes which parameters are identifiable from T.

- **Singular models:** When parameters are non-identifiable, the Fisher information matrix becomes singular. Watanabe's algebraic geometry of singular learning [1] characterizes this; we provide the differential-geometric complement.

- **Embedding theorems:** Classical results (Whitney, Takens) bound the dimension needed to embed manifolds without self-intersection. Recent work [2] established analogous bounds for statistical manifolds; the quotient structure explains what happens below those bounds.

1.1 Main Contributions

We establish three main results:

1. **Fiber Structure Theorem** (Theorem 4): A collapse map π with constant rank r foliates $\mathcal{M}$ into $(n - r)$-dimensional fibers. The Fisher metric is degenerate along fibers: points on the same fiber yield identical projected distributions and cannot be distinguished by any estimator based on projected observations.

2. **Quotient Metric Theorem** (Theorem 8): Under regularity conditions, the Fisher metric descends to a well-defined Riemannian metric on the quotient space $\mathcal{M}/\sim_\pi$. We characterize when this descent is compatible with the α-connection structure and identify obstructions arising from non-trivial holonomy.

3. **Covering Number Bounds** (Theorem 14): The number of ε-distinguishable classes scales as $N(\varepsilon) \sim C \cdot \mathrm{Vol}(\mathcal{M})/\varepsilon^r$, where r is the projection rank. This quantifies the "effective alphabet size" induced by dimensional collapse.

1.2 Relation to Prior Work

This paper and a companion work [2] form a two-step program:

- **Companion paper:** establishes a specific phenomenon—cyclic processes with monotone meta-time require embedding dimension $k \geq 3$ to avoid self-intersection; at $k \leq 2$, temporally distinct states are necessarily conflated.

- **This paper:** provides the general quotient-geometric framework that makes that phenomenon inevitable. The minimal embedding result drops out of Theorem 14 as a special case (Corollary 16).

The genuinely new contributions of the present work are:

- **Quotient metric descent criteria:** characterizing when the Fisher metric descends to the quotient via bundle-like or totally geodesic conditions (Theorem 8).

- **α-connection obstruction framing:** identifying when dual connections survive collapse (Proposition 11).

- **Covering numbers as capacity:** the scaling $N(\varepsilon) \sim \varepsilon^{-r}$ provides an operational measure of distinguishable equivalence classes (Theorem 14).

The framework also connects to:

- Chentsov's uniqueness theorem [3]: the Fisher metric is the unique Riemannian metric (up to scale) invariant under sufficient statistics. Our quotient construction respects this invariance.

- Amari's α-geometry [4]: we show how dual connections transform under collapse maps and identify conditions for the dually flat structure to survive.

- Watanabe's singular learning theory [1]: our fiber structure provides the differential-geometric setting for his algebraic singularities.

2 Preliminaries

2.1 Statistical Manifolds

Let $\mathcal{M} = \{p_\theta : \theta \in \Theta\}$ be a parametric family of probability distributions on a sample space $\mathcal{X}$, where $\Theta \subseteq \mathbb{R}^n$ is an open subset. Under standard regularity conditions (the family is identifiable, $\theta \mapsto p_\theta$ is smooth, expectations can be differentiated under the integral sign), $\mathcal{M}$ carries the structure of a Riemannian manifold with the *Fisher information metric*:

$$g_{ij}(\theta) = \mathbb{E}_{p_\theta}\left[\frac{\partial \log p_\theta}{\partial \theta^i} \frac{\partial \log p_\theta}{\partial \theta^j}\right] = -\mathbb{E}_{p_\theta}\left[\frac{\partial^2 \log p_\theta}{\partial \theta^i \partial \theta^j}\right]. \tag{1}$$

The Fisher metric has fundamental statistical meaning: the geodesic distance $d_g(\theta, \theta')$ measures the difficulty of distinguishing p_θ from $p_{\theta'}$ based on finite samples. Specifically, the Cramér-Rao bound states that any unbiased estimator $\hat{\theta}$ of θ satisfies $\text{Cov}(\hat{\theta}) \geq g^{-1}(\theta)$.

2.2 The α-Connection Family

Beyond the metric, statistical manifolds carry a one-parameter family of affine connections. The *α-connection* $\nabla^{(\alpha)}$ is defined by:

$$\Gamma_{ij,k}^{(\alpha)} = \mathbb{E}_{p_\theta}\left[\left(\frac{\partial^2 \log p_\theta}{\partial \theta^i \partial \theta^j} + \frac{1-\alpha}{2} \frac{\partial \log p_\theta}{\partial \theta^i} \frac{\partial \log p_\theta}{\partial \theta^j}\right) \frac{\partial \log p_\theta}{\partial \theta^k}\right]. \tag{2}$$

Key special cases:

- $\alpha = 0$: the Levi-Civita connection of the Fisher metric.

- $\alpha = 1$: the exponential (e-)connection, flat in exponential families.

- $\alpha = -1$: the mixture (m-)connection, flat in mixture families.

The connections $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are *dual* with respect to g:

$$Xg(Y, Z) = g(\nabla_X^{(\alpha)} Y, Z) + g(Y, \nabla_X^{(-\alpha)} Z). \tag{3}$$

When both $\nabla^{(1)}$ and $\nabla^{(-1)}$ are flat (as in exponential families), the manifold is *dually flat*, enabling powerful computational tools.

2.3 Chentsov's Theorem

A fundamental result of Chentsov [3] states that the Fisher metric is essentially unique:

Theorem 1 (Chentsov). *The Fisher information metric is the unique Riemannian metric on finite sample spaces (up to a positive constant) that is invariant under sufficient statistics. That is, if $T : \mathcal{X} \to \mathcal{Y}$ is a sufficient statistic for $\{p_\theta\}$, and $\tilde{g}$ is the Fisher metric on the induced family $\{\tilde{p}_\theta\}$ on $\mathcal{Y}$, then the natural map $\mathcal{M} \to \tilde{\mathcal{M}}$ is an isometry.*

This theorem constrains how the Fisher metric can transform under dimensional reduction: any "natural" collapse must respect the sufficient statistic structure.

2.4 Notation Summary

For clarity, we summarize the key objects introduced in this paper:

Symbol	Description
$\mathcal{M}$	Full statistical manifold with parameters $\theta \in \Theta \subset \mathbb{R}^n$
g	Fisher metric on $\mathcal{M}$ (the full model)
$\pi : \mathcal{M} \to \mathbb{R}^k$	Collapse map ($k < n$, smooth surjective submersion)
$\tilde{p}_y$	Collapsed family: $p_\theta = \tilde{p}_{\pi(\theta)}$
$\tilde{g}$	Fisher metric of the collapsed model on $\pi(\mathcal{M})$
$g_{\mathrm{obs}} = \pi^* \tilde{g}$	Pullback (observed) metric on $\mathcal{M}$; has rank $r \leq k$
$\mathcal{V} = \ker(d\pi)$	Vertical distribution (tangent to fibers)
$\mathcal{H} = \mathcal{V}^\perp$	Horizontal distribution (g-orthogonal to fibers)
$\bar{g}$	Quotient metric on $\mathcal{M}/\sim_\pi$ (when it exists)

3 Collapse Maps and Fiber Structure

3.1 Statistical Collapse via Observation

Dimensional collapse in statistical models arises from observation: when we observe a system through a lower-dimensional statistic, parameters that produce identical observed distributions become indistinguishable.

Definition 2 (Observation-Induced Equivalence). *Let $T : \mathcal{X} \to \mathcal{Y}$ be a measurable map (a statistic or observation). This induces an equivalence relation on the parameter space:*

$$\theta_1 \sim_T \theta_2 \iff T_* p_{\theta_1} = T_* p_{\theta_2}, \tag{4}$$

where $T_* p_\theta$ denotes the pushforward (induced distribution) of p_θ under T.

When this equivalence relation has "regular" level sets, the quotient $\mathcal{M}/\sim_T$ inherits the structure of a lower-dimensional statistical manifold $\widetilde{\mathcal{M}}$, and we obtain a natural projection $F : \mathcal{M} \to \widetilde{\mathcal{M}}$.

Definition 3 (Collapse Map). *A collapse map is a smooth surjective submersion $\pi : \mathcal{M} \to \mathbb{R}^k$ with $k < \dim(\mathcal{M})$ such that the statistical family factors through π: there exists a family $\{\tilde{p}_y : y \in \pi(\mathcal{M})\}$ with $p_\theta = \tilde{p}_{\pi(\theta)}$ for all $\theta \in \mathcal{M}$.*

Equivalently, $\pi(\theta_1) = \pi(\theta_2)$ implies $p_{\theta_1} = p_{\theta_2}$. The map π serves as a coordinate representation of the quotient structure.

Collapse maps arise naturally in several contexts:

- **Sufficient statistics:** If $T : \mathcal{X} \to \mathcal{Y}$ is sufficient, the family factors through $T_* p_\theta$, inducing a collapse map on parameter space.

- **Marginalization:** Projecting a joint distribution to a marginal: $(\theta_X, \theta_Y) \mapsto \theta_X$.

- **Curved exponential families:** When a full exponential family is restricted to a submanifold, the natural parameters outside the submanifold become non-identifiable.

3.2 Vertical and Horizontal Distributions

Given a collapse map $\pi : \mathcal{M} \to \mathbb{R}^k$, the *vertical distribution* is the subbundle of $T\mathcal{M}$:

$$\mathcal{V} = \ker(d\pi) = \{v \in T_\theta \mathcal{M} : d\pi_\theta(v) = 0\}. \tag{5}$$

This consists of tangent directions along which π—and hence the observed distribution—is locally constant.

The *horizontal distribution* is the g-orthogonal complement:

$$\mathcal{H} = \mathcal{V}^\perp = \{v \in T_\theta \mathcal{M} : g(v, w) = 0 \text{ for all } w \in \mathcal{V}_\theta\}. \tag{6}$$

At each point, $T_\theta \mathcal{M} = \mathcal{V}_\theta \oplus \mathcal{H}_\theta$.

3.3 Fiber Structure Theorem

Theorem 4 (Fiber Structure and Non-Identifiability). *Let $\pi : \mathcal{M} \to \mathbb{R}^k$ be a collapse map (Definition 3) with constant rank $r \leq k$ on a neighborhood $U \subset \mathcal{M}$. Let g denote the Fisher metric of the full model on $\mathcal{M}$, and let $\tilde{g}$ denote the Fisher metric of the collapsed model $\{\tilde{p}_y\}$ on $\pi(U) \subset \mathbb{R}^k$. Then:*

(i) ***Foliation:*** *The vertical distribution $\mathcal{V} = \ker(d\pi)$ has dimension $n - r$ at each point and is integrable. Its integral submanifolds—the* fibers $\pi^{-1}(y)$ for $y \in \pi(U)$—foliate U.

(ii) ***Non-identifiability:*** *For θ_1, θ_2 on the same fiber, the distributions coincide: $p_{\theta_1} = p_{\theta_2}$. Equivalently, for any statistic measurable with respect to the σ-algebra generated by π, the induced distributions are identical. No estimator can distinguish parameters on the same fiber.*

(iii) ***Observed Fisher metric:*** *The pullback metric $g_{obs} := \pi^* \tilde{g}$ on $\mathcal{M}$ has rank r with $\ker(g_{obs}) = \mathcal{V}$. In coordinates, $(g_{obs})_{ij} = \sum_{a,b} \frac{\partial \pi^a}{\partial \theta^i} \frac{\partial \pi^b}{\partial \theta^j} \tilde{g}_{ab}$.*

(iv) ***Singularity:*** *The observed model $\{\tilde{p}_y\}$ viewed in the original θ-coordinates is* singular *in the sense of Watanabe [1]: the effective Fisher information has rank $r < n$, and classical asymptotics (Cramér-Rao bounds, BIC) fail along vertical directions.*

Proof. (i) By the constant rank theorem, $\ker(d\pi_\theta)$ has constant dimension $n - r$ on U. Integrability follows from Frobenius: $\mathcal{V} = \ker(d\pi)$ is the tangent bundle to the level sets of π, hence involutive.

(ii) By Definition 3, $\pi(\theta_1) = \pi(\theta_2)$ implies $p_{\theta_1} = \tilde{p}_{\pi(\theta_1)} = \tilde{p}_{\pi(\theta_2)} = p_{\theta_2}$.

(iii) The pullback $\pi^*\tilde{g}$ has kernel equal to $\ker(d\pi) = \mathcal{V}$ by the chain rule. The coordinate formula follows from the definition of pullback.

(iv) Vectors in $\mathcal{V}$ represent parameter changes that leave $\pi(\theta)$—and hence p_θ—unchanged, so they contribute zero to the Fisher information of the observed model. □

Remark 5 (Connection to Watanabe's Theory)**.** *Watanabe [1] characterizes singularities algebraically: the set where the Fisher matrix drops rank is an algebraic variety. His approach resolves singularities via blow-ups to compute the learning coefficient λ. Our framework is complementary: rather than resolving the singularity, we accept it as the feature that generates the quotient structure. The fiber structure provides the differential-geometric picture—singularity arises because the parameter-to-distribution map factors through a lower-dimensional space. The learning coefficient relates to the codimension of the singular set, which in our framework is $\dim(\mathcal{V}) = n - r$.*

4 Quotient Geometry

4.1 The Quotient Space

A collapse map $\pi : \mathcal{M} \to \mathbb{R}^k$ induces an equivalence relation:

$$\theta_1 \sim_\pi \theta_2 \iff \pi(\theta_1) = \pi(\theta_2). \tag{7}$$

The equivalence classes are exactly the fibers $\pi^{-1}(y)$.

Definition 6 (Quotient Statistical Manifold)**.** *The quotient $\mathcal{M}/\sim_\pi$ is the set of equivalence classes, equipped with the quotient topology. When π is a submersion (rank k everywhere), the quotient inherits a smooth manifold structure of dimension k.*

4.2 Descent of the Fisher Metric

For the Fisher metric to descend to a well-defined metric on $\mathcal{M}/\sim_\pi$, we need the metric to be constant along fibers in an appropriate sense.

Definition 7 (Bundle-Like Metric)**.** *A Riemannian metric g on $\mathcal{M}$ is* bundle-like *with respect to a foliation $\mathcal{F}$ if for any horizontal vector fields X, Y (orthogonal to the leaves), the function $g(X, Y)$ is constant along leaves.*

Theorem 8 (Quotient Metric). *Let $\pi : \mathcal{M} \to \mathbb{R}^k$ be a submersion (constant rank k). The Fisher metric g descends to a Riemannian metric $\bar{g}$ on $\mathcal{M}/\sim_\pi$ if and only if the metric is bundle-like with respect to the fiber foliation. When this holds, the quotient metric is defined by $\bar{g}(d\pi(X), d\pi(Y)) = g(X, Y)$ for $X, Y \in \mathcal{H}$.*

Proof. The metric $\bar{g}$ is well-defined if and only if $g(X, Y)$ depends only on the equivalence classes, not on the choice of fiber representatives. For horizontal lifts X, Y of vectors in the quotient, this requires $\mathcal{L}_V g(X, Y) = 0$ for all vertical vector fields V—i.e., $g(X, Y)$ is constant along fibers. This is precisely the bundle-like condition. $\qquad\square$

Proposition 9 (Totally Geodesic Implies Bundle-Like). *If the fibers are totally geodesic with respect to the Levi-Civita connection of g, then g is bundle-like. The converse is false in general.*

Proof. Totally geodesic fibers imply that the second fundamental form vanishes. By O'Neill's formulas for Riemannian submersions, this forces $g(X, Y)$ to be constant along fibers for basic (projectable) horizontal vector fields. Thus totally geodesic is a sufficient but not necessary condition for metric descent. $\qquad\square$

Example 10 (Gaussian Families). *Consider $\mathcal{M} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$ with Fisher metric $g = diag(1/\sigma^2, 2/\sigma^2)$.*

The projection $\pi(\mu, \sigma) = \mu$ has fibers $\{\sigma > 0\}$ for each fixed μ. The horizontal distribution is $\mathcal{H} = span(\partial_\mu)$. The quotient is $\mathbb{R}$ with metric $\bar{g} = 1/\sigma^2$—but this depends on σ, so the fibers are not totally geodesic and the quotient metric is not well-defined globally. However, on each fiber, we get a consistent 1D metric.

The projection $\pi(\mu, \sigma) = \sigma$ has fibers $\{\mu \in \mathbb{R}\}$ for each fixed σ. The fibers are totally geodesic (straight lines in the half-plane), and the quotient metric is $\bar{g} = 2/\sigma^2$ on $(0, \infty)$.

4.3   Behavior of α-Connections

Proposition 11 (α-Connection Under Collapse). *Let $\pi : \mathcal{M} \to \mathbb{R}^k$ be a collapse map with totally geodesic fibers (for $\nabla^{(0)}$). Then:*

(i) *The Levi-Civita connection $\nabla^{(0)}$ descends to the quotient.*

(ii) *The α-connections $\nabla^{(\alpha)}$ descend if and only if the fibers are also totally geodesic with respect to $\nabla^{(\alpha)}$.*

(iii) *For exponential families, the e-connection ($\alpha = 1$) descends when projecting onto natural parameters; the m-connection ($\alpha = -1$) descends when projecting onto expectation parameters.*

Proof. (i) follows from the Riemannian submersion theory. (ii) requires checking that parallel transport along fibers preserves horizontal vectors. (iii) follows from the dually flat structure: in natural coordinates, e-geodesics are straight lines, so any linear projection preserves them. $\qquad\square$

Corollary 12 (Preservation of Dual Flatness). *If $\mathcal{M}$ is dually flat and π is a linear projection in natural coordinates, then $\mathcal{M}/\sim_\pi$ inherits a dually flat structure.*

This explains why projecting exponential families onto subsets of natural parameters yields well-behaved quotients: the dual flatness structure is preserved.

5 Covering Numbers and Forced Discretization

5.1 Finite Resolution

In practice, we distinguish points only up to finite resolution $\varepsilon > 0$. Rather than working with equivalence classes (which require an equivalence relation), we use covering numbers.

Definition 13 (Covering Number). *The ε-covering number of $\pi(K) \subset \mathbb{R}^k$ is:*

$$N(\varepsilon) = \min\{m : \exists y_1, \ldots, y_m \text{ such that } \pi(K) \subset \bigcup_{i=1}^{m} B_\varepsilon(y_i)\}. \tag{8}$$

Each covering ball induces a cluster in $\mathcal{M}$: the preimage $\pi^{-1}(B_\varepsilon(y_i)) \cap K$. These clusters are the resolution-ε equivalence classes.

5.2 Covering Number Bounds

Theorem 14 (Covering Number Bound). *Let $\pi : \mathcal{M} \to \mathbb{R}^k$ have constant rank r on a compact region $K \subset \mathcal{M}$. Suppose:*

(R1) π is L-Lipschitz: $\|\pi(x) - \pi(y)\| \leq L \cdot d_g(x, y)$.

(R2) The Jacobian restricted to $\mathcal{H}$ has singular values bounded below by $\sigma_{\min} > 0$.

Then:
$$N(\varepsilon) \geq C_K \cdot \varepsilon^{-r}, \tag{9}$$
where the constant $C_K = \omega_r^{-1} \sigma_{\min}^r \cdot Vol_g(K)/\bar{V}_{fiber}$ absorbs the Fisher volume of K, the minimal singular value, and the mean fiber volume $\bar{V}_{fiber} = Vol_g(K)/\mathcal{H}^r(\pi(K))$. The bound is valid when fiber volumes are bounded away from zero and infinity on K.

Proof. The image $\pi(K)$ has Hausdorff dimension r. By the area formula, $\mathcal{H}^r(\pi(K)) \geq \sigma_{\min}^r \cdot Vol_g(K)/\bar{V}_{fiber}$. Any ε-cover of $\pi(K)$ requires at least $\mathcal{H}^r(\pi(K))/(\omega_r \varepsilon^r)$ balls, where ω_r is the volume of the unit r-ball. $\square$

Corollary 15 (Scaling Law). *As resolution improves ($\varepsilon \to 0$), the number of distinguishable classes grows as $N(\varepsilon) \sim \varepsilon^{-r}$. Lower-rank projections (more severe collapse) yield fewer classes at any resolution.*

5.3 Connection to Minimal Embedding

The minimal embedding theorem of [2] is a special case:

Corollary 16 (Minimal Embedding as Covering Collapse). *Let $\gamma : S^1 \to \mathcal{M}$ be a cyclic process with monotone meta-time (a coordinate that increases strictly along the trajectory and resets at cycle completion). Consider the projection $\pi : \mathcal{M} \to \mathbb{R}^k$.*

(i) $k = 2$*: Any such trajectory must self-intersect: the curve must return to its starting spatial position while meta-time has advanced. At each self-intersection, distinct temporal states (different meta-time values) map to the same projected point. This forces the* effective *image rank to* $r = 0$ *at intersection loci. By Theorem 14, the number of* ε*-distinguishable classes is bounded:* $N(\varepsilon) = O(1)$ *as* $\varepsilon \to 0$.

(ii) $k = 3$*: The Whitney embedding theorem guarantees that generic embeddings of* S^1 *into* $\mathbb{R}^3$ *are injective (no self-intersections). The projection has full rank* $r = 1$ *along the curve, and Theorem 14 yields* $N(\varepsilon) \sim \varepsilon^{-1} \to \infty$.

The $k = 3$ threshold marks the transition from "too few classes" (forced discretization) to "enough classes" (preserved continuity). In the language of the present framework: the meta-time coordinate becomes part of the vertical distribution $\mathcal{V}$ when $k = 2$, because self-intersections collapse temporally distinct states to the same fiber. At $k = 3$, meta-time remains in the horizontal distribution $\mathcal{H}$ and contributes to distinguishability.

6 Examples

6.1 Exponential Families

Let $\mathcal{M}$ be an exponential family:

$$p_\theta(x) = \exp(\theta \cdot T(x) - A(\theta))h(x), \tag{10}$$

with natural parameters $\theta \in \mathbb{R}^n$ and sufficient statistics $T : \mathcal{X} \to \mathbb{R}^n$. The Fisher metric is $g_{ij} = \partial^2 A / \partial \theta^i \partial \theta^j$.

Projection to subset of parameters. Let $\pi(\theta) = (\theta_1, \ldots, \theta_k)$. The fibers are affine subspaces $\{(\theta_1, \ldots, \theta_k, \theta_{k+1}, \ldots, \theta_n) : \theta_{k+1}, \ldots, \theta_n \text{ vary}\}$.

Since A is convex, the Hessian restricted to fibers is positive definite, so fibers have positive volume. The quotient is the marginal exponential family on $(T_1, \ldots, T_k)$.

Example 17 (Multinomial)**.** *For a* K*-category multinomial,* $\mathcal{M}$ *has dimension* $K - 1$*. Projecting to the first* k *category probabilities yields a quotient where distributions agreeing on* $p_1, \ldots, p_k$ *are equivalent. The fiber over* $(p_1, \ldots, p_k)$ *consists of all ways to distribute the remaining probability* $1 - \sum_{i=1}^{k} p_i$ *among categories* $k + 1, \ldots, K$.

6.2 Location-Scale Families

Consider the location-scale family $p_{\mu,\sigma}(x) = \sigma^{-1} f((x - \mu)/\sigma)$ for a fixed density f. The Fisher metric depends on f but has the general form:

$$g = \begin{pmatrix} a/\sigma^2 & b/\sigma^2 \\ b/\sigma^2 & c/\sigma^2 \end{pmatrix} \tag{11}$$

for constants a, b, c determined by f.

Projection to location. $\pi(\mu, \sigma) = \mu$. The fibers $\{\sigma > 0\}$ are not totally geodesic in general. The quotient metric would be $\bar{g} = (a - b^2/c)/\sigma^2$, which depends on the fiber coordinate σ—so no well-defined global quotient metric exists.

Projection to scale. $\pi(\mu, \sigma) = \sigma$. The fibers $\{\mu \in \mathbb{R}\}$ are geodesics when $b = 0$ (e.g., for symmetric f). In this case, the quotient metric $\bar{g} = c/\sigma^2$ is well-defined.

6.3 Gaussian Mixtures

Consider 2-component Gaussian mixtures $p_\theta = wN(\mu_1, \sigma_1^2) + (1 - w)N(\mu_2, \sigma_2^2)$ with $\theta = (w, \mu_1, \sigma_1, \mu_2, \sigma_2) \in (0, 1) \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$.

This is a singular model: when $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, the weight w is non-identifiable. The singular set is a 3-dimensional submanifold.

In our framework: the "identity" map (no explicit projection) has fibers along the non-identifiable locus. The effective Fisher information has rank 3 (not 5) on this locus, consistent with Theorem 4(iii).

7 Discussion

7.1 Geometric Interpretation of Singularity

The fiber structure theorem provides a geometric interpretation of singular statistical models: singularity arises when the parameter-to-distribution map factors through a lower-dimensional space. The "redundant" parameters span the fibers, and the effective model lives on the quotient.

This complements Watanabe's algebraic approach. Where he characterizes singularities via resolution of singularities and learning coefficients, we characterize them via fiber dimension and quotient geometry. The two perspectives are related: Watanabe's learning coefficient λ measures how the log-likelihood degenerates near singularities, while our fiber dimension measures the local rank drop.

7.2 Sufficient Statistics and Chentsov's Theorem

Chentsov's theorem states that the Fisher metric is invariant under sufficient statistics. Our framework explains this geometrically: a sufficient statistic defines a collapse map whose fibers are exactly the parameter sets with identical sufficient statistic distributions. The quotient inherits the Fisher metric precisely because the original metric is constant along these fibers.

Conversely, collapse maps that do *not* correspond to sufficient statistics have fibers that are not metrically homogeneous, obstructing metric descent.

7.3 Implications for Dimensional Reduction

The covering number bounds quantify what is lost under dimensional reduction. The key parameter is the *rank* of the collapse map:

- Rank r projections yield $O(\varepsilon^{-r})$ distinguishable classes.

- Lower rank means coarser quotient: fewer classes at any resolution.

- The minimal embedding dimension is the smallest k such that the projection has full rank generically.

This provides a geometric foundation for understanding when low-dimensional representations preserve essential structure. These resolution-dependent equivalence classes provide a natural substrate on which discrete structures may be defined, without invoking any additional assumptions beyond the geometry of the collapse map.

7.4 Open Problems

1. **Optimal quotients:** Given a target dimension k, which collapse maps minimize information loss (in a rate-distortion sense) while respecting the Fisher geometry?

2. **Curvature under collapse:** How do sectional curvatures transform under quotient maps? Are there obstructions to negative curvature in low-dimensional quotients?

3. **Holonomy:** When the horizontal distribution $\mathcal{H}$ is not integrable, parallel transport around loops in the quotient may not close. This is analogous to Berry phase in quantum mechanics. The statistical interpretation: the inferred state after a cycle in the projected space may depend on the unobserved trajectory through the fibers—a form of path-dependent inference.

4. **Higher-order geometry:** How do higher-order tensors (cubic forms, etc.) in information geometry behave under collapse?

8 Conclusion

We have developed the quotient geometry of statistical manifolds under collapse maps. The main results are:

1. **Fiber Structure Theorem:** Collapse maps foliate the manifold into fibers along which the Fisher metric degenerates, with points on the same fiber being statistically non-identifiable.

2. **Quotient Metric Theorem:** The Fisher metric descends to a quotient metric if and only if it is bundle-like (constant along fibers on horizontal vectors); totally geodesic fibers provide a sufficient condition. The α-connection structure descends under additional conditions.

3. **Covering Number Bounds:** The number of distinguishable classes at resolution ε scales as ε^{-r} where r is the projection rank.

The framework unifies dimensional reduction, sufficient statistics, and singular models under a common geometric structure. The minimal embedding theorem for recurrent processes emerges as a special case: the $k = 3$ threshold marks where covering numbers transition from bounded to unbounded.

Acknowledgments

Statements and Declarations

References

[1] Sumio Watanabe. *Algebraic Geometry and Statistical Learning Theory*. Cambridge University Press, 2009.

[2] Ian Todd. Minimal embedding dimension for self-intersection-free recurrent processes on statistical manifolds. *Information Geometry*, 2025. Under review. Preprint and code available at `https://github.com/todd866/minimalembeddingdimension`.

[3] N. N. Chentsov. *Statistical Decision Rules and Optimal Inference*. American Mathematical Society, 1982. Translation of the 1972 Russian original.

[4] Shun-ichi Amari. *Information Geometry and Its Applications*. Springer, 2016.