Unified Singular Information Geometry of Deep Neural Networks
Inverse Geometric Deep Learning

Hanwen Ge
Chalmers University of Technology

November 2025

Abstract

We introduce a unified geometric framework for deep neural networks that combines information geometry (via the Fisher–Rao metric on statistical manifolds) with singular Riemannian geometry (via degenerate pullback metrics, null foliations, and quotient spaces). Both parameter and representation spaces arise as singular pullbacks of a common Fisher–Rao metric on a space of probability distributions. Degeneracy of these pullback metrics encodes functional equivalences and symmetries, and the associated quotient constructions yield effective, low-dimensional geometries for both parameters and representations.

1 Introduction

We introduce a unified geometric framework for deep neural networks that combines *information geometry* (via the Fisher–Rao metric on statistical manifolds) with *singular Riemannian geometry* (via degenerate pullback metrics, null foliations, and quotient spaces). The key point is that both the geometry of *parameter space* and the geometry of *representation spaces* arise as *singular pullbacks of the same Fisher–Rao metric* on an underlying space of probability distributions. Degeneracy of these pullback metrics encodes functional equivalences and symmetries, and the associated quotient constructions yield effective, low-dimensional geometries for both parameters and representations.

Throughout, bold symbols such as $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ denote vectors, $\Theta \subset \mathbb{R}^P$ denotes parameter space, and f_θ denotes the neural network map with parameters $\theta \in \Theta$.

Conceptually, this follows a function-space viewpoint: only the input–output behaviour of the model matters, and parameters and hidden representations are coordinate descriptions of an underlying map into a space of probability distributions. The contribution here is to *internalise* this viewpoint in geometric terms: we fix a canonical information-geometric metric (Fisher–Rao) on the output side and pull it back to parameters and representations, taking the resulting singular pullbacks, their null foliations and their Kolmogorov quotients as the primary objects of study. We refer to this perspective as *inverse geometric deep learning*: instead of endowing input, latent or parameter spaces with ad hoc geometries and pushing them forward, we fix an information geometry on task outputs and pull it back through the network. In this note we focus on the *static* geometry of a fixed network (or a fixed training snapshot); how these structures evolve along optimisation is taken up elsewhere.

Statistical Model and Fisher–Rao Base Geometry

Let $X \in \mathbb{R}^d$ be a random variable representing inputs and Z the associated labels. A neural network with parameters $\theta \in \Theta$ induces, via a likelihood model, a conditional distribution

$$p_\theta(\mathbf{z} \mid \mathbf{x}) \;=\; p(\mathbf{z} \mid \mathbf{x}; \theta),$$

for example by composing $f_\theta(\mathbf{x})$ with a softmax or a Gaussian likelihood. We write

$$p_\theta(\mathbf{x}, \mathbf{z}) \;=\; p(\mathbf{x})\, p_\theta(\mathbf{z} \mid \mathbf{x})$$

for the joint model distribution, and consider the model family

$$\mathcal{S} \;=\; \{p_\theta(\mathbf{x}, \mathbf{z}) \mid \theta \in \Theta\}.$$

Under standard regularity assumptions, $\mathcal{S}$ is locally a finite-dimensional statistical manifold embedded in the infinite-dimensional space of densities. In singular models (such as overparameterised neural networks), this manifold structure holds only on regular strata, but the Fisher–Rao metric on the ambient space of densities remains well defined and induces the constructions below.

Information geometry endows $\mathcal{S}$ with the *Fisher–Rao metric*. At a point $p \in \mathcal{S}$, tangent vectors can be represented by score-like functions, and the Fisher–Rao metric G_{FR} is defined by the $L^2(p)$ inner product of these tangent vectors. In a particular parametrisation $\theta \mapsto p_\theta$, this yields the familiar coordinate expression: for $u, v \in T_\theta\Theta$,

$$G_{\mathrm{FR},\theta}(u, v) \;:=\; \mathbb{E}_{(\mathbf{x},\mathbf{z})\sim p_\theta}\Big[\big\langle u, \nabla_\theta \log p_\theta(\mathbf{x}, \mathbf{z})\big\rangle \, \big\langle v, \nabla_\theta \log p_\theta(\mathbf{x}, \mathbf{z})\big\rangle\Big]. \tag{1}$$

Equivalently, the population Fisher information matrix is the matrix representation of the pullback metric

$$g^{(\Theta)} := \Psi^* G_{\mathrm{FR}}, \qquad \Psi : \Theta \to \mathcal{S},\; \theta \mapsto p_\theta(\mathbf{x}, \mathbf{z}).$$

In practice, we regard $p(\mathbf{x})$ as fixed (for instance as the empirical input distribution), so the Fisher geometry is effectively determined by the conditional $p_\theta(\mathbf{z} \mid \mathbf{x})$.

In applications, we typically observe a finite dataset $\mathcal{D} = \{(\mathbf{x}_n, \mathbf{z}_n)\}_{n=1}^N$ and work with an empirical approximation. For a conditional model $p_\theta(\mathbf{z} \mid \mathbf{x})$, the empirical Fisher information matrix (FIM) is

$$\mathbf{F}(\theta) \;:=\; \frac{1}{N} \sum_{n=1}^{N} \Big[\nabla_\theta \log p_\theta(\mathbf{z}_n \mid \mathbf{x}_n)\Big] \Big[\nabla_\theta \log p_\theta(\mathbf{z}_n \mid \mathbf{x}_n)\Big]^\top. \tag{2}$$

As we will see below, this empirical FIM can itself be interpreted as a pullback of Fisher–Rao through a *dataset map*, so that both population and empirical Fisher geometries fit into the same singular pullback framework.

Abstract Singular Pullback Geometry

We now formulate a general geometric construction that underlies both representation- and parameter-space geometries.

Definition 1 (Pullback metric and singular Riemannian structure)**.** Let (Y, G) be a smooth Riemannian manifold with metric G (positive definite), and let $\Phi : Z \to Y$ be a smooth map between manifolds. The *pullback metric* $g := \Phi^* G$ is the smooth field of symmetric bilinear forms on TZ defined by

$$g_z(u, v) \;:=\; G_{\Phi(z)}\big(d\Phi_z(u), d\Phi_z(v)\big), \qquad u, v \in T_z Z.$$

In general g_z is only positive semi-definite, so the induced metric may be degenerate. We will say that (Z, g) carries a *singular Riemannian structure* when g is positive semidefinite and may have variable rank across Z.

Lemma 1 (Null space and differential)**.** *Under the above assumptions,*

$$\ker g_z \;=\; \ker d\Phi_z \;\subset\; T_z Z, \qquad \forall z \in Z.$$

Proof. If $v \in \ker d\Phi_z$ then $d\Phi_z(v) = 0$, so for any $u \in T_z Z$,

$$g_z(v, u) = G_{\Phi(z)}\big(d\Phi_z(v), d\Phi_z(u)\big) = G_{\Phi(z)}(0, d\Phi_z(u)) = 0.$$

In particular $g_z(v, v) = 0$, so $v \in \ker g_z$. Conversely, if $v \in \ker g_z$, then $g_z(v, v) = 0$, i.e.

$$G_{\Phi(z)}\big(d\Phi_z(v), d\Phi_z(v)\big) = 0.$$

Since G is positive definite, this implies $d\Phi_z(v) = 0$, so $v \in \ker d\Phi_z$. $\square$

Thus the degeneracy of the pullback metric is *exactly* controlled by the rank of the map Φ: null directions coincide with directions tangent to the fibers of Φ.

The pullback metric defines a natural *pseudo-length* and *pseudo-distance* on Z.

Definition 2 (Pseudometric induced by a pullback)**.** Let (Z, g) be as above. For a piecewise C^1 curve $\gamma : [0, 1] \to Z$ define

$$L_g(\gamma) := \int_0^1 \sqrt{g_{\gamma(t)}\big(\dot\gamma(t), \dot\gamma(t)\big)}\; dt.$$

The induced *pseudo-distance* is

$$d_g(z_0, z_1) := \inf_{\gamma} L_g(\gamma),$$

where the infimum is taken over all piecewise C^1 curves γ with $\gamma(0) = z_0$ and $\gamma(1) = z_1$. In general d_g is a pseudometric: it is symmetric and satisfies the triangle inequality, but $d_g(z_0, z_1)$ may be zero for $z_0 \neq z_1$.

Definition 3 (Kolmogorov quotient)**.** Let (Z, d_g) be a pseudometric space. The *Kolmogorov quotient* associated with d_g is the set of equivalence classes $Z/\!\sim$, where

$$z_0 \sim z_1 \quad\Longleftrightarrow\quad d_g(z_0, z_1) = 0,$$

equipped with the induced metric $\overline{d}_g([z_0], [z_1]) := d_g(z_0, z_1)$.

The key question is how the pseudometric d_g relates to the *functional* structure carried by Φ. The next result provides a clean answer under mild regularity assumptions.

Assumption 1 (Constant rank and regular fibers)**.** The map $\Phi : Z \to Y$ has locally constant rank on an open subset $U \subset Z$. Equivalently, for each $z \in U$ there is a neighbourhood U_z on which $\operatorname{rank}(d\Phi)$ is constant. In this case, by the constant rank theorem, each fiber $\Phi^{-1}(y) \cap U_z$ is a smoothly embedded submanifold of U_z, and its tangent space at z equals $\ker d\Phi_z$.

Proposition 1 (Metric and functional equivalence)**.** *Let (Y, G) be Riemannian, let $\Phi : Z \to Y$ be smooth, and let $g = \Phi^* G$. Assume 1 holds on an open subset $U \subset Z$, and that each fiber $\Phi^{-1}(y) \cap U$ is path-connected. Then for $z_0, z_1 \in U$,*

$$d_g(z_0, z_1) = 0 \quad\Longleftrightarrow\quad \Phi(z_0) = \Phi(z_1).$$

Proof sketch. If z_0 and z_1 lie in the same fiber $\Phi^{-1}(y)\cap U$ for some $y \in Y$, then by path-connectedness there exists a continuous curve $\gamma : [0,1] \to U$ with $\gamma(0) = z_0$, $\gamma(1) = z_1$ and $\gamma(t) \in \Phi^{-1}(y)$ for all t. This curve can be approximated by piecewise C^1 curves within the same fiber, and for each such curve we have $d\Phi_{\gamma(t)}(\dot\gamma(t)) = 0$, hence $g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t)) = 0$ by Lemma 1. Consequently $L_g(\gamma) = 0$, and thus $d_g(z_0, z_1) = 0$.

Conversely, assume $d_g(z_0, z_1) = 0$. By definition there exists a sequence of curves γ_k from z_0 to z_1 with $L_g(\gamma_k) \to 0$. A standard compactness and reparameterisation argument in finite dimensions yields a limiting curve γ in U with $g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t)) = 0$ almost everywhere. By Lemma 1, this implies $d\Phi_{\gamma(t)}(\dot\gamma(t)) = 0$ a.e., hence $(\Phi \circ \gamma)'(t) = 0$ a.e. Therefore $\Phi(\gamma(t))$ is constant, so $\Phi(z_0) = \Phi(z_1)$ and $\gamma([0,1]) \subset \Phi^{-1}(\Phi(z_0)) \cap U$. □

Under the assumptions of Proposition 1, the Kolmogorov quotient associated with d_g collapses exactly those points that are *functionally equivalent* under the map Φ. Locally on U, the quotient $Z/\!\sim$ carries a smooth Riemannian structure inherited from G via the identification $Z/\!\sim\, \cong \Phi(U) \subset Y$.

Remark 1 (SRG terminology and locality). In this work we use "singular Riemannian geometry" in a minimal sense: Z remains a smooth manifold, but the pullback metric g is only positive semidefinite and its rank can vary across Z. On regions where the rank is locally constant, Assumption 1 applies and fibers form smooth null leaves; globally, variation in rank and nontrivial fiber topology induce a stratified metric structure. We do not attempt to develop a full theory of singular Riemannian spaces in the sense of metric geometry; rather, SRG here refers to the analysis of these degenerate pullback metrics and their quotients on such constant-rank patches. In deep networks, global fibers can decompose into multiple symmetry-related components; our assumptions are therefore understood locally on an open region U and are not intended to describe the global topology of the quotient.

In this framework, the null foliations of the pullback metrics and their Kolmogorov quotients are the primary geometric objects; scalar summaries such as eigenvalue spectra or effective ranks are secondary reflections of how these foliations sit inside Z.

Representation-Space Singular Geometry

We first apply the abstract construction to the representation manifolds M_ℓ of a fixed network f_θ.

Deep networks as layered maps. We model a deep network (for fixed parameters θ) as a sequence of smooth maps between finite-dimensional manifolds:

$$M_0 \xrightarrow{\Lambda_1} M_1 \xrightarrow{\Lambda_2} \cdots \xrightarrow{\Lambda_L} M_L, \tag{3}$$

where

- M_0 is the input space (e.g. an open subset of $\mathbb{R}^d$),
- M_ℓ is the representation manifold at layer ℓ,
- $\Lambda_\ell : M_{\ell-1} \to M_\ell$ is the layer map (affine map composed with a nonlinearity),
- M_L is the final pre-likelihood output space.

We write $f_\theta : M_0 \to M_L$ for the full network map, so that $f_\theta = \Lambda_L \circ \cdots \circ \Lambda_1$. For $0 \le \ell < L$ we denote the *tail* map by

$$N_\ell \;:=\; \Lambda_L \circ \cdots \circ \Lambda_{\ell+1} \;:\; M_\ell \to M_L,$$

so $f_\theta = N_0$ and $N_{L-1} = \Lambda_L$.

Given a likelihood model from M_L into the statistical manifold $\mathcal{S}$, we obtain a map

$$\Phi_\ell : M_\ell \longrightarrow \mathcal{S}, \qquad h_\ell \mapsto p_\theta(\cdot \mid h_\ell),$$

where $h_\ell \in M_\ell$ is a representation at layer ℓ and $p_\theta(\cdot \mid h_\ell)$ is the induced output distribution through the tail N_ℓ and the likelihood.

Pullback metric and null directions. Pulling back the Fisher–Rao metric G_{FR} along Φ_ℓ yields a singular metric $g^{(\ell)} = \Phi_\ell^* G_{\mathrm{FR}}$ on M_ℓ: for $v, w \in T_{h_\ell} M_\ell$,

$$g^{(\ell)}_{h_\ell}(v, w) = G_{\mathrm{FR},\, \Phi_\ell(h_\ell)}\big(d\Phi_\ell(v), d\Phi_\ell(w)\big).$$

In an empirical setting on a dataset $\mathcal{D}$, this metric may be averaged over the empirical distribution of representations $h_\ell(\mathbf{x}_n)$, yielding an empirical Fisher-type metric that is still a pullback of G_{FR} through Φ_ℓ restricted to the empirical law.

By Lemma 1, the null space of $g^{(\ell)}$ at h_ℓ equals $\ker d\Phi_\ell$. Geometrically, a tangent vector $v \in T_{h_\ell} M_\ell$ is null if and only if infinitesimal movement in direction v does not change the induced distribution $p_\theta(\cdot \mid h_\ell)$. Curves tangent to $\ker d\Phi_\ell$ therefore encode *representation-side invariances and redundancies* with respect to the task. For instance, in a classifier where subsequent normalisation and pooling render the logits invariant to global additive brightness shifts, moving along a direction in M_ℓ that simply adds a constant offset to all pixels can be a null direction for the task.

On regions where Φ_ℓ has locally constant rank and connected fibers, Proposition 1 implies that the pseudometric induced by $g^{(\ell)}$ identifies precisely those representations h_ℓ, h'_ℓ that yield the same output distribution under the tail of the network. Equivalence classes are the connected components of fibers of Φ_ℓ; they form the leaves of a null foliation of M_ℓ.

Representation quotients. We denote the induced equivalence relation by

$$h_\ell \sim_\ell h'_\ell \quad \Longleftrightarrow \quad d_{g^{(\ell)}}(h_\ell, h'_\ell) = 0.$$

The corresponding quotient space

$$\overline{M}_\ell := M_\ell / {\sim_\ell}$$

may be interpreted as the *effective representation manifold* at layer ℓ: points in the same equivalence class are indistinguishable to the task, given the fixed tail network and likelihood.

Locally, on regions where Φ_ℓ has constant rank and connected fibers, the quotient $\overline{M}_\ell$ inherits a smooth Riemannian structure from G_{FR} and is diffeomorphic to the image of Φ_ℓ. As one moves across regions where the rank of Φ_ℓ changes (for example, at activation boundaries in piecewise-linear networks), new null directions can appear and the dimension of the quotient can drop. Heuristically, each $(M_\ell, g^{(\ell)})$ thus supports a stratified singular metric structure, with strata determined by the rank profile of Φ_ℓ and the induced foliation by null leaves.

Parameter-Space Singular Information Geometry

We now apply the same construction to parameter space Θ. The base manifold remains the statistical manifold $\mathcal{S}$ with Fisher–Rao metric.

Parameter-to-distribution map and Fisher metric. The statistical model induces a smooth map

$$\Psi : \Theta \longrightarrow \mathcal{S}, \qquad \theta \mapsto p_\theta(\cdot, \cdot).$$

Restricting to a fixed input distribution $p(\mathbf{x})$ (or to an empirical input distribution on $\mathcal{D}$), we may equally regard Ψ as the map $\theta \mapsto p_\theta(\mathbf{z} \mid \mathbf{x})$. Pulling back G_{FR} along Ψ yields a (possibly singular) metric on Θ:

$$g^{(\Theta)} := \Psi^* G_{\mathrm{FR}}.$$

In coordinates, the matrix representation of $g^{(\Theta)}$ in the basis $\{\partial_{\theta_i}\}$ is the population Fisher information matrix. Thus the familiar FIM is reinterpreted as the *parameter-space pullback* of the Fisher–Rao metric.

Null directions and functional equivalence of parameters. By Lemma 1,

$$\ker g_\theta^{(\Theta)} = \ker d\Psi_\theta.$$

A tangent vector $v \in T_\theta\Theta$ is null if and only if it belongs to the kernel of the differential $d\Psi_\theta$, i.e. if infinitesimal movement in direction v leaves the model distribution p_θ unchanged to first order. Such null directions encode

- *reparameterisation symmetries*, such as neuron permutations or layerwise rescalings that leave p_θ exactly invariant, and
- *task-specific redundancies*, where p_θ changes in regions of the sample space to which the chosen metric (e.g. Fisher–Rao on a given data distribution) is insensitive.

For example, in a ReLU network, jointly scaling the incoming weights to a neuron and inversely scaling its outgoing weights leaves p_θ unchanged and can define a null direction in $T_\theta\Theta$.

To connect the metric degeneracy to functional equivalence on a finite dataset, consider the restricted map

$$\Psi_{\mathcal{D}} : \Theta \longrightarrow \mathcal{S}^N, \qquad \theta \mapsto \big(p_\theta(\cdot \mid \mathbf{x}_1), \dots, p_\theta(\cdot \mid \mathbf{x}_N)\big).$$

Equip the product $\mathcal{S}^N$ with the product Fisher–Rao metric $G_{\mathrm{FR}}^{\otimes N}$. Then the empirical FIM (2) is the pullback

$$g_{\mathcal{D}}^{(\Theta)} := \Psi_{\mathcal{D}}^* G_{\mathrm{FR}}^{\otimes N}.$$

Under a constant-rank assumption on $\Psi_{\mathcal{D}}$ in a region $U \subset \Theta$, and path-connected fibers in U, Proposition 1 yields:

Corollary 1 (Parameter-space singular information geometry). *Let $g_{\mathcal{D}}^{(\Theta)}$ be the empirical Fisher metric on Θ induced by $\Psi_{\mathcal{D}}$. Assume $\Psi_{\mathcal{D}}$ has locally constant rank on an open subset $U \subset \Theta$ and that each fiber $\Psi_{\mathcal{D}}^{-1}(y) \cap U$ is path-connected. Define an equivalence relation on U by*

$$\theta \sim_{\mathcal{D}} \theta' \quad \Longleftrightarrow \quad d_{g_{\mathcal{D}}^{(\Theta)}}(\theta, \theta') = 0.$$

Then for $\theta, \theta' \in U$,

$$\theta \sim_{\mathcal{D}} \theta' \quad \Longleftrightarrow \quad p_\theta(\cdot \mid \mathbf{x}_n) = p_{\theta'}(\cdot \mid \mathbf{x}_n) \quad \forall n = 1, \dots, N.$$

The corresponding quotient

$$\overline{\Theta}_{\mathcal{D}} := \Theta/\sim_{\mathcal{D}}$$

may be interpreted as the *effective parameter manifold* seen through the lens of the empirical Fisher geometry. Locally on U, it inherits a smooth Riemannian structure and is diffeomorphic to the image of $\Psi_{\mathcal{D}}$. Global singularities—arising from changes in rank of $\Psi_{\mathcal{D}}$, or from non-trivial topology of fibers—lead to a stratified singular geometry in parameter space, in line with singular statistical models in the sense of **?**.

Population vs empirical Fisher. Formally, the same construction applies to the population map $\Psi : \Theta \to \mathcal{S}$, yielding the (possibly singular) population Fisher metric $g^{(\Theta)} = \Psi^* G_{\mathrm{FR}}$. In overparameterised networks this metric typically has non-trivial null directions due to symmetries and non-identifiability under the true data-generating distribution. The empirical Fisher metric $g_{\mathcal{D}}^{(\Theta)}$ induced by $\Psi_{\mathcal{D}}$ is the finite-sample object that appears in practical training and whose quotient captures functional equivalence on the dataset $\mathcal{D}$. Understanding how the empirical quotients $\overline{\Theta}_{\mathcal{D}}$ approximate the population quotient determined by $g^{(\Theta)}$, and how discrepancies relate to generalisation, is an important question left for future work.

Unified View: Singular Pullbacks of a Common Base Metric

The constructions in Sections 1 and 1 can be summarised as instances of the same pattern. Let G_{FR} denote the Fisher–Rao metric on $\mathcal{S}$, and equip $\mathcal{S}^N$ with the product Fisher–Rao metric $G_{\mathrm{FR}}^{\otimes N}$. For any of the maps

$$\Phi \in \{\Phi_\ell : M_\ell \to \mathcal{S},\ \Psi : \Theta \to \mathcal{S},\ \Psi_{\mathcal{D}} : \Theta \to \mathcal{S}^N\},$$

write Z for its domain and $\widetilde{G}$ for the corresponding Fisher–Rao (or product Fisher–Rao) metric on the target. Then we obtain:

- a singular metric $g = \Phi^* \widetilde{G}$ on the source manifold Z;
- a null distribution $\ker g = \ker d\Phi$ describing task-induced invariances and redundancies;
- a Kolmogorov quotient $Z/\sim$ (Definition 3) whose points correspond to equivalence classes of functionally indistinguishable parameters or representations on the region where Φ has locally constant rank.

Thus, both parameter and representation spaces can be seen as manifolds equipped with *singular Riemannian structures* induced by pullback from a common Fisher–Rao base. Their null foliations and quotient spaces describe, respectively,

- families of functionally equivalent parameterisations (reparameterisation symmetries, redundant directions, sparse sub-networks), and
- families of functionally equivalent representations (layer-wise invariances, effective submersions, representation collapse at the level of the task).

In this unified picture, the various objects considered elsewhere—the Fisher information matrix, its spectrum and effective dimension, Cauchy–Green-type deformation tensors, and NTK-like operators—appear as different coordinate expressions or summaries of the same underlying singular pullback geometry. This is the sense in which the framework realises an *inverse geometric deep learning* perspective: starting from the task-induced Fisher–Rao geometry on output distributions, one pulls this geometry back through the network to obtain a layered, singular Riemannian structure on both representation and parameter spaces, and studies the resulting null foliations and quotient manifolds as precise geometric encodings of degeneracy, invariance, and effective degrees of freedom in deep neural networks.

Geometric Challenges, Novelty, and Research Directions

The unified SRG–information-geometric framework above is deliberately local and static: it describes the singular geometry of a fixed network (or a fixed training snapshot), analysed on regions where the relevant maps have locally constant rank. This reflects not a flaw of the construction, but genuine geometric features of modern deep networks. Several of the “limitations” of the framework can be reinterpreted as concrete research directions:

- **Stratified structure and rank profiles.** The metric–functional equivalence and smooth quotient structure are guaranteed only on regions where the relevant map Φ (either Φ_ℓ or $\Psi_{\mathcal{D}}$) has locally constant rank and path-connected fibers. In realistic deep networks, the rank profile changes across activation boundaries or parameter-space symmetries, leading to a stratified singular geometry. A natural research direction is to characterise these strata, their dimensions, and how training trajectories move between them.
- **Piecewise-smooth activations.** The presentation above assumes smooth maps. For piecewise-smooth activations such as ReLU, the construction applies on each smooth region, yielding a piecewise SRG structure whose strata are glued along activation boundaries. Extending the analysis to fully nonsmooth settings, for example via Clarke differentials or metric-geometry tools, is an open technical direction.
- **Empirical versus population geometry.** The parameter-space geometry seen in practice is typically induced by empirical Fisher metrics on finite datasets. Understanding how the empirical quotients $\overline{\Theta}_{\mathcal{D}}$ and $\overline{M}_\ell$ converge (or fail to converge) to their population counterparts as $|\mathcal{D}| \to \infty$ is a natural bridge between finite-sample training dynamics and asymptotic information geometry.
- **Symmetries, non-identifiability, and subnetworks.** Reparameterisation symmetries and non-identifiability create non-trivial fiber topology and singularities in the quotients. Rather than obstacles, these are precisely the phenomena SRG is designed to describe. A key research direction is to relate the null foliations and quotient strata to practically observed structures such as lottery-ticket subnetworks, low-dimensional training subspaces, and mode connectivity in parameter space.
- **Approximate computation and geometric diagnostics.** Exact Fisher–Rao pullbacks are infeasible for large modern networks. Low-rank and structured approximations to the Fisher information suggest practical ways to estimate local null spaces and effective dimensions. Developing such approximations within the SRG framework, and relating the resulting geometric diagnostics to standard notions of flatness, capacity and generalisation, is a promising direction.
- **Dynamics and training trajectories.** Although the present note is static, the same pullback and quotient constructions can be applied to successive snapshots along an optimisation trajectory. Tracking how ranks, null spaces and quotient dimensions evolve during training—how the quotient geometry folds and unfolds as invariances are learned or discarded—would connect the SRG perspective directly to optimisation and learning dynamics.

Finally, it is worth situating this framework relative to prior work. Singular Riemannian approaches to deep networks pull back a metric from the output space through layered maps and study the induced pseudometrics and quotient structures. Pullback information geometry applies Fisher–Rao pullbacks to generative models in order to understand latent geometry. Singular learning theory develops asymptotic statistics for models with degenerate Fisher information and non-identifiable parameters. The present framework fixes Fisher–Rao as a canonical base metric and treats both parameter and representation spaces as singular pullbacks of this common base, taking null distributions and Kolmogorov quotients as primary objects. In this way it links SRG-style layered pullbacks, information geometry, and singular statistical models into a single geometric language for degeneracy and effective degrees of freedom in deep neural networks.

References omitted in this draft.