Information Geometry of Learned Latent Representations

Part I: How does information geometry reveal latent structures?

ML4Jets Vienna 2026

Rebecca Maria Kuntz | Astronomisches Rechen-Institut, Heidelberg University
Tilman Plehn, Björn Malte Schäfer, Benedikt Schosser, Sophia Vent

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Motivation

"The Geometry of Decision Making"

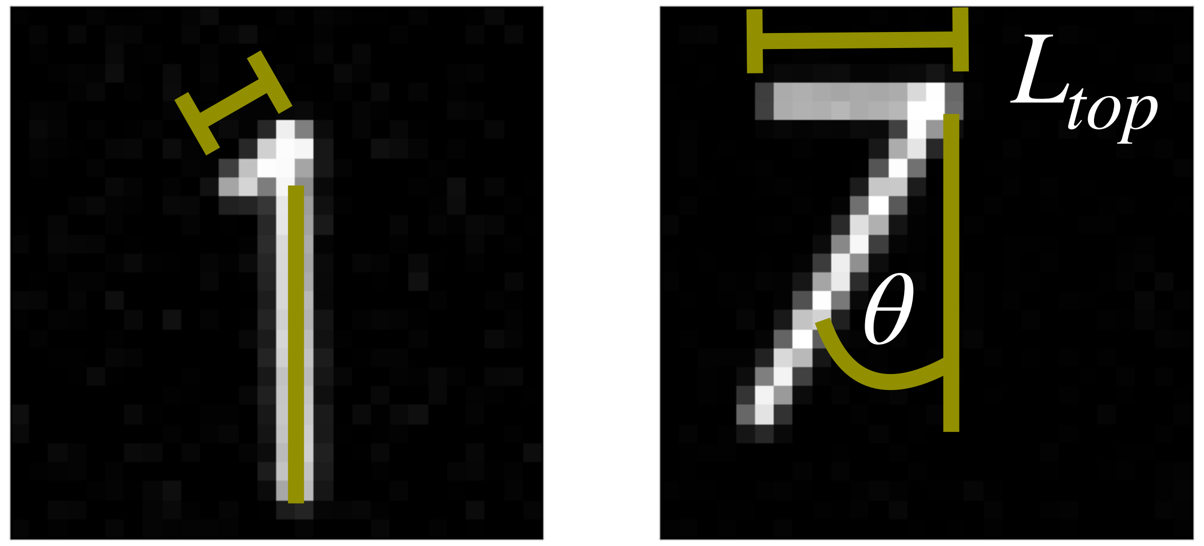
Goal: find a geometric link between a networks classification decision and (jet-)physics observables.

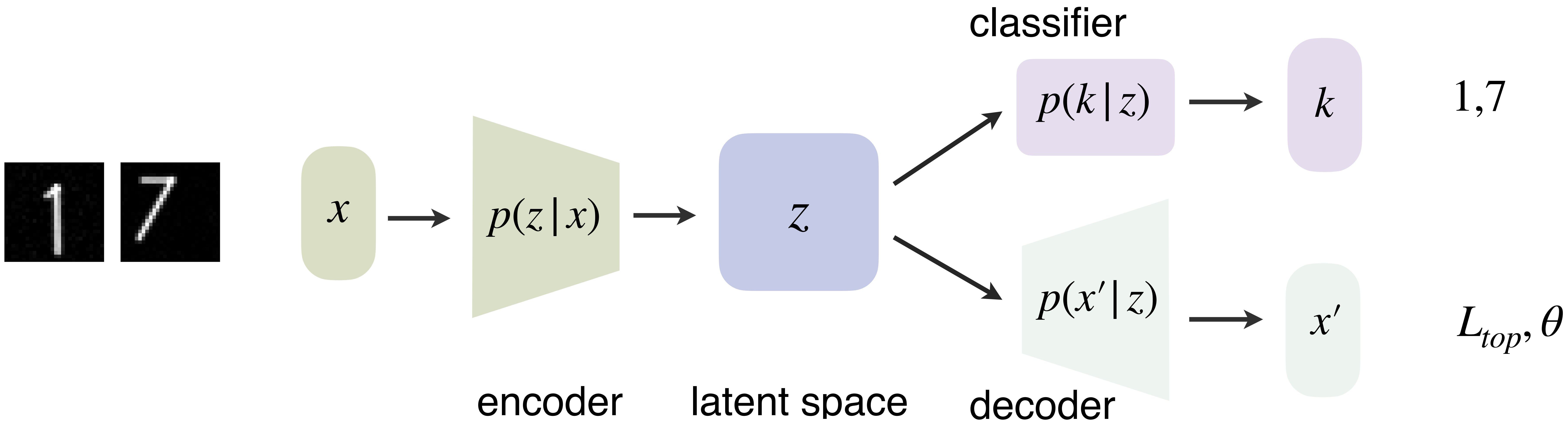
Question: Given a task, embed high-dimensional data in a low-dimensional latent space. Do geometric structures emerge?

Information geometry!

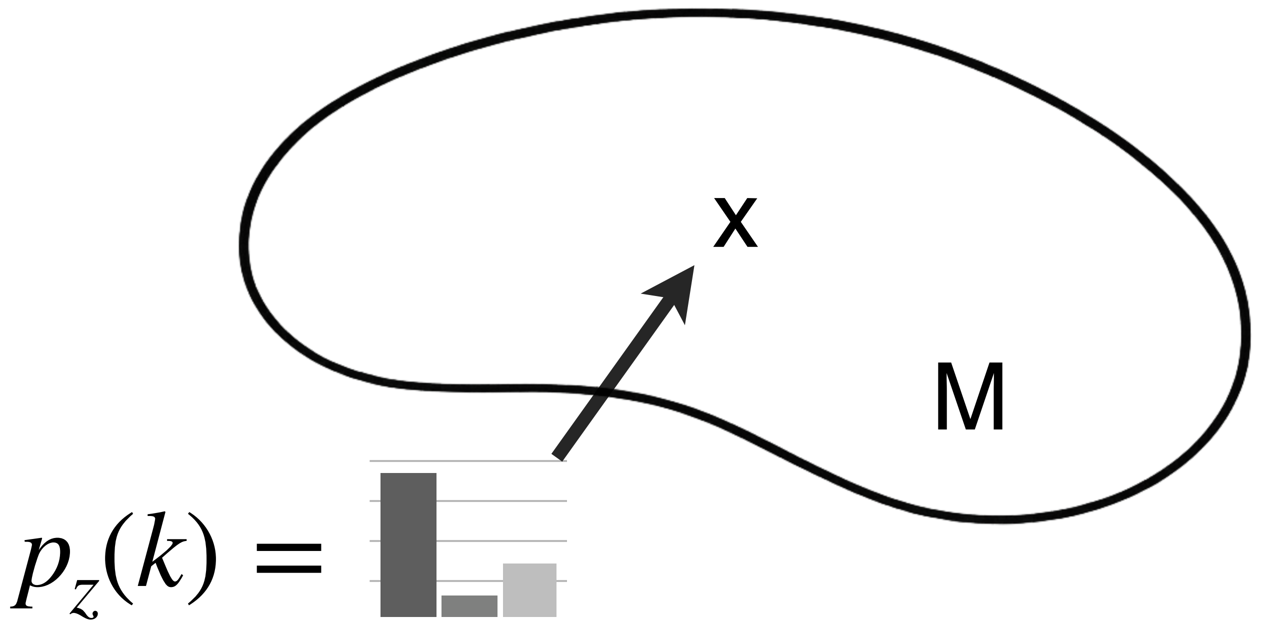
Network and Latent Space

- Variational autoencoder (VAE) with classifier head
- Toy data: MNIST with controlled features
- Training: reconstruction & classification
- Regularization + probabilistic network → information geometry





Construction of a Latent Information Geometry

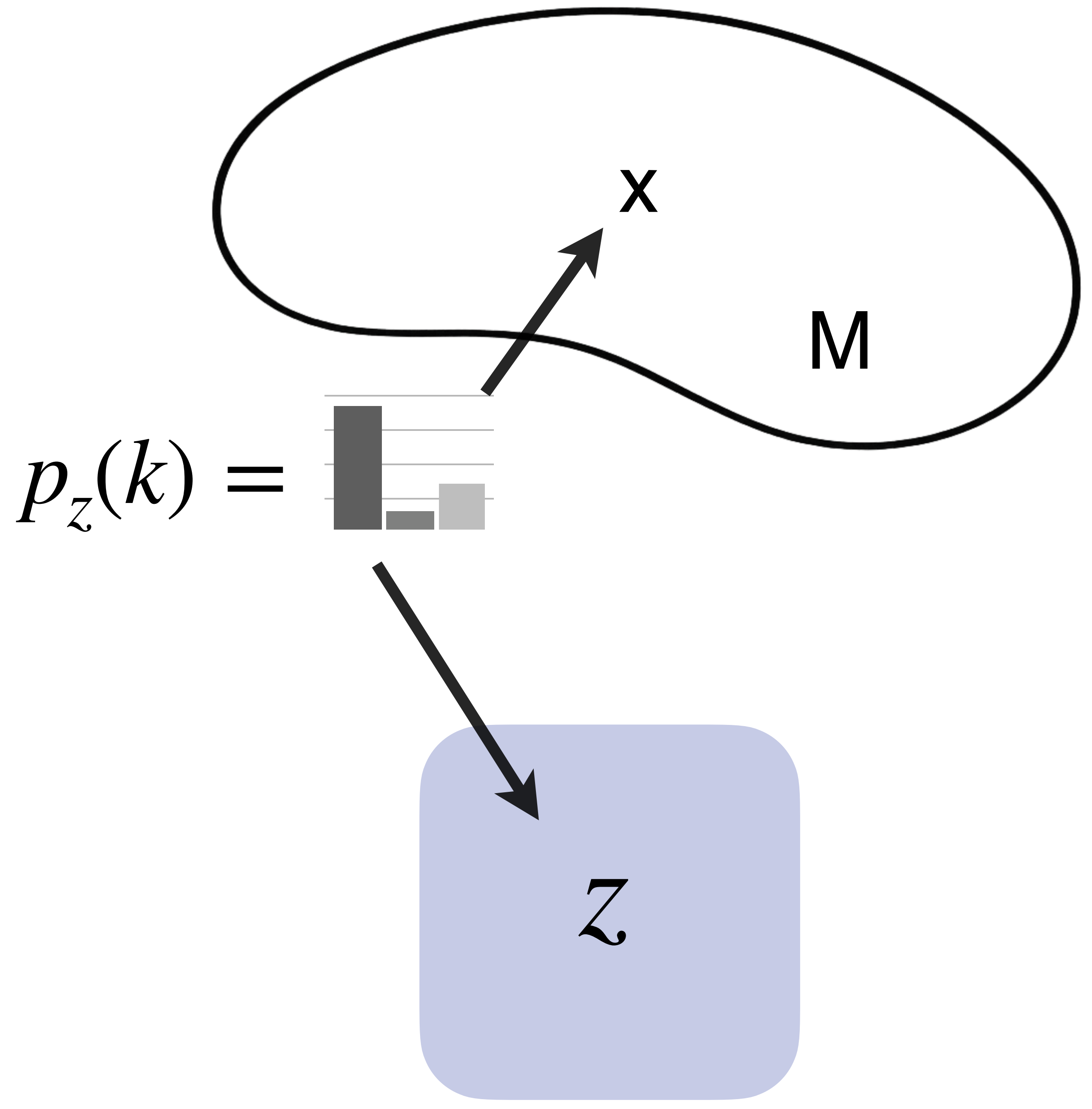


Each distribution p_z is identified with a point on a **statistical manifold** M

Construction of a Latent Information Geometry

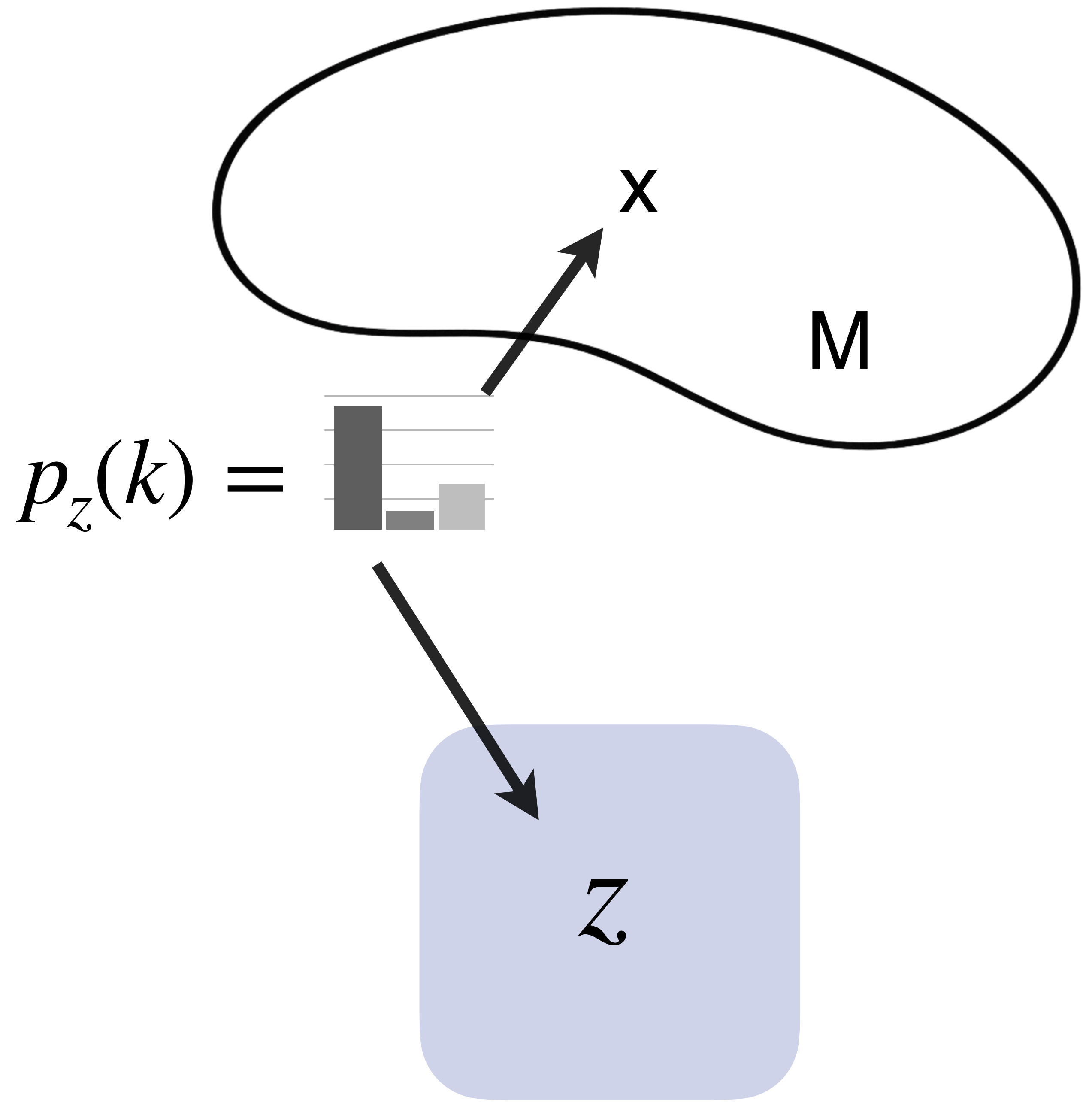
Each distribution p_z is identified with a point on a **statistical manifold** M

Latent $\{z^i\}$ are a special, learned coordinate frame



Construction of a Latent Information Geometry

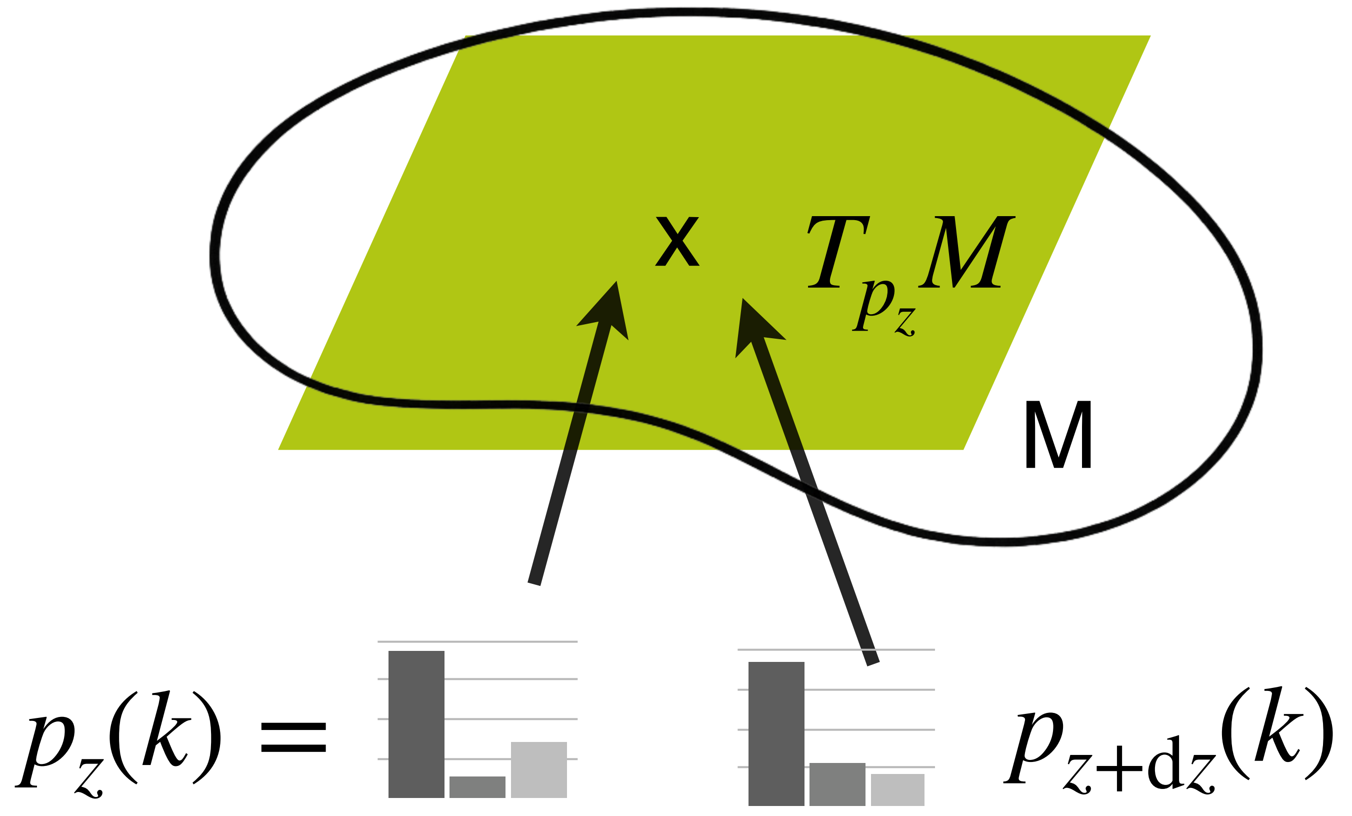
Each distribution p_z is identified with a point on a **statistical manifold** M



Latent $\{z^i\}$ are a special, learned coordinate frame

Fisher information metric

$$F_{ij}(z) = -\partial_i \partial_{j'} D_{\mathrm{KL}}[p_z, p_{z'}] = \mathbb{E}_{p_z}[\partial_i \log p_z \partial_j \log p_z], \quad \partial_i \equiv \frac{\partial}{\partial z^i}$$

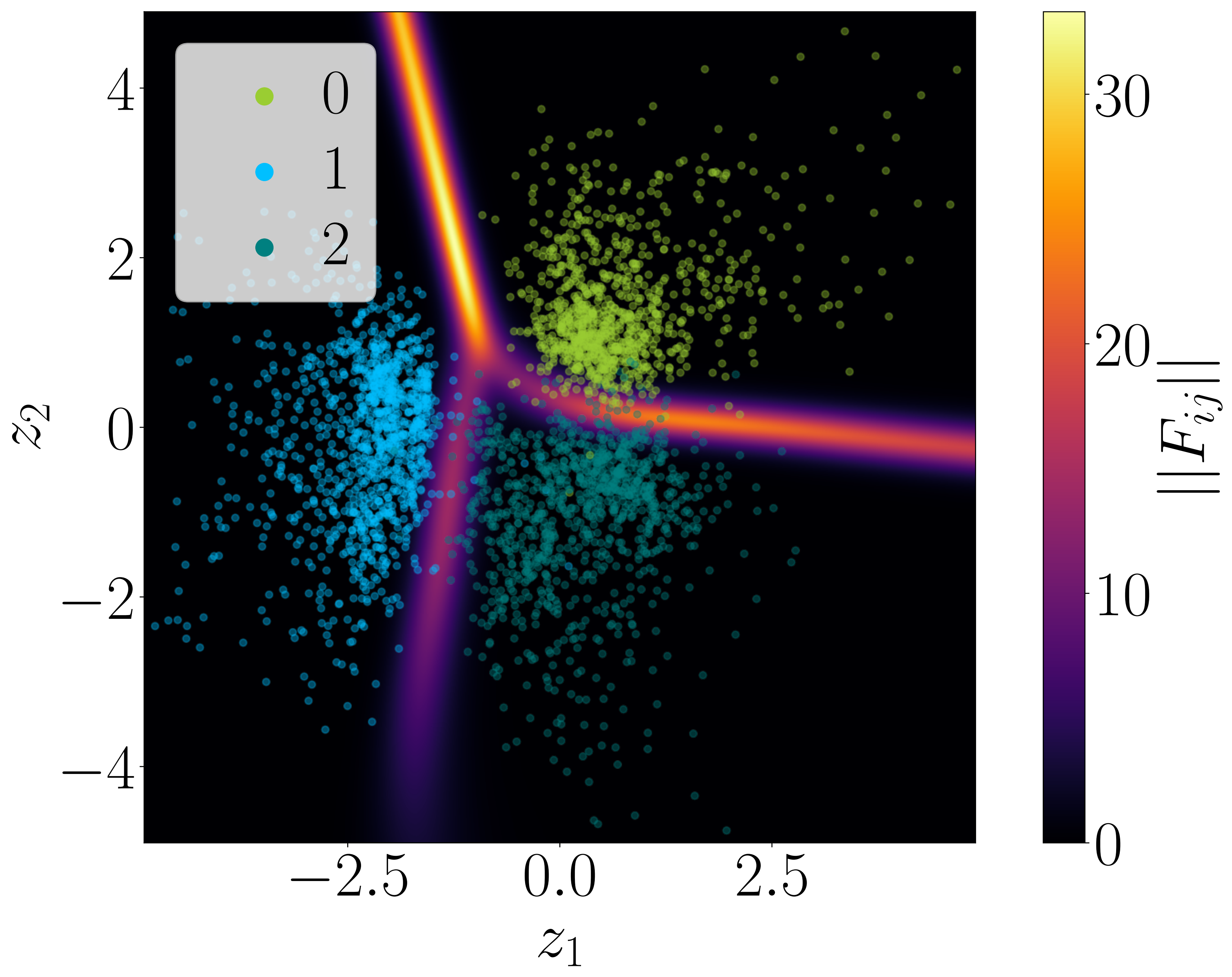


Classifier Fisher information

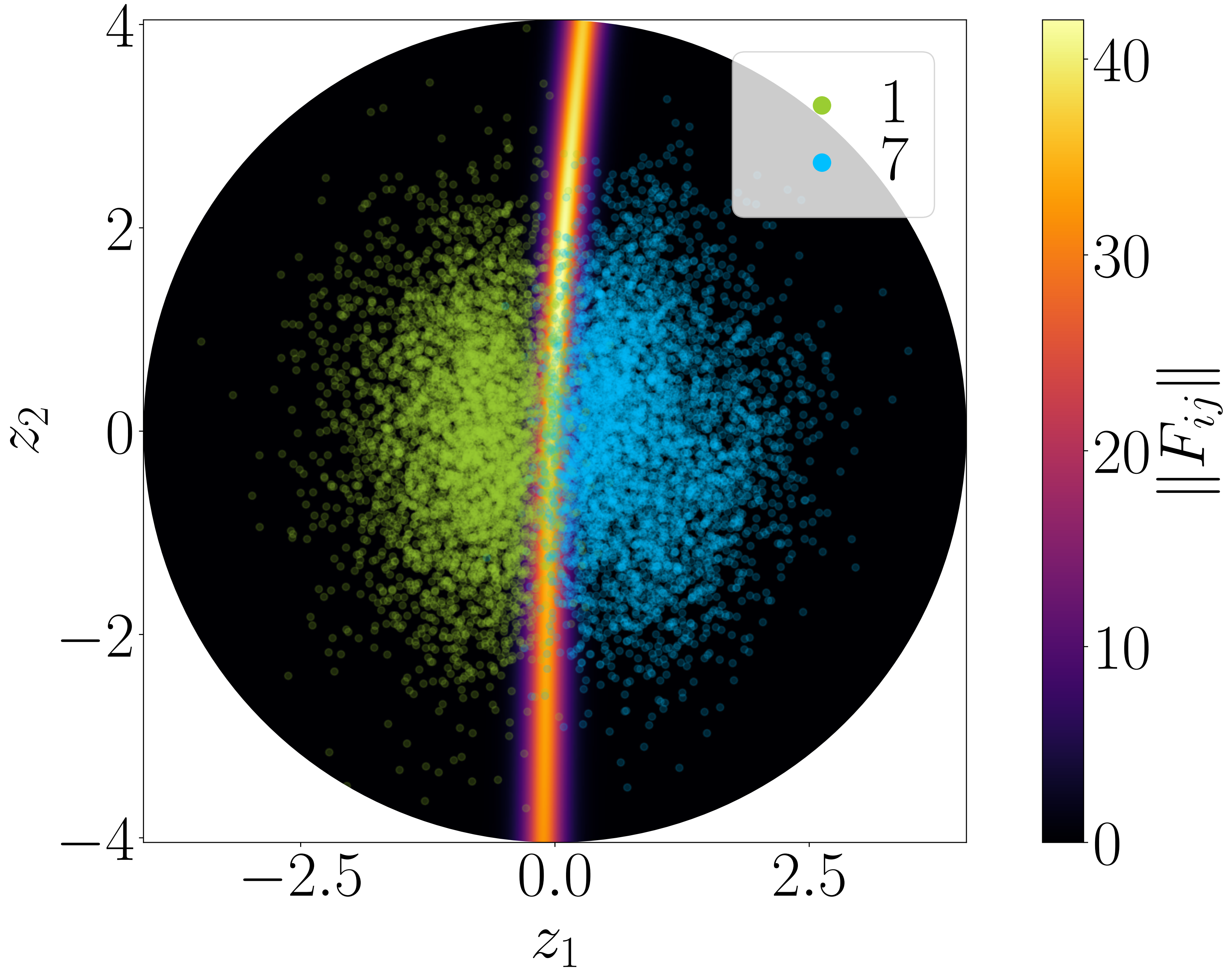
Fisher norm $||F_{ij}|| = \sum F_{ij}F_{ij}$

Local measure of event distinguishability / classifier sensitivity.

MNIST



1 — 7



Classifier vs. Decoder Geometry

Network Head	**Classifier** $p(k\,\vert\,z)$	**Decoder/ Reconstruction** $p(x'\,\vert\,z)$
Distribution Family	Categorical: $p_z(k) \propto \exp(\eta_k(z))$	Gaussian: $p_z(x') = \mathcal{N}(x'; \mu(z), \sigma)$
Moving in latent space	Change in class probabilities (p_1, p_2, p_3).	(Correlated) change in $\vec{\mu} = (\mu_1, \ldots, \mu_N)$.

Geometric Link between Classifier and Decoder

Leading Observable

A. Most sensitive direction of classifier metric

B. Directional derivative of reconstructed feature

$$D_v \mu(z) = \partial_\lambda \mu(z + \lambda v)\,\big|_{\lambda=0}$$

↪ Observable which most effectively changes classifier outcome.

C. (Pattern stability under retraining.)

Geometric Link between Classifier and Decoder

Leading Observable

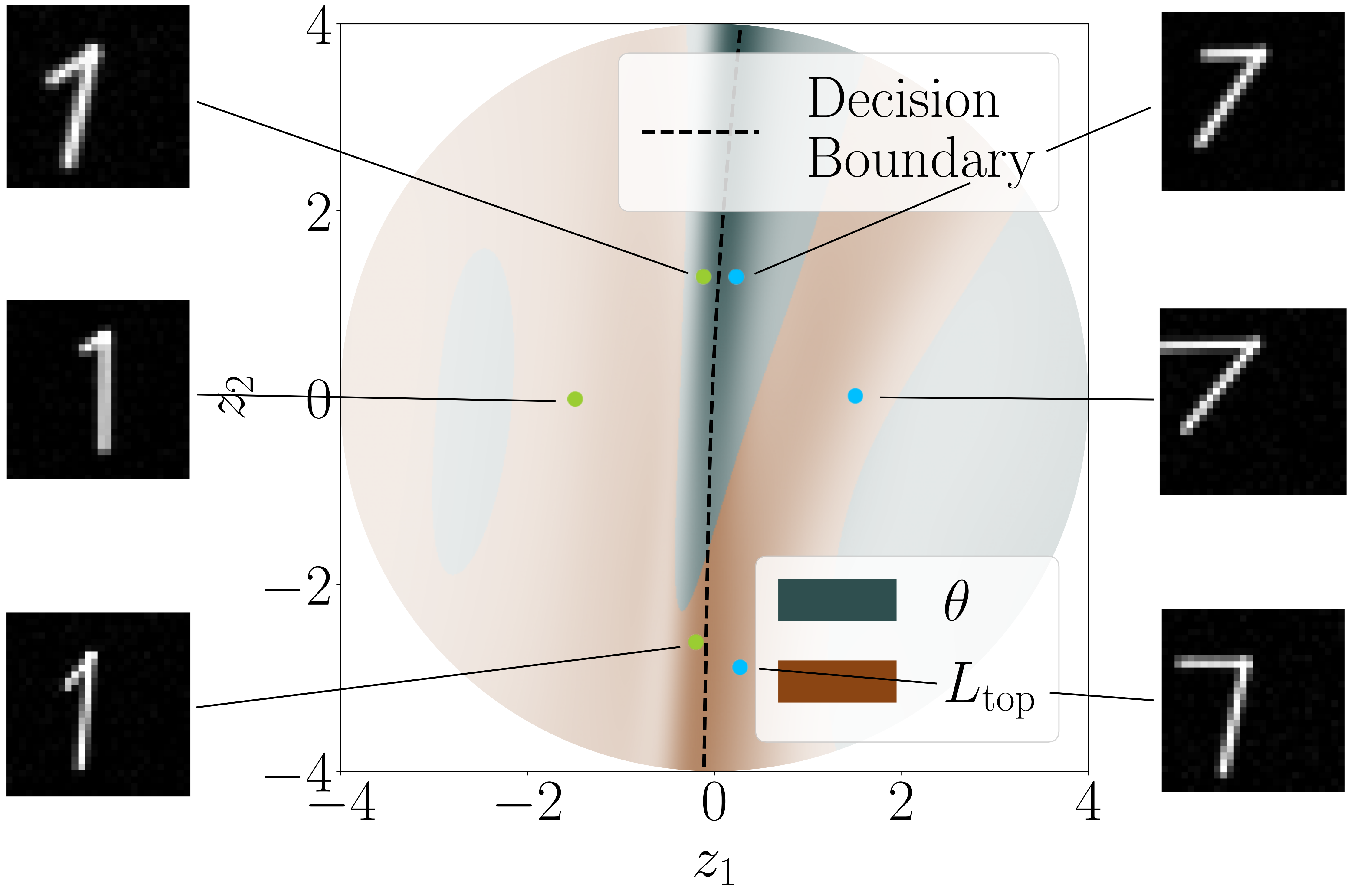
A. Most sensitive direction of classifier metric

B. Directional derivative of reconstructed feature

$$D_v \mu(z) = \partial_\lambda \mu(z + \lambda v)\,|_{\lambda=0}$$

↪ Observable which most effectively changes classifier outcome.

C. (Pattern stability under retraining.)



> Is interpretable physics reflected in classifier geometry?

↪ *For most events, θ is the most decisive feature.*

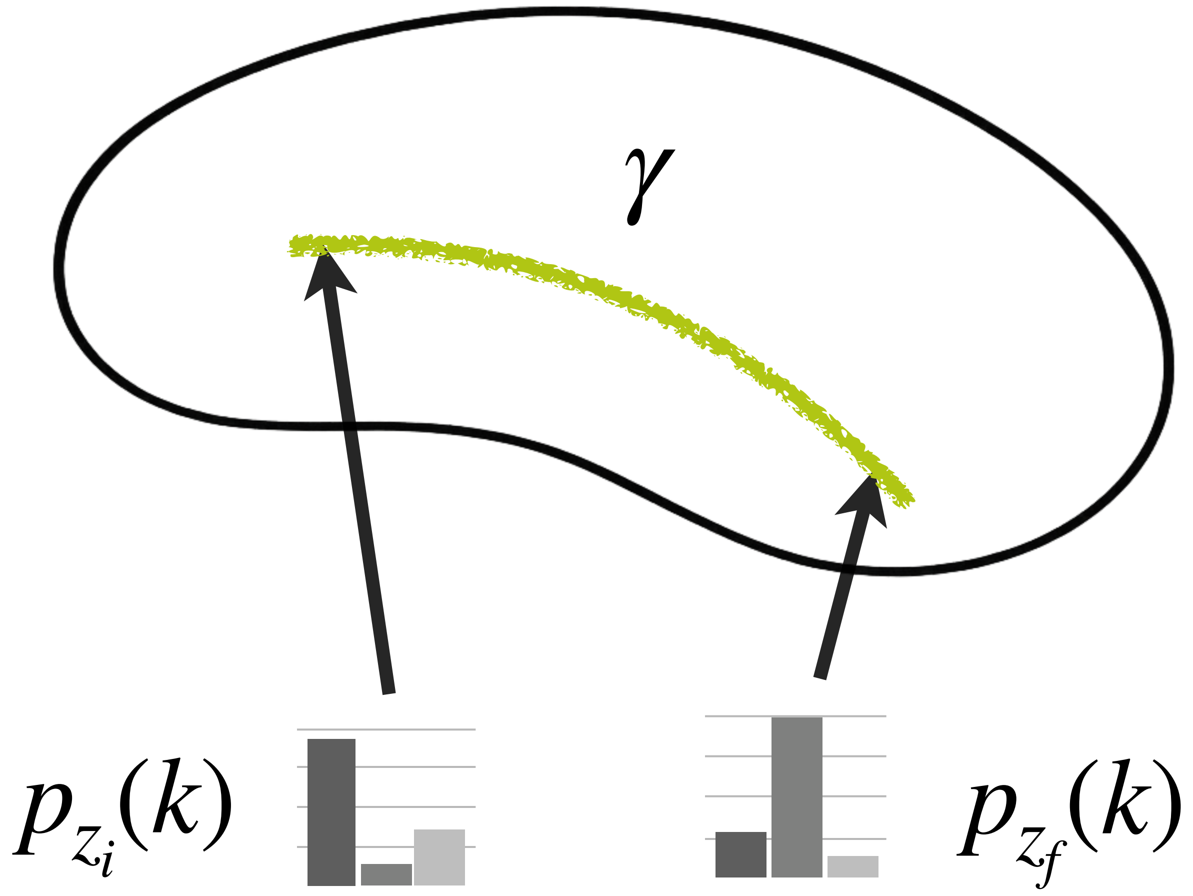
↪ *Class distinction is also reflected in feature geometry.*

→ Beyond Local Information

Geodesic Paths

Geodesics:

- Geometrically: locally shortest paths $\gamma(t)$ with $\gamma : [0,1] \to M$.
- Optimal (meaning shortest) path to morphe one event into another acc. to the classifier.



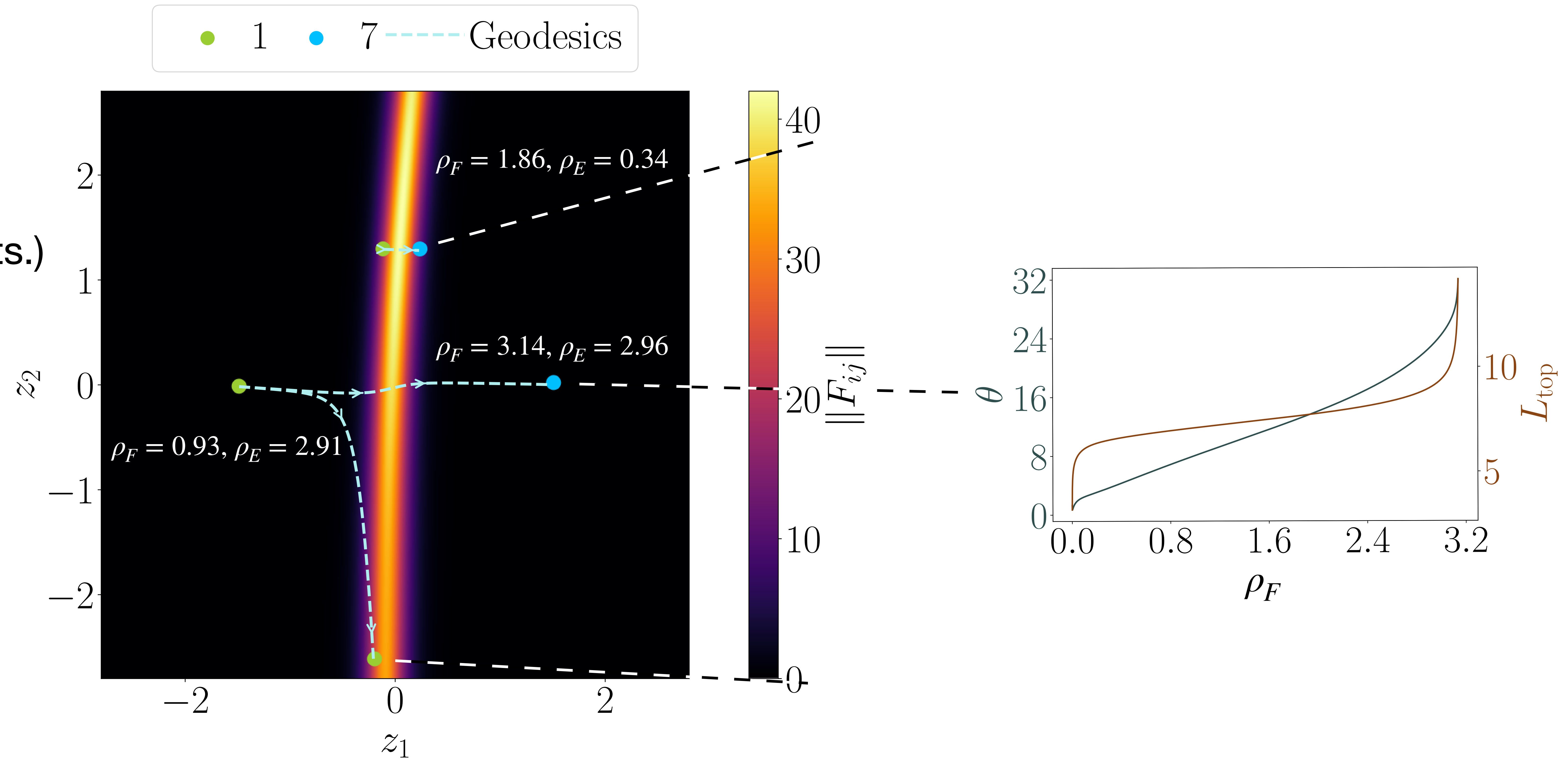
Two geometries IV

Link between Classifier Geodesic and Features

Leading features along classifier paths

A. Change of μ along classifier geodesic

B. (Geodesic distances $\rho(z_i, z_f)$ between events.)



→ Beyond Second-Order information

Third-Order in the Classifier Geometry

Divergence-Induced Geometry: $D_{\mathsf{KL}}(p_z, p_{z+\mathrm{d}z}) \approxeq \quad \frac{1}{2} F_{ij}(z) \mathrm{d}z^i \, \mathrm{d}z^j \quad + \quad \frac{1}{3!} C_{ijk}(z) \, \mathrm{d}z^i \mathrm{d}z^j \mathrm{d}z^k \quad (+^{(+1)}\Gamma \ldots) \quad + \quad \mathcal{O}((\mathrm{d}z)^4)$

Third-Order in the Classifier Geometry

decoder

Divergence-Induced Geometry: $D_{\mathsf{KL}}(p_z, p_{z+\mathrm{d}z}) \approxeq \frac{1}{2} F_{ij}(z) \mathrm{d}z^i \, \mathrm{d}z^j + \frac{1}{3!} C_{ijk}(z) \, \mathrm{d}z^i \mathrm{d}z^j \mathrm{d}z^k \quad (+^{(+1)}\Gamma \ldots) \quad + \quad \mathcal{O}((\mathrm{d}z)^4)$

Second-order = Riemannian Geometry.

Everything we saw thus far

(Fisher metric, geodesics,…)

Third-Order in the Classifier Geometry

Divergence-Induced Geometry: $D_{\mathrm{KL}}(p_z, p_{z+\mathrm{d}z}) \approxeq \frac{1}{2} F_{ij}(z)\mathrm{d}z^i\, \mathrm{d}z^j + \frac{1}{3!} C_{ijk}(z)\, \mathrm{d}z^i \mathrm{d}z^j \mathrm{d}z^k\ (+{}^{(+1)}\Gamma \ldots) + \mathcal{O}((\mathrm{d}z)^4)$

Second-order = Riemannian Geometry.

Everything we saw thus far

(Fisher metric, geodesics,…)

Third-order = Non-Riemannian Geometry

Third-Order in the Classifier Geometry

Divergence-Induced Geometry: $D_{\mathrm{KL}}(p_z, p_{z+\mathrm{d}z}) \approxeq \frac{1}{2} F_{ij}(z)\mathrm{d}z^i\, \mathrm{d}z^j + \frac{1}{3!} C_{ijk}(z)\, \mathrm{d}z^i \mathrm{d}z^j \mathrm{d}z^k \; (+ {}^{(+1)}\Gamma \dots) + \mathcal{O}((\mathrm{d}z)^4)$

Third-order = Non-Riemannian Geometry

> Statistically: skewness-/ Amari-Chentsov tensor: $C_{ijk} = \mathbb{E}_{p_z}\left[\partial_i \log p_z \partial_j \log p_z \partial_k \log p_z\right]$
>
> Geometrically: nonmetricity tensor: $\pm C_{ijk} = {}^{(\pm 1)}\nabla_i F_{jk}$

Classifier Nonmetricity

New: Nonmetricity scalars

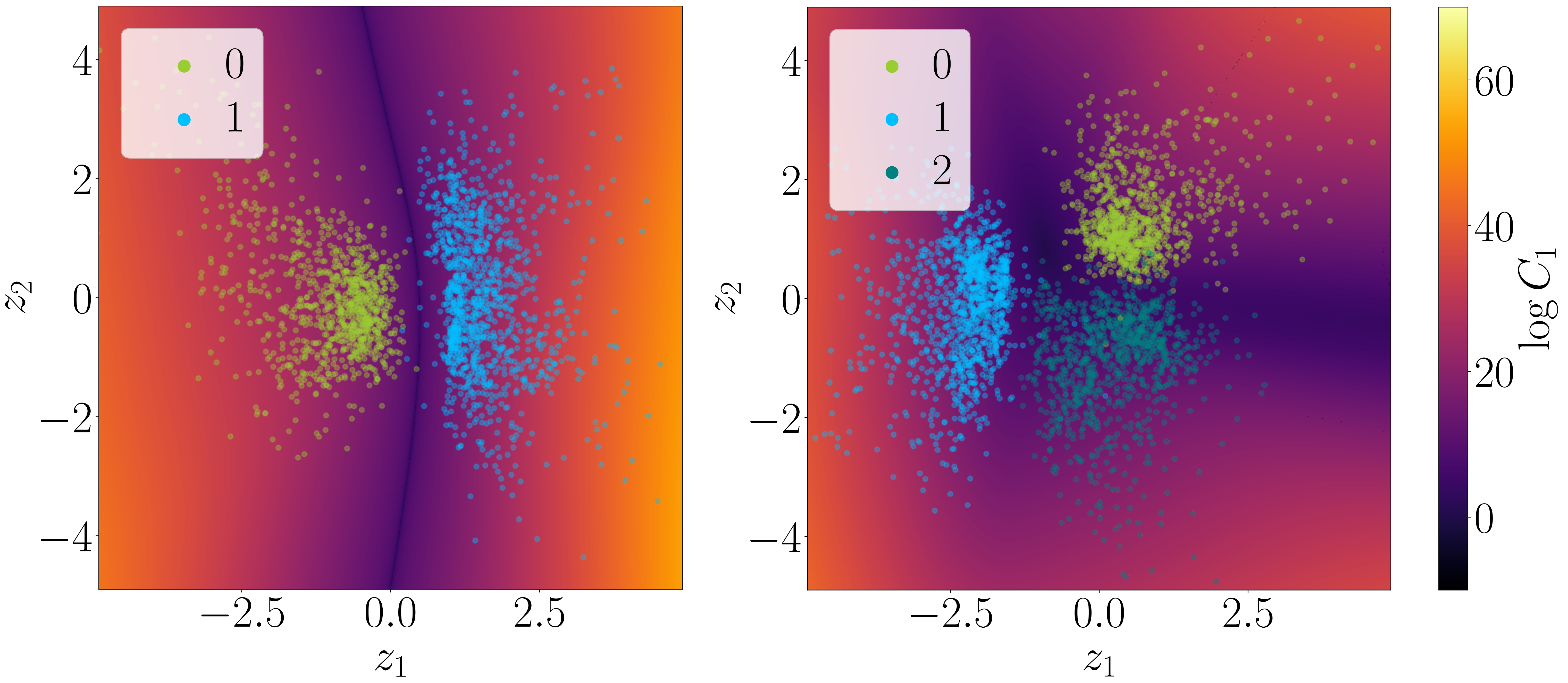
- E.g. full contraction $C_1 = C_{ijk}C^{ijk}$
- Inspiration: metric affine gravity, Weyl gravity.

Classifier Nonmetricity

New: Nonmetricity scalars

- E.g. full contraction $C_1 = C_{ijk}C^{ijk}$
- Inspiration: metric affine gravity, Weyl gravity.

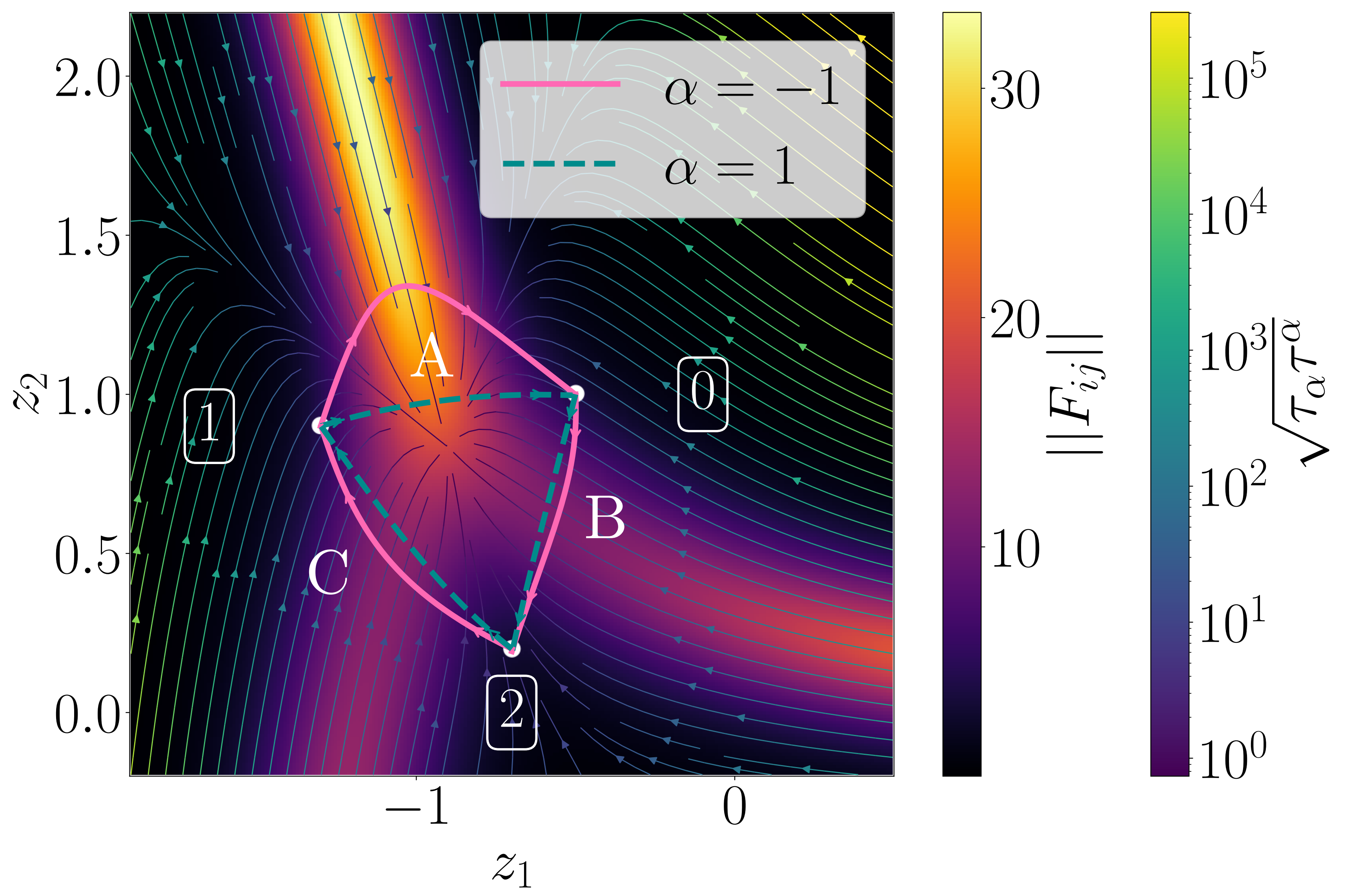
> Reflects a different part of the classifier decision
>
> Coordinate-invariant
>
> No curvature



"On average, how asymmetric is the pull coming from the different classes?"

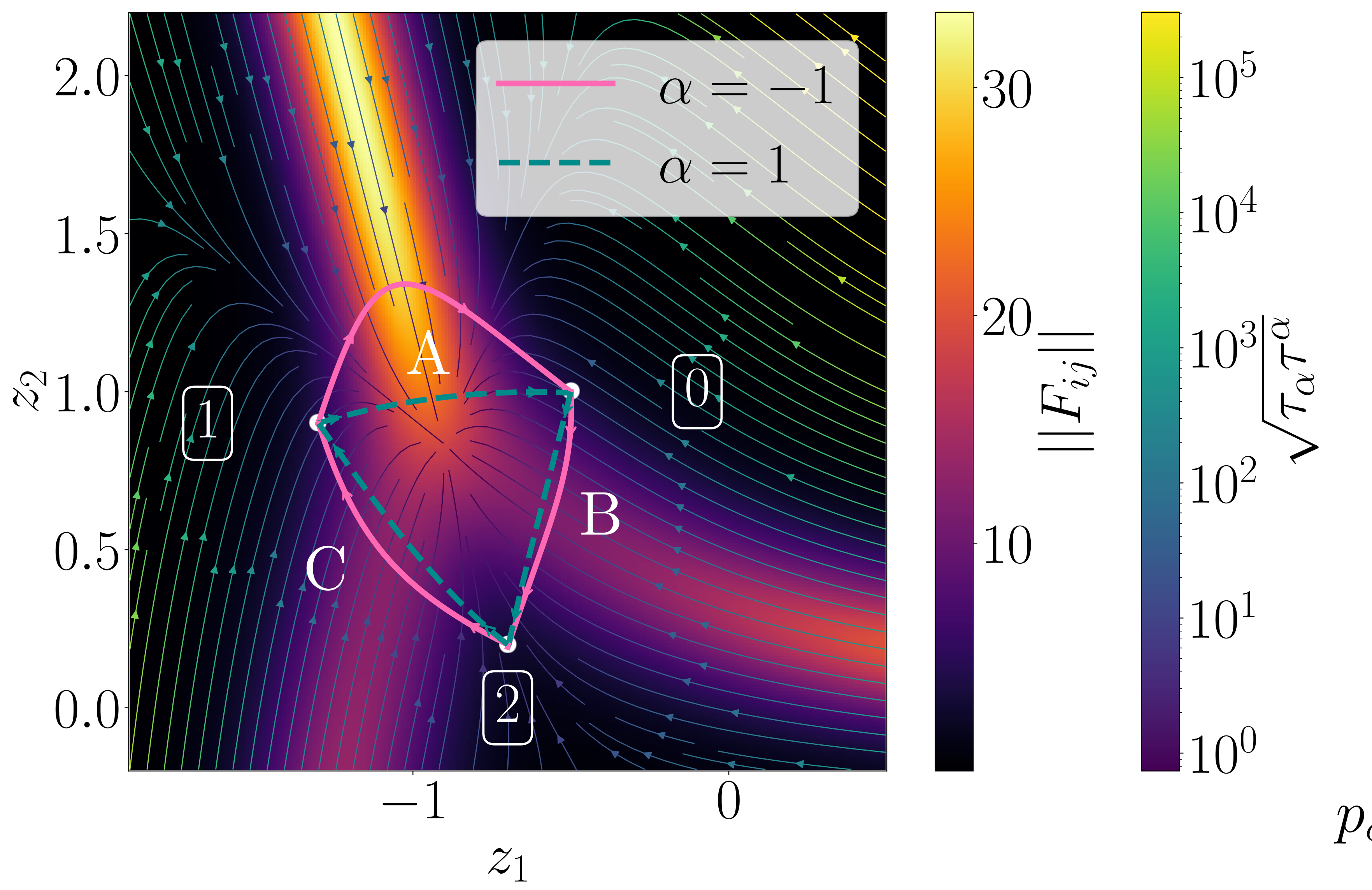
Geometry VI

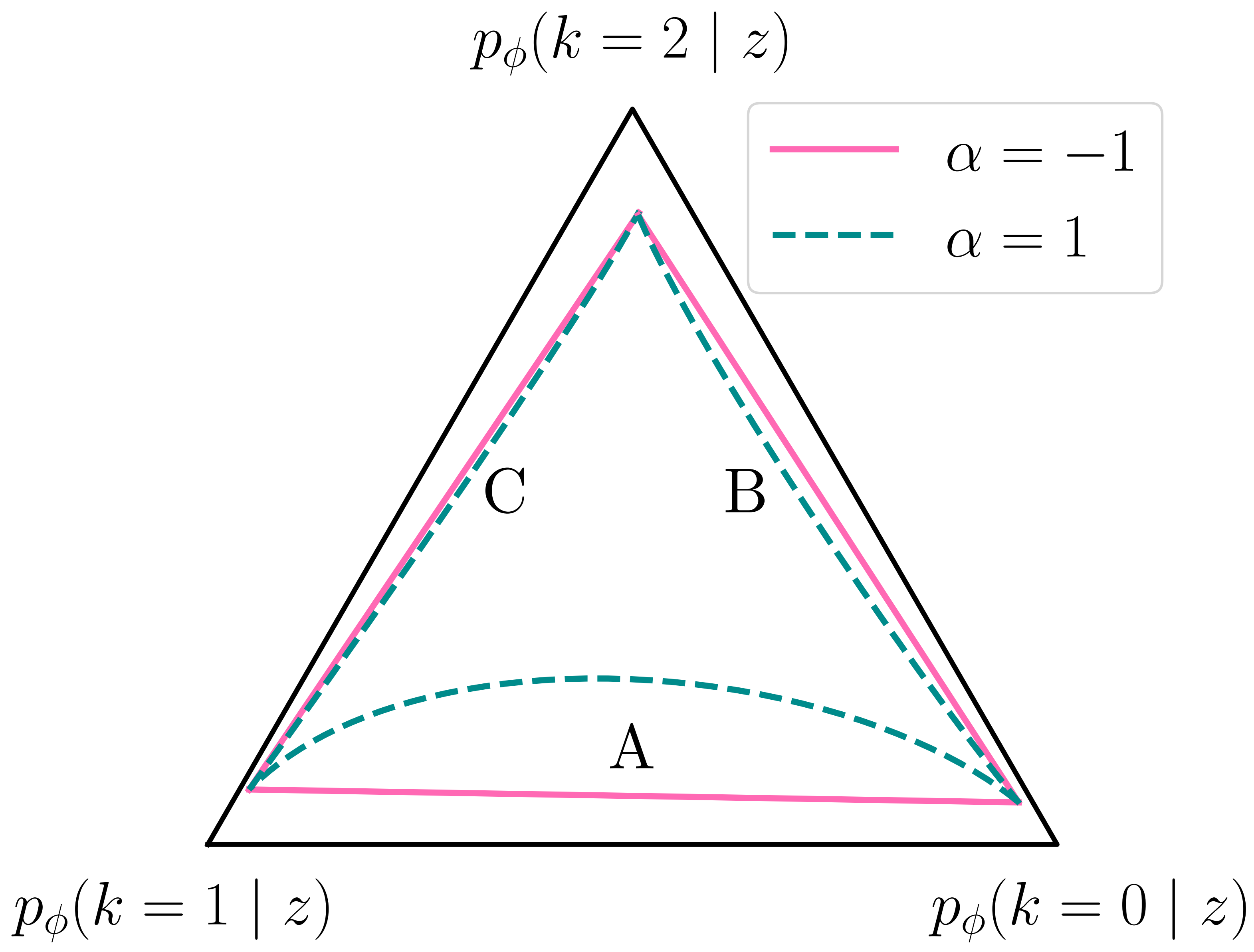
Dual Paths in the Classifier Geometry

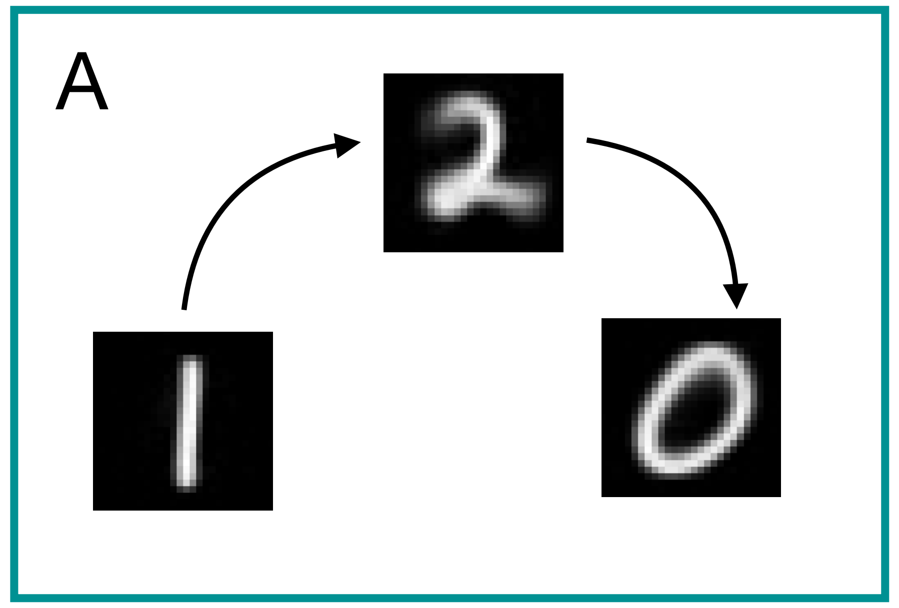


Geometry VI

Dual Paths in the Classifier Geometry

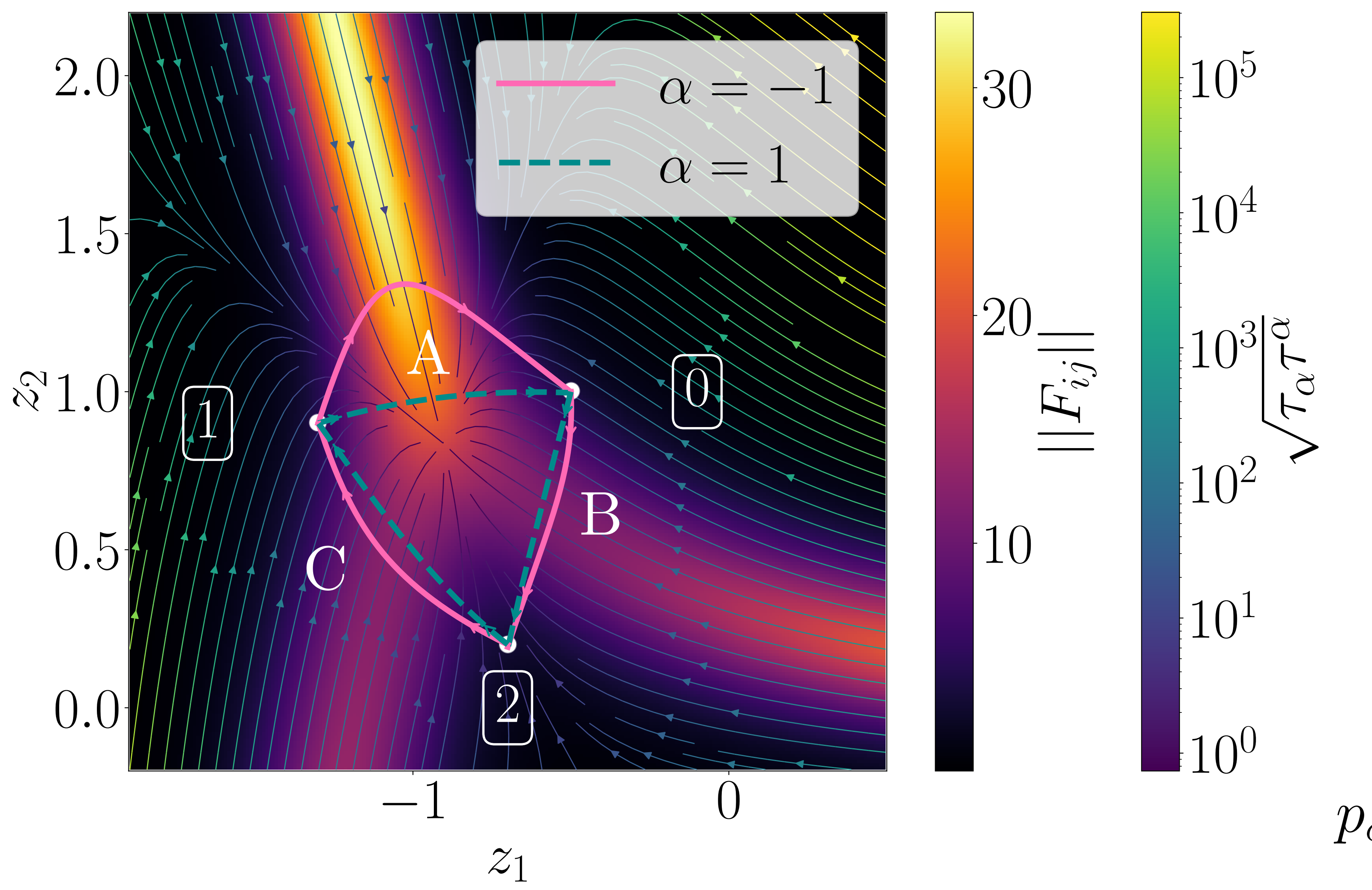


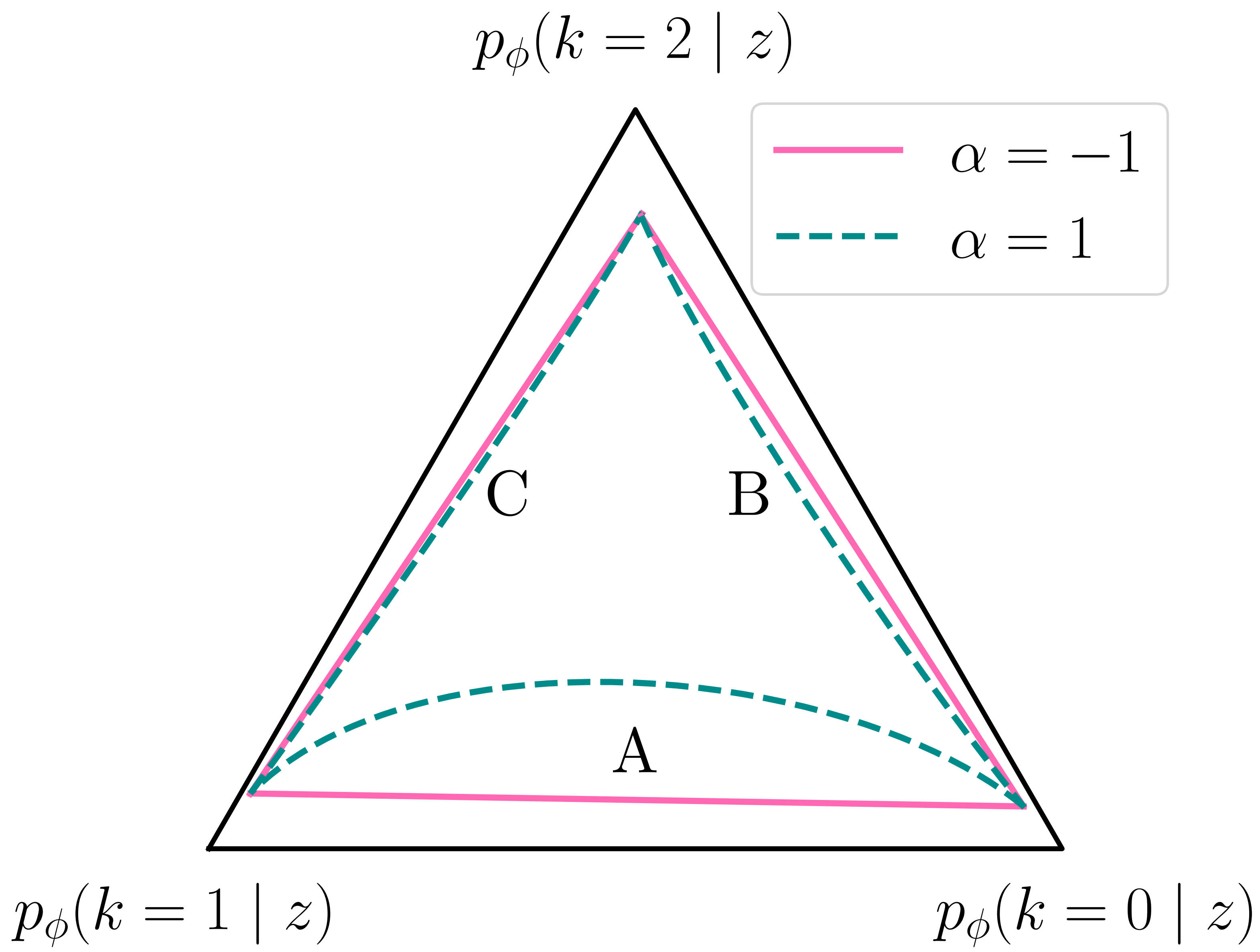


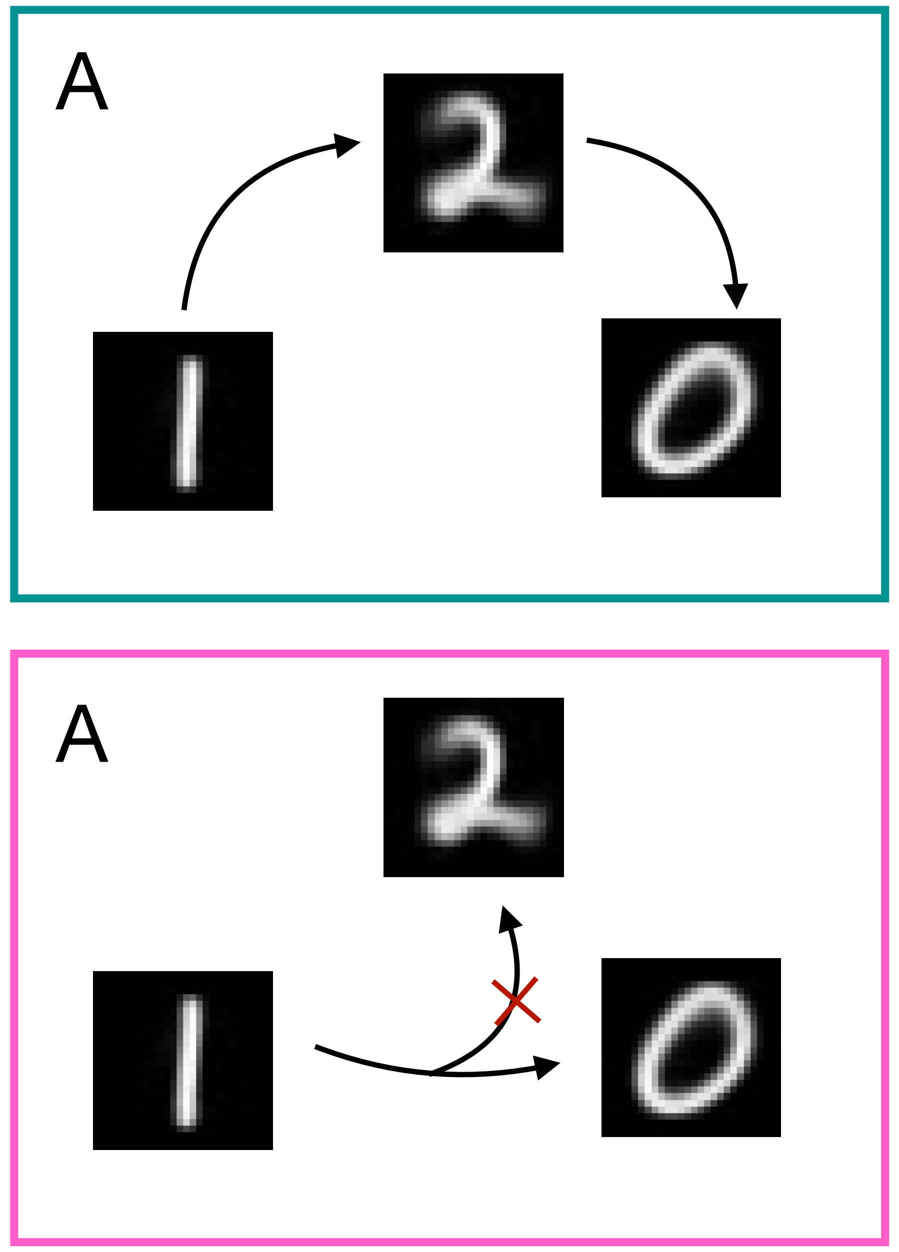


A: (+1)-path has intermediate state: $1 \rightarrow 2 \rightarrow 0$

Dual Paths in the Classifier Geometry







A: (+1)-path has intermediate state: $1 \to 2 \to 0$

A: (-1)-path avoids intermediate state: $1 \to 0$

Take-Home

What can the geometry tell us?

Read the paper :)

- ★ Leading observables which most efficiently change the classification.
- ★ Show the alignment between classifier and feature geometry.
- ★ Reveal interpretable paths and intermediate states between events.

Thank you!

References

Kuntz, Plehn, Schäfer, Schosser, Vent, “The Latent Information Geometry of Jet Classification”, arXiv:2603.02310 (2026).

Matsuzoe, Takeuchi, Amari, “Equiaffine structures on statistical manifolds and bayesian statistics”, Differential Geometry and its Applications 24(6), 567(2006).

Calin, Udriste, Geometric Modeling in Probability and Statistics, Springer, (2014).

Nielsen, “An Elementary Introduction to Information Geometry”, Entropy 22(10), 1100 (2020).

Amari, Information Geometry and Its Applications, Springer, (2016).

Ay, Jost, Vân Lê, Schwachhöfer, Information Geometry, Springer, (2017).

Nielsen, “The many faces of information geometry”, Notices of the American Mathematical Society 69(1), 36 (2022).