Geometric Inductive Biases of Deep Networks: The Role of Data and Architecture

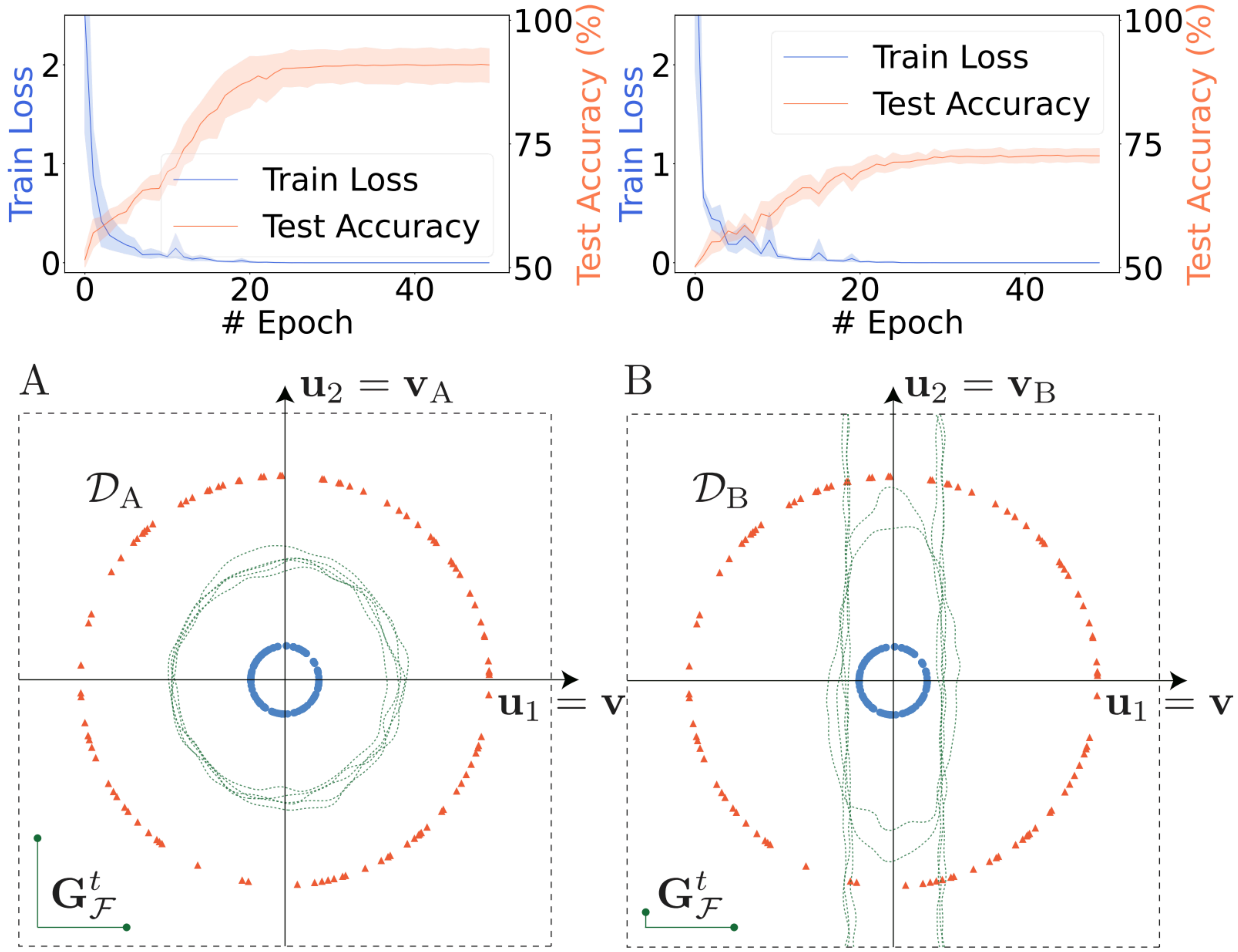
ICLR 2025 - Spotlight (top 5%)

Sajad Movahedi[1,2], Antonio Orvieto[1,2], Seyed-Mohsen Moosavi-Dezfooli[3]
[1]ELLIS Institute Tuebingen, [2]MPI for Intelligent Systems, [3]Apple (Work done while at Imperial College London)

A Motivating Example

The Doughnut

- Each circle corresponds to a class.
- The circles reside on a plane.
- The plane on **high-dimensional** space.
- A ResNet generalizes only when the plane is on a specific subspace.
- An MLP generalizes regardless of the orientation of the plane.



Definitions

Compact architecture-dependent summary of geometry

- **Average geometry** of family of models $\mathcal{F}$ (*architecture*) with $f_\theta(\,.\,) \in \mathcal{F}$:

$$\mathbf{G}_{\mathcal{F}}(x) = \mathbb{E}_\theta \left[\nabla_x f_\theta(x) \, \nabla_x f_\theta(x)^\top \right].$$

  - Geometry of a neural network
  - Related to loss Hessian w.r.t. the input and local decision boundaries

- **Average geometry evolution** $\Delta_{\mathcal{F}}(x)$: the derivative of average geometry w.r.t. time (gradient-flow model).
  - **Changes** in the geometry of a neural network

The Role of Data and Architecture

At initialization

- Conjecture 1 **(at initialization)**:

$$\mathbb{E}_{x\sim\mathcal{N}(0,\mathbf{I})}\left[\Delta_{\mathcal{F}}(x)\right] \propto \mathbb{E}_{x\sim\mathcal{N}(0,\mathbf{I})}\left[\mathbf{G}_{\mathcal{F}}(x)\right] \quad \overbrace{\mathbb{E}_{x_\mu\in\mathcal{D}}\left[x_\mu x_\mu^\top\right]}^{\textbf{Data}} \quad \underbrace{\mathbb{E}_{x\sim\mathcal{N}(0,\mathbf{I})}\left[\mathbf{G}_{\mathcal{F}}(x)\right]}_{\textbf{Architecture}}$$

- At initialization, changes in geometry is the data support projected onto initial geometry.

- What if this remains the case later in training as well?

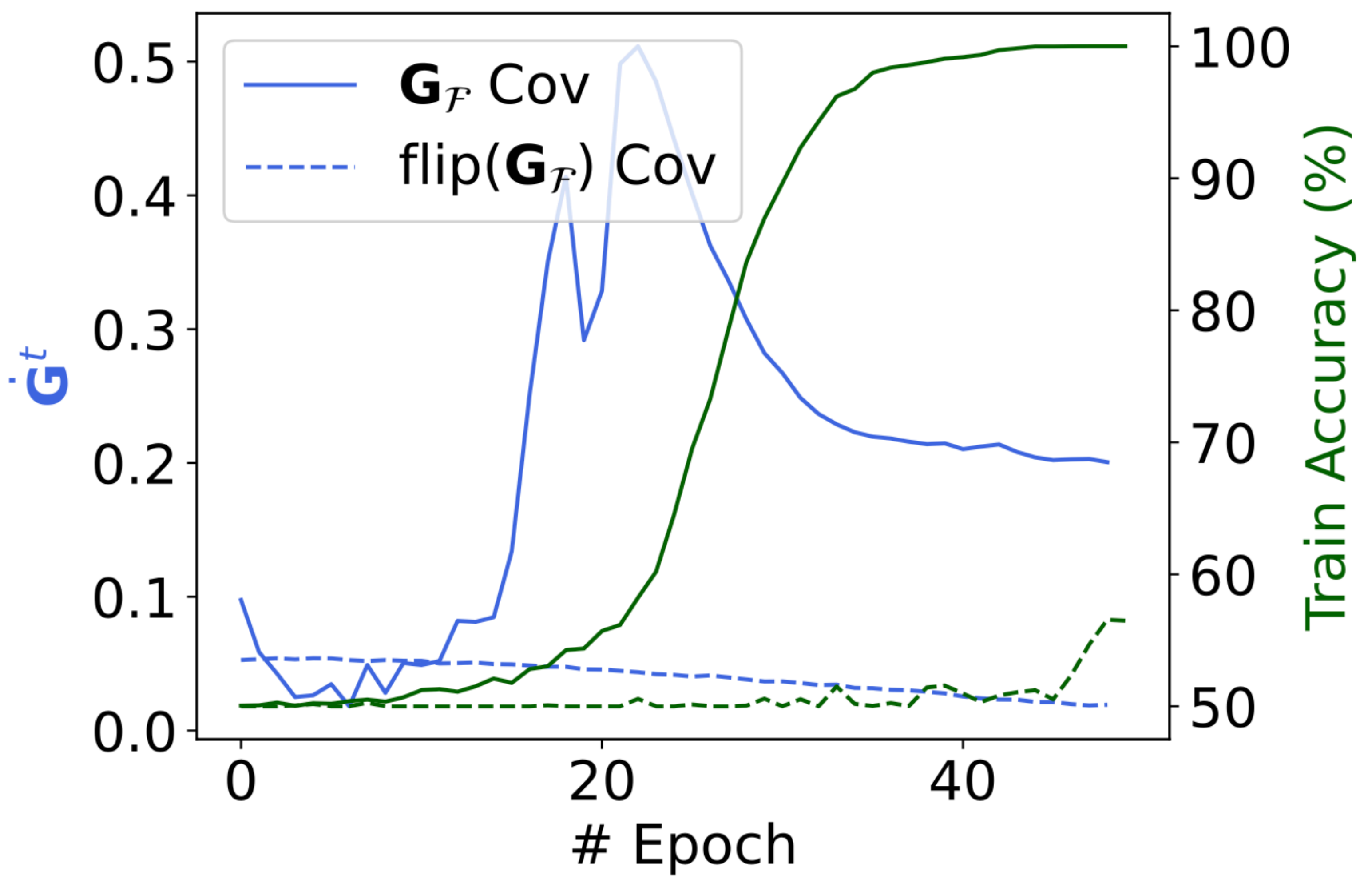
Geometric Invariance Hypothesis

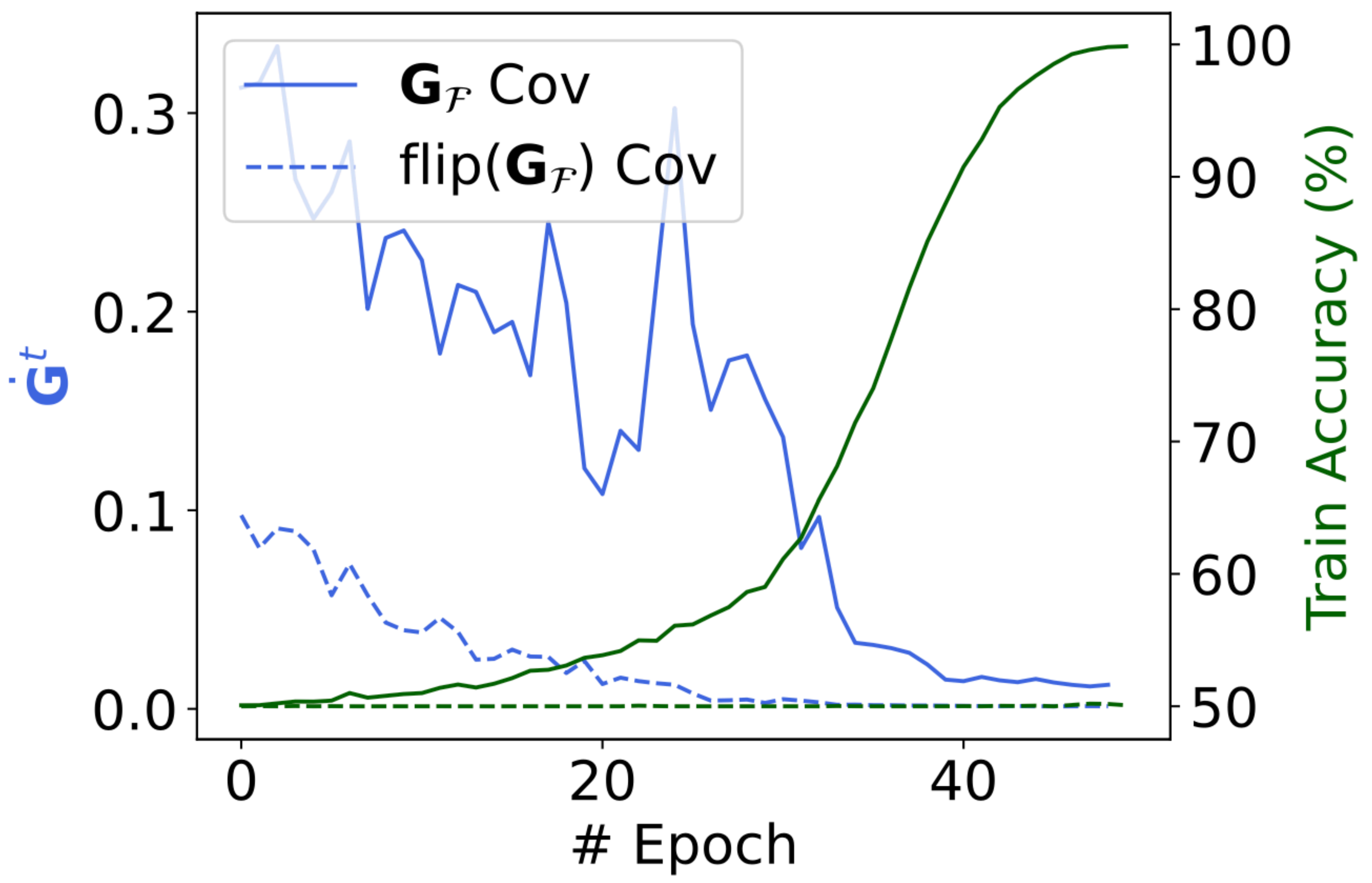
Linear projection of data support onto average geometry

- Let **Eig**$(\mathbf{A})$ be subspace defined by top eigenvalues of $\mathbf{A}$.
- The geometry remains **invariant** in directions not in **Eig** $\left(\mathbb{E}_{x\sim\mathcal{N}(0,\mathbf{I})}\left[\mathbf{G}_{\mathcal{F}}(x)\right]\right)$.

ResNet18

ViT





A Motivating Example

The Doughnut

- The plane defined by orthogonal vectors $(\mathbf{u}_1, \mathbf{u}_2)$.
- $\mathscr{D}_A$: both $\mathbf{u}_1$ and $\mathbf{u}_2$ are in **Eig** $\left(\mathbb{E}_{x\sim\mathcal{N}(0,\mathbf{I})}\left[\mathbf{G}_{\mathcal{F}}(x)\right]\right)$.
- $\mathscr{D}_B$: only $\mathbf{u}_1$ is in **Eig** $\left(\mathbb{E}_{x\sim\mathcal{N}(0,\mathbf{I})}\left[\mathbf{G}_{\mathcal{F}}(x)\right]\right)$.