Neural Tangent Kernels Under Stochastic Data Augmentation

Joshua DeOliveira, Sajal Chakroborty, Walter Gerych, Elke Rundensteiner

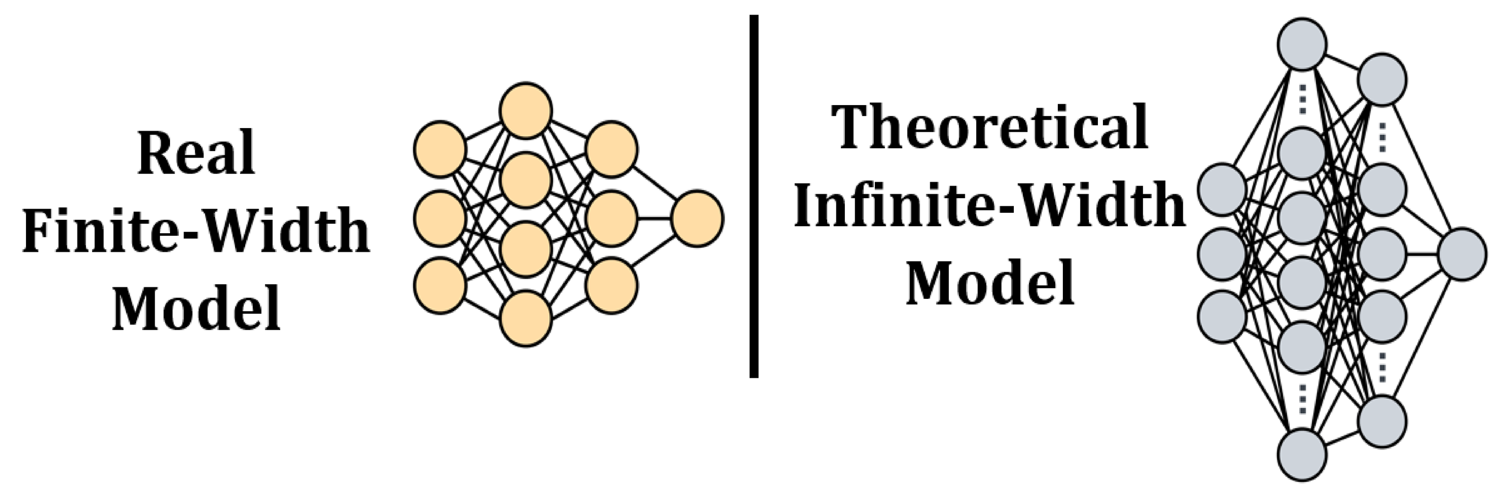
WPI

Association for the Advancement of Artificial Intelligence

Background

Understanding training remains a fundamental problem

Neural Tangent Kernels offer a tractable approach



Modern training involves more complicated training assumptions beyond the foundational assumptions

There is a need to analyzing NTKs under modern training regimes to better practically understand real training

Neural Tangent Kernels (NTKs)

A **similarity measure** of a neural net's sensitivity

$$\Theta_f(\mathbf{z}, \mathbf{x}; \theta_t) := \langle \nabla_{\boldsymbol{\theta}} f(\mathbf{z}; \theta_t), \nabla_{\boldsymbol{\theta}} f(\mathbf{x}; \theta_t) \rangle \quad [1]$$

For a loss *C*, the NTK describes function dynamics

$$d_t f(\cdot; \theta_t) = -\nabla_{\Theta_f} \mathcal{C}|_{f(\cdot;\theta_t)} \quad [1]$$

Kernel-Preserving Augmentations

*How can we leverage **NTKs** to understand how data augmentations during training can perturb dynamics?*

$$g : \mathbb{R}^n \to \mathbb{R} \quad g(u; z, x, f) := \Theta_f(z, u) - \Theta_f(z, x)$$

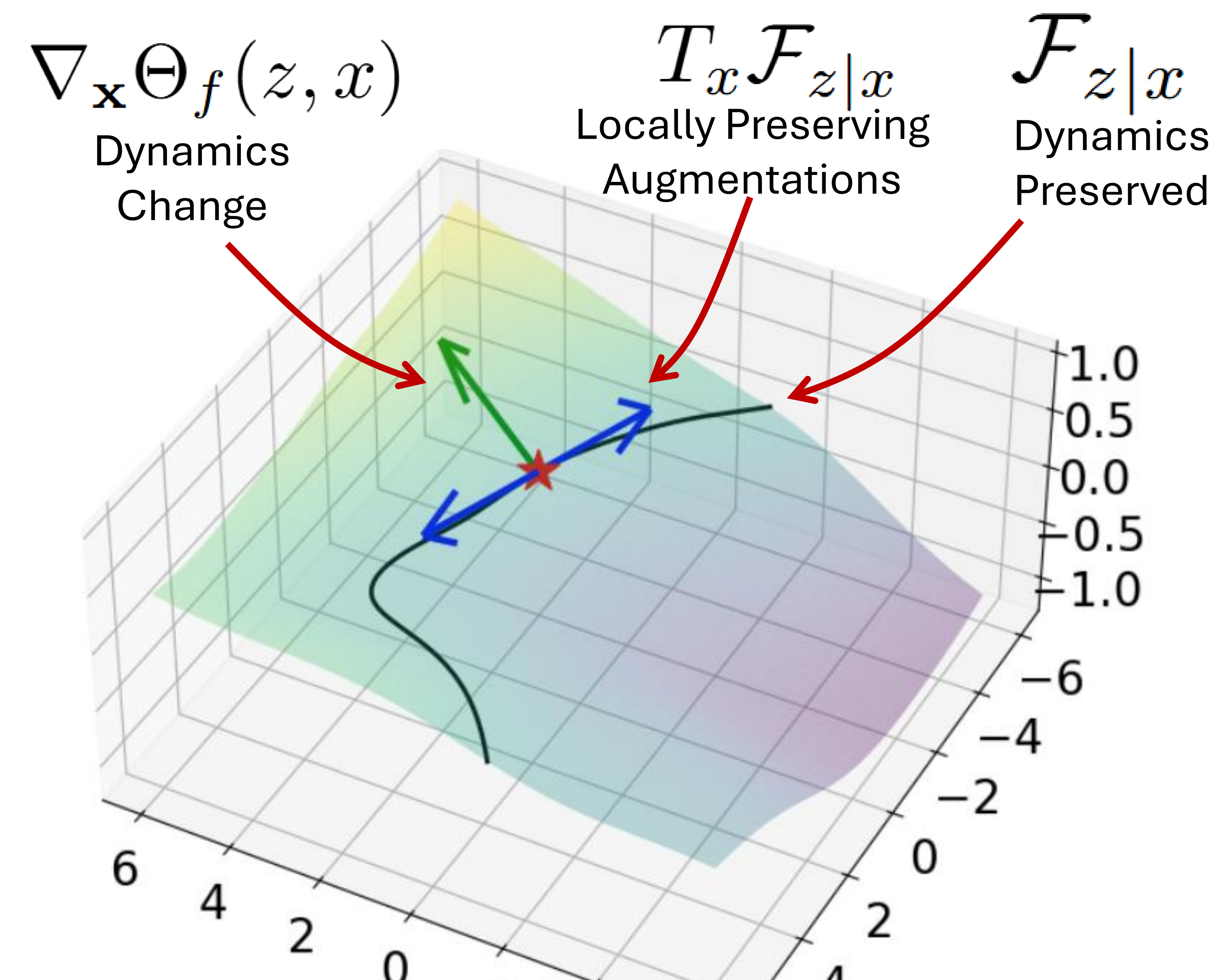
Definition 4.1. Let $z, x \in \mathbb{R}^n$. A *kernel fiber* under the NTK Θ_f is the pullback of the level set $\{0\} \subset \mathbb{R}$ via the map g, defined as $\mathcal{F}_{z|x} = g^{-1}(0; z, x, f)$.

By looking at the inverse of *g*, we can investigate elements on the same level-set that act identically under the NTK

Definition 4.3. The *local kernel fiber tangent space* is the pullback of the tangent distribution under Θ_f, given by $T_x\mathcal{F}_{z|x} = \{v \in \mathbb{R}^n : v^\top \nabla_x \Theta_f(z, x) = 0\}$ which defines local perturbation directions that preserve the kernel.

We also can consider the local directions to augment points to either stay on the level set or perturb off of it.

Empirical Results



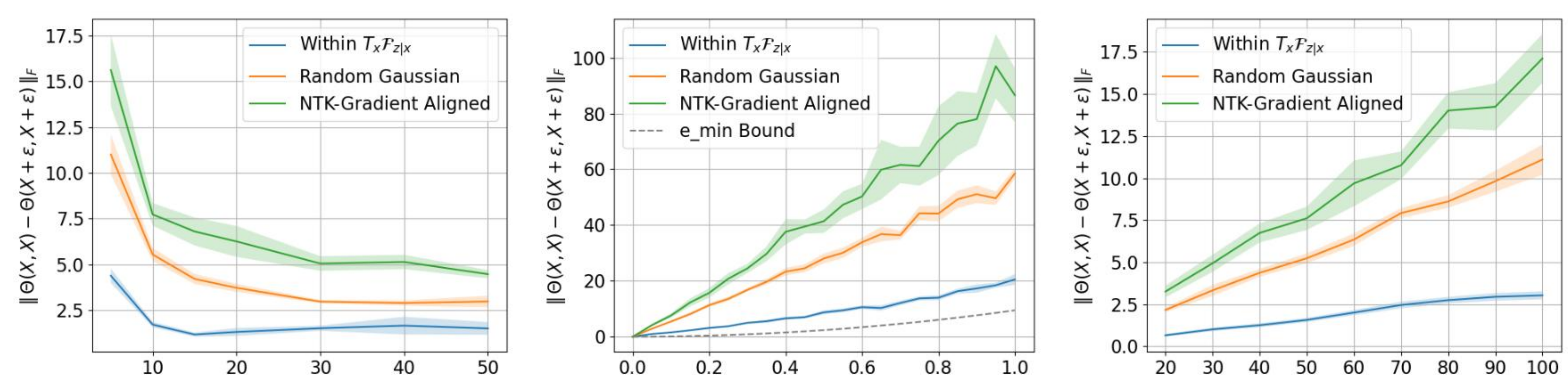
Visualization of g for a 2-layer GeLU network

$$\text{Relative Level Set Deviation} = \frac{\|\Theta(z,x) - \Theta(z,\hat{x})\|}{\Theta(z,x)}$$

Augmentation Technique	$\nabla_x\Theta$ Alignment	Relative Level Set Deviation
Rotation	$-0.062644_{\pm 0.023}$	$0.092406_{\pm 0.085}$
Hflip	$-0.080352_{\pm 0.020}$	$0.023625_{\pm 0.021}$
Vflip	$-0.092738_{\pm 0.021}$	$0.042702_{\pm 0.034}$
Cutout	$-0.008870_{\pm 0.016}$	$1.708538_{\pm 1.121}$
Mixup	$-0.104793_{\pm 0.031}$	$0.227315_{\pm 0.239}$
Uniform Gaussian	$-0.003340_{\pm 0.006}$	$0.211346_{\pm 0.086}$
$T_x\mathcal{F}_{z\vert x}$ (Eq. 10)	$-0.001090_{\pm 0.003}$	$0.022298_{\pm 0.010}$
$\nabla_x\Theta$ (Ground Truth)	$0.994879_{\pm 0.015}$	$0.189625_{\pm 0.149}$

Conventional augmentations lie near dynamic-preserving directions but yield high relative level set deviation leading to chaotic control over dynamics, whereas **NTK-based** augmentations yield **consistent** and **predictable** control.

Data augmentations that are conducted along first-order approximations of the local kernel fiber **(blue)** can produce novel datasets that yield **near-identical** training dynamics



(a) Input Dimension (b) Noise Magnitude (c) Training Set Volume

Takeaways

Data augmentations via NTKs exhibit uniquely tunable properties of neural network function dynamics beyond what conventional data augmentation regimes provide

NTK-Inspired data augmentations offer a promising avenue to expand the capabilities of more sophisticated augmentation strategies that control training for user-specified intents

Acknowledgements
DAISY Research Lab
NSF Grant DMS-1337943
NRT-HDR-2021871

References:

[1] Jacot, A., Gabriel, F., & Hongler, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems*, *31*.