Staying on the Manifold: Geometry-Aware Noise Injection

Albert Kjøller Jacobsen*, Johanna Marie Gegenfurtner* and Georgios Arvanitidis
Section for Cognitive Systems – Technical University of Denmark
*Equal contribution, listed in arbitrary order

Motivation

Training ML models with input noise can **improve robustness** and influence **generalisation**.

Bishop [1] proves that training with **Gaussian input noise penalises the trace of the Hessian** in expectation:

$$\mathbb{E}_{\boldsymbol{\epsilon}}\left[\mathcal{L}\left(\boldsymbol{x}+\boldsymbol{\epsilon}\right)\right]=\mathcal{L}\left(\boldsymbol{x}\right)+\frac{\sigma^2}{2}\Delta_x\mathcal{L}\left(\boldsymbol{x}\right),$$

where $\boldsymbol{\epsilon}\sim\mathcal{N}\left(\boldsymbol{0},\eta^2\mathbb{I}_D\right).$

RQ: Can we improve **generalization** by adding **geometry-aware noise** to the data?

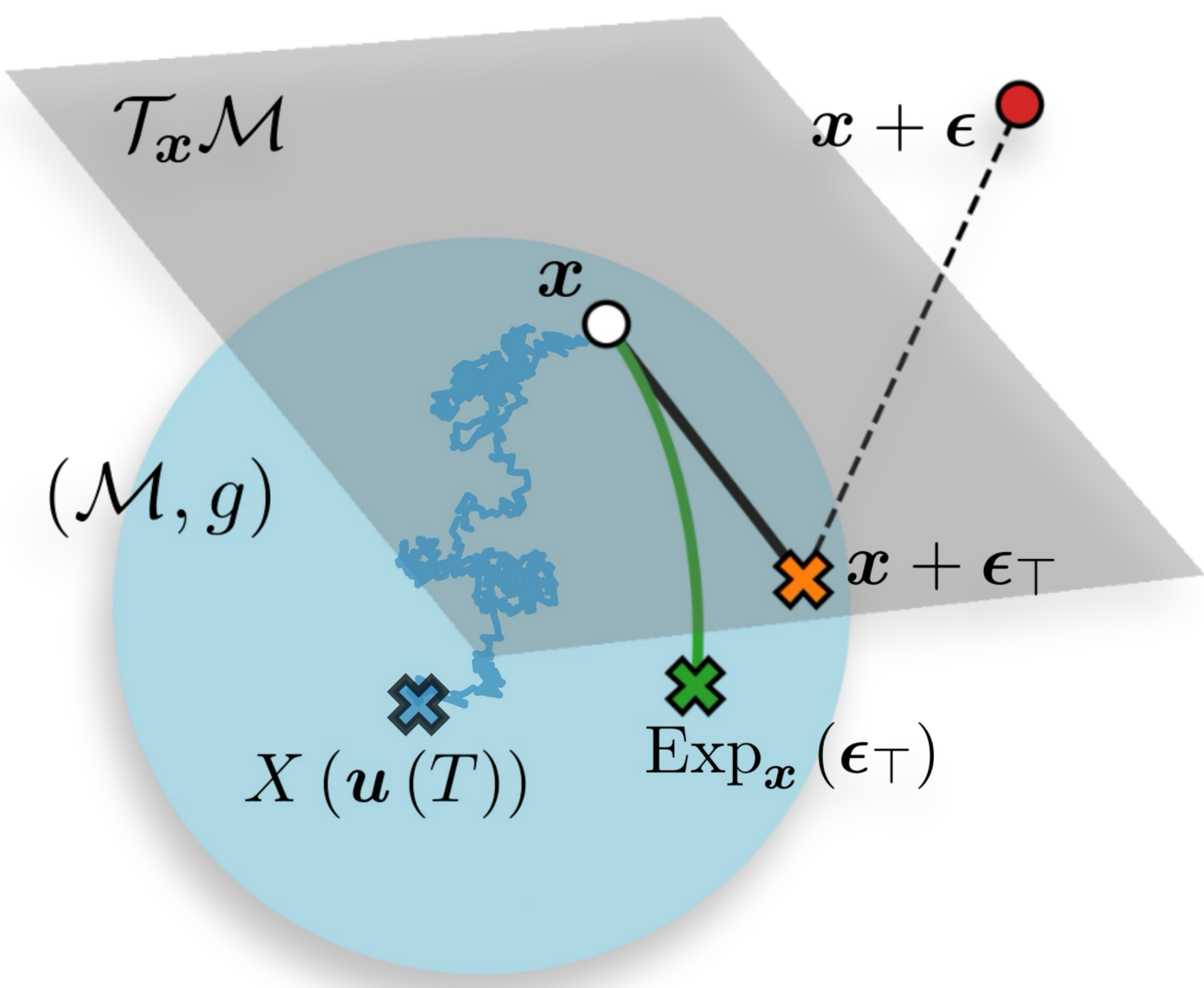
TL;DR: **Yes** – constraining noisy inputs to lie on the data manifold **can improve generalization**!

Noise Injection Strategies [2,3]

- **tangential**: Gaussian noise projected to the tangent space.
- **geodesic**: mapping tangent vectors to the manifold.
- **Brownian motion**: walking along the manifold randomly.

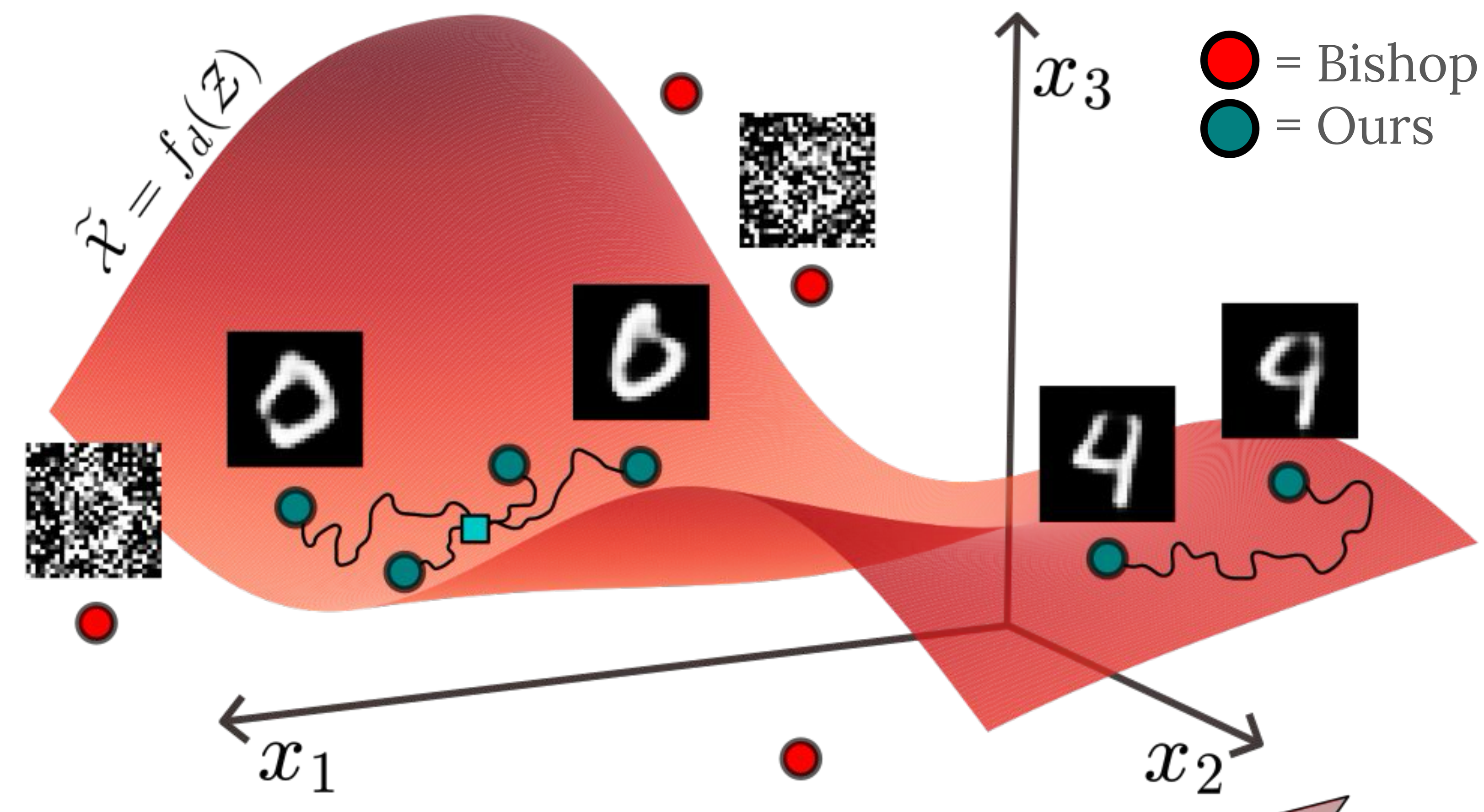
This penalises the tangential part of the Tikhonov regulariser:

$$R(\boldsymbol{x},\boldsymbol{\theta})=\frac{\sigma^2}{2}\sum_{n=1}^{N}\|\nabla_x f_\theta(\boldsymbol{x}_n)_\top\|^2$$



1. Bishop, "*Training with Noise is equivalent to Tikhonov regularization*", Neural Computation (1995)
2. Hsu, "*Brownian motion and Riemannian geometry.*" Contemp. Math 73 (1988)
3. Girolami & Calderhead, "*Riemann manifold Langevin and Hamiltonian Monte Carlo methods*". 2011

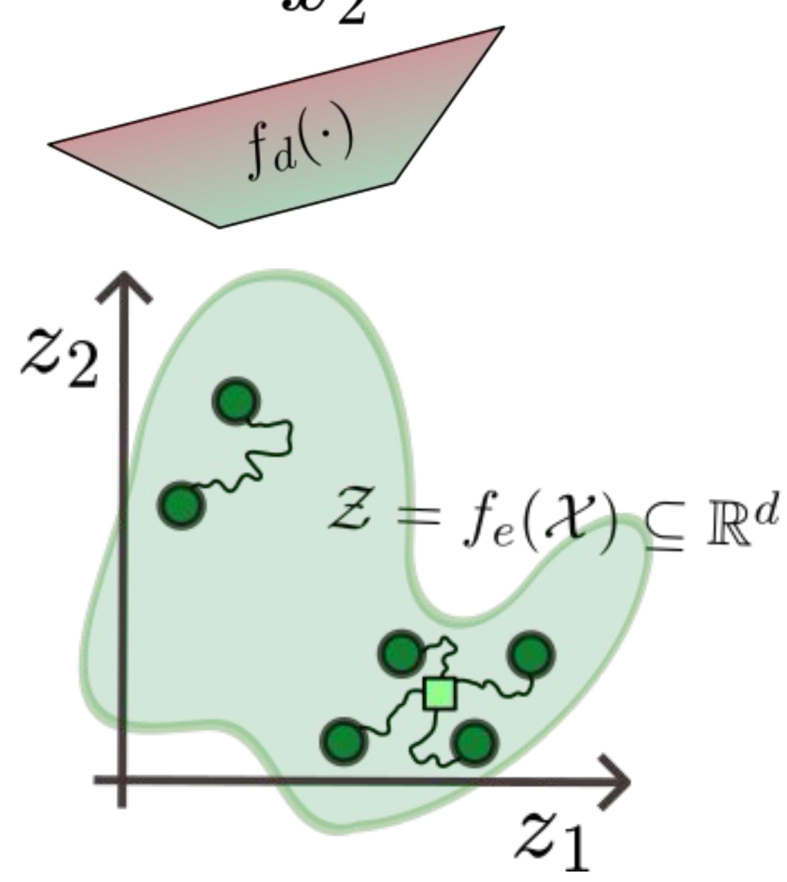
Approximating the Data Manifold

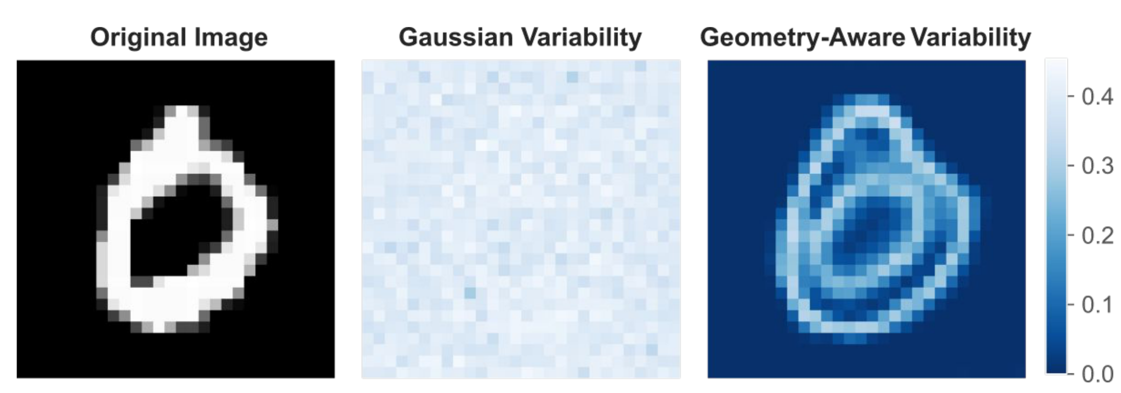


We don't know the real data manifold, but we can approximate it:

- using a **generative model**, or
- using an **autoencoder.**

Our **augmented samples** are **natural variations** of the original sample!





Results

On MNIST, our geometry-aware strategy:

- **improves** in highly overparameterised settings,
- suffers from a **manifold approximation gap,**
- **yet consistently improves** over training on **reconstructions**.

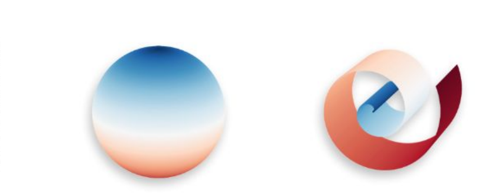
MNIST: Performance when subsampling the training data

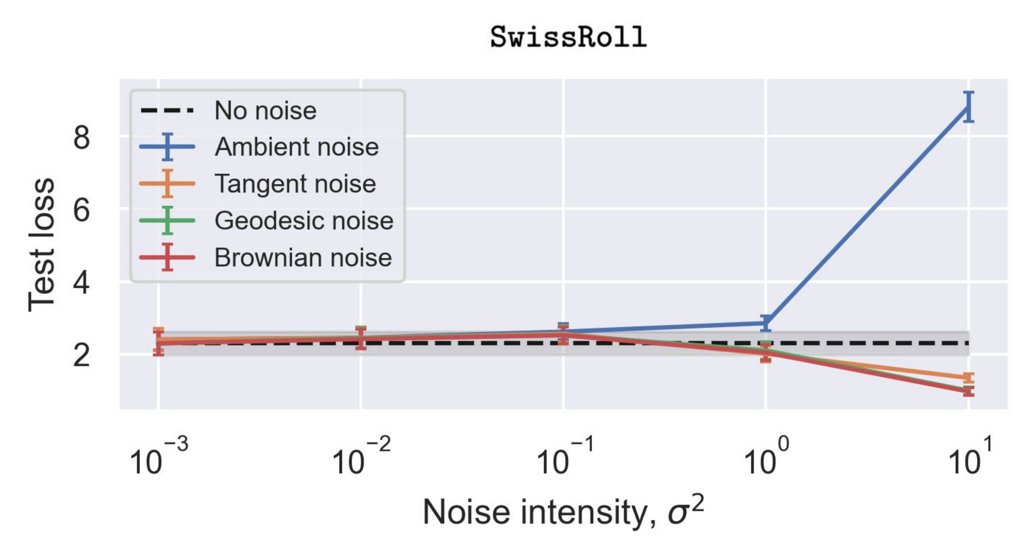
	1%	10%	50%
O	0.883 ± 0.008	0.956 ± 0.002	$\mathbf{0.981\pm0.001}$
A	0.883 ± 0.008	$\mathbf{0.965\pm0.001}$	$\mathbf{0.981\pm0.001}$
R	0.877 ± 0.005	0.943 ± 0.002	0.967 ± 0.002
BM	$\mathbf{0.896\pm0.008}$	0.959 ± 0.002	0.971 ± 0.001

Improved performance on "wiggly"/curved toy manifolds:

	Sphere	SwissRoll
B	1.00 ± 0.16	1.00 ± 0.18
A	$\mathbf{0.91\pm0.10}$	1.00 ± 0.19
T	0.98 ± 0.14	0.62 ± 0.07
G	1.00 ± 0.16	0.47 ± 0.06
BM	1.00 ± 0.16	$\mathbf{0.46\pm0.06}$

Manifold





Northern Lights Deep Learning Conference 2026

OTTO MØNSTEDS FOND
INDEPENDENT RESEARCH FUND DENMARK
Danish Data Science Academy