Deep Neural Networks as Finite-Step Hopfield Dynamics

Wooyul Jung, Hochan Bang, Hee Bin Yoo, Byoung−Tak Zhang
Interdisciplinary Program in Artificial Intelligence, Seoul National University

ICLR

Deep Neural Networks as Finite-Step Hopfield Dynamics

Assumption $T(v) = M^T \sigma(Mv), \quad v_{k+1} = T(v_k)$ Weight-shared iterative 2-layer Network

Hopfield Energy $E(v) = \frac{1}{2} \| v \|_2^2 - \Phi(Mv), \quad \sigma(u) = \nabla \Phi(u)$ Activation function σ

Energy Descent $\nabla E(v) = v - \alpha T(v)$ Step size $\alpha > 0$

$$v_{k+1} = v_k - \alpha \nabla E(v_k) = (1-\alpha) v_k + \alpha T(v_k).$$

Activation $\sigma(u)$	Potential $\Phi(u)$	Hessian $H_\Phi(u)$	L_Φ
$\tanh(\beta u)$	$\sum_i \frac{1}{\beta} \log\cosh(\beta u_i)$	$\mathrm{diag}(\beta(1-\tanh^2(\beta u_i)))$	β
$\mathrm{sigmoid}(\beta u)$	$\sum_i \frac{1}{\beta} \log(1+e^{\beta u_i})$	$\mathrm{diag}(\beta s_i(1-s_i))$	$\beta/4$
$\mathrm{ReLU}(u)$	$\sum_i \frac{1}{2} \mathrm{ReLU}(u_i)^2$	$\mathrm{diag}(\mathbb{I}[u_i > 0])$	1

$$\| J_T(v) \|_2 \leq \| M \|_2^2 \| H_\Phi(Mv) \|_2 \leq L_\Phi \| M \|_2^2$$

$$\nabla^2 E(v) = I - J_T(v)$$

Conditions for Stable Energy Descent

$$J_T(v) = M^T H_\Phi(Mv) M$$

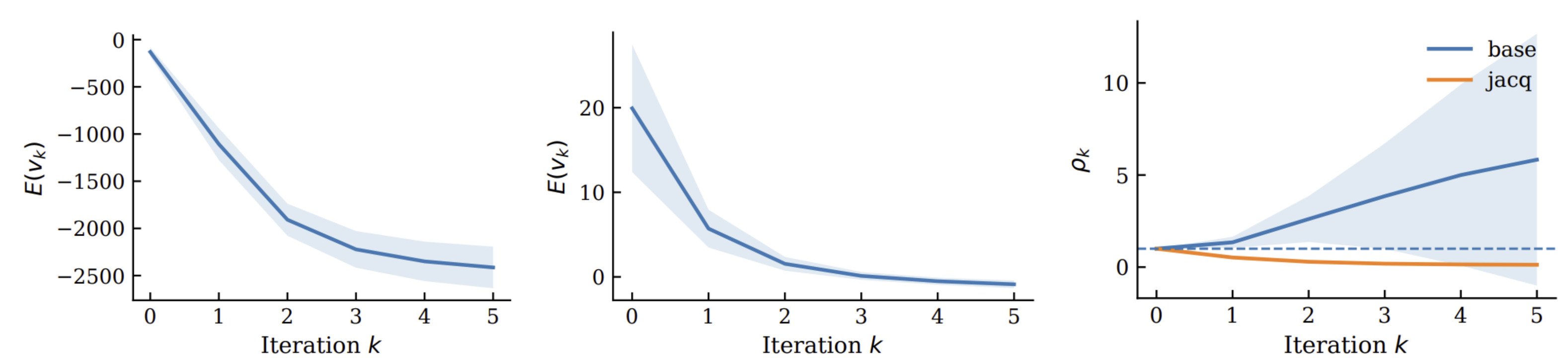
$$\| J_T(v_k) \|_2 \leq m, \quad 0 < \alpha \leq \frac{1}{1+m}$$

$$\Rightarrow E(v_{k+1}) \leq E(v_k)$$

$$R_{JacQ} = \lambda \left[\max(J_q - m, 0)\right]^2$$

- **The energy is jointly defined by the weight matrix *W* and the activation-induced potential Φ** encodes learned interactions, while the activation shapes the nonlinear energy landscape.
- Under this formulation, each layer update is interpreted not as a static mapping but as **a gradient descent step on the induced energy**, so network depth becomes finite-step dynamical evolution in state space.
- This perspective shifts **stability analysis from fixed-point convergence to trajectory-level behavior**, where the Jacobian determines local smoothness, perturbation amplification, and whether the forward trajectory follows monotone energy descent.

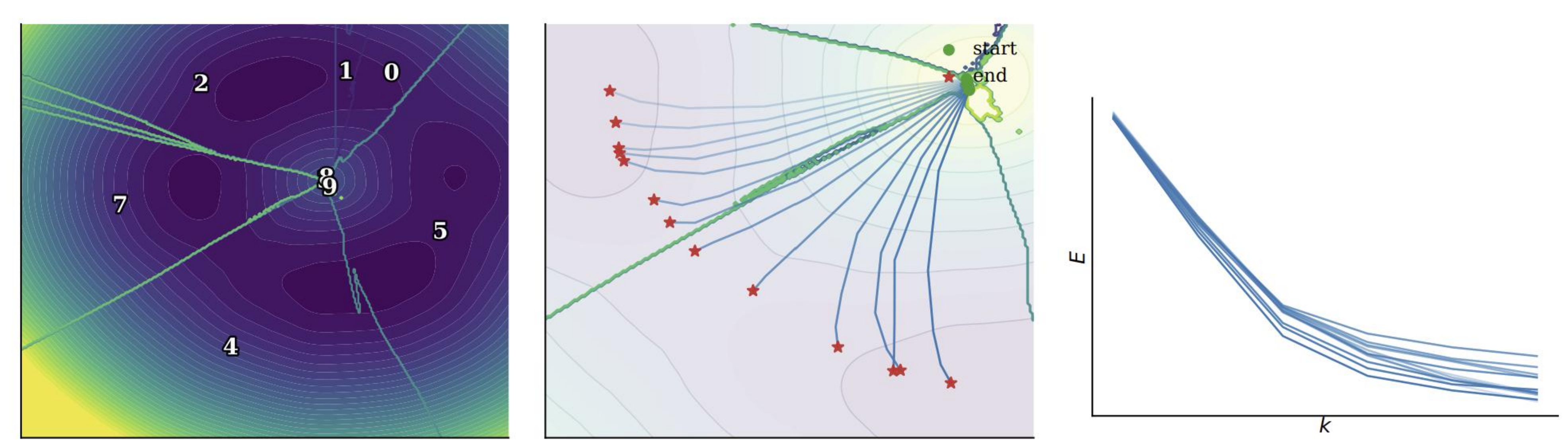
Classification Performance



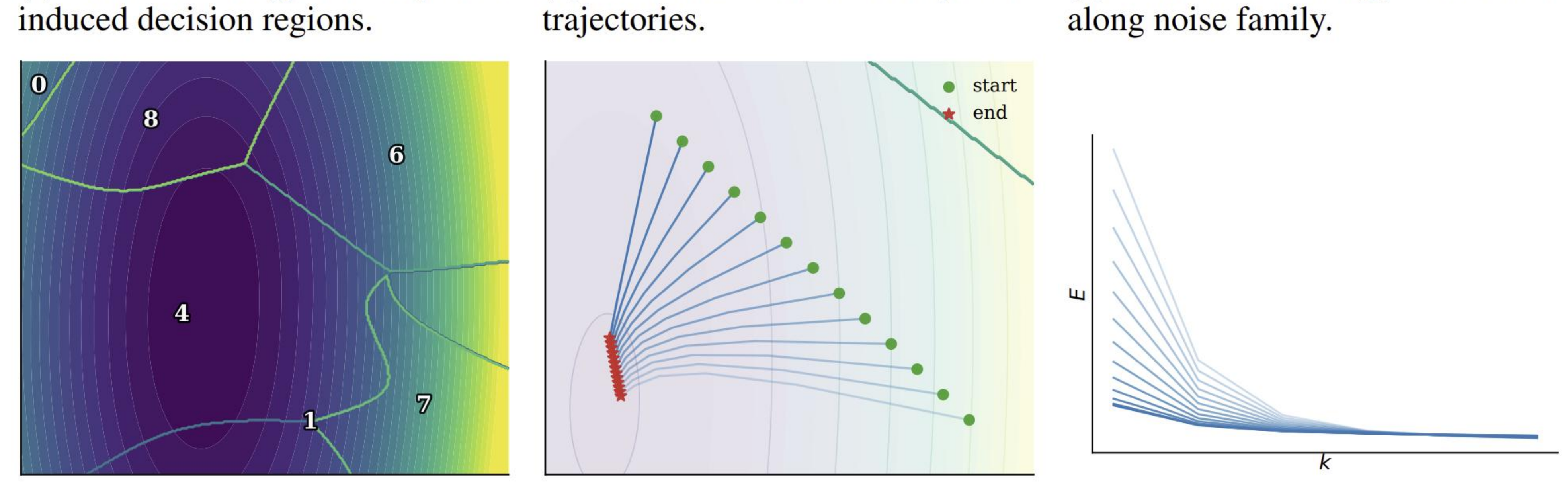
(a) Energy: baseline (b) Energy: JacQ (c) Distance ratio ρ_k

Activation	Method	$K=1$ Test	$K=1$ Noisy	$K=3$ Test	$K=3$ Noisy	$K=5$ Test	$K=5$ Noisy
ReLU	Base	97.90 ± 0.07	68.75 ± 2.70	97.51 ± 0.20	80.32 ± 3.63	97.17 ± 0.20	79.40 ± 3.51
	JacF	97.63 ± 0.22	90.13 ± 1.69	97.39 ± 0.14	65.60 ± 2.61	97.23 ± 0.17	66.93 ± 2.65
	Lip	98.06 ± 0.14	87.67 ± 2.87	97.78 ± 0.33	81.95 ± 2.71	97.33 ± 0.38	81.64 ± 2.23
	JacQ(Ours)	98.04 ± 0.33	**93.97** ± 1.02	97.84 ± 0.24	**90.56** ± 2.52	97.73 ± 0.16	**92.23** ± 1.28
Sigmoid	Base	97.74 ± 0.05	67.59 ± 3.29	97.52 ± 0.13	62.26 ± 2.71	97.04 ± 0.18	62.48 ± 2.71
	JacF	97.06 ± 0.23	72.79 ± 3.00	96.33 ± 0.20	66.68 ± 4.10	95.82 ± 0.42	64.79 ± 3.61
	Lip	98.06 ± 0.08	**83.88** ± 2.74	97.64 ± 0.21	72.33 ± 2.72	96.98 ± 0.33	65.58 ± 1.88
	JacQ(Ours)	97.30 ± 0.21	78.41 ± 5.45	96.75 ± 0.34	**74.18** ± 4.17	96.43 ± 0.31	**75.46** ± 5.63
Tanh	Base	97.64 ± 0.54	66.03 ± 2.43	97.26 ± 0.26	71.28 ± 3.38	96.30 ± 0.50	72.26 ± 2.08
	JacF	96.59 ± 0.29	79.19 ± 5.19	96.58 ± 0.37	71.39 ± 3.11	96.44 ± 0.28	72.99 ± 1.88
	Lip	98.05 ± 0.13	**85.12** ± 3.04	97.37 ± 0.27	76.48 ± 3.12	96.59 ± 0.40	72.84 ± 3.72
	JacQ(Ours)	97.25 ± 0.34	83.12 ± 3.34	96.74 ± 0.20	**77.99** ± 5.18	96.91 ± 0.33	**80.91** ± 5.37

Energy Geometry and Trajectories



(a) **Baseline** Energy landscape and induced decision regions. (b) **Baseline** Local landscape and trajectories. (c) **Baseline** Energy evolution along noise family.



(d) **JacQ** Energy landscape and induced decision regions. (e) **JacQ** Local landscape and trajectories. (f) **JacQ** Energy evolution along noise family.

Reference

Daniel Jakubovitz and Raja Giryes. Improving DNN Robustness to Adversarial Attacks using Jacobian Regularization. ECCV, 2018.
Beren Millidge, Tommaso Salvatori, Yuhang Song, Thomas Lukasiewicz, and Rafal Bogacz. *Universal Hopfield Networks: A General Framework for Single-Shot Associative Memory Models.* ICML, 2022.
Hubert Ramsauer et al. *Hopfield Networks is All You Need.* arXiv, 2020.

Contact