Pacific Northwest
NATIONAL LABORATORY

Dissipative Deep Neural Dynamical Systems

Third Symposium on Machine Learning and Dynamical Systems

Fields Institute, Toronto, Canada

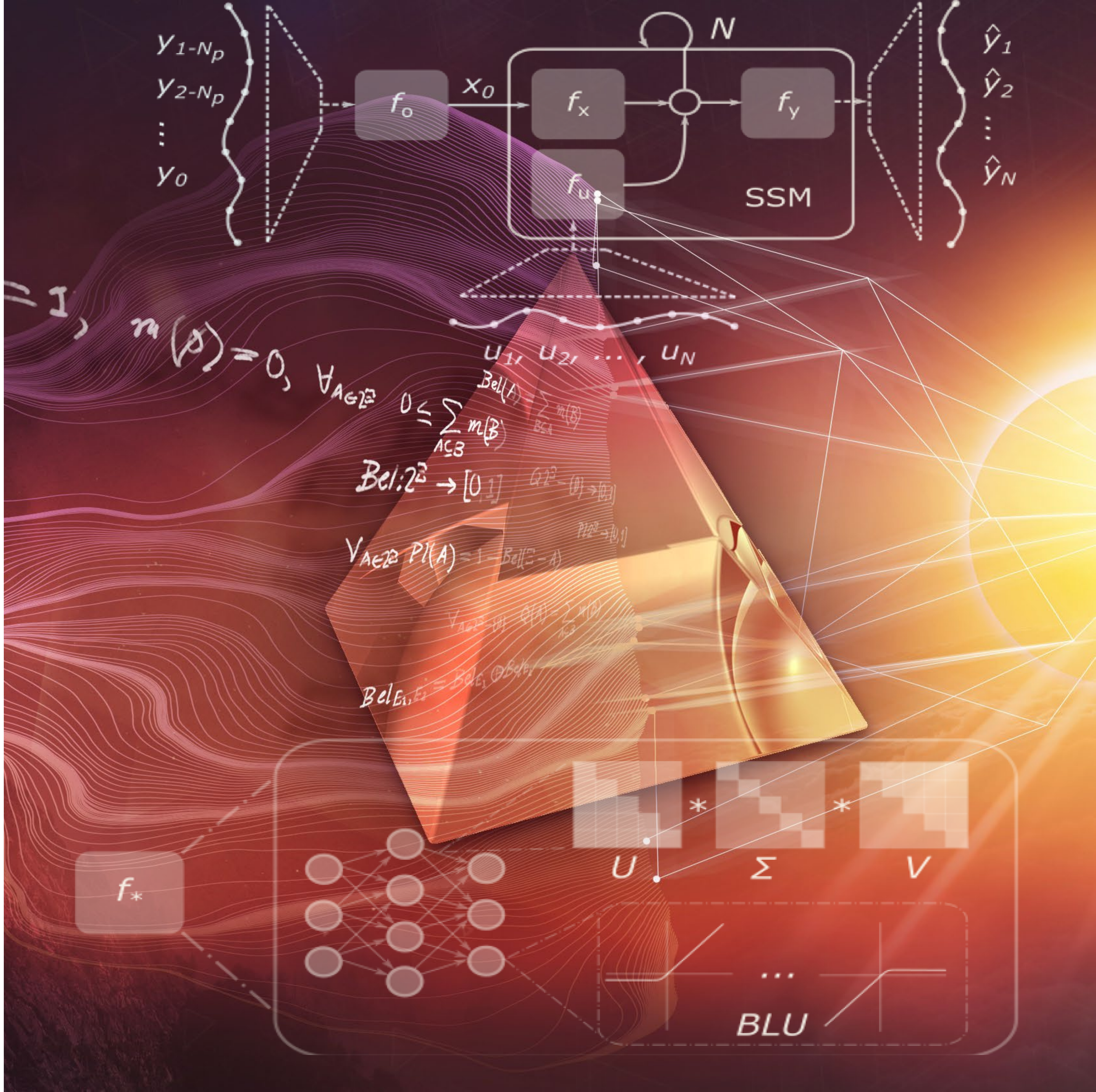
9/28/2022

Jan Drgona

Collaborators: Aaron Tuor, Soumya Vasisht, Mia Skomski, Draguna Vrabie

U.S. DEPARTMENT OF ENERGY BATTELLE

PNNL is operated by Battelle for the U.S. Department of Energy

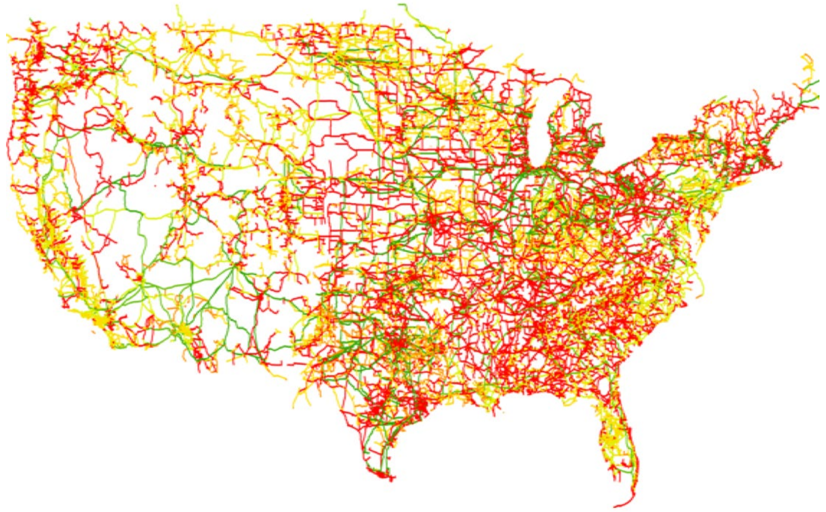


Motivation

- **Simulations** are crucial for many areas of decision-making and scientific discovery
- **Need**: Improve computational efficiency and scalability for heterogenous scientific simulations
- **Methods:** Integration of machine learning with physics-based models.
- **Applications**: Optimization, modeling and control of dynamical systems
- **Challenges**: Analysis, interpretability, rigorous performance and safety guarantees

Latest Neural Nets Solve World's Hardest Equations Faster Than Ever Before

Quantamagazine



Stability of Deep Neural Networks

Spectral approaches

- Eldad Haber and Lars Ruthotto, Stable Architectures for Deep Neural Networks, IOP, 2017
- S. Wang et al., “Analysis of deep neural networks with extended data Jacobian matrix”, ICML, 2016
- Z. Liao and R. Couillet, “The dynamics of learning: A random matrix approach”, ICML, 2018

Contraction and Lipschitz approaches

- M. Fazlyab, et al., “Efficient and accurate estimation of Lipschitz constants for deep neural networks”, NeurIPS, 2019,
- M. Revay and I. R. Manchester, “Contracting implicit recurrent neural networks: Stable models with improved trainability”, L4DC, 2020
- P. Pauli, et al., “Training robust neural networks using Lipschitz bounds”, IEEE Con. Sys. Lett., 2022

Lyapunov and Potential function approaches

- G. Manek and J. Z. Kolter, “Learning stable deep dynamics models”, NeurIPS, 2019
- Spencer M. Richards, Felix Berkenkamp, Andreas Krause, “The Lyapunov Neural Network: Adaptive Stability Certification for Safe Learning of Dynamical Systems”, CoRL 2018
- A. Sosanya and S. Greydanus, “Dissipative Hamiltonian neural networks: Learning dissipative and conservative dynamics separately”, CoRR, vol. abs/2201.10085, 2022.

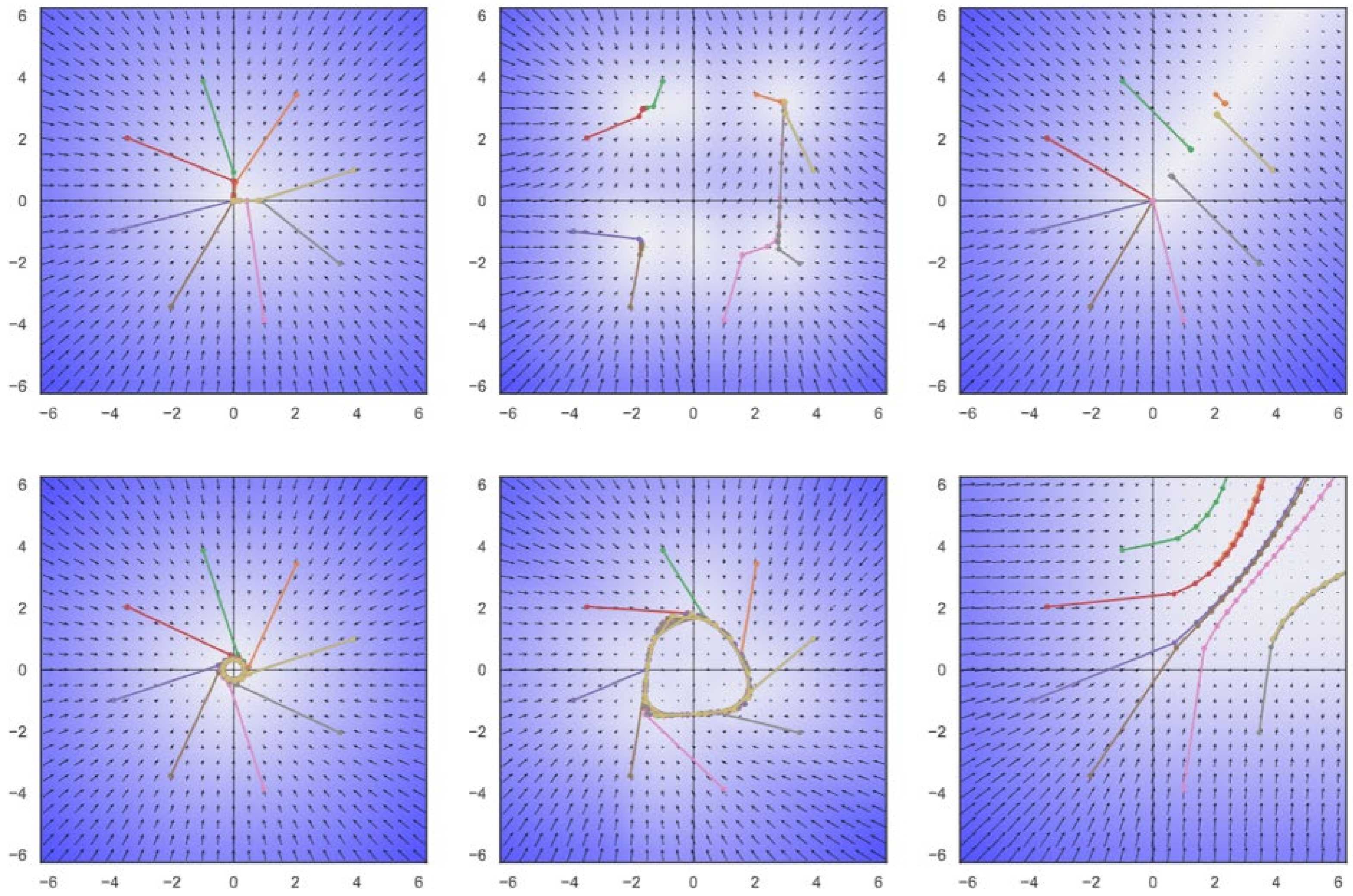
Deep Neural Dynamical System

$$\mathbf{x}_{t+1} = \mathbf{f}_\theta(\mathbf{x}_t)$$

$$\begin{aligned} \mathbf{f}_\theta(\mathbf{x}) &= \mathbf{A}_L \mathbf{z}_L + \mathbf{b}_L \\ \mathbf{z}_{i+1} &= \boldsymbol{\sigma}(\mathbf{A}_i \mathbf{z}_i + \mathbf{b}_i) \\ \mathbf{z}_0 &= \mathbf{x} \end{aligned}$$

Explore connections between:

- Deep neural network components
- Piecewise affine (PWA) maps
- Contraction of PWA maps
- Dissipativity of dynamical systems



2D attractors generated by deep neural dynamics.

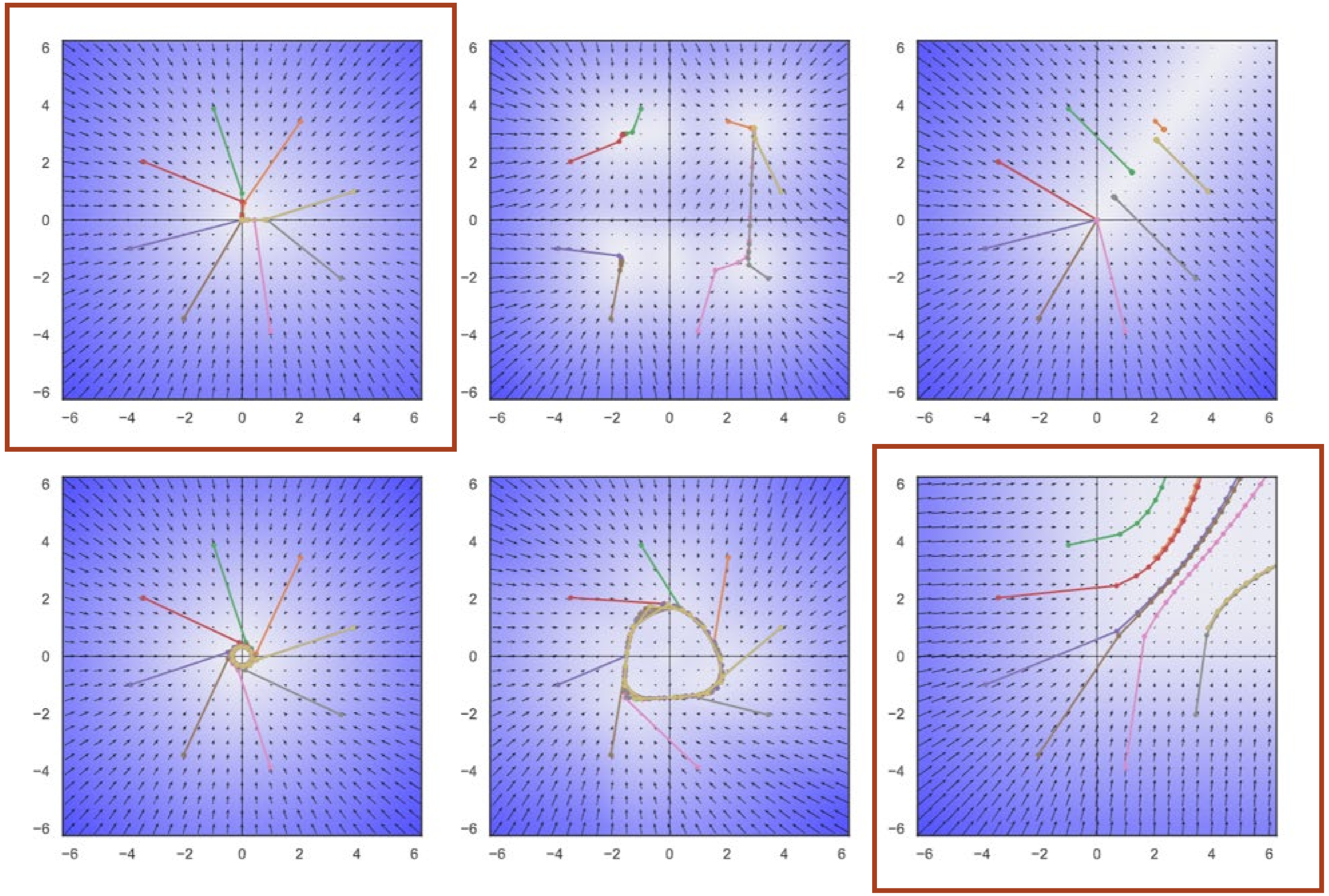
Deep Neural Dynamical System

$$\mathbf{x}_{t+1} = \mathbf{f}_\theta(\mathbf{x}_t)$$

$$\begin{aligned} \mathbf{f}_\theta(\mathbf{x}) &= \mathbf{A}_L \mathbf{z}_L + \mathbf{b}_L \\ \mathbf{z}_{i+1} &= \boldsymbol{\sigma}(\mathbf{A}_i \mathbf{z}_i + \mathbf{b}_i) \\ \mathbf{z}_0 &= \mathbf{x} \end{aligned}$$

Explore connections between:

- Deep neural network components
- Piecewise affine (PWA) maps
- Contraction of PWA maps
- Dissipativity of dynamical systems



2D attractors generated by deep neural dynamics.

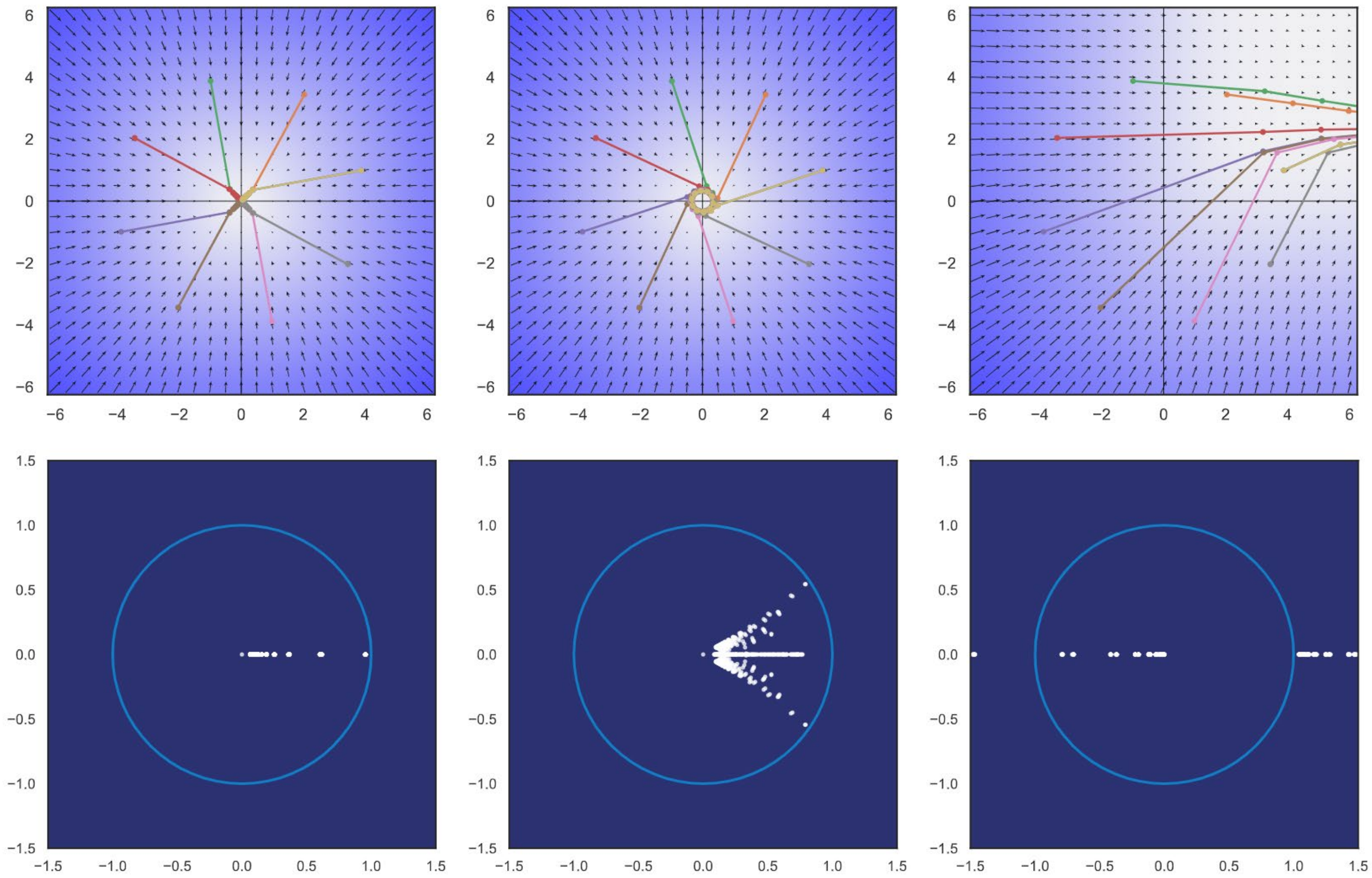
Pacific Northwest NATIONAL LABORATORY

Dynamical Effects of Weights

$$\mathbf{x}_{t+1} = \mathbf{f}_\theta(\mathbf{x}_t)$$

$$\begin{aligned} \mathbf{f}_\theta(\mathbf{x}) &= \boxed{\mathbf{A}_L}\mathbf{z}_L + \mathbf{b}_L \\ \mathbf{z}_{i+1} &= \sigma(\boxed{\mathbf{A}_i}\mathbf{z}_i + \mathbf{b}_i) \\ \mathbf{z}_0 &= \mathbf{x} \end{aligned}$$

Intuition: weight eigenvalues determine stability of DNN. Weight eigenvectors determine eigenvector basis of DNN.

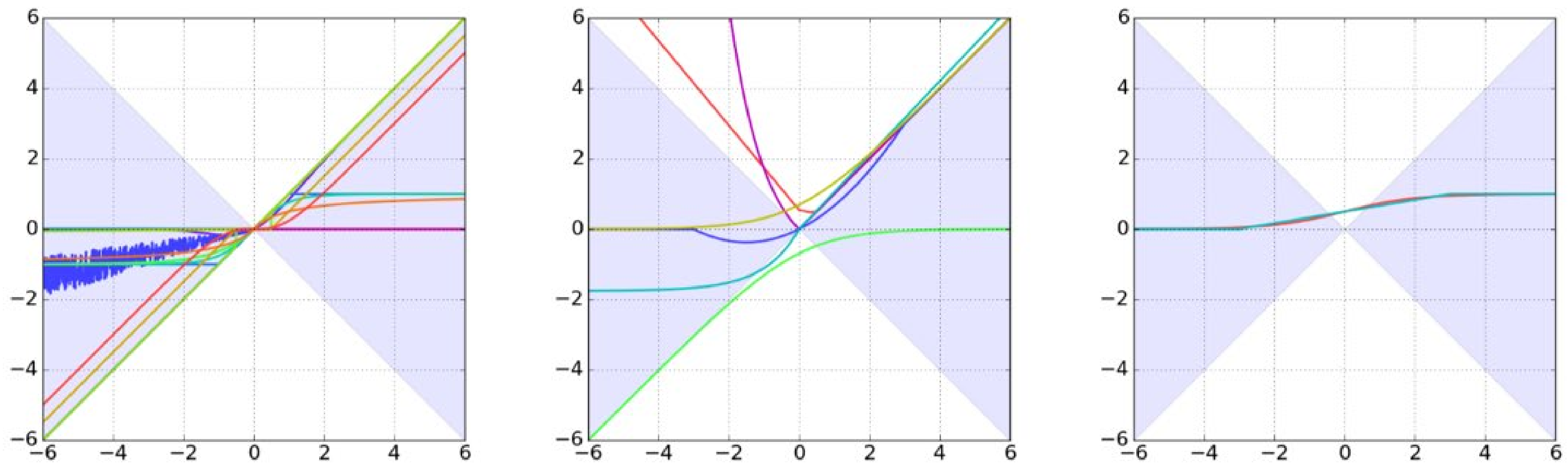


J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

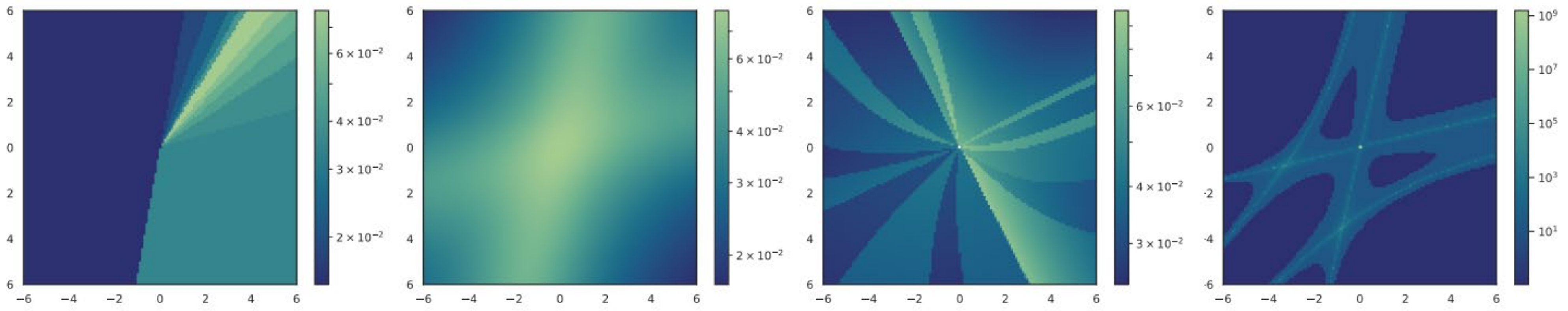
Dynamical Effects of Activation Functions

$$\mathbf{x}_{t+1} = \mathbf{f}_\theta(\mathbf{x}_t)$$

$$\mathbf{f}_\theta(\mathbf{x}) = \mathbf{A}_L\mathbf{z}_L + \mathbf{b}_L$$
$$\mathbf{z}_{i+1} = \boxed{\sigma(}\mathbf{A}_i\mathbf{z}_i + \mathbf{b}_i)$$
$$\mathbf{z}_0 = \mathbf{x}$$



Most commonly used activation functions such as ReLU, GELU, or tanh are contractive.



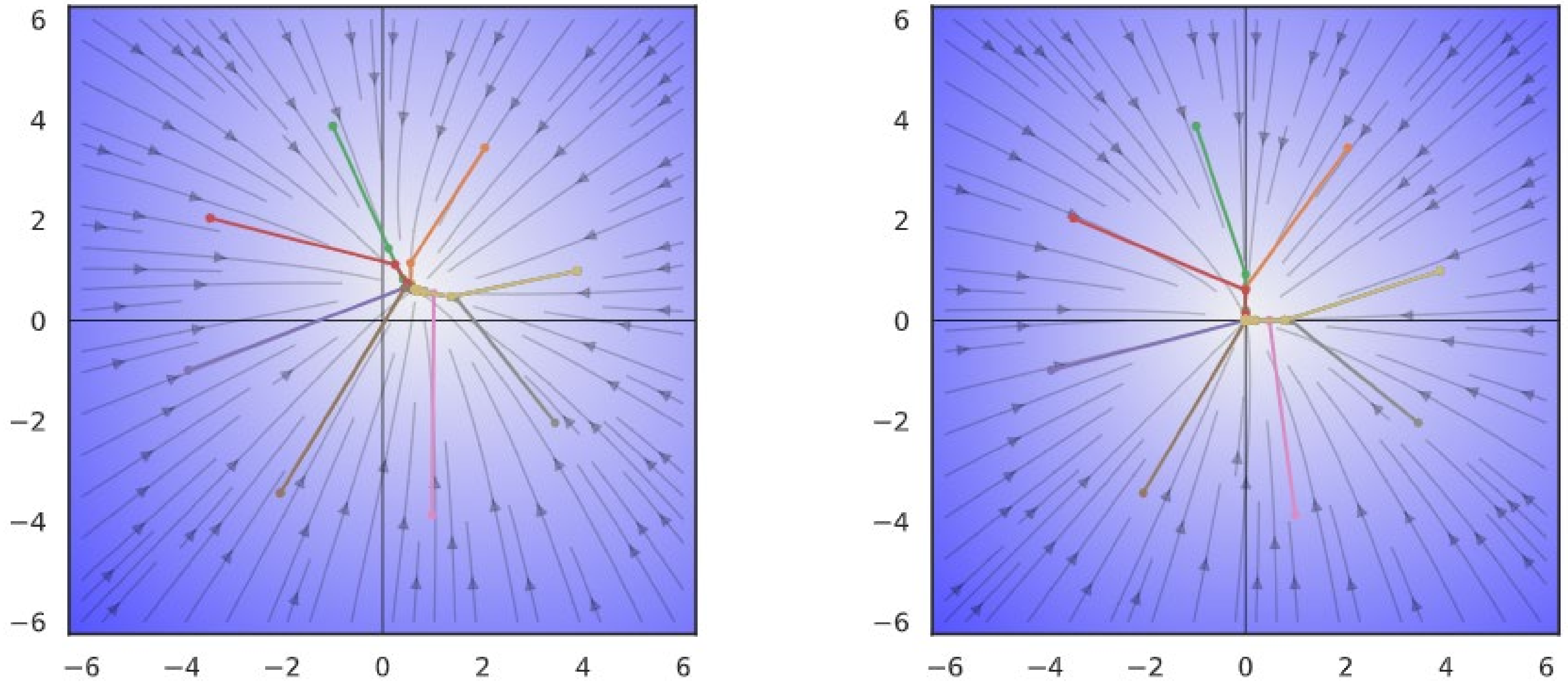
Intuition: Activation functions determine state space partitioning and scaling of local eigenvalues.

J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

Dynamical Effects of Biases

$$\mathbf{x}_{t+1} = \mathbf{f}_\theta(\mathbf{x}_t)$$

$$\mathbf{f}_\theta(\mathbf{x}) = \mathbf{A}_L \mathbf{z}_L + \boxed{\mathbf{b}_L}$$
$$\mathbf{z}_{i+1} = \boldsymbol{\sigma}(\mathbf{A}_i \mathbf{z}_i + \boxed{\mathbf{b}_i})$$
$$\mathbf{z}_0 = \mathbf{x}$$



Intuition: biases determine state space partitioning and the location of region of attraction via coordinate shifts.
Zero bias with stable weights leads to steady state at the origin.
Non-zero biases define fixed points of deep neural dynamics with stable weights.

J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

Dynamical Effects of Network Depth

$$\mathbf{x}_{t+1} = \mathbf{f}_\theta(\mathbf{x}_t)$$

$$\mathbf{f}_\theta(\mathbf{x}) = \mathbf{A}_L\mathbf{z}_L + \mathbf{b}_L$$
$$\mathbf{z}_{\boxed{i+1}} = \boldsymbol{\sigma}(\mathbf{A}_i\mathbf{z}_i + \mathbf{b}_i)$$
$$\mathbf{z}_0 = \mathbf{x}$$

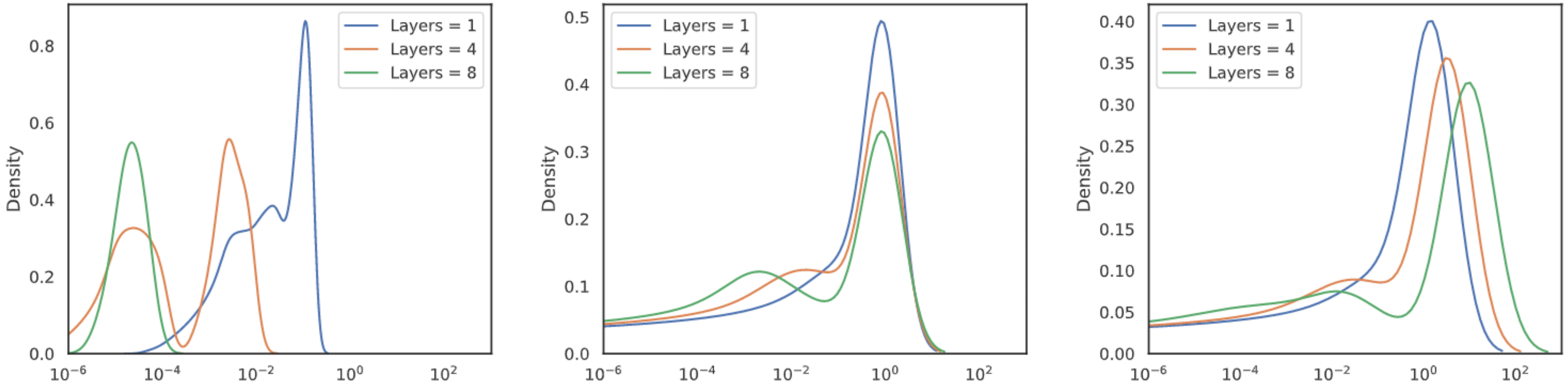


Figure 5: Real eigenvalue distributions of neural networks with varying depth using GELU *layers with stable (first column), on the edge of stability (second column), and unstable dynamics (third column), respectively.*

Deeper networks shift their eigenvalue distribution towards heavy tail distributions and exacerbate the dynamical effects of their layers leading to vanishing and exploding gradient problem.

J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

Deep Neural Networks as Piecewise Affine Maps

Lemma 1: [58] Let $\mathbf{f}_{\boldsymbol{\theta}}$ (1) be a deep neural network with activation function $\boldsymbol{\sigma}$, then there exists a pointwise affine map $\mathbf{A}^{\star}(\mathbf{x})\mathbf{x} + \mathbf{b}^{\star}(\mathbf{x})$ parametrized by $\mathbf{x}$ which satisfies the following:

$$\mathbf{f}_{\theta}(\mathbf{x}) := \mathbf{A}^{\star}(\mathbf{x})\mathbf{x} + \mathbf{b}^{\star}(\mathbf{x}) \tag{2}$$

where $\mathbf{A}^{\star}(\mathbf{x})$ is a state-dependent matrix given as:

$$\mathbf{A}^{\star}(\mathbf{x})\mathbf{x} = \mathbf{W}_L \boldsymbol{\Lambda}_{\mathbf{z}_{L-1}} \mathbf{W}_{L-1} \dots \boldsymbol{\Lambda}_{\mathbf{z}_0} \mathbf{W}_0 \mathbf{x} \tag{3}$$

and $\mathbf{b}^{\star}(\mathbf{x})$ is a state-dependent vector given as:

$$\mathbf{b}^{\star}(\mathbf{x}) = \mathbf{b}_L^{\star}, \quad \mathbf{b}_l^{\star} := \mathbf{W}_i \boldsymbol{\Lambda}_{\mathbf{z}_{l-1}} \mathbf{b}_{l-1}^{\star} \tag{4}$$

$$+ \mathbf{W}_i \boldsymbol{\sigma}_{l-1}(\mathbf{0}) + \mathbf{b}_l, \quad l \in \mathbb{N}_1^L \tag{5}$$

Here $\boldsymbol{\Lambda}_{\mathbf{z}_l}$ represents a diagonal matrix of activation patterns dependent on a hidden states $\mathbf{z}_l$ at l-th layer defined as:

$$\boldsymbol{\sigma}(\mathbf{z}) = \boldsymbol{\Lambda}_{\mathbf{z}}\mathbf{z} + \boldsymbol{\sigma}(\mathbf{0}) \tag{6a}$$

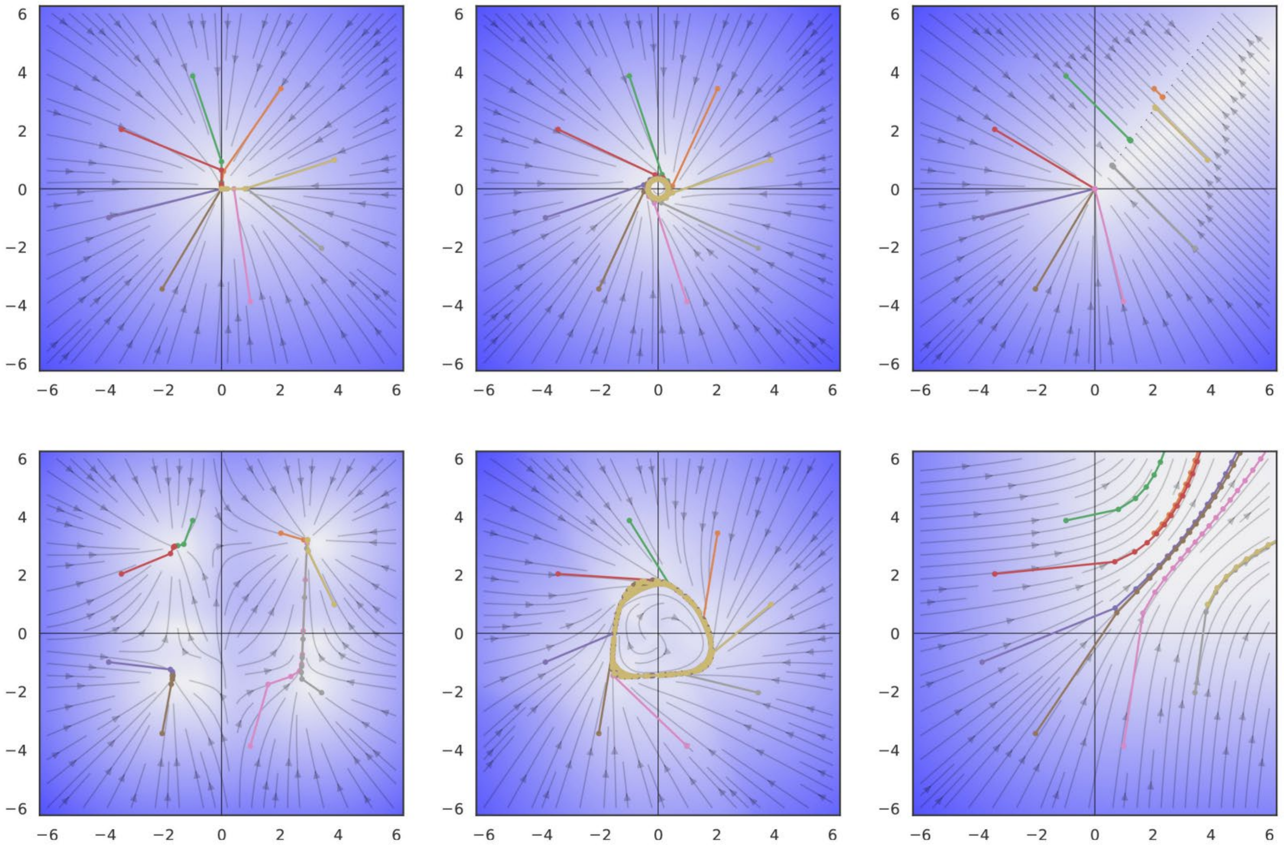
$$\boldsymbol{\sigma}(\mathbf{z}) = \begin{bmatrix} \frac{\sigma(z_1)-\sigma(0)}{z_1} & & \\ & \ddots & \\ & & \frac{\sigma(z_n)-\sigma(0)}{z_n} \end{bmatrix} \mathbf{z} + \begin{bmatrix} \sigma(0) \\ \vdots \\ \sigma(0) \end{bmatrix} \tag{6b}$$

Refactoring neural networks to expose their local affine operators suitable for analysis.

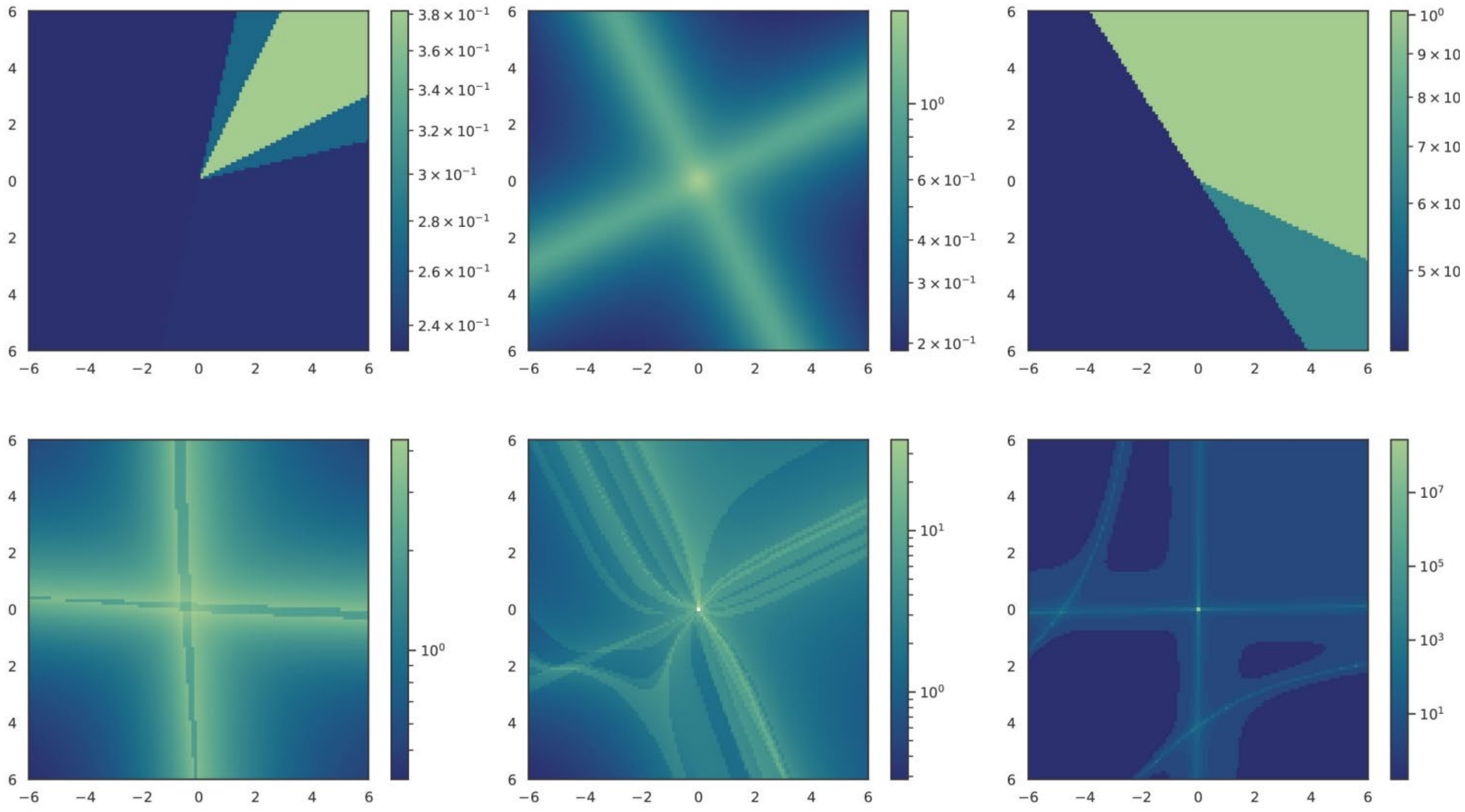
J. Drgona, S. Mukherjee, J. Zhang, F. Liu, and M. Halappanavar, "On the stochastic stability of deep Markov models," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2021.

Analysis of Local Operators of Autonomous Neural Dynamics

$$\mathbf{x}_{t+1} = \mathbf{A}^{\star}(\mathbf{x}_t)\mathbf{x}_t + \mathbf{b}^{\star}(\mathbf{x}_t)$$



2D attractors generated by deep neural dynamics.

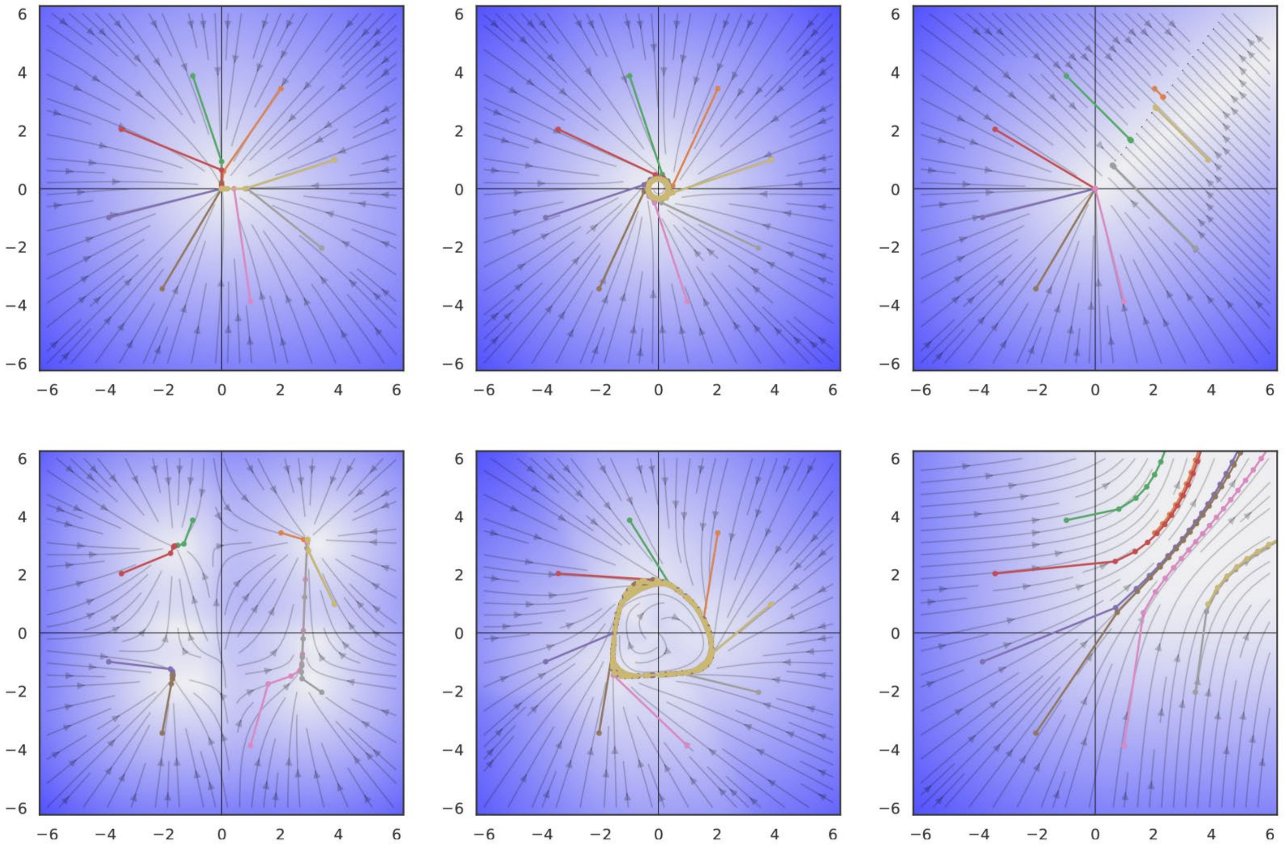


State space partitioning.

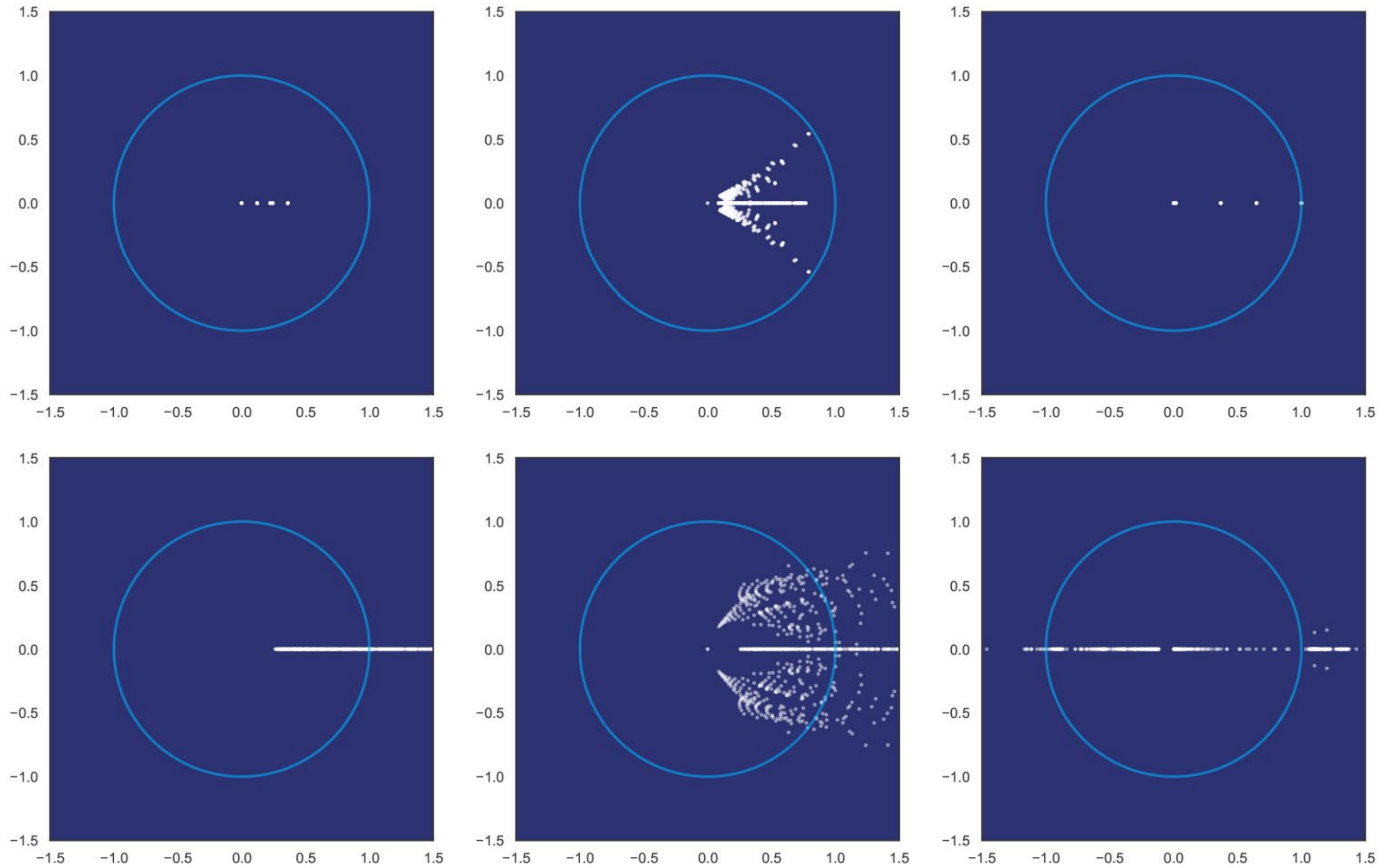
J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

Analysis of Local Operators of Autonomous Neural Dynamics

$$\mathbf{x}_{t+1} = \mathbf{A}^{\star}(\mathbf{x}_t)\mathbf{x}_t + \mathbf{b}^{\star}(\mathbf{x}_t)$$



2D attractors generated by deep neural dynamics.



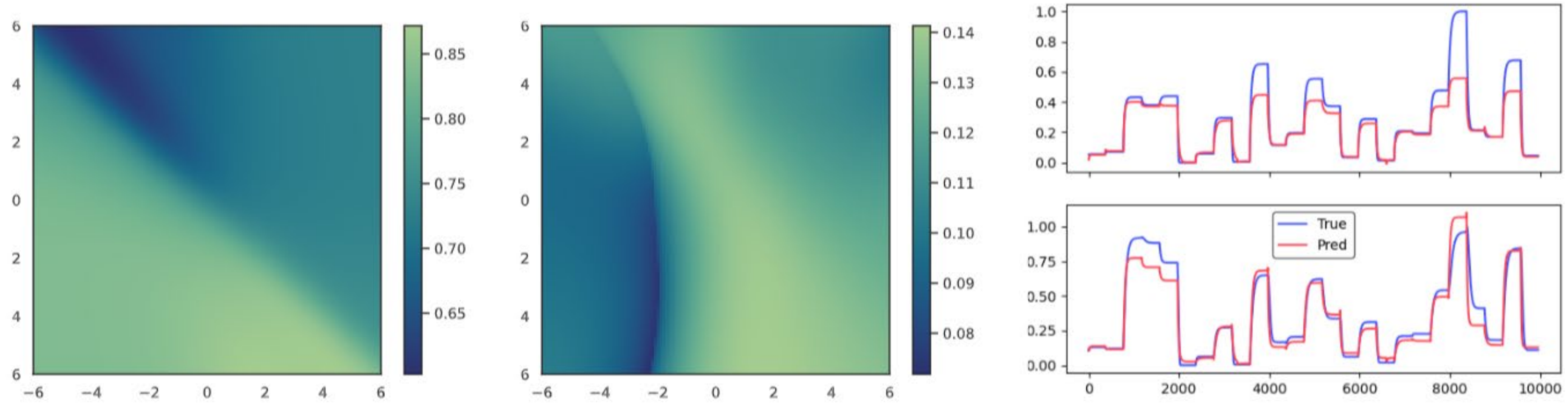
Eigenvalue distributions.

J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

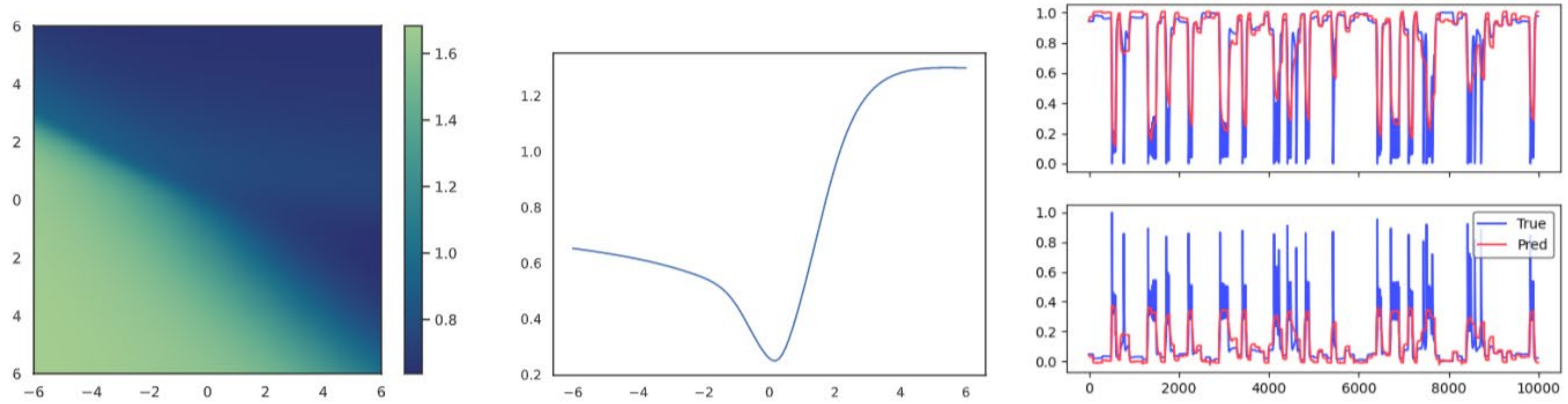
Analysis of Local Operators of Non-autonomous Neural Dynamics

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t) + \mathbf{g}(\mathbf{u}_t)$$

Two tank system: 2 states x, 2 inputs u



Exothermic stirred tank reactor system: 2 states x, 1 input u



J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

Dissipativity of Dynamical Systems

Definition III.1 Dissipative Discrete-time Dynamical System [1]: A discrete-time dynamical system (7) is said to be *dissipative* if the following condition holds:

$$\mathbf{V}(\mathbf{x}_{t+1}) - \mathbf{V}(\mathbf{x}_t) \leq \mathbf{s}(\mathbf{x}_t), \ \forall t \in \{0, 1, 2, \ldots\} \tag{8}$$

Where $\mathbf{V}(\mathbf{x}_t) : \mathbb{R}^{n_x} \to \mathbb{R}$ such that $\mathbf{V}(0) = 0$, and $\mathbf{V}(\mathbf{x}_t) \geq 0$ represents a non-negative storage function quantifying the energy stored internally in the system, and $\mathbf{s}(\mathbf{x}_t) : \mathbb{R}^{n_x} \to \mathbb{R}$ is the so-called supply rate representing energy supplied to the system from the external environment.

Dissipativity is an extension of Lyapunov stability for open systems.

C. Byrnes and W. Lin, "Losslessness, feedback equivalence, and the global stabilization of discrete-time nonlinear systems," IEEE Trans. Autom. Control, vol. 39, no. 1, pp. 83–98, Jan. 1994.

Dissipative Deep Neural Dynamical System

Theorem 1. Dissipative Deep Neural Dynamical Systems: the neural dynamical system (7) is dissipative over a state-space region $\mathcal{X} \subseteq \mathbb{R}^{n_x}$ with respect to the supply rate $\mathbf{s}(\mathbf{x}_t) = ||\mathbf{b}^\star(\mathbf{x}_t)||_2$ if the local linear dynamics $\|\mathbf{A}^\star(\mathbf{x})\|_2$ of the equivalent PWA form (2) is a contractive map over the entire region $\mathcal{X}$. Or more formally the following must hold:

$$\|\mathbf{A}^\star(\mathbf{x})\|_2 < 1, \quad \forall \mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^{n_x}. \tag{10}$$

Proof: Consider the dissipativity condition (8) with a choosen storage function $\mathbf{V}(\mathbf{x}) = \sqrt{\mathbf{x}^T \mathbf{x}}$ and supply rate $\mathbf{s}(\mathbf{x}_t) = ||\mathbf{b}^\star(\mathbf{x}_t)||_2$ we will prove the following dissipativity condition:

$$||\mathbf{x}_{t+1}||_2 - ||\mathbf{x}_t||_2 \le ||\mathbf{b}^\star(\mathbf{x}_t)||_2 \tag{11}$$

Leveraging the equivalence of DNN with PWA (2) we get:

$$\mathbf{x}_{t+1} = \mathbf{A}^\star(\mathbf{x}_t)\mathbf{x}_t + \mathbf{b}^\star(\mathbf{x}_t) \tag{12}$$

Applying 2-norms to (12) we get:

$$||\mathbf{x}_{t+1}||_2 = ||\mathbf{A}^\star(\mathbf{x}_t)\mathbf{x}_t + \mathbf{b}^\star(\mathbf{x}_t)||_2 \tag{13}$$

Then applying norm subadditivity (43) and submultiplicativity (44) of the norms we have:

$$||\mathbf{x}_{t+1}||_2 \le ||\mathbf{A}^\star(\mathbf{x}_t)||_2||\mathbf{x}_t||_2 + ||\mathbf{b}^\star(\mathbf{x}_t)||_2 \tag{14}$$

We can substitute (14) into the dissipativity condition (11):

$$||\mathbf{A}^\star(\mathbf{x}_t)||_2||\mathbf{x}_t||_2 + ||\mathbf{b}^\star(\mathbf{x}_t)||_2 - ||\mathbf{x}_t||_2 \le ||\mathbf{b}^\star(\mathbf{x}_t)||_2 \tag{15}$$

Leading to:

$$||\mathbf{A}^\star(\mathbf{x}_t)||_2||\mathbf{x}_t||_2 - ||\mathbf{x}_t||_2 \le 0 \tag{16}$$

Now its clear that the condition (10) must hold $\forall \mathbf{x}_t \in \mathcal{X}$ to satisfy the dissipativity (16) locally over the set $\mathcal{X}$. ■

J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

Steady States of Deep Neural Dynamics

$$\mathbf{x}_{t+1} = \mathbf{f}_\theta(\mathbf{x}_t) \tag{7}$$

Corollary 3: Assume the conditions of Theorem 1 with bounded supply rate and deep neural dynamics (7) that converge to equilibrium $\bar{\mathbf{x}}$ (18). Then given the system (7), there exists an equilibrium point $\mathbf{x}_{\mathrm{lb}} \leq ||\bar{\mathbf{x}}||_p \leq \mathbf{x}_{\mathrm{ub}}$ with the bounds:

$$\mathbf{x}_{\mathrm{lb}} = \frac{\|\mathbf{b}^\star(\mathbf{x})\|_p}{||\mathbf{I} - \mathbf{A}^\star(\mathbf{x})||_p}, \quad \mathbf{x}_{\mathrm{ub}} = \frac{\|\mathbf{b}^\star(\mathbf{x})\|_p}{1 - \|\mathbf{A}^\star(\mathbf{x})\|_p}. \tag{23}$$

Intuition: for stable systems, region of attraction shrinks with smaller bias terms and stronger contractivity of neural network layers and vice versa.

$$\bar{\mathbf{x}} = \mathbf{f}_\theta(\bar{\mathbf{x}}) = \lim_{t \to \infty} \mathbf{f}_\theta(\mathbf{x}_t) \tag{18}$$

Corollary 4: Neural neural dynamics (7) satisfies the dissipativity condition (10) if the norms of all the weights $\mathbf{W}_i$ and activation matrices $\mathbf{\Lambda}_{\mathbf{z}_j}$ (6) of $\mathbf{f}_\theta(\mathbf{x})$ are contractive:

$$\|\mathbf{W}_i\|_2 < 1, \; i \in \mathbb{N}_0^L, \; \|\mathbf{\Lambda}_{\mathbf{z}_j}\|_2 \leq 1, \; \forall j \in \mathbb{N}_1^L \tag{33}$$

Intuition: submultiplicativity of layer norms allows to design globally dissipative neural dynamics by choosing the weights and activation functions.

J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

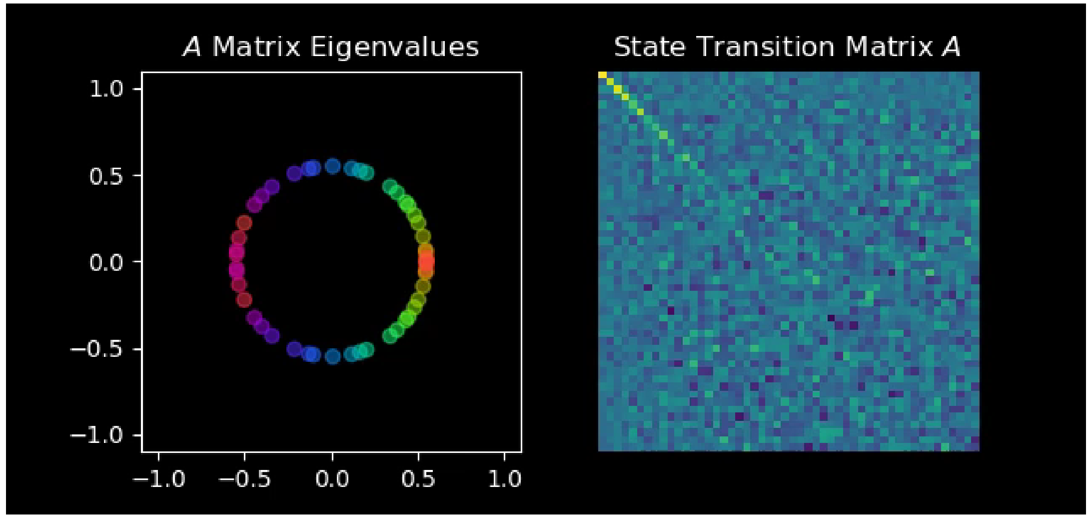
Practical Eigenvalue Constraints for Weights

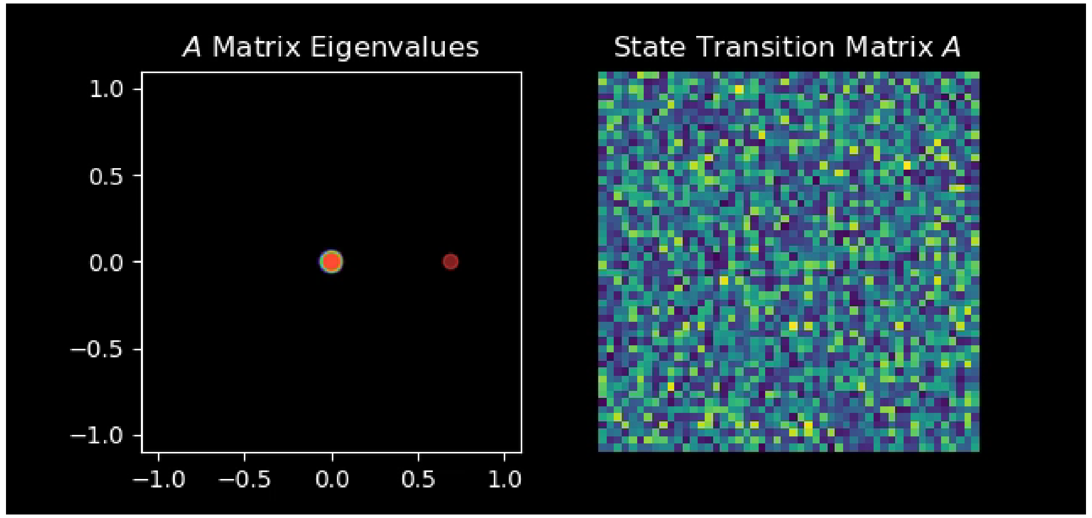
SVD factorization

$$\mathbf{\Sigma} = \mathrm{diag}(\lambda_{\max} - (\lambda_{\max} - \lambda_{\min}) \cdot \sigma(\Sigma))$$
$$\mathbf{W} = \mathbf{U\Sigma V}$$

Perron-Frobenius map

$$\mathbf{M} = \lambda_{\max} - (\lambda_{\max} - \lambda_{\min}) g(\mathbf{M}')$$
$$\mathbf{W}_{i,j} = \frac{\exp(\mathbf{A'}_{ij})}{\sum_{k=1}^{n_x} \exp(\mathbf{A'}_{ik})} \mathbf{M}_{i,j}$$





Pytorch implementation: https://github.com/pnnl/slim

[1] J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022
[2] Jiong Zhang, et al., Stabilizing Gradients for Deep Neural Networks via Efficient SVD Parameterization, 2018.

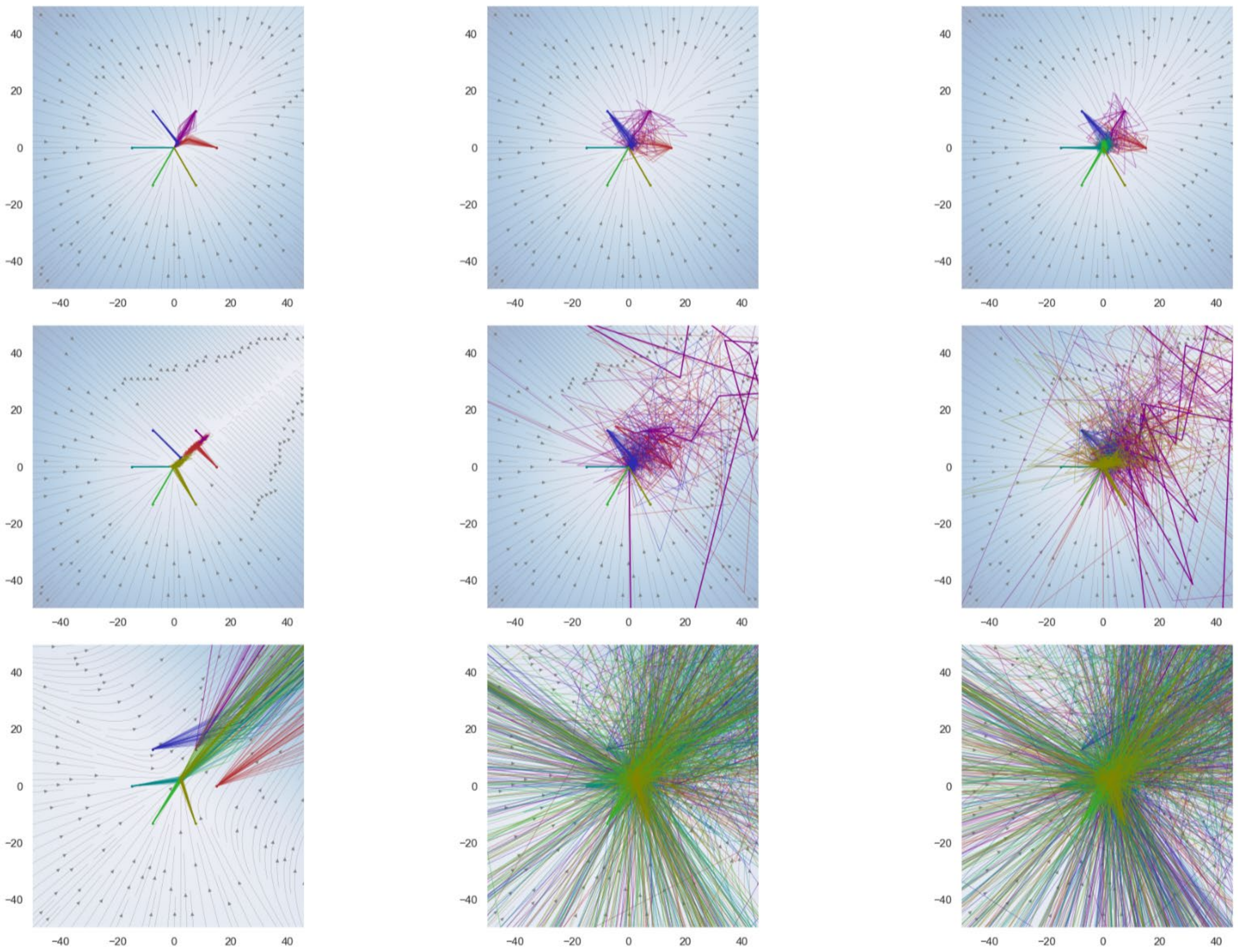
Extension: Stability of Deep Markov Models

$$\mathbf{x}_{t+1} \sim \mathcal{N}(K_\alpha(\mathbf{x}_t, \Delta t), L_\beta(\mathbf{x}_t, \Delta t))$$
$$\mathbf{y}_t \sim \mathcal{M}(F_\kappa(\mathbf{x}_t))$$

$$K_\alpha(\mathbf{x}_t, \Delta t) = \mathbf{f}_{\theta_\mathbf{f}}(\mathbf{x}_t)$$
$$\mathrm{vec}(L_\beta(\mathbf{x}_t, \Delta t)) = \mathbf{g}_{\theta_\mathbf{g}}(\mathbf{x}_t)$$

Exploring connections between:

- Stability of stochastic systems
- Deep neural networks (DNNs)
- Deep Markov models (DMMs)
- Contraction of DMMs



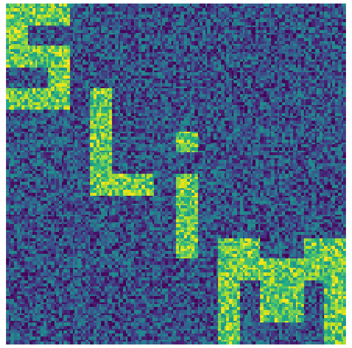
2D attractors generated by deep Markov models.

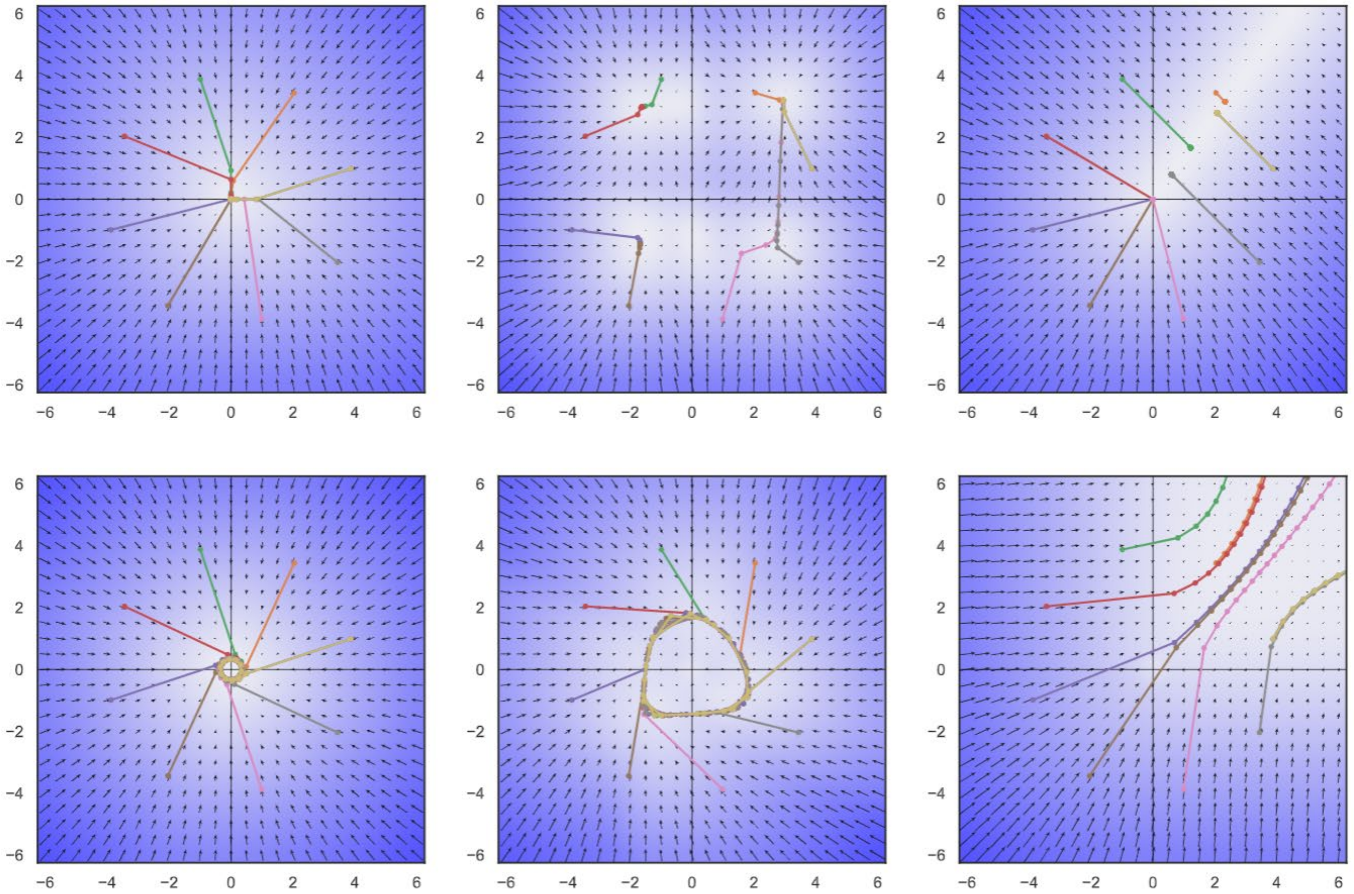
Jan Drgona, et al., On the stochastic stability of deep Markov models, NeurIPS 2021.

Pacific Northwest
NATIONAL LABORATORY

Conclusion

- **Dissipativity of Deep Neural Networks**
  - Deep neural networks (DNNs)
  - Piecewise affine (PWA) maps
  - Dissipativity of PWA
- **Weight Constraints**
  - Structured linear maps (SLIM) in Pytorch
  - https://pnnl.github.io/slim/

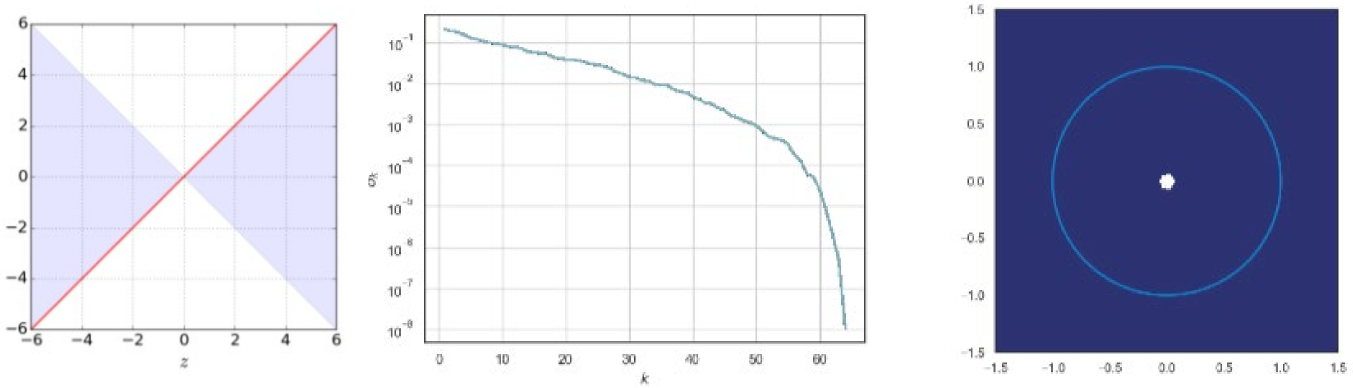




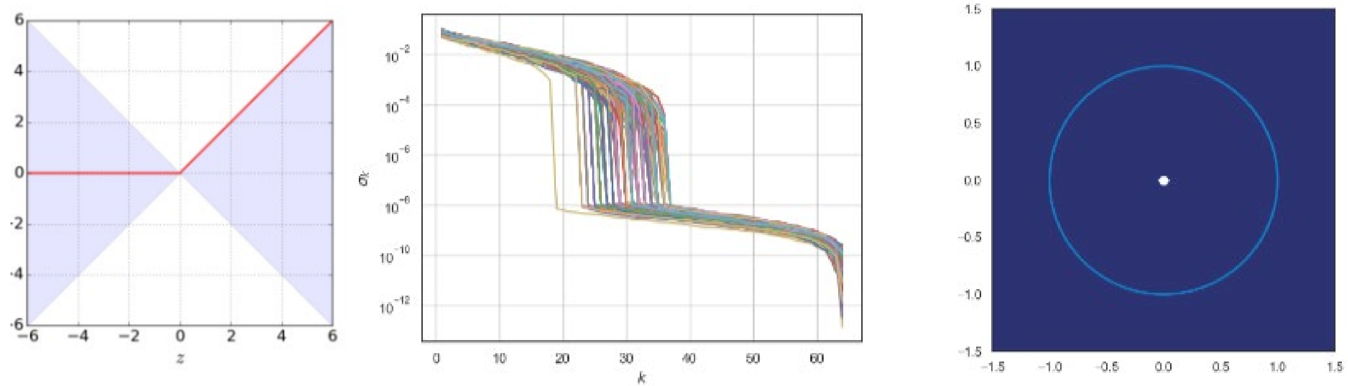
- **Contact**
  - jan.drgona@pnnl.gov
  - https://www.linkedin.com/in/drgona/

U.S. DEPARTMENT OF ENERGY | Office of Science

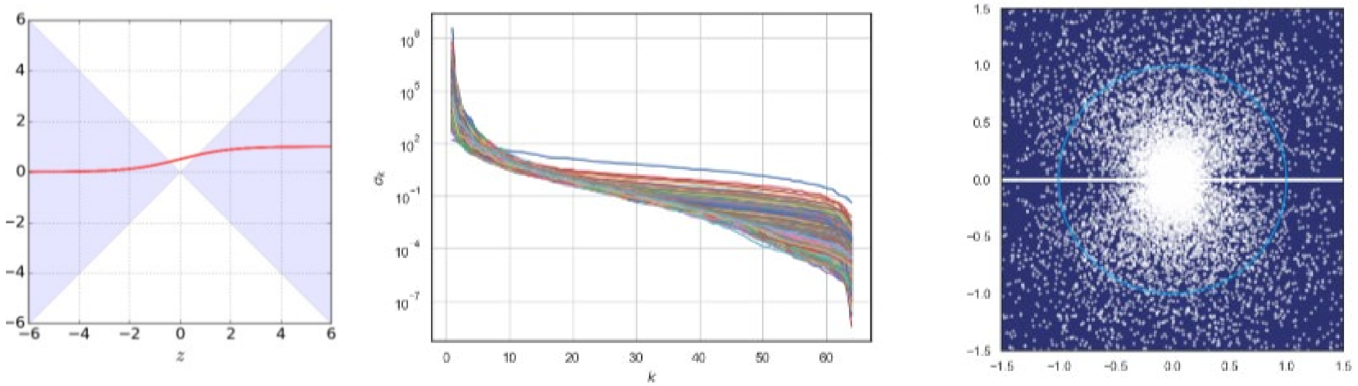
Nonlinearity and Eigenvalues of Deep Neural Networks with Different Activation Functions



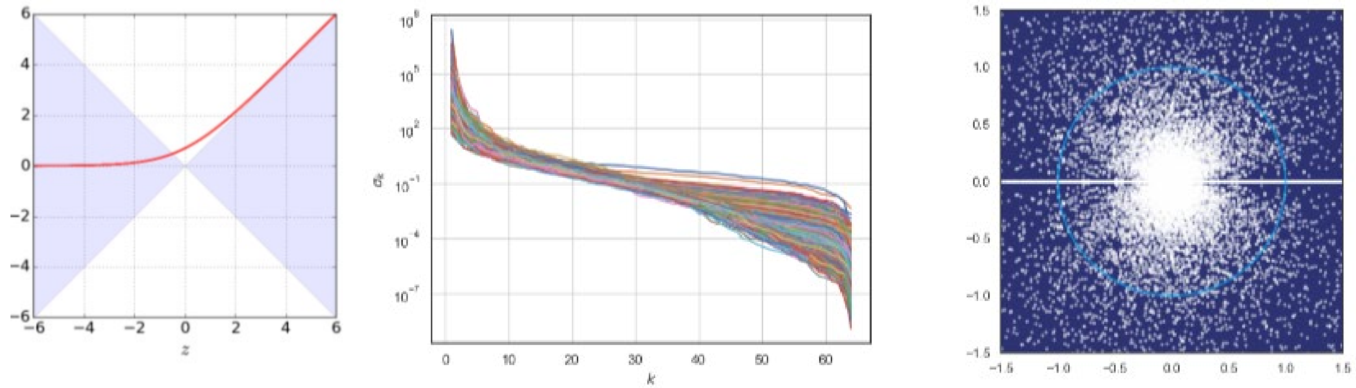
Identity



ReLU



Sigmoid



SoftExponential

Intuition: Activation functions shape eigenvalue distributions of deep neural networks.

J. Drgoňa, A. Tuor, S. Vasisht and D. Vrabie, "Dissipative Deep Neural Dynamical Systems," in *IEEE Open Journal of Control Systems*, vol. 1, pp. 100-112, 2022

Spectra of Deep Neural Network with Different Weight Factorizations