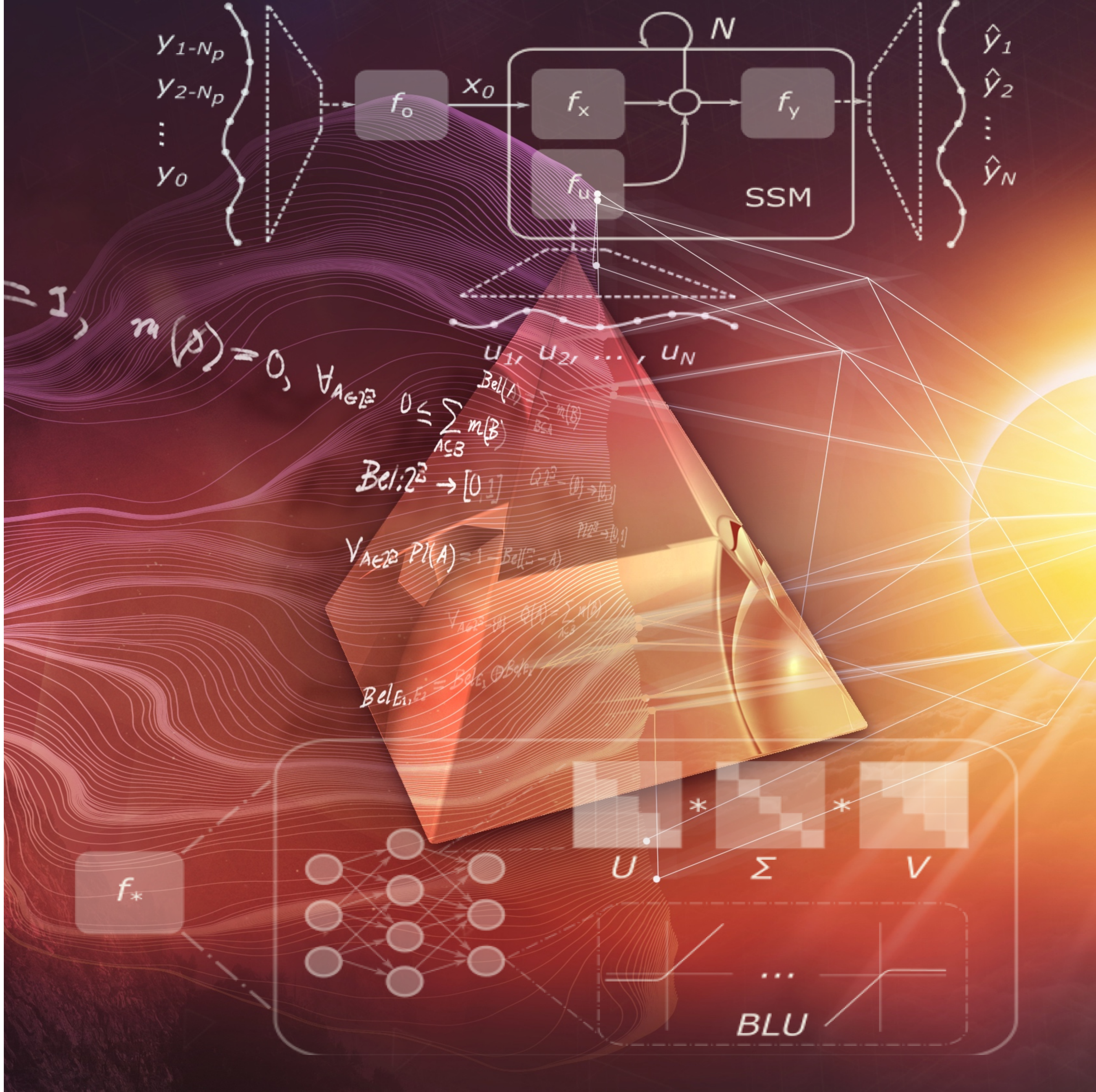
On Stochastic Stability of Deep Markov Models

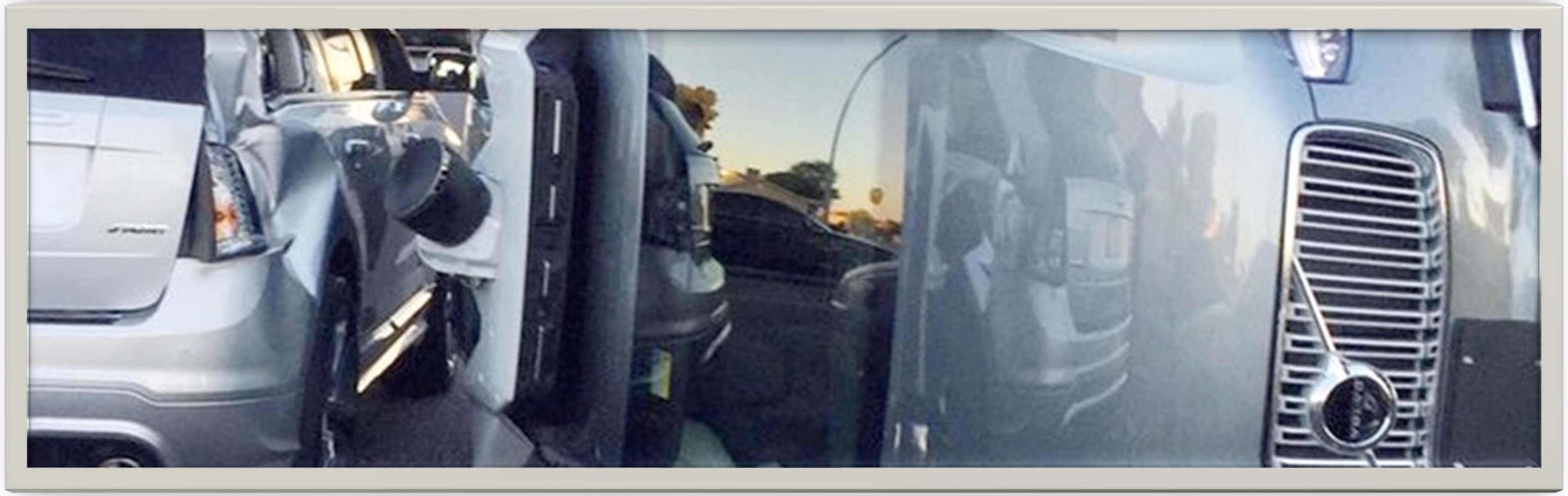
NeurIPS, December 6-14, 2021

Jan Drgona (PNNL), Sayak Mukherjee (PNNL),
Jiaxin Zhang (ORNL), Frank Liu (ORNL),
Mahantesh Halappanavar (PNNL)



Stability Analysis of Deep Markov Models

- **Motivation**
  - Safety-critical systems call for formal verification methods to ensure safe operation.
  - Properties like stability and robustness are crucial for reliable modeling and control.
- **Objectives**
  - Sufficient conditions for stochastic stability of deep Markov models (DMMs).
- **Approaches**
  - Apply system-theoretic analysis methods on DMMs.

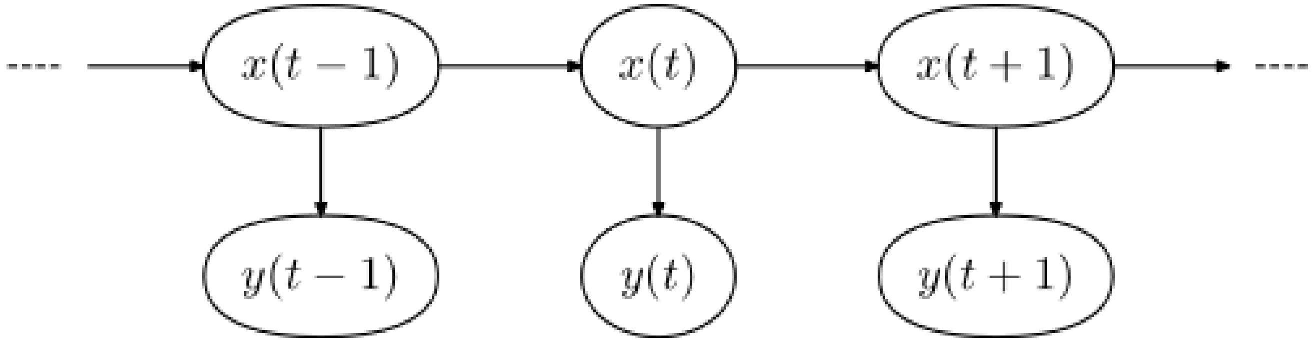


Deep Markov Models

$$P(\mathbf{x}_{0:T}, \mathbf{y}_{0:T}) = P(\mathbf{x}_0)P(\mathbf{y}_0|\mathbf{x}_0) \prod_{0}^{T-1} P(\mathbf{x}_{t+1}|\mathbf{x}_t)P(\mathbf{y}_t|\mathbf{x}_t).$$

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

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



https://en.wikipedia.org/wiki/Hidden_Markov_mode

Deep Markov Models:
- Probabilistic graphical model (PGM)
- Generative model of sequential data
- Applications:
  - Economics, finance
  - Pattern recognition
  - Signal processing

Exploring connections between:
- Stability of stochastic systems
- Deep Markov models (DMMs)
- Contraction of DMM transitions
- Operator norms
- Banach fixed point theorem

Deep Neural Networks as Piecewise Affine Maps

DNN

$$\psi_{\theta_\psi}(\mathbf{x}) = \mathbf{A}_L^\psi \mathbf{h}_L^\psi + \mathbf{b}_L$$
$$\mathbf{h}_l^\psi = \boldsymbol{v}(\mathbf{A}_{l-1}^\psi \mathbf{h}_{l-1}^\psi + \mathbf{b}_{l-1})$$

PWA map

$$\psi_{\theta_\psi}(\mathbf{x}) = \mathbf{A}_\psi(\mathbf{x})\mathbf{x} + \mathbf{b}_\psi(\mathbf{x}).$$

PWA activation map

$$\boldsymbol{v}(\mathbf{z}) = \begin{bmatrix} \frac{v(z_1)-v(0)}{z_1} & & \\ & \ddots & \\ & & \frac{v(z_n)-v(0)}{z_n} \end{bmatrix} \mathbf{z} + \begin{bmatrix} v(0) \\ \vdots \\ v(0) \end{bmatrix} = \boldsymbol{\Lambda}_{\mathbf{z}}^\psi \mathbf{z} + \boldsymbol{v}(\mathbf{0})$$

Local linear dynamics of DNN

- At every point $\mathbf{x}$, DNN can be represented as a product of PWA maps:

$$\mathbf{A}_\psi(\mathbf{x})\mathbf{x} = \mathbf{A}_L^\psi \boldsymbol{\Lambda}_{\mathbf{z}_L}^\psi \mathbf{A}_{L-1}^\psi \ldots \boldsymbol{\Lambda}_{\mathbf{z}_1}^\psi \mathbf{A}_0^\psi \mathbf{x}$$

$$\mathbf{b}_{\psi,i} = \mathbf{A}_i^\psi \boldsymbol{\Lambda}_{\mathbf{z}_i}^\psi \mathbf{b}_{\psi,i-1} + \mathbf{b}_i, \;\; i \in \mathbb{N}_1^L, \;\; \mathbf{b}_{\psi,0} = \mathbf{b}_0.$$

Stochastic Stability of Deep Markov Models

Definition 3. *The stochastic process* $\mathbf{x}_t \in \mathbb{R}^n$ *is mean-square stable (MSS) if and only if there exists* $\mu \in \mathbb{R}^n, \mathbf{\Sigma} \in \mathbb{R}^{n \times n}$, *such that* $\lim_{t\to\infty} \mathbb{E}(\mathbf{x}_t) = \mu$, *and* $\lim_{t\to\infty} \mathbb{E}(\mathbf{x}_t \mathbf{x}_t^T) = \mathbf{\Sigma}$.

Sufficient stability conditions of DMM:

$$||\mathbf{A_f}(\mathbf{x})||_p < 1$$

$$||\mathbf{A_g}(\mathbf{x})||_p + \frac{||\mathbf{b_g}(\mathbf{x})||_p}{||\mathbf{x}||_p} < K, K > 0,$$

$$\forall \mathbf{x} \in Domain(\mathbf{f}_{\theta_\mathbf{f}}(\mathbf{x}), \mathbf{g}_{\theta_\mathbf{g}}(\mathbf{x})).$$

Local Lipschitz constant of DNN:

$$\mathcal{K}^{\mathbf{g}}(\mathbf{x}) = ||\mathbf{A_g}(\mathbf{x})||_p + \frac{||\mathbf{b_g}(\mathbf{x})||_p}{||\mathbf{x}||_p}.$$

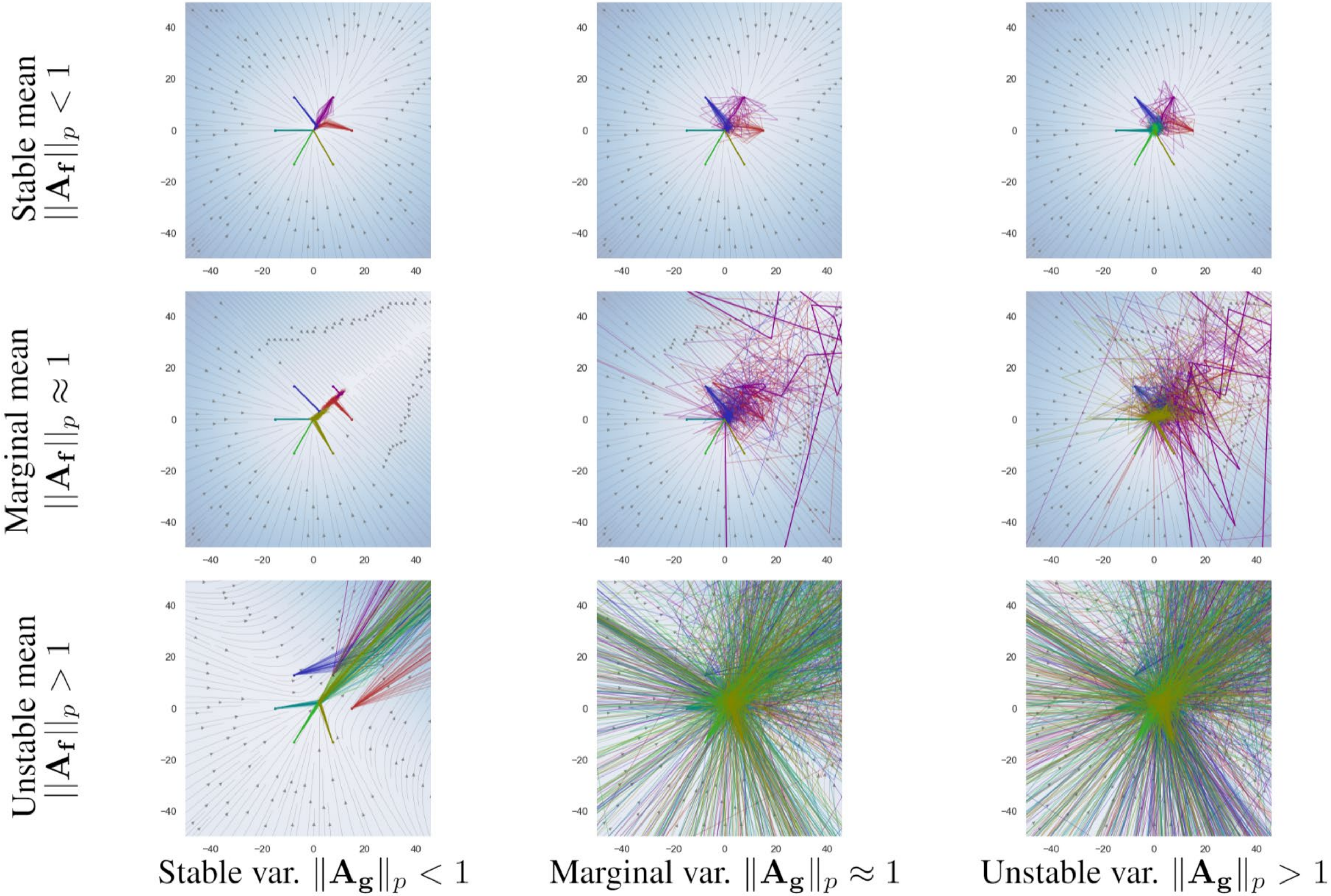
Contractive weights and activations imply DMM stability:

$$||\mathbf{A}_i^{\mathbf{f}}||_p < 1, \ ||\mathbf{\Lambda}^{\mathbf{f}}_{\mathbf{z}_i}||_p \leq 1 \ i \in \mathbb{N}_1^{L_\mathbf{f}},$$

$$||\mathbf{A}_j^{\mathbf{g}}||_p < c^{\mathbf{A}}, \ ||\mathbf{\Lambda}^{\mathbf{g}}_{\mathbf{z}_j}||_p \leq c^{\mathbf{\Lambda}}, \ j \in \mathbb{N}_1^{L_\mathbf{g}},$$

$$\forall \mathbf{x} \in Domain(\mathbf{f}_{\theta_\mathbf{f}}(\mathbf{x}), \mathbf{g}_{\theta_\mathbf{g}}(\mathbf{x})).$$

Effect of Mean and Variance on Stochastic Stability of Deep Markov Models



Effect of Biases and Depth on the Stability of Deep Markov Models

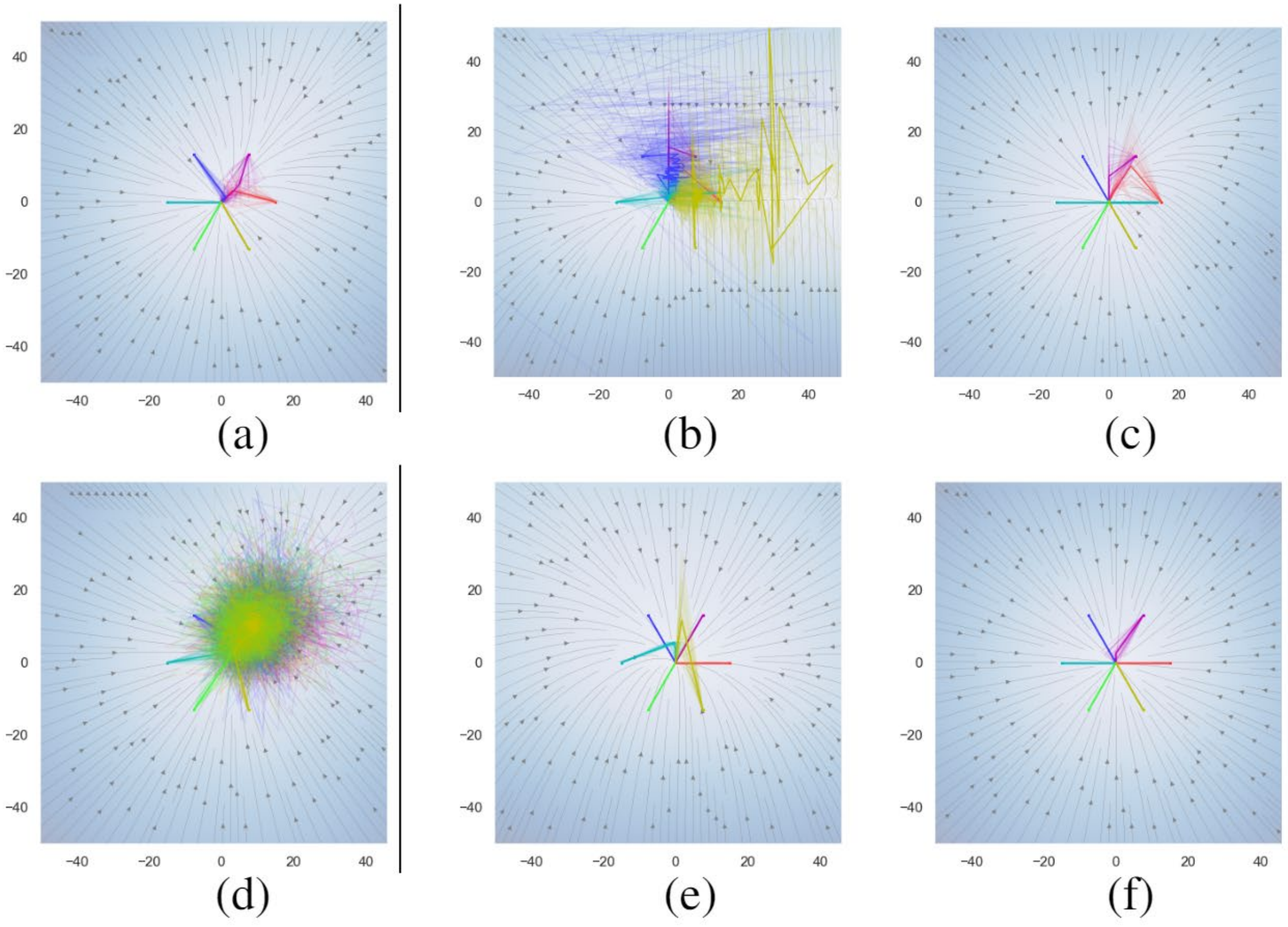
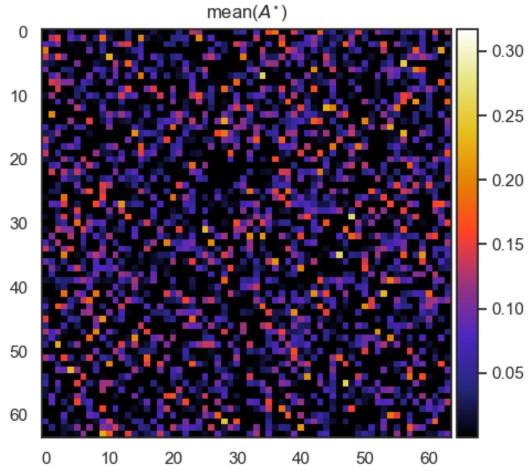


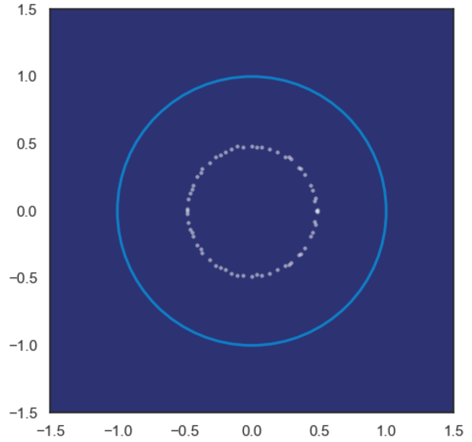
Figure 2: Left panels show the effect of biases using PF regularization and `ReLU` *activation ((a) w/o bias, (d) w bias). Right panels show the effect of network* **f** *depths with SVD regularization and* `ReLU` *: (b) 1 layer, (c) 2 layers, (e) 4 layers, (f) 8 layers.*

Practical Stability Constraints for DNN and DMM

SVD factorization

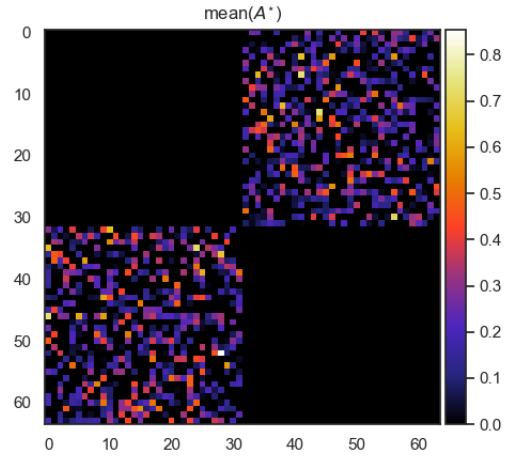
$$\tilde{\Sigma} = \mathrm{diag}(\lambda_{\max} - (\lambda_{\max} - \lambda_{\min}) \cdot \sigma(\Sigma))$$
$$\tilde{\mathbf{A}} = \mathbf{U}\tilde{\Sigma}\mathbf{V}$$

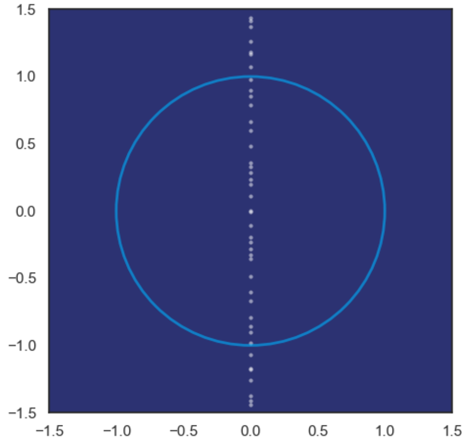




Hamiltonian weight

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{0} & \mathbf{A} \\ -\mathbf{A}^\top & \mathbf{0} \end{bmatrix}$$



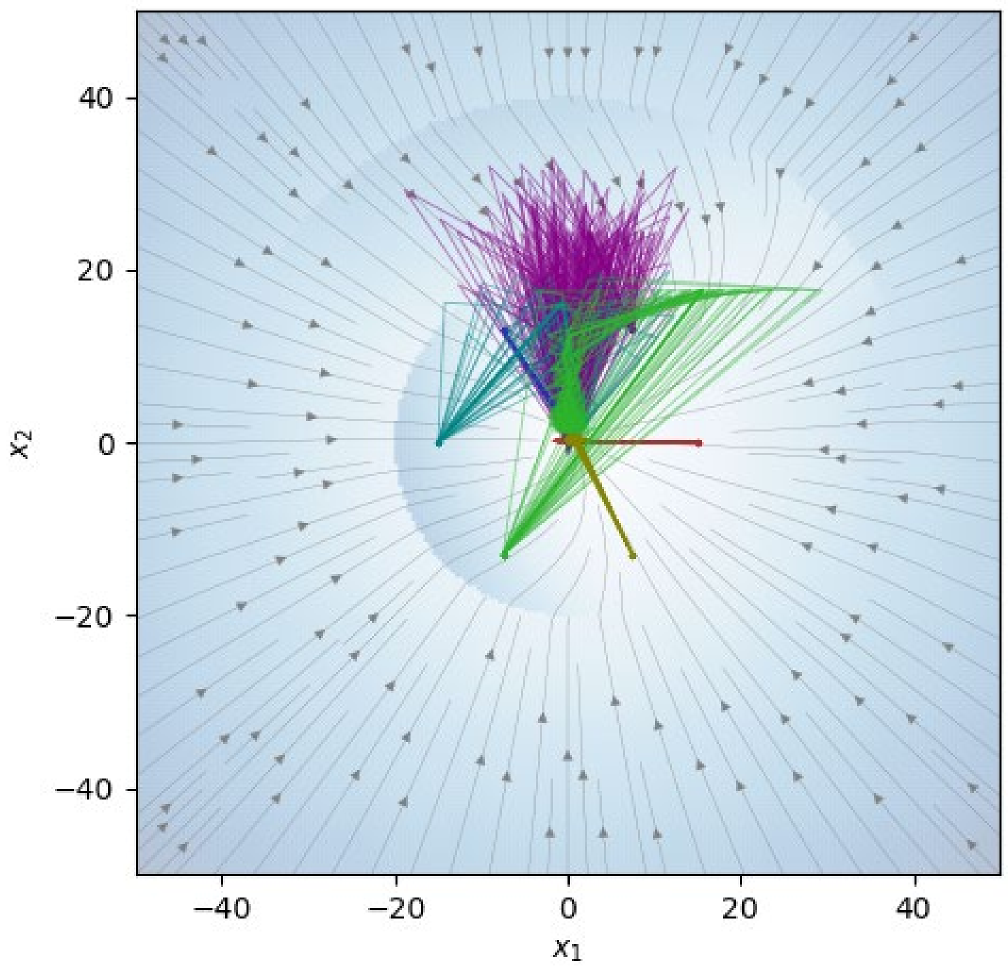


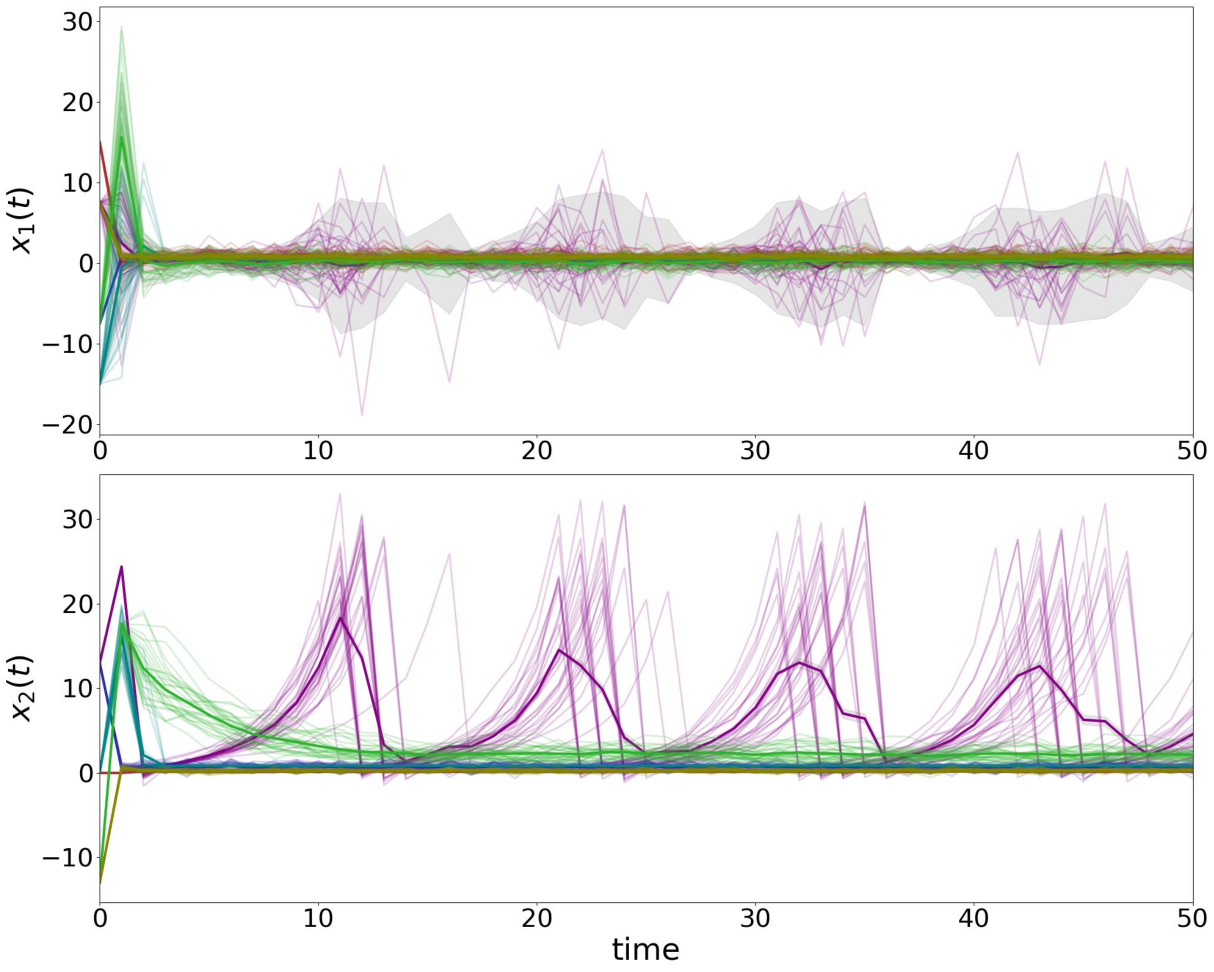
Pytorch implementation: https://github.com/pnnl/slim

Jiong Zhang, et al., Stabilizing Gradients for Deep Neural Networks via Efficient SVD Parameterization, 2018.
Eldad Haber and Lars Ruthotto, Stable Architectures for Deep Neural Networks 2017

Parametric Stability Constraints for DMMs

Constrained operator norms of DMMs: $\underline{\mathbf{p}}(\mathbf{x}) < \|\mathbf{A_f}(\mathbf{x})\|_p < \overline{\mathbf{p}}(\mathbf{x})$





Conclusion

- **Stability of Deep Markov Models**
  - Neural Networks as PWA maps
  - Contraction of PWA maps
  - Banach fixed point theorem
  - Operator norm constraints
  - Contraction of DMMs
- **Stable Weights**
  - Structured linear maps in Pytorch
  - https://pnnl.github.io/slim/
- **Contact**
  - jan.drgona@pnnl.gov
  - https://www.linkedin.com/in/drgona/
  - https://twitter.com/jan_drgona