

---

# On the Stochastic Stability of Deep Markov Models

---

Ján Drgoňa<sup>1</sup>, Sayak Mukherjee<sup>1</sup>, Jiaxin Zhang<sup>2</sup>, Frank Liu<sup>2</sup>, Mahantesh Halappanavar<sup>1</sup>

<sup>1</sup> Pacific Northwest National Laboratory  
Richland, Washington, USA

<sup>2</sup> Oak Ridge National Laboratory  
Oak Ridge, Tennessee, USA

{jan.drgona, sayak.mukherjee, mahantesh.halappanavar}@pnnl.gov  
{zhangj, liufy}@ornl.gov

## 1 Supplementary Material

### 1.1 Background

In this section, we recall few important mathematical definitions and theorems used in the paper.

**Definition 1.** *Induced operator norm of a matrix  $\mathbf{A} \in \mathbb{R}^{n \times m}$  is defined as:*

$$\|\mathbf{A}\|_p = \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p} = \max_{\|\mathbf{x}\|_p=1} \|\mathbf{Ax}\|_p, \quad \forall \mathbf{x} \in \mathcal{X}, \quad (1)$$

where  $\mathcal{X}$  is a compact normed vector space, and  $\|\cdot\|_p : \mathbb{R}^n \rightarrow \mathbb{R}$  represents vector  $p$ -norm inducing the matrix norm  $\|\mathbf{A}\|_p : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$ . In case of Euclidean norm  $\|\cdot\|_2$ , the operator norm corresponds to the largest singular value  $\sigma_{\max}(\mathbf{A})$ :

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 = \sigma_{\max}(\mathbf{A}). \quad (2)$$

The matrix norm is sub-additive:

$$\|\mathbf{A} + \mathbf{B}\|_p \leq \|\mathbf{A}\|_p + \|\mathbf{B}\|_p. \quad (3)$$

Induced  $p$ -norm  $\|\cdot\|_p : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$  is called submultiplicative if it satisfies [Malek-Shahmirzadi, 1983]:

$$\|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{B}\|_p. \quad (4)$$

**Remark 2.** For commonly used norms  $\|\cdot\|_p$  where  $p \in \{1, 2, \infty\}$  the submultiplicativity (4) holds.

**Definition 3.** *Asymptotic stability of the dynamical system for a bounded initial condition  $\mathbf{x}_0$  implies that its states converge to the equilibrium point  $\bar{\mathbf{x}}_e$ :*

$$\|\mathbf{x}_0 - \bar{\mathbf{x}}_e\| < \delta \implies \lim_{t \rightarrow \infty} \|\bar{\mathbf{x}}_t\| = \bar{\mathbf{x}}_e \quad (5)$$

**Definition 4.** *Given a metric space  $(\mathcal{X}, d)$ , a mapping  $T : \mathcal{X} \rightarrow \mathcal{X}$  is called contractive if there exist a constant  $c \in [0, 1)$  and a metric  $d$  such that following holds:*

$$d(T(\mathbf{x}_1), T(\mathbf{x}_2)) \leq cd(\mathbf{x}_1, \mathbf{x}_2), \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \quad (6)$$

Contraction generalizes to maps  $T : \mathcal{X} \rightarrow \mathcal{Y}$  between two metric spaces  $(\mathcal{X}, d)$ , and  $(\mathcal{Y}, d')$  as:

$$d'(T(\mathbf{x}_1), T(\mathbf{x}_2)) \leq cd(\mathbf{x}_1, \mathbf{x}_2), \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \quad (7)$$

**Definition 5.** *An affine map  $T(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}$  with  $\mathbf{x} \in \mathbb{R}^n$  and metric  $d = \|\cdot\|_p$  is a contraction mapping if the spectral norm of its linear component  $\mathbf{A}$  is less than one, i.e.  $\|\mathbf{A}\|_2 < 1$ .*

**Theorem 6.** *Banach fixed-point theorem. Lets have non-empty complete metric space  $(\mathcal{X}, d)$  then every contraction (6) is converging towards a unique fixed point  $T(\mathbf{x}_{ss}) = \mathbf{x}_{ss}$ .*

## 1.2 Stable Deep Markov Models with Bounded Equilibria

**Corollary 7.** *An equilibrium  $\bar{\mathbf{x}}_e$  of a mean-square stable deep Markov model is bounded by  $\underline{\mathbf{x}} \leq \|\bar{\mathbf{x}}_e\|_p \leq \bar{\mathbf{x}}$  if either of the conditions in Theorem 3 or Corollary 5 are satisfied, and the following holds  $\forall \mathbf{x} \in \text{dom}(\mathbf{f}_{\theta_f}(\mathbf{x}))$ .*

$$\bar{\mathbf{x}} \leq \frac{\|\mathbf{b}_f(\mathbf{x})\|_p}{1 - \|\mathbf{A}_f(\mathbf{x})\|_p}, \quad \frac{\|\mathbf{b}_f(\mathbf{x})\|_p}{1 + \|\mathbf{A}_f(\mathbf{x})\|_p} \leq \underline{\mathbf{x}} \quad (8)$$

*Proof.* The conditions in Theorem 3 and Corollary 5 imply mean square stability of the DMM with a mean dynamics converging towards a stable equilibrium  $\bar{\mathbf{x}}_e$ :

$$\bar{\mathbf{x}}_e = \mathbf{f}_{\theta_f}(\bar{\mathbf{x}}_e) = \lim_{t \rightarrow \infty} \mathbf{f}_{\theta_f}(\bar{\mathbf{x}}_t) \quad (9)$$

Now by substitution of the PWA form of the DNN into (9) we get:

$$\bar{\mathbf{x}}_e = \mathbf{A}_f(\bar{\mathbf{x}}_e)\bar{\mathbf{x}}_e + \mathbf{b}_f(\bar{\mathbf{x}}_e) \quad (10)$$

Where the matrix  $\mathbf{A}_f(\bar{\mathbf{x}}_e)$  and bias vector  $\mathbf{b}_f(\bar{\mathbf{x}}_e)$  uniquely define the affine equilibrium dynamics.

For the upper bound in (8), we apply operator norm (1) to the equation (10) to obtain:

$$\|\bar{\mathbf{x}}_e\|_p = \|\mathbf{A}_f(\bar{\mathbf{x}}_e)\bar{\mathbf{x}}_e + \mathbf{b}_f(\bar{\mathbf{x}}_e)\|_p \quad (11)$$

Then applying triangle inequality (3) and operator upper bound  $\|\mathbf{A}\mathbf{x}\|_p \leq \|\mathbf{A}\|_p\|\mathbf{x}\|_p$  we get:

$$\|\bar{\mathbf{x}}_e\|_p \leq \|\mathbf{A}_f(\bar{\mathbf{x}}_e)\|_p\|\bar{\mathbf{x}}_e\|_p + \|\mathbf{b}_f(\bar{\mathbf{x}}_e)\|_p \quad (12)$$

And by applying straightforward algebra we have:

$$(1 - \|\mathbf{A}_f(\bar{\mathbf{x}}_e)\|_p)\|\bar{\mathbf{x}}_e\|_p \leq \|\mathbf{b}_f(\bar{\mathbf{x}}_e)\|_p \quad (13)$$

With resulting equilibrium upper bound given as:

$$\|\bar{\mathbf{x}}_e\|_p \leq \frac{\|\mathbf{b}_f(\bar{\mathbf{x}}_e)\|_p}{1 - \|\mathbf{A}_f(\bar{\mathbf{x}}_e)\|_p} \quad (14)$$

For deriving the lower bound in (8), we start with straightforward algebraic operations on (10) to obtain equilibrium dynamics in a form:

$$(\mathbf{I} - \mathbf{A}_f(\bar{\mathbf{x}}_e))\bar{\mathbf{x}}_e = \mathbf{b}_f(\bar{\mathbf{x}}_e) \quad (15)$$

Again for two equivalent vectors their norms must be equal:

$$\|(\mathbf{I} - \mathbf{A}_f(\bar{\mathbf{x}}_e))\bar{\mathbf{x}}_e\|_p = \|\mathbf{b}_f(\bar{\mathbf{x}}_e)\|_p \quad (16)$$

Now applying operator norm upper bound inequality  $\|\mathbf{A}\mathbf{x}\|_p \leq \|\mathbf{A}\|_p\|\mathbf{x}\|_p$  to (16) we have:

$$\|(\mathbf{I} - \mathbf{A}_f(\bar{\mathbf{x}}_e))\|_p\|\bar{\mathbf{x}}_e\|_p \geq \|\mathbf{b}_f(\bar{\mathbf{x}}_e)\|_p \quad (17)$$

$$\|\bar{\mathbf{x}}_e\|_p \geq \frac{\|\mathbf{b}_f(\bar{\mathbf{x}}_e)\|_p}{\|\mathbf{I} - \mathbf{A}_f(\bar{\mathbf{x}}_e)\|_p} \quad (18)$$

Then from triangle inequality  $\|\mathbf{I} - \mathbf{A}_f(\bar{\mathbf{x}}_e)\|_p \leq \|\mathbf{I}\| + \|\mathbf{A}_f(\bar{\mathbf{x}}_e)\|_p$  we obtain a lower bound on the equilibrium norm as follows:

$$\|\bar{\mathbf{x}}_e\|_p \geq \frac{\|\mathbf{b}_f(\bar{\mathbf{x}}_e)\|_p}{1 + \|\mathbf{A}_f(\bar{\mathbf{x}}_e)\|_p} \quad (19)$$

Now clearly if the conditions of Corollary 5 are satisfied then the conditions (14) and (19) hold.  $\square$

## 1.3 Stability Regularizations for Deep Markov Models

This section proposes additional regularization methods for learning stable deep Markov models. The most direct approach is to include the stability conditions as extra penalties in the DMM loss function.

$$\mathcal{L}_{\text{stable}} = \max(1, \|\mathbf{A}_f(\mathbf{x})\|_p) + \max(K, \|\mathbf{A}_g(\mathbf{x})\|_p + \frac{\|\mathbf{b}_g(\mathbf{x})\|_p}{\|\mathbf{x}\|_p}) \quad (20)$$

Then the stability will be enforced by assigning large penalty weights to (20). Additionally, we can bound the norms of the equilibria of DMMs by enforcing constraints (8) via following penalties.

$$\mathcal{L}_{\text{bounded}} = \max(0, \underline{\mathbf{x}} - \frac{\|\mathbf{b}_f(\mathbf{x})\|_p}{1 - \|\mathbf{A}_f(\mathbf{x})\|_p}) + \max(0, \frac{\|\mathbf{b}_f(\mathbf{x})\|_p}{1 - \|\mathbf{A}_f(\mathbf{x})\|_p} - \bar{\mathbf{x}}) \quad (21)$$

We leave the empirical validation of the proposed regularizations (20) and (21) for future work.

