Stability Analysis of Discrete time Recurrent Neural Networks

Dr. Nikita Barabanov, Jayant Singh*

Department of Mathematics
North Dakota State University
Fargo,ND

November 8, 2014
2014 AMS Fall Southeastern Sectional Meeting
University of North Carolina, Greensboro
Special Session on Difference Equations and its Applications

Outline

- Problem Statement
- Absolute stability approach
- Novel Approach-Reduction of Dissipativity domain
- Summary

Problem Setting

Consider an example of discrete time Recurrent Neural Network (RNN)

$$\begin{aligned} x_1^{k+1} &= \tanh(W_1 x_1^k + V_n x_n^k + b_1) \\ x_2^{k+1} &= \tanh(W_2 x_2^k + V_1 x_1^{k+1} + b_2) \\ &\dots \\ x_n^{k+1} &= \tanh(W_n x_n^k + V_{n-1} x_{n-1}^{k+1} + b_n) \end{aligned} \tag{1}$$

where x_n^k is the state vector of *nth* layer at step k,
W_n, V_n are weight matrices, and b_n represents the bias vector.

Objective: Find the stability criterion for the system above.

- Consider a discrete time MIMO system:

$$\begin{aligned} x^{k+1} &= Ax^k + B\xi^k, \sigma^k = Cx^k \\ \xi_i^k &= \varphi_i(\sigma_i^k), i = 1 \dots m, \end{aligned} \tag{2}$$

  where, A ,B, C are matrices, $\xi^k = (\xi_1^k, \dots, \xi_m^k)$, and $\sigma^k = (\sigma_1^k, \dots, \sigma_m^k)$.
- Develop stability criterion for (2), where $\varphi(\cdot)$ is such that
  (i) $\varphi_i(0) = 0$, and
  (ii) $0 \leq \frac{\varphi_i(s)}{s} \leq \mu_i$, for some μ_i.

Lyapunov function approach

- Consider $V(x) = x^* H x$, where $H = H^* > 0$.
  Then,

  $$V(x^{k+1}) - V(x^k) = (Ax^k + B\xi^k)^* H (Ax^k + B\xi^k) - (x^k)^* H x^k$$

- We want $V(x^{k+1}) - V(x^k) < 0$ for all $(x^k, \xi^k) \neq 0$ such that $\xi^k = \varphi(x^k)$.

Reformulated Problem

Subproblem: Suppose F is a quadratic function. Moreover, assume there exists matrix L such that $A + BL$ is stable (i.e.(A,B) is stabilizable), and $F(x, Lx) \geq 0$. Find necessary and sufficient conditions for the existence of $H = H^* > 0$ s.t.

$$(Ax + B\xi)^* H(Ax + B\xi) - x^* Hx + F(x, \xi) < 0 \tag{3}$$

for all $(x, \xi) \neq 0$.
Solution: Necessary condition is given by
$\mathbb{R}e(F((e^{i\omega} I - A)^{-1} Bw, w)) < 0$ for all $\omega \in [0, \pi]$ and $w \neq 0$, called the Frequency domain condition.
And sufficient condition is provided by Kalman Szegö Lemma.

Kalman Szegö Lemma

Lemma 1.

Assume (A, B) is stabilizable. Moreover, $\mathbb{R}e(F((e^{i\omega}I - A)^{-1}Bw, w)) < 0$ for all $\omega \in [0, \pi]$ and $w \neq 0$. Then there exists $H = H^*$ such that

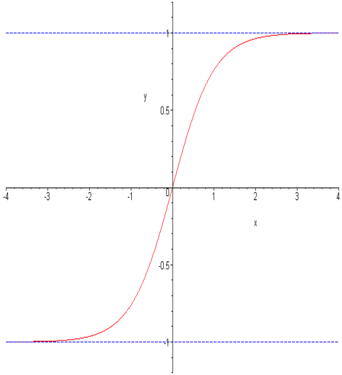
$$(Ax + B\xi)^* H(Ax + B\xi) - x^* Hx + F(x, \xi) < 0$$

for all $(x, \xi) \neq 0$.

- As a consequence, there exists $H = H^* > 0$ such that $x^* Hx$ is a Lyapunov function.

Application to RNN Stability

- In case of RNN,$\varphi(\cdot) = \tanh(\cdot)$,
- $0 \leq \frac{\tanh(\sigma)}{\sigma} \leq 1$(sector condition)
- $\varphi(\sigma)(\sigma - \varphi(\sigma)) \geq 0$ is the quadratic function, F.



System Transformation

Automatic Control form:

$$\begin{aligned} x^{k+1} &= Ax^k + B\xi^k, \\ \sigma^k &= \Theta x^k, \end{aligned} \tag{4}$$

where

$$\xi^k = \begin{pmatrix} \xi_1^k \\ \cdots \\ \xi_m^k \end{pmatrix}, \sigma^k = \begin{pmatrix} \sigma_1^k \\ \cdots \\ \sigma_m^k \end{pmatrix}, \xi_i = \varphi_i(\sigma_i) \text{ for all } i = 1 \ldots m.$$

State Space Extension

- Consider a two layer RNN:

$$\begin{aligned} x_1^{k+1} &= \tanh(W_1 x_1^k + V_2 x_2^k) \\ x_2^{k+1} &= \tanh(W_2 x_2^k + V_1 x_1^{k+1}). \end{aligned} \tag{5}$$

where x_1, x_2 define the state vectors for the layers.

- Transformed form:

$$\begin{aligned} x_{11}^{k+1} &= \tanh(W_1 x_{12}^k + V_2 x_{21}^k) \\ x_{12}^{k+1} &= x_{11}^k \\ x_{21}^{k+1} &= \tanh(W_2 x_{22}^k + V_1 x_{11}^k) \\ x_{22}^{k+1} &= x_{21}^k. \end{aligned} \tag{6}$$

Independent Processes

The extended system can be decomposed into two independent processes given by

- $$\begin{aligned} x_{12}^{k+2} &= \phi_1(x_{12}^k, x_{21}^k) \\ x_{21}^{k+2} &= \phi_2(x_{21}^k, \phi_1(x_{12}^k, x_{21}^k)) \end{aligned} \tag{7}$$

- $$\begin{aligned} x_{22}^{k+2} &= \phi_2(x_{11}^k, x_{22}^k) \\ x_{11}^{k+2} &= \phi_2(x_{11}^k, \phi_1(x_{11}^k, x_{22}^k)) \end{aligned} \tag{8}$$

- It can be checked that systems (7) and (8) are identical to original system (5)

Discussion about Theory of Absolute stability

To recapitulate,

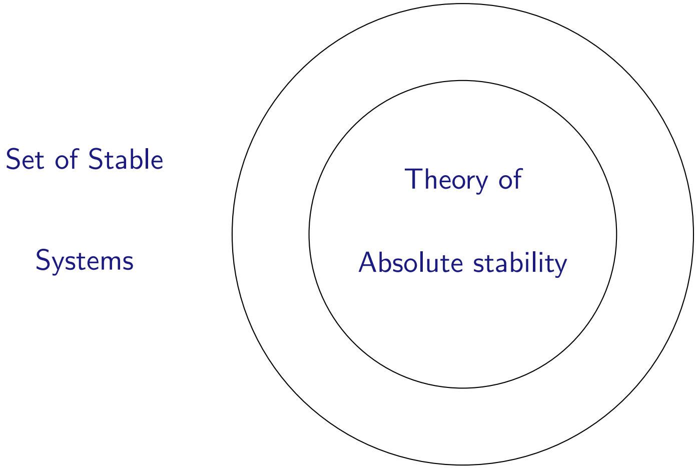
- Consider the system defined by

$$\begin{aligned} x^{k+1} &= Ax^k + B\xi^k, \sigma^k = Cx^k \\ \xi^k &= \varphi(\sigma^k) \end{aligned} \tag{9}$$

  where $\varphi(\cdot)$ is a function such that $\varphi(0) = 0$, and $0 \leq \frac{\varphi(\sigma)}{\sigma} \leq \mu$, for some μ.

- The above system is globally asymptotically stable if there exists $H = H^* > 0$ such that $(Ax + B\xi)^* H(Ax + B\xi) - x^* Hx + F(x, \xi) < 0$ for all $(x, \xi) \neq 0$ where $F(x, \xi) = \xi(Cx - \frac{1}{\mu}\xi) \geq 0$.

Shortcomings in absolute stability approach



- A more general stability criteria should be developed.

Method of reduction of Dissipativity domain

- $x^{k+1} = \phi(x^k)$, $\phi(\cdot)$ is bounded non-linear function.
- Construct $\{D_k\}$ such that $D_{k+1} \subsetneq D_k, \phi(D_k) \subset D_{k+1}$ then $x^k \in D_k$, provided that $x^0 \in D_0$. Thus if $\{D_k\} \to 0$, then $x^k \to 0$, as $k \to \infty$.
- $D_{k+1} = \{x \in D_k : f_{k,j}(x) \leq \alpha_{k+1,j}, j = 1 \dots m_{k+1}\}$ where m_{k+1} defines the number of constraints at time step $k+1$, $f_{k,j}$ defines the linear function, $\alpha_{k+1,j} := \max_{x \in D_k} f_{k,j}(\phi(x))$.

Convex Lyapunov function and constrained optimization problem

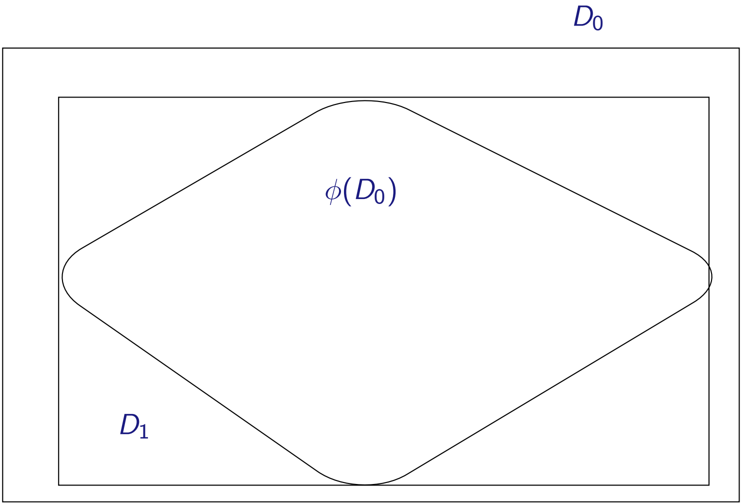
Theorem 2.

Define $\alpha_j^{k+1} = \max_{y \in D_k}(f_j^k(\phi(y)))$. *Assume system* $x^{k+1} = \phi(x^k)$ *has a convex Lyapunov function.Then there exists linear functions* $f_1, f_2, \ldots f_m$ *such that* $D_{k+1} = \{y : f_j^{k+1}(y) \leq \alpha_j^{k+1}, j = 1 \ldots m\}$, *and* $\{D_k\} \to 0$.

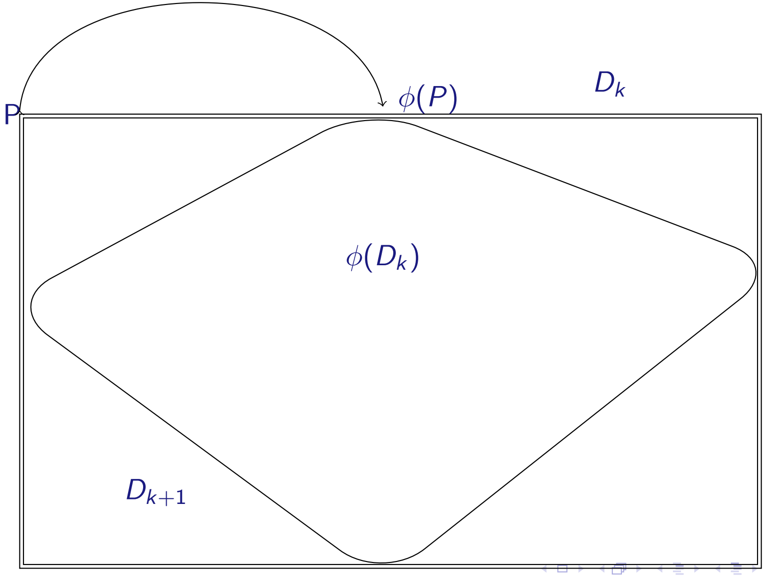
Application to RNN stability problem

- In case of RNN, $\phi(\cdot) = \tanh(\cdot)$.
- We define $D_{k+1} = \{y : f_j(y) = \langle l_j, y \rangle \leq \alpha_j^{k+1},\ j = 1 \dots m\}$, and l is unit normal vector
- Algorithm:
  1. Define $D_k = \{x : \mid x \mid \leq \alpha_j^k,\ j = 1 \dots m\}$ when $k = 0$.
  2. Find $\max_{x \in D_k} \langle l_j, (\tanh(x)) \rangle := \alpha_j^{k+1}$ for all j and define $D_{k+1} = \{y : \langle l_j, y \rangle \leq \alpha_j^{k+1}\}$
  3. If $\max_j(\alpha_j^k - \alpha_j^{k+1}) > \varepsilon > 0$, increase k by 1 and go to step 2 and repeat. Here ε is some fixed threshold.

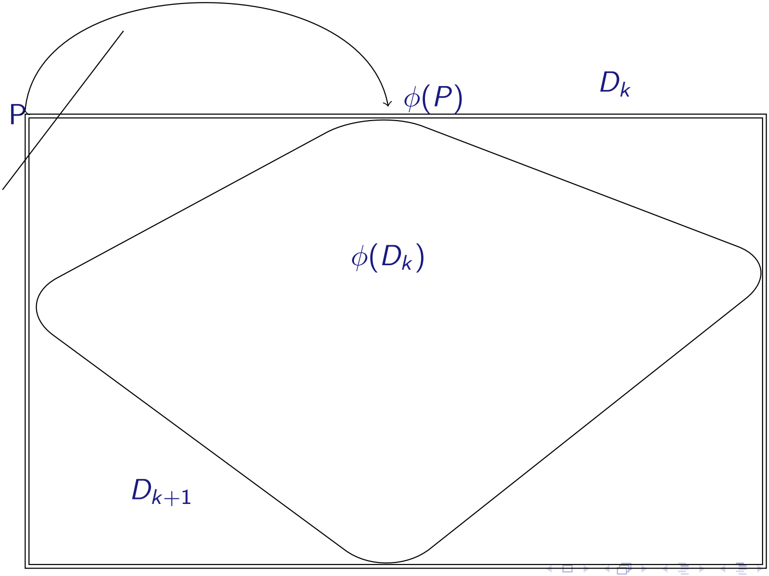
Construction of sets



Addition of new constraints



Addition of new constraints contd.



Problem Under Consideration

Problem Statement: Given the function $f(x) = \sum_{i=1}^{n} c_i \phi(x_i)$, where $c_i \neq 0$ for all i. How to locate the points of local maxima for $f(\cdot)$ over a convex set(for our case it is a rectangle)?

Identify point of local maxima

Subproblem: Consider the hyperplane $P = \{x : l^T x = b\}$ where l is unit normal vector and $b \in \mathbb{R}$.Suppose the function $f(x) = \sum_{i=1}^{n} c_i \phi(x_i)$ defined on P has a critical point x_0. Is x_0 a point of local maximum or not?

Solution: Necessary condition :

Theorem 3.

Suppose $l \| \frac{\partial f}{\partial x}(x_0) := \overrightarrow{\nabla} f(x_0)$*. Then* x_0 *is a point of local maximum of* $f(\cdot)$ *over* P *only if* $K \leq 0$*, where* $K = (I - \frac{ll^T}{\|l\|^2}) D (I - \frac{ll^T}{\|l\|^2})$ *is the projection matrix, and* $D = D^T = \frac{\partial^2 f}{\partial x^2}(x_0) = diag(d_j)_{j=1}^{n}$.

Point of Local maximum on Hyperplane

Theorem 4 (-).

Consider the hyperplane $P = \{x : l^T x = b\}$ where l is unit normal vector and $b \in \mathbb{R}$. Suppose the function $f(x) = \sum_{i=1}^{n} c_i \phi(x_i)$ defined on P. Assume the function $\phi(\cdot)$ satisfies the following conditions:

(i) $\phi(\cdot) \in C^2, \phi(-x) = -\phi(x), \phi'(x) > 0, x\phi''(x) < 0,$ *for all* $x \neq 0$, *and* $\lim_{x\to\infty} \phi(x) < \infty$. *Denote* $\psi(\cdot) = (\phi'(\cdot))^{-1}$.

(ii) $x(\ln|\psi'(x)|)'$ *is a monotonically increasing function of* x.

(iii) *Set* $h(\beta q_j) = \frac{\psi'(\beta q_j)}{\psi'(\beta q_n)}$. *Then* $\frac{d}{d\beta}\left[\frac{h'(\beta q_j)}{h'(\beta q_l)}\right] \neq 0$, *where* $q_j < q_n < q_l$.

(iv) *For all* $p > q$, *we have* $\frac{d}{d\beta}\left(\frac{\psi(\beta p)}{\psi(\beta q)}\right) < 0$.

(v) *For all* $x > 0$, *we have* $\frac{d}{dx}\left(x\frac{d}{dx}\left(\frac{\psi(x)}{x\psi'(x)}\right)\right) \geq 0$.

Thm. contd.

Theorem 4 (-).

Then, the function $f(x)$ has at most one point of local maximum on the hyperplane, P.

Points of local maxima on Intersection of Hyperplanes

- *Problem Setting*: Given the function $f(x) = \sum_{i=1}^{n} c_i \phi(x_i)$, where $c_i \neq 0$ for all i. Consider the intersection of m hyperplanes in n- dimensional space. How many points of local maxima does $f(\cdot)$ have?
- We need to develop necessary conditions for existence of local maxima.

Subproblem: Consider the function, $f(x) = \sum_{i=1}^{n} c_i \phi(x_i)$. Suppose the pair (L, α) defines the hyperplanes, where $L \sim n \times m$, and $\alpha \sim m \times 1$. Assume that x_0 is a critical point for $f(\cdot)$ on $n - m$ dimensional plane. Then $\overrightarrow{\nabla} f(x_0) = L \cdot \beta$. Is x_0 a point of local maximum?

Theorem 5 (-).

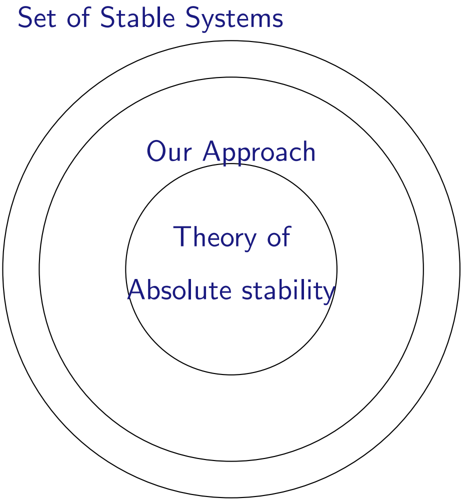
The point x_0 is a point of local maximum only if the following conditions are met,

- $Q' \cdot C \cdot \psi'(Q \cdot \beta) \cdot Q > 0$, *and*
- $Q' \cdot C \cdot \psi(Q \cdot \beta) = \alpha$

where, $\psi(\cdot) = (\phi')^{-1}, C = diag(c_j)_{j=1}^{n}, Q = C^{-1} \cdot L, \psi'(\cdot) = diag[\psi'(Q \cdot \beta)_j]_{j=1}^{n}, \psi(\cdot) = \{\psi(Q \cdot \beta)_j\}_{j=1}^{n}$.

Discussion about new approach

- We check existence of convex Lyapunov function. On the other hand, in theory of absolute stability approach, we considered the problem of existence of quadratic Lyapunov function(a type of convex function).



Summary

- The stability problem of discrete time RNN is studied
- Results from Theory of Absolute stability has been discussed
- The method of Reduction of Dissipativity domain has been presented
- Some new results have been discussed