Stable learning

Wen Yu

Departamento de Control Automático
CINVESTAV-IPN
A.P. 14-740, Av.IPN 2508, México D.F., 07360, México

Input-to-state stability (ISS)

Definition

A system is said to be globally *input-to-state stability* if there exists a $\mathcal{K}$-function $\gamma(\cdot)$ (continuous and strictly increasing $\gamma(0) = 0$) and $\mathcal{KL}$ -function $\beta(\cdot)$ ($\mathcal{K}$-function and $\lim_{s_k \to \infty} \beta(s_k) = 0$) , such that, for each $u \in L_\infty$ ($\sup\{\|u(k)\|\} < \infty$) and each initial state $x^0 \in R^n$, it holds that

$$\left\|x\left(k, x^0, u(k)\right)\right\| \leq \beta\left(\left\|x^0\right\|, k\right) + \gamma\left(\|u(k)\|\right)$$

Definition

A smooth function $V : \Re^n \to \Re \geq 0$ is called a smooth ISS-Lyapunov function for system if:

(a) there exists a $\mathcal{K}_\infty$-function $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ ($\mathcal{K}$-function and $\lim_{s_k \to \infty} \beta(s_k) = \infty$) such that

$$\alpha_1(s) \leq L(s) \leq \alpha_2(s), \quad \forall s \in \Re^n$$

Input-to-state stability (ISS)

Stable training law without dead-zone modification is

$$W_{k+1} = W_k - s_k \eta_k \phi \left[V_k X(k) \right] e(k)$$
$$V_{k+1} = V_k - s_k \eta_k W^0 \phi' X(k)\, e(k)$$

where

$$\eta_k = \frac{\eta_0}{1 + \left\| \phi V_k X(k) \right\|^2 + \left\| W^0 \phi' X(k) \right\|^2}, \qquad 0 \le \eta_0 \le 1$$

When

$$2\eta_k e(k)[\delta(k)+\xi(k)] \le \eta_k e^2(k) + \eta_k [\delta(k)+\xi(k)]^2$$
$$\le \eta_k e^2(k) + \eta_k \bar{\xi} + \bar{\delta}$$

$$\Delta L_k \le \eta_k \left[\frac{\left(\|\phi V_k X(k)\|^2 + \left\|W^0 \phi' X(k)\right\|^2\right)}{1+\|\phi V_k X(k)\|^2 + \|W^0 \phi' X(k)\|^2} \eta_0 - 1 \right] e^2(k) + \eta_k [\delta(k)+\xi(k)]^2$$
$$\le -\pi_k e^2(k) + \eta_k \epsilon^2(k)$$

π_k is an $\mathcal{K}_\infty$-function, $\eta_k [\delta(k)+\xi(k)]^2$ is a $\mathcal{K}$-function. L_k is the

function of $e(k)$ and $\epsilon(k) = \delta(k) + \xi(k)$, $n\left[\min\left(\widetilde{w}_i^2\right) + \min\left(\widetilde{v}_i^2\right)\right]$ and $n\left[\max\left(\widetilde{w}_i^2\right) + \max\left(\widetilde{v}_i^2\right)\right]$ are $\mathcal{K}_\infty$-functions, so L_k admits the smooth ISS-Lyapunov function. If the input

$$\|\epsilon(k)\|_2 = \|\delta(k)\|_2 + \|\xi(k)\|_2 \le \bar{\xi} + \bar{\delta}$$

is bounde, then the state $e^2(k)$ is bounded

Recurrent neural networks

- Plant: discrete-time state-space nonlinear system

$$x(k+1) = f\left[x\left(k\right), u\left(k\right)\right]$$

- NN model

$$\widehat{x}\left(k+1\right) = A\widehat{x}\left(k\right) + \sigma\left[W_1\left(k\right)x\left(k\right)\right] + \phi\left[W_2\left(k\right)x\left(k\right)\right]u\left(k\right)$$

- Learning methods
  1. BPTT
  2. Stable learning

Recurrent neural network

Discrete-time recurrent neural networks

$$\hat{x}(k+1) = A\hat{x}(k) + W_1\phi[V_1\hat{x}(k)] + W_2\phi[V_2\hat{x}(k)]\,u(k)$$

Simplest case, RNN is

$$\hat{x}(k+1) = A\hat{x}(k) + W\phi[x(k)]$$

Unknown nonlinear system is

$$x(k+1) = f[x(k)]$$

From generalized Stone-Weierstrass theorem, the nonlinear system can be written as

$$x(k+1) = Ax(k) + W^{*}\phi[x(k)]$$

Recurrent neural networks

The modeling error is

$$e = \hat{x}\left(k\right) - x\left(k\right)$$

The error dynamic is

$$e\left(k+1\right) = Ae\left(k\right) + \tilde{W}\phi\left[x\left(k\right)\right]$$

We use the updating law as

$$W\left(k+1\right) = W\left(k\right) - \eta\phi e\left(k\right)$$

Lyapunov function is

$$L(k) = \left\| \tilde{W}(k) \right\|^2$$

So

$$\Delta L(k) = \left\| \tilde{W}(k) - \eta\phi e(k) \right\|^2 - \left\| \tilde{W}(k) \right\|^2$$
$$= \eta^2 \left\| e(k) \right\|^2 \phi^2 - 2\eta \left\| \tilde{W}\phi \right\| \left\| e(k) \right\|$$

The error dynamic

$$e(k+1) = Ae(k) + \tilde{W}\phi\left[x(k)\right]$$

$$\Delta L(k) = \eta^2 e^2(k)^2 \phi^2 - 2\eta \left\| e(k+1) - Ae(k) \right\| \left\| e(k) \right\|$$
$$= \eta^2 \left\| e(k) \right\|^2 \phi^2 - 2\eta \left\| e(k)^T e(k+1) - e(k)^T Ae(k) \right\|$$
$$\leq \eta^2 \left\| e(k) \right\|^2 \phi^2 - 2\eta \left\| e(k)^T e(k+1) \right\| + 2\eta \left\| e(k) \right\|^2_{\lambda_{\max}(A)}$$
$$\leq -2 \left\| e(k)^T e(k+1) \right\| + \left[2\lambda_{\max}(A) + \eta\phi^2\right] \left\| e(k) \right\|^2$$

We need

$$\left\| e(k)^T e(k+1) \right\| \to \| e(k) \|^2$$

Define RNN as

$$\beta \hat{x}(k+1) = A\hat{x}(k) + W\phi[x(k)]$$

where $\beta > 0$ is a design parameter. So the error dynamic is

$$\beta e(k+1) = Ae(k) + \tilde{W}\phi[x(k)]$$

and

$$\Delta L(k) \leq -2 \left\| e(k)^T e(k+1) \right\| + \left[2\lambda_{\max}(A) + \eta\phi^2 \right] \| e(k) \|^2$$

If $\|\beta e(k+1)\| \geq \|e(k)\|$

$$\begin{aligned}\Delta L(k) &\leq -2\|e(k)\|^2 + \left[2\lambda_{\max}(A) + \eta\phi^2\right]\|e(k)\|^2 \\ &= -2\left\{1 - \left[\lambda_{\max}(A) + \tfrac{1}{2}\eta\phi^2\right]\right\}\|e(k)\|^2\end{aligned}$$

We need

$$\lambda_{\max}(A) + \tfrac{1}{2}\eta\phi^2 < 1$$

$$\eta < 2\frac{1-\lambda_{\max}(A)}{\phi^2}, \qquad \eta_0 \leq 2\left[1 - \lambda_{\max}(A)\right]$$

$$\eta(k) = \frac{\eta_0}{1+\phi^2} \leq \frac{2\left[1-\lambda_{\max}(A)\right]}{1+\phi^2} < \frac{2\left(1-\lambda_{\max}(A)\right)}{\phi^2}$$

The RNN is

$$\beta \hat{x}(k+1) = A\hat{x}(k) + W\phi[x(k)], \qquad \beta > 0$$

The updating law as

$$W(k+1) = W(k) - \eta(k)\phi e(k)$$

where $\eta(k) = \frac{\eta_0}{1+\phi^2[x(k)]}$, $\eta_0 \leq 2[1 - \lambda_{\max}(A)]$,
If $\|\beta e(k+1)\| < \|e(k)\|$ we stop the updating, finally

$$\eta(k) = \begin{cases} \frac{\eta_0}{1+\phi^2[x(k)]} & \text{if } \beta\|e(k+1)\| \geq \|e(k)\| \\ 0 & \text{if } \beta\|e(k+1)\| < \|e(k)\| \end{cases}$$

Unmodeled dynamic

RNN is

$$\beta\hat{x}\left(k+1\right)=A\hat{x}\left(k\right)+W\phi\left[x\left(k\right)\right]$$

Unknown nonlinear system is

$$x\left(k+1\right)=f\left[x\left(k\right)\right]$$

The nonlinear system can be written as

$$x(k+1)=Ax\left(k\right)+W^{*}\phi\left[x\left(k\right)\right]+\xi\left(k\right)$$

where $\left\|\xi\left(k\right)\right\|^{2}\leq\bar{\xi}$

So the error dynamic is

$$\beta e(k+1) = Ae(k) + \tilde{W}\phi[x(k)] - \xi(k)$$

Lyapunov function is

$$L(k) = \left\|\tilde{W}(k)\right\|^2$$

then

$$\begin{aligned}
\Delta L(k) &= \eta^2 e^2(k)^2 \phi^2 - 2\eta \left\|-\beta e(k+1) + Ae(k) + \xi(k)\right\| \left\|e(k)\right\| \\
&= \eta^2 \left\|e(k)\right\|^2 \phi^2 - 2\eta \left\|-e(k)^T e(k+1) + e(k)^T Ae(k) + \xi(k) e(k)\right\| \\
&\leq \eta^2 \left\|e(k)\right\|^2 \phi^2 - 2\eta \left\|e(k)^T e(k+1)\right\| + 2\eta \left\|e(k)\right\|^2_{\lambda_{\max}(A)} + \eta \left\|\xi(k)\right\|^2 \\
&\leq -2\left\|e(k)^T e(k+1)\right\| + \left[1 + 2\lambda_{\max}(A) + \eta\phi^2\right] \left\|e(k)\right\|^2 + \left\|\xi(k)\right\|^2
\end{aligned}$$

Unmodeled dynamic

If $\|\beta e(k+1)\| \geq \|e(k)\|$

$$\begin{aligned}\Delta L(k) &\leq -2\|e(k)\|^2 + \left[1 + 2\lambda_{\max}(A) + \eta\phi^2\right]\|e(k)\|^2 + \|\xi(k)\|^2 \\ &= -\left[1 - 2\lambda_{\max}(A) - \eta\phi^2\right]\|e(k)\|^2 + \|\xi(k)\|^2\end{aligned}$$

We need

$$\begin{gathered}2\lambda_{\max}(A) + \eta\phi^2 < 1 \\ \eta < \tfrac{1-2\lambda_{\max}(A)}{\phi^2}, \qquad \eta_0 \leq 1 - 2\lambda_{\max}(A) \\ \eta(k) = \tfrac{\eta_0}{1+\phi^2} \leq \tfrac{1-2\lambda_{\max}(A)}{1+\phi^2} < \tfrac{1-2\lambda_{\max}(A)}{\phi^2}\end{gathered}$$

then

$$\Delta L(k) \leq -\frac{\eta_0}{1+\phi^2}\|e(k)\|^2 + \|\xi(k)\|^2 \leq \frac{\eta_0}{1+\bar{\phi}}\|e(k)\|^2 + \bar{\xi}$$

The average of the identification error satisfies

$$J = \limsup_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T} e^2(k) \leq \frac{(1+\bar{\phi})}{\eta_0}\bar{\xi}$$

Single-layer neural network

Discrete-time recurrent neural networks

$$\hat{x}(k+1) = A\hat{x}(k) + W_1\phi_1[\hat{x}(k)] + W_2\phi_2[\hat{x}(k)]U(k)$$

Theorem

If the single-layer neural network is used to identify nonlinear plant, the eigenvalues of A is selected as $-1 < \lambda(A) < 0$, the following gradient updating law can make the identification error $e(k)$ bounded (stable in an L_∞ sense)

$$\begin{aligned} W_1(k+1) &= W_1(k) - \eta(k)\phi_1[\hat{x}(k)]e^T(k) \\ W_2(k+1) &= W_2(k) - \eta(k)U(k)\phi_2[\hat{x}(k)]e^T(k) \end{aligned}$$

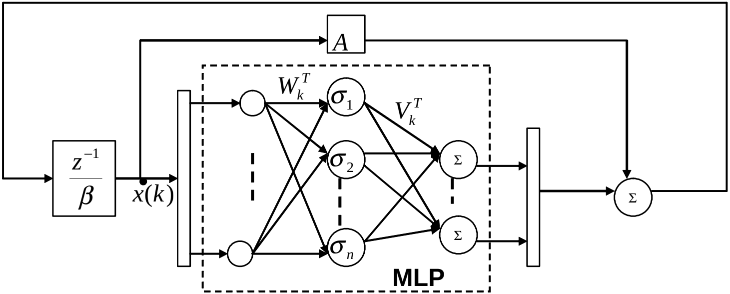
where $\eta(k)$ satisfies

$$\eta(k) = \begin{cases} \dfrac{\eta}{1+\|\phi_1\|^2+\|U(k)\phi_2\|^2} & \text{if } \beta\|e(k+1)\| \geq \|e(k)\| \\ 0 & \text{if } \beta\|e(k+1)\| < \|e(k)\| \end{cases}$$

Multilayer RNN

The discrete-time multilayer recurrent neural networks

$$\beta\hat{x}\left(k+1\right)=A\hat{x}\left(k\right)+W_k\sigma\left[V_k x\left(k\right)\right]$$



The smooth function f has Taylor formula as

$$f(x_1, x_2) = \sum_{k=0}^{l-1} \frac{1}{k!} \left[\left(x_1 - x_1^0\right) \frac{\partial}{\partial x_1} + \left(x_2 - x_2^0\right) \frac{\partial}{\partial x_2} \right]_0^k f + R_l$$

where R_l is the remainder of the Taylor formula.
Using Taylor series around the point of $W_k X(k)$ and V_k, the identification error $e(k) = \widehat{x}(k) - x(k)$ can be represented as

$$\beta e(k+1) = Ae(k) + \tilde{W}_k \phi [V_k x(k)] + W^* \phi' \tilde{V}_k x(k) + \zeta(k)$$

$\zeta(k) = R_1 + \mu(k)$, here R_1 is second order approximation error of the Taylor series.

Multilayer RNN

Theorem

For multilayer RNN, A is selected as $-1 < \lambda(A) < 0$, the following gradient updating law can make identification error $e(k)$ bounded

$$\begin{array}{l} W_{k+1} = W_k - \eta_k e(k)\,\phi \\ V_{k+1} = V_k - \eta_k e(k)\,\phi' W^0 x(k) \end{array}$$

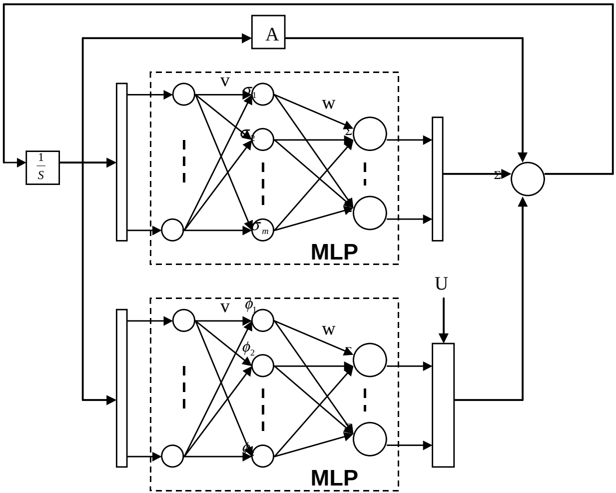
where $0 < \eta \leq 1$

$$\eta_k = \begin{cases} \dfrac{\eta}{1 + \|\phi' W^0\|^2 + \|\phi\|^2} & \text{if } \beta\,\|e(k+1)\| \geq \|e(k)\| \\ 0 & \text{if } \beta\,\|e(k+1)\| < \|e(k)\| \end{cases}$$

The average of the identification error satisfies

$$J = \limsup_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} e^2(k) \leq \frac{\eta}{\pi}\bar{\zeta}$$

Recurrent MultiLayer Perceptrons



Countinous-time RNN (Dynamic neural networks)

Plant

$$\dot{x}_t = f(x_t, u_t, t), \quad x_t \in \Re^n,\ u_t \in \Re^m$$

Model

$$\frac{d}{dt}\hat{x} = A\hat{x}_t + W_1\sigma(\hat{x}_t) + W_2\phi(\hat{x}_t)\gamma(u_t)$$

$\hat{x} \in \Re^n, u \in \Re^k$. The matrix $A \in \Re^{n\times n}$ is a stable matrix. The matrices $W_1 \in \Re^{n\times m}$ and $W_2 \in \Re^{n\times m}$ are the weights output layers. $V_1 \in \Re^{m\times n}$ and $V_2 \in \Re^{m\times n}$ are the weights of hidden layers. $\sigma(\cdot) \in \Re^m$ is sigmoidal vector functions, $\phi(\cdot)$ is $\Re^{m\times m}$ diagonal matrix, i.e.,

$$\phi(\cdot) = \text{diag}\left[\phi_1(V_2\hat{x})_1 \cdots \phi_m(V_2\hat{x})_m\right].$$

$\sigma_i(\cdot)$ and $\phi_i(.)$ are as sigmoid functions, i.e.

$$\sigma_i(x_i) = \frac{a_i}{1+e^{-b_i x_i}} - c_i.$$

Simplest case

The plant is

$$\dot{x} = f(x)$$

Continuous-time recurrent NN (series-parallel)

$$\frac{d}{dt}\hat{x} = A\hat{x} + W\sigma(x)$$

The nonlinear system is complete described by following neural network

$$\dot{x} = Ax + W^{*}\sigma(x)$$

The modeling error is

$$\Delta = \hat{x} - x$$

The error dynamic is

$$\dot{\Delta} = \frac{d}{dt}\hat{x} - \dot{x} = A\Delta + \tilde{W}\sigma(x)$$

where $\tilde{W} = W - W^{*}$

Lyapunov

Define Lyapunov function candidate as

$$L = \Delta^T P \Delta + \frac{1}{k} tr \left[\tilde{W}^T \tilde{W} \right]$$

then

$$\dot{L} = \dot{\Delta}^T P \Delta + \Delta^T P \dot{\Delta} + \frac{2}{k} tr \left[\left(\frac{d}{dt} \tilde{W} \right) \tilde{W} \right]$$

Using error dynamic $\dot{\Delta} = A\Delta + \tilde{W}\sigma(x)$

$$\dot{L} = \Delta^T \left(A^T P + PA \right) \Delta + 2\Delta^T P \tilde{W}^T \sigma(x) + \frac{2}{k} tr \left[\left(\frac{d}{dt} \tilde{W} \right) \tilde{W}^T \right]$$

If A is stabel

$$A^T P + PA = -Q, \qquad Q = Q^T > 0$$

so

$$\dot{L} = -\Delta^T Q \Delta + \left[2\Delta^T P \sigma(x) + \frac{2}{k} \left(\frac{d}{dt} \tilde{W} \right) \right] \tilde{W}^T$$

Lyapunov

If we let

$$2\Delta^T P\sigma(x) + \frac{2}{k}\left(\frac{d}{dt}\tilde{W}\right) = 0$$

The leaning law is

$$\dot{W} = \frac{d}{dt}\tilde{W} = -k\Delta^T P\sigma(x)$$

then

$$\dot{L} = -\Delta^T Q\Delta$$

So $\Delta^T \in L_2 \cap L_\infty$, from the error dynamic, $\dot{\Delta} \in L_\infty$. Using Barlalat's Lemma, we have

$$\lim_{t\to\infty} \Delta = 0$$

The plant is

$$\dot{x} = f(x, u)$$

Continuous-time recurrent NN (series-parallel)

$$\frac{d}{dt}\hat{x} = A\hat{x} + W_1\sigma(x) + W_2\phi(x)\gamma(u)$$

The nonlinear system is complete described by following neural network

$$\dot{x} = Ax + W_1^*\sigma(x) + W_2^*\phi(x)\gamma(u)$$

The error dynamic is

$$\dot{\Delta} = \frac{d}{dt}\hat{x} - \dot{x} = A\Delta + \tilde{W}_1\sigma(x) + \tilde{W}_2\phi(x)\gamma(u)$$

where $\tilde{W}_1 = W_1 - W_1^* \tilde{W}_2 = W_2 - W_2^*$

With control

Lyapunov function candidate as

$$L = \Delta^T P \Delta + \frac{1}{k_1} tr\left[\tilde{W}_1^T \tilde{W}_1\right] + \frac{1}{k_2} tr\left[\tilde{W}_2^T \tilde{W}_2\right]$$

then

$$\dot{L} = \dot{\Delta}^T P \Delta + \Delta^T P \dot{\Delta} + \frac{2}{k_1} tr\left[\left(\frac{d}{dt}\tilde{W}_1\right)\tilde{W}_1^T\right] + \frac{2}{k_2} tr\left[\left(\frac{d}{dt}\tilde{W}_2\right)\tilde{W}_2^T\right]$$

Using error dynamic $\dot{\Delta} = A\Delta + \tilde{W}_1\sigma(x) + \tilde{W}_2\phi(x)\gamma(u)$

$$\dot{L} = -\Delta^T Q \Delta + \left[2\Delta^T P\sigma(x) + \tfrac{2}{k_1}\left(\tfrac{d}{dt}\tilde{W}_1\right)\right]\tilde{W}_1^T$$
$$+ \left[2\Delta^T P\phi(x)\gamma(u) + \tfrac{2}{k_2}\left(\tfrac{d}{dt}\tilde{W}_2\right)\right]\tilde{W}_2^T$$

With control

The leaning law is

$$\dot{W}_1 = -k_1 \Delta^T P \sigma(x)$$
$$\dot{W}_2 = -k_2 \Delta^T P \phi(x) \gamma(u)$$

then

$$\dot{L} = -\Delta^T Q \Delta$$

Using Barlalat's Lemma, we have

$$\lim_{t \to \infty} \Delta = 0$$

Parallel model

The plant is

$$\dot{x} = f(x)$$

Continuous-time recurrent NN (series-parallel)

$$\frac{d}{dt}\hat{x} = A\hat{x} + W\sigma(\hat{x})$$

The nonlinear system is complete described by following neural network

$$\dot{x} = Ax + W^{*}\sigma(x)$$

The error dynamic is

$$\begin{aligned}
\dot{\Delta} &= A\Delta + W\sigma(\hat{x}) - W^{*}\sigma(x) - W^{*}\sigma(\hat{x}) + W^{*}\sigma(\hat{x}) \\
&= A\Delta + \tilde{W}\sigma(\hat{x}) + W^{*}\sigma(\hat{x}) - W^{*}\sigma(x) \\
&= A\Delta + \tilde{W}\sigma(\hat{x}) + W^{*}\left[\sigma(\hat{x}) - \sigma(x)\right]
\end{aligned}$$

Parallel model

The function $\tilde{\sigma} = \sigma(\hat{x}) - \sigma(x)$ fulfill generalized Lipshitz condition

$$\|\sigma(\hat{x}) - \sigma(x)\|_{\Lambda_\sigma} = \tilde{\sigma}^T \Lambda_\sigma \tilde{\sigma} \leq (\hat{x} - x)^T D_\sigma (\hat{x} - x) = \Delta^T D_\sigma \Delta$$

Define Lyapunov function candidate as

$$L = \Delta^T P \Delta + \frac{1}{k} tr \left[\tilde{W}^T \tilde{W} \right]$$

then

$$\dot{L} = \dot{\Delta}^T P \Delta + \Delta^T P \dot{\Delta} + \frac{2}{k} tr \left[\left(\frac{d}{dt} \tilde{W} \right) \tilde{W} \right]$$

Using error dynamic $\dot{\Delta} = A\Delta + \tilde{W}\sigma(\hat{x}) + W^* \tilde{\sigma}$,

$$\begin{gathered} \dot{L} = \Delta^T \left(A^T P + PA \right) \Delta \\ +2\Delta^T P \left[\tilde{W}^T \sigma(x) + W^* \tilde{\sigma} \right] + \tfrac{2}{k} tr \left[\left(\tfrac{d}{dt} \tilde{W} \right) \tilde{W}^T \right] \end{gathered}$$

Parallel model

We use matrix inequality

$$X^T Y + \left(X^T Y\right)^T \leq X^T \Lambda^{-1} X + Y^T \Lambda Y$$

where $X, Y, \Lambda \in \Re^{n\times k}$ and for any positive defined matrix $\Lambda = \Lambda^T > 0$, Becasue

$$2\Delta^T P W^* \tilde{\sigma} \leq \Delta^T P W^* \Lambda^{-1} W^{*T} P\Delta + \tilde{\sigma}^T \Lambda \tilde{\sigma} \leq \Delta^T \left(P\overline{W}_1 P + D_\sigma\right)\Delta$$

So

$$\dot{L} \leq \Delta^T \left(A^T P + PA\right)\Delta + \left[2\Delta^T P\sigma(x) + \frac{2}{k}\left(\frac{d}{dt}\tilde{W}\right)\right]\tilde{W}^T$$
$$+\Delta^T \left(P\overline{W}_1 P + D_\sigma\right)\Delta$$

Parallel model

with the leaning law is

$$\dot{W} = -k\Delta^T P\sigma(x)$$

Then

$$\dot{L} \leq \Delta^T \left(PA + A^T P + P\overline{W}_1 P + D_\sigma + Q_0\right) \Delta - \Delta^T Q_0 \Delta$$
$$\dot{L} \leq -\Delta^T Q_0 \Delta + \Delta^T \left(A^T P + PA + PRP + Q\right) \Delta$$

Parallel model

If A is stable, the pair $(A, R^{1/2})$ is controllable, the pair $(Q^{1/2}, A)$ is observable, and a local frequency condition

$$A^T R^{-1} A - Q \geq \frac{1}{4} \left[A^T R^{-1} - R^{-1} A \right] R \left[A^T R^{-1} - R^{-1} A \right]^T$$

is fulfilled, then the matrix Riccati equation

$$A^T P + PA + PRP + Q = 0$$

has a solution $P = P^T > 0$.
Then

$$\dot{L} \leq -\Delta^T Q \Delta \leq 0$$

Using Barlalat's Lemma, we have

$$\lim_{t \to \infty} \Delta = 0$$

Unmodeled dynamics present

The plant is

$$\dot{x} = f(x)$$

Continuous-time recurrent NN (series-parallel)

$$\frac{d}{dt}\hat{x} = A\hat{x} + W\sigma(x)$$

The nonlinear system is complete described by following neural network

$$\dot{x} = Ax + W^{*}\sigma(x) + \Delta f$$

Error dynamic

$$\dot{\Delta} = \frac{d}{dt}\hat{x} - \dot{x} = A\Delta + \tilde{W}\sigma(x) + \Delta f$$

Define Lyapunov function candidate as

$$L = \Delta^T P \Delta + \frac{1}{k} tr \left[\tilde{W}^T \tilde{W} \right]$$

Using error dynamic $\dot{\Delta} = A\Delta + \tilde{W}\sigma(x) + \Delta f$.

$$\|\Delta f\|_\Lambda^2 = \Delta f^T \Lambda \Delta f \leq \overline{f}$$

and

$$2\Delta^T P \Delta f \leq \Delta^T P \Lambda^{-1} P \Delta + \Delta f^T \Lambda \Delta f \leq \Delta^T P \Lambda^{-1} P \Delta + \bar{f}$$

then

$$\dot{L} \leq \Delta^T \left(PA + A^T P + P\Lambda^{-1} P + Q_0 \right) \Delta - \Delta^T Q_0 \Delta + \bar{f}$$

from the matrix Riccati equation

$$A^T P + PA + PRP + Q = 0$$

Unmodeled dynamics present

we have

$$\dot{L} \leq -\Delta^T Q_0 \Delta + \bar{f}$$

$$\dot{L} \leq -\Delta^T Q_0 \Delta + \bar{f} \leq -\lambda_{\min}\left(Q_0\right) \|\Delta\|^2 + \bar{f}$$

The learning law is

$$\dot{W} = -sk\Delta^T P\sigma(x)$$

where

$$s = \begin{cases} 1 & \text{if} \quad \|\Delta\|^2 > \frac{\bar{f}}{\lambda_{\min}(Q_0)} \\ 0 & \text{if} \quad \|\Delta_t\|^2 \leq \frac{\bar{f}}{\lambda_{\min}(Q_0)} \end{cases}$$

if $\|\Delta\|^2 > \frac{\bar{f}}{\lambda_{\min}(Q_0)}$, with the updating law $\dot{L} < 0$. If $\|\Delta_t\|^2 \leq \frac{\bar{f}}{\lambda_{\min}(Q_0)}$, W is constant. L is bounded.

Unmodeled dynamics present

Integrating $\dot{L}$ from 0 to T yields

$$L(T) - L(0) \leq -\int_0^T \Delta^T Q_0 \Delta dt + T\overline{f}$$

so

$$\begin{array}{c} \frac{1}{T}\int_0^T \|\Delta\|_{Q_0}^2 \, dt \leq \overline{f} + \frac{V_0}{T} \\ \lim_{T\to\infty}\int_0^T \|\Delta\|_{Q_0}^2 \, dt \leq \overline{f} \end{array} \tag{1}$$