# “Shallow” Neural Networks

## Part I: What are Neural Networks

### I. Introduction to Neural Networks

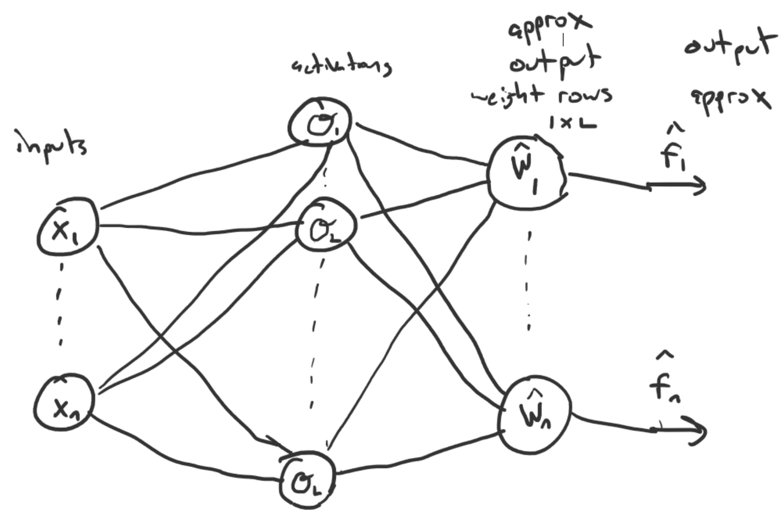
What are neural networks (NN) and more specifically what makes some networks “shallow”? At the core some we can think of them as generalized function approximators that allow us to estimate systems that have parametric uncertainty that is nonlinear in the unknown parameters or just an entirely unknown structure. Take for example these simple nonlinear relationships

$$\dot{x} = \sin(bx) + u$$
$$\dot{x} = a\sin(bx + c) + u$$

where $a, b, c \in \mathbb{R}$ are unknown constants and $u \in \mathbb{R}$ is the input. The first one we could still do some tricks to get a linear in the unknown relationship

$$\sin^{-1}(\dot{x} - u) = bx$$

and we could use this to find $b$. However, if we have an unknown amplitude and phase shift we can’t form this type of relationship since we can’t get the parameters into a linear in the parameters form. So how do we approach these types of systems? Well, we will look at various adaptive control methods that use NN starting with the most basic, single layer NN.




The general idea of using a NN stems from results like *M. H. Stone, "The Generalized Weierstrass Approximation Theorem", Math. Mag. 21, 4 (1948), pp. 167-184.* which shows this idea is not that new. You might ask well why is it still so popular today and the main reason is recent advances have enabled the use of much more complex nonlinear systems and more complex inputs such as images or speech (which is beyond the scope of this course) but this is remains a very important and popular area of research.

Now, lets look at the simplest general form of a NN approximate of the dynamics for our state $x(t) \in \mathbb{R}^n$, $f(x) \in \mathbb{R}^n$

$$f(x) = W^\top \sigma(x) + \varepsilon(x)$$

where $W \in \mathbb{R}^{L\times n}$ is the output weights, $\sigma(x) \in \mathbb{R}^L$ is the output function, and $\varepsilon(x) \in \mathbb{R}^n$ is the residual error of the approximation. In the single layer case, we usually choose a large $L$ and choose the basis directly. We don’t know $\varepsilon(x)$ or $W$ though so our estimate of $f(x)$, call it $\widehat{f}(x)$ becomes

$$\widehat{f}(x) \triangleq \widehat{W}^\top \sigma(x)$$

where $\widehat{W}(t) \in \mathbb{R}^{L\times n}$. Another way to view this is as above pictorially or as a summation for each state element

$$\widehat{f}_i(x) = \sum_{j=1}^{L} \widehat{W}_{ji}\sigma_j(x).$$

## II. Estimator and Controller Design

Now the big question becomes how do we choose $\sigma(x)$? Well there are several approaches popular activation functions to try for example check out pytorchs https://pytorch.org/docs/stable/nn.html or more specific to nonlinear activations https://pytorch.org/docs/stable/nn.html#non-linear-activations-weighted-sum-nonlinearity. You could try various approaches though for example a polynomial (based on idea behind Taylor series expansion) or you could try some bounded basis functions like Gaussian functions (https://en.wikipedia.org/wiki/Gaussian_function) or any of the various trigonometric functions.

For

$$\dot{x} = a\sin(bx + c) + u$$

lets say we knew the form so we just guessed several basis based on some information about it based on knowledge about the constants. For example, assume we knew the bounds on $a, b, c \in \mathbb{R}$ were $\underline{a} \leq a \leq \overline{a}$, $\underline{b} \leq b \leq \overline{b}$, and $\underline{c} \leq c \leq \overline{c}$. Then we could choose

$$\sigma_j(x) = \sin\left(\widehat{b}_j x + \widehat{c}_j\right), \widehat{b}_j \sim \mathcal{U}_{[\underline{b},\overline{b}]}, \widehat{c}_j \sim \mathcal{U}_{[\underline{c},\overline{c}]}, \forall j \in \{1, \ldots, L\}$$

or we could generate random numbers from a normal distribution. A random basis is often a good choice if you are choosing a custom basis for a single layer NN regardless of the activation function. These last basis could also be chosen as 1 to account for a bias. No lets use the function approximator to rewrite the dynamics

$$\dot{x} = W^\top \sigma(x) + \varepsilon(x) + u$$

Now lets examine a tracking scenario where we want to track $x_d, \dot{x}_d$

$$\begin{aligned}
e &= x - x_d \\
\dot{e} &= \dot{x} - \dot{x}_d \\
\dot{e} &= W^\top \sigma(x) + \varepsilon(x) + u - \dot{x}_d
\end{aligned}$$

Similar to when we estimate the parametric uncertainty, we will estimate the output weights as

$$\begin{aligned}
\widetilde{W} &= W - \widehat{W} \\
\dot{\widetilde{W}} &= -\dot{\widehat{W}}
\end{aligned}$$

Now in this case $W \in \mathbb{R}^L$ so we can treat it like we did $\theta$ before let $\zeta = \begin{bmatrix} e \\ \widetilde{W} \end{bmatrix}$

$$\begin{aligned}
V &= \frac{1}{2}e^2 + \frac{1}{2}\widetilde{W}^\top \Gamma^{-1} \widetilde{W} \\
\dot{V} &= e\dot{e} + \widetilde{W}^\top \Gamma^{-1} \dot{\widetilde{W}}
\end{aligned}$$

and using the above we get

$$\dot{V} = e\left(W^\top \sigma(x) + \varepsilon(x) + u - \dot{x}_d\right) - \widetilde{W}^\top \Gamma^{-1} \dot{\widehat{W}}$$

so lets design the input as

$$u = \dot{x}_d - \widehat{W}^\top \sigma(x) - \beta_e e - \beta_\varepsilon \operatorname{sgn}(e)$$

implying

$$\begin{aligned}
\dot{V} &= e\left(W^\top \sigma(x) + \varepsilon(x) + \dot{x}_d - \widehat{W}^\top \sigma(x) - \beta_e e - \beta_\varepsilon \operatorname{sgn}(e) - \dot{x}_d\right) - \widetilde{W}^\top \Gamma^{-1} \dot{\widehat{W}} \\
\dot{V} &= e\left(-\beta_e e + \widetilde{W}^\top \sigma(x) + \varepsilon(x) - \beta_\varepsilon \operatorname{sgn}(e)\right) - \widetilde{W}^\top \Gamma^{-1} \dot{\widehat{W}} \\
\dot{V} &= -\beta_e e^2 + \widetilde{W}^\top \sigma(x) e + \varepsilon(x) e - \beta_\varepsilon \operatorname{sgn}(e) e - \widetilde{W}^\top \Gamma^{-1} \dot{\widehat{W}}
\end{aligned}$$

Design the update law using ICL

$$
\begin{aligned}
\dot{x} - u &= \sigma^{\top}(x)\, W + \varepsilon(x) \\
\underbrace{\dot{x}(t) - u(t)}_{\mathcal{U}_{CL}(t)} &= \underbrace{\sigma^{\top}(x(t))}_{\mathcal{Y}_{CL}(t)} W + \underbrace{\varepsilon(x(t))}_{\mathcal{E}_{CL}(t)} \\
\underbrace{\int_{t-\Delta t}^{t} \mathcal{U}_{CL}(\iota)\, d\iota}_{\mathcal{U}(t)} &= \underbrace{\int_{t-\Delta t}^{t} \mathcal{Y}_{CL}(\iota)\, d\iota}_{\mathcal{Y}(t)} W + \underbrace{\int_{t-\Delta t}^{t} \mathcal{E}_{CL}(\iota)\, d\iota}_{\mathcal{E}(t)} \\
\mathcal{Y}^{\top}(t)\, \mathcal{U}(t) &= \mathcal{Y}^{\top}(t)\, \mathcal{Y}(t)\, W + \mathcal{Y}^{\top}(t)\, \mathcal{E}(t) \\
\sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{U}(t_j) &= \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{Y}(t_j)\, W + \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{E}(t_j)
\end{aligned}
$$

Now we can design the update as

$$
\dot{\widehat{W}} = \Gamma \sigma(x)\, e + \Gamma k_{CL} \left( \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{U}(t_j) - \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{Y}(t_j)\, \widehat{W}(t) \right)
$$

which for analysis we can write as

$$
\begin{aligned}
\dot{\widehat{W}} &= \Gamma \sigma(x)\, e + \Gamma k_{CL} \left( \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{Y}(t_j)\, W + \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{E}(t_j) - \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{Y}(t_j)\, \widehat{W}(t) \right) \\
\dot{\widehat{W}} &= \Gamma \sigma(x)\, e + \Gamma k_{CL} \left( \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{Y}(t_j)\, \widetilde{W}(t) + \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{E}(t_j) \right)
\end{aligned}
$$

and sub back in

$$
\begin{aligned}
\dot{V} &= -\beta_e e^2 + \widetilde{W}^{\top} \sigma(x)\, e + \varepsilon(x)\, e - \beta_{\varepsilon} \operatorname{sgn}(e)\, e - \widetilde{W}^{\top} \Gamma^{-1} \left( \Gamma \sigma(x)\, e + \Gamma k_{CL} \left( \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{Y}(t_j)\, \widetilde{W} + \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{E}(t_j) \right) \right) \\
\dot{V} &= -\beta_e e^2 + \widetilde{W}^{\top} \sigma(x)\, e + \varepsilon(x)\, e - \beta_{\varepsilon} \operatorname{sgn}(e)\, e - \widetilde{W}^{\top} \sigma(x)\, e - \widetilde{W}^{\top} k_{CL} \left( \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{Y}(t_j)\, \widetilde{W} + \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{E}(t_j) \right) \\
\dot{V} &= -\beta_e e^2 + \varepsilon(x)\, e - \beta_{\varepsilon} \operatorname{sgn}(e)\, e - \widetilde{W}^{\top} k_{CL} \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{Y}(t_j)\, \widetilde{W} + \widetilde{W}^{\top} k_{CL} \sum_{j=1}^{N} \mathcal{Y}^{\top}(t_j)\, \mathcal{E}(t_j) \\
\dot{V} &\leq -\beta_e e^2 + \overline{\varepsilon} |e| - \beta_{\varepsilon} |e| - \underline{k_{CL}} \lambda_{CL} \|\widetilde{W}\|^2 + N \overline{k_{CL} \mathcal{E} \mathcal{Y}} \|\widetilde{W}\|, \forall t > T_{CL} \\
\dot{V} &\leq -\beta_e e^2 + \overline{\varepsilon} |e| - \beta_{\varepsilon} |e| - \frac{1}{2} \underline{k_{CL}} \lambda_{CL} \|\widetilde{W}\|^2 - \frac{1}{2} \underline{k_{CL}} \lambda_{CL} \|\widetilde{W}\|^2 + N \overline{k_{CL} \mathcal{E} \mathcal{Y}} \|\widetilde{W}\|, \forall t > T_{CL}
\end{aligned}
$$

choose $\beta_{\varepsilon} > \overline{\varepsilon}$ and upper bound using nonlinear damping

$$
-\frac{1}{2} \underline{k_{CL}} \lambda_{CL} \|\widetilde{W}\|^2 + N \overline{k_{CL} \mathcal{E} \mathcal{Y}} \|\widetilde{W}\| \leq \frac{\left(N \overline{k_{CL} \mathcal{E} \mathcal{Y}}\right)^2}{4 \frac{1}{2} \underline{k_{CL}} \lambda_{CL}} \leq \frac{\left(N \overline{k_{CL} \mathcal{E} \mathcal{Y}}\right)^2}{2 \underline{k_{CL}} \lambda_{CL}}
$$

Using this yields

$$
\begin{aligned}
\dot{V} &\leq -\beta_e e^2 - \frac{1}{2} \underline{k_{CL}} \lambda_{CL} \|\widetilde{W}\|^2 + \frac{\left(N \overline{k_{CL} \mathcal{E} \mathcal{Y}}\right)^2}{2 \underline{k_{CL}} \lambda_{CL}}, \forall t > T_{CL} \\
\dot{V} &\leq -\min\left\{ \beta_e, \frac{1}{2} \underline{k_{CL}} \lambda_{CL} \right\} \|\zeta\|^2 + \frac{\left(N \overline{k_{CL} \mathcal{E} \mathcal{Y}}\right)^2}{2 \underline{k_{CL}} \lambda_{CL}}, \forall t > T_{CL}
\end{aligned}
$$

and using the bounds on $V$

$$
\begin{aligned}
V &\leq \frac{1}{2} \max\left\{ 1, \lambda_{\max}\left\{ \Gamma^{-1} \right\} \right\} \|\zeta\|^2 \\
-\frac{2}{\max\left\{ 1, \lambda_{\max}\left\{ \Gamma^{-1} \right\} \right\}} V &\geq -\|\zeta\|^2
\end{aligned}
$$

implying

$$\dot{V} \leq -\underbrace{\frac{2\min\left\{\beta_e, \frac{1}{2}\underline{k_{CL}}\lambda_{CL}\right\}}{\max\left\{1, \lambda_{\max}\left\{\Gamma^{-1}\right\}\right\}}}_{\beta_v} V + \underbrace{\frac{\left(N\overline{k_{CL}}\mathcal{E}\mathcal{Y}\right)^2}{2\underline{k_{CL}}\lambda_{CL}}}_{\delta_v}, \forall t > T_{CL}$$

$$\dot{V} \leq -\beta_v V + \delta_v, \forall t > T_{CL}$$

$$V\left(\zeta\left(t\right)\right) \leq V\left(\zeta\left(T_{CL}\right)\right)\exp\left(-\beta_v\left(t - T_{CL}\right)\right) + \frac{\delta_v}{\beta_v}\left(1 - \exp\left(-\beta_v\left(t - T_{CL}\right)\right)\right)$$

and since

$$V\left(\zeta\left(T_{CL}\right)\right) \leq V\left(\zeta\left(0\right)\right), \forall t \leq T_{CL}$$

$$V\left(\zeta\left(t\right)\right) \leq V\left(\zeta\left(0\right)\right)\exp\left(-\beta_v\left(t - T_{CL}\right)\right) + \frac{\delta_v}{\beta_v}\left(1 - \exp\left(-\beta_v\left(t - T_{CL}\right)\right)\right)$$

## III. Two-Link System

Consider the two-link dynamics with an unknown force acting on the system

$$M\left(\phi\right)\ddot{\phi} + C\left(\phi, \dot{\phi}\right) + G\left(\phi\right) + \tau_d\left(\phi, \dot{\phi}, \ddot{\phi}\right) = \tau$$

$$M\left(\phi\right) \triangleq \begin{bmatrix} m_1 l_1^2 + m_2\left(l_1^2 + 2l_1l_2c_2 + l_2^2\right) & m_2\left(l_1l_2c_2 + l_2^2\right) \\ m_2\left(l_1l_2c_2 + l_2^2\right) & m_2l_2^2 \end{bmatrix},$$

$$C\left(\phi, \dot{\phi}\right) \triangleq \begin{bmatrix} -2m_2l_1l_2s_2\dot{\phi}_1\dot{\phi}_2 \\ 0 \end{bmatrix} + \begin{bmatrix} -m_2l_1l_2s_2\dot{\phi}_2^2 \\ m_2l_1l_2s_2\dot{\phi}_1^2 \end{bmatrix},$$

$$G\left(\phi\right) \triangleq \begin{bmatrix} \left(m_1 + m_2\right)gl_1c_1 + m_2gl_2c_{12} \\ m_2gl_2c_{12} \end{bmatrix},$$

where $\tau_d\left(\phi, \dot{\phi}, \ddot{\phi}\right) \in \mathbb{R}^2$ is a function of the state and could be a force of an object the system picked up, unknown friction at the joints, air resistance, any combination of these, or various other types of unstructured dynamics. There are a few ways to approach this problem but we will leverage the known structure for the inertia, centripetal Coriolis, and gravity matrices but use various basis to approximate $\tau_d\left(\phi, \dot{\phi}, \ddot{\phi}\right)$

$$\tau_d\left(\phi, \dot{\phi}, \ddot{\phi}\right) = W^\top\sigma\left(\phi, \dot{\phi}, \ddot{\phi}\right) + \varepsilon\left(\phi, \dot{\phi}, \ddot{\phi}\right)$$

where $W \in \mathbb{R}^{L\times 2}$, $\sigma\left(\phi, \dot{\phi}, \ddot{\phi}\right) \in \mathbb{R}^L$, and $\varepsilon\left(\phi, \dot{\phi}, \ddot{\phi}\right)$.

### A. Error Systems

Now we can develop the error systems

$$\begin{aligned} e &= \phi_d - \phi \\ \dot{e} &= \dot{\phi}_d - \dot{\phi} \\ \ddot{e} &= \ddot{\phi}_d - \ddot{\phi} \\ r &= \dot{e} + \alpha e \\ \dot{r} &= \ddot{e} + \alpha\dot{e} \\ \dot{r} &= \ddot{\phi}_d - \ddot{\phi} + \alpha\dot{e} \\ M\left(\phi\right)\dot{r} &= M\left(\phi\right)\ddot{\phi}_d - M\left(\phi\right)\ddot{\phi} + M\left(\phi\right)\alpha\dot{e} \\ M\left(\phi\right)\dot{r} &= M\left(\phi\right)\left(\ddot{\phi}_d + \alpha\dot{e}\right) - M\left(\phi\right)\ddot{\phi} \end{aligned}$$

Using the dynamics we know

$$M\left(\phi\right)\ddot{\phi} = -C\left(\phi, \dot{\phi}\right) - G\left(\phi\right) - \tau_d\left(\phi, \dot{\phi}, \ddot{\phi}\right) + \tau$$

implying

$$M\left(\phi\right)\dot{r} = M\left(\phi\right)\left(\ddot{\phi}_d + \alpha\dot{e}\right) - \left(-C\left(\phi, \dot{\phi}\right) - G\left(\phi\right) - \tau_d\left(\phi, \dot{\phi}, \ddot{\phi}\right) + \tau\right)$$

$$M\left(\phi\right)\dot{r} = M\left(\phi\right)\left(\ddot{\phi}_d + \alpha\dot{e}\right) + C\left(\phi, \dot{\phi}\right) + G\left(\phi\right) - \tau_d\left(\phi, \dot{\phi}, \ddot{\phi}\right) - \tau$$

And recall from the analysis we get an extra term from the inertia derivative so lets add and subtract $\frac{1}{2}\dot{M}\left(\phi,\dot{\phi}\right)r$ now

$$M\left(\phi\right)\dot{r}=M\left(\phi\right)\left(\ddot{\phi}_d+\alpha\dot{e}\right)+C\left(\phi,\dot{\phi}\right)+G\left(\phi\right)+\tau_d\left(\phi,\dot{\phi},\ddot{\phi}\right)-\tau\pm\frac{1}{2}\dot{M}\left(\phi,\dot{\phi}\right)r$$

We can write the

$$\dot{M}\left(\phi,\dot{\phi}\right)r=Y_{\dot{M}}\left(\phi,\dot{\phi},r\right)\theta$$

where

$$Y_{\dot{M}}\left(\phi,\dot{\phi},r\right)\triangleq\begin{bmatrix}0 & -\left(2s_2\dot{\phi}_2r_1+s_2\dot{\phi}_2r_2\right) & 0 & 0 & 0\\ 0 & -s_2\dot{\phi}_2r_1 & 0 & 0 & 0\end{bmatrix}.$$

Then we can write the above as

$$M\left(\phi\right)\left(\ddot{\phi}_d+\alpha\dot{e}\right)+C\left(\phi,\dot{\phi}\right)+G\left(\phi\right)+\frac{1}{2}\dot{M}\left(\phi,\dot{\phi}\right)r=Y\theta$$

with $\varphi=\ddot{\phi}_d+\alpha\dot{e}$

$$Y\left(\phi,\dot{\phi},\ddot{\phi}_d,\dot{e},r\right)=\left(Y_M\left(\phi,\varphi\right)+Y_C\left(\phi,\dot{\phi}\right)+Y_G\left(\phi\right)+Y_{\dot{M}}\left(\phi,\dot{\phi},r\right)\right)$$

$$Y_M\left(\phi,\varphi\right)\triangleq\begin{bmatrix}\varphi_1 & \left(2c_2\varphi_1+c_2\varphi_2\right) & \varphi_2 & 0 & 0\\ 0 & c_2\varphi_1 & \left(\varphi_1+\varphi_2\right) & 0 & 0\end{bmatrix},$$

$$Y_C\left(\phi,\dot{\phi}\right)\triangleq\begin{bmatrix}0 & -\left(2s_2\dot{\phi}_1\dot{\phi}_2+s_2\dot{\phi}_2^2\right) & 0 & 0 & 0\\ 0 & s_2\dot{\phi}_1^2 & 0 & 0 & 0\end{bmatrix},$$

$$Y_G\left(\phi\right)\triangleq\begin{bmatrix}0 & 0 & 0 & gc_1 & gc_{12}\\ 0 & 0 & 0 & 0 & gc_{12}\end{bmatrix},$$

$$Y_{\dot{M}}\left(\phi\right)\triangleq\begin{bmatrix}0 & -\left(2s_2\dot{\phi}_2r_1+s_2\dot{\phi}_2r_2\right) & 0 & 0 & 0\\ 0 & -s_2\dot{\phi}_2r_1 & 0 & 0 & 0\end{bmatrix}$$

$$\begin{bmatrix}\theta_1\\ \theta_2\\ \theta_3\\ \theta_4\\ \theta_5\end{bmatrix}=\begin{bmatrix}m_1l_1^2+m_2l_1^2+m_2l_2^2\\ m_2l_1l_2\\ m_2l_2^2\\ \left(m_1+m_2\right)l_1\\ m_2l_2\end{bmatrix}$$

and we can use the NN

$$M\left(\phi\right)\dot{r}=Y\left(\phi,\dot{\phi},\ddot{\phi}_d,\dot{e},r\right)\theta+W^\top\sigma\left(\phi,\dot{\phi},\ddot{\phi}\right)+\varepsilon\left(\phi,\dot{\phi},\ddot{\phi}\right)-\tau-\frac{1}{2}\dot{M}\left(\phi,\dot{\phi}\right)r$$

which we will write as

$$M\dot{r}=Y\theta+W^\top\sigma+\varepsilon-\tau-\frac{1}{2}\dot{M}r$$

### *B. Analysis*

Let

$$\zeta\triangleq\begin{bmatrix}e\\ r\\ \widetilde{\theta}\\ \text{vec}\left(\widetilde{W}\right)\end{bmatrix}$$

$$\text{vec}\left(\widetilde{W}\right)=\begin{bmatrix}\widetilde{W}_{11}\\ \vdots\\ \widetilde{W}_{L1}\\ \widetilde{W}_{12}\\ \vdots\\ \widetilde{W}_{L2}\end{bmatrix}$$

and

$$V\left(\zeta,t\right)\triangleq\frac{1}{2}e^\top e+\frac{1}{2}r^\top M\left(\phi\right)r+\frac{1}{2}\widetilde{\theta}^\top\Gamma_\theta^{-1}\widetilde{\theta}+\frac{1}{2}\text{tr}\left(\widetilde{W}^\top\Gamma_W^{-1}\widetilde{W}\right)$$

taking the time derivative

$$\dot{V}(\zeta,t) = e^\top \dot{e} + \frac{1}{2} r^\top \dot{M}(\phi) r + r^\top M(\phi)\dot{r} + \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widetilde{\theta}} + \text{tr}\left(\widetilde{W}^\top \Gamma_W^{-1} \dot{\widetilde{W}}\right)$$

then we can use

$$\begin{aligned} r &= \dot{e} + \alpha e \\ \implies \dot{e} &= r - \alpha e \end{aligned}$$

and

$$M\dot{r} = Y\theta + W^\top \sigma + \varepsilon - \tau - \frac{1}{2}\dot{M} r$$

$$\begin{aligned} \dot{V}(\zeta,t) &= e^\top (r - \alpha e) + \frac{1}{2} r^\top \dot{M}(\phi) r + r^\top \left(Y\theta + W^\top \sigma + \varepsilon - \tau - \frac{1}{2}\dot{M} r\right) - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left(-\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}}\right) \\ \dot{V}(\zeta,t) &= -e^\top \alpha e + r^\top e + r^\top \left(Y\theta + W^\top \sigma + \varepsilon - \tau\right) - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left(-\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}}\right) \end{aligned}$$

Now we can design the input

$$\tau = Y\widehat{\theta} + \widehat{W}^\top \sigma + e + \beta_r r + \beta_\varepsilon \text{sgn}(r)$$

which yields

$$\begin{aligned} \dot{V}(\zeta,t) &= -e^\top \alpha e + r^\top e + r^\top \left(Y\theta + W^\top \sigma + \varepsilon - \left(Y\widehat{\theta} + \widehat{W}^\top \sigma + e + \beta_r r + \beta_\varepsilon \text{sgn}(r)\right)\right) - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left(-\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}}\right) \\ \dot{V}(\zeta,t) &= -e^\top \alpha e + r^\top \left(Y\widetilde{\theta} + \widetilde{W}^\top \sigma + \varepsilon - \beta_r r - \beta_\varepsilon \text{sgn}(r)\right) - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left(-\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}}\right) \\ \dot{V}(\zeta,t) &= -e^\top \alpha e - r^\top \beta_r r + r^\top Y\widetilde{\theta} + r^\top \widetilde{W}^\top \sigma + r^\top \varepsilon - r^\top \beta_\varepsilon \text{sgn}(r) - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left(-\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}}\right) \end{aligned}$$

*1) Gradient-Based Approximation:* First we will examine gradient-based estimators for this system. Similar to before we can design $\dot{\widehat{\theta}}$

$$\dot{\widehat{\theta}} = \Gamma_\theta Y^\top r$$

but how about the trace? We can use a trick for this which says for two vectors $a, b \in \mathbb{R}^n$

$$\text{tr}\left(ba^\top\right) = a^\top b$$

but how does this help us?

$$\underbrace{r^\top}_{a^\top} \underbrace{\widetilde{W}^\top \sigma}_{b} = \text{tr}\left(\underbrace{\widetilde{W}^\top \sigma}_{b} \underbrace{r^\top}_{a^\top}\right)$$

so we can move this into the trace term $\text{tr}\left(-\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}}\right)$ and get

$$\text{tr}\left(-\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}} + \widetilde{W}^\top \sigma r^\top\right)$$

and we can design $\dot{\widehat{W}}$

$$\dot{\widehat{W}} = \Gamma_W \sigma r^\top$$

With these designs we get

$$\begin{aligned} \dot{V}(\zeta,t) &= -e^\top \alpha e - r^\top \beta_r r + \widetilde{\theta}^\top Y^\top r + r^\top \varepsilon - r^\top \beta_\varepsilon \text{sgn}(r) - \widetilde{\theta}^\top \Gamma_\theta^{-1} \Gamma_\theta Y^\top r + \text{tr}\left(-\widetilde{W}^\top \Gamma_W^{-1} \Gamma_W \sigma r^\top + \widetilde{W}^\top \sigma r^\top\right) \\ \dot{V}(\zeta,t) &= -e^\top \alpha e - r^\top \beta_r r + r^\top \varepsilon - r^\top \beta_\varepsilon \text{sgn}(r) \end{aligned}$$

and we select $\beta_\varepsilon > \overline{\varepsilon}$ where $\left\|\varepsilon\left(\phi(t), \dot{\phi}(t), \ddot{\phi}(t)\right)\right\| \leq \overline{\varepsilon}$ then

$$\begin{aligned} \dot{V}(\zeta,t) &\leq -\underline{\alpha}\|e\|^2 - \underline{\beta}_r \|r\|^2 + \|\varepsilon\|\|r\| - \underline{\beta}_\varepsilon \|r\| \\ \dot{V}(\zeta,t) &\leq -\underline{\alpha}\|e\|^2 - \underline{\beta}_r \|r\|^2 \end{aligned}$$

and we can show asymptotic tracking.

# PART II: TWO LAYER NETWORKS

## IV. INTRODUCTION TO TWO LAYER NETWORKS

Before we looked at several methods of selecting $\sigma(\cdot)$ to approximate a function but now we will look at a way of doing it using a two layer NN with the general structure

$$f(x) = W^\top \sigma(\Phi(\xi)) + \varepsilon(x)$$

where $x \in \mathbb{R}^n$ is the state, $\xi = \begin{bmatrix} x \\ 1 \end{bmatrix} \in \mathbb{R}^{n+1}$, $W \in \mathbb{R}^{L+1\times n}$ are the output weights, $\sigma(\cdot) \in \mathbb{R}^{L+1}$ are the activation functions and bias, $\Phi(\xi) \in \mathbb{R}^L$ is the inner weights and typically are

$$\Phi(\xi) = V^\top \xi,$$

$V^\top \in \mathbb{R}^{L\times n+1}$, and $\varepsilon(x) \in \mathbb{R}^n$ is the approximation error. Now before we would select several basis randomly but now we generally select $\sigma(\cdot)$ as activation functions and a bias. This form is more similar to the original *Stone-Weierstrass* Theorem which is captured in the following

**Theorem 1.** *Stone-Weierstrass*
*Suppose $f$ is a continuous real function defined on a compact set $[a,b]$, for every $\bar{\varepsilon} > 0$ $\exists$ a polynomial $p(x)$ such that $\forall x \in [a,b]$ we have $\|f(x) - p(x)\| \leq \bar{\varepsilon}$.*

**Property 1.** Universal Function Approximation Property
Let $f(x)$ be a smooth function from $\mathbb{R}^n$ then $\exists$ weights and thresholds such that

$$f(x) = W^\top \sigma(\Phi(\xi)) + \varepsilon(x)$$

for some number $L$ of hidden layer neurons where $\varepsilon(x) < \bar{\varepsilon}$ is the function approximation error where $\bar{\varepsilon}$ decreases as $L \to \infty$.

*Remark* 1. Now we have seen that as we increase the number of random basis it doesn't necessarily guarantee we get closer but this goes into a much broader question of what basis improve approximation. For now we will just take this as it is and the main takeaway is what we have been showing, we get better or worse approximation depending on the the basis choice. Often you will also see that the bias term is not needed or ignored but this will often be application specific.

Now you might say aren't we just back where we started since this is a nonlinear function of the unknown parameters? The answer is yes but we will take this a step further and this approach also works for a more broad class of systems even if we don't know the basis. Let the applied approximate of $f(x)$ be

$$\widehat{f}(x) = \widehat{W}^\top \sigma\left(\widehat{\Phi}(\xi)\right)$$

where $\widehat{\Phi}(\xi) = \widehat{V}^\top \xi$. That implies that

$$f(x) - \widehat{f}(x) = W^\top \sigma(\Phi(\xi)) + \varepsilon(x) - \widehat{W}^\top \sigma\left(\widehat{\Phi}(\xi)\right).$$

Then recall from a Taylor series approximations

$$\sigma(\Phi(\xi)) = \sigma\left(\widehat{\Phi}(\xi)\right) + \frac{\partial \sigma}{\partial \Phi}\,|_{\widehat{\Phi}} \left(\Phi(\xi) - \widehat{\Phi}(\xi)\right) + \varepsilon_\sigma\left(\widetilde{\Phi}(\xi)^2\right)$$
$$\sigma(\Phi(\xi)) = \sigma\left(\widehat{\Phi}(\xi)\right) + \frac{\partial \sigma}{\partial \Phi}\,|_{\widehat{\Phi}} \left(V^\top \xi - \widehat{V}\xi\right) + \varepsilon_\sigma\left(\widetilde{\Phi}(\xi)^2\right)$$
$$\sigma(\Phi(\xi)) = \sigma\left(\widehat{\Phi}(\xi)\right) + \frac{\partial \sigma}{\partial \Phi}\,|_{\widehat{\Phi}}\, \widetilde{V}^\top \xi + \varepsilon_\sigma\left(\widetilde{\Phi}(\xi)^2\right)$$

Using this above we get

$$f(x) - \widehat{f}(x) = W^\top \left(\sigma\left(\widehat{\Phi}(\xi)\right) + \frac{\partial \sigma}{\partial \Phi}\,|_{\widehat{\Phi}}\, \widetilde{V}^\top \xi + \varepsilon_\sigma\left(\widetilde{\Phi}(\xi)^2\right)\right) + \varepsilon(x) - \widehat{W}^\top \sigma\left(\widehat{\Phi}(\xi)\right)$$
$$f(x) - \widehat{f}(x) = W^\top \sigma\left(\widehat{\Phi}(\xi)\right) + W^\top \frac{\partial \sigma}{\partial \Phi}\,|_{\widehat{\Phi}}\, \widetilde{V}^\top \xi + W^\top \varepsilon_\sigma\left(\widetilde{\Phi}(\xi)^2\right) + \varepsilon(x) - \widehat{W}^\top \sigma\left(\widehat{\Phi}(\xi)\right)$$
$$f(x) - \widehat{f}(x) = \widetilde{W}^\top \sigma\left(\widehat{\Phi}(\xi)\right) + W^\top \frac{\partial \sigma}{\partial \Phi}\,|_{\widehat{\Phi}}\, \widetilde{V}^\top \xi + W^\top \varepsilon_\sigma\left(\widetilde{\Phi}(\xi)^2\right) + \varepsilon(x)$$

and now we add and subtract $\widehat{W}^\top \frac{\partial \sigma}{\partial \Phi} \widetilde{V}^\top \xi$ which yields

$$f(x) - \widehat{f}(x) = \widetilde{W}^\top \sigma\left(\widehat{\Phi}(\xi)\right) + W^\top \frac{\partial \sigma}{\partial \Phi}\,|_{\widehat{\Phi}}\, \widetilde{V}^\top \xi + W^\top \varepsilon_\sigma\left(\widetilde{\Phi}(\xi)^2\right) + \varepsilon(x) \pm \widehat{W}^\top \frac{\partial \sigma}{\partial \Phi} \widetilde{V}^\top \xi$$
$$f(x) - \widehat{f}(x) = \widetilde{W}^\top \sigma\left(\widehat{\Phi}(\xi)\right) + \widehat{W}^\top \frac{\partial \sigma}{\partial \Phi}\,|_{\widehat{\Phi}}\, \widetilde{V}^\top \xi + \widetilde{W}^\top \frac{\partial \sigma}{\partial \Phi} \widetilde{V}^\top \xi + W^\top \varepsilon_\sigma\left(\widetilde{\Phi}(\xi)^2\right) + \varepsilon(x)$$

## V. First Order System

Now lets think back to the first order dynamics

$$\dot{x} = a\sin\left(bx + c\right) + u.$$

Before we selected the basis as several sine waves but this time we can really just a single weight so lets look at the next form of $W^\top \sigma\left(\Phi\left(\xi\right)\right)$

$$W^\top = \begin{bmatrix} w_x & w_b \end{bmatrix},$$
$$\Phi\left(\xi\right) = V^\top \xi,$$
$$V^\top = \begin{bmatrix} v_x & v_b \end{bmatrix},$$
$$\xi = \begin{bmatrix} x \\ 1 \end{bmatrix}$$

and

$$\sigma\left(\cdot\right) = \begin{bmatrix} \sin\left(\cdot\right) \\ 1 \end{bmatrix},$$

where $w_x \in \mathbb{R}$ is the output weight associated with the state, $w_b \in \mathbb{R}$ is the output weight associated with the bias, $v_x \in \mathbb{R}$ is the inner weight associated with the state, $v_b \in \mathbb{R}$ is the inner weight bias. So the total approximation is then

$$f\left(x\right) = W^\top \sigma\left(\Phi\left(\xi\right)\right) + \varepsilon\left(x\right)$$

Now as before we must do the approximate of the basis as

$$\sigma\left(\Phi\left(\xi\right)\right) = \widehat{\sigma} + \widehat{\sigma}' \widetilde{V}^\top \xi + \varepsilon_\sigma\left(\widetilde{\Phi}\left(\xi\right)^2\right)$$
$$\widehat{\sigma} = \sigma\left(\widehat{\Phi}\left(\xi\right)\right)$$
$$\widehat{\sigma}' = \frac{\partial \sigma}{\partial \Phi} \mid_{\widehat{\Phi}}$$

Now since

$$\sigma\left(\Phi\left(\xi\right)\right) = \begin{bmatrix} \sin\left(V^\top \xi\right) \\ 1 \end{bmatrix}$$
$$\sigma\left(\Phi\left(\xi\right)\right) = \begin{bmatrix} \sin\left(\begin{bmatrix} v_x & v_b \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}\right) \\ 1 \end{bmatrix}$$
$$\sigma\left(\Phi\left(\xi\right)\right) = \begin{bmatrix} \sin\left(v_x x + v_b\right) \\ 1 \end{bmatrix}.$$

Now calculate $\sigma\left(\widehat{\Phi}\left(\xi\right)\right)$

$$\sigma\left(\widehat{\Phi}\left(\xi\right)\right) = \begin{bmatrix} \sin\left(\widehat{V}^\top \xi\right) \\ 1 \end{bmatrix}$$
$$\sigma\left(\widehat{\Phi}\left(\xi\right)\right) = \begin{bmatrix} \sin\left(\widehat{v}_x x + \widehat{v}_b\right) \\ 1 \end{bmatrix}$$

Now calculate $\frac{\partial \sigma}{\partial \Phi} \mid_{\widehat{\Phi}}$ and since $\Phi$ is just a scalar

$$\frac{\partial \sigma}{\partial \Phi} \mid_{\widehat{\Phi}} = \begin{bmatrix} \cos\left(\widehat{\Phi}\right) \\ 0 \end{bmatrix}$$
$$\frac{\partial \sigma}{\partial \Phi} \mid_{\widehat{\Phi}} = \begin{bmatrix} \cos\left(\widehat{v}_x x + \widehat{v}_b\right) \\ 0 \end{bmatrix}$$

Now using the approximation, we get the dynamics are

$$\dot{x} = W^\top \sigma\left(\Phi\left(\xi\right)\right) + \varepsilon\left(x\right) + u$$
$$\dot{x} = W^\top \left(\widehat{\sigma} + \widehat{\sigma}' \widetilde{V}^\top \xi + \varepsilon_\sigma\left(\widetilde{\Phi}\left(\xi\right)^2\right)\right) + \varepsilon\left(x\right) + u$$
$$\dot{x} = W^\top \widehat{\sigma} + W^\top \widehat{\sigma}' \widetilde{V}^\top \xi + W^\top \varepsilon_\sigma\left(\widetilde{\Phi}\left(\xi\right)^2\right) + \varepsilon\left(x\right) + u$$

Now add and subtract $\widehat{W}^\top \frac{\partial \sigma}{\partial \Phi} \widetilde{V}^\top \xi$

$$\dot{x} = W^\top \widehat{\sigma} + W^\top \widehat{\sigma}' \widetilde{V}^\top \xi + W^\top \varepsilon_\sigma \left( \widetilde{\Phi} \left( \xi \right)^2 \right) + \varepsilon \left( x \right) + u \pm \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi$$
$$\dot{x} = W^\top \widehat{\sigma} + \underbrace{\widetilde{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + W^\top \varepsilon_\sigma \left( \widetilde{\Phi} \left( \xi \right)^2 \right) + \varepsilon \left( x \right)}_{\delta} + u + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi$$

With $\delta \triangleq \widetilde{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + W^\top \varepsilon_\sigma \left( \widetilde{\Phi} \left( \xi \right)^2 \right) + \varepsilon \left( x \right)$

$$\dot{x} = W^\top \widehat{\sigma} + u + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + \delta$$

Let the tracking error be

$$e = x - x_d$$
$$\dot{e} = \dot{x} - \dot{x}_d$$
$$\dot{e} = W^\top \widehat{\sigma} + u + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + \delta - \dot{x}_d$$

And the weight errors are

$$\widetilde{W} = W - \widehat{W}$$
$$\dot{\widetilde{W}} = -\dot{\widehat{W}}$$
$$\widetilde{V} = V - \widehat{V}$$
$$\dot{\widetilde{V}} = -\dot{\widehat{V}}$$

Now let the Lyapunov candidate be defined as

$$V_L = \frac{1}{2} e^2 + \frac{1}{2\gamma_w} \widetilde{W}^\top \widetilde{W} + \frac{1}{2\gamma_v} \widetilde{V}^\top \widetilde{V}$$

Taking the time derivative yields

$$\dot{V}_L = e \left( W^\top \widehat{\sigma} + u + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + \delta - \dot{x}_d \right) + \frac{1}{\gamma_w} \widetilde{W}^\top \dot{\widetilde{W}} + \frac{1}{\gamma_v} \widetilde{V}^\top \dot{\widetilde{V}}$$

Now we can design the input

$$u = -\widehat{W}^\top \widehat{\sigma} + \dot{x}_d - \alpha_e e - \alpha_\delta \mathrm{sgn} \left( e \right)$$

Substitute this in

$$\dot{V}_L = e \left( W^\top \widehat{\sigma} - \widehat{W}^\top \widehat{\sigma} + \dot{x}_d - \alpha_e e - \alpha_\delta \mathrm{sgn} \left( e \right) + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + \delta - \dot{x}_d \right) + \frac{1}{\gamma_w} \widetilde{W}^\top \dot{\widetilde{W}} + \frac{1}{\gamma_v} \widetilde{V}^\top \dot{\widetilde{V}}$$
$$\dot{V}_L = e \left( \widetilde{W}^\top \widehat{\sigma} - \alpha_e e - \alpha_\delta \mathrm{sgn} \left( e \right) + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + \delta \right) - \frac{1}{\gamma_w} \widetilde{W}^\top \dot{\widehat{W}} - \frac{1}{\gamma_v} \widetilde{V}^\top \dot{\widehat{V}}$$
$$\dot{V}_L = e \widetilde{W}^\top \widehat{\sigma} - \alpha_e e^2 - \alpha_\delta \mathrm{sgn} \left( e \right) e + e \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + \delta e - \frac{1}{\gamma_w} \widetilde{W}^\top \dot{\widehat{W}} - \frac{1}{\gamma_v} \widetilde{V}^\top \dot{\widehat{V}}$$

Design the weight updates

$$\dot{\widehat{W}} = \mathrm{proj} \left( \gamma_w e \sigma \left( \widehat{\Phi} \left( \xi \right) \right) \right)$$
$$\dot{\widehat{V}} = \mathrm{proj} \left( \gamma_v e \widehat{W}^\top \widehat{\sigma}' \xi \right)$$

Substitute these in

$$\dot{V}_L = e \widetilde{W}^\top \widehat{\sigma} - \alpha_e e^2 - \alpha_\delta \mathrm{sgn} \left( e \right) e + e \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + \delta e - \frac{1}{\gamma_w} \widetilde{W}^\top \gamma_w e \sigma \left( \widehat{\Phi} \left( \xi \right) \right) - \frac{1}{\gamma_v} \widetilde{V}^\top \gamma_v e \widehat{W}^\top \widehat{\sigma}' \xi$$
$$\dot{V}_L = e \widetilde{W}^\top \widehat{\sigma} - \alpha_e e^2 - \alpha_\delta \mathrm{sgn} \left( e \right) e + e \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + \delta e - e \widetilde{W}^\top \sigma \left( \widehat{\Phi} \left( \xi \right) \right) - e \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi$$
$$\dot{V}_L = -\alpha_e e^2 - \alpha_\delta \mathrm{sgn} \left( e \right) e + \delta e$$

and using the bounds on $\delta$ to choose $\alpha_\delta > \overline{\delta}$

$$-\alpha_\delta |e| + |\delta||e| \leq 0$$

implying

$$\dot{V}_L \leq -\alpha_e e^2$$

and following the signals we can show asymptotic tracking using Barbalats Lemma.

### A. *If more than one basis*

Now for more than one basis we cant actually use the above exactly and now we will look over why. Recall that

$$\sigma\left(\Phi\left(\xi\right)\right)=\widehat{\sigma}+\widehat{\sigma'}\widetilde{\Phi}+\varepsilon_{\sigma}$$

where $\widehat{\sigma}=\sigma\left(\widehat{\Phi}\left(\xi\right)\right)$, $\widehat{\sigma'}=\frac{\partial\sigma}{\partial\Phi}\mid_{\widehat{\Phi}}$, $\widetilde{\Phi}=\widetilde{V}^{\top}\xi$, $\varepsilon_{\sigma}=\varepsilon_{\sigma}\left(\widetilde{\Phi}^{2}\left(\xi\right)\right)$. Now let $L=l+1$ where $W\in\mathbb{R}^{L}$ and $\Phi\left(\xi\right)\in\mathbb{R}^{l}$ with $\Phi\left(\xi\right)=V^{\top}\xi$ , $\xi\in\mathbb{R}^{n+1}$, $V^{\top}\in\mathbb{R}^{l\times n+1}$, and since $n=1$, $n+1=2$. This implies

$$\begin{aligned}
\Phi\left(\xi\right)&=V^{\top}\xi\\
\Phi\left(\xi\right)&=\begin{bmatrix}v_{x1}&v_{b1}\\ \vdots&\vdots\\ v_{xl}&v_{bl}\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}\\
\Phi\left(\xi\right)&=\begin{bmatrix}\begin{bmatrix}v_{x1}&v_{b1}\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}\\ \vdots\\ \begin{bmatrix}v_{xl}&v_{bl}\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}\end{bmatrix}\\
\Phi\left(\xi\right)&=\begin{bmatrix}v_{x1}x+v_{b1}\\ \vdots\\ v_{xl}x+v_{bl}\end{bmatrix}\\
\Phi\left(\xi\right)&=\begin{bmatrix}\Phi_{1}\\ \vdots\\ \Phi_{l}\end{bmatrix}
\end{aligned}$$

Now calculate $\sigma\left(\Phi\left(\xi\right)\right)$ using the estimates $\widehat{\Phi}\left(\xi\right)=\widehat{V}^{\top}\xi$

$$\sigma\left(\Phi\left(\xi\right)\right)=\begin{bmatrix}\sigma_{1}\left(\Phi_{1}\right)\\ \vdots\\ \sigma_{l}\left(\Phi_{l}\right)\\ \sigma_{L}\end{bmatrix}$$

Now calculate $\frac{\partial\sigma}{\partial\Phi}\mid_{\widehat{\Phi}}$ and since

$$\begin{aligned}
\frac{\partial\sigma}{\partial\Phi}&=\begin{bmatrix}\frac{\partial\sigma_{1}}{\partial\Phi_{1}}&\cdots&\frac{\partial\sigma_{1}}{\partial\Phi_{l}}\\ \vdots&\ddots&\vdots\\ \frac{\partial\sigma_{l}}{\partial\Phi_{1}}&\cdots&\frac{\partial\sigma_{l}}{\partial\Phi_{l}}\\ \frac{\partial\sigma_{L}}{\partial\Phi_{1}}&\cdots&\frac{\partial\sigma_{L}}{\partial\Phi_{l}}\end{bmatrix}\in\mathbb{R}^{L\times l}\\
\frac{\partial\sigma}{\partial\Phi}&=\begin{bmatrix}\frac{\partial\sigma_{1}}{\partial\Phi_{1}}&0&\cdots\\ 0&\ddots&\vdots\\ \vdots&\cdots&\frac{\partial\sigma_{l}}{\partial\Phi_{l}}\\ 0&\cdots&0\end{bmatrix}
\end{aligned}$$

And if we select

$$\sigma\left(\Phi\left(\xi\right)\right)=\begin{bmatrix}\sin\left(\Phi_{1}\right)\\ \vdots\\ \sin\left(\Phi_{l}\right)\\ 1\end{bmatrix}$$

then

$$\frac{\partial\sigma}{\partial\Phi}=\begin{bmatrix}\cos\left(\Phi_1\right) & 0 & \dots\\ 0 & \ddots & \vdots\\ \vdots & \dots & \cos\left(\Phi_l\right)\\ 0 & \dots & 0\end{bmatrix}$$

$$\frac{\partial\sigma}{\partial\Phi}\mid_{\widehat{\Phi}}=\begin{bmatrix}\cos\left(\widehat{\Phi}_1\right) & 0 & \dots\\ 0 & \ddots & \vdots\\ \vdots & \dots & \cos\left(\widehat{\Phi}_l\right)\\ 0 & \dots & 0\end{bmatrix}$$

Now for the analysis we must use the trace for $\widetilde{V}$ but not for $\widetilde{W}$ so we choose a candidate $V_L$ (this $L$ is not $l+1$ just stands for Lyapunov)

$$V_L\triangleq\frac{1}{2}e^2+\frac{1}{2}\widetilde{W}^\top\Gamma_w^{-1}\widetilde{W}+\frac{1}{2}\text{tr}\left(\widetilde{V}^\top\Gamma_v^{-1}\widetilde{V}\right)$$

Taking the time derivative yields

$$\dot{V}_L=e\dot{e}+\widetilde{W}^\top\Gamma_w^{-1}\dot{\widetilde{W}}+\text{tr}\left(\widetilde{V}^\top\Gamma_v^{-1}\dot{\widetilde{V}}\right)$$

Now we can design the input the same as before and everything reduces down to the relationship

$$\dot{V}_L=e\widetilde{W}^\top\widehat{\sigma}-\alpha_e e^2-\alpha_\delta\text{sgn}\left(e\right)e+e\widehat{W}^\top\widehat{\sigma}'\widetilde{V}^\top\xi+\delta e-\widetilde{W}^\top\Gamma_w^{-1}\dot{\widehat{W}}+\text{tr}\left(-\widetilde{V}^\top\Gamma_v^{-1}\dot{\widehat{V}}\right)$$

Design the weight updates for $W$

$$\dot{\widehat{W}}=\text{proj}\left(\Gamma_w e\widehat{\sigma}\right)$$

but we must do the trick for two vectors $a,b\in\mathbb{R}^n$

$$\text{tr}\left(ba^\top\right)=a^\top b$$

$$\underbrace{e\widehat{W}^\top\widehat{\sigma}'}_{a^\top}\underbrace{\widetilde{V}^\top\xi}_{b}=\text{tr}\left(\underbrace{\widetilde{V}^\top\xi}_{b}\underbrace{e\widehat{W}^\top\widehat{\sigma}'}_{a^\top}\right)$$

$$\dot{\widehat{V}}=\text{proj}\left(\Gamma_v\xi e\widehat{W}^\top\widehat{\sigma}'\right)$$

Substitute these in

$$\begin{aligned}\dot{V}_L&=e\widetilde{W}^\top\widehat{\sigma}-\alpha_e e^2-\alpha_\delta\text{sgn}\left(e\right)e+e\widehat{W}^\top\widehat{\sigma}'\widetilde{V}^\top\xi+\delta e-\widetilde{W}^\top\Gamma_w^{-1}\Gamma_w e\widehat{\sigma}+\text{tr}\left(-\widetilde{V}^\top\Gamma_v^{-1}\Gamma_v\xi e\widehat{W}^\top\widehat{\sigma}'\right)\\ \dot{V}_L&=-\alpha_e e^2-\alpha_\delta\text{sgn}\left(e\right)e+\delta e\\ \dot{V}_L&=-\alpha_e e^2-\alpha_\delta\text{sgn}\left(e\right)e+\delta e\end{aligned}$$

and using the bounds on $\delta$ to choose $\alpha_\delta>\overline{\delta}$

$$-\alpha_\delta|e|+|\delta||e|\leq 0$$

implying

$$\dot{V}_L\leq-\alpha_e e^2$$

## VI. Two-Link

Consider the two-link dynamics with an unknown force acting on the system

$$M\left(\phi\right)\ddot{\phi}+C\left(\phi,\dot{\phi}\right)+G\left(\phi\right)+\tau_d\left(\phi,\dot{\phi}\right)=\tau$$

$$\begin{aligned}M\left(\phi\right)&\triangleq\begin{bmatrix}m_1l_1^2+m_2\left(l_1^2+2l_1l_2c_2+l_2^2\right) & m_2\left(l_1l_2c_2+l_2^2\right)\\ m_2\left(l_1l_2c_2+l_2^2\right) & m_2l_2^2\end{bmatrix},\\ C\left(\phi,\dot{\phi}\right)&\triangleq\begin{bmatrix}-2m_2l_1l_2s_2\dot{\phi}_1\dot{\phi}_2\\ 0\end{bmatrix}+\begin{bmatrix}-m_2l_1l_2s_2\dot{\phi}_2^2\\ m_2l_1l_2s_2\dot{\phi}_1^2\end{bmatrix},\\ G\left(\phi\right)&\triangleq\begin{bmatrix}\left(m_1+m_2\right)gl_1c_1+m_2gl_2c_{12}\\ m_2gl_2c_{12}\end{bmatrix},\end{aligned}$$

where as before $\tau_d\left(\phi,\dot{\phi}\right)\in\mathbb{R}^2$ is a function of the state and could be a force of an object the system picked up, unknown friction at the joints, air resistance, any combination of these, or various other types of unstructured dynamics.

$$\tau_d\left(\phi,\dot{\phi}\right)=W^{\top}\sigma\left(\Phi\left(\xi\right)\right)+\varepsilon\left(\phi,\dot{\phi}\right)$$

where since the output is of dimension 2, $W\in\mathbb{R}^{L\times 2}$, $\sigma\left(\Phi\left(\xi\right)\right)\in\mathbb{R}^{L}$, $\Phi\left(\xi\right)\in\mathbb{R}^{l}$ is $\Phi\left(\xi\right)=V^{\top}\xi$, $V\in\mathbb{R}^{5\times l}$, $\xi\in\mathbb{R}^{5}$ is $\xi=\begin{bmatrix}\phi\\ \dot{\phi}\\ 1\end{bmatrix}$, and $\varepsilon\left(\phi,\dot{\phi}\right)\in\mathbb{R}^2$.

### A. Error Systems

Now lets develop the error systems with everything basically being the same as before

$$\begin{aligned}
e&=\phi_d-\phi\\
\dot{e}&=\dot{\phi}_d-\dot{\phi}\\
\ddot{e}&=\ddot{\phi}_d-\ddot{\phi}\\
r&=\dot{e}+\alpha e\\
\dot{r}&=\ddot{e}+\alpha\dot{e}\\
\dot{r}&=\ddot{\phi}_d-\ddot{\phi}+\alpha\dot{e}\\
M\left(\phi\right)\dot{r}&=M\left(\phi\right)\ddot{\phi}_d-M\left(\phi\right)\ddot{\phi}+M\left(\phi\right)\alpha\dot{e}\\
M\left(\phi\right)\dot{r}&=M\left(\phi\right)\left(\ddot{\phi}_d+\alpha\dot{e}\right)-M\left(\phi\right)\ddot{\phi}
\end{aligned}$$

Using the dynamics we know

$$M\left(\phi\right)\ddot{\phi}=-C\left(\phi,\dot{\phi}\right)-G\left(\phi\right)-\tau_d\left(\phi,\dot{\phi}\right)+\tau$$

implying

$$\begin{aligned}
M\left(\phi\right)\dot{r}&=M\left(\phi\right)\left(\ddot{\phi}_d+\alpha\dot{e}\right)-\left(-C\left(\phi,\dot{\phi}\right)-G\left(\phi\right)-\tau_d\left(\phi,\dot{\phi}\right)+\tau\right)\\
M\left(\phi\right)\dot{r}&=M\left(\phi\right)\left(\ddot{\phi}_d+\alpha\dot{e}\right)+C\left(\phi,\dot{\phi}\right)+G\left(\phi\right)-\tau_d\left(\phi,\dot{\phi}\right)-\tau
\end{aligned}$$

And recall from the analysis we get an extra term from the inertia derivative so lets add and subtract $\frac{1}{2}\dot{M}\left(\phi,\dot{\phi}\right)r$ now

$$M\left(\phi\right)\dot{r}=M\left(\phi\right)\left(\ddot{\phi}_d+\alpha\dot{e}\right)+C\left(\phi,\dot{\phi}\right)+G\left(\phi\right)+\tau_d\left(\phi,\dot{\phi}\right)-\tau\pm\frac{1}{2}\dot{M}\left(\phi,\dot{\phi}\right)r$$

We can write the

$$\dot{M}\left(\phi,\dot{\phi}\right)r=Y_{\dot{M}}\left(\phi,\dot{\phi},r\right)\theta$$

where

$$Y_{\dot{M}}\left(\phi,\dot{\phi},r\right)\triangleq\begin{bmatrix}0 & -\left(2s_2\dot{\phi}_2r_1+s_2\dot{\phi}_2r_2\right) & 0 & 0 & 0\\ 0 & -s_2\dot{\phi}_2r_1 & 0 & 0 & 0\end{bmatrix}.$$

Then we can write the above with $\varphi=\ddot{\phi}_d+\alpha\dot{e}$

$$M\left(\phi\right)\varphi+C\left(\phi,\dot{\phi}\right)+G\left(\phi\right)+\frac{1}{2}\dot{M}\left(\phi,\dot{\phi}\right)r=Y\theta$$

$$Y\left(\phi,\dot{\phi},\ddot{\phi}_d,\dot{e},r\right)=Y_M\left(\phi,\varphi\right)+Y_C\left(\phi,\dot{\phi}\right)+Y_G\left(\phi\right)+Y_{\dot{M}}\left(\phi,\dot{\phi},r\right)$$

$$\begin{aligned}
Y_M\left(\phi,\varphi\right)&\triangleq\begin{bmatrix}\varphi_1 & \left(2c_2\varphi_1+c_2\varphi_2\right) & \varphi_2 & 0 & 0\\ 0 & c_2\varphi_1 & \left(\varphi_1+\varphi_2\right) & 0 & 0\end{bmatrix},\\
Y_C\left(\phi,\dot{\phi}\right)&\triangleq\begin{bmatrix}0 & -\left(2s_2\dot{\phi}_1\dot{\phi}_2+s_2\dot{\phi}_2^2\right) & 0 & 0 & 0\\ 0 & s_2\dot{\phi}_1^2 & 0 & 0 & 0\end{bmatrix},\\
Y_G\left(\phi\right)&\triangleq\begin{bmatrix}0 & 0 & 0 & gc_1 & gc_{12}\\ 0 & 0 & 0 & 0 & gc_{12}\end{bmatrix},\\
Y_{\dot{M}}\left(\phi\right)&\triangleq\begin{bmatrix}0 & -\left(2s_2\dot{\phi}_2r_1+s_2\dot{\phi}_2r_2\right) & 0 & 0 & 0\\ 0 & -s_2\dot{\phi}_2r_1 & 0 & 0 & 0\end{bmatrix}
\end{aligned}$$

$$\begin{bmatrix}\theta_1\\\theta_2\\\theta_3\\\theta_4\\\theta_5\end{bmatrix}=\begin{bmatrix}m_1l_1^2+m_2l_1^2+m_2l_2^2\\m_2l_1l_2\\m_2l_2^2\\(m_1+m_2)\,l_1\\m_2l_2\end{bmatrix}$$

and we can use the NN

$$M(\phi)\,\dot{r}=Y\left(\phi,\dot{\phi},\ddot{\phi}_d,\dot{e},r\right)\theta+W^\top\sigma\left(\Phi\left(\xi\right)\right)+\varepsilon\left(\phi,\dot{\phi}\right)-\tau-\frac{1}{2}\dot{M}\left(\phi,\dot{\phi}\right)r$$

which we will write as

$$M\dot{r}=Y\theta+W^\top\sigma+\varepsilon-\tau-\frac{1}{2}\dot{M}r.$$

Now lets substitute $\sigma=\widehat{\sigma}+\widehat{\sigma'}\widetilde{\Phi}+\varepsilon_\sigma$

$$M\dot{r}=Y\theta+W^\top\left(\widehat{\sigma}+\widehat{\sigma'}\widetilde{\Phi}+\varepsilon_\sigma\right)+\varepsilon-\tau-\frac{1}{2}\dot{M}r$$

Now add and subtract $\widehat{W}^\top\frac{\partial\sigma}{\partial\Phi}\widetilde{\Phi}$

$$\begin{aligned}M\dot{r}&=Y\theta+W^\top\left(\widehat{\sigma}+\widehat{\sigma'}\widetilde{\Phi}+\varepsilon_\sigma\right)+\varepsilon-\tau-\frac{1}{2}\dot{M}r\pm\widehat{W}^\top\widehat{\sigma'}\widetilde{\Phi}\\M\dot{r}&=Y\theta+W^\top\widehat{\sigma}+\underbrace{\widetilde{W}^\top\widehat{\sigma'}\widetilde{\Phi}+W^\top\varepsilon_\sigma+\varepsilon}_{\delta}-\tau-\frac{1}{2}\dot{M}r+\widehat{W}^\top\widehat{\sigma'}\widetilde{\Phi}\end{aligned}$$

With $\delta\triangleq\widetilde{W}^\top\widehat{\sigma'}\widetilde{\Phi}+W^\top\varepsilon_\sigma+\varepsilon$

$$M\dot{r}=Y\theta+W^\top\widehat{\sigma}+\delta-\tau-\frac{1}{2}\dot{M}r+\widehat{W}^\top\widehat{\sigma'}\widetilde{V}^\top\xi$$

### *B. Analysis*

Let

$$\zeta\triangleq\begin{bmatrix}e\\r\\\widetilde{\theta}\\\mathrm{vec}\left(\widetilde{W}\right)\\\mathrm{vec}\left(\widetilde{V}\right)\end{bmatrix}$$

$$\mathrm{vec}\left(\widetilde{W}\right)=\begin{bmatrix}\widetilde{W}_{11}\\\vdots\\\widetilde{W}_{L1}\\\widetilde{W}_{12}\\\vdots\\\widetilde{W}_{L2}\end{bmatrix}$$

$$\mathrm{vec}\left(\widetilde{V}\right)=\begin{bmatrix}\widetilde{V}_{11}\\\vdots\\\widetilde{V}_{1l}\\\vdots\\\widetilde{V}_{51}\\\vdots\\\widetilde{V}_{5l}\end{bmatrix}$$

and

$$V(\zeta,t)\triangleq\frac{1}{2}e^\top e+\frac{1}{2}r^\top M(\phi)\,r+\frac{1}{2}\widetilde{\theta}^\top\Gamma_\theta^{-1}\widetilde{\theta}+\frac{1}{2}\mathrm{tr}\left(\widetilde{W}^\top\Gamma_W^{-1}\widetilde{W}\right)++\frac{1}{2}\mathrm{tr}\left(\widetilde{V}^\top\Gamma_V^{-1}\widetilde{V}\right)$$

taking the time derivative

$$\dot{V}(\zeta,t)=e^\top\dot{e}+\frac{1}{2}r^\top\dot{M}(\phi)\,r+r^\top M(\phi)\,\dot{r}+\widetilde{\theta}^\top\Gamma_\theta^{-1}\dot{\widetilde{\theta}}+\mathrm{tr}\left(\widetilde{W}^\top\Gamma_W^{-1}\dot{\widetilde{W}}\right)+\mathrm{tr}\left(\widetilde{V}^\top\Gamma_V^{-1}\dot{\widetilde{V}}\right)$$

then we can use

$$\begin{aligned} r &= \dot{e} + \alpha e \\ \implies \dot{e} &= r - \alpha e \end{aligned}$$

and

$$M\dot{r} = Y\theta + W^\top \sigma + \varepsilon - \tau - \frac{1}{2}\dot{M}r$$

$$\begin{aligned} \dot{V}(\zeta, t) = e^\top (r - \alpha e) + \frac{1}{2} r^\top \dot{M}(\phi) r + r^\top \left( Y\theta + W^\top \widehat{\sigma} + \delta - \tau - \frac{1}{2}\dot{M}r + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi \right) \\ - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left( -\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}} \right) + \text{tr}\left( -\widetilde{V}^\top \Gamma_V^{-1} \dot{\widehat{V}} \right) \end{aligned}$$

$$\begin{aligned} \dot{V}(\zeta, t) = -e^\top \alpha e + r^\top e + r^\top \left( Y\theta + W^\top \widehat{\sigma} + \delta - \tau + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi \right) \\ - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left( -\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}} \right) + \text{tr}\left( -\widetilde{V}^\top \Gamma_V^{-1} \dot{\widehat{V}} \right) \end{aligned}$$

*1) Input Design:* Now we can design the input

$$\tau = Y\widehat{\theta} + \widehat{W}^\top \widehat{\sigma} + e + \beta_r r + \beta_\delta \text{sgn}(r)$$

which yields

$$\begin{aligned} \dot{V}(\zeta, t) = -e^\top \alpha e + r^\top e + r^\top \left( Y\theta + W^\top \widehat{\sigma} + \delta - \left( Y\widehat{\theta} + \widehat{W}^\top \widehat{\sigma} + e + \beta_r r + \beta_\delta \text{sgn}(r) \right) + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi \right) \\ - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left( -\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}} \right) + \text{tr}\left( -\widetilde{V}^\top \Gamma_V^{-1} \dot{\widehat{V}} \right) \end{aligned}$$

$$\begin{aligned} \dot{V}(\zeta, t) = -e^\top \alpha e + r^\top e + r^\top \left( Y\widetilde{\theta} + \widetilde{W}^\top \widehat{\sigma} + \delta - e - \beta_r r - \beta_\delta \text{sgn}(r) + \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi \right) \\ - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left( -\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}} \right) + \text{tr}\left( -\widetilde{V}^\top \Gamma_V^{-1} \dot{\widehat{V}} \right) \end{aligned}$$

$$\begin{aligned} \dot{V}(\zeta, t) = -e^\top \alpha e + r^\top e + r^\top Y\widetilde{\theta} + r^\top \widetilde{W}^\top \widehat{\sigma} + r^\top \delta - r^\top e - r^\top \beta_r r - r^\top \beta_\delta \text{sgn}(r) + r^\top \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi \\ - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} + \text{tr}\left( -\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}} \right) + \text{tr}\left( -\widetilde{V}^\top \Gamma_V^{-1} \dot{\widehat{V}} \right) \end{aligned}$$

$$\begin{aligned} \dot{V}(\zeta, t) = &-e^\top \alpha e - r^\top \beta_r r \\ &- r^\top \beta_\delta \text{sgn}(r) + r^\top \delta \\ &+ r^\top Y\widetilde{\theta} - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} \\ &+ r^\top \widetilde{W}^\top \widehat{\sigma} - \text{tr}\left( \widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}} \right) \\ &+ r^\top \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi - \text{tr}\left( \widetilde{V}^\top \Gamma_V^{-1} \dot{\widehat{V}} \right) \end{aligned}$$

*2) Bound Approximation Errors:* Now we can bound

$$-r^\top \beta_\delta \text{sgn}(r) + r^\top \delta$$

as

$$\begin{aligned} -\underline{\beta_\delta} \|r\| + \overline{\delta} \|r\| \leq 0 \\ \underline{\beta_\delta} > \overline{\delta} \end{aligned}$$

Yielding

$$\begin{aligned} \dot{V}(\zeta, t) = &-e^\top \alpha e - r^\top \beta_r r \\ &+ r^\top Y\widetilde{\theta} - \widetilde{\theta}^\top \Gamma_\theta^{-1} \dot{\widehat{\theta}} \\ &+ r^\top \widetilde{W}^\top \widehat{\sigma} - \text{tr}\left( \widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}} \right) \\ &+ r^\top \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi - \text{tr}\left( \widetilde{V}^\top \Gamma_V^{-1} \dot{\widehat{V}} \right) \end{aligned}$$

*3) Design Structured Update Law:* Now we can design $\dot{\widehat{\theta}}$ using

$$r^\top Y\widetilde{\theta} - \widetilde{\theta}^\top \Gamma_\theta^{-1}\dot{\widehat{\theta}} = 0$$

$$\widetilde{\theta}^\top Y^\top r = \widetilde{\theta}^\top \Gamma_\theta^{-1}\dot{\widehat{\theta}}$$

$$\dot{\widehat{\theta}} = \text{proj}\left(\Gamma_\theta Y^\top r\right)$$

yielding

$$\begin{aligned}\widetilde{\theta}^\top Y^\top r &= \widetilde{\theta}^\top \Gamma_\theta^{-1}\Gamma_\theta Y^\top r\\ \widetilde{\theta}^\top Y^\top r &= \widetilde{\theta}^\top Y^\top r\end{aligned}$$

$$\begin{aligned}\dot{V}\left(\zeta,t\right) = &-e^\top \alpha e - r^\top \beta_r r\\ &+ r^\top \widetilde{W}^\top \widehat{\sigma} - \text{tr}\left(\widetilde{W}^\top \Gamma_W^{-1}\dot{\widehat{W}}\right)\\ &+ r^\top \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi - \text{tr}\left(\widetilde{V}^\top \Gamma_V^{-1}\dot{\widehat{V}}\right)\end{aligned}$$

*4) Design Unstructured Outer Weight Update Law:* Now we can design $\dot{\widehat{W}}$ using the trace trick for two vectors $a, b \in \mathbb{R}^n$

$$\text{tr}\left(ba^\top\right) = a^\top b$$

$$\begin{aligned}r^\top \widetilde{W}^\top \widehat{\sigma} + \text{tr}\left(-\widetilde{W}^\top \Gamma_W^{-1}\dot{\widehat{W}}\right) &= 0\\ r^\top \widetilde{W}^\top \widehat{\sigma} &= \text{tr}\left(\widetilde{W}^\top \Gamma_W^{-1}\dot{\widehat{W}}\right)\end{aligned}$$

$$\underbrace{r^\top}_{a^\top}\underbrace{\widetilde{W}^\top \widehat{\sigma}}_{b} = \text{tr}\left(\underbrace{\widetilde{W}^\top \widehat{\sigma}}_{b}\underbrace{r^\top}_{a^\top}\right)$$

$$\text{tr}\left(\widetilde{W}^\top \widehat{\sigma} r^\top\right) = \text{tr}\left(\widetilde{W}^\top \Gamma_W^{-1}\dot{\widehat{W}}\right)$$

$$\dot{\widehat{W}} = \text{proj}\left(\Gamma_W \widehat{\sigma} r^\top\right)$$

implying

$$\text{tr}\left(\widetilde{W}^\top \widehat{\sigma} r^\top\right) = \text{tr}\left(\widetilde{W}^\top \Gamma_W^{-1}\Gamma_W \widehat{\sigma} r^\top\right)$$

$$\text{tr}\left(\widetilde{W}^\top \widehat{\sigma} r^\top\right) = \text{tr}\left(\widetilde{W}^\top \widehat{\sigma} r^\top\right)$$

yielding

$$\begin{aligned}\dot{V}\left(\zeta,t\right) = &-e^\top \alpha e - r^\top \beta_r r\\ &+ r^\top \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi - \text{tr}\left(\widetilde{V}^\top \Gamma_V^{-1}\dot{\widehat{V}}\right)\end{aligned}$$

*5) Design Unstructured Inner Weight Update Law:* Last we can design $\dot{\widehat{V}}$

$$\begin{aligned}r^\top \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi + \text{tr}\left(-\widetilde{V}^\top \Gamma_V^{-1}\dot{\widehat{V}}\right) &= 0\\ r^\top \widehat{W}^\top \widehat{\sigma}' \widetilde{V}^\top \xi &= \text{tr}\left(\widetilde{V}^\top \Gamma_V^{-1}\dot{\widehat{V}}\right)\end{aligned}$$

$$\underbrace{r^\top \widehat{W}^\top \widehat{\sigma}'}_{a^\top}\underbrace{\widetilde{V}^\top \xi}_{b} = \text{tr}\left(\underbrace{\widetilde{V}^\top \xi}_{b}\underbrace{r^\top \widehat{W}^\top \widehat{\sigma}'}_{a^\top}\right)$$

$$\text{tr}\left(\widetilde{V}^\top \xi r^\top \widehat{W}^\top \widehat{\sigma}'\right) = \text{tr}\left(\widetilde{V}^\top \Gamma_V^{-1}\dot{\widehat{V}}\right)$$

$$\dot{\widehat{V}} = \text{proj}\left(\Gamma_V \xi r^\top \widehat{W}^\top \widehat{\sigma}'\right)$$

implying

$$\mathrm{tr}\left(\widetilde{V}^{\top}\xi r^{\top}\widehat{W}^{\top}\widehat{\sigma'}\right) = \mathrm{tr}\left(\widetilde{V}^{\top}\Gamma_V^{-1}\Gamma_V\xi r^{\top}\widehat{W}^{\top}\widehat{\sigma'}\right)$$
$$\mathrm{tr}\left(\widetilde{V}^{\top}\xi r^{\top}\widehat{W}^{\top}\widehat{\sigma'}\right) = \mathrm{tr}\left(\widetilde{V}^{\top}\xi r^{\top}\widehat{W}^{\top}\widehat{\sigma'}\right)$$

yielding

$$\dot{V}\left(\zeta,t\right) = -e^{\top}\alpha e - r^{\top}\beta_r r$$

*6) Final Bound on System:* And we can bound these as

$$\dot{V}\left(\zeta,t\right) \leq -\underline{\alpha}\|e\|^2 - \underline{\beta_r}\|r\|^2$$

then we can show asymptotic tracking using Barbalats Lemma