CHAPTER 15
Dynamically Driven Recurrent Networks

Problem 15.1

Referring to the simple recurrent neural network of Fig. 15.3, let the vector $\mathbf{u}(n)$ denote the input signal, the vector $\mathbf{x}(n)$ denotes the signal produced at the output of the hidden layer, and the vector $\mathbf{y}(n)$ denotes the output signal of the whole network. Then, treating $\mathbf{x}(n)$ as the state of the network, we may describe the state-space model of the network as follows:

$$\begin{aligned}\mathbf{x}(n+1) &= \mathbf{f}(\mathbf{x}(n), \mathbf{u}(n)) \\ \mathbf{y}(n) &= \mathbf{g}(\mathbf{x}(n))\end{aligned}$$

where $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are vector-valued functions of their respective arguments.

Problem 15.2

Referring to the recurrent MLP of Fig. 15.4, we note the following:

$$\mathbf{x}_{\mathrm{I}}(n+1) = \mathbf{f}_1(\mathbf{x}_{\mathrm{I}}(n), \mathbf{u}(n)) \tag{1}$$

$$\mathbf{x}_{\mathrm{II}}(n+1) = \mathbf{f}_2(\mathbf{x}_{\mathrm{II}}(n), \mathbf{x}_{\mathrm{I}}(n+1)) \tag{2}$$

$$\mathbf{x}_0(n+1) = \mathbf{f}_3(\mathbf{x}_0(n)\mathbf{x}_{\mathrm{II}}(n+1)) \tag{3}$$

where $\mathbf{f}_1(\cdot)$, $\mathbf{f}_2(\cdot)$, and $\mathbf{f}_3(\cdot)$ are vector-valued functions of their respective arguments. Substituting (1) into (2), we write

$$\mathbf{x}_{\mathrm{II}} = \mathbf{f}_2(\mathbf{x}_{\mathrm{II}}, \mathbf{f}_1(\mathbf{x}_{\mathrm{I}}, \mathbf{u}(n))) \tag{4}$$

Define the state of the system at time n as

$$\mathbf{x}_{\mathrm{II}}(n+1) = \begin{bmatrix} \mathbf{x}_{\mathrm{II}}(n) \\ \mathbf{x}_0(n-1) \end{bmatrix} \tag{5}$$

Then, from (4) and (5) we immediately see that

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(n), \mathbf{u}(n)) \tag{6}$$

where $\mathbf{f}$ is a new vector-valued function. Define the output of the system as

$$\mathbf{y}(n) = \mathbf{x}_0(n) \tag{7}$$

With $\mathbf{x}_0(n)$ included in the definition of the state $\mathbf{x}(n+1)$ and with $\mathbf{x}(n)$ dependent on the input $\mathbf{u}(n)$, we thus have

$$\mathbf{y}(n) = \mathbf{g}(\mathbf{x}(n), \mathbf{u}(n)) \tag{8}$$

where $\mathbf{g}(\cdot,\cdot)$ is another vector valued function. Equations (6) and (8) define the state-space model of the recurrent MLP.

Problem 15.3

It is indeed possible for a dynamic system to be controllable but unobservable, and vice versa. This statement is justified by virtue of the fact that the conditions for controllability and observability are entirely different, which means that there are situations where the conditions are satisfied for one and not for the other.

Problem 15.4

(a) We are given the process equation

$$\mathbf{x}(n+1) = \phi(\mathbf{W}_a\mathbf{x}(n) + \mathbf{w}_b u(n))$$

Hence, iterating forward in time, we write

$$\begin{aligned}\mathbf{x}(n+2) &= \phi(\mathbf{W}_a\mathbf{x}(n+1) + \mathbf{w}_b u(n+1)) \\ &= \phi(\mathbf{W}_a\phi(\mathbf{W}_a\mathbf{x}(n) + \mathbf{w}_b u(n)) + \mathbf{w}_b u(n+1))\end{aligned}$$

$$\begin{aligned}\mathbf{x}(n+3) &= \phi(\mathbf{W}_a\mathbf{x}(n+2) + \mathbf{w}_b u(n+2)) \\ &= \phi(\mathbf{W}_a\phi\mathbf{W}_a\phi(\mathbf{W}_a\mathbf{x}(n) + \mathbf{w}_b u(n)) + \mathbf{w}_b u(n+1)) + \mathbf{w}_b u(n+2))\end{aligned}$$

and so on. By induction, we may state that the state $\mathbf{x}(n+q)$ is a nested nonlinear function of $\mathbf{x}(n)$ and $\mathbf{u}_q(n)$, where

$$\mathbf{u}_q(n) = [u(n), u(n+1), \ldots, u(n+q-1)]^T$$

(b) The Jacobian of $\mathbf{x}(n+q)$ with respect to $\mathbf{u}_q(n)$ at the origin, is

$$\mathbf{J}_q(n) = \left[\frac{\partial\mathbf{x}(n+q)}{\partial\mathbf{u}_q(n)}\right]_{\substack{\mathbf{x}(n) = 0 \\ u(n) = 0}}$$

As an illustrative example, consider the cast of $q = 3$. The Jacobian of $\mathbf{x}(n+3)$ with respect to $\mathbf{u}_3(n)$ is

$$\mathbf{J}_3(n) = \left[\frac{\partial \mathbf{x}(n+3)}{\partial u(n)}, \frac{\partial \mathbf{x}(n+2)}{\partial u(n+1)}, \frac{\partial \mathbf{x}(n+3)}{\partial u(n+2)}\right]_{\substack{\mathbf{x}(n) = 0 \\ u(n) = 0}}$$

From the defining equation of $\mathbf{x}(n+3)$, we find that

$$\begin{aligned}\frac{\partial \mathbf{x}(n+3)}{\partial u(n)} &= \phi'(0)\mathbf{W}_a\phi'(0)\mathbf{W}_a\phi'(0)\mathbf{w}_b \\ &= \mathbf{AAb} \\ &= \mathbf{A}^2\mathbf{b}\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{x}(n+3)}{\partial u(n+1)} &= \phi'(0)\mathbf{W}_a\phi'(0)\mathbf{w}_b \\ &= \mathbf{Ab}\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{x}(n+3)}{\partial u(n+2)} &= \phi'(0)\mathbf{w}_b \\ &= \mathbf{b}\end{aligned}$$

All these partial derivatives have been evaluated at $\mathbf{x}(n) = 0$ and $u(n) = 0$. The Jacobian $\mathbf{J}_3(n)$ is therefore

$$\mathbf{J}_3(n) = [\mathbf{A}^2\mathbf{b}, \mathbf{Ab}, \mathbf{b}]$$

We may generalize this result by writing

$$\mathbf{J}_q(n) = [\mathbf{A}^{q-1}\mathbf{b}, \mathbf{A}^{q-2}\mathbf{b}, \ldots, \mathbf{Ab}, \mathbf{b}]$$

Problem 15.5

We start with the state-space model

$$\begin{aligned}\mathbf{x}(n+1) &= \phi(\mathbf{W}_a\mathbf{x}(n) + \mathbf{w}_b u(n)) \\ y(n) &= \mathbf{c}^T\mathbf{x}(n)\end{aligned} \tag{1}$$

where $\mathbf{c}$ is a column vector. We thus write

$$y(n+1) = \mathbf{c}^T\mathbf{x}(n+1)$$

$$= \mathbf{c}^T \phi(\mathbf{W}_a \mathbf{x}(n) + \mathbf{w}_b u(n)) \tag{2}$$

$$\begin{aligned} y(n+2) &= \mathbf{c}^T \mathbf{x}(n+2) \\ &= \mathbf{c}^T \phi(\mathbf{W}_a \phi(\mathbf{W}_a \mathbf{x}(n) + \mathbf{w}_b u(n)) + \mathbf{w}_b u(n+1)) \end{aligned} \tag{3}$$

and so on. By induction, we may therefore state that $y(n+q)$ is a nested nonlinear function of $\mathbf{x}(n)$ and $\mathbf{u}_q(n)$, where

$$\mathbf{u}_q(n) = [u(n), u(n+1), \ldots, u(n+q-1)]^T$$

Define the q-by-1 vector

$$\mathbf{y}_q(n) = [y(n), y(n+1), \ldots, y(n+q-1)]^T$$

The Jacobian of $\mathbf{y}_q(n)$ with respect to $\mathbf{x}(n)$, evaluated at the origin, is defined by

$$\mathbf{J}_q(n) = \left[\frac{\partial \mathbf{y}_q^T(n)}{\partial \mathbf{x}(n)}\right]_{\substack{\mathbf{x}(n) = \mathbf{0} \\ u(n) = 0}}$$

As an illustrative example, consider the case of $q = 3$, for which we have

$$\mathbf{J}_3(n) = \left[\frac{\partial y(n)}{\partial \mathbf{x}(n)}, \frac{\partial y(n+1)}{\partial \mathbf{x}(n)}, \frac{\partial y(n+2)}{\partial \mathbf{x}(n)}\right]_{\substack{\mathbf{x}(n) = 0 \\ u(n) = 0}}$$

From (1), we readily find that

$$\frac{\partial y(n)}{\partial \mathbf{x}(n)} = \mathbf{c}$$

From (2), we find that

$$\begin{aligned} \frac{\partial y(n+1)}{\partial \mathbf{x}(n)} &= \mathbf{c}(\phi'(0)\mathbf{W}_a)^T \\ &= \mathbf{c}\mathbf{A}^T \end{aligned}$$

From (3), we finally find that

$$\frac{\partial y(n+2)}{\partial \mathbf{x}(n)} = \mathbf{c}(\phi'(0)\mathbf{W}_a)^T(\phi'(0)\mathbf{W}_a)$$

$$= \mathbf{c}\mathbf{A}^T\mathbf{A}^T$$

$$= \mathbf{c}(\mathbf{A}^T)^2$$

All these partial derivatives have been evaluated at the origin. We thus write

$$\mathbf{J}_3(n) = [\mathbf{c}, \mathbf{c}(\mathbf{A}^T, \mathbf{c}\mathbf{A}^T)^2]$$

By induction, we may now state that the Jacobian $\mathbf{J}_q(n)$ for observability is, in general,

$$\mathbf{J}_q(n) = [\mathbf{c}, \mathbf{c}\mathbf{A}^T, \mathbf{c}(\mathbf{A}^T)^2, \ldots, \mathbf{c}(\mathbf{A}^T)^{q-1}]$$

where $\mathbf{c}$ is a column vector and $\mathbf{A} = \phi'(0)\mathbf{W}_a$.

Problem 15.6

We are given a nonlinear dynamic system described by

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(n), \mathbf{u}(n)) \tag{1}$$

Suppose $\mathbf{x}(n)$ is N-dimensional and $\mathbf{u}(n)$ is m-dimensional. Define a new nonlinear dynamic system in which the input is of additive form, as shown by

$$\mathbf{x}'(n+1) = \mathbf{f}'(\mathbf{x}'(n)) + \mathbf{u}'(n) \tag{2}$$

where

$$\mathbf{x}'(n) = \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{u}(n-1) \end{bmatrix} \tag{3}$$

$$\mathbf{u}'(n) = \begin{bmatrix} \mathbf{0} \\ \mathbf{u}(n) \end{bmatrix} \tag{4}$$

and

$$\mathbf{f}'(\mathbf{x}'(n)) = \begin{bmatrix} \mathbf{f}(\mathbf{x}(n), \mathbf{u}(n)) \\ \mathbf{0} \end{bmatrix} \tag{5}$$

Both $\mathbf{x}'(n)$ and $\mathbf{u}'(n)$ are $(N + m)$-dimensional, and the first N elements of $\mathbf{u}'(n)$ are zero. From these definitions, we readily see that

$$\mathbf{x}'(n+1) = \begin{bmatrix} \mathbf{x}(n+1) \\ \mathbf{u}(n) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{f}(\mathbf{x}(n), \mathbf{u}(n)) \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{u}(n) \end{bmatrix}$$

which is in perfect agreement with the description of the original nonlinear dynamic system defined in (1).

Problem 15.7

(a) The state-space model of the local activation feedback system of Fig. P15.7a depends on how the linear dynamic component is described. For example, we may define the input as

$$\mathbf{z}(n) = \begin{bmatrix} x(n-1) \\ \mathbf{B}\mathbf{u}(n) \end{bmatrix} \tag{1}$$

where $\mathbf{B}$ is a (p-1)-by-(p-1) matrix and

$$\mathbf{u}(n) = [u(n), u(n-1), \ldots, u(n-p+2)]^T$$

Let $\mathbf{w}$ denote the synaptic weight vector of the single neuron in Fig. P15.7a, with w_1 being the first element and $\mathbf{w}_0$ denoting the rest. we may then write

$$\begin{aligned} x(n) &= \mathbf{w}^T\mathbf{z}(n) + b \\ &= [w_1, \mathbf{w}_0^T]\begin{bmatrix} x(n-1) \\ \mathbf{B}\mathbf{u}(n) \end{bmatrix} + b \\ &= w_1 x(n-1) + \mathbf{B}'\mathbf{u}'(n) \end{aligned} \tag{2}$$

where

$$\mathbf{u}'(n) = \begin{bmatrix} \mathbf{u}'(n) \\ 1 \end{bmatrix}$$

and

$$\mathbf{B}' = [\mathbf{w}_0^T\mathbf{B}, b]$$

The output $y(n)$ is defined by

$$y(n) = \varphi(x(n)) \tag{3}$$

Equations (2) and (3) define the state-space model of Fig. P15.7a, assuming that its linear dynamic component is described by (1).

(b) Consider next the local output feedback system of Fig. 15.7b. Let the linear dynamic component of this system be described by (1). The output of the whole system in Fig. 15.7b is then defined by

$$\begin{aligned} x(n) &= \phi(\mathbf{w}^T\mathbf{z}(n) + b) \\ &= \phi\left([w_1, \mathbf{w}_0^T]\begin{bmatrix} x(n-1) \\ \mathbf{Bu}(n) \end{bmatrix} + b\right) \\ &= \phi(w_1 x(n-1) + \mathbf{B}'\mathbf{u}'(n)) \end{aligned} \tag{4}$$

where w_1, $\mathbf{w}_0$, $\mathbf{B}'$, and $\mathbf{u}'(n)$ are all as defined previously. The output $y(n)$ of Fig. P15.7b is

$$y(n) = x(n) \tag{5}$$

Equations (4) and (5) define the state-space model of the local output feedback system of Fig. P15.7b, assuming that its linear dynamic component is described by (1).

The process (state) equation of the local feedback system of Fig. P15.7a is linear but its measurement equation is nonlinear, and conversely for the local feedback system of Fig. P15.7b. These two local feedback systems are controllable and observable, because they both satisfy the conditions for controllability and observability.

Problem 15.8

We start with the state equation

$$\mathbf{x}(n+1) = \phi(\mathbf{W}_a\mathbf{x}(n) + \mathbf{w}_b u(n))$$

Hence, we write

$$\begin{aligned} \mathbf{x}(n+2) &= \phi(\mathbf{W}_a\mathbf{x}(n+1) + \mathbf{w}_b u(n+1)) \\ &= \phi(\mathbf{W}_a\phi(\mathbf{W}_a\mathbf{x}(n) + \mathbf{w}_b u(n)) + \mathbf{w}_b u(n+1)) \end{aligned}$$

$$\begin{aligned}\mathbf{x}(n+3) &= \phi(\mathbf{W}_a\mathbf{x}(n+2) + \mathbf{w}_b u(n+2)) \\ &= \phi(\mathbf{W}_a\phi(\mathbf{W}_a\phi(\mathbf{W}_a\mathbf{x}(n) + \mathbf{w}_b u(n)) + \mathbf{w}_b u(n+1)) + \mathbf{w}_b u(n+2))\end{aligned}$$

and so on.

By induction, we may now state that $\mathbf{x}(n+q)$ is a nested nonlinear function of $\mathbf{x}(n)$ and $\mathbf{u}_q(n)$, and thus write

$$\mathbf{x}(n+q) = \mathbf{g}(\mathbf{x}(n)\mathbf{u}_q(n))$$

where $\mathbf{g}$ is a vector-valued function, and

$$\mathbf{u}_q(n) = [u(n), u(n+1), \ldots, u(n+q-1)]^T$$

By definition, the output is correspondingly given by

$$\begin{aligned}y(n+q) &= \mathbf{c}^T\mathbf{x}(n+q) \\ &= \mathbf{c}^T\mathbf{g}(\mathbf{x}(n)\mathbf{u}_q(n)) \\ &= \Phi(\mathbf{x}(n), \mathbf{u}_q(n))\end{aligned}$$

where Φ is a new scalar-valued nonlinear function.

Problem 15.11

Consider a state-space model described by

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{x}(n), \mathbf{u}(n)) \tag{1}$$

$$\mathbf{y}(n) = \mathbf{g}(\mathbf{x}(n)) \tag{2}$$

Using (1), we may readily write

$$\begin{aligned}
\mathbf{x}(n) &= \mathbf{f}(\mathbf{x}(n-1), \mathbf{u}(n-1)) \\
\mathbf{x}(n-1) &= \mathbf{f}(\mathbf{x}(n-2), \mathbf{u}(n-2)) \\
\mathbf{x}(n-2) &= \mathbf{f}(\mathbf{x}(n-3), \mathbf{u}(n-3))
\end{aligned}$$

and so on. Accordingly, the simple recurrent network of Fig. 15.3 may be unfolded in time as follows:

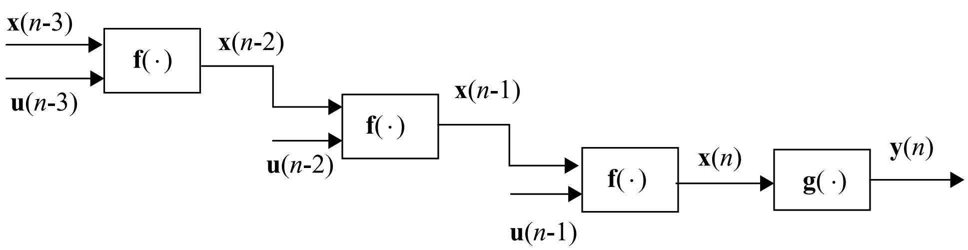

Figure : Problem 15.11

Problem 15.12

The local gradient for the hybrid form of the BPTT algorithm is given by

$$\delta_j(l) = \begin{cases} \varphi'(v_j(l))e_j(l) & \text{for } l = n \\ \varphi'(v_j(l))\left(e_j(l) + \sum_k w_{kj}(l)\delta_l(l+1)\right) & \text{for } n-h < l < n \\ \varphi'(v_j(l))\sum_k w_{kj}(l)\delta_l(l+1) & \text{for } n-h < l < j-h' \end{cases}$$

where h' is the number of additional steps taken before performing the next BPTT computation, with $h' < h$.

Problem 15.13

(a) The nonlinear state dynamics of the real-time recurrent learning algorithm of described in (15.48) and (15.52) olf the text may be reformulated in the equivalent form:

$$\frac{\partial y_j(n+1)}{\partial w_{kl}(n)} = \varphi'(v_j(n)) \sum_{i \in A \in B} w_{ji}(n) \frac{\partial \xi_i(n)}{\partial w_{kl}(n)} + \delta_{kj}\xi_l(n) \tag{1}$$

where δ_{kj} is the Kronecker delta and $y_j(n + 1)$ is the output of neuron j at time $n + 1$. For a teacher-forced recurrent network, we have

$$\xi_i(n) = \begin{cases} u_i(n) \text{ if } i \in A \\ d_i(n) \text{ if } i \in C \\ y_i(n) \text{ if } i \in B\text{-}C \end{cases} \tag{2}$$

Hence, substituting (2) into (1), we get

$$\frac{\partial y_j(n+1)}{\partial w_{kl}(n)} = \varphi'(v_j(n)) \sum_{i \in B\text{-}C} w_{ji}(n) \frac{\partial y_i(n)}{\partial w_{kl}(n)} + \delta_{kj}\xi_l(n) \tag{3}$$

(b) Let

$$\pi_{kl}^j(n) = \frac{\partial y_i(n)}{\partial w_{kl}(n)}$$

Provided that the learning-rate parameter η is small enough, we may put

$$\pi_{kl}^j(n+1) = \frac{\partial y_i(n+1)}{\partial w_{kl}(n+1)} \approx \frac{\partial y_i(n+1)}{\partial w_{kl}(n)}$$

Under this condition, we may rewrite (3) as follows:

$$\pi_{kl}^j(n+1) = \phi'(v_j(n)) \sum_{i \in B\text{-}C} w_{ji}(n) \pi_{kl}^j(n) + \delta_{kj}\xi_l(n) \tag{4}$$

This nonlinear state equation is the centerpiece of the RTRL algorithm using teacher forcing.