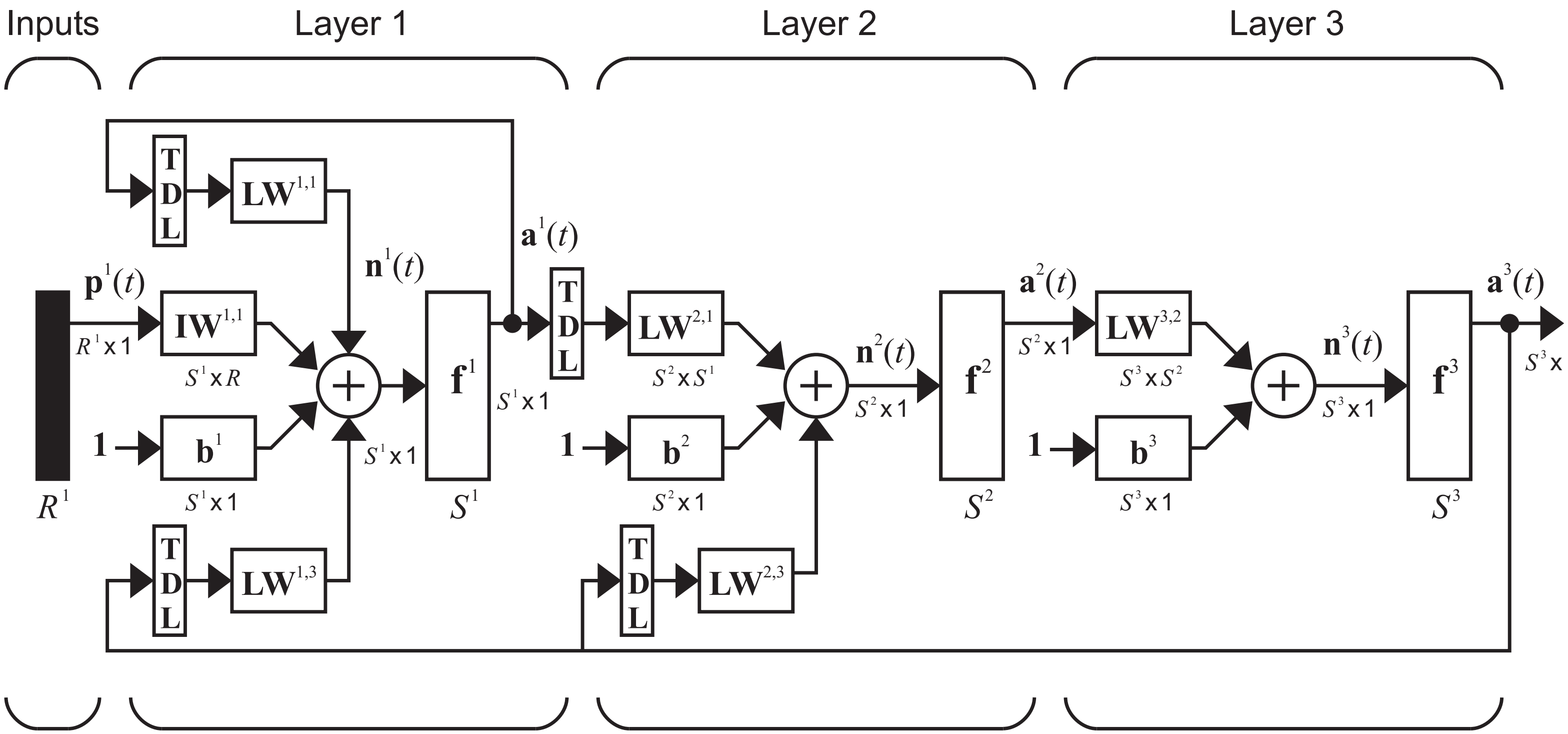
Dynamic Networks

Dynamic Networks

- Dynamic networks are networks that contain delays and that operate on a sequence of inputs.
- The ordering of the inputs is important to the operation of the network.
- In dynamic networks, the output depends not only on the current input to the network, but also on the current or previous inputs, outputs or states of the network.

Layered Digital Dynamic Networks



$$\mathbf{n}^m(t) = \sum_{l \in L_m^f} \sum_{d \in DL_{m,l}} \mathbf{LW}^{m,l}(d)\mathbf{a}^l(t-d) + \sum_{l \in I_m} \sum_{d \in DI_{m,l}} \mathbf{IW}^{m,l}(d)\mathbf{p}^1(t-d) + \mathbf{b}^m$$

$$\mathbf{a}^m(t) = \mathbf{f}^m(\mathbf{n}^m(t))$$

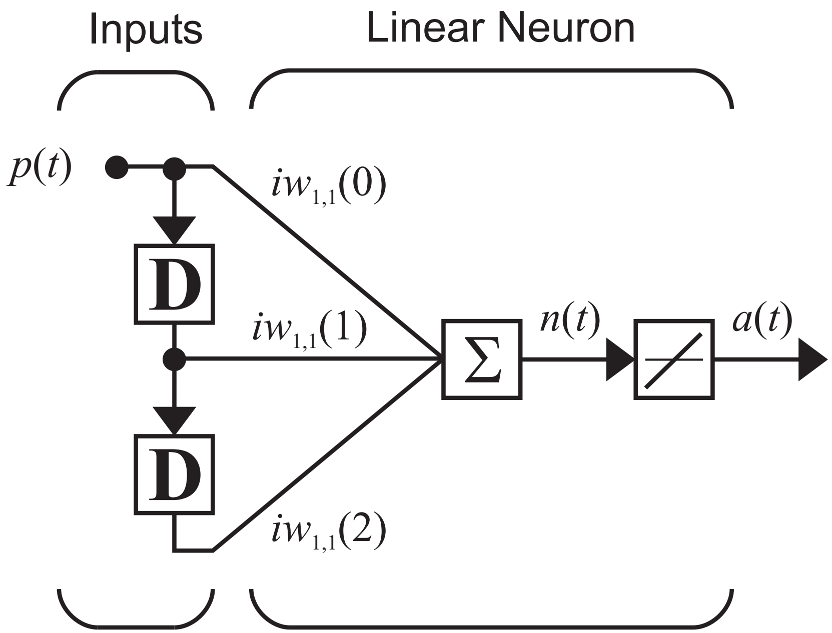
Layer Components

- A set of weight matrices that come into that layer (which may connect from other layers or from external inputs),
- Any tapped delay lines that appear at the input of a set of weight matrices,
- A bias vector,
- A summing junction, and
- A transfer function.

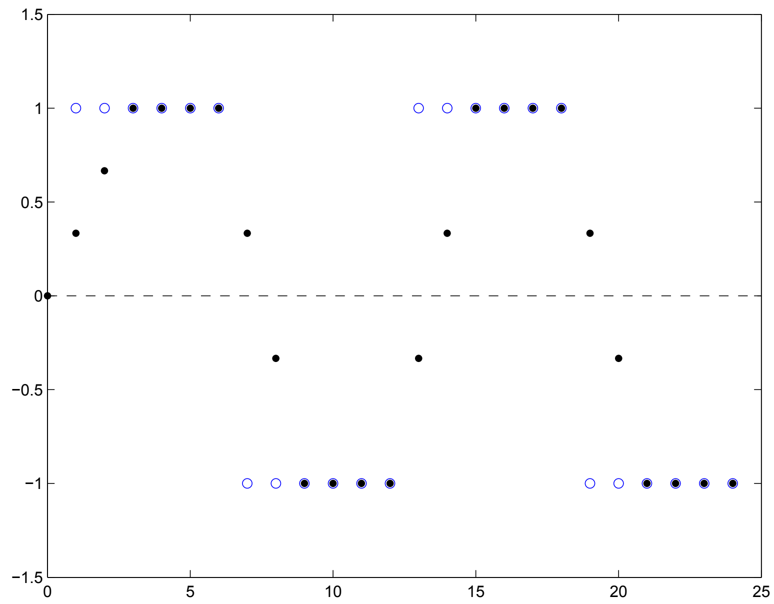
Definitions

- Simulation Order
- Backpropagation Order
- Input Layer (has an input weight, or contains any delays with any of its weight matrices)
- Output Layer (its output will be compared to a target during training, or it is connected to an input layer through a matrix that has any delays associated with it)

Example Feedforward Dynamic Network

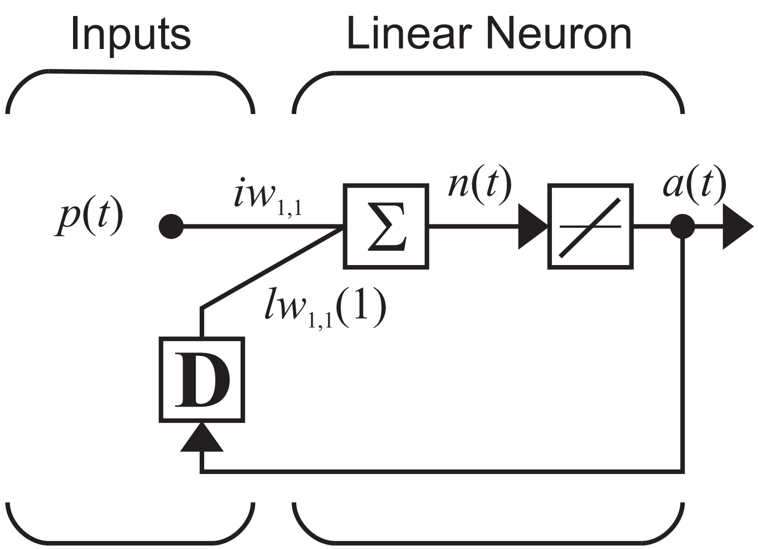


$a(t) = iw_{1,1}(0)p(\mathrm{t})+iw_{1,1}(1)p(\mathrm{t-1})+iw_{1,1}(2)p(\mathrm{t-2})$

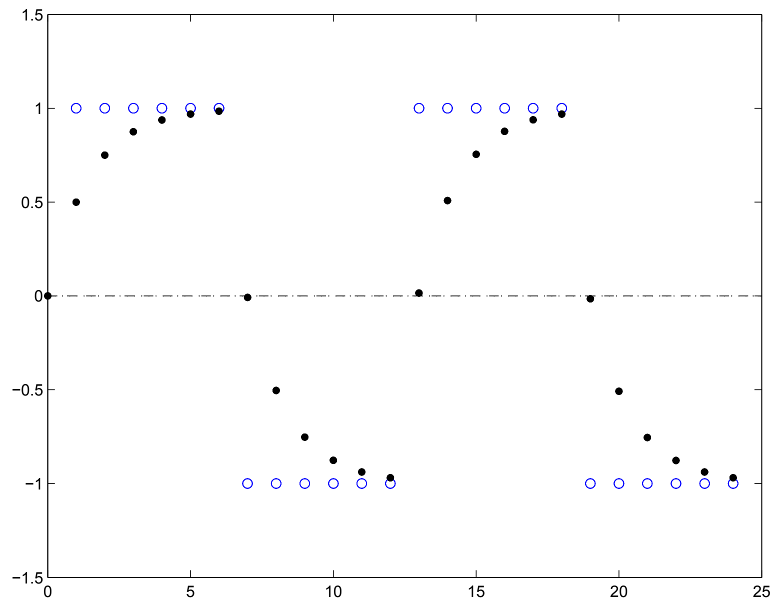


$$iw_{1,1}(0) = \frac{1}{3} \qquad iw_{1,1}(1) = \frac{1}{3} \qquad iw_{1,1}(2) = \frac{1}{3}$$

Example Recurrent Dynamic Network



$$a(t) = iw_{1,1}p(t) + lw_{1,1}(1)a(t\text{-}1)$$

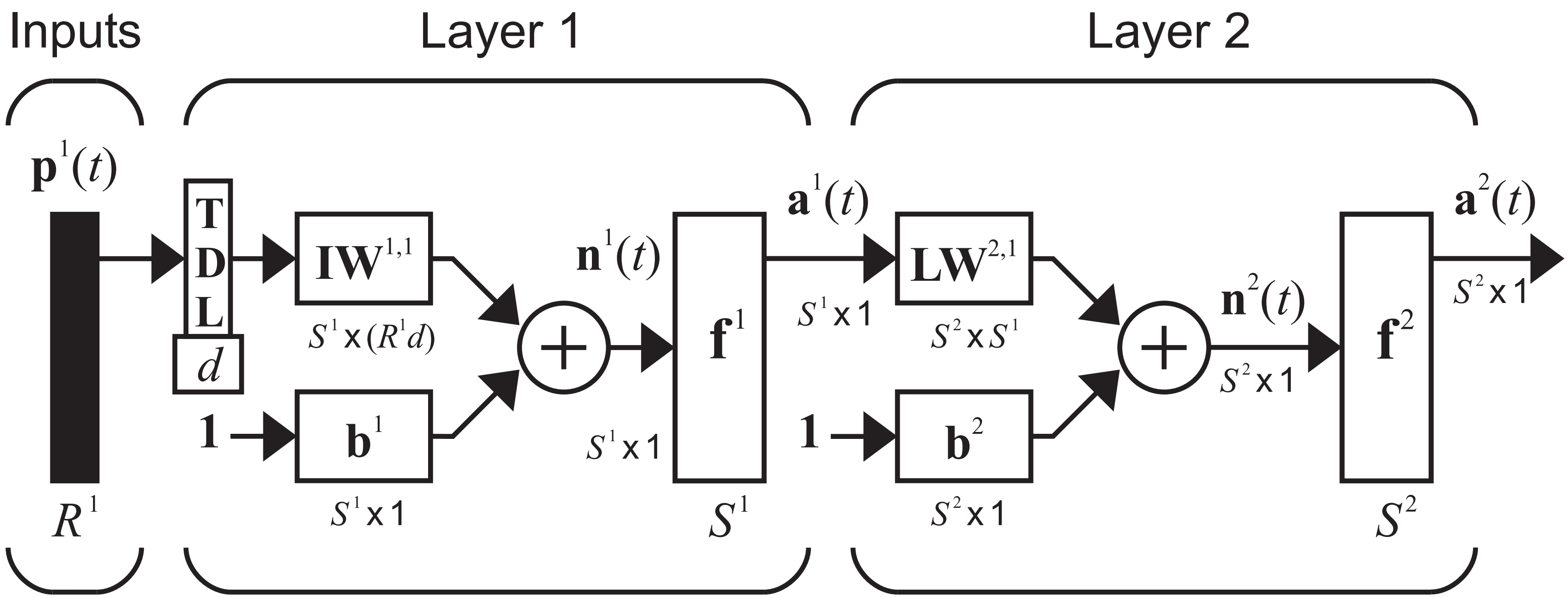


$$lw_{1,1}(1) = \frac{1}{2} \qquad iw_{1,1} = \frac{1}{2}$$

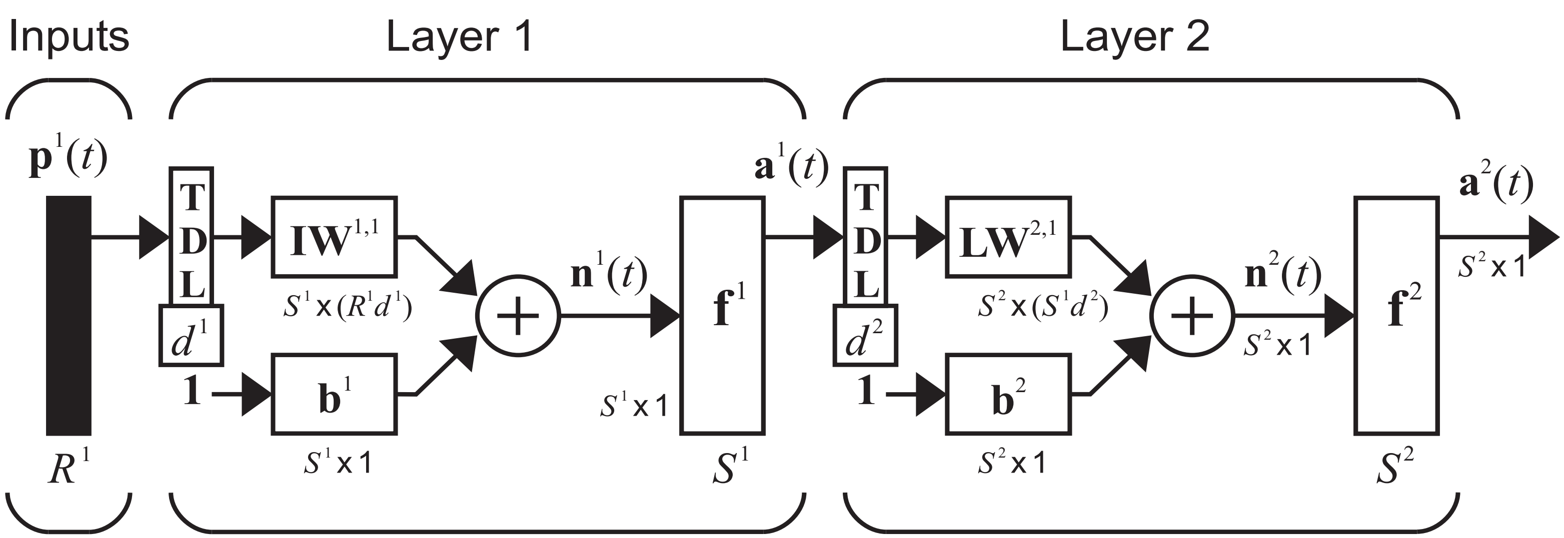
Applications of Dynamic Networks

- Prediction in financial markets
- Channel equalization in communication systems
- Phase detection in power systems
- Nonlinear filtering/fusion of sensor signals
- Fault detection
- Speech recognition
- Prediction of protein structure in genetics

Focused Time Delay Network

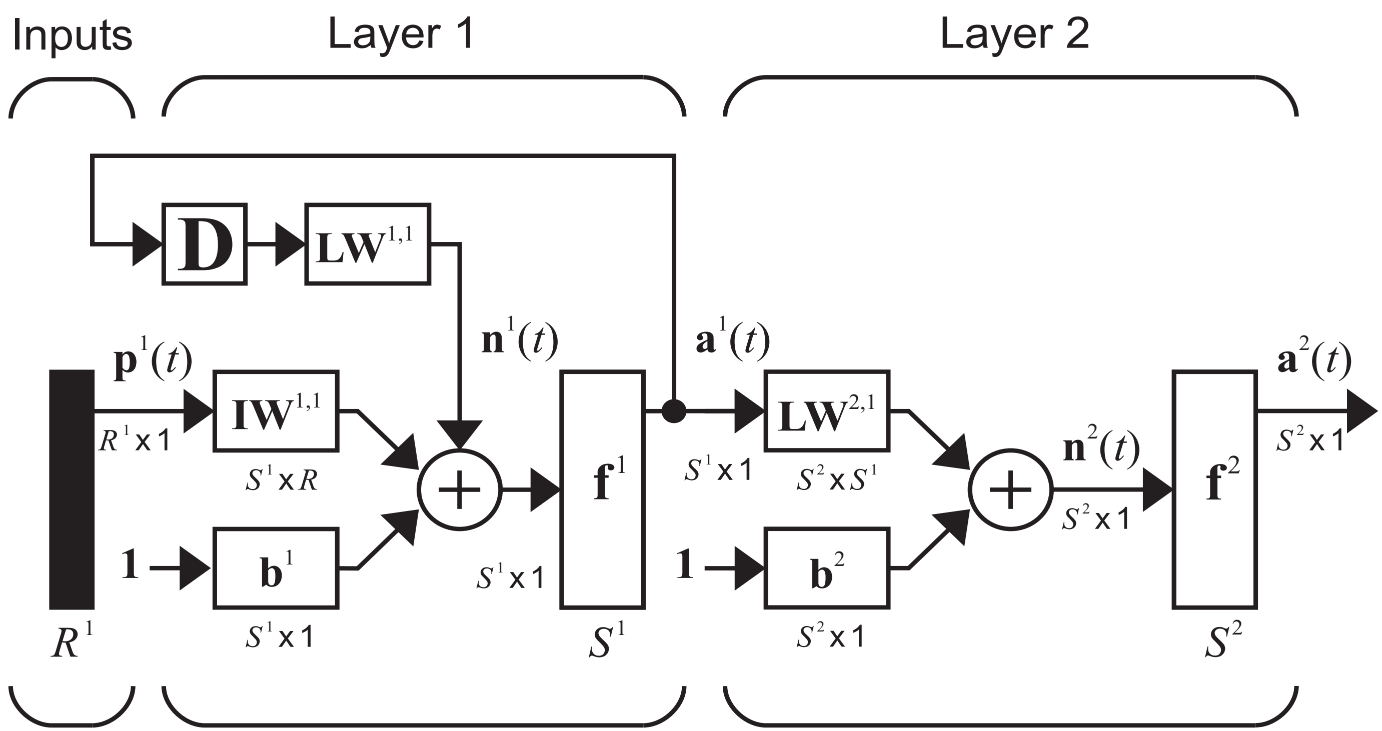


Distributed Time Delay Network



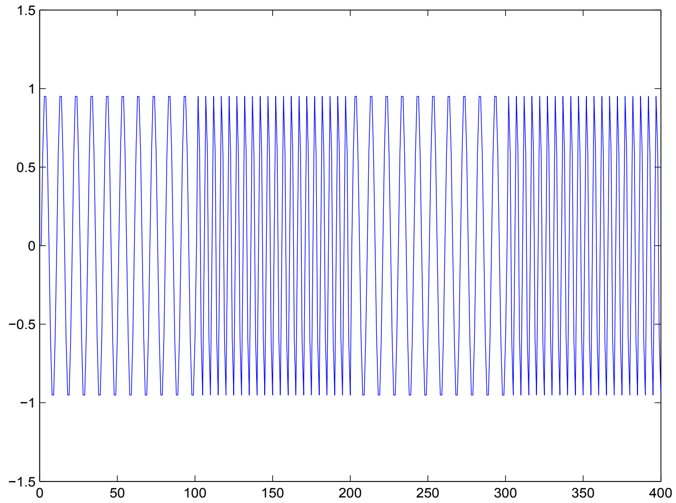
Layered Recurrent Network

(Elman)

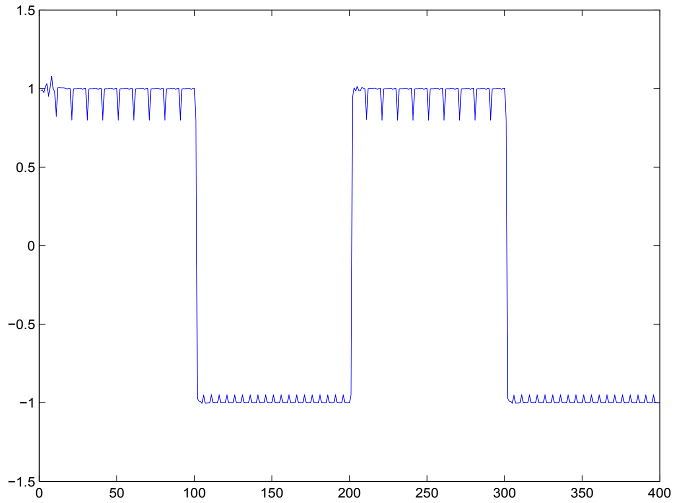


LRN Example

Network Input

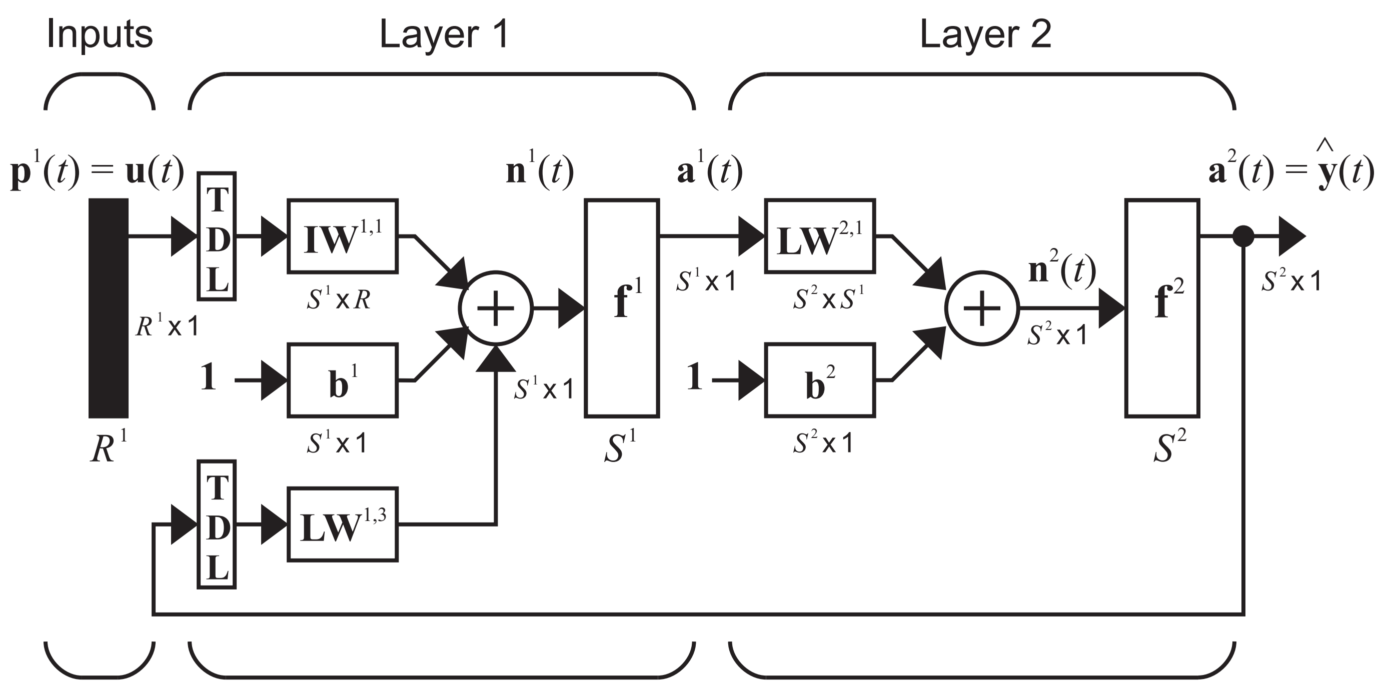


Network Response

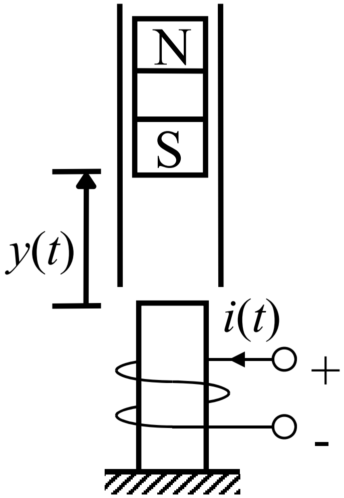


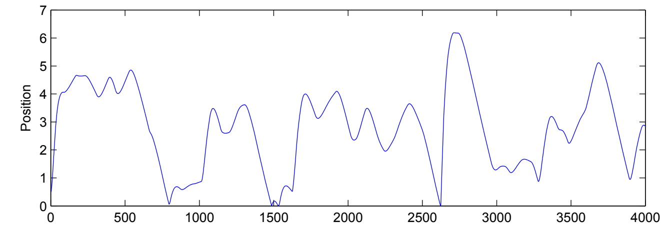
Nonlinear ARX

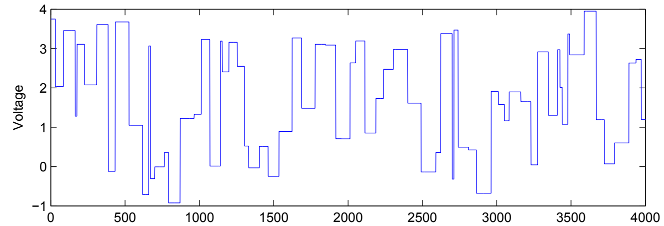
$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-n_y), u(t-1), u(t-2), \ldots, u(t-n_u))$$

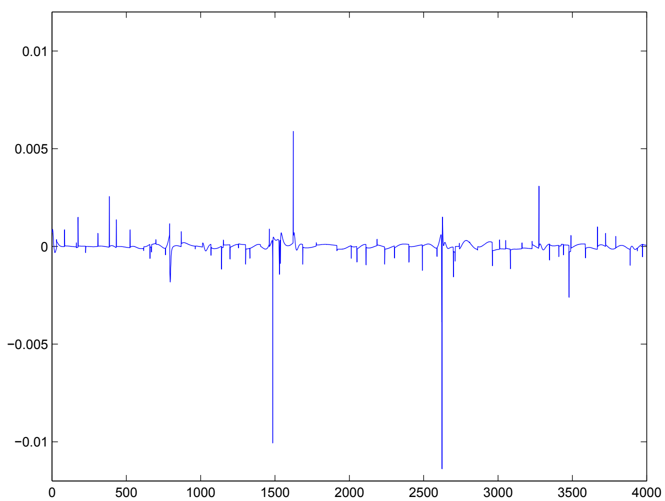


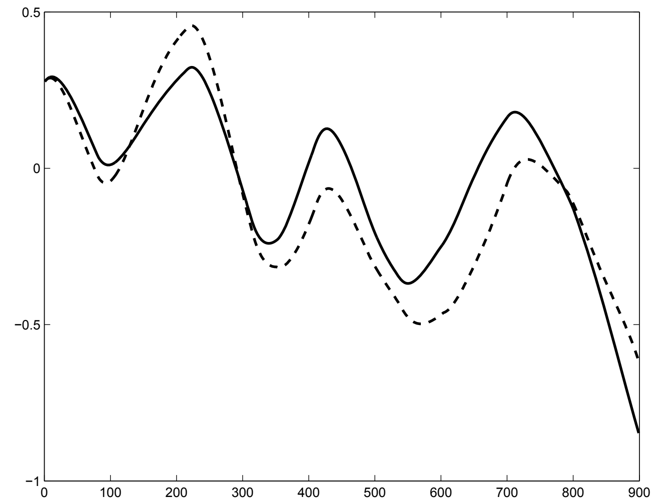
Modeling with NARX



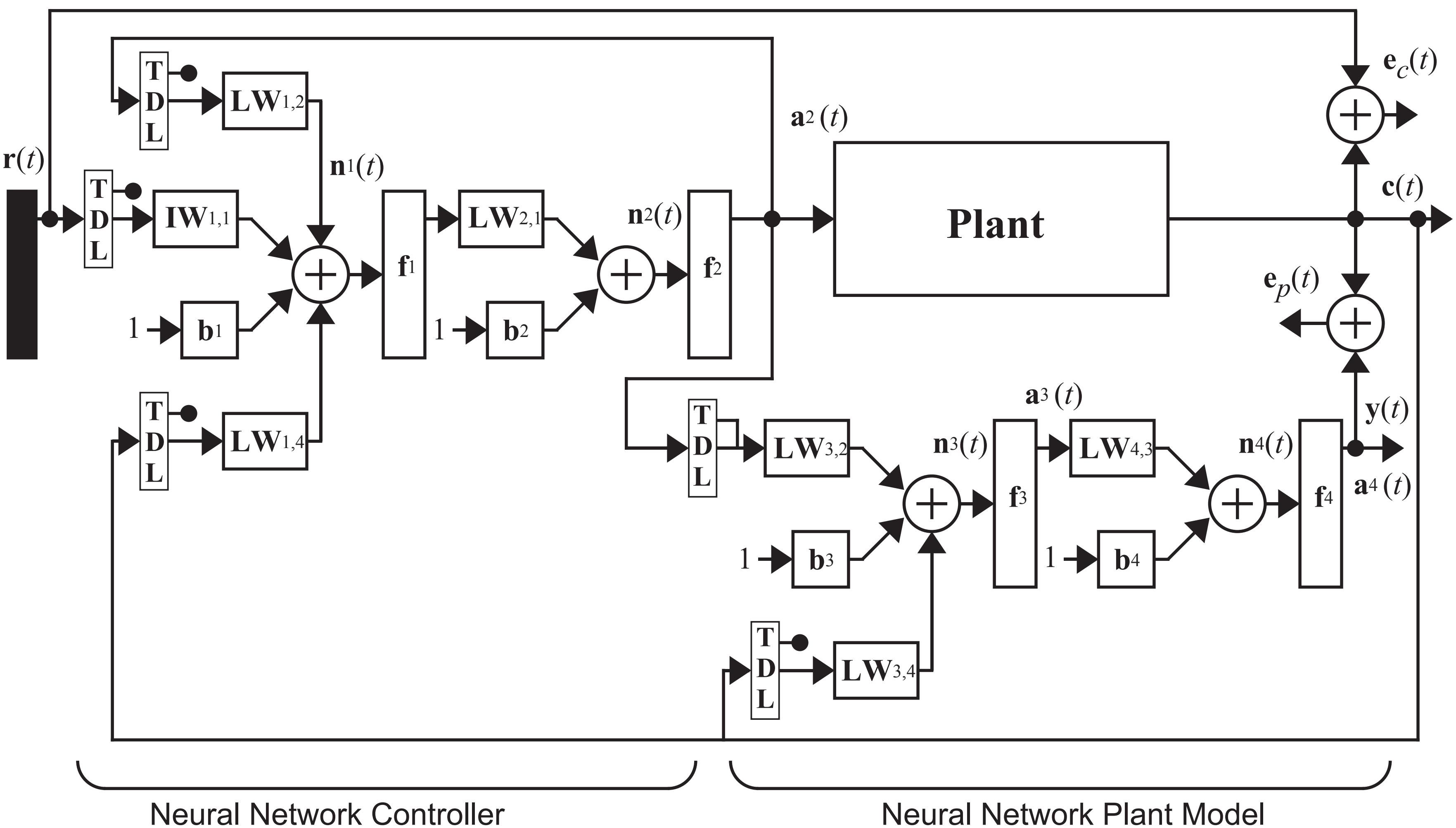




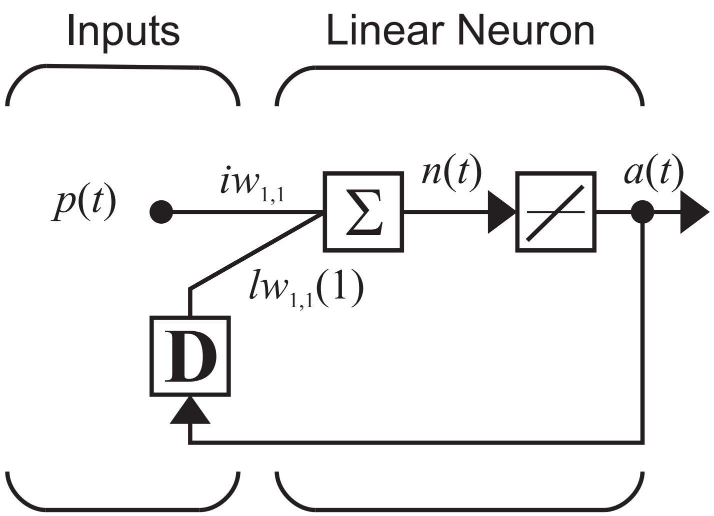




Model Reference Control



Example of Dynamic Derivatives



$$a(t) = iw_{1,1}p(t) + lw_{1,1}(1)a(t-1)$$

$$F(\mathbf{x}) = \sum_{t=1}^{Q} e^2(t) = \sum_{t=1}^{Q} (t(t) - a(t))^2$$

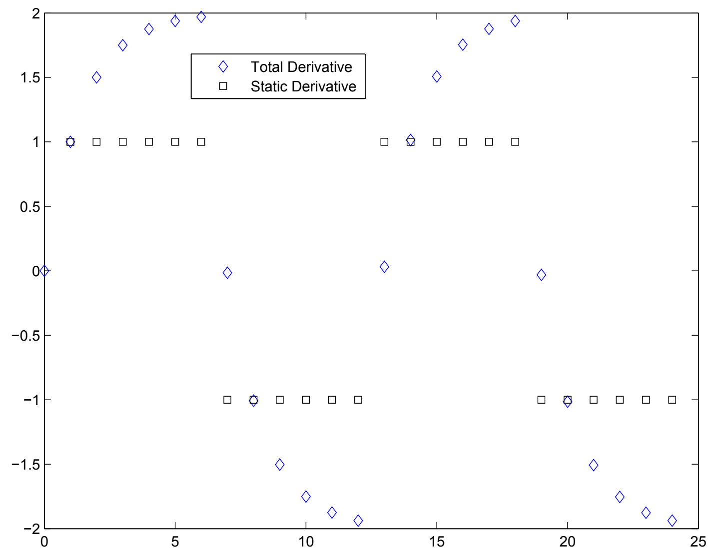
$$\frac{\partial F(\mathbf{x})}{\partial iw_{1,1}} = \sum_{t=1}^{Q} \frac{\partial e^2(t)}{\partial iw_{1,1}} = -2\sum_{t=1}^{Q} e(t)\frac{\partial a(t)}{\partial iw_{1,1}}$$

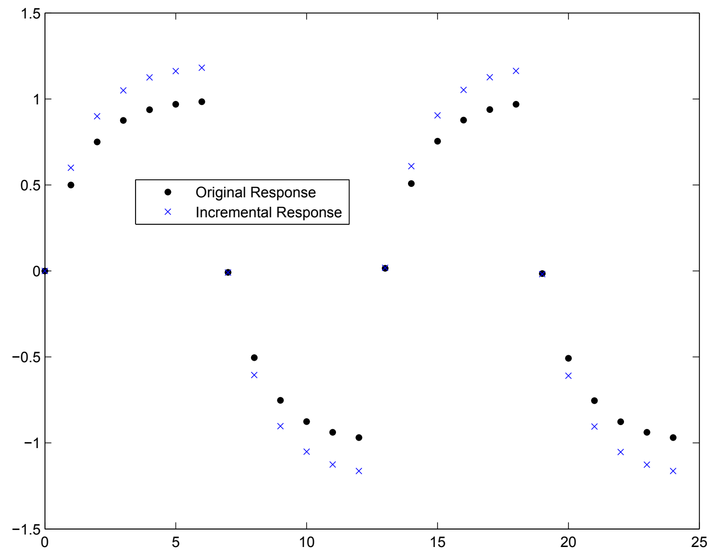
$$\frac{\partial F(\mathbf{x})}{\partial lw_{1,1}(1)} = \sum_{t=1}^{Q} \frac{\partial e^2(t)}{\partial lw_{1,1}(1)} = -2\sum_{t=1}^{Q} e(t)\frac{\partial a(t)}{\partial lw_{1,1}(1)}$$

$$\frac{\partial a(t)}{\partial lw_{1,1}(1)} = a(t-1) + lw_{1,1}(1)\frac{\partial a(t-1)}{\partial lw_{1,1}(1)} \qquad \frac{\partial a(t)}{\partial iw_{1,1}} = p(t) + lw_{1,1}(1)\frac{\partial a(t-1)}{\partial iw_{1,1}}$$

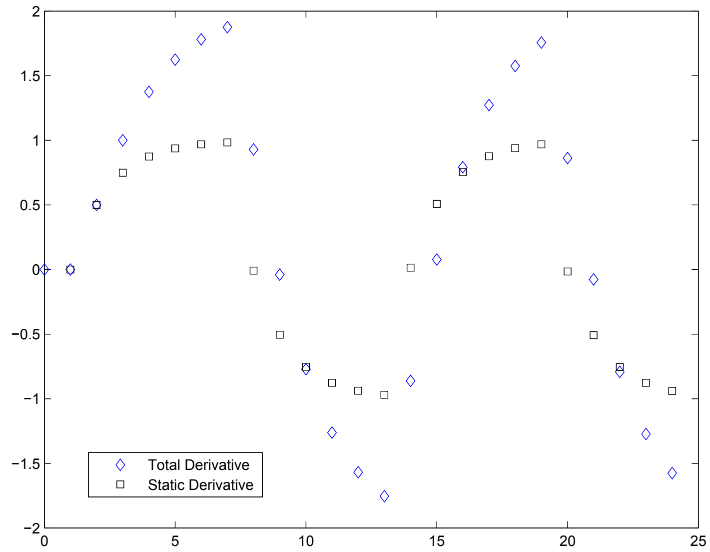
14

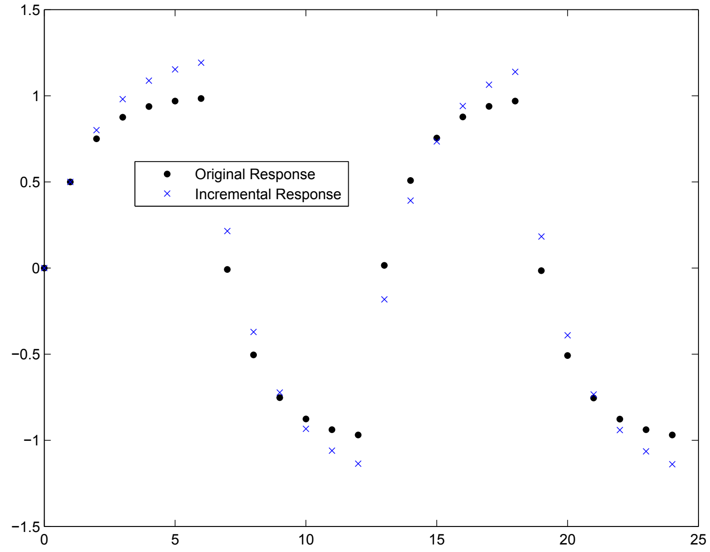
$iw_{1,1}$ Effect





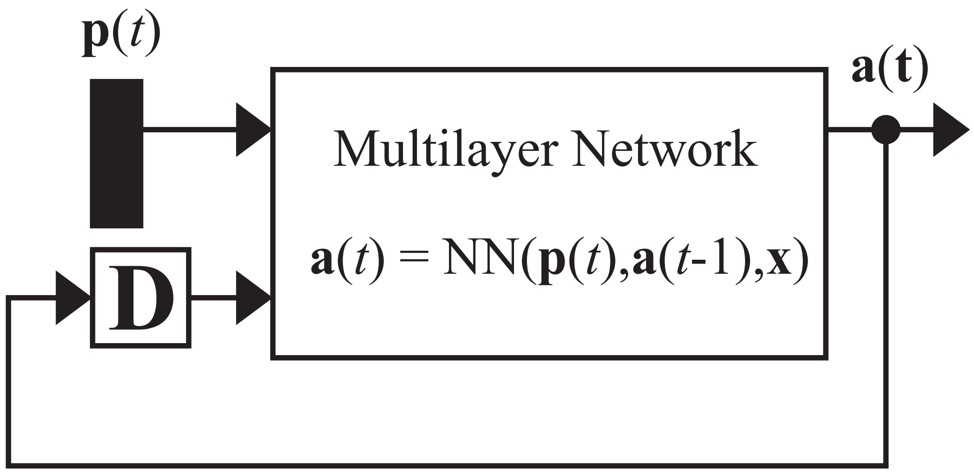
$lw_{1,1}(1)$ Effect





Basic Concepts of Dynamic Training

Simple Recurrent Network



Performance Index

$$F(\mathbf{x}) = \sum_{t=1}^{Q} (\mathbf{t}(t) - \mathbf{a}(t))^T (\mathbf{t}(t) - \mathbf{a}(t))$$

Dynamic Backpropagation

Real Time Recurrent Learning (RTRL)

$$\frac{\partial F}{\partial \mathbf{x}} = \sum_{t=1}^{Q} \left[\frac{\partial \mathbf{a}(t)}{\partial \mathbf{x}^T} \right]^T \times \frac{\partial^e F}{\partial \mathbf{a}(t)}$$

$$\frac{\partial \mathbf{a}(t)}{\partial \mathbf{x}^T} = \frac{\partial^e \mathbf{a}(t)}{\partial \mathbf{x}^T} + \frac{\partial^e \mathbf{a}(t)}{\partial \mathbf{a}^T(t-1)} \times \frac{\partial \mathbf{a}(t-1)}{\partial \mathbf{x}^T}$$

Backpropagation Through Time (BPTT)

$$\frac{\partial F}{\partial \mathbf{x}} = \sum_{t=1}^{Q} \left[\frac{\partial^e \mathbf{a}(t)}{\partial \mathbf{x}^T} \right]^T \times \frac{\partial F}{\partial \mathbf{a}(t)}$$

$$\frac{\partial F}{\partial \mathbf{a}(t)} = \frac{\partial^e F}{\partial \mathbf{a}(t)} + \frac{\partial^e \mathbf{a}(t+1)}{\partial \mathbf{a}^T(t)} \times \frac{\partial F}{\partial \mathbf{a}(t+1)}$$

14

General RTRL

$$\frac{\partial F}{\partial \mathbf{x}} = \sum_{t=1}^{Q} \left[\frac{\partial \mathbf{a}(t)}{\partial \mathbf{x}^T}\right]^T \times \frac{\partial^e F}{\partial \mathbf{a}(t)} \quad \Longrightarrow \quad \frac{\partial F}{\partial \mathbf{x}} = \sum_{t=1}^{Q} \sum_{u \in U} \left[\left[\frac{\partial \mathbf{a}^u(t)}{\partial \mathbf{x}^T}\right]^T \times \frac{\partial^e F}{\partial \mathbf{a}^u(t)}\right]$$

$$\frac{\partial \mathbf{a}(t)}{\partial \mathbf{x}^T} = \frac{\partial^e \mathbf{a}(t)}{\partial \mathbf{x}^T} + \frac{\partial^e \mathbf{a}(t)}{\partial \mathbf{a}^T(t-1)} \times \frac{\partial \mathbf{a}(t-1)}{\partial \mathbf{x}^T}$$

$$\Downarrow$$

$$\frac{\partial \mathbf{a}^u(t)}{\partial \mathbf{x}^T} = \frac{\partial^e \mathbf{a}^u(t)}{\partial \mathbf{x}^T} + \sum_{u' \in U} \sum_{x \in X} \sum_{d \in DL_{x,u'}} \frac{\partial^e \mathbf{a}^u(t)}{\partial \mathbf{n}^x(t)^T} \times \frac{\partial^e \mathbf{n}^x(t)}{\partial \mathbf{a}^{u'}(t-d)^T} \times \frac{\partial \mathbf{a}^{u'}(t-d)}{\partial \mathbf{x}^T}$$

Sensitivities

$$\frac{\partial^e \mathbf{a}^u(t)}{\partial \mathbf{n}^x(t)^T} \times \frac{\partial^e \mathbf{n}^x(t)}{\partial \mathbf{a}^{u'}(t-d)^T} = \mathbf{S}^{u,x}(t) \times \mathbf{LW}^{x,u'}(d)$$

$$\mathbf{S}^{u,m}(\mathrm{t}) = \frac{\partial^e \mathbf{a}^u(t)}{\partial \mathbf{n}^m(t)^T} = \begin{bmatrix} s^{u,m}_{1,1}(t) & s^{u,m}_{1,2}(t) & \dots & s^{u,m}_{1,S_m}(t) \\ s^{u,m}_{2,1}(t) & s^{u,m}_{2,2}(\mathrm{t}) & \dots & s^{u,m}_{2,S_m}(t) \\ \vdots & \vdots & & \vdots \\ s^{u,m}_{S_u,1}(t) & s^{u,m}_{S_u,2}(t) & \dots & s^{u,m}_{S_u,S_m}(t) \end{bmatrix}$$

$$\mathbf{S}^{u,m}(t) = \left[\sum_{l \in E_S(u) \cap L^b_m} \mathbf{S}^{u,l}(t)\mathbf{LW}^{l,m}(0) \right] \dot{\mathbf{F}}^m(\mathbf{n}^m(t)) \qquad \mathbf{S}^{u,u}(t) = \dot{\mathbf{F}}^u(\mathbf{n}^u(t))$$

14

Explicit Derivatives

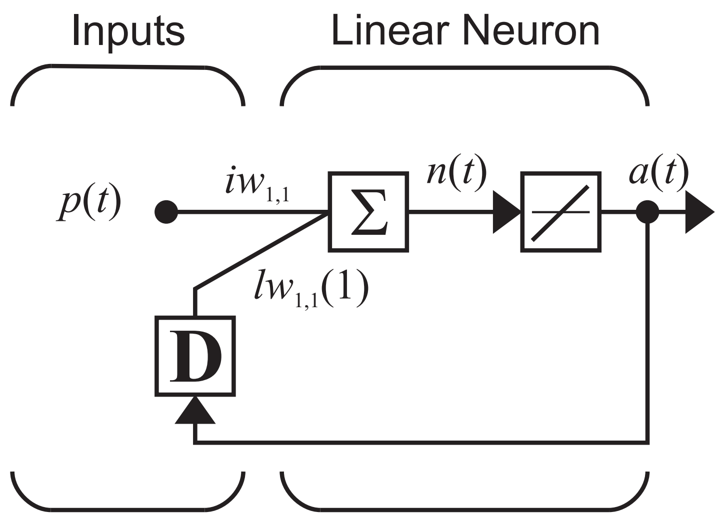
$$\frac{\partial^e a_k^u(t)}{\partial iw_{i,j}^{m,l}(d)} = \frac{\partial^e a_k^u(t)}{\partial n_i^m(t)} \times \frac{\partial^e n_i^m(t)}{\partial iw_{i,j}^{m,l}(d)} = s_{k,i}^{u,m}(t) \times p_j^l(t-d)$$

$$\frac{\partial^e \mathbf{a}^u(t)}{\partial vec(\mathbf{IW}^{m,l}(d))^T} = [\mathbf{p}^l(t-d)]^T \otimes \mathbf{S}^{u,m}(t)$$

$$\frac{\partial^e \mathbf{a}^u(t)}{\partial vec(\mathbf{LW}^{m,l}(d))^T} = [\mathbf{a}^l(t-d)]^T \otimes \mathbf{S}^{u,m}(t)$$

$$\frac{\partial^e \mathbf{a}^u(t)}{\partial(\mathbf{b}^m)^T} = \mathbf{S}^{u,m}(t)$$

RTRL Example (1)



$$a(t) = iw_{1,1}p(t) + lw_{1,1}(1)a(t\text{-}1)$$

$$F = \sum_{t=1}^{Q}(t(t) - a(t))^2 = \sum_{t=1}^{3} e^2(t) = e^2(1) + e^2(2) + e^2(3)$$

$$\{p(1), t(1)\}, \{p(2), t(2)\}, \{p(3), t(3)\}$$

$$a(1) = lw_{1,1}(1)a(0) + iw_{1,1}p(1)$$

RTRL Example (2)

$$\mathbf{S}^{1,1}(1) = \dot{\mathbf{F}}^1(\mathbf{n}^1(1)) = 1$$

$$\frac{\partial^e \mathbf{a}^1(1)}{\partial vec(\mathbf{IW}^{1,1}(0))^T} = \frac{\partial^e a(1)}{\partial iw_{1,1}} = [\mathbf{p}^1(1)]^T \otimes \mathbf{S}^{1,1}(1) = p(1)$$

$$\frac{\partial^e \mathbf{a}^1(1)}{\partial vec(\mathbf{LW}^{1,1}(1))^T} = \frac{\partial^e a(1)}{\partial lw_{1,1}(1)} = [\mathbf{a}^1(0)]^T \otimes \mathbf{S}^{1,1}(1) = a(0)$$

$$\frac{\partial \mathbf{a}^1(t)}{\partial \mathbf{x}^T} = \frac{\partial^e \mathbf{a}^1(t)}{\partial \mathbf{x}^T} + \mathbf{S}^{1,1}(t)\mathbf{LW}^{1,1}(1)\frac{\partial \mathbf{a}^1(t-1)}{\partial \mathbf{x}^T}$$

$$\frac{\partial a(1)}{\partial iw_{1,1}} = p(1) + lw_{1,1}(1)\frac{\partial a(0)}{\partial iw_{1,1}} = p(1)$$

$$\frac{\partial a(1)}{\partial lw_{1,1}(1)} = a(0) + lw_{1,1}(1)\frac{\partial a(0)}{\partial lw_{1,1}(1)} = a(0)$$

RTRL Example (3)

$$\frac{\partial^e a(2)}{\partial iw_{1,1}} = p(2) \qquad \frac{\partial^e a(2)}{\partial lw_{1,1}(1)} = a(1)$$

$$\frac{\partial a(2)}{\partial iw_{1,1}} = p(2) + lw_{1,1}(1)\frac{\partial a(1)}{\partial iw_{1,1}} = p(2) + lw_{1,1}(1)p(1)$$

$$\frac{\partial a(2)}{\partial lw_{1,1}(1)} = a(1) + lw_{1,1}(1)\frac{\partial a(1)}{\partial lw_{1,1}(1)} = a(1) + lw_{1,1}(1)a(0)$$

$$\frac{\partial^e a(3)}{\partial iw_{1,1}} = p(3) \qquad \frac{\partial^e a(3)}{\partial lw_{1,1}(1)} = a(2)$$

$$\frac{\partial a(3)}{\partial iw_{1,1}} = p(3) + lw_{1,1}(1)\frac{\partial a(2)}{\partial iw_{1,1}} = p(3) + lw_{1,1}(1)p(2) + (lw_{1,1}(1))^2 p(1)$$

$$\frac{\partial a(3)}{\partial lw_{1,1}(1)} = a(2) + lw_{1,1}(1)\frac{\partial a(2)}{\partial lw_{1,1}(1)} = a(2) + lw_{1,1}(1)a(1) + (lw_{1,1}(1))^2 a(0)$$

RTRL Example (4)

$$\frac{\partial F}{\partial \mathbf{x}} = \sum_{t=1}^{Q} \sum_{u \in U} \left[\left[\frac{\partial \mathbf{a}^u(t)}{\partial \mathbf{x}^T} \right]^T \times \frac{\partial^e F}{\partial \mathbf{a}^u(t)} \right] = \sum_{t=1}^{3} \left[\left[\frac{\partial \mathbf{a}^1(t)}{\partial \mathbf{x}^T} \right]^T \times \frac{\partial^e F}{\partial \mathbf{a}^1(t)} \right]$$

$$\frac{\partial F}{\partial iw_{1,1}} = \frac{\partial a(1)}{\partial iw_{1,1}}(-2e(1)) + \frac{\partial a(2)}{\partial iw_{1,1}}(-2e(2)) + \frac{\partial a(3)}{\partial iw_{1,1}}(-2e(3))$$

$$= -2e(1)[p(1)] - 2e(2)[p(2) + lw_{1,1}(1)p(1)]$$

$$-2e(3)[p(3) + lw_{1,1}(1)p(2) + (lw_{1,1}(1))^2 p(1)]$$

$$\frac{\partial F}{\partial lw_{1,1}(1)} = \frac{\partial a(1)}{\partial lw_{1,1}(1)}(-2e(1)) + \frac{\partial a(2)}{\partial lw_{1,1}(1)}(-2e(2)) + \frac{\partial a(3)}{\partial lw_{1,1}(1)}(-2e(3))$$

$$= -2e(1)[a(0)] - 2e(2)[a(1) + lw_{1,1}(1)a(0)]$$

$$-2e(3)[a(2) + lw_{1,1}(1)a(1) + (lw_{1,1}(1))^2 a(0)]$$

General BPTT

$$\frac{\partial F}{\partial \mathbf{x}} = \sum_{t=1}^{Q} \left[\frac{\partial^e \mathbf{a}(t)}{\partial \mathbf{x}^T} \right]^T \times \frac{\partial F}{\partial \mathbf{a}(t)}$$

$$\frac{\partial F}{\partial lw_{i,j}^{m,l}(d)} = \sum_{t=1}^{Q} \left[\sum_{u \in U} \sum_{k=1}^{S_u} \frac{\partial F}{\partial a_k^u(t)} \times \frac{\partial^e a_k^u(t)}{\partial n_i^m(t)} \right] \frac{\partial^e n_i^m(t)}{\partial lw_{i,j}^{m,l}(d)}$$

$$\frac{\partial F}{\partial \mathbf{a}(t)} = \frac{\partial^e F}{\partial \mathbf{a}(t)} + \frac{\partial^e \mathbf{a}(t+1)}{\partial \mathbf{a}^T(t)} \times \frac{\partial F}{\partial \mathbf{a}(t+1)}$$

$$\frac{\partial F}{\partial \mathbf{a}^u(t)} = \frac{\partial^e F}{\partial \mathbf{a}^u(t)} + \sum_{u' \in U} \sum_{x \in X} \sum_{d \in DL_{x,u}} \left[\frac{\partial^e \mathbf{a}^{u'}(t+d)}{\partial \mathbf{n}^x(t+d)^T} \times \frac{\partial^e \mathbf{n}^x(t+d)}{\partial \mathbf{a}^u(t)^T} \right]^T \times \frac{\partial F}{\partial \mathbf{a}^{u'}(t+d)}$$

$$\frac{\partial^e \mathbf{a}^{u'}(t+d)}{\partial \mathbf{n}^x(t+d)^T} \times \frac{\partial^e \mathbf{n}^x(t+d)}{\partial \mathbf{a}^u(t)^T} = \mathbf{S}^{u',x}(t+d) \times \mathbf{LW}^{x,u}(d)$$

14

BPTT

$$\frac{\partial F}{\partial lw_{i,j}^{m,l}(d)} = \sum_{t=1}^{Q}\left[\sum_{u \in U}\sum_{k=1}^{S_u}\frac{\partial F}{\partial a_k^u(t)} \times \frac{\partial^e a_k^u(t)}{\partial n_i^m(t)}\right]\frac{\partial^e n_i^m(t)}{\partial lw_{i,j}^{m,l}(d)}$$

$$\mathbf{d}^m(t) = \sum_{u \in U}[\mathbf{S}^{u,m}(t)]^T \times \frac{\partial F}{\partial \mathbf{a}^u(t)}$$

$$\frac{\partial F}{\partial \mathbf{LW}^{m,l}(d)} = \sum_{t=1}^{Q}\mathbf{d}^m(t) \times [\mathbf{a}^l(t-d)]^T$$

$$\frac{\partial F}{\partial \mathbf{IW}^{m,l}(d)} = \sum_{t=1}^{Q}\mathbf{d}^m(t) \times [\mathbf{p}^l(t-d)]^T$$

$$\frac{\partial F}{\partial \mathbf{b}^m} = \sum_{t=1}^{Q}\mathbf{d}^m(t)$$

BPTT Example (1)

$$a(1) = lw_{1,1}(1)a(0) + iw_{1,1}p(1)$$

$$a(2) = lw_{1,1}(1)a(1) + iw_{1,1}p(2)$$

$$a(3) = lw_{1,1}(1)a(2) + iw_{1,1}p(3)$$

$$\mathbf{S}^{1,1}(3) = \dot{\mathbf{F}}^{1}(\mathbf{n}^{1}(3)) = 1$$

$$\frac{\partial F}{\partial \mathbf{a}^{1}(t)} = \frac{\partial^{e} F}{\partial \mathbf{a}^{1}(t)} + \mathbf{LW}^{1,1}(1)^{T}\mathbf{S}^{1,1}(t+1)^{T} \times \frac{\partial F}{\partial \mathbf{a}^{1}(t+1)}$$

$$\frac{\partial F}{\partial \mathbf{a}^{1}(3)} = \frac{\partial^{e} F}{\partial \mathbf{a}^{1}(3)} + lw_{1,1}(1)\mathbf{S}^{1,1}(4)^{T} \times \cancelto{0}{\frac{\partial F}{\partial \mathbf{a}^{1}(4)}} = \frac{\partial^{e} \mathrm{F}}{\partial \mathbf{a}^{1}(3)} = -2e(3)$$

$$\mathbf{d}^{1}(3) = [\mathbf{S}^{1,1}(3)]^{T} \times \frac{\partial F}{\partial \mathbf{a}^{1}(3)} = -2e(3)$$

14

BPTT Example (2)

$$\mathbf{S}^{1,1}(2) = \dot{\mathbf{F}}^{1}(\mathbf{n}^{1}(2)) = 1$$

$$\frac{\partial F}{\partial \mathbf{a}^{1}(2)} = \frac{\partial^{e} F}{\partial \mathbf{a}^{1}(2)} + lw_{1,1}(1)\mathbf{S}^{1,1}(3)^{T} \times \frac{\partial F}{\partial \mathbf{a}^{1}(3)}$$

$$= -2e(2) + lw_{1,1}(1)(-2e(3))$$

$$\mathbf{d}^{1}(2) = [\mathbf{S}^{1,1}(2)]^{T} \times \frac{\partial F}{\partial \mathbf{a}^{1}(2)} = -2e(2) + lw_{1,1}(1)(-2e(3))$$

$$\mathbf{S}^{1,1}(1) = \dot{\mathbf{F}}^{1}(\mathbf{n}^{1}(1)) = 1$$

$$\frac{\partial F}{\partial \mathbf{a}^{1}(1)} = \frac{\partial^{e} F}{\partial \mathbf{a}^{1}(1)} + lw_{1,1}(1)\mathbf{S}^{1,1}(2)^{T} \times \frac{\partial F}{\partial \mathbf{a}^{1}(2)}$$

$$= -2e(1) + lw_{1,1}(1)(-2e(2)) + (lw_{1,1}(1))^{2}(-2e(3))$$

$$\mathbf{d}^{1}(1) = [\mathbf{S}^{1,1}(1)]^{T} \times \frac{\partial F}{\partial \mathbf{a}^{1}(1)} = -2e(1) + lw_{1,1}(1)(-2e(2)) + (lw_{1,1}(1))^{2}(-2e(3))$$

BPTT Example (3)

$$\frac{\partial F}{\partial \mathbf{LW}^{1,1}(1)} = \frac{\partial F}{\partial lw_{1,1}(1)} = \sum_{t=1}^{3} \mathbf{d}^1(t) \times [\mathbf{a}^1(t-1)]^T$$

$$= a(0)[-2e(1) + lw_{1,1}(1)(-2e(2)) + (lw_{1,1}(1))^2(-2e(3))]$$

$$+ a(1)[-2e(2) + lw_{1,1}(1)(-2e(3))] + a(0)[-2e(3)]$$

$$\frac{\partial F}{\partial \mathbf{IW}^{1,1}(0)} = \frac{\partial F}{\partial iw_{1,1}} = \sum_{t=1}^{3} \mathbf{d}^1(t) \times [\mathbf{p}^1(t)]^T$$

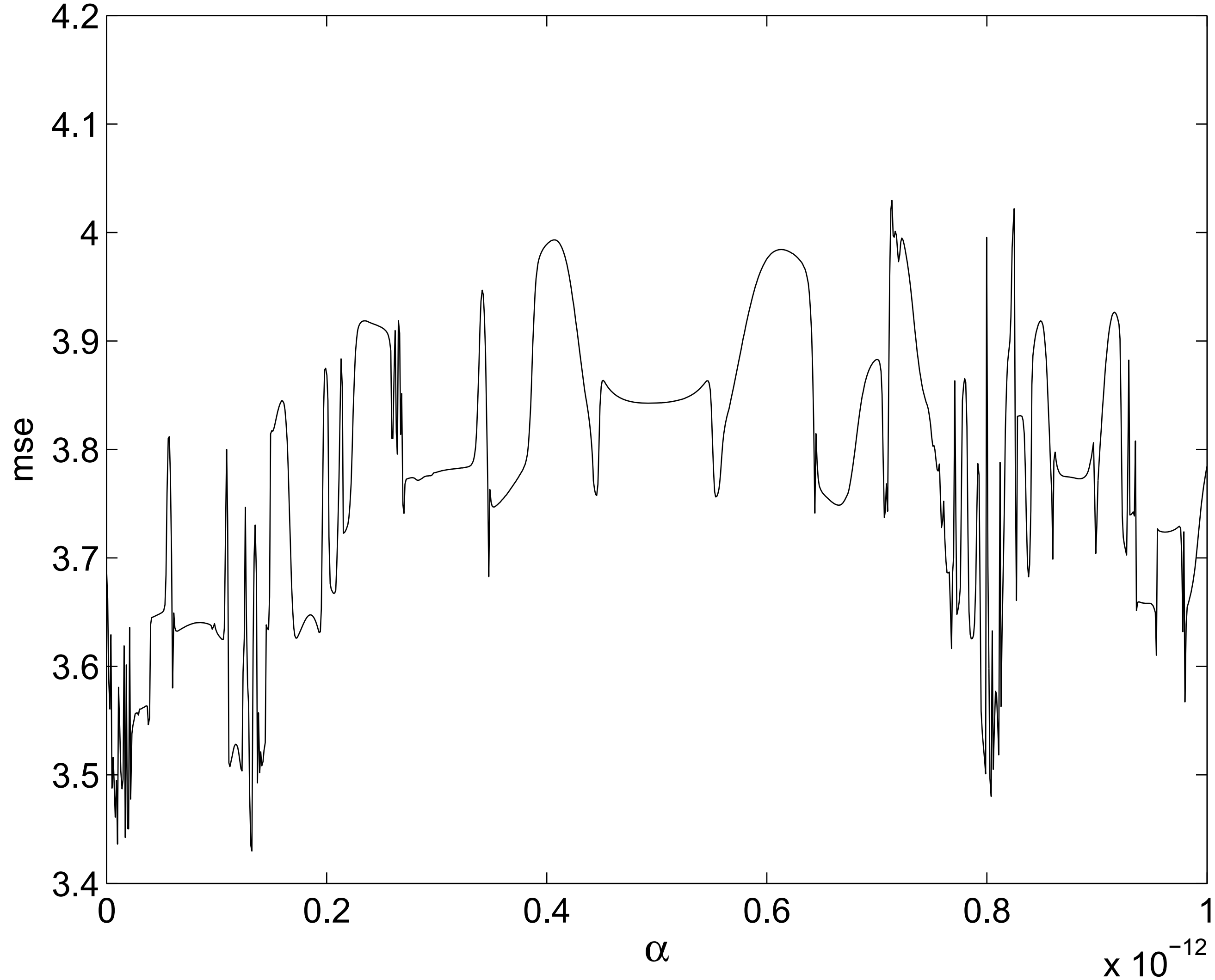
$$= p(1)[-2e(1) + lw_{1,1}(1)(-2e(2)) + (lw_{1,1}(1))^2(-2e(3))]$$

$$+ p(2)[-2e(2) + lw_{1,1}(1)(-2e(3))] + p(3)[-2e(3)]$$

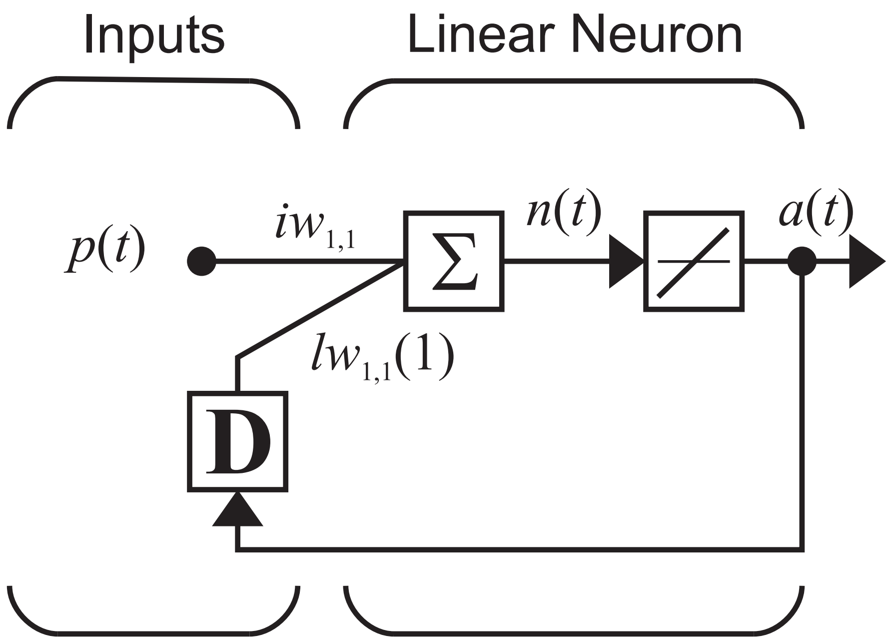
A Problem with Recurrent Network Training

Spurious Valleys

14 Recurrent Net Error Surface Profile



Simple Recurrent Network

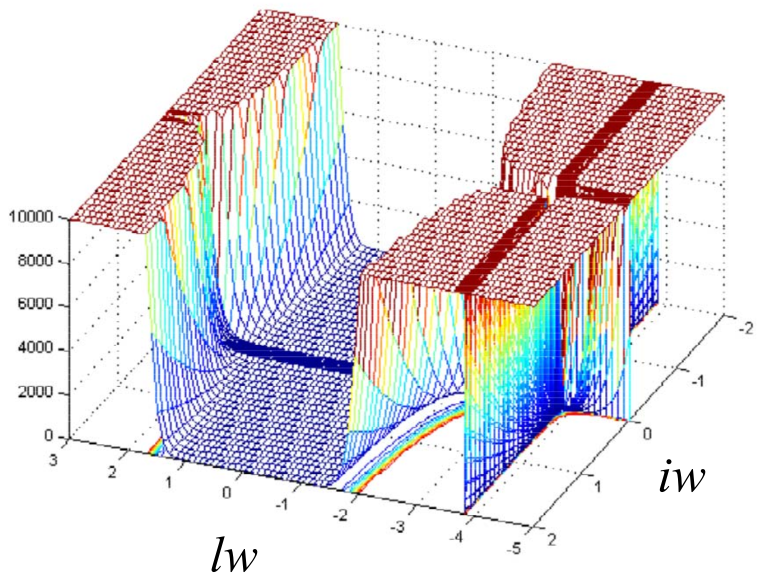


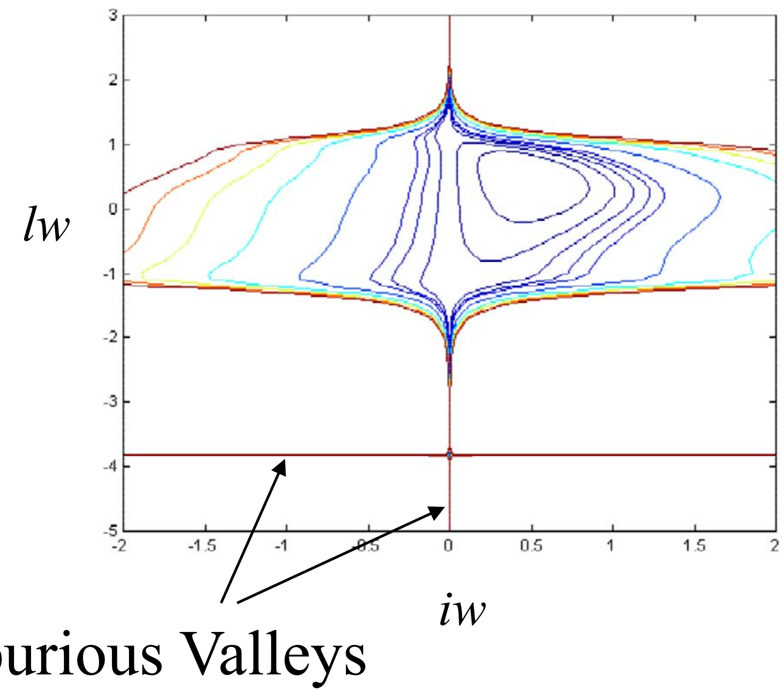
$$a(t) = iw_{1,1}p(t) + lw_{1,1}(1)a(t-1)$$

Mean Square Error Surface

Training data generated with weight values:

$$lw = 0.5 \qquad iw = 0.5$$





14

Network Response

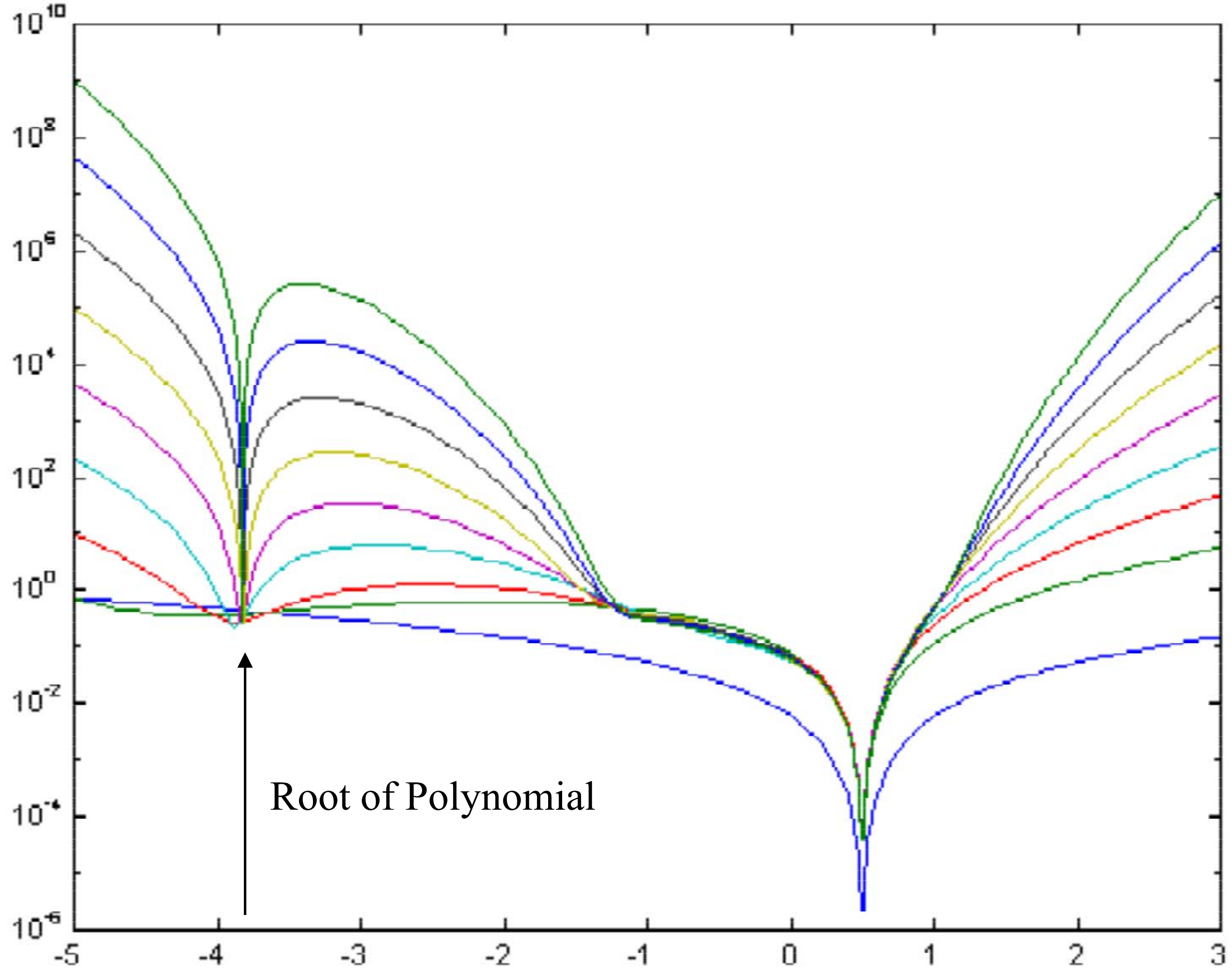
$$a(t) = iw_{1,1}\,p(t) + lw_{1,1}(1)a(t-1)$$

$$a(t) = iw\left\{p(t) + lw\ p(t-1) + (lw)^2\ p(t-2) + \ldots + (lw)^{t-1}\ p(1)\right\} + (lw)^t\, a(0)$$

- The response can be considered a polynomial in *lw*.
- The coefficients of the polynomial involve the input sequence and the initial conditions.
- The network output will be zero at the root of the polynomial.
- Roots greater than 1 (unstable system) tend to remain constant with sequence length (increasing order).

14

Effect of Sequence Length



Cause of Spurious Valleys

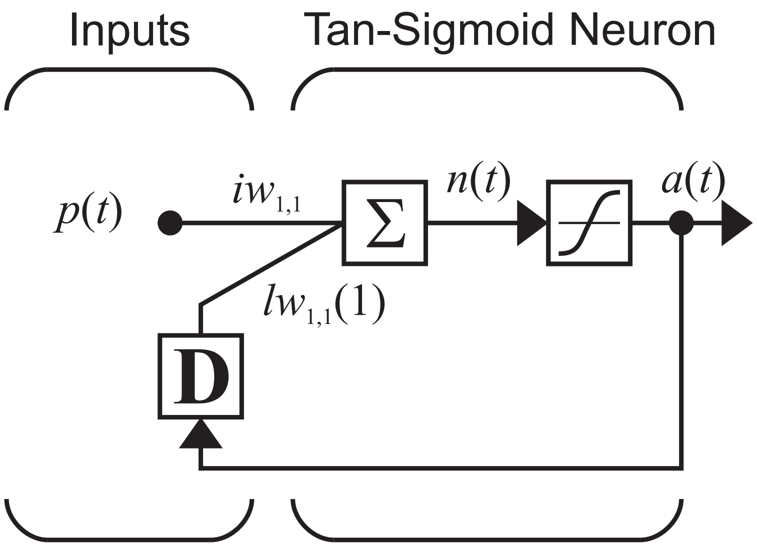
Initial Conditions

If some initial conditions (neuron outputs) are zero, then there are certain combinations of weights that will produce zero outputs for all time.

Input Sequence

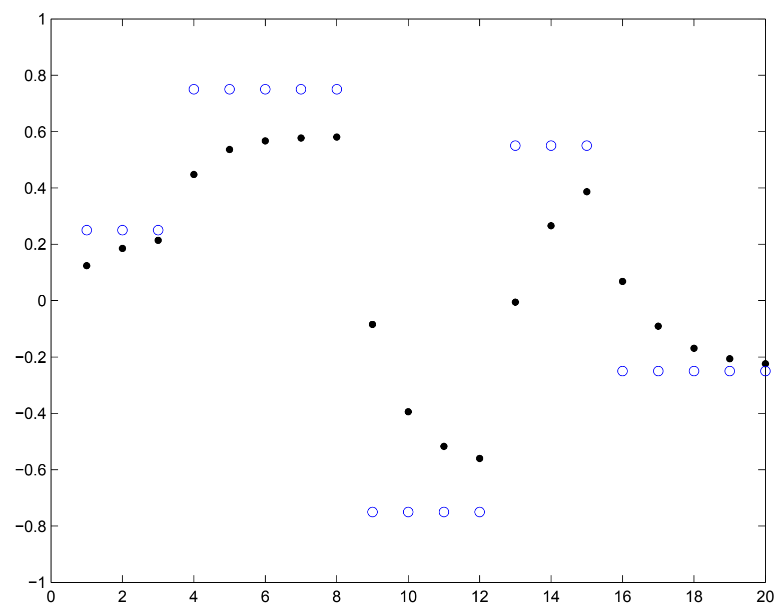
There are values for the weights that produce an unstable network, but for which the output remains small for a particular input sequence. If the input sequence is modified, it may produce a valley in a different location.

Nonlinear Recurrent Network

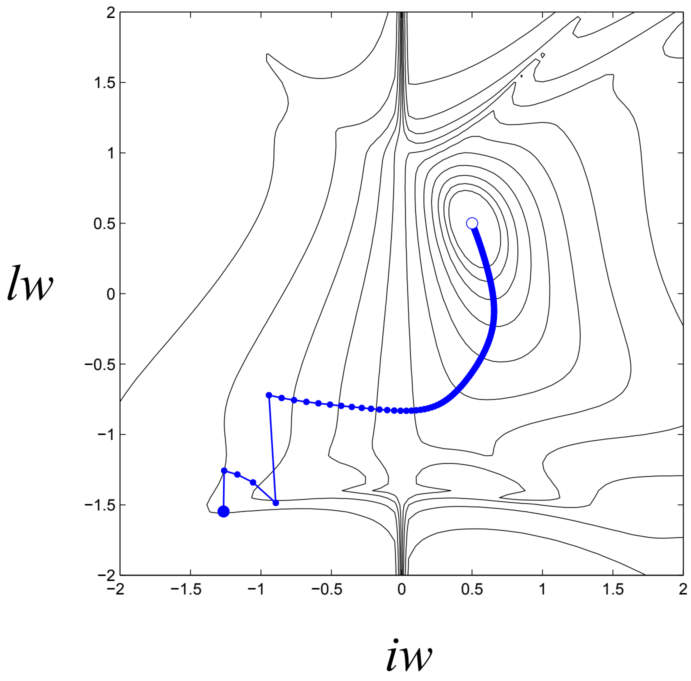


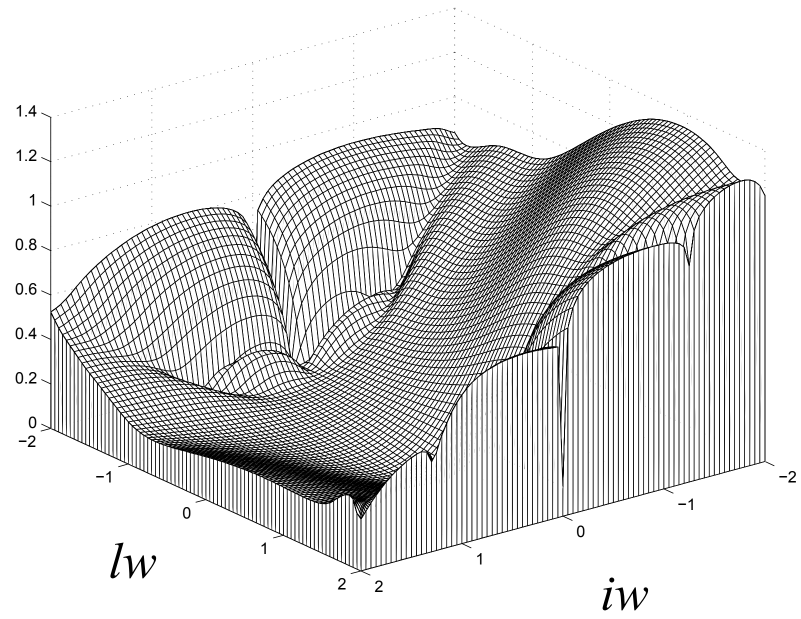
$$a(t) = tansig(\ iw_{1,1} p(t) + lw_{1,1}(1)\, a(t\text{-}1)\ )$$

Training Data (Skyline)

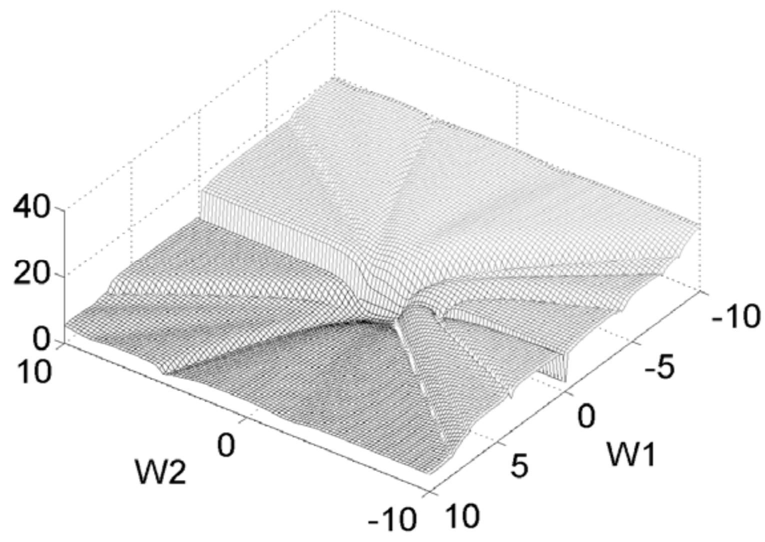


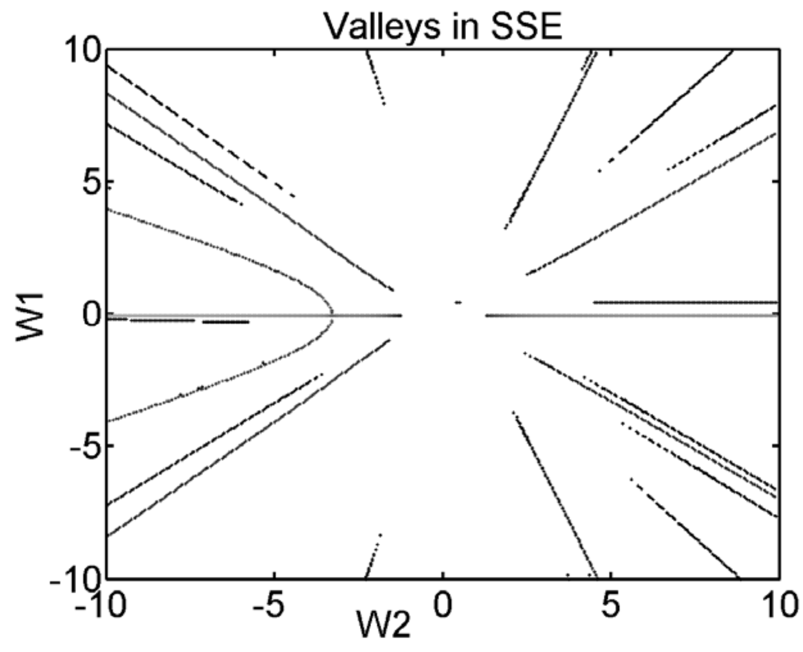
Steepest Descent Trajectory



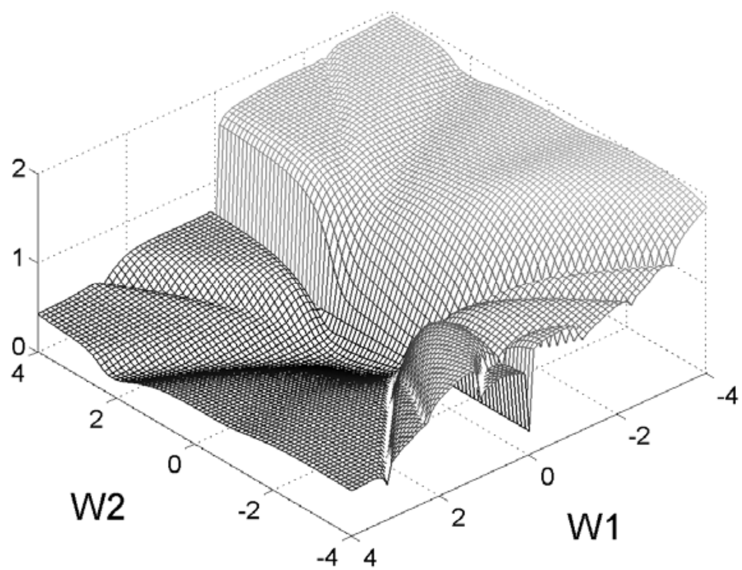


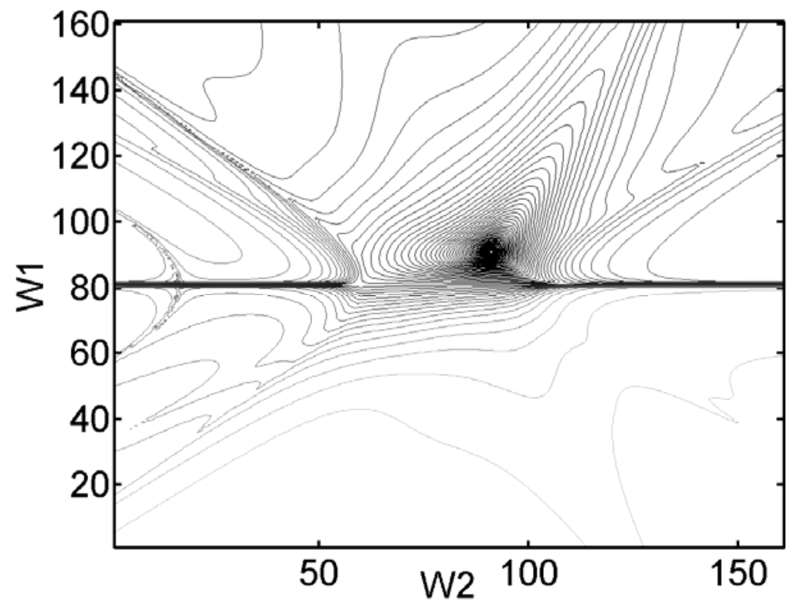
Nonlinear Error Surface



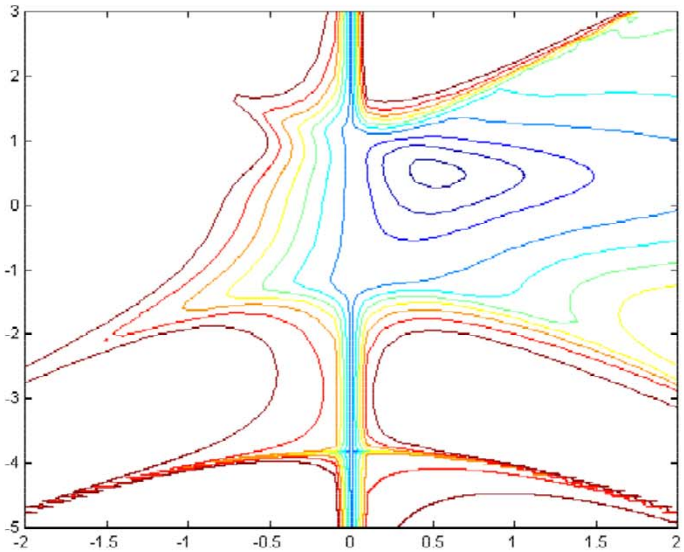


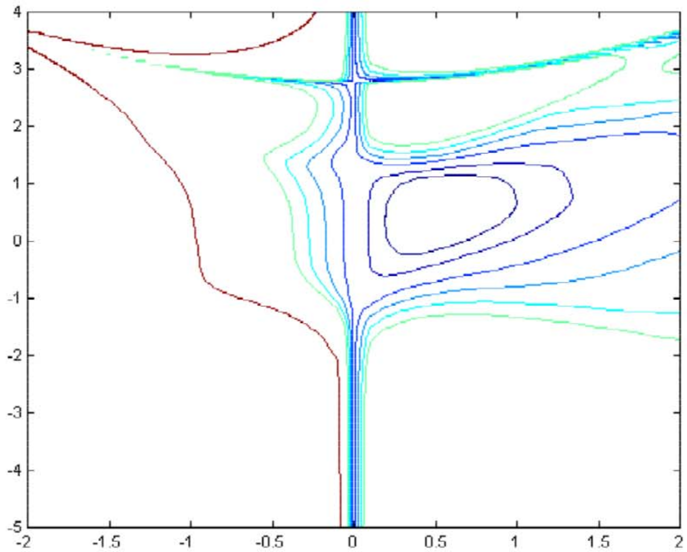
Different Input Sequence



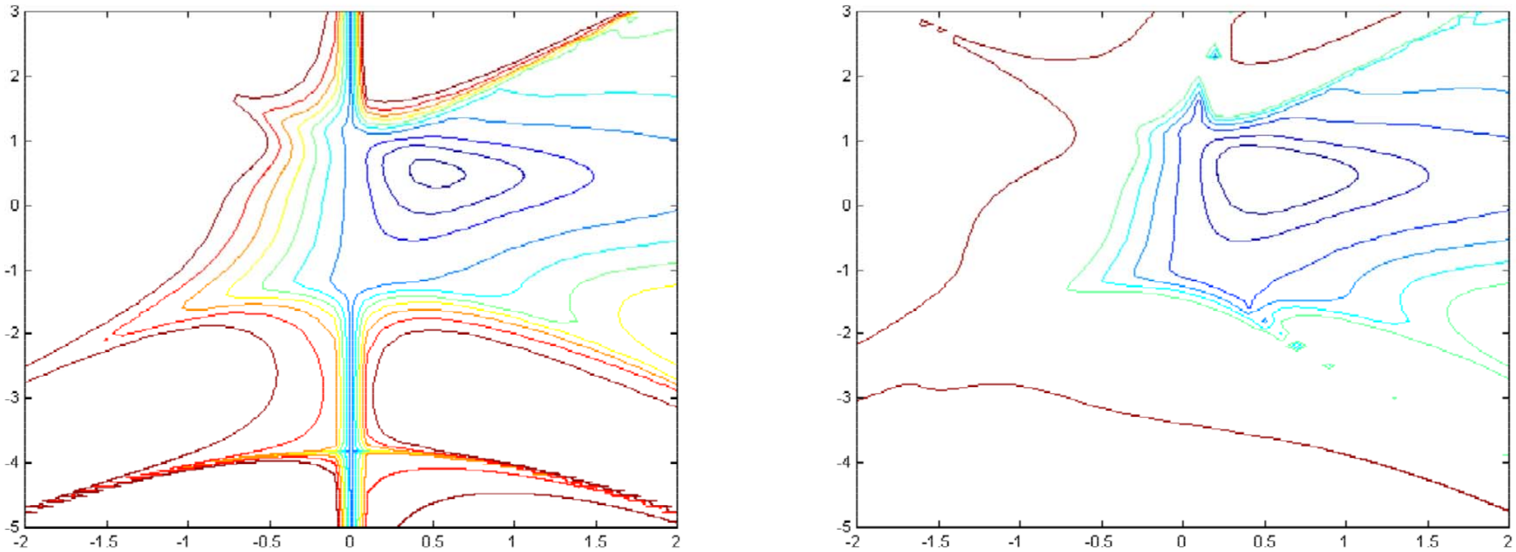


Changing Input Sequence





14 Changing Initial Condition



14 # Procedures for Training Recurrent Nets

- Switch training sequences often during training.
- Use small random initial conditions for neuron outputs and change periodically during training.
- Use a regularized performance index to force weights into stable region. Decay regularization factor during training.

$$J(\mathbf{w}) = SSE + \alpha SSW$$

Summary

- Recurrent networks can be used for a variety of filtering and control applications.
- The gradient calculations for recurrent networks require dynamic backpropagation.
- The error surfaces of recurrent networks have spurious valleys, which require modified training procedures.