

# Learning Sequence Attractors in Recurrent Networks with Hidden Neurons: Supplementary Material

## A Numerical Results for Figure 7 and 8 in the Main Paper

In the main paper, we only showed bar charts (Figure 7 and 8) of the results in Section 6.2. Here, for more information, we provide the numerical results for Figure 7 in Table 1 and Figure 8 in Table 2.

$T$	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
Learning only $\mathbf{V}$	100	100	100	91	66	19	8	2	0	0	0	0	0	0	0
Learning $\mathbf{U}$ and $\mathbf{V}$	100	100	100	100	100	99	88	52	20	1	0	0	0	0	0

Table 1: Successful retrievals out of 100 trials with different sequence period lengths  $T$ .

$M$	100	200	300	400	500	600	700	800	900	1000
Learning only $\mathbf{V}$	0	0	1	3	6	16	14	27	30	37
Learning $\mathbf{U}$ and $\mathbf{V}$	10	52	85	90	94	88	95	96	96	97

Table 2: Successful retrievals out of 100 trials with different numbers of hidden neurons  $M$ .

## B Ablation Experiments: Joint Learning of $\mathbf{U}$ and $\mathbf{V}$

To verify the effective of the proposed learning algorithm in Section 5, we show additional experimental results in which three methods for the recurrent networks of hidden units in learning the sequences in Section 5.3 are compared.

1. Fixing  $\mathbf{U}$  and learning  $\mathbf{V}$  by the temporal asymmetric Hebbian algorithm

$$V_{ji} = \sum_t x_j(t+1)y_i(t)$$

where

$$y_i(t) = \text{sign}\left(\sum_{k=1}^N U_{ik}x_k(t)\right).$$

2. Fixing  $\mathbf{U}$  and learning  $\mathbf{V}$  with the three-factor rule (7)(8)(9) in Section 5.
3. Learning both  $\mathbf{U}$  and  $\mathbf{V}$  with the three-factor rule (4)(5)(6)(7)(8)(9) in Section 5.

The experimental settings are the same as in Section 6.3. The results are shown in Figure 1-5, from which we can see the algorithm proposed in Section 5 is indeed effective.

## C Ablation Experiments: Sparsity

We provide some further ablation study of our algorithm on the effect of sparsity under the experimental settings of Section 6.2 in the main paper.

**Figure 6: Sparse Random Projected Inputs** We compare our method (learning both  $\mathbf{U}$  and  $\mathbf{V}$  with the three-factor rule) with using fixed random  $\mathbf{U}$  whose elements are sampled i.i.d. from the standard Gaussian distribution and learning only with the three-factor rule. The sparse random projected inputs are defined as

$$y_i(t) = \text{sign}\left(\sum_{k=1}^N U_{ik}x_k(t) - \theta\right)$$

where  $\theta > 0$  controls the sparsity level.

**Figure 7: Sparse Random Projected Targets** We test different levels of sparsity in the random projected targets defined as

$$z_i(t=1) = \text{sign}\left(\sum_{k=1}^N P_{ik}x_k(t+1) - \theta\right)$$

where  $\theta > 0$  controls the sparsity level.

From both sets of experiments, we do not find sparsity enlarges significantly the capacity of the networks in learning sequences as attractors.



Figure 1: Silhouette sequence for  $t = 1, \dots, 10$ .



(a) Ground truth



(b) Fixing  $\mathbf{U}$  and learning only  $\mathbf{V}$  with temporal asymmetric Hebbian algorithm



(c) Fixing  $\mathbf{U}$  and learning only  $\mathbf{V}$  with three-factor rule



(d) Learning  $\mathbf{U}$  and  $\mathbf{V}$  with three-factor rule

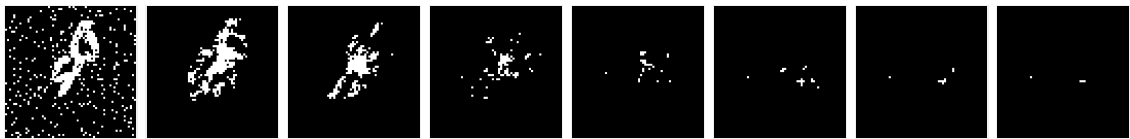
Figure 2: Handwriting sequence 1 for  $t = 1, \dots, 8$ .



(a) Ground truth



(b) Fixing  $\mathbf{U}$  and learning only  $\mathbf{V}$  with temporal asymmetric Hebbian algorithm



(c) Fixing  $\mathbf{U}$  and learning only  $\mathbf{V}$  with three-factor rule



(d) Learning  $\mathbf{U}$  and  $\mathbf{V}$  with three-factor rule

Figure 3: Handwriting sequence 6 for  $t = 1, \dots, 8$ .



(a) Ground truth



(b) Fixing  $\mathbf{U}$  and learning only  $\mathbf{V}$  with temporal asymmetric Hebbian algorithm



(c) Fixing  $\mathbf{U}$  and learning only  $\mathbf{V}$  with three-factor rule



(d) Learning  $\mathbf{U}$  and  $\mathbf{V}$  with three-factor rule

Figure 4: Handwriting sequence 11 for  $t = 1, \dots, 8$ .



(a) Ground truth



(b) Fixing  $\mathbf{U}$  and learning only  $\mathbf{V}$  with temporal asymmetric Hebbian algorithm



(c) Fixing  $\mathbf{U}$  and learning only  $\mathbf{V}$  with three-factor rule



(d) Learning  $\mathbf{U}$  and  $\mathbf{V}$  with three-factor rule

Figure 5: Handwriting sequence 16 for  $t = 1, \dots, 8$ .

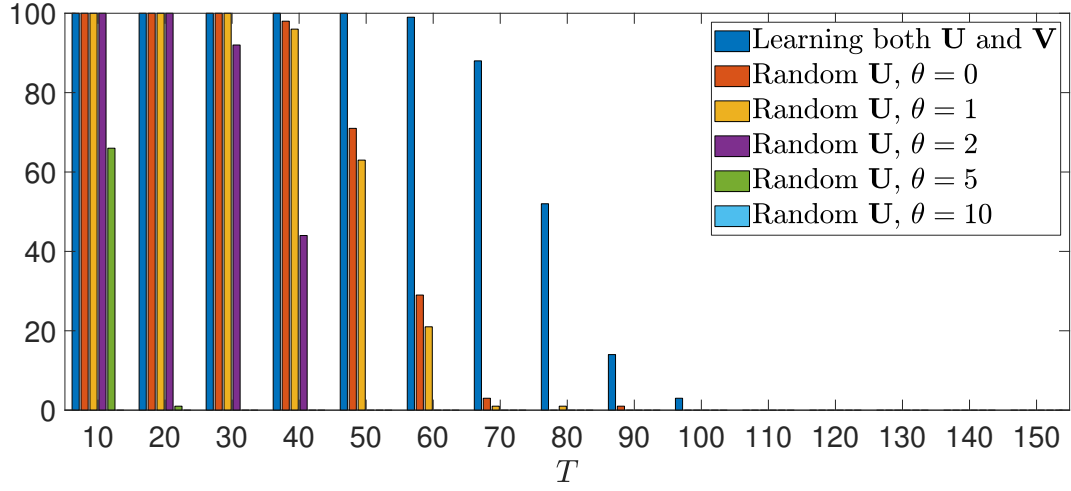


Figure 6: Successful retrievals out of 100 trials with different sequence period lengths.

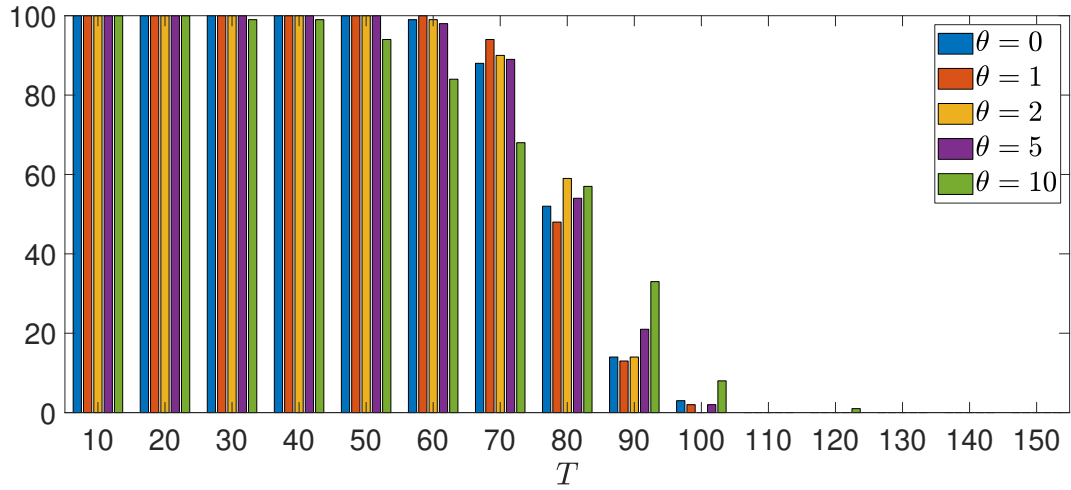


Figure 7: Successful retrievals out of 100 trials with different sequence period lengths