

# 1 Discussion

## Initial model for symmetric RNNs

The previous work from M. Kaschube lab and D. Fitzpatrick lab [?] introduced the feedforward recurrent alignment hypothesis first. Their experimental results compared neural activities in visually naive and visually experienced ferrets' visual cortexes, demonstrating that experience plays a critical role in increasing neural response consistency both across and within trials. Moreover, they found out that the initial evoked responses at eye-opening consist of novel patterns, distinct from spontaneous activity of the endogenous network and that visual experience drives the development of low-dimensional, reliable representations aligned with spontaneous activity.

A computational model was developed with symmetric RNNs to give the proposal that high alignment of feedforward and recurrent networks can effectively amplify the novel activity patterns induced by the onset of visual experience [?]. The modeling results (section ??) support the experimental observations, indicating the potential of the feedforward recurrent alignment hypothesis. Since symmetric RNNs are simple conceptual models for cortical networks, the investigation of broader network structures could improve the modeling and even give new insights into the understanding of underlying mechanisms during experience-driven development.

## Extension modifications for asymmetric RNNs

The closest extension of symmetric RNNs can be the asymmetric RNNs. Without symmetry, the recurrent interactions can generally enable the alignment of feedforward and recurrent networks in a complex plane. Complex patterns are however difficult to interpret in the context of neuroscience. Therefore, modifications to the original model are necessary to overcome this problem. We consider three possible modifications and evaluate them based on the experimental and modeling results from the prior work from M. Kaschube lab and D. Fitzpatrick lab [?]. After the verification, the modification considering the alignment of feedforward and symmetrized network has the best performance.

This indicates that in the case of general asymmetric RNNs, the alignment in the complex plane can be well approximated through symmetrization of the original RNNs. Besides, for full-rank asymmetric RNNs, symmetrization keeps the most of complex-plane information contained when transformed to a real number among all three modifications.

## Test with low-rank RNNs

Beyond the full-rank RNNs, some recent work [?, ?, ?] introduced low-rank interaction structures, which was predicted to capture the actual cortical connectivity more realistically than full-rank RNNs. Thus, we test the modeling on both symmetric and asymmetric RNNs with two possible constructions. The difference between the two constructions is the dispersion from full-rank random noise. Symmetric low-rank RNNs coincide with the experimental observations with both constructions. If applying the best modification from full-rank asymmetric RNNs on low-rank asymmetric RNNs without random noise, the symmetrization leads to a duplication of alignment.

If considering the disturbance of random noise, the dynamics from low-rank RNNs will be overwritten and the results did not differ from full-rank RNNs. For asymmetric low-rank RNNs without random noise, other modifications would be necessary to guarantee the functionality of the model on low-rank RNNs in general.

## Modification with white-noise evoked activity for black box RNNs

Considering the difficult accessibility of whole cortical interactions during the experimental environment, we try to modify the modeling of the feedforward recurrent alignment hypothesis with white-noise-evoked activity as feedforward input to align with the recurrent network. It turns out, that the experimental results can be roughly reflected. Besides, we test out the influence of repeatedly applying prior response activity as input on feedforward recurrent alignment. The dynamic of alignment indicates that the reuse of response as feedforward input to the recurrent network can the response to a stable oscillation.

Therefore, white-noise-evoked activity can be a good candidate for the design of experimental methods to verify the feedforward recurrent alignment hypothesis. The underlying mechanism for stable oscillation when reusing prior response activity still needs to be analytically and experimentally explored. It can give new insight into the feedforward recurrent alignment hypothesis and more understanding of feedforward recurrent network structure.

## Modeling under dynamic feedforward network with Hebbian learning

Since the development of the ferrets' visual cortex was an experience-driven development, plasticity, and learning mechanisms cannot be ignored. For simplicity, start modeling the feedforward recurrent alignment with a dynamic feedforward network. The classic Hebbian learning rule was considered. The feedforward recurrent alignment increases over time until the steady state.

This indicates that during the learning, the feedforward network adapts itself to unsupervised learning such that its outputs align better with the recurrent network. According to the feedforward recurrent alignment hypothesis, this could imply that the responses become robust during learning. Therefore, modeling with plasticity and learning rules can be a promising direction to explore the underlying mechanism for experimental observations of ferrets' visual cortex.

## 2 Outlook

### Possible constructions of feedforward recurrent alignment that does not be influenced by complex eigenvectors

In the case of general asymmetric RNNs, the modifications of the original model emerge from the problem that the alignment is in the complex plane. The modifications considered in this work focus on turning the aligned feedforward input from the complex plane into the real-number plane, while possibly keeping the original feedforward recurrent alignment definition from [?].

Another possible way for adapting the modeling of the feedforward recurrent alignment hypothesis could be a new construction of alignment score. The new construction could even allow the feedforward recurrent input to align in the complex plane resulting in the final score being a real number. For example, in a form similar to

$$\nu := (Jh)^*(Jh) = h^*J^*Jh = h^*JJh, \quad (2.1)$$

where  $h$  is the feedforward input and  $J$  the recurrent interaction matrix. The operator  $*$  denotes the conjugate transformed of vectors or matrices. The last equation is followed by the fact that interactions between neurons are represented by real numbers.

With the new construction eq.(2.1), for both symmetric RNNs and asymmetric RNNs, the feedforward recurrent alignment is a real number when feedforward inputs are aligned to eigenvectors of the interaction matrix. However, the monotony between eigenvalues and alignment score cannot be kept. Although this construction is not a candidate, it could provide some ideas for new construction of feedforward recurrent alignment to overcome the problems and bring new insights to the feedforward recurrent alignment hypothesis in further studies.

### Learning in recurrent network and learning rules with normalization

The simple tryout with only basic Hebbian learning in feedforward recurrent delivers the results of increased feedforward recurrent alignment during unsupervised learning. A next possible step to verify the feedforward recurrent hypothesis combined with plasticity could be also embedding learning in a recurrent network. This could

help to understand the influence of recurrent dynamics on alignment and bring a closer understanding of how feedforward and recurrent networks cooperate to generate reliable responses.

Besides, the basic Hebbian rule is unstable and allows unbounded growth of feedforward network interaction strengths [?]. Therefore, further studies can consider normalization rules to limit the growth of network interactions, especially when considering more multiple input neurons. Moreover, other rules developed based on Hebbian rules can also be implemented to explore the performance of different rules in the model.

### **Interaction structure considering excitatory and inhibitory neuron populations**

A further important perspective that influences the activity dynamics and feedforward recurrent alignment is neuron population. Neurons are typically classified as either excitatory or inhibitory, meaning that they have either excitatory or inhibitory effects on all of their postsynaptic targets [?]. Excitatory-inhibitory networks with local heterogeneity were recently found to raise long-range neuron correlations and global organizations [?]. Moreover, a recent work [?] proposed that synapse-type-specific plasticity enables the joint development of stimulus selectivity and excitatory-inhibitory balance. Their model allows even implementation combining excitatory-inhibitory networks and Hebbian learning. Expanding the feedforward recurrent alignment hypothesis with excitatory and inhibitory neuron populations can be a potential focus for further studies.