1 Methods

In this chapter, we will give an overview about the recurrent network (RNN) models for exploration of the feedforward recurrent alignment hypothesis that are evolved in this work. The firstly introduced symmetric network model builds the basis for modifications and extensions in other further models. The modified models will be introduced subsequently. Finally, we consider the role of learning could play in the feedforward recurrent alignment hypothesis.

1.1 Symmetric Recurrent Network Model

Due to the well understood mathematical characters of symmetric RNNs, they are often applied in models for neuroscience for a better understanding of certain dynamics. Therefore, we firstly consider the basic case of having a symmetric recurrent network, which has the symmetric interaction matrix. For symmetric RNNs, if there is a connection between two neurons n_i and n_j, the strength of the directed connection from the neuron n_i to the neuron n_j equals the directed connection from n_j to n_i.

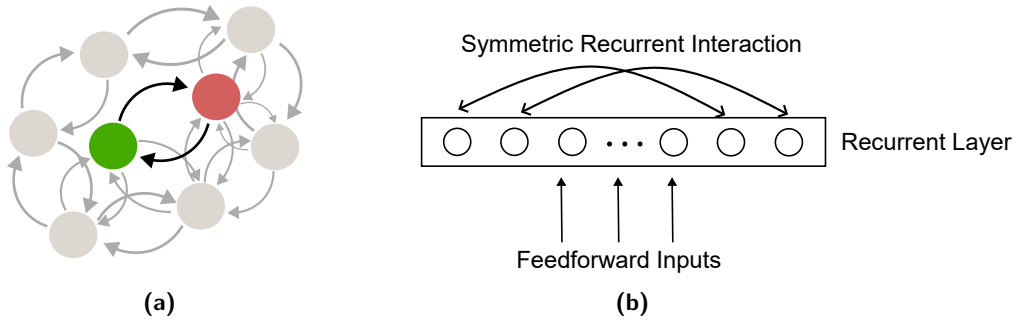

Figure 1.1 Illustration of symmetric recurrent networks (symmetric RNNs). **(a)** An example of symmetric connections between two neurons in a network with multiple neurons. If there are connections between two neurons, here for example the green and red neurons, the directed connection from the green neuron to the red neuron has the same strength as the directed connection from the red to the green. **(b)** Structure of a symmetric RNN with feedforward inputs as the inputs for the recurrent layer. The connections between neurons inside the recurrent layer are symmetric, as illustrated in figure (a).

1.1.1 Symmetric Recurrent Interaction

In the model, we consider a full rank real symmetric recurrent interaction matrix J with Gaussian distributed entries with mean 0 and variance 1,

$$J_{ij} \sim \mathcal{N}(0, 1). \tag{1.1}$$

Besides, J has full rank equals the number of neurons,

$$\text{rank}(J) = n\,, \tag{1.2}$$

where n is the number of neurons involved in the RNN. The eigenvalues $\{\lambda_i\}_{i=1,...,n}$ of J are limited by parameter $R < 1$ through

$$\lambda_i = \frac{R\tilde{\lambda}_i}{\tilde{\lambda}_{\max}} \quad \forall i\,, \tag{1.3}$$

$\tilde{\lambda}_i$ are the original eigenvalues of J. As a result, the maximal eigenvalues after the re-scaling would take value $R < 1$.

1.1.2 Response Steady State

Existence of Steady State When considering the relationship between firing rate and synaptic current as linear, the dynamic system of the RNN illustrated in 1.1 could be described as [**?**]:

$$\tau_r \frac{\mathrm{d}r}{\mathrm{d}t} = -r + J \cdot r + h \overset{\tau_r=1}{\Rightarrow} \frac{\mathrm{d}r}{\mathrm{d}t} = -r + J \cdot r + h\,, \tag{1.4}$$

with the vector $r \in \mathbb{R}^{n\times 1}$ describing responses of neurons in the recurrent layer, the vector $h \in \mathbb{R}^{n\times 1}$ as feedforward inputs ,and τ_r the time constant controlling the speed of dynamic. The steady state of the dynamic system 1.4 can be received by setting the ordinary differential equation to zero. For simplicity, the time constant is set to one. We then have

$$\frac{\mathrm{d}r}{\mathrm{d}t} = -r + J \cdot r + h = 0 \Rightarrow r = (I_n - J)^{-1} \cdot h =: r^*\,, \tag{1.5}$$

r^* the steady state for responses. $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix. Since J is full rank, the matrix $(I_n - J)$ is invertible. Therefore, the steady state exists.

Stability of Steady State The dynamic 1.4 could also be written in an elementary expression:

$$f_i(r_1, ..., r_n) := \frac{\mathrm{d}r_i}{\mathrm{d}t} = -r_i + \sum_{j=1}^{n} J_{ij} r_j + h_i \text{ for } i = 1, ..., n\,. \tag{1.6}$$

The derivative of f_i to r_j is

$$\frac{\partial f_i}{\partial r_j} = \begin{cases} -1 + J_{ij} & \text{if } i = j \\ J_{ij} & \text{if } i \neq j \end{cases}\,. \tag{1.7}$$

The Jacobian matrix A of the dynamic system $\frac{\mathrm{d}r}{\mathrm{d}t}$ is then

$$A := \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} = -I_n + J\,. \tag{1.8}$$

Therefore, the Jacobian matrix A is a linear transformation of the symmetric recurrent interaction matrix J, which is independent of the steady state response. So, A has the same set of eigenvectors [1] as J. With $E := \{e_i\}_{i=1,\dots,n}$ the matrix containing eigenvectors of J column-wise,

$$(-I_n + J)E = -I_n \cdot E + J \cdot E = -I_n \cdot E + \Lambda \cdot E = (-I_n + \Lambda)E\,, \tag{1.9}$$

Λ the diagonal matrix with eigenvalues $\{\lambda_i\}_{i=1,\dots,n}$ of J on its diagonal. This means, $\{-1 + \lambda_i\}_{i=1,\dots,n}$ are eigenvalues for the Jacobian matrix A.

The eigenvalues of the Jacobian matrix A determines the stability of steady states. Here, since the matrix A is symmetric, all its eigenvalues $-1 + \lambda_i, i = 1, ..., n$ are from $\mathbb{R}$. Because the eigenvalues λ_i of matrix J is limited by the parameter $R < 1$, defined in 1.3, we have

$$-1 + \lambda_i \overset{(1.3)}{<} -1 + 1 = 0\,. \tag{1.10}$$

That is, all eigenvalues of the Jacobian matrix A are negative. This indicates that the steady state r^* is stable. Under the assumption that the system reaches its steady state quick enough, we could apply the steady state r^* for further analysis.

1.1.3 Feedforward Recurrent Alignment for Symmetric Interactions

Generally, the feedforward inputs can be considered as a firing rate distribution with certain mean value. The mean firing rate is essential for the strength of inputs. We therefore consider mainly the mean firing rate of inputs. For the rest of work, if mentioning feedforward inputs without further definition, we mean the mean firing rate of inputs.

The alignment of a feedforward input $h \in \mathbb{R}^{n\times 1}$ with the recurrent network J is defined as [**?**]

$$\nu := \frac{h^T J h}{\| h \|_2^2} \tag{1.11}$$

If the inputs are aligned to the eigenvectors e_i of the recurrent interaction J, i.e.,

$$h \propto e_i\,, \tag{1.12}$$

[1] For a symmetric matrix, the set of left eigenvectors equal the set of right eigenvectors

the feedforward recurrent alignment ν is proportional to the eigenvalues λ_i, because inserting the proportionality (1.12) in (1.11) leads to

$$\nu = \frac{h^T J h}{\parallel h \parallel_2^2} \propto \frac{{e_i}^T J e_i}{\parallel e_i \parallel_2^2} = \frac{\lambda_i {e_i}^T e_i}{\parallel e_i \parallel_2^2} = \lambda_i \,. \tag{1.13}$$

It was therefore observed that the maximal alignment was attained when the input was proportional to the eigenvector $e_{\max}$ with maximal eigenvalue $\lambda_{\max}$ [**?**].

1.1.4 Response Properties for Evaluation

Trial-to-trial correlation Given the feedforward inputs that are from the same distribution for multiple trials, the correlation between responses from different trials indicates the reliability of the responses. Large correlation implies high reliability of the response generated by the RNN.

Model the inputs $h \in \mathbb{R}^{n\times 1}$ as multivariate normal distributions with mean vector $\mu \in \mathbb{R}^{n\times 1}$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$

$$h \sim \mathcal{N}(\mu, \Sigma) \;\text{ with } \Sigma := \sigma_{\text{trial}} I_n. \tag{1.14}$$

Then, the steady state response $r^* = (I_n - J)^{-1} \cdot h$ from (1.5) has the linearly transformed normal distribution

$$r^* \sim \mathcal{N}\left((I_n - J)^{-1}\mu, (I_n - J)^{-1}\Sigma(I_n - J)^{-T}\right) , \tag{1.15}$$

where the mean vector and covariance matrix are linearly transformed. The property could be proved analogously as in [**?**] with the moment-generating function of the multivariate normal distribution

$$M_h(t) = \mathbb{E}\left[\exp(t^T h)\right] = \exp\left[t^T\mu + \frac{1}{2}t^T\Sigma t\right] . \tag{1.16}$$

Therefore, the moment-generating function of the random vector r^* becomes

$$\begin{aligned} M_{r^*} &= M_h\left((I_n - J)^{-T}t\right) \\ &= \exp\left[t^T\left((I_n - J)^{-1}\mu\right) + \frac{1}{2}t^T(I_n - J)^{-1}\Sigma(I_n - J)^{-T}t\right] , \end{aligned} \tag{1.17}$$

which indicates the linearly transformed distribution of r^* as in (1.15).

As calculated in [**?**], the trial to trial correlation β for one stimulus s is calculated by taking the mean of correlations between N response trials that evoked by this stimulus. That is

$$\beta_s = \frac{2}{N(N-1)} \sum_{i=1, j=i+1}^{N} corr(r_i^s, r_j^s) \,, \tag{1.18}$$

where r_i^s is the i-th response trial that evoked by stimulus s.

Intra-trial stability It was observed that presenting ongoing visual grating stimuli, the responses in the visually naive cortex has a stronger variation than they are after visual experience. In order to reflect the variation of responses during the stimulation period, the quantity of "intra-trial stability" was defined [**?**].

To model the time dependent input $h(t) \in \mathbb{R}^{n\times 1}$ distributed as (1.14) and its evoked steady state responses $r(t) \in \mathbb{R}^{n\times 1}$, the following stochastic differential equations were formulated

$$\mathrm{d}h = \mu \mathrm{dt} + \sigma_{\text{time}} \mathrm{d}W \tag{1.19a}$$

$$\mathrm{d}r = (-r + J \cdot \mu)\mathrm{d}t + \sigma_{\text{time}} \mathrm{d}W\,, \tag{1.19b}$$

with W the Wiener process, which is a continuous-time stochastic process with independent Gaussian increments.

To approximate the evoked response $r(t)$, the equation (1.19b) is solved numerically with Euler-Maruyama scheme

$$r_{t+1} = r_t + (-r_t + J \cdot \mu)\Delta t + \sigma_{\text{time}} \sqrt{\Delta t} \Delta \tilde{W}_t\,, \tag{1.20}$$

with r_t the response at time point t, Δt the step width for iteration, and $\Delta \tilde{W}_t \in \mathbb{R}^{n\times 1}$ the Gaussian increment at time point t defined as the multivariate normal distribution with mean vector 0_v and covariance matrix I_n

$$\Delta \tilde{W}_t \sim \mathcal{N}(0_v, I_n)\,. \tag{1.21}$$

For a certain step width $\tilde{\Delta} t$, the intra-trial stability $c(\Delta \tilde{t})$ was defined by the correlation between normalized response at time t and its delayed response at time $t + \Delta \tilde{t}$

$$c(t, \Delta \tilde{t}) := \bar{r}(t)^T \bar{r}(t + \Delta t)\,, \tag{1.22}$$

where the normalized response is defined as

$$\bar{r}(t) := \frac{r - \langle r \rangle}{\sigma_r}\,, \tag{1.23}$$

with mean value of r denoted by $\langle r \rangle$ and standard deviation by σ_r.

The final intra-trial stability for a time period T is the time-averaged value over all time points $0 \leq t \leq T - \Delta \tilde{t}$

$$\begin{aligned} \bar{c}(\Delta \tilde{t}) &= \frac{1}{T - \Delta \tilde{t}} \int_0^{T - \Delta \tilde{t}} c(t, \Delta \tilde{t}) \mathrm{d}t \\ &= \frac{1}{T - \Delta \tilde{t}} \int_0^{T - \Delta \tilde{t}} \bar{r}(t)^T \bar{r}(t + \Delta t) \mathrm{d}t\,. \end{aligned} \tag{1.24}$$

Dimensionality The dimensionality of neuron responses reflect the complexity of the informatbetaion they encoded. A diverse response pattern corresponds with a broader distribution of the variance over principal components, leading to a higher-dimensional linear manifold. The corresponding dimensionality for more diverse and variable response pattern is therefore higher [**?**]. Given the multivariate normal distributed inputs $h \in \mathbb{R}^{n\times 1}$

$$h \sim \mathcal{N}(0_v, \Sigma^{\mathrm{Dim}})\,, \tag{1.25}$$

the linear transformed responses (analogously as (1.15)) are

$$r \sim \mathcal{N}(0_v, (I_n - J)^{-1}\Sigma^{\mathrm{Dim}}(I_n - J)^{-T}) \tag{1.26}$$

with

$$\Sigma^{\mathrm{Dim}} := \sum_{i=L}^{L+M} \exp\left(\frac{-2(i-L)}{\beta}\right) e_i e_i^T\,, \tag{1.27}$$

in which the parameter $M := \kappa\beta$ and β reflects the dimensionality [**?**] and κ for determining the number of directions e_i that contribute to the dimensionality. Since the eigenvectors of J build a set of basis for $\mathbb{R}^n$, they could be chosen as e_i for (1.27). Hereby, the eigenvectors are ordered according to their eigenvalues in descending order. The exponential factor in (1.27) simulates the exponential decay of variance ratio observed in prior data [**?**].

The linear effective dimensionality based on participation ratio was defined to quantify the tendency of dimensionality during visual maturation. The participation ratio is defined as

$$d_{\mathrm{eff}} := \frac{(\sum_i \lambda_i)^2}{(\sum_i \lambda_i^2)}\,, \tag{1.28}$$

where λ_i the eigenvalues of a certain response pattern with covariance $\mathbf{\Sigma}$. Since as defined in (1.27), Σ^{Dim} has the same eigenvectors (aka. principal components)as J. Therefore, the eigenvalues λ_i^{Dim}(aka. variance ratio or variance explained) for Σ^{Dim} are transformed eigenvalues λ_i of J expressed as

$$\lambda_i^{\mathrm{Dim}} = \exp\left(\frac{-2(i-L)}{\beta}\right)\lambda_i\,. \tag{1.29}$$

The covariance of the responses share the same eigenvectors as Σ^{Dim} based on its distribution (1.26) and therefore also the same as J. The eigenvalues (aka. variance ratio) for the responses could be constructed analogously with

$$\lambda_i^{\mathrm{Act}} = \exp\left(\frac{-2(i-L)}{\beta}\right)\frac{1}{(1-\lambda_i)^2}\,, \tag{1.30}$$

for $i = L, ..., L + M$, due to the inverse transformation.

Insert the eigenvalues of responses (1.30) in the formula for effective dimensionality to get the final formulation of dimensionality for responses

$$d_{\text{eff}}^{r} = \frac{\left(\sum_{i=L}^{L+M} \exp\left(-2\frac{i-L}{\beta}\right)(1-\lambda_i)^{-2}\right)^2}{\sum_{i=L}^{L+M} \exp\left(-4\frac{i-L}{\beta}\right)(1-\lambda_i)^{-4}} \tag{1.31}$$

Since the vector of explained variance ratios in the principal component analysis (PCA) is the normalized vector containing eigenvales of the covariance matrix rescaled by the largest eigenvalue in descending order, which then explains how much variance does the corresponding principal component contribute. Therefore, another way to access the dimensionality is to empirically determine the explained ratio of generated data samples through PCA and insert the variance ratio into the definition of effective dimentionality, i.e.,

$$d_{\text{eff}} = \frac{\left(\sum_{i=L}^{L+M} \text{var}_i\right)^2}{\sum_{i=L}^{L+M} \text{var}_i^2} \tag{1.32}$$

with var_i the i-th variance ratio.

Alignment with spontaneous activity The alignment of activity patterns to spontaneous activity reflects in principle the size of overlaps between activity patterns and spontaneous activity pattern. Assuming having the evoked response pattern as R and the spontaneous activity pattern as S, the projection of R to S could be quantified as the covariance matrix of R explained by the principal components (a.k.a. eigenvectors of covariance matrix) of S, which results a variance ratio vector

$$\mathbf{v}_i = \frac{\mathbf{p}_{i,S}^T \cdot \mathbf{\Sigma_R} \cdot \mathbf{p}_{i,S}}{\text{Tr}(\mathbf{\Sigma_R})}\,, \tag{1.33}$$

for $i = 1, ..., n$. $\mathbf{p}_{i,S}$ are the principal component of spontaneous activity and $\mathbf{\Sigma_R}$ the covariance matrix of evoked activity. Consider the projection of activity pattern R in all directions of spontaneous pattern S together to reflect the overall overlaps between two patterns, we calculate firstly the alignment between a response trial $r_{i,R}$ form R to the spontaneous pattern S

$$\gamma_i = \frac{r_{i,R}^T \cdot \Sigma_S \cdot r_{i,R}}{\|r_{i,R}\|^2 \text{Tr}(\Sigma_S)}\,, \tag{1.34}$$

where Σ_S is the covariance of pattern S. The final alignment between S and R is then the average value of alignment between S and all trials of R.

To model the inputs and responses, we assumed that spontaneous activity was evoked by inputs from broad sources. Besides, since the spontaneous activity already

exists almost a week before eye opening, we assume that they already fit to the activity space generated by recurrent network [**?**]. Therefore, the inputs would be explained by more directions (or eigenvectors) than stimuli evoked responses as modeled before with (1.25), that is higher dimensionality. Since the parameter β in (1.25) indicates the dimensionality, we could set for spontaneous activity higher $\beta_{\text{spont}} > \beta$ to generate high dimensional inputs. Therefor, we than have the broad inputs $h^{\text{spont}} \in \mathbb{R}^{n\times 1}$ and spontaneous activity $r^{\text{spont}} \in \mathbb{R}^n$, which are multivariate distributed vectors

$$\mathbf{h}^{\text{spont}} \sim \mathcal{N}(\mathbf{0_v}, \mathbf{\Sigma}^{\text{spont}}) \tag{1.35}$$

and

$$\mathbf{r}^{\text{spont}} \sim \mathcal{N}\left(\mathbf{0_v}, (\mathbf{I_n} - \mathbf{J})^{-1}\mathbf{\Sigma}^{\text{spont}}(\mathbf{I_n} - \mathbf{J})^{-T}\right) . \tag{1.36}$$

The covariance matrix $\mathbf{\Sigma}^{\text{spont}}$ is constructed in the same way as $\mathbf{\Sigma}^{\text{Dim}}$ only with $L = 1$ and larger β_{spont}, that is

$$\mathbf{\Sigma}^{\text{spont}} := \sum_{i=1}^{\kappa\beta_{\text{spont}}} \exp\left(\frac{-2(i-1)}{\beta_{\text{spont}}}\right) e_i e_i^T . \tag{1.37}$$

1.2 Asymmetric Recurrent Network Model

Different as symmetric recurrent networks, asymmetric recurrent networks do not necessary have symmetric interaction strength between two neurons. That is, if we have two neurons n_i and n_j, the interaction strength from n_i to n_j can generally differ from the interaction strength from n_j to n_i. Therefore, the network has a more complex neural interactions and dynamics. Compare to symmetric recurrent network, the asymmetric recurrent network reflects more general neural interactions and thus more biologically realistic.

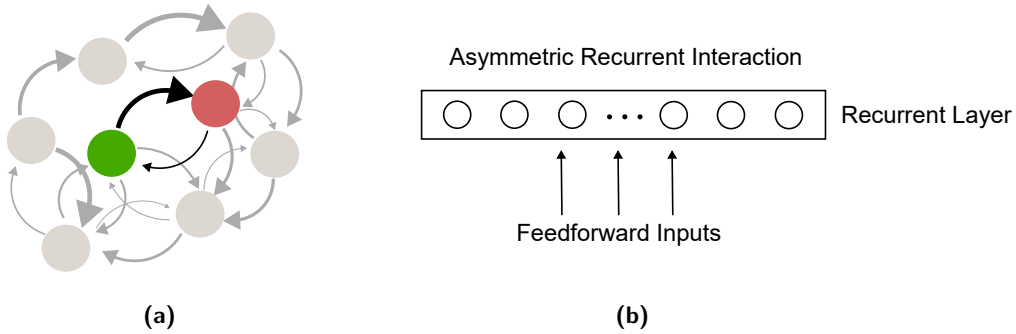

Figure 1.2 Illustration of asymmetric recurrent networks (asymmetric RNNs). In general, asymmetric recurrent networks do not have symmetric interaction strength between to neurons. **(a)** An example of asymmetric connections between two neurons in a network with multiple neurons. There are connections between the green and red neurons. The connection from green one to red one is stronger than the connection from the red to green. There are also connections that only from one neuron to the other but no connection back from the other neuron. **(b)** Structure of a asymmetric RNN with feedforward inputs as the inputs for the recurrent layer. The connections between neurons inside the recurrent layer are asymmetric, as illustrated in figure (a).

1.2.1 Asymmetric Recurrent Interaction

For the modeling, we construct the asymmetric interaction matrix J through disturbing symmetric interaction matrix by a general asymmetric random matrix

$$J = aJ_{\text{sym}} + (1-a)J_{\text{asym}}\,. \tag{1.38}$$

The symmetric part is generated as described in section 1.1.1. The asymmetric part has also Gaussian distributed entries with mean 0 and 1 as variance (1.1). Parameter a therefore indicates the degree of symmetry in the network.

The recurrent network dynamics remains the same as in symmetric case, described by (1.4). Therefore, the steady state also keeps its form as (1.5). Stability analysis of the steady state follows the same procedure as for symmetric interaction

matrix (section 1.1.2). However, here we have to mind that the asymmetric interaction matrix J now have left and right eigenvectors, which are not identical to each other any more. The Jacobian matrix $A = -I_n + J$ defined from (1.8) still has the same set of eigenvectors as J. The eigenvalues are found in the similar way as (1.9) but with left and right eigenvectors. If E_l and E_r the matrices containing the left and right eigenvectors column-wise, it follows

$$\begin{aligned} E_l^*(-I_n + J) &= -E_l^* + E_l^* J = -E_l^* + \Lambda E_l^* = (-I_n + \Lambda)E_l^* \\ (-I_n + J)E_r &= -E_r + JE_r = -E_r + E_r\Lambda = E_r(-I_n + \Lambda)\,, \end{aligned} \tag{1.39}$$

with E_l^* the conjugate transport matrix of E_l, I_n the identity matrix, and Λ the diagonal matrix that contains eigenvalues of J on its diagonal. The eigenvalues $\{\lambda_i\}$ for J are in general complex numbers

$$\{\lambda_i \in \mathbb{C} \mid \lambda_i \text{ eigenvalues of } J\}\,. \tag{1.40}$$

Therefore, the eigenvalues for the Jacobian matrix A is now $-I_n + \Lambda \in \mathbb{C}^{n\times n}$.

To determine the stability of steady state, we now have to consider the real part of $(-I_n + \Lambda)$. If $\exists i : -1 + \mathrm{Re}(\lambda_i) > 0$, the steady state will be unstable. However, if $\forall i : -1 + \mathrm{Re}(\lambda_i) < 0$, the steady state is then stable and we will have

$$\forall i : -1 + \mathrm{Re}(\lambda_i) < 0 \;\Rightarrow\; \lim_{t\to\infty} r(t) = r^* \tag{1.41}$$

Thus, we need to limit the range of eigenvalues for J such that the real part of eigenvalues is limited by $R < 1$. The limitation could be achieved by dividing the complex eigenvalues by maximal absolute magnitude.

$$\tilde{\lambda}_i = \frac{R\lambda_i}{\|\lambda\|_{\max}} \quad \forall i\,, \tag{1.42}$$

The contribution of eigenvalues is influenced by the parameter a, which determines the proportion of symmetric interaction. In the case of only having symmetric interaction network, all eigenvalues are real and therefore the distribution forms a line in the complex plane. On the other hand, if J is only a general asymmetric interaction network, the distribution forms a circle [**?**]. For a between 0 and 1, the distribution is a symmetric ellipse.

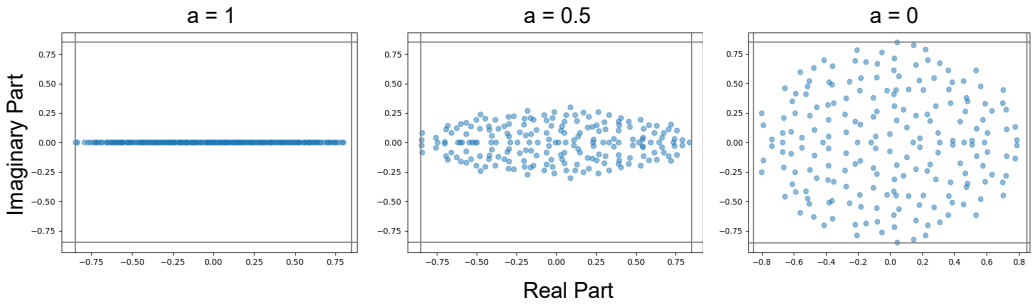

Figure 1.3 Eigenvalue distribution in dependence of parameter a in (1.38). In general, an asymmetric matrix has eigenvalues in complex plane. The degree of symmetry determines the form of distribution from a line ($a = 1$, only real eigenvalues) to a full circle ($a = 0$) with radius $R < 1$. Between $a = 0$ and $a = 1$, the distribution is a symmetric ellipse along imaginary part $= 0$. Sub figures from left to right show the eigenvalue distribution in the complex plane from $a = 1, 0.5$, and 1. The gray line marks the radius $-R$ and R. Here $R = 0.85 < 1$.

1.2.2 Modifications of Feedforward Recurrent Alignment for Asymmetric Interactions

We want to have the similar quantification of feedforward recurrent alignment as for the symmetric interaction matrix (section 1.1.3). However, when aligning the feedforward input to eigenvectors of asymmetric recurrent interaction network J like (1.13), the result calculated by (1.11) is a complex number and can therefore be hardly interpreted. To embed such a sore with analogical idea, we deliberate the following modifications and try to discover the outcomes.

Aligning feedforward input $h \in \mathbb{C}^{n\times 1}$ to the asymmetric interaction J, we consider:

1. Only the real part of input h and calculate the feedforward recurrent alignment with $\tilde{h} = \mathrm{Re}(h) \in \mathbb{R}^{n\times 1}$.

$$\nu_{\mathrm{Re}} := \frac{\tilde{h}^T J \tilde{h}}{\|\tilde{h}\|^2} = \frac{\mathrm{Re}(h)^T J \mathrm{Re}(h)}{\|\mathrm{Re}(h)\|^2}\,. \tag{1.43}$$

2. Consider the magnitude of all entries in feedforward input h and take $\tilde{h}_i = |h_i| \in \mathbb{R}\ \forall i$. Therefore $\tilde{h} \in \mathbb{R}^{n\times 1}$.

$$\nu_{\mathrm{mag}} := \frac{\tilde{h}^T J \tilde{h}}{\|\tilde{h}\|^2} \text{ with } \tilde{h}_i = |h_i| \in \mathbb{R}\,. \tag{1.44}$$

3. Symmetrize J through

$$\tilde{J} = \frac{J + J^T}{2}\,. \tag{1.45}$$

Instead of align the feedforward input directly to J, we align indirectly to $\tilde{J}$ with its eigenvectors $\tilde{e}_i \in \mathbb{R}^{n\times 1}$. If having feedforward input aligned to eigenvectors to J as in (1.13), the feedforward recurrent alignment will be calculated with eigenvectors of $\tilde{J}$ instead:

$$\nu_{\text{sym}} = \frac{\tilde{e}_i^T J \tilde{e}_i}{\|\tilde{e}_i\|^2} \in \mathbb{R}\,, \tag{1.46}$$

with $\tilde{\lambda}_i$ eigenvalues of $\tilde{J}$.

1.2.3 Related Modification for Evaluation

As for symmetric interaction network, four activity properties are taken into account to evaluate the feedforward recurrent alignment hypothesis with measurement of the alignment scores. In the case of asymmetric interaction, the modified alignment scores (section 1.2.2). For repetition, the four properties are

1. Trial-to-trial correlation with (1.18).
2. Intra-trial stability with (1.24).
3. Dimensionality.

   With symmetric interaction networks, the covariance matrix for generation of inputs and responses is constructed with eigenvectors of interaction matrix since they build up a set of basis for $\mathbb{R}^{n\times n}$. But with asymmetric interaction matrix, the eigenvectors are basis for $\mathbb{C}^{n\times n}$. If using complex eigenvectors, the inputs and responses will be complex without plausible interpretations. Therefore, the construction of the covariance matrix need to be modified synchronously to have at least vectors from $\mathbb{R}^{n\times 1}$.

   The same problem exists also for the analytical calculation for effective dimensionality (1.28): we now have complex eigenvalues that lead to the dimensionality also complex. To overcome this problem, we work along the same modifications as above for covariance matrix.

   As a result, having e_i eigenvectors and λ_i of asymmetric interaction network J

   - For modification 1 (1.43), applying $\tilde{e}_i = \text{Re}(e_i)$ for construction covariance matrix and $\tilde{\lambda}_i = \text{Re}\lambda_i$ for calculation of effective dimensionality.
   - For modification 2 (1.44), applying magnitude for all entries in e_i to formulate $\tilde{e}_i$ and $\tilde{\lambda}_i = |\lambda_i|$ also magnitude of eigenvalues.
   - For modification 3 (1.46), applying eigenvectors $\tilde{e}_i$ and eigenvalues $\tilde{\lambda}_i$ from $\tilde{J}$ (1.45).

The covariance matrix for generating inputs is constructed similar to it with symmetric interaction network (1.27) but with $\tilde{e}_i$ defined above,

$$\Sigma^{\text{Dim}} := \sum_{i=L}^{L+M} \exp\left(\frac{2(i-L)}{\beta}\right) \tilde{e}_i \tilde{e}_i^T \,. \tag{1.47}$$

Analogously, calculating the effective dimensionality defined by (1.31) but with $\tilde{\lambda}_i$,

$$d_{\text{eff}}^r = \frac{\left(\sum_{i=L}^{L+M} \exp\left(-2\frac{i-L}{\beta}\right)(1-\tilde{\lambda}_i)^{-2}\right)^2}{\sum_{i=L}^{L+M} \exp\left(-4\frac{i-L}{\beta}\right)(1-\tilde{\lambda}_i)^{-4}} \tag{1.48}$$

4. Alignment to spontaneous activity.

   Since the same formulation of covariance matrix with a larger parameter β_{spont} is used for generation of broader endogenous inputs, the same modifications above for dimensionality (1.47) can be taken over into covariance matrix of endogenous inputs like (1.37),

$$\mathbf{\Sigma}^{\text{spont}} := \sum_{i=1}^{\kappa\beta_{\text{spont}}} \exp\left(\frac{2(i-1)}{\beta_{\text{spont}}}\right) \tilde{e}_i \tilde{e}_i^T \,. \tag{1.49}$$

1.3 Low Rank Recurrent Network Model

Until now, we considered full ranked random recurrent networks in both symmetric and asymmetric cases. Those fully recurrent connectivity structure is one of the most popular and best-studied classes of network models. However, randomly connected network display only very stereotyped responses to external inputs, can implement only a limited range of iniput-output computations [**?**].

Furthermore, experimental large-scale neural recordings have established that the transformation of sensory stimuli into motor outputs relies on low-dimensional dynamics at the population level [**?**]. Besides, actual cortical connectivity appears to be neither fully random nor fully structured [**?**]. Understanding how low-dimensional computations on mixed, distributed representations emerge from the structure of the recurrent connectivity and inputs to cortical networks is a major challenge [**?**]. The neural computations can also be understood at the level of dynamical systems that govern low-dimensional trajectories of collective neural activity [**?**].

Low-rank recurrent neural networks (RNNs) rely on connectivity matrices that are restricted to be low rank, which directly generate low-dimensional dynamics. The rank of the network determines the number of collective variables needed to provide a full description of the collective dynamics [**?**].

1.3.1 Construction of Low Rank Interactions

Low-rank networks has the rank smaller than the number neurons. We found two possible constructions of low rank RNNs, which differs if the network is disturbed with Gaussian distributed random noise.

Low-rank RNNs without random noise [?, ?]. Here, neurons in low-rank RNNs are organized in distinct populations that correspond to clusters in the space of low-rank connectivity patterns. Each population is defined by its statistics of connectivity, described by a multi-variate Gaussian distribution, so that the full network is specified by a mixture of Gaussians [**?**].

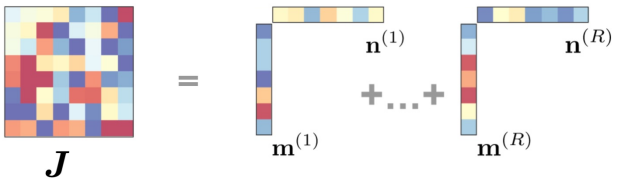

Figure 1.4 Low-rank recurrent networks (RNNs) constructed with distinct Gaussian distribution [**?**]. A low-rank matrix could be written as a sum of outer products of vectors that are Gaussian distributed. As a result, the connectivity matrix is a mixture of Gaussians.

The connection matrix is constructed as a $n \times n$ dimensional matrix J where n is the number of neurons. Using singular value decomposition, the connectivity matrix with rank $G \ll n$ can be expressed as the sum of G unit rank terms

$$J_{ij} = \frac{1}{n} \sum_{g=1}^{G} l_i^{(g)} r_j^{(g)} \text{ or } J = \frac{1}{n} \sum_{g=1}^{G} l^{(g)} r^{(g)T} . \tag{1.50}$$

The connectivity is therefore characterized by a set of G n-dimensional vectors, or the g-th connectivity patterns, $l^{(g)} = \left\{ l_i^{(g)} \right\}_{i=1,...,n} \in \mathbb{R}^{n \times 1}$ and $r^{(g)} = \left\{ r_i^{(g)} \right\}_{i=1,...,n} \in \mathbb{R}^{n \times 1}$ for $g = 1, ..., G$. $\left\{ l^{(g)} \right\}$ are the left singular vectors of the connectivity matrix and $\left\{ r^{(g)} \right\}$ the right. The vectors $\left\{ l^{(g)} \right\}$ and $\left\{ r^{(g)} \right\}$ are mutually orthogonal and randomly multi-variate Gaussian distributed [**?**].

Low-rank RNNs with random noise [?]. Similar to the construction of low-rank connection above (figure 1.4 defined by (1.50)), [**?**] suggested that the connectivity matrix can be constructed with a part P to be fixed and known and a random uncorrelated part.

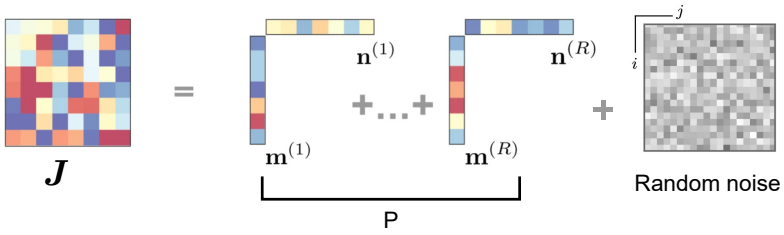

Figure 1.5 Low-rank recurrent networks (RNNs) constructed with fixed part and random noise [?]. The connectivity matrix is given by a sum of an uncontrolled, random matrix and a structured, controlled matrix P. Except for the fixed and know part P, which is a mixture of uncorrelated Gaussians, the RNN is disturbed by random noise.

The fixed part P has the same construction as the network without noise (1.50). The uncontrolled random matrix can be constructed in a complex way with certain current-to-rate transfer function [**?**]. That is

$$J = \frac{1}{n} \sum_{g=1}^{G} l^{(g)} r^{(g)T} + J_{\text{rand}} , \tag{1.51}$$

with J_{rand} a $n \times n$ random matrix.

1.3.2 Feedforward Recurrent Alignment Hypothesis with Low-rank RNNs

To test the feedforward recurrent alignmnent hypothesis applying the low-rank RNNs, we firstly consider the simple case of having symmetric low-rank connectivity, because symmetric structure is easier to interpret. Then we tried the asymmetric low-rank RNNs to discover the influence of rank on response properties.

Symmetric low-rank RNN This could be achieved through choosing the left connectivity vectors $\left\{l^{(g)}\right\}$ equal the right connectivity vectors $\left\{r^{(g)}\right\}$. As a result the low-rank RNN without noise can be formulated by sum of symmetric matrices and therefore also symmetric:

$$J = \frac{1}{n}\sum_{g=1}^{G} l^{(g)} l^{(g)T} \text{ or } J = \frac{1}{n}\sum_{g=1}^{G} r^{(g)} r^{(g)T}\,. \tag{1.52}$$

If considering the connectivity matrix with noise, for simplicity, we add a random $n \times n$ symmetrized Gaussian distributed matrix to J from (1.52).

The dynamics and conditions for stable stability of the response in the symmetric low-rank RNNs should be kept the same as with symmetric full rank RNNs (1.4, 1.10). In order to keep the steady state of responses stable, the eigenvalues $\lambda_i \in \mathbb{R}$ of the low-rank RNNs J (1.52) are limited by $R < 1$ through normalizing with the maximal eigenvalue (1.3).

Asymmetric low-rank RNN When the set of left connectivity vector $\left\{l^{(g)}\right\}$ is different from the set of right connectivity vector $\left\{r^{(g)}\right\}$, we receive the asymmetric low-rank with (1.50) for the case without random noise. If considering the random noise, we again add a simple $n \times n$ Gaussian distributed full rank asymmetric matrix as figure 1.5 illustrated (1.51).

To enable the stable steady state of responses, as shown in asymmetric RNNs (1.41), the real part of eigenvalues $\mathrm{Re}(\lambda_i)$ of asymmetric RNNs J (1.51) are limited by $R < 1$ through normalization with the maximal magnitude of eigenvalues (1.42).

1.3.3 Evaluation of Feedforward Recurrent Alignment Hypothesis Based on Response Properties

Similar to the analysis on full rank networks, we also want to evaluate the modulation of feedforward recurrent alignment for low-rank RNNs based on the four response properties in correlation with feedforward recurrent alignment:

- Trial-to-trial correlation
- Intra-trial stability

- Dimensionality
- Alignment of evoked activity pattern to spontaneous activity pattern

Symmetric low-rank RNN For symmetric low-rank RNN, the methods for modulations and evaluations are traced back to the full rank symmetric RNN before. The feedforward recurrent alignment is defined by (1.11). Moreover, the trial-to-trial correlation is quantified with (1.18), the intra-trial stability with (1.24), the dimensionality with (1.31) and (1.32), and at the end the alignment to spontaneous activity with (1.34).

Asymmetric low-rank RNN Since for asymmetric low-rank RNNs, the problem of complex eigenvectors and eigenvalues still exists, the modifications undertaken at asymmetric full rank RNNs (section 1.2.2 and 1.2.3) can therefore be directly applied here, when align the inputs to the RNN. Generally, the modifications that were considered can be roughly described as: 1) only consider real part of eigenvalues and eigenvectors, 2) consider for each neuron the magnitude of inputs and variance ratio, 3) align the inputs to symmetrized network to approximate.

1.4 Black Box Recurrent Network Model

Until now, we assume that we already know the structure of recurrent networks and evaluate the feedforward recurrent alignment hypothesis on the networks. However, in reality during the experiments, mostly only a small part of the network is known and the total structure of the network keeps mostly like a black box. As a result, the eigenvectors of the networks are generally unknown. The feedforward recurrent alignment cannot use the dominant modes to characterize the development of feedforward recurrent network system.

It was pointed out that the reliability of evoked dynamics in recurrent networks is dependent on the stimulus used. As a consequence, to a recurrent network would correspond a set of stimuli which are more efficiently transmitted than others [**?**]. Especially the stimulus inputs that align with the structure of endogenous subnetworks would be recurrently amplified, leading to more reliable evoked responses [**?**].

Besides, the similarity between spontaneous and evoked activity in sensory cortical areas could be a signature of efficient transmission and propagation across cortical networks. Based on a better recall caused by a match between spontaneous activity and input statistics, it was hypothesized that the recurrent connectivity could have been shaped by a learning process so that the spontaneous activity matches the natural input statistics [**?**].

Therefore, we wonder if we could apply the experimental measurable spontaneous activity to characterize the feedforward recurrent alignment. Without knowing the eigenvectors of the recurrent networks, align feedforward inputs to the spontaneous activity instead of the recurrent networks. This can be a further modification for modeling feedforward recurrent alignment for general asymmetric recurrent networks.

Furthermore, we consider to repeatedly apply the recurrently amplified spontaneous stimuli as inputs to discover if the amplified spontaneous activity could increase the alignment to the recurrent network.

1.4.1 Approximation with White Noise Evoked Activity

With unknown recurrent structure, which we assume to be asymmetric generally, it is then difficult to find the stimuli pattern such that the trial-to-trial correlation, intra-trial stability and alignment between evoked activity to spontaneous activity to be high while keeping dimensionality low. In other words, we cannot apply the eigenvalues of the recurrent network to align with inputs and then characterize the development of feedforward inputs leading to stable response properties.

Alternatively, we align the inputs to spontaneous pattern and explore the response properties correlation with feedforward recurrent alignment.

Spontaneous activity is evoked by white noise $h \in \mathbb{R}^{n\times 1}$, which is modulated by multivariate normal distribution with mean zero vector $0_v \in \mathbb{R}^{n\times 1}$ and covariance matrix the identity matrix $I_n \in \mathbb{R}^{n\times n}$,

$$h_{\text{white}} \sim \mathcal{N}(0_v, I_n)\,. \tag{1.53}$$

The spontaneous activity is then modeled by the transformed steady state response $r \in \mathbb{R}^{n\times 1}$, which is also multivariate normal distributed with transformed covariance matrix,

$$r_{\text{spont}} \sim \mathcal{N}\left(0_v, (1-J)^{-1}(1-J)^{-T}\right)\,. \tag{1.54}$$

The response pattern is determined by the covariance matrix. If align the inputs to response patterns, the eigenvectors of covariance matrices are aligned. The eigenvectors of a covariance matrix are also known as principal components for the distribution.

To model the feedforward recurrent alignment hypothesis, the inputs are aligned to principal components and the feedforward alignment score is formulated with principal components instead of with modified eigenvectors of asymmetric recurrent network like section 1.2.2. For a input h aligned to a principal component p, the feedforward recurrent alignment is constructed with

$$\nu := \frac{p^T J p}{\|p\|^2}\,. \tag{1.55}$$

The original recurrent network is now approximated by the spontaneous response pattern and the inputs are aligned to principal components of the spontaneous response pattern. As for the recurrent networks, some properties of the new formulated feedforward recurrent alignment score have to be evaluated. The perspectives that we take into account are:

- Monotonously growing correlation between feedforward recurrent alignment score and eigenvalues of covariance matrix from spontaneous activity pattern.
- Positive correlation between feedforward recurrent alignment score and trial-to-trial correlation.
- Positive correlation between feedforward recurrent alignment score and intra-trial stability.
- Negative correlation between feedfroward recurrent alignment score and dimensionality.
- Feedforward recurrent alignment score is positive correlated with alignment of evoked activity to spontaneous activity.

Monotony The feedforward recurrent alignment should reflect how well the input pattern is aligned with the considered recurrent network. The more the input pattern is aligned with the dominant projection direction in activity space, the stronger should be the evoked response due to response amplification. The response strength is determined by the corresponding eigenvalue of aligned direction. Since the inputs are aligned to the spontaneous activity pattern, the feedforward recurrent alignment should be monotonusly positive correlated with the eigenvalues of spontaneous activity pattern.

Trial-to-trial correlation We consider the case of general asymmetric recurrent network from section 1.2.1. Thus, the modification is similar to section 1.2.2. The input pattern h is aligned to the principal component p of the covariance matrix from spontaneous activity pattern and therefore modeled by

$$h \sim \mathcal{N}(p, \sigma_{\text{trial}} I_n) \,. \tag{1.56}$$

The steady state response evoked by the inputs are transformed multivariate normal distribution

$$r \sim \mathcal{N}\left((1-J)^{-1}p, \sigma_{\text{trial}}(1-J)^{-1}(1-J)^{-T}\right) \,. \tag{1.57}$$

The trial-to-trial correlation reflects the variation between different trials. It is again the average of pairwise Pearson correlation between response trials as defined by β_s (1.18).

Intra-trial stability Intra-trial stability quantifies the variation inside one response trial evoked by input. One time-dependent input and evoked steady state response trial is approximated by Euler-Maruyama scheme 1.19. The input pattern h is aligned to principal components p of activity pattern. Therefore, the mean vector for h is the aligned principal component p. The input and evoked response are therefore formulated as

$$\mathrm{d}h = p\mathrm{d}t + \sigma_{\text{time}}\mathrm{d}W \tag{1.58a}$$

$$\mathrm{d}r = (-r + J \cdot p)\mathrm{d}t + \sigma_{\text{time}}\mathrm{d}W \,. \tag{1.58b}$$

The intra-trial stability is the time average of delayed response correlation defined by (1.25).

Dimensionality For modulation of change in dimensionality along alignment, the covariance matrix for input distribution is constructed as 1.27 but with principal components p_i of spontaneous activity as an approximation to the eigenvectors of the original recurrent network.

$$\Sigma^{\text{Dim}} := \sum_{i=L}^{L+M} \exp\left(\frac{-2(i-L)}{\beta}\right) p_i p_i^T \,. \tag{1.59}$$

The input and the evoked activity are modeled by multivariate normal distribution

$$h \sim \mathcal{N}(0_v, \Sigma^{\text{Dim}}) \tag{1.60a}$$

$$r \sim \mathcal{N}(0_v, (I_n - J)^{-1}\Sigma^{\text{Dim}}(I_n - J)^{-T}) \,. \tag{1.60b}$$

The effective dimensionality (1.28) is approximated empirically by the eigenvalues of covariance matrix from spontaneous activity pattern (1.32).

Alignment between evoked activity and spontaneous activity To grantee the spontaneous has a broader input than the evoked activity, the spontaneous activity here for alignment to evoked activity is constructed similar to (1.60) but with a higher dimensioality $\beta_{\text{spont}} > \beta$. For the formulation of covariance matrix for spontaneous activity, the principal components from white noise evoked activity pattern is used in (1.49).

$$\Sigma^{\text{spont}} := \sum_{i=L}^{M+1} \exp\left(\frac{-2(i-1)}{\beta_{\text{spont}}}\right) p_i p_i^T \,. \tag{1.61}$$

The spontaneous activity is then modeled by

$$r \sim \mathcal{N}(0_v, (I_n - J)^{-1}\Sigma^{\text{spont}}(I_n - J)^{-T}) \,. \tag{1.62}$$

The amount of overlap between evoked activity pattern (1.60b) and the principal components of spontaneous activity (1.62) quantifies the alignment (1.34). The average alignment over all N evoked response trials is the final alignment score for alignment to spontaneous activity.

$$\gamma = \frac{1}{N}\left(\frac{r_i^T \cdot \Sigma^{\text{spont}} \cdot r_i}{\|r_i\|^2 \text{Tr}(\Sigma^{\text{spont}})}\right) \tag{1.63}$$

1.4.2 Iterative Approximation with Low Dimensional Inputs

Low dimensional inputs can be generated experimentally, while the high dimensional inputs are almost not possible to be generated. So, we wonder if the feedforward recurrent alignment can also be adapted to represent the development and better alignment under the settings that, 1) only low dimensional inputs are offered and 2) the original recurrent network is asymmetric but unknown.

To model the low dimensional input, random orthonormal basis vectors for construction of covariance matrix e_i are obtained through Gram-Schmidt process. The same scheme as Σ^{Dim} is applied,

$$\Sigma_{\text{Low}} := \sum_{i=L}^{1+M} \exp\left(\frac{-2(i-1)}{\beta}\right) e_i e_i^T \,. \tag{1.64}$$

The low dimensional inputs are then modeled as multivariate normal distribution with mean the zero vector $0_v \in \mathbb{R}^{n\times 1}$ and covariance matrix Σ^{Low},

$$h_{\text{Low}} \sim \mathcal{N}(0_v, \Sigma_{\text{Low}})\,. \tag{1.65}$$

The response evoked by low dimensional input (1.65) is the linearly transformed multivariate normal distribution

$$r_0 \sim \mathcal{N}\left(0_v, (1-J)^{-1}\Sigma_{\text{Low}}(1-J)^{-T}\right)\,. \tag{1.66}$$

At this step, the feedforward recurrent alignment ν can be calculated with low dimensional input h

$$\nu_0 = \frac{h^T J h}{\|h\|^2}\,. \tag{1.67}$$

Prior knowledge predicts stimulus inputs can be recurrently amplified and lead to more reliable responses [**?**]. Besides, the recurrent connectivity pattern shapes so that the inputs matches better with spontaneous activity during development [**?**]. the response can be better aligned with the spontaneous activity than the input

1.5 Hebbian Learning in Feedforward Recurrent Networks

1.5.1 Model Setting

1.5.2 Update Rules for Feedforward Network

1.5.3 Projection of the Feedforward Weights on Eigenvectors