# 1 Results

## 1.1 Response Properties from symmetrical Recurrent Interaction Networks in Correlation with Feedforward Recurrent Alignment

The feedforward recurrent alignment, defined in the method under section **??** quantify the degree of how much the feedforward input is aligned with the direction that spanned the recurrent network. If the input is well aligned with the maximal eigenvector of the recurrent interaction $J$, the random noise in the response would be suppressed by the evoked response due to the selective amplification. That is, with response amplification

$$\mathbf{r}^* = \sum_{i=1}^{n} \frac{(\mathbf{e}_i \cdot \mathbf{h})\mathbf{e}_i}{1 - \lambda_i} \, , \tag{1.1}$$

the steady state response is then dominated by the projection of the input vector $\mathbf{h}$ along the axis defined by the eigenvector $\mathbf{e}_{\max}$ whose eigenvalue $\lambda_{\max}$ is maximal and near one [**?**].

$$\mathbf{r}^* \approx \frac{(\mathbf{e}_{\max} \cdot \mathbf{h})\mathbf{e}_{\max}}{1 - \lambda_{\max}} \tag{1.2}$$

The projection of input on eigenvector $\mathbf{e}_{\max}$ reaches its maximal when $\mathbf{h}$ is approximately $\mathbf{e}_{\max}$ itself. Therefore, the steady state responses reaches its maximum and the feedforward recurrent alignment equals the maximal eigenvalue $\lambda_{\max}$ of $J$ because of (**??**).

On the other hand, if the input is not well aligned with the dominant eigenvectors, the random noise is large relative to the response. For an extreme example, if the input is aligned with the eigenvector with minimal eigenvalue $\lambda_{\min}$, the response will almost not contribute to the steady state response at all due to the response amplification (1.1). This could also be reflected by the feedforward recurrent alignment, which equals $\lambda_{\min}$ in this case because of (**??**).

We can thus conclude that the feedforward recurrent alignment (**??**) reflects the alignment between the feedforward input and the dominant eigenvectors of the recurrent interaction network $J$. In the following sections, the four properties introduced at section **??** will be evaluated in the model and compare to the tendency observed in [**?**].

### 1.1.1 Trial-to-Trial Correlation increases with larger Alignment

As defined in the paragraph **??**, the inputs are constructed by multivariate normal distribution (**??**). To model the alignment of inputs with eigenvectors, the eigenvectors of the interaction matrix $J$ are chosen to be the the mean vector $\mu$ for input

1

distribution. That is

$$\mathbf{h_i} \sim \mathcal{N}(\mathbf{e}_i, \sigma_i^2 \mathbf{I_n}) \,. \tag{1.3}$$

The degree of alignment of input on the dominant eigenvector $\mathbf{e}_{\max}$ is then determined approximately by the size of eigenvalue, to which the eigenvector is corresponded.

We want to find out the correlation between the feedforward recurrent and the trial-to-trial correlation $\beta_s$, defined by (**??**). Therefore, sort the eigenvectors in the order such that their corresponding eigenvalues are in ascending order. That is

$$\mathbf{e}_{\min}, ..., \mathbf{e}_i, \mathbf{e}_j, ..., \mathbf{e}_{\max} \text{ such that } \lambda_{\min} < ... < \lambda_i < \lambda_j < ... < \lambda_{\max} \,, \tag{1.4}$$

with $\lambda_i$ the corresponding eigenvalue for eigenvector $\mathbf{e}_i$. As a result, the inputs that aligned with eigenvectors in this order have an increasing feeedforward recurrent alignment approximately.

Generating the results with (**??**) for $N$ trials. The trial-to-trial correlation can be calculated with (**??**).
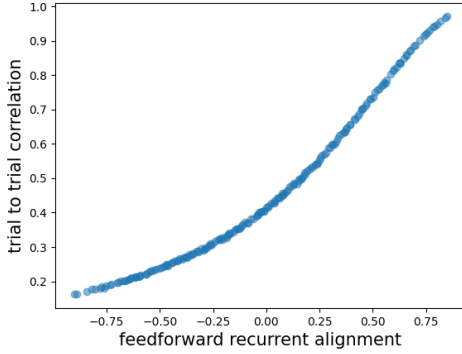


**Figure 1.1 Correlation between feedforward recurrent alignment and trial to trial correlation.** Inputs aligned to eigenvectors $\mathbf{e}_i$ of interaction matrix $J$ in the ascending order of eigenvalues (1.4), resulting the feedforward recurrent alignment varies approximately between $\lambda_{\min}$ and $\lambda_{\max}$. For each input alignment to an eigenvector, $N$ trials of evoked responses were generated for calculation of the trial-to-trial correlation.

We considered visually naive cortex as feedforward recurrent alignment equals zero, which could be interpreted as responses evoked by random inputs. The trial-to-trial correlation with random inputs is then smaller than the case, when the feedforward recurrent alignment takes the maximal value under inputs aligned with $\mathbf{e}_{\max}$. This coincides with the experimental observations od responses from visually naive and experienced primary visual cortex of ferrets [**?**].

Moreover, the modulation suggests a positive correlation between the feedforward recurrent alignment and the trial-to-trial correlation over the whole alignment range. The result confirms the idea that with the feedforward inputs more and more aligned with the dominant eigenvector of recurrent network, the stability between trials increases also simultaneously and almost continuously. The process of reaching higher trial-to-trial stability is therefore a process of becoming more aligned with the dominant eigenvector.

### 1.1.2 Intra-Trial Stability increases with larger Alignment

Now we want to see if our modeling could capture the change in intra-trial stability during the development observed in the primary visual cortex of ferrets [**?**]. The intra-trial stability increased after the eye opening and a couple of days. The feedforward recurrent alignment hypothesis suggests that the visual experienced cortex should have a better alignment between feedforward inputs and the dominant modes in recurrent network. To confirm this idea, we would expect the intra-trial stability would be larger with better alignment between input and the dominant eigenvector.

Analogous to trial-to-trial correlation, the eigenvectors are sorted in the descending order according to the eigenvalues (1.4). The intra-trial stability is calculated with (**??**).
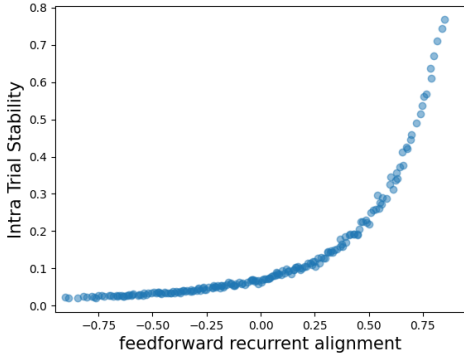


**Figure 1.2 Correlation between feedforward recurrent alignment and intra-trial stability.** Inputs aligned to eigenvectors $\mathbf{e}_i$ of interaction matrix $J$ in the ascending order of eigenvalues (1.4), resulting the feedforward recurrent alignment varies approximately between $\lambda_{\min}$ and $\lambda_{\max}$. For each input alignment to an eigenvector, the intra-trial stability is calculated with the evoked steady state response.

The result (figure 1.2) indicates also a positive correlation between the feedforward recurrent alignment and the intra-trial stability. With random feedforward inputs, the feedforward recurrent alignment take the value near zero. So the eye opening happens somewhere between feedforward recurrent equals zero and reaches maximal. Thus, before the eye opening, there is already certain alignment that lead to a certain degree of intra-trial correlation.

Furthermore, the correlation is almost exponential. So, the enhancement of the input alignment to the dominant eigenvector is more rapid after the eye opening than before the eye opening. One assumption for this phenomenon could be that after the eye opening, the environment provides more training data for the network so that the alignment between inputs and dominant eigenvector could be improved more efficiently. As a result, the responses get more intense and drive a better alignment forward. A positive loop could arise and speed up until the optimum is reached.

### 1.1.3 Dimensionality decreases with larger Alignment

Dimensionality is a generally important property of neural representations and could help to understand processes for example in learning and controlling [**?**, **?**]. A low dimensional representation will encode a diverse range of inputs into a small set of common, orthogonal activity patterns. In other words, low dimensional activity pattern require small number of basis vectors from the response space to represent itself. On the other hand, a high dimensional representation will separate even similar inputs into orthogonal activity patterns. Compare to low dimensional representation, a high dimensional activity pattern represents with a large set of basis vectors [**?**].

Therefore, we also aim to take a look at the change of dimensionality during the increase of alignment in the modeling. In ferrets primary visual cortex, the dimensionality decreased from days before eye opening to eye opening and then to days after eye opening [**?**]. We would then expect that the model should also suggest a decrease of dimensionality with increase of alignment between inputs and dominant eigenvector of recurrent network.

For a certain alignment, the principal component analysis reflects actually directly the dimensionality of the evoked activity pattern under this alignment. Because the principal components are the eigenvectors of response covariance, they also build up a set of basis vector for the activity pattern space. The variance ratio of principal components reflect the weight that each eigenvector takes to represent the activity pattern. Thus, if only a small number of principal components contributes the most, the activity pattern is then low dimensional. While if a broad set of principal components are similarly important, the activity pattern is high dimensional.

With the idea of earning the linear dimensionality defined as participation ratio based on the principal component analysis, the eigenvectors here are ordered in descending order, in the opposite order of ascending order defined above (1.4). That is

$$\mathbf{e}_{\max}, ..., \mathbf{e}_i, \mathbf{e}_j, ..., \mathbf{e}_{\min} \text{ such that } \lambda_{\max} > ... > \lambda_i > \lambda_j > ... > \lambda_{\min}. \qquad (1.5)$$

For the generation of inputs with covariance matrix $\mathbf{\Sigma}^{\mathrm{Dim}}$ (**??**), a subset of eigenvectors $\{\mathbf{e}_i\}_{i=L,...,L+M}$ will be chosen for $L = 1, ..., \frac{n}{2}$. $M$ then determines how many eigenvectors will contribute to generate inputs and evoked activity. In each such subset of eigenvectors, the leading eigenvector is $\mathbf{e}_L$. Approximate here the feedforward recurrent alignment with the leading eigenvector only. Since $L$ is considered only in range of the first half of eigenvectors ordered as (1.5), the range of feedforward recurrent alignment is between around 0 and $\lambda_{\max}$. The linear dimensionality analytically and empirically will be calculated according to (**??**) and (**??**).
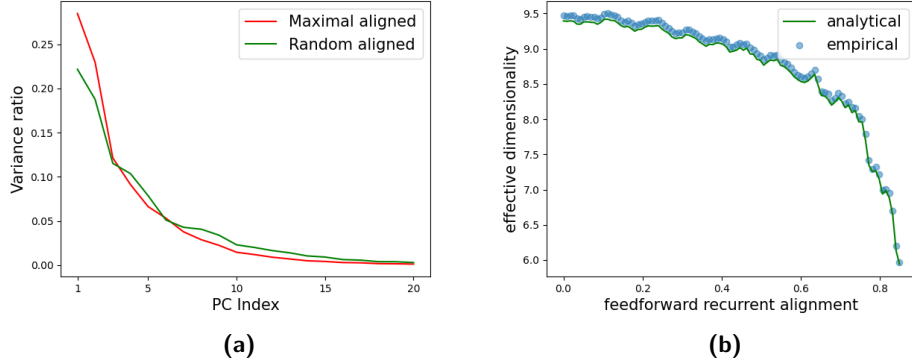
**Figure 1.3 The correlation between dimensionality and feedforward recurrent alignment.** Prior experimental observation suggests that the dimensionality decreases from prior to post eye opening [**?**]. With the feedforward recurrent alignment hypothesis, dimensionality property of the neural representation could be captured. **(a)** Principal component analysis for the evoked activity under best alignment between inputs and dominant eigenvector and spontaneous random alignment. Red line for maximal alignment and green line for spontaneous alignment. (b) Achieve the correlation between dimensionality and feedforward recurrent alignment with analytical (**??**) and empirically (**??**). The green line displays the analytical approximation for dimensionality and the blue dots for empirical approximation.

As explained above, the principal component analysis helps to decode the dimensionality. The curve in figure 1.3 (a) of the variance ratio reflects the dimensionality. Spontaneous random alignment has a flatter variance ratio curve, indicating a broader range of eigenvector contribution. Therefore, the spontaneous random alignment has a higher dimensionality. With the feedforward recurrent alignment hypothesis, the spontaneous random alignment matches the state before the eye opening and the maximal alignment for the state after eye opening. Therefore, the hypothesis results the same tendency of dimensionality decrease that observed in ferrets[**?**].

In total, a negative correlation between the dimensionality and feedforward recurrent alignment is shown in figure 1.3 (b). The correlation forms nearly a flipped logarithmic function. The error between analytical and empirical results are small, which confirmed the well approximation formulated analytically by (**??**).

Besides, the flipped logarithmic correlation suggests the similar principle for intra-trial stability (figure 1.2). After eye opening, the reduction of dimensionality becomes larger when the feedforward inputs aligns better with dominant eigenvector. It could be the case, that after the eye opening, the recurrent network tries to encode the environment information with large number of eigenvectors, which is costing for the system. After a period of time of getting used to the stimuli, the system speed up to find out a more efficient encoding with fewer eigenvectors and the information becomes more determined.

### 1.1.4 Alignment to spontaneous activity increases with larger Alignment

Spontaneous activity in neural systems is defined as neural activity that is not driven by an external stimulus. The activity patterns of spontaneous activity is not completely random and have often unique spatiotemporal patterns that instruct neural circuit development in the developing brain. Moreover, normal and aberrant patterns of spontaneous activity underline behavioral states and diseased conditions in adult brains. Therefore, the spontaneous activity is essential for the understanding of brain development [**?**]. The alignment between activity pattern and spontaneous activity pattern could show the structural relation between patterns. Besides, it could also reflect the presentation of spontaneous activity in other activity patterns.

In baby ferrets' brain, a transformation from visual responses that are loosely aligned with spontaneous activity in the cortex before eye opening to the reliable and well aligned responses several days after eye opening were observed [**?**]. Thus, we will expect that the feedforward recurrent hypothesis can reflect the tendency that evoked activity aligned better to spontaneous activity during the development of brain.



(a)  (b)
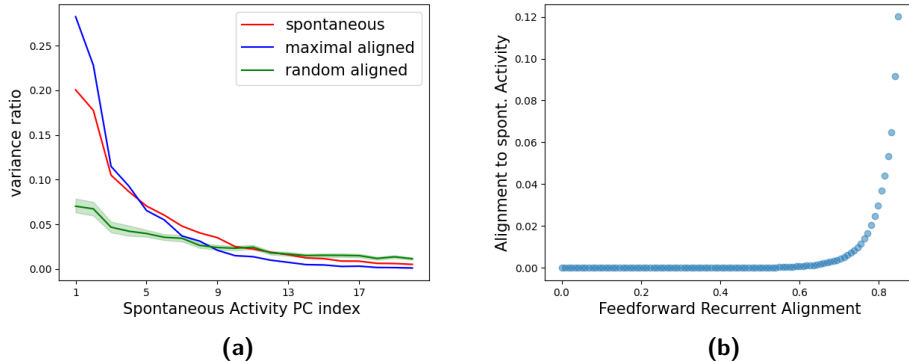
**Figure 1.4 Correlation between alignment to spontaneous activity and feedforward recurrent alignment.** Spontaneous activity reflects inputs from wide range of different sources and considered to already aligned to the recurrent network[**?**]. Aligning activity patterns to spontaneous activity is in principle to explain the activity pattern by the principal components of spontaneous activity. **(a)** Variance ratio (**??**) of spontaneous activity, evoked activity by feedforward input maximally aligning to recurrent network, and evoked activity by randomly aligning to recurrent network explained by principal components of spontaneous activity. The red line illustrates the variance ratio of spontaneous activity, blue line the maximal alignment, and green line the random alignment. Shadow shows the 95% confident interval for 50 networks. (b) The correlation between final alignment score to spontaneous activity and feedforward recurrent alignment better than random alignment calculated with (**??**). The eigenvalues are ordered in the same way as (1.5). Only the first half eigenvectors are taken into account to determine the correlation.

Under the assumption that at eye opening, the patterns of feedforward inputs are not as well aligned to the recurrent network as the spontaneous activity, the evoked and spontaneous pattern overlaps only little (figure 1.4(a), visually comparing green line to red line.) It could be observed here that only the last few principal components has the similar variance ratio, while the first few dominant eigenvectors differs a lot. On the contrary, experience-driven changes that optimize the feedforward-recurrent alignment results in a stronger overlap between distributions of evoked and spontaneous activity pattern (figure 1.4(a), visually comparing red line to blue line.). Most of the principal components has the similar explained variance ratio. In both cases, the hypothesis matches experimental observations in baby ferrets' brain [**?**].

To visualize and quantify the overlap between activity patterns from evoked and endogenous pattern, we considered the summarized alignment score (1.4). A exponential correlation between alignment to spontaneous activity and feedforward recurrent alignment (figure 1.4(b)) is suggested by the modeling. The strong growth of overlaps between evoked and endogenous activity starts only shortly before the optimal experience-driven alignment, indicating that the alignment could require a large amount of experience and training. The costly process to optimal alignment to spontaneous activity could on the other hand reflect the importance of the connection between evoked and endogenous activity patterns.

## 1.2 Evaluation of Feedforward Recurrent Alignment Modulations for asymmetric Recurrent Interaction Networks

For symmetric interaction networks, the feedforward recurrent alignment hypothesis and the modulation based on it can demonstrate the response properties observed in ferrets [**?**] (section 1.1). With better alignment between inputs and recurrent network, key results of the response properties were

- The trial-to-trial correlation increases.

- The intra-trial stability increases.

- The Dimensionality decreases.

- The Alignment of evoked activity to spontaneous activity increases.

However, since the symmetric interaction matrices simplify the neural connection dramatically, we try to embed the more biological realistic asymmetric interaction matrices. Considering the modifications listed in methods (section **??**), we want to evaluate the results based on the key results we got with symmetric interaction matrices.

Firstly, we check if the feedforward recurrent alignment score keeps the proportionality to eigenvalues as shown with (**??**). We expect that a suitable modification could keep the positive correlation with eigenvalues. Then, we go through the four response properties to verify if the tendency above are still kept with increased alignment between inputs and recurrent network.

### 1.2.1 Monotony of Feedforward Recurrent Alignment Score in dependence of Eigenvalues

During the development, suggested by feedforward recurrent alignment hypothesis for symmetric interactions, the inputs align from eigenvectors with small eigenvalues to eigenvectors with large eigenvalues. Due to the formulation (**??**), feedforward recurrent alignment is proportional to the corresponding eigenvalue of the eigenvector that is aligned with input (**??**).

When considering the asymmetric interaction network, keeping the hypothesis that the input align more and more to the eigenvector with maximal eigenvalue, we examine if the modified feedforward recurrent alignment score could still keep the proportionality that the modified variations increase with larger eigenvalues. If the positive correlation between the feedforward recurrent alignment and eigenvalues could be kept, the alignment score could at least quantify how well input align to the dominant eigenvector and thus quantify the alignment between input and recurrent network.
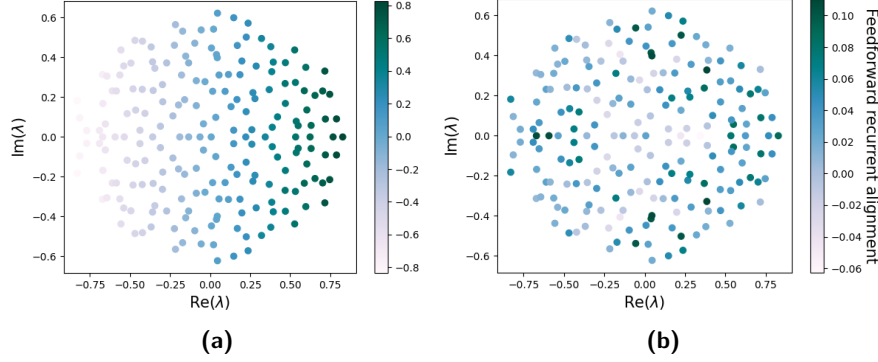
**Figure 1.5 Correlation between eigenvalues and feedforward recurrent alignment.**
To verify the considered modifications, a positive correlation between eigenvalues and feed-
forward recurrent alignment is advantageous. Represent the correlation in the complex
plane, since eigenvalues are complex. Color bar indicate the size of feedforward recurrent
alignment. The darker the color, the larger the alignment score. **(a)** Representation of
alignment score calculated with modification **??** (**??**) in complex plane. **(b)** Representation
of alignment score calculated with modification **??** (**??**) in complex plane.

**Modification 1**   : Apply the real part of complex inputs (**??**).

With the feedforward recurrent alignment, we mainly consider the correlation of
alignment score and eigenvalues under alignment between inputs and eigenvectors
of recurrent network $J$. Inserting inputs equal the eigenvectors $h = e_i \in \mathbb{C}^{n \times 1}$ into
(**??**), it leads to

$$\nu_{\text{Re}} = \frac{\text{Re}(e_i)^T J \text{Re}(e_i)}{\|\text{Re}(e_i)\|^2} \,. \tag{1.6}$$

Because $\|e_i\| = c\|\text{Re}(e_i)\|$ with $c \in \mathbb{R}_+$ a general positive constant, we could get the
proportionality between alignment score and the real part of eigenvalues $\text{Re}(\lambda_i)$ of
$J$ :

$$
\begin{aligned}
\nu_{\text{Re}} &= c\text{Re}\left(\frac{e_i}{\|e_i\|}\right)^T J \,\text{Re}\left(\frac{e_i}{\|e_i\|}\right) \\
&= c\text{Re}\left(\frac{e_i^T}{\|e_i\|} J \frac{e_i}{\|e_i\|}\right) \\
&= c\text{Re}(\lambda_i)\,,
\end{aligned} \tag{1.7}
$$

with $c$ general positive constant.

The positive correlation between alignment and real part of eigenvalues $\lambda_i$ could
also be observed in representation in complex plane (figure 1.5a). With increasing
real part of eigenvalue, the size of alignment score also becomes larger. Furthermore,
there is no correlation found in the direction of imaginary part of eigenvalues, since
the alignment score only depends on the real part.

**Modification 2** : Apply the magnitude for each neuron input (**??**).

When the inputs are aligned to eigenvectors $e$ of $J$, the feedforward recurrent alignment could be expressed with $|e| := (|e_i|)_{i=1,\ldots,n} \in \mathbb{C}^{n \times 1}$ after inserting into (**??**). As the euclidean norm of vector $|e|$ is the same as the norm directly on the complex eigenvector $\|e\|_2$, the final feedforward recurrent alignment score can be then formulated as following:

$$
\begin{aligned}
\nu_{\mathrm{mag}} &= \frac{|e|^T J |e|}{\|e\|^2} \\
&= \frac{\sum_{j=1}^n |e_j| \left( \sum_{i=1}^n |e_i| J_{ij} \right)}{\|e\|^2} ,
\end{aligned}
\tag{1.8}
$$

where $e_i, e_j$ are the $i$-th and $j$-th element of the vector $e$, and $J_{ij}$ the matrix element at $i$-th row and $j$-th column.

Therefore, no direct correlation between the alignment score and eigenvalues $\lambda$ can be established. The lack of correlation is also represented in the complex plane (figure 1.5b).

**Modification 3** : Align the inputs to eigenvectors of symmetrized network (**??**).

Instead of aligning the inputs to eigenvectors of original asymmetric recurrent network and thinking about modifications of eigenvectors to calculate the feedforward recurrent alignment score, we now directly align the inputs to real eigenvectors $\tilde{e}$ of symmetrized recurrent network $\tilde{J}$ as an approximation. However, the alignment score is projected to the original asymmetric recurrent network (**??**). There is still a proportionality between the eigenvalues of symmetrized network and alignment score.
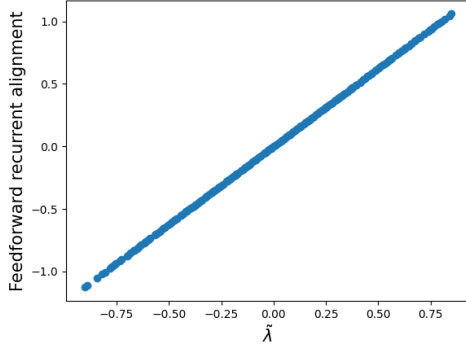


**Figure 1.6 Positive correlation between feedforward recurrent alignment score and eigenvalues of symmetrized network.** Project the inputs to the eigenvectors of symmetrized network to refer the development instead of the projection to the original asymmetric network. The correlation between the alignment score (y-axis) based on the original recurrent network to the eigenvalues of symmetrized network (x-axis) remains positive.

Since the symmetrized recurrent network $\tilde{J}$ is received by (**??**), the proportion-

10

ality shown in figure 1.6 can be obtained as the following:

$$\frac{\tilde{e}^T \tilde{J} \tilde{e}}{\|\tilde{e}\|^2} = \frac{\tilde{e}^T \frac{J+J^T}{2} \tilde{e}}{\|\tilde{e}\|^2} = \frac{1}{2} \left( \frac{\tilde{e}^T J \tilde{e}}{\|\tilde{e}\|^2} + \frac{\tilde{e}^T J^T \tilde{e}}{\|\tilde{e}\|^2} \right)$$

$$\Rightarrow 2\frac{\tilde{e}^T \tilde{J} \tilde{e}}{\|\tilde{e}\|^2} - \frac{\tilde{e}^T J^T \tilde{e}}{\|\tilde{e}\|^2} = \frac{\tilde{e}^T J \tilde{e}}{\|\tilde{e}\|^2}$$

$$\Rightarrow 2\tilde{\lambda} - c = \frac{\tilde{e}^T J \tilde{e}}{\|\tilde{e}\|^2} \text{ with } c := \frac{\tilde{e}^T J^T \tilde{e}}{\|\tilde{e}\|^2} \in \mathbb{R} \text{ and } \tilde{\lambda} \text{ the corresponding eigenvalue of } \tilde{e}$$

$$\Rightarrow \frac{\tilde{e}^T J \tilde{e}}{\|\tilde{e}\|^2} \propto \tilde{\lambda} \,.$$

$$(1.9)$$

**Conclusion**  After the analysis above for all modifications, our expectation of contained positive correlation are kept when considering the real part (modification 1) and alignment with symmetrized recurrent network (modification 3). The variant with magnitude (modification 2) fails to fulfill the correlation. Thus, we will leave the modification 2 out without further analysis for response properties in the following section.

### 1.2.2 Representing Response Properties with modified Feedforward Recurrent Alignment

After verifying the expected correlation between feedforward recurrent alignment score and the network that the inputs are aligned to, we will check out if the remaining modifications could still reflect the four experimental observations listed at the beginning of the section 1.2 with remaining modification 1 and 3.

- When aligning the inputs to the asymmetirc recurrent network to test the feedforward recurrent alignment hypothesis, the modification 1 considers the complex input only with real part of eigenvectors.

- The modification 3 aligns the inputs to symmetrized recurrent network but calculate the alignment score still with original asymmetric network.

Besides, since the asymmetirc recurrent network is constructed with certain degree of symmetry (**??**), we also controlled how much the degree of symmetry influences the results.

**Trial-to-trial Correlation**  Trial-to-trial correlation is the averaged correlations between trials (**??**). The expectation is a positive correlation between feedforward recurrent alignment and trial-to-trial correlation, independent of the degree of symmetry.
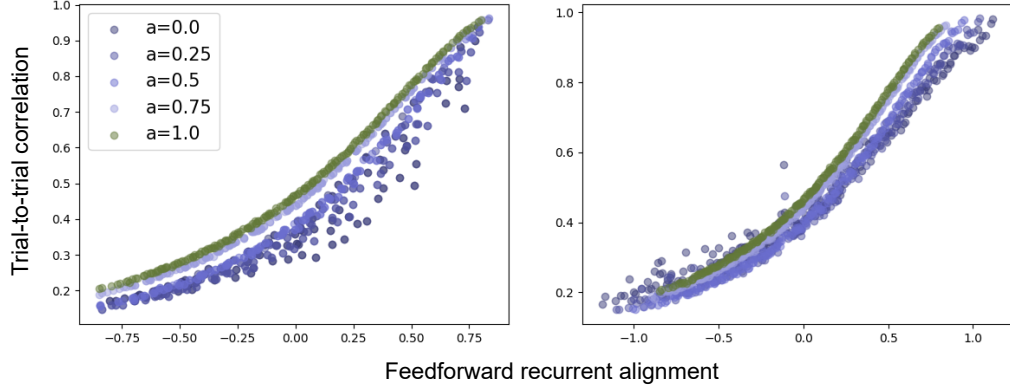
**Figure 1.7 Correlation between trial-to-trial correlation and feedforward recurrent alignment with asymmetric recurrent network inclusive the influence of symmetry.** For different degree of symmetry $a$, different color dots are applied shown in legend. The green color with $a = 1.0$ illustrates the case of symmetric network. The most dark purple $a = 0.0$ represents the asymmetric network without any influence of symmetry. The x-axis shows the alignment score calculated with considered modifications and the y-axis indicates the trial-to-trial correlation calculated with aligned inputs through (**??**) **Left**: Result generated with modification 1 (only consider real part). **Right**: Result with modification 3 (symmetrized recurrent network).

The results (figure 1.7) point out that the trendy of positive correlation keeps while the network increases its asymmetry for both modifications 1 and 3.

However, when the network is fully asymmetric, the correlation has larger dispersion with modification 1 (only real part). We suspect that the reason is that in the case of full asymmetry, more information got lost because of only considering information from real part of aligned inputs. If there is certain degree of symmetry, a part of information during the alignment originates from the symmetric recurrent network and therefore for this part of alignment, no information get lost while only taking the real part of aligned inputs. Aligning to fully asymmetric recurrent network and only taking real part of aligned inputs, all neurons receive as a result incomplete information. Therefore, the correlation between evoked patterns will be disturbed mostly.

In the case of modification 3 (align inputs to symmetrized network), the correlation is almost maintained over all degree of symmetry.

**Intra-trial Stability**  Intra-trial stability is the averaged time-delayed activity correlation inside one single trial. As the name indicates, it shows how stable the information is represented inside one trial. With the result from symmetric recurrent network, we expect a similar exponential positive correlation as in figure 1.2 also

with the asymmetric recurrent network. The degree of symmetry should also not influence the result significantly since we assume that the hypothesis should also work well with asymmetric recurrent networks generally.
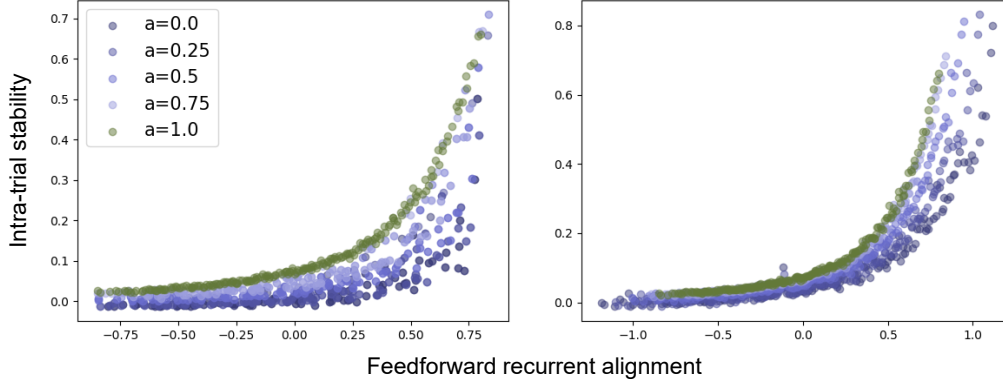


**Figure 1.8 Correlation between intra-trial stability and feedforward recurrent alignment with asymmetric recurrent network considering the influence from degree of symmetry.** For multiple degrees of symmetry $a$, different color dots are applied shown in legend. From complete symmetric $a = 1.0$ (as control) to full asymmetric recurrent network, the corresponding feedforward recurrent alignment (x-axis) is plotted against the intra-trial stability (y-axis). **Left**: Result with modification 1 (only real part of aligned inputs). **Right**: Result with modification 3 (align input to symmetrized network).

We suspect with the similar reason as in trial-to-trial correlation that for modification 1, the lost of imaginary part information of aligned inputs could lead to increased dispersion with larger asymmetry of network (figure 1.8 left panel).

For modification 3, except for the expected exponential positive correlation between intra-trial stability and feedforward recurrent alignment score, we also observe that the dispersion is more noticeable with increased asymmetry. So, the degree of symmetry influence here more significant than in trial-to-trial correlation. Furthermore, similar to modification 1, although align inputs to the symmetrized network, the lost of information also got lost with the increased asymmetry. Imagine if having a total symmetric network, the result of symmetrization is the network itself and therefore no information is lost in eigenvectors. However, if having a fully asymmetric recurrent network, after symmetrization, information that stored in complex eigenvectors get transformed to lower dimension real eigenvectors completely. Thus, when align inputs to the symmetrized asymmetric network, transformed information could be the reason for larger dispersion.

**Dimensionality** Dimensionality reflects the complexity of the information they encoded. A high dimensional activity pattern needs more orthogonal activity pattern for representation then a low dimensional activity pattern and thus indicates larger variability and higher complexity of contained information. Modification 1 and 3 change the eigenvectors for construction of input variance and the eigenvalues for calculation of effective dimensionality (section **??** eq. (**??**), (**??**)).

If both modifications work well, the results should be similar to result from symmetric recurrent network (figure 1.3b). Low degree of symmetry is allowed to lead to small range of dispersion, which however should still keep the tendency of decreased dimensionality with increased alignment score.
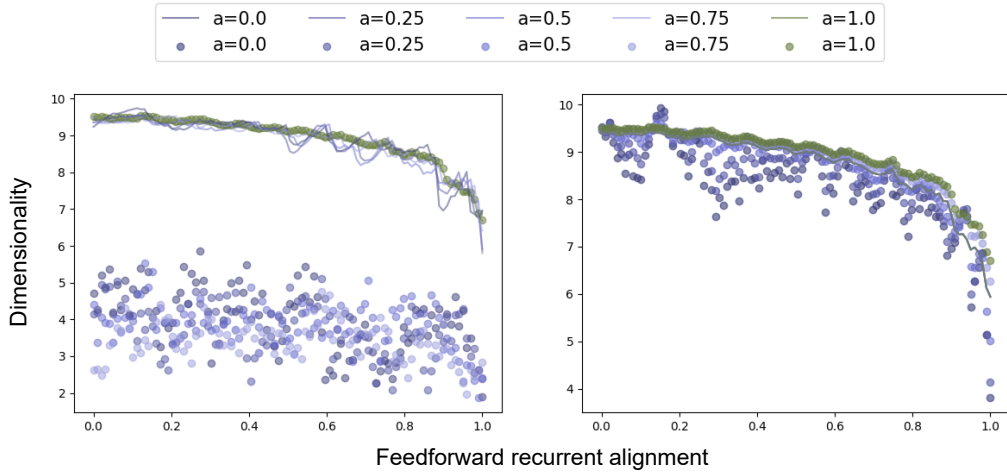


**Figure 1.9 Analytical and empirical effective dimensionality in correlation with feedforward recurrent alignment with asymmetric recurrent network. Impact on the correlation from the degree of symmetry in network.** The full symmetric recurrent network ($a = 1.0$) is control for other cases with $a < 1.0$. When $a = 0.0$, the recurrent network is fully asymmetric. The dots represents the empirical approximation for effective dimensionality (**??**). The lines are for the analytical calculation for dimensionality adjusted to modification 1 and 3 (section **??** eq. (**??**), (**??**)). **Left**: Result with modification 1 (only real part of aligned inputs and eigenvalues for analytical dimensionality). **Right**: Result with modification 3 (align inputs and to symmetrized recurrent network).

With modification 1 (figure 1.9 left panel), the empirical dimensionality differs a lot from the analytical calculation of dimensionality. Besides, there is no significant correlation between dimensionality and feedforward recurrent alignment with empirical data. The reason for that could be the lost of orthogonality between the real part of eiegenvectors, that is

$$e_i \perp e_j \;\nRightarrow \mathrm{Re}(e_i) \perp \mathrm{Re}(e_j) \; \forall i, j = 1, ..., n \,. \tag{1.10}$$

As a result, the covariance matrix $\Sigma^{\text{Dim}}$ (**??**) is not necessarily constructed with orthogonal vectors anymore, which contradicts the assumption of reflecting dimensionality by projecting activity pattern on orthogonal patterns that are also applied to construct covariance matrix.

With modification 3, the results fulfill largely the expectations: the negative correlation between dimensionality and feedforward recurrent alignment is kept in both analytical and empirical approximations. Little dispersion is due to the structural information lost when symmetrize the asymmetric recurrent network and align the inputs to symmetrized eigenvectors.

Thus, until this step, we would also drop the modification 1 and only further consider the modification 3.

**Alignment to spontaneous activity**   The alignment of evoked activity pattern to the endogenous pattern measures the overlap between their variance ratio curves explained by endogenous pattern (**??**), for example the overlaps between curves in figure 1.4a. Only modification 3 is left for the evaluation.

Similar to prior cases, if the modification works, we expect the correlation tendency between alignment to spontaneous activity and feedforward recurrent alignment score would also be kept as with symmetric recurrent network (figure 1.4b). The alignment to spontaneous activity is modified with (**??**) and calculate with (**??**).
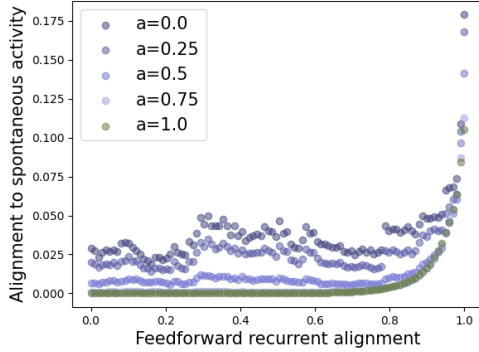


**Figure 1.10 The correlation between alignment to spontaneous activity and feedforward recurrent alignment considering influence from degree of symmetry in the recurrent network.** As control, fully symmetric recurrent network $a = 1.0$ is represented with dark green dots. For different degree of symmetry from $a = 0.75$ to 0, the darker the dot color, the more asymmetric is the recurrent network.

The dispersion increases with the degree of asymmetry due to the increased information lost, but the general tendency that alignment of evoked activity pattern to spontaneous activity becomes larger when the alignment of inputs to symmetrized recurrent network is larger. With high degree of symmetry, e.g. $a = 0.75$, the correlation is significantly similar to it with total symmetric recurrent network.

**Conclusion**   After evaluating the rest modifications 1 and 3 with four perspectives of response properties, the modification 1 is filtered out because of the lack of or-

thogonality between eigenvectors after modification. The modification 3 performs at the end the best and could generally fulfill the expectations of correlations between the response properties and feedforward recurrent alignment.

Therefore, the modification that align the inputs to symmetrized asymmetric recurrent network (modification 3) is a good candidate for modeling feedforward recurrent alignment hypothesis in asymmetric recurrent networks.

## 1.3 Modeling Feedforward Recurrent Alignment Hypothesis on Low-rank Recurrent Neural Networks (RNNs)

Although fully recurrent connectivity structure is one of the most popular of network models for theoretical neuroscience, there are experimental recordings suggesting that the transformation of sensory stimuli into motor outputs relies on low-dimensional dynamics at the population level [**?**]. Therefore, the low rank connectivity structure can be a good candidate to understand the neural activity better.

We hence also try to model the feedforward recurrent alignment hypothesis also on the low-rank RNNs to discover if the hypothesis modeling adapted from full rank RNNs also works with low-rank RNNs. Hereby, we consider both symmetric and asymmetric RNNs with two constructions described in section **??** by (**??**) and (**??**). Under symmetric or asymmetric condition, we go through both constructions with four response properties in correlation with feedforward recurrent alignment as for full rank RNNs: trial-to-trial correlation, intra-trial stability, dimensionality, and alignment to spontaneous activity. The method will be dropped out here if the expectations are not fulfilled and the rest of results are contained in the Appendix.

### 1.3.1 Evaluation of Feedforward Recurrent Alignment in symmetric Low-rank RNNs based on response properties considering different Constructions

We firstly consider symmetric low-rank networks in both constructions. With each case, the results of response property analysis based on the modeling of feedforward recurrent alignment hypothesis for symmetric networks are evaluated.

**Low-rank RNNs without random noise**  The formulation of low-rank RNNs is followed by **??** with the rank $G$ significantly smaller than the number of neurons $n$. The eigenvalues of symmetric low-rank RNNs are real numbers.
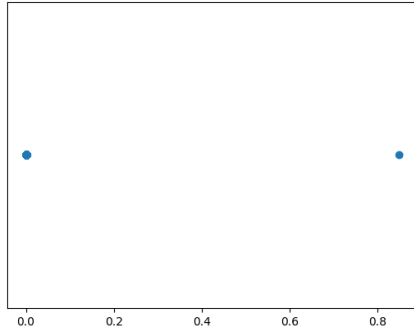


**Figure 1.11 Eigenvalue distribution of low-rank RNNs without random noise.** The eigenvalues of symmetric low-rank RNNs are real numbers (x-axis). With rank $G = 1 \ll n = 200$ the number of neurons, number of $G = 1$ eigenvalues takes the value of normalization factor $R = 0.85 < 1$ and the rest $n - G = 199$ number of eigenvalues equal 0.

As shown in figure **??**, only two values are presented in real eigenvalues. Number of ranks $G$ eigenvalues equal the normalization factor $R < 1$ that limits the value range (**??**). The rest of $n - G$ eigenvalues take the value 0.

The phenomenon of bi-polarized eigenvalue distribution is due to the construction of low-rank RNNs here. If considering the case of having the left connectivity vectors as orthonormal basis for construction of RNNs from (**??**), which is

$$J = \frac{1}{n} \sum_{g=1}^{G} l^{(g)} l^{(g)T} = \sum_{g=1}^{G} \frac{1}{n} l^{(g)} l^{(g)T} . \tag{1.11}$$

If the rank equals the number of neurons, $G = n$, the formulation of low-rank matrix (1.11) is at the same time a symmetric full rank matrix with eigenvectors $\left\{ l^{(g)} \right\}$ and all eigenvectors correspond to the same eigenvalue $\frac{1}{n}$. Since the eigenvalues are re-scaled by parameter $R < 1$ to enable the stable steady state , the eigenvalues are equal $R$ based on (**??**).

Now, if the rank $G$ is smaller than the number of neurons $n$, the formulation of low-rank RNN (1.11) can be rewritten as

$$J = \sum_{g=1}^{G} \frac{1}{n} l^{(g)} l^{(g)T} + 0 = \sum_{g=1}^{G} \frac{1}{n} l^{(g)} l^{(g)T} + \sum_{g=G}^{n-G} 0 l^{(g)} l^{(g)T} . \tag{1.12}$$

So, there are $G$ basis vectors that have eigenvalue $\frac{1}{n}$, which is further re-scaled to $R < 1$. The rest of $n - G$ eigenvectors have eigenvalue 0. This then results the eigenvalue distribution shown in figure 1.11.

We firstly look at the trial-to-trial correlation and except a positive correlation with the feedforward recurrent alignment. When align the inputs to the symmetric recurrent alignment, the inputs have the mean value matched to the eigenvectors of the network. Due to the feedforward recurrent alignment formulation, the alignment scores are the corresponding eigenvalues (**??**).

Since there are only two eigenvalues $R$ and 0 for the rank smaller than number of neurons $G < n$, we assume that there is no continuous correlation but two groups of trial-to-trial correlation values. However, low feedforward recurrent alignment should still correlated with small trail-to-trial correlation and large alignment score with big trial-to-trial correlation. We expect therefore here a discontinuous positive correlation between trial-to-trial correlation and feedforward recurrent alignment.

As the result (figure 1.12a) show, the trial-to-trial correlation distributes separately in two groups due to the number of eigenvalues. When the feedforward recurrent alignment is low, the trial-to-trial correlation concentrates also at the small number range. The trial-to-trial correlation increases when the feedforward recurrent alignment takes the larger eigenvalue.

18
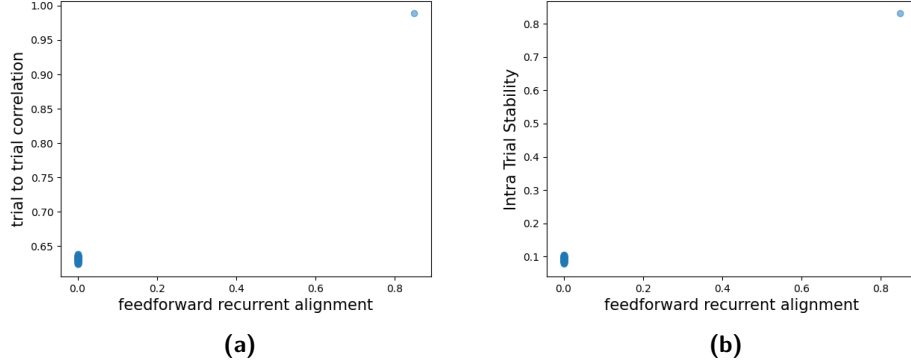
**Figure 1.12 Trial-to-trial correlation and intra-trial stability in correlation with feedforward recurrent alignment score for low-rank symmetric RNNs. (a)** Trial-to-trial correlation in relationship with feedforward recurrent alignment. **(b)** Intra-trial stability in relationship with feedforward recurrent alignment.
The positive correlation between feedforward recurrent alignment and trial-to-trial correlation **(a)** or intra-trial stability **(b)** is discontinuous due to the number of eigenvalues under low rank. If having rank $G = 1$, only one feedforward recurrent alignment score take the value $R = 0.85$, correlating with large trial-to-trial correlation **(a)** or intra-trial stability **(b)**. The rest concentrates at alignment score 0 with low trial-to-trial correlation **(a)** or low intra-trial stability **(b)**.

Analogous to the trial-to-trial correlation, we suspect the result for intra-trial stability also be a discontinuous positive correlation to feedforward recurrent alignment. Due to the same reason of existing only two eigenvalues $R$ and 0, the intra-trial stability value distributed also separately in two groups while keeping the positive correlation with feedforward recurrent alignment. The result (figure 1.12b) confirmed our assumption.

Although the positive correlation of feedforward recurrent alignment with both trial-to-trial correlation and intra-trial stability is kept. Since the simple distribution of eigenvalues leads to trivial dynamics between response properties and feedforward recurrent alignment as well as some unwanted effects in dimensionality (Appendix), we would rather abandon the formulation of low-rank RNNs for trying modeling feedforward recurrent alignment hypothesis.

**Low-rank RNNs with random noise**   Low-rank symmetric RNNs with random noise includes extra part of symmetrized Gaussian distributed matrix (**??**, section **??**). If analog to (1.11) considering the right connectivity vectors equal the left connectivity vectors, the symmetric low-rank RNNs with random noise can be formulated with a part of symmetric low-rank RNNs and a part of symmetrized Gaussian

distributed matrix $J_{\text{rand}}$.

$$J = \frac{1}{n} \sum_{g=1}^{G} l^{(g)} l^{(g)T} + J_{\text{rand}} \,. \tag{1.13}$$

The symmetric random part $J_{\text{rand}}$ should provide more dynamics in the network.
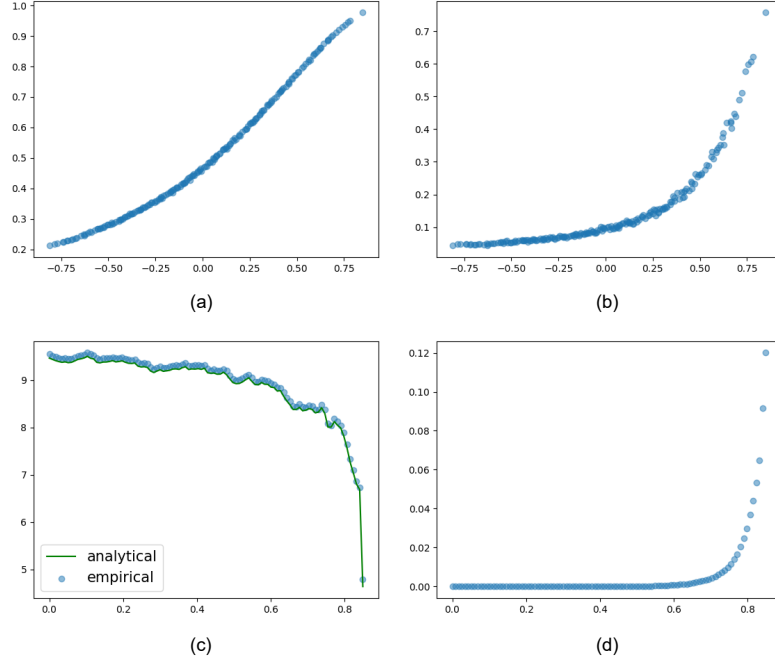


**Figure 1.13 Correlation between response properties and feedforward recurrent alignment in low-rank RNNs with random noise.** Besides the part of a low-rank matrix with rank $G$, RNNs with noise includes an extra part of symmetrized random Gaussian matrix (1.13). The rank of low rank part has the rank $G = 1$.
**(a)** Trial-to-trial correlation (y-axis) against feedforward recurrent alignmnet score (x-axis). **(b)** Intra-trial stability (y-axis) against feedforward recurrent alignment score (x-axis). **(c)** Dimensionality (y-axis) in dependence of feedforward recurrent alignment score (x-axis). **(d)** Alignment to spontaneous activity (y-axis) against feedforward recurrent alignment (x-axis).

Since $J_{\text{rand}}$ is a full rank symmetric matrix, the final recurrent network $J$ is also a full rank symmetric matrix in spite of the low-rank symmetric part with rank $G$. Therefore, the correlations between the response properties and feedforward recurrent alignment score are similar to the case of symmetric matrix (figure 1.1, 1.2, 1.3b, and 1.4b).

Therefore, the feedforward recurrent alignment hypothesis can be modeled. The expected relationships between feedforward recurrent alignment and response properties (section 1.1) are fulfilled. However, the influence of low-rank construction is therefore overwritten by the random noise, which leads to the final dynamics not much different from the case with general random symmetric RNNs.

**Conclusion**  Considering symmetric low-rank RNNs without random noise, the feedforward recurrent alignment keeps the positive correlation to trial-to-trial correlation and intra-trial stability. However, the network construction leads to simple distribution of eigenvalues (1.11) and in turn also the straightforward distribution of feedforward recurrent alignment score. As a result, only a simple discontinuous positive correlation of feedforward recurrent alignment to trial-to-trial correlation and intra-trial stability can be observed (figure 1.12).

If adding a random noise to disturb the simple dynamics of low-rank RNNs, the final construction (1.13) is a full rank symmetric RNN. As a result, the correlations of feedforawrd recurrent alignment to response properties are similar to the results observed in general symmetric RNNs. Although the expected phenomenons are seen, the effect of low-rank RNNs cannot be significantly detected.

After considering the symmetric low-rank RNNs with and without random noise, although the feedforward recurrent alignment hypothesis can be modeled in both cases. The relationships between feedforward recurrent alignment and response properties meet also the expectations. But either can have significant complex dynamics. Besides, the influence of the rank is also not observable.

### 1.3.2  Different Constructions influence the impact of rank in asymmetric Low-rank RNNs based on response properties

Asymmetric low-rank RNNs can have more complex dynamics than symmetric low-rank RNNs. Section **??** also introduces the construction of asymmetric low-rank RNNs with or without random noise. Generally, if taking the set of left connectivity vectors different from the set of right connectivity vectors, asymmetric matrices can be formulated (**??**), (**??**).

Since asymmetric low-rank RNNs are only a certain case of asymmetric RNNs with specific constructions, they also have the problem of complex eigenvectors and eigenvalues when modeling the feedforward recurrent alignment hypothesis. Therefore, from the previous results with general asymmetric RNNs, we choose the modification of aligning inputs to symmetrized RNN because of the overall best performance (**??**).

**Low-rank RNNs without random noise**

**Low-rank RNNs with random noise**

**Conclusion**