Biologically plausible heavy-tailed connectivity enhances generalizations on cognitive tasks in recurrent neural networks

Zhe Jiao [1] Xiaodong He [1] Shanglin Zhou [2]

[1]Northwestern Polytechnical University, China [2]Fudan University, China

Problem

While heavy-tailed synaptic weight distributions are pervasive in biological neural networks, their computational role—particularly in relation to generalization—remains poorly understood.

1 Train RNNs with heavy-tailed connectivity

To address this, we develop a novel optimal-transport-based optimization algorithm that incorporates key **biologically plausible features**: (C1) Non-reciprocal interactions and no self-coupling; (C2) Dale's principle; (C3) Heavy-tailed connectivity, to train recurrent neural networks (RNNs) on a wide range of cognitive tasks, including Cue-Set-Go (CSG) task, Measure-Wait-Go (MWG) task, Flip-flop task, Mante task, Romo task and perceptual decision making task.

Algorithm 1 Asymptotic DSco-SGD

Input: initialization of the weight matrix (J^E, J^I), task trials $\{u^{(q)}, \hat{z}^{(q)}\}_{q=1,\dots,N_{\mathrm{tr}}}$
Parameter: network size N, learning rate (η_1, η_2), max iteration, task parameters
Output: readout z of the network

Sign-condition = False
while $z^{(q)} \neq \hat{z}^{(q)}$ *for some trial q* **do**
  compute empirical risk $\mathcal{L}(J)$.
  step along the gradient:
  $J_n^E \leftarrow J_n^E - \eta_1^E \nabla\mathcal{L}, \quad J_n^I \leftarrow J_n^I - \eta_1^I \nabla\mathcal{L}$
  if not Sign-condition **then**
    if $\frac{\sharp\{J_n^E \neq 0\}}{\sharp\{J_{n-1}^E \neq 0\}}$ and $\frac{\sharp\{J_n^I \neq 0\}}{\sharp\{J_{n-1}^I \neq 0\}}$ belong to $(1-\iota, 1)$ given a sufficiently small $\iota > 0$ **then**
      Sign-condition ← True
    end if
  end if
  if Sign-condition **then**
    $\theta_n^E \leftarrow \mathrm{MLE}\{J_n^E\}, \theta_n^I \leftarrow \mathrm{MLE}\{J_n^I\}$
    $\{\hat{J}_n^E\} \leftarrow \rho^m(\theta_n^E), \{\hat{J}_n^I\} \leftarrow \rho^I(\theta_n^I)$
    sort the weights in ascending order:
    $\{\hat{J}_{(n)}^E : \hat{J}_{(1)}^E \leq \cdots \leq \hat{J}_{(fN)}^E\} \leftarrow \{\hat{J}_n^E\}$
    $\{\hat{J}_{(n)}^I : \hat{J}_{(1)}^I \leq \cdots \leq \hat{J}_{((1-f)N)}^I\} \leftarrow \{\hat{J}_n^I\}$
    $\{J_{(n)}^E : J_{(1)}^E \leq \cdots \leq J_{(fN)}^E\} \leftarrow \{J_n^E\}$
    $\{J_{(n)}^I : J_{(1)}^I \leq \cdots \leq J_{((1-f)N)}^I\} \leftarrow \{J_n^I\}$
    step along the optimal transport map:
    $J_{(n)}^E \leftarrow J_{(n)}^E - \eta_2^E (J_{(n)}^E - \hat{J}_{(n)}^E)$
    $J_{(n)}^I \leftarrow J_{(n)}^I - \eta_2^I (J_{(n)}^I - \hat{J}_{(n)}^I)$
  end if
  if *exceeds max iteration* **then**
    break
  end if
end while

The RNNs used to implement cognitive tasks consisted of N units, with the **dynamics** of the total input current x_i to the i-th unit are governed by

$$\tau \dot{x}_i(t) = -x_i(t) + \sum_{j=1}^{N} J_{ij}\varphi(x_j(t)) + \sum_{k=1}^{N_{\mathrm{in}}} I_{ik}u_k + \xi_i.$$

Each readout of the network is a linear combination of the firing rates of all units along the output vector.

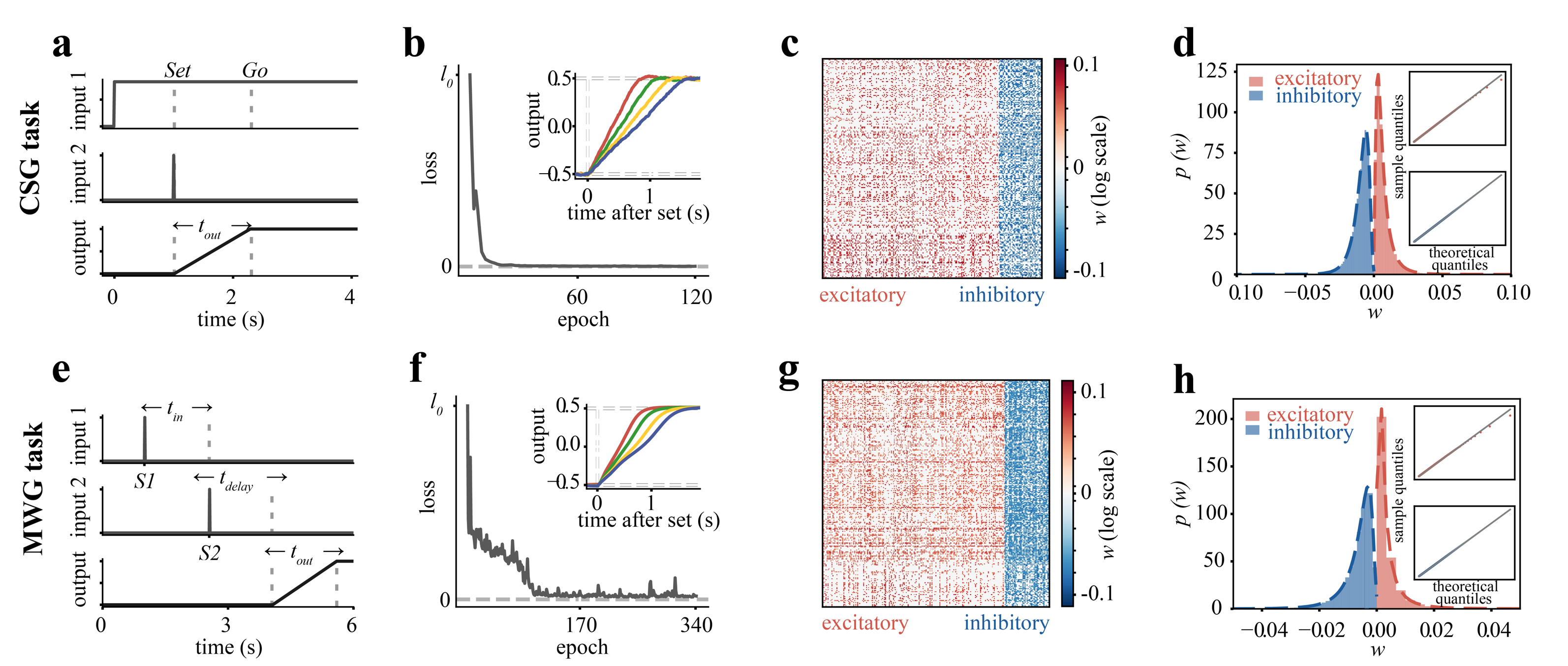


Figure 1. (a, e) Structures of the tasks. (b, f) Loss function with respect to the epochs. Here, l_0 means the initial loss value. Inset: Output of heavy-tailed RNNs on four training samples. (c, g) Heat map of neural connectivity matrices. (d, h) Weight distributions of excitatory and inhibitory units and the corresponding Quantile-Quantile plots for the insets.

2 Heavy-tailed connectivity improves generalization

To assess the influence of heavy-tailed connectivity on generalization, we compared the performance of RNNs trained with heavy-tailed connectivity on CSG and MWG tasks, with that of four baseline models.

RNN	Connectivity	C1	C2	C3	References
Unconstrained	/	✗	✗	✗	Haykin, 2009
l^1-regularization	Sparse	✗	✗	✗	Haykin, 2009
Low-rank	Spiked-covariance	✗	✗	✗	Mastrogiuseppe and Ostojic, 2018
E-I	Dale's principle	✓	✓	✗	Song, Yang and Wang, 2016
Heavy-tailed	Log-Normal	✓	✓	✓	This work

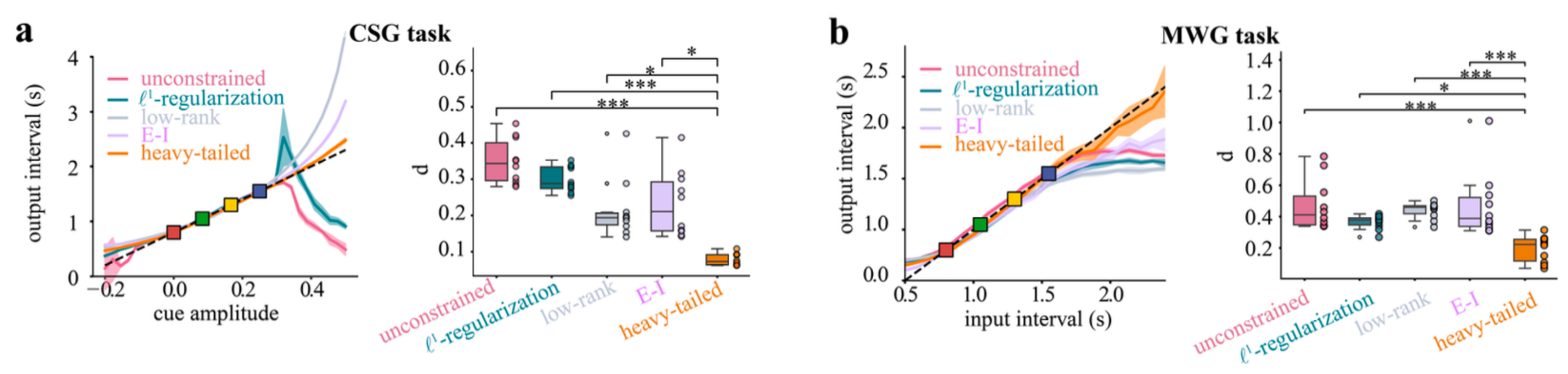


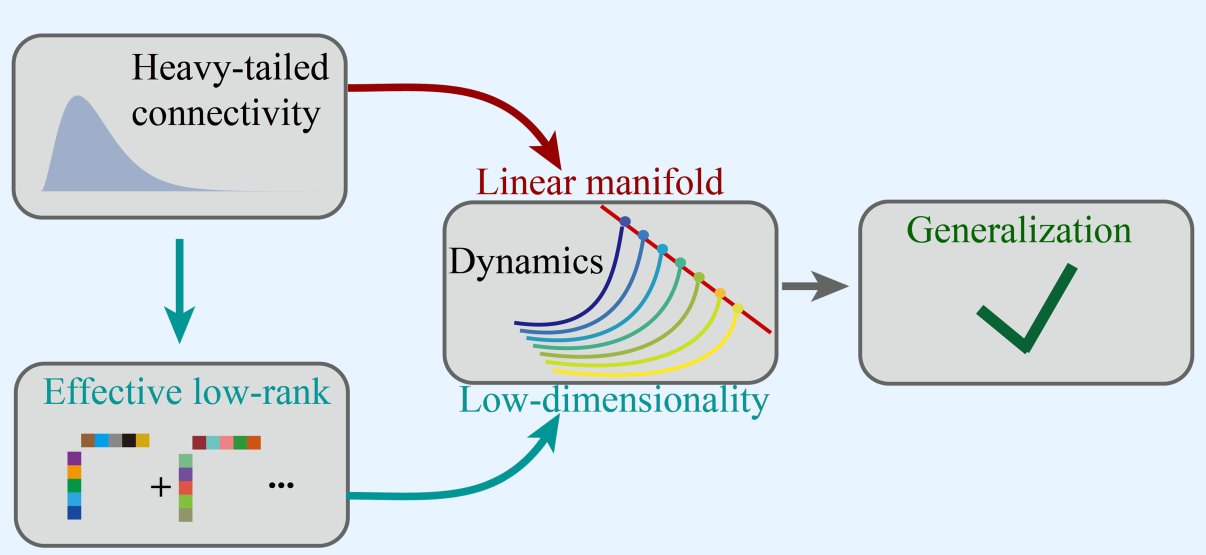
Figure 2. The dashed line denotes the linear target for the input interval, and proximity to the dashed line corresponds to superior generalization capability. The distances d to the dashed lines computed to quantify generalization. Dunn's test to show statistically significant difference, in which $*$ means p-value < 0.05, $**$ means p-value < 0.01, and $***$ means p-value < 0.001.

Highlights

Our theoretical analysis and numerical experiments reveal two complementary mechanisms underlying this generalization enhancement.

- Topologically, heavy-tailed connectivity induces an effectively low-rank structure, which in turn yields low-dimensional neural dynamics.
- Geometrically, heavy-tailed connectivity intrinsically shapes task variable representations to lie near a linear manifold, thereby improving generalization for a linear readout strategy.

Together, these results identify heavy-tailed connectivity as a biologically grounded intrinsic mechanism that promotes low-rank structure and favorable representational geometry, leading to improved generalization in flexible cognitive tasks.



How heavy-tailed connectivity enhances generalization.

3 Heavy-tailed connectivity induces low-dimensional dynamics through effective low-rank structure

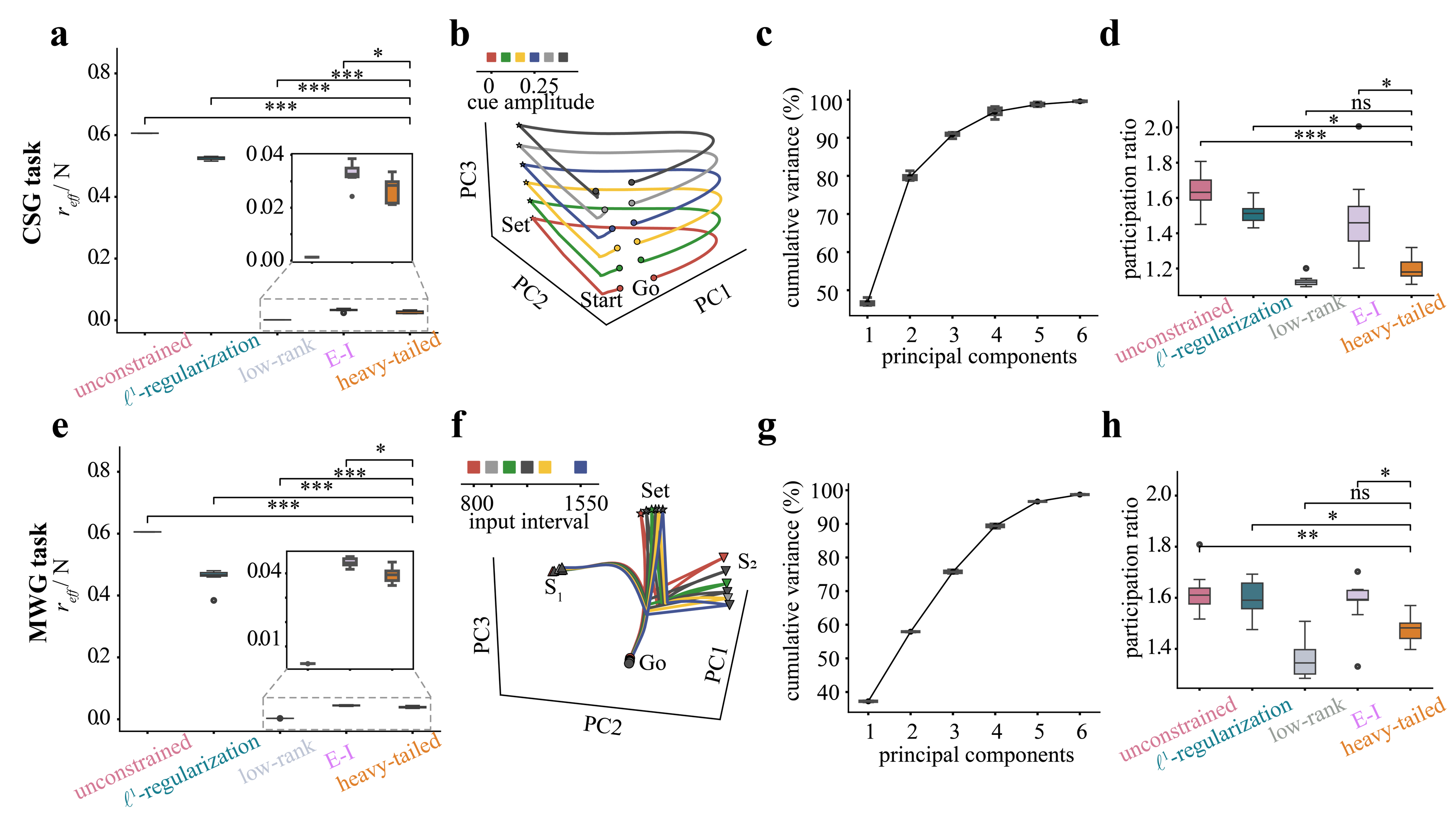


Figure 3. (a, e) **Effective ranks** of different neural connections with mean and SD over 10 trained RNNs. (b, f) Neural activities generated by the heavy-tailed RNNs in the subspace spanned by the three leading principal components in the standard principal component analysis (PCA). (c, g) Cumulative percentage of explained variance in the PCA. (d, h) **Participation ratio** of activity during 'Set' and 'Go', with mean and SD over 10 trained RNNs. Wilcoxon rank-sum test: $*$ means p-value < 0.05, $**$ means p-value < 0.01, $***$ means p-value < 0.001.

4 Heavy-tailed connectivity induces linear manifold representations

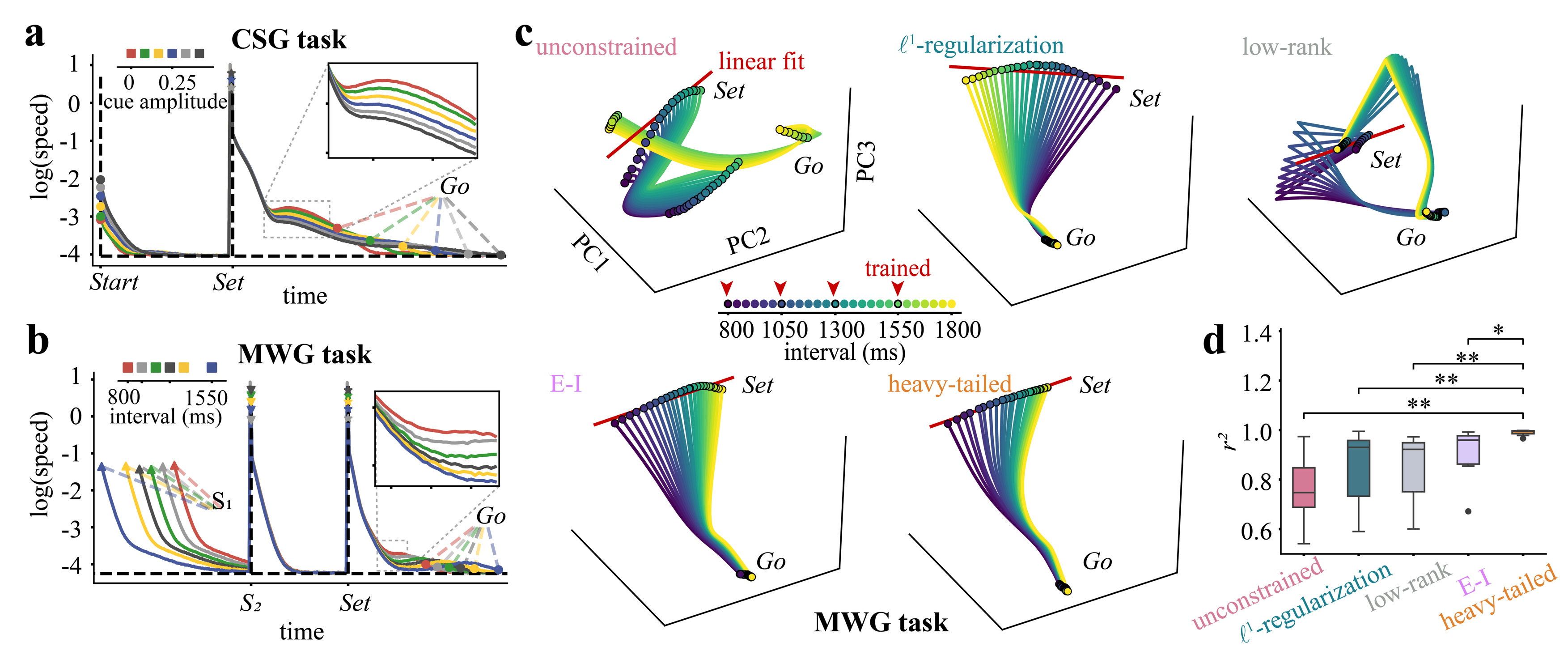


Figure 4. (a, b) The speed during the 'Set' and 'Go' signals is scaled by the cue amplitude (CSG task) or input interval (MWG task), even beyond the training range. (c) The dynamics in the embedding subspace of trained RNNs with different connectivities during the 'Set' and 'Go' signals exhibit distinct manifold geometry. Red lines denote the linear fits of the states at the 'Set' phase. (d) Quantification of the linear representational geometry at the 'Set' phase. r^2 is the coefficient of determination of the linear fits, i.e. the square of the Pearson coefficient, with mean and SD over 10 trained RNNs. Wilcoxon rank-sum test: $*$ means p-value < 0.05, $**$ means p-value < 0.01, $***$ means p-value < 0.001.

Email: zjiao@nwpu.edu.cn and zhoushanglin@fudan.edu.cn