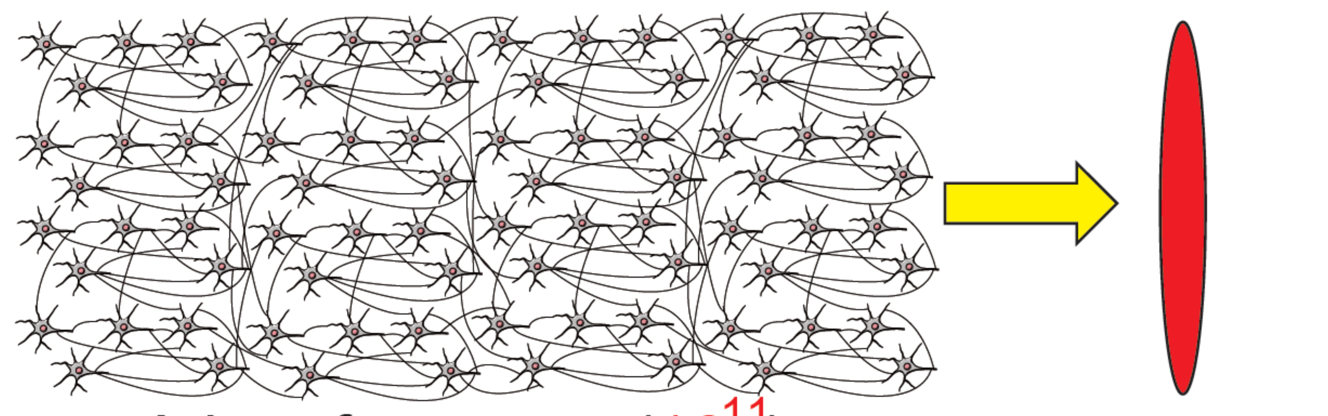
Ⅱ-27 Maximization of learning speed in motor cortex due to neuron redundancy

THE UNIVERSITY OF TOKYO
Ken Takiyama[A] and Masato Okada[B]
A … The University of Tokyo, B … RIKEN



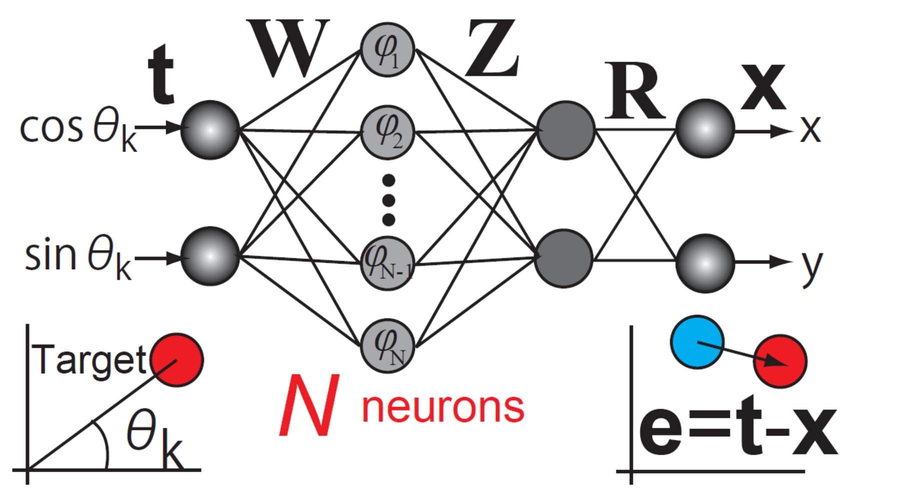
What are functional roles of neuron redundancy ?

1. Introduction

Motor system involves many kinds of redundancies: kinematic, muscle, and neuron redundancy. Many studies have investigated functional roles of kinematic[1] and muscle[2] redundancies. However, there remains a question as to *what are functional roles of neuron redundancy*. Our analysis on redundancy neural network model suggests that *one of the roles is to maximize learning speed* in motor learning.

2. Model

We analyze a linear rate model that can reproduce neurophysiological data and can be easily analyzed.[3]



Neural activity
$$A^t = W^t t$$
Planning force
$$F^t = ZA^t$$
Output (*R*…rotation)
$$x^t = RF^t$$

The encoder *W* is adaptable. $W^{t+1} = W^t + NB\frac{\partial E^t}{\partial W^t}$

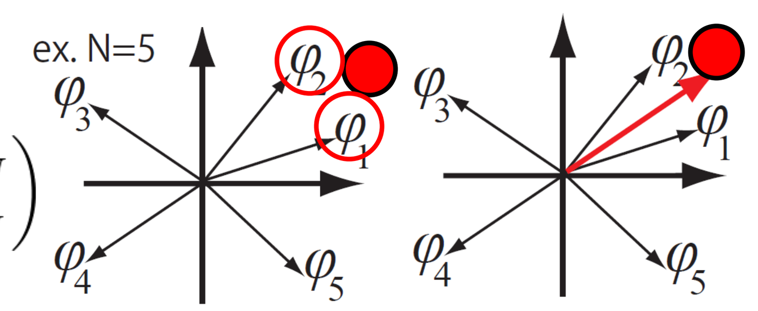
Task…***t*** = ***x*** (2-dim). Adaptable ***W***…2*N*-dim.

Learning process does not explicitly depend on N.

The decoder **Z** is fixed.

$$Z = \frac{1}{N}\begin{pmatrix}\cos\varphi_1 & \cos\varphi_2 & \dots & \cos\varphi_N \\ \sin\varphi_1 & \sin\varphi_2 & \dots & \sin\varphi_N\end{pmatrix}$$

φ…uniform distribution.



3. Neuron redundancy maximizes learning speed

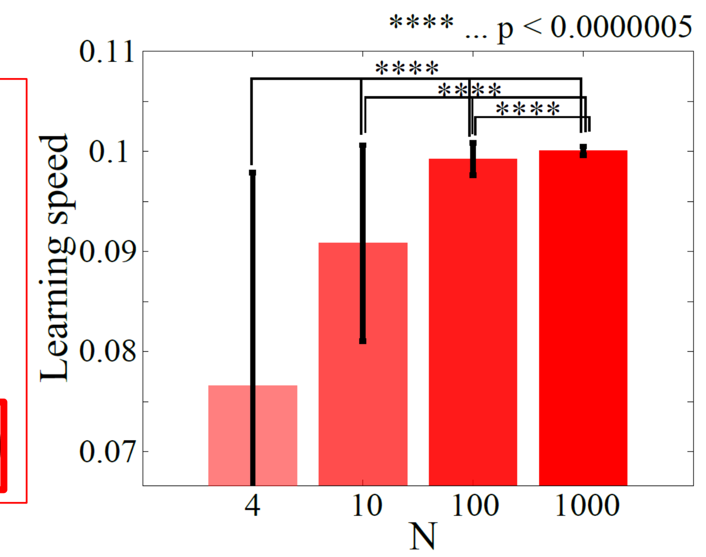
Learning curve can be calculated as (see Appendix in detail)

$$E^{t+1} = \frac{1}{2}(t - x^{t+1})^T(t - x^{t+1}) = \frac{1}{2}(t - x^t)^T\Lambda(t - x^t) = \frac{1}{2}(\lambda_1^t u_1 + \lambda_2^t u_2)$$

The condition to maximize learning speed is $\frac{1}{N}\sum_{i=1}^N \cos\varphi_i \sin\varphi_i = 0$.

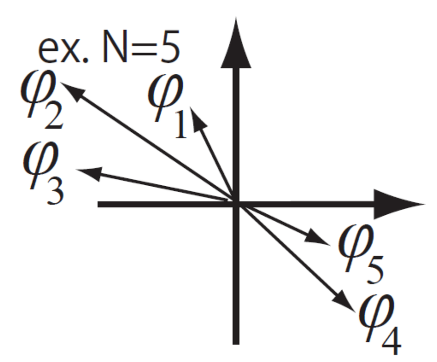
Thus, the necessary and sufficient condition to maximize learning speed is *neuron redundancy* because it lets the neural network be self-averaging,[4] that is

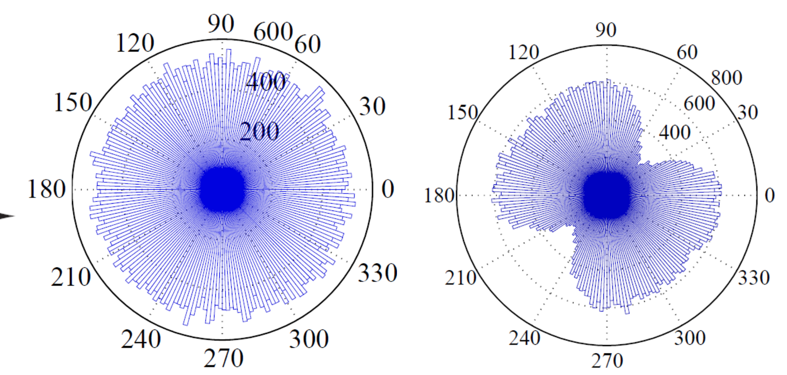
$$\frac{1}{N}\sum_{i=1}^N \cos\varphi_i \sin\varphi_i = \int d\varphi P(\varphi)\cos\varphi\sin\varphi = 0$$



4. Consistency with neurophysiological experiment

Generally, the necessary and sufficient conditions to maximize learning speed are *neuron redundancy* and $\mathrm{Cov}(Z_1, Z_2) + \langle Z_1\rangle\langle Z_2\rangle = 0$





Histograms of preferred direction when P(Z) satisfies the above constraint.

Some studies suggested the distribution of PD is skewed and bimodal.[5]
Other studies suggested the distribution of PD is uniform.[6]
Our hypothesis is consistent with both results.

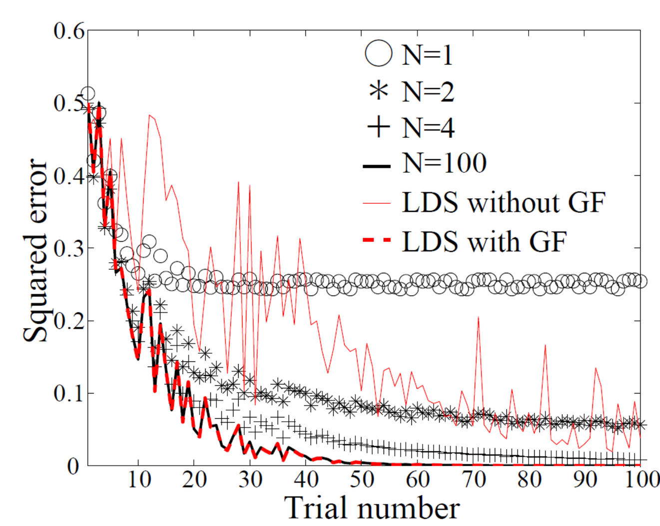
5. Consistency with behavioral experiment

A linear dynamical system (LDS) can explain results of behavioral experiments.[7]

$$x^{t+1} = x^t + B'e^t$$

Neuron redundancy yields the update rule of the neural network as

$$x^{t+1} = x^t + B\zeta e^t \qquad \zeta = (Z_1) + (\langle Z_1\rangle)^2$$



Because the equivalence is ensured only by neuron redundancy, our hypothesis is consistent with results of behavioral experiments.

6. Construction of optimal perturbation

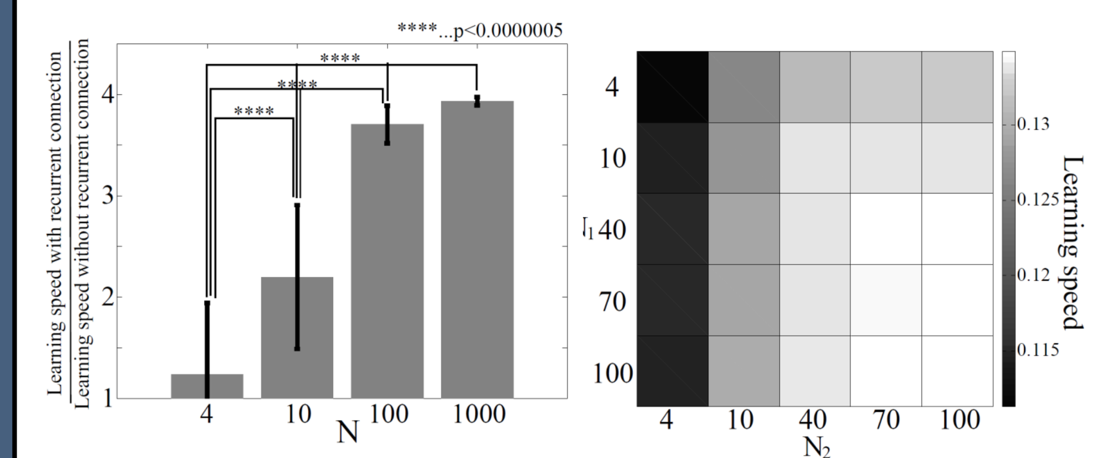
Our hypothesis yields the condition where the perturbation $R = \begin{pmatrix} a & b \\ c & d\end{pmatrix}$ maximizes learning speed: *ac*+*bd*=0. If learning speed is the same in the x- and y- coordination, the optimal perturbations are

$$R = \begin{pmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix} \qquad R = \begin{pmatrix}\cos\theta & \sin\theta \\ \sin\theta & -\cos\theta\end{pmatrix}$$

which coincide with often-used rotational and saddle perturbation.

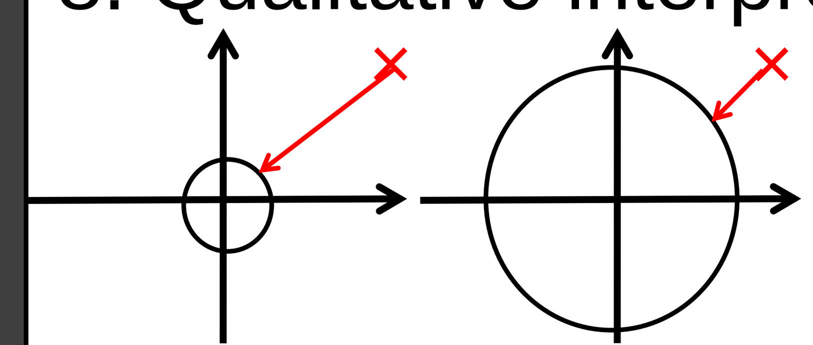
7. Analysis of biologically plausible network

These analysis confirmed that our hypothesis is invariant if a neural network includes recurrent connections[7] or two-layer structure.[8]



Left: ratio of learning speed between with and without recurrent connections.
Right: learning speed when the 1st and the 2nd layer include N1 and N2 neurons.

8. Qualitative interpretation of our hypothesis



There are only two equations to be satisfied: ***t***=***x***. Adaptable ***W*** are 2*N* dimensional. Thus, (2N-2)-dimensional subspace of ***W*** yields ***t***=***x***.

The more redundant a neural network becomes, the more the fraction of the subspace grows; (2N-2)/2N→1. Since neuron redundancy shortens the distance between an initial value of ***W*** and the subspace, learning speed gets faster.

9. Summary

・Neuron redundancy maximizes learning speed.

・Our hypothesis is consistent with the results of neurophysiological and behavioral experiments.

・Our hypothesis is invariant in biologically plausible network models.