Cerebellar-driven cortical dynamics enable task acquisition, switching and consolidation Supplementary note 1

Contents

1 Number of granule cells needed is inversely proportional to square root of signal-to-noise ratio

1.1 Problem statement

The problem can be summarised as follows: given a particular ratio of task-dependent and task-independent input - i.e. *signal-to-noise ratio* (SNR) - what is the number of granule cells (GCs) required to ensure (with reasonable confidence) that the population vector will have distinct activations over the different task conditions?

Formally, we make the assumption that the input a to the GC population can be expressed as a sum of two Gaussians of zero mean

$$a = \omega + \zeta; \;\; \omega \sim \mathcal{N}(0, \sigma_\omega^2), \; \zeta \sim \mathcal{N}(0, \sigma_\zeta^2) \tag{1}$$

Where we assume ω might be "general" input (which comes from the task-agnostic environment) and ζ is task specific. We assume ω and ζ to be independent, and therefore a is also normally distributed. To avoid confusion we clarify that Equation 1 describes the input across all neurons, i.e. the total input distribution. For a given neuron we will have one sample of this and ω will be fixed given the same task-agnostic environment, and ζ will be fixed given the same task-specific condition. Note that we use the notation a, ω, ζ in place of I_{GC}, I_ω, I_ζ in Equation 17 of the main methods section to avoid clutter.

We make the assumption that each granule cell uses a spiking activation function f_{spike} where $f_{\text{spike}}(I) = 1$ if $I > 0$ and 0 otherwise. To obtain a task-encoding granule cell we would therefore like to know, assuming the same general input ω, what is the chances that two random samples of ζ produce two values of a of different signs. Assuming a threshold of 0, this is like asking what are the chances of a random neuron spiking for exactly one of the two task specific inputs. We label this case *unique spike.* We first derive an expression for the probability of a unique spike, before then considering how many neurons we would need to ensure with reasonable confidence for a unique spike to occur.

1.2 Derivation of P(unique spike)

Without loss of generality we focus on the case whereby there is no spike for task condition 1 and a spike for task condition 2. Assume that the same general input ω applies to each condition. The probability of this occurrence can then be expressed as

$$P(\text{no spike then spike}) = P(\omega + \zeta_1 < 0 \cup \omega + \zeta_2 > 0); \;\; \omega \sim \mathcal{N}(0, \sigma_\omega^2), \; \zeta_1, \zeta_2 \sim \mathcal{N}(0, \sigma_\zeta^2)$$

Where ζ_1, ζ_2 denote the task-specific input for task 1 and task 2 respectively. We apply the sum rule to the different possible values of this general input ω and solve as follows

$$\begin{aligned}
P(\text{no spike then spike}) &= \int p(\omega = x) P(x + \zeta_1 < 0) P(x + \zeta_2 > 0) dx \\
&= \int p(\omega = x) P(\zeta_1 < -x) P(\zeta_2 > -x) dx \\
&= \int p(\omega = x) P(\zeta_1 < x) P(\zeta_2 < -x) dx \\
&= \int p(\omega = x) \frac{1}{2}[1 + \operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}})] \frac{1}{2}[1 + \operatorname{erf}(\frac{-x}{\sigma_\zeta \sqrt{2}})] dx \\
&= \frac{1}{4} \int p(\omega = x) \Big(1 + \operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}}) + \operatorname{erf}(\frac{-x}{\sigma_\zeta \sqrt{2}}) + \operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}}) \operatorname{erf}(\frac{-x}{\sigma_\zeta \sqrt{2}}) \Big) dx \\
&= \frac{1}{4} \int p(\omega = x) dx + \frac{1}{4} \int p(\omega = x) \left[\operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}}) + \operatorname{erf}(\frac{-x}{\sigma_\zeta \sqrt{2}}) \right] dx \\
&+ \frac{1}{4} \int p(\omega = x) \operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}}) \operatorname{erf}(\frac{-x}{\sigma_\zeta \sqrt{2}}) dx
\end{aligned}$$

Where $p(\omega)$ is the probability density function (pdf) of ω and erf denotes the error function. By the nature of pdfs, for the first term in the last expression above we have

$$\int p(\omega = x) dx = 1$$

For the second term we use the fact that the error function is odd, i.e. $\operatorname{erf}(-y) = -\operatorname{erf}(y)$.

$$\begin{aligned}
\int p(\omega = x)[\operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}}) + \operatorname{erf}(\frac{-x}{\sigma_\zeta \sqrt{2}})] dx &= \int p(\omega = x)[\operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}}) - \operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}})] dx \\
&= \int p(\omega = x)[0] dx \\
&= 0
\end{aligned}$$

Finally, for the third term we take $y = \frac{x}{\sigma_\zeta \sqrt{2}}$ and apply the approximation[1]:

$$\operatorname{erf}^2(y) = 1 - \exp\left(-\frac{\pi^2}{8} y^2 \right) + \varepsilon(y)$$

For some small error term $\varepsilon(y)$. This yields

$$\begin{aligned}
\int p(\omega = x) \operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}}) \operatorname{erf}(\frac{-x}{\sigma_\zeta \sqrt{2}}) dx &= -\int p(\omega = x) \left(\operatorname{erf}(\frac{x}{\sigma_\zeta \sqrt{2}}) \right)^2 dx \\
&\approx -\int p(\omega = x) \left[1 - \exp\left(-\frac{\pi^2}{8} (\frac{x}{\sigma_\zeta \sqrt{2}})^2 \right) \right] dx \\
&= -\int p(\omega = x) dx + \int p(\omega = x) \exp\left(-\frac{1}{2} \left(\frac{\pi}{2\sqrt{2}\sigma_\zeta} \right)^2 x^2 \right) dx \\
&= -1 + \int p(\omega = x) \exp\left(-\frac{1}{2} \left(\frac{\pi}{2\sqrt{2}\sigma_\zeta} \right)^2 x^2 \right) dx
\end{aligned}$$

We compute the term inside the integral above and see it is actually proportional to the product of two Gaussian pdfs:

[1] `https://math.stackexchange.com/questions/1172474/erf-squared-approximation`

$$\begin{aligned}\int p(\omega = x)\exp\left(-\frac{1}{2}\left(\frac{\pi}{2\sqrt{2}\sigma_\zeta}\right)^2 x^2\right)dx &= \frac{1}{\sigma_\omega\sqrt{2\pi}}\int \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_\omega}\right)^2 x^2\right)\exp\left(-\frac{1}{2}\left(\frac{\pi}{2\sqrt{2}\sigma_\zeta}\right)^2 x^2\right)dx\\ &= \frac{1}{\sigma_\omega\sqrt{2\pi}}\int \exp\left(-\frac{1}{2}\left(\left(\frac{1}{\sigma_\omega}\right)^2 + \left(\frac{\pi}{2\sqrt{2}\sigma_\zeta}\right)^2\right)x^2\right)dx\\ &= \frac{1}{\sigma_\omega\sqrt{2\pi}}\int \exp(-ax^2)dx\\ &= \frac{1}{\sigma_\omega\sqrt{2\pi}}\sqrt{\frac{\pi}{a}}\end{aligned}$$

Where

$$\begin{aligned}a &= \frac{1}{2}\left(\left(\frac{1}{\sigma_\omega}\right)^2 + \left(\frac{\pi}{2\sqrt{2}\sigma_\zeta}\right)^2\right)\\ &= \frac{1}{2}\left(\frac{1}{\sigma_\omega^2} + \frac{\pi^2}{8\sigma_\zeta^2}\right)\\ &= \frac{1}{2}\left(\frac{8\sigma_\zeta^2}{\sigma_\omega^2 8\sigma_\zeta^2} + \frac{\pi^2\sigma_\omega^2}{8\sigma_\zeta^2\sigma_\omega^2}\right)\\ &= \frac{1}{2}\left(\frac{8\sigma_\zeta^2 + \pi^2\sigma_\omega^2}{8\sigma_\zeta^2\sigma_\omega^2}\right)\end{aligned}$$

So

$$\begin{aligned}\frac{1}{\sigma_\omega\sqrt{2\pi}}\sqrt{\frac{\pi}{a}} &= \frac{1}{\sigma_\omega\sqrt{2\pi}}\sqrt{\pi}\sqrt{2}\sqrt{\frac{8\sigma_\zeta^2\sigma_\omega^2}{8\sigma_\zeta^2 + \pi^2\sigma_\omega^2}}\\ &= \frac{1}{\sigma_\omega}\sqrt{\frac{8\sigma_\zeta^2\sigma_\omega^2}{8\sigma_\zeta^2 + \pi^2\sigma_\omega^2}}\\ &= \frac{2\sqrt{2}\sigma_\zeta}{\sqrt{8\sigma_\zeta^2 + \pi^2\sigma_\omega^2}}\\ &= \frac{\sigma_\zeta}{\sqrt{\sigma_\zeta^2 + \frac{\pi^2}{8}\sigma_\omega^2}}\end{aligned}$$

Combining these results finally produces the approximation

$$\begin{aligned}P(\text{no spike then spike}) &= \frac{1}{4}\int p(\omega = x)dx + \frac{1}{4}\int p(\omega = x)\left[\operatorname{erf}(\frac{x}{\sigma_\zeta\sqrt{2}}) + \operatorname{erf}(\frac{-x}{\sigma_\zeta\sqrt{2}})\right]dx\\ &+ \frac{1}{4}\int p(\omega = x)\operatorname{erf}(\frac{x}{\sigma_\zeta\sqrt{2}})\operatorname{erf}(\frac{-x}{\sigma_\zeta\sqrt{2}})dx\\ &\approx \frac{1}{4}[1] + \frac{1}{4}[0] + \frac{1}{4}\left[-1 + \frac{\sigma_\zeta}{\sqrt{\sigma_\zeta^2 + \frac{\pi^2}{8}\sigma_\omega^2}}\right]\\ &= \frac{1}{4}\sqrt{\frac{\sigma_\zeta^2}{\sigma_\zeta^2 + \frac{\pi^2}{8}\sigma_\omega^2}}\end{aligned}$$

We can then use this probability to compute the more general case that there is exactly one spike for the two different task conditions

$$\begin{aligned} P(\text{unique spike}) &= P(\text{no spike then spike}) + P(\text{spike then no spike}) \\ &= 2P(\text{no spike then spike}) \\ &\approx \frac{1}{2}\sqrt{\frac{\sigma_\zeta^2}{\sigma_\zeta^2 + \frac{\pi^2}{8}\sigma_\omega^2}} \end{aligned} \tag{2}$$

Let's suppose $\sigma_\zeta^2 = r\sigma_\omega^2$ for some $r > 0$, then this turns into

$$\begin{aligned} \frac{1}{2}\sqrt{\frac{\sigma_\zeta^2}{\sigma_\zeta^2 + \frac{\pi^2}{8r}\sigma_\zeta^2}} &= \frac{1}{2}\sqrt{\frac{\sigma_\zeta^2}{\sigma_\zeta^2\left[1 + \frac{\pi^2}{8r}\right]}} \\ &= \frac{1}{2}\sqrt{\frac{1}{\frac{8r+\pi^2}{8r}}} \\ &= \frac{1}{2}\sqrt{\frac{8r^2}{8r+\pi^2}} \end{aligned}$$

Note that we can test toy examples to gain an intuition. For example, if the variance of non-task specific variance is very large (or the task-specific variance is very small), such that $\sigma_\omega \gg \sigma_\zeta$ ($r \ll 1$), then the denominator will dominate and we will have $P(\text{unique spike}) \to 0$. If, however, the task-specific variance dominates, $\sigma_\zeta \gg \sigma_\omega$ ($r \gg 1$), then the denominator will be approximate to the numerator and we will be left with $P(\text{unique spike}) \approx \frac{1}{2}$. If we have equal variance between the task agnostic and task specific distributions, i.e. $\sigma_\zeta = \sigma_\omega$, then we will have $P(\text{unique spike}) \approx \frac{1}{2}\sqrt{\frac{8}{8+\pi^2}} \approx 0.334$ (approximately one in three chance).

More generally, we can consider the Taylor expansion of the above and use $y = \frac{8r}{\pi^2}$.

$$\begin{aligned} \frac{1}{2}\sqrt{\frac{8r}{8r+\pi^2}} &= \frac{1}{2}\sqrt{\frac{8r}{\pi^2}\frac{1}{\frac{8r^2}{\pi^2}+1}} \\ &= \frac{1}{2}\sqrt{y\frac{1}{y+1}} \\ &\approx \frac{1}{2}\sqrt{y\sum_{n=0}(-y)^n} \\ &= \frac{1}{2}\sqrt{\sum_{n=1}(-1)^{n-1}y^n} \\ &= \frac{1}{2}\sqrt{y - y^2 + y^3 - y^4 + \cdot} \end{aligned}$$

Now, for small enough r we will get small y and be able to ignore the power terms (y^n for $n > 1$) above and finally achieve

$$\begin{aligned} P(\text{unique spike}) &= \frac{1}{2}\sqrt{\frac{8r}{8r+\pi^2}} \\ &\approx \frac{1}{2}\sqrt{y} \\ &= \frac{1}{2}\sqrt{\frac{8r}{\pi^2}} \\ &= \frac{\sqrt{2}}{\pi}\sqrt{r} \end{aligned} \tag{3}$$

Which is Equation 18 in the main methods section.

1.3 # granule cells needed

Given the approximate probability of a unique spike given by Equation 3, we now ask: how many "trials" - or equivalently different neurons - do we need to ensure probability θ that there will be distinct spiking in at least one trial (or one neuron). Let p be the probability of distinct spiking for a given trial, and suppose there are n trials. The probability of at least one distinct trial is then:

$$P(\geq \text{one unique spike}) = 1 - P(\text{no unique spikes}) = 1 - (1-p)^n$$

We are interested in when this value $> \theta$. This is easy to solve

$$\begin{aligned} 1 - (1-p)^n &\geq \theta \\ (1-p)^n &\leq 1 - \theta \\ n \log(1-p) &\leq \log(1-\theta) \\ n &\geq \frac{\log(1-\theta)}{\log(1-p)} \end{aligned}$$

Where in the last line the inequality sign changes direction since $\log(1-p) < 0$. Now, the Taylor expansion of $\log(1+x)$ is

$$\log(x+1) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^3}{5} + \cdots$$

Finally, we apply Equation 3 and see that for small p (which is the case for small r) the power terms disappear and we have

$$n \geq \frac{\log(1-\theta)}{-p} = \frac{\log(1-\theta)}{-(\frac{\sqrt{2}}{\pi}\sqrt{r})} = -\frac{\pi}{\sqrt{2}} \log(1-\theta) \frac{1}{\sqrt{r}} \tag{4}$$

Which is Equation 19 in the main methods section.

2 Granule cell input retains RNN SNR and can be approximated by normal distribution

In this section we demonstrate that it is reasonable to apply Equation 1 to the cortico-cerebellar model considered in the main text.

Let h denote the random variable for the activity of an RNN neuron. We assume the distribution of this variable as unknown but that its variance can be expressed as $\text{Var}(h) = \sigma^2_{\text{RNN}} = \sigma^2_\omega + \sigma^2_\zeta$, where σ^2_ω, σ^2_ζ denote the variance task-agnostic and task-dependent variance respectively. Suppose that the RNN population is of size k.

For given (fixed) RNN activity $\mathbf{h} = (h_1, h_2, \cdots, h_k)$ we show that the input to a given GC follows a Gaussian distribution. To show this we use the fact that for a given GC the RNN-GC weights $W_{\text{MF}} = \{w_1, w_2, \cdots, w_k\}$ are themselves normally distributed around zero with variance σ^2_{MF}. Thus for given (fixed) RNN activity $\mathbf{h} = (h_1, h_2, \cdots, h_k)$ the input to a given GC is

$$I_{\text{GC}} = w_1 h_1 + w_2 h_2 + \cdot + w_k h_k; \quad w_i \sim \mathcal{N}(0, \sigma^2_{\text{MF}})$$

That is, the distribution of input currents to the GC population I_{GC} for a given RNN population activity can be expressed as a linear sum of (independent) normally distributed, and is itself therefore normally distributed.

We now consider the mean and variance of this distribution? Since each w_i has mean zero, I_{GC} has mean zero. For the variance, we apply the variance laws $\text{Var}(\lambda X) = \lambda^2 \text{Var}(X)$ and $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ for independent variables X, Y to see that

$$\begin{aligned} \text{Var}(I_{\text{GC}}) &= \text{Var}(w_1 h_1) + \text{Var}(w_2 h_2) + \cdot + \text{Var}(w_k h_k) \\ &= h_1^2 \text{Var}(w_1) + h_2^2 \text{Var}(w_2) + \cdot + h_k^2 \text{Var}(w_k) \\ &= (h_1^2 + h_2^2 + \cdot + h_k^2) \sigma^2_{\text{MF}} \end{aligned}$$

Then assuming that the mean RNN activity is zero (which we empirically observe as reasonable), we note that $(h_1^2 + h_2^2 + \cdot + h_k^2)$ is actually a sampled approximation for $k\text{Var}(h)$ so that

$$\text{Var}(I_{\text{GC}}) = \sigma_{\text{MF}}^2(h_1^2 + h_2^2 + \cdot + h_k^2) \approx \sigma_{\text{MF}}^2 k\text{Var}(h) = k\sigma_{\text{MF}}^2\sigma_{\text{RNN}}^2 = k\sigma_{\text{MF}}^2(\sigma_\omega^2 + \sigma_\zeta^2) \tag{5}$$

We can therefore write the input to the GC population for a given RNN population activity as

$$I_{\text{GC}} = \omega + \zeta; \quad \omega \sim \mathcal{N}(0,\, k\sigma_{\text{MF}}^2\sigma_\omega^2),\ \zeta \sim \mathcal{N}(0,\, k\sigma_{\text{MF}}^2\sigma_\zeta^2) \tag{6}$$

Assuming that other instances of RNN population activity **h** also provide good approximations of its true variance as in Equation 5, Equation 6 can apply. That is, the marginal distribution of I_{GC}, taken over all possible **h**, also follows Equation 6.

We have therefore shown that it is reasonable to apply Equation 1 to I_{GC} as set out in our original problem statement. Crucially, we see that the ratio between variances produced by task-agnostic and task-relevant activity - the SNR - remains the same as in the RNN, with

$$\underbrace{\frac{k\sigma_{\text{MF}}^2\sigma_\zeta^2}{k\sigma_{\text{MF}}^2\sigma_\omega^2}}_{\text{SNR}(\text{I}_{\text{GC}})} = \underbrace{\frac{\sigma_\zeta^2}{\sigma_\omega^2}}_{\text{SNR(RNN)}} \tag{7}$$

Thus Equation 3 can be used to estimated the probability of a unique spike for a given GC, and therefore Equation 4 is a valid prediction for the number of GCs required for distinct population vectors.