General Model

Circular Data Analysis

Our previous work reveals that the heading direction $s_m,\ m=1,2$ is subject to the normal distribution, that is, $s_m \sim \mathcal{N}(\mu, \sigma^2)$, where μ this the *mean direction* of s_m. The *mean resultant length* ρ is defined as

$$\rho = \mathbb{E}[\cos(s_m - \mu)]. \tag{1}$$

Assuming $\sigma \ll \pi$ ($\kappa > 1$), the *mean resultant length* is calculated by

$$\rho = \int_{-\pi}^{\pi} \cos(s_m) \frac{\exp[-\frac{(s_m-\mu)^2}{2\sigma^2}]}{\sigma\sqrt{2\pi}} ds_m \approx \int_{-\infty}^{\infty} \cos(s_m) \frac{\exp[-\frac{(s_m-\mu)^2}{2\sigma^2}]}{\sigma\sqrt{2\pi}} ds_m = e^{-\frac{\sigma^2}{2}}. \tag{2}$$

Definition of Circular Variance

If the *circular variance* $\hat{\sigma}^2$ is defined as

$$\hat{\sigma}^2 \equiv -2\ln\rho, \tag{3}$$

then $\hat{\sigma}^2$ is equivalent to σ^2. We are using $\frac{1}{\hat{\sigma}^2}$ to represent the *concentration parameter* $\hat{\kappa}$.

Zhang et al. assume s_m can be represented by a von Mises distribution, that is, $s_m \sim \mathcal{M}(\hat{s}_m, \hat{\kappa})$. The maximum likelihood of $\hat{\kappa}$ is the solution to

$$\rho = A(\hat{\kappa}), \tag{4}$$

where $A(x) \equiv I_1(x)/I_0(x)$. This solution is not easy to find for a large $\hat{\kappa}$ unless we are using the following approximation (see Fisher (1993) page 88 and Mardia & Jupp (2000) pages 85-86)

$$\hat{\kappa} = \begin{cases} 2\rho + \rho^3 + \frac{5\rho^5}{6}, & \rho < 0.53 \\ -0.4 + 1.39\rho + \frac{0.43}{1-\rho}, & 0.53 \le \rho < 0.85 \\ \frac{1}{3\rho - 4\rho^2 + \rho^3}, & \rho \ge 0.85. \end{cases} \tag{5}$$

Consider the Taylor expansion of $\hat{\kappa}$ centered at $\rho = 1$ for a large $\hat{\kappa}$, then $\hat{\kappa} \approx \frac{1}{3\rho-4\rho^2+\rho^3} \approx \frac{1}{2(1-\rho)} = \frac{1}{2(1-e^{-\sigma^2/2})} \approx \frac{1}{\sigma^2}$ is fine for $\rho \to 1$. That is reason why we still get right answer while using the Gaussian noise. However, after introducing the output-dependent noise, the concentration parameter might be smaller so that we can no longer use the von Mises distribution to approximate a normal distribution. In that case, we shall use Eq. (3) to calculate the variance of circular data.

Bayesian Inference

According to Bayes rule, the posterior $p(s_1, s_2|x_1, x_2)$ is given by

$$p(s_1, s_2|x_1, x_2) = \frac{p(x_1|s_1)p(x_2|s_2)p(s_1, s_2)}{p(x_1)p(x_2)}. \tag{6}$$

$p(x_m)$ and $p(s_m)$ are uniform distributions. Take the derivative of $p(s_1, s_2|x_1, x_2)$ with respect to s_1 to be zero

$$\kappa_s \sin(s_2 - s_1) + \kappa_1 \sin(x_1 - s_1) = 0, \tag{7}$$

that will allow us to find the peak position. We project the dynamic equation to the position mode, then we get

$$HJ_{rp} \sin(s_2 - s_1) + \frac{F}{u_1} \sin(x_1 - s_1) = 0. \tag{8}$$

Compared with Eq. (7), our network will encode the prior information in the following way

$$\frac{\kappa_s}{\kappa_1} = \frac{HJ_{rp}}{F} u_1. \tag{9}$$

Discussion

If we project the dynamic equation to the height mode $\cos(y_m - s_m)$ and position mode $\sin(y_m - s_m)$, these two equations can be combined together

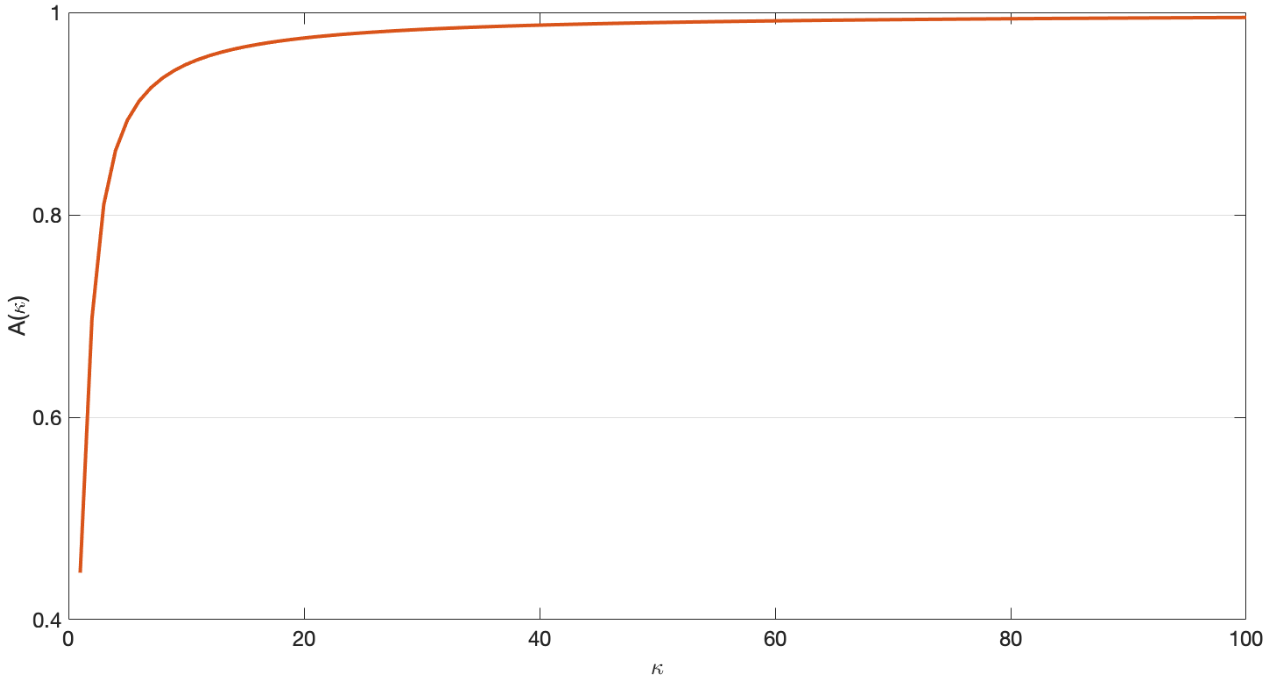
$$u_1 e^{js_1} = u_1 \frac{HJ_{rp}}{1 - HJ_{rc}} e^{js_2} + \frac{F}{1 - HJ_{rc}} e^{jx_1}, \tag{10}$$

where j is the imaginary unit. If we assume u_1 on the right hand side to be $\frac{F}{1-HJ_{rc}} e^{jx_1}$ and $s_2 \approx x_2$, we have $\hat{\kappa}_1 e^{js_1}|_{I_1, I_2} = \hat{\kappa}_1 e^{jx_1}|_{I_1} + \hat{\kappa}_{2s} e^{jx_2}$ in our CCN paper.

The ratio will be

$$\frac{\kappa_s}{\kappa_1} \approx \frac{HJ_{rp}}{1 - HJ_{rc}}. \tag{11}$$

However, the bias $\frac{\hat{s}_1 - x_1}{x_2 - x_1}|_{x_2 \to 0}$ and $\frac{\hat{\kappa}_{2s}}{\hat{\kappa}_1}$ are also approaching $\frac{HJ_{rp}}{1-HJ_{rc}}$. We maginalize s_2 then posterior will give us $\kappa_{2s} = A^{-1}[A(\kappa_2)A(\kappa_s)]$. When κ_2 is large ($\kappa_2 > 50$), we have $\kappa_s \approx \kappa_{2s}$.



We project the dynamic equation to the height mode

$$1 = HJ_{rc} + HJ_{rp}\cos(\Delta s) + \frac{F}{u_1}\cos(x_1 - s_1). \quad (12)$$

If we take derivative of both sides with respect to s_1, that will give us Eq. (8). We maginalize s_2

$$p(s_1|x_1, x_2) \propto p(s_1|x_1)p(s_1|x_2), \quad (13)$$

according to vector diagram, we have

$$\kappa_1|_{I_1,I_2} = \kappa_{12}\cos(x_2 - s_1) + \kappa_1|_{I_1}\cos(x_1 - s_1). \quad (14)$$

Compared with Eq. (12), if we further assume $x_2 \approx s_2$, then

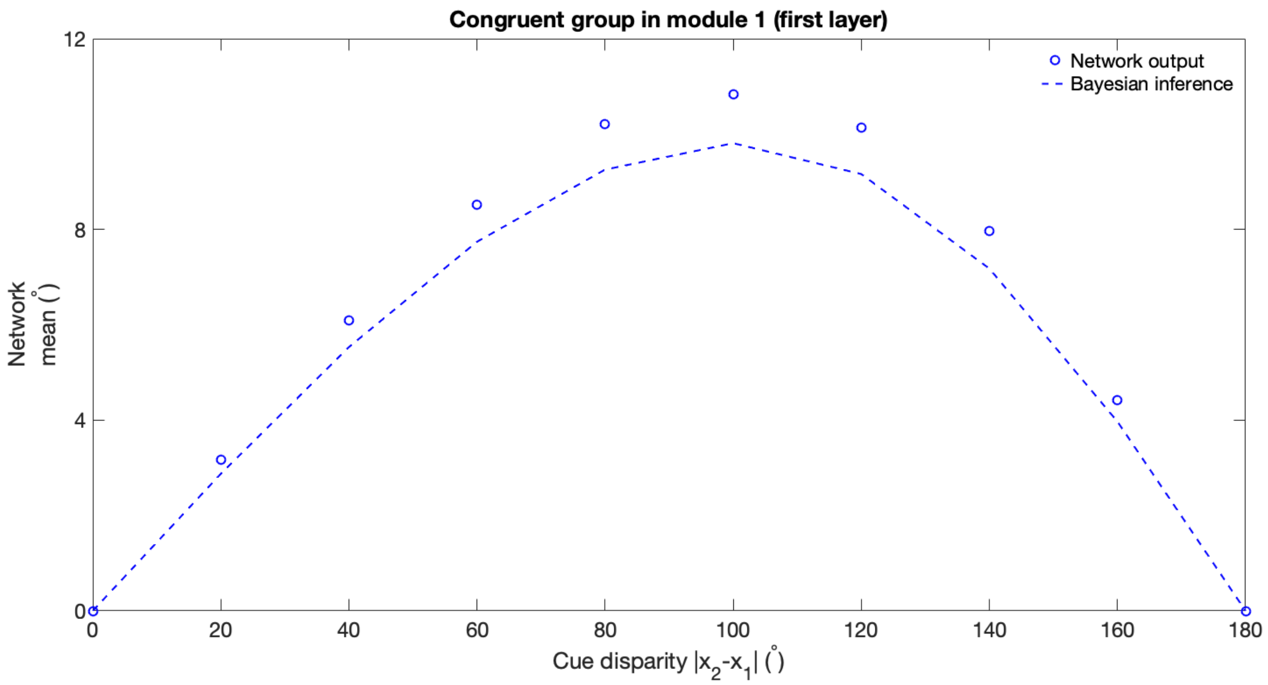
$$\begin{aligned} \hat{\kappa}_1|_{I_1,I_2} &\propto u_1, \\ \hat{\kappa}_{12} &\propto u_1\frac{HJ_{rp}}{1 - HJ_{rc}}, \\ \hat{\kappa}_1|_{I_1} &\propto \frac{F}{1 - HJ_{rc}}. \end{aligned} \quad (15)$$

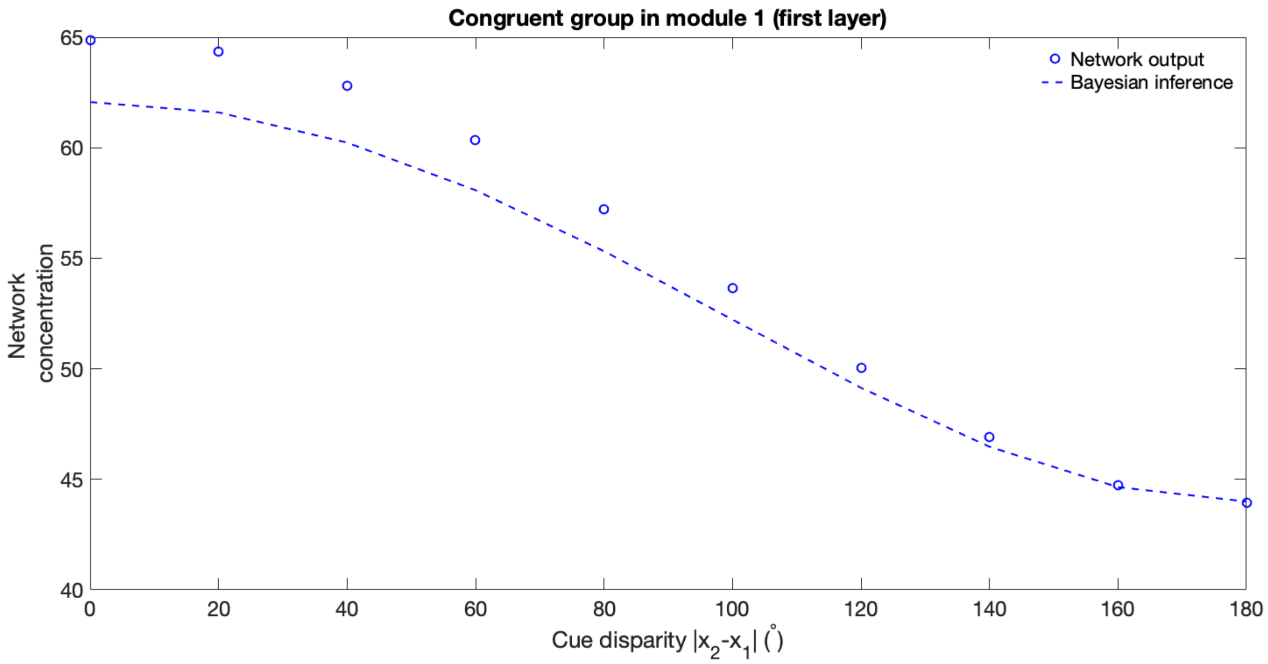
The coefficient should be $g_{mm}c_0$ based on the derivation of the noise variance

$$g_{mm}c_0 \approx \frac{\pi\tau ND}{F_0 u_0} - \frac{2\pi\tau\rho NJ_{rc}B_1(a_0)}{F_0}. \quad (16)$$

Results

Our purpose is to encode the prior information in our network model, we compare our network outputs with Bayes' theorem.





Note that we take the average of $g_{mm}c_0$ among 10 trials. Besides, we have $\langle u_1 \rangle$ and $\langle H \rangle$ in the same way. κ_1 and κ_{12} are calculated by

$$\begin{aligned} \kappa_1 &\approx g_{mm}c_0 \frac{F}{1 - HJ_{rc}}, \\ \kappa_s/\kappa_1 &= \frac{HJ_{rp}}{F} u_1, \\ \kappa_{12} &= A^{-1}[A(\kappa_1)A(\kappa_s)]. \end{aligned} \tag{17}$$

Prior with Independent Component

The two-component prior is as follows

$$p(s_1', s_2') = \frac{p_0}{2\pi} V(s_1' - s_2', \kappa_s) + \frac{1 - p_0}{(2\pi)^2}. \tag{18}$$

The marginal posterior $p(s_1'|x_1,x_2)$ is given by

$$p(s_1'|x_1,x_2) = p_0 CV(s_1'-x_1,\kappa_1)V(s_1'-x_2,\kappa_{2s}) + (1-p_0)V(s_1'-x_1,\kappa_1), \tag{19}$$

where $C=\frac{2\pi I_0(\kappa_1)I_0(\kappa_{2s})}{I_0(\sqrt{\kappa_1^2+\kappa_{2s}^2})}$, differentiation with respect to s_1 yields

$$\begin{aligned}\frac{\partial}{\partial s_1'}p(s_1'|x_1,x_2) =& -p_0CV(s_1'-x_1,\kappa_1)V(s_1'-x_2,\kappa_{2s})[\kappa_1\sin(s_1'-x_1)+\kappa_{2s}\sin(s_1'-x_2)]\\ &-(1-p_0)V(s_1'-x_1,\kappa_1)\kappa_1\sin(s_1'-x_1).\end{aligned} \tag{20}$$

We project the dynamic equation of the second layer to the height mode and the position mode as below

$$u_1' =(1-p_0)\frac{F}{1-H'J_{rc}}\cos(x_1-s_1') + p_0\frac{c_k H u_1}{B_1(a_0)(1-H'J_{rc})}\cos(s_1-s_1'), \tag{21}$$

$$0 =(1-p_0)\frac{F}{1-H'J_{rc}}\sin(x_1-s_1') + p_0\frac{c_k H u_1}{B_1(a_0)(1-H'J_{rc})}\sin(s_1-s_1'). \tag{22}$$

Taking the logarithm of $p(s_1'|x_1,x_2)$ yields

$$\ln[p(s_1'|x_1,x_2)] = \ln[1-p_0+p_0CV(s_1'-x_2,\kappa_{2s})] + \kappa_1\cos(s_1'-x_1) - \ln[2\pi I_0(\kappa_1)], \tag{23}$$

we are on the wrong track.

Prior Encoding

~~The independent component~~ $\cancel{\frac{1-p_0}{(2\pi)^2}}$ ~~won't be encoded in our model unless we make the following assumption~~

$$p(x_1,s_1') \propto [p(x_1|s_1)]^{1-p_0}, \tag{24}$$

~~that is,~~ $p(x_1|s_1')\propto V[x_1-s_1',(1-p_0)\kappa_1]$~~, which makes sense since the input from cue 1 has been rescaled by~~ $1-p_1$ ~~according to the current network structure. Furthermore, we will get another vector diagram based on~~

$$p(s_1'|s_1,x_1) \propto p(s_1'|x_1)p(s_1'|s_1). \tag{25}$$

~~Actually,~~

$$\begin{aligned}p(s_1'|x_1,x_2) &= \int ds_1 p(s_1',s_1|x_1,x_2)\\ &\propto p(s_1'|x_1)\int ds_1 p(s_1'|s_1)p(s_1|x_1,x_2)\end{aligned} \tag{26}$$

~~we can assume~~ $p(s_1'|s_1)=V(s_1'-s_1,\kappa_s')$~~.~~

~~We start from Eq.~~ (23)~~, with assuming~~ $p_0CV(s_1'-x_2,\kappa_{2s})$ ~~is dominant, the logarithm of~~ $p(s_1'|x_1,x_2)$ ~~becomes~~

$$\ln[p(s_1'|x_1,x_2)] \approx \kappa_{2s}\cos(s_1'-x_2) + \kappa_1\cos(s_1'-x_1) + \ln[\frac{p_0C}{(2\pi)^2I_0(\kappa_1)I_0(\kappa_{2s})}]. \tag{27}$$

~~In Eq.~~ (20)~~, we take the derivative of~~ $p(s'_1|x_1, x_2)$ ~~to be zero,~~ $V(s'_1 - x_2, \kappa_{2s})$ ~~is more like the Dirac delta function centered at~~ $s'_1 = x_2$ ~~when~~ κ_{2s} ~~is large. That requires~~

$$\kappa_1 \sin(s'_1 - x_1) + \kappa_{2s} \sin(s'_1 - x_2) \approx 0, \tag{28}$$

~~where we assume~~ $s'_1 - x_1$ ~~is sufficiently small.~~

Firstly we start from the inverse probrem. Eqs. (21) and (22) reveal a vector diagram. In Eq. (23), when p_0 is small,

$$\ln[1 - p_0 + p_0 CV(s'_1 - x_2, \kappa_{2s})] \simeq p_0 \exp[\kappa_{2s} \cos(s'_1 - x_2)] - p_0, \tag{29}$$

with the condition $\kappa_1 \gg \kappa_{2s}$. However, $s'_1 - x_2$ varies from $-\pi$ to π, which means this prior encoding only works in a narrow range around $\frac{\pi}{2}$. Here we assume $|s'_1 - x_2| \to \frac{\pi}{2}$, then Eq. (23) will be

$$\ln[p(s'_1|x_1, x_2)] \approx p_0[\kappa_{2s} \cos(s'_1 - x_2) + \kappa_1 \cos(s'_1 - x_1)] + (1 - p_0)\kappa_1 \cos(s'_1 - x_1) - \ln[2\pi I_0(\kappa_1)]. \tag{30}$$

Hence we obtain the same form of the product of two von Mises functions which have been used to illustrate a vector diagram.

The next part is to eliminate x_2 in Eq. (30). Let $\frac{c_k H}{B_1(a_0)(1-H' J_{rc})} = 1$, c_k depends on the outputs of the first and second layers. However, in weak input limit $c_k \approx \frac{J_{rc}+J_{rp}}{1-\sqrt{1-4\rho(J_{rc}+J_{rp})}}$. That is, the terms inside the square brackets in Eq. (30) is the vector which has been used in Eq. (14).

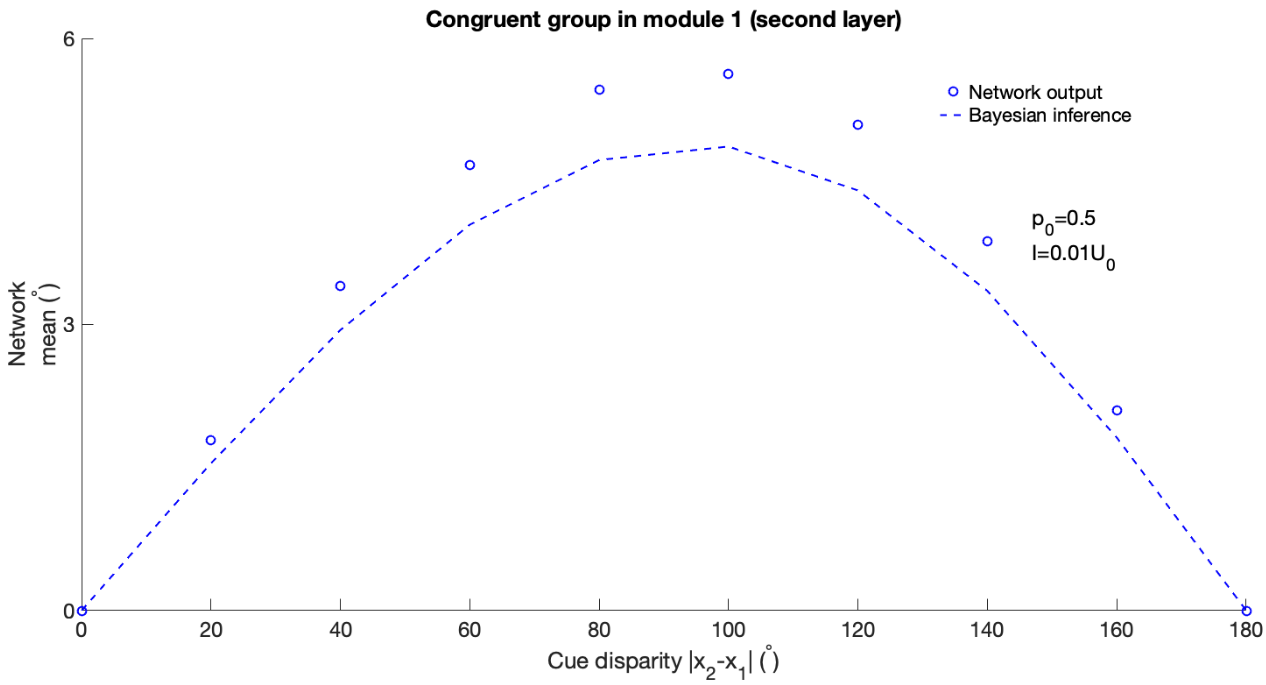
Results

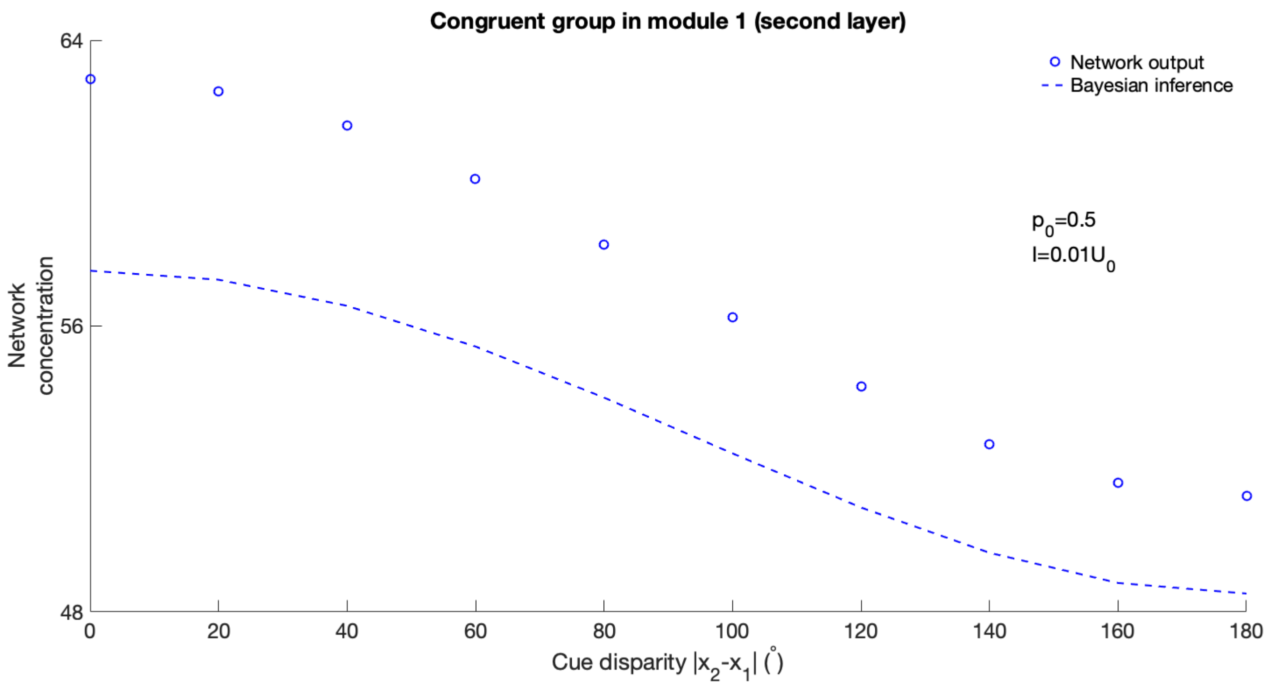
Note that

$$\begin{aligned} g'_{mm} c'_0 &\approx g_{mm} c_0, \\ \kappa'_1/(1 - p_0) &\approx g'_{mm} c'_0 \frac{F}{1 - H' J_{rc}} \approx g_{mm} c_0 \frac{F}{1 - H J_{rc}}, \end{aligned} \tag{31}$$

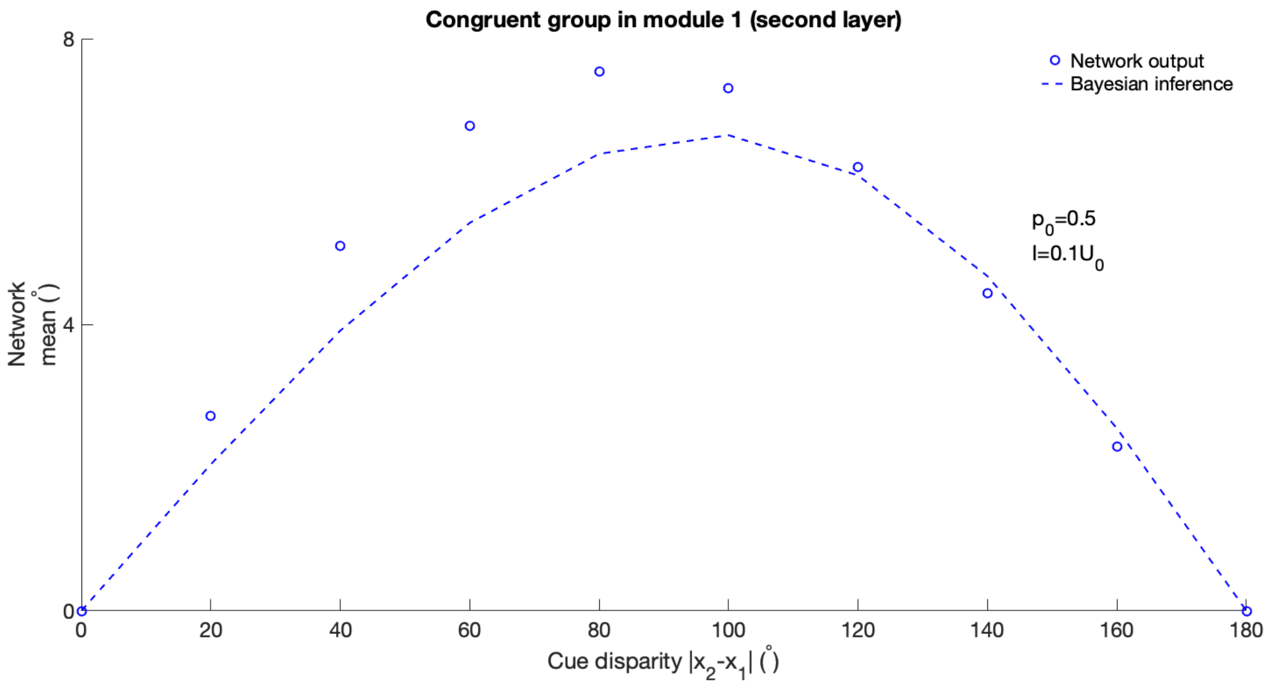
and add $p_0 \kappa_1$ to our Bayesian model, which κ_1 is a vector sum of κ_1 and κ_{12} from the previous discussion.

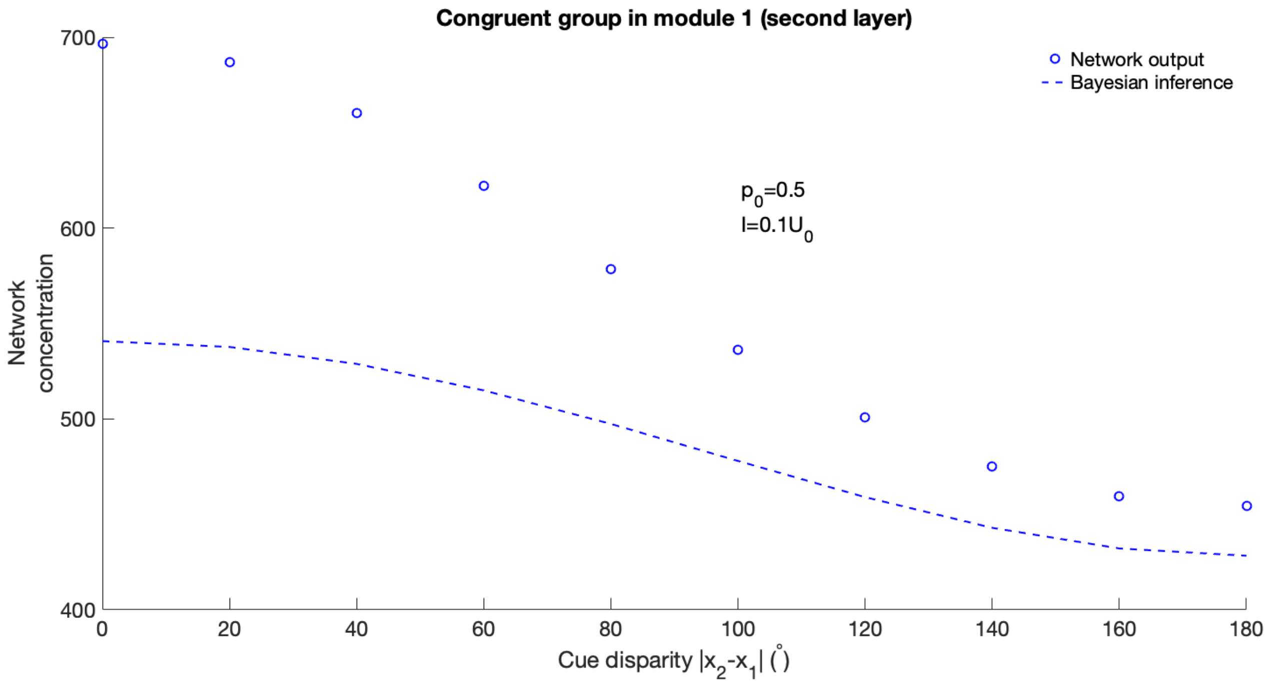
Again, in weak input limit, the result is quite satisfactory.





However, if we increase the input strength, things are getting worse which is not surprising due to the limitation of our assumptions.





Bayes Factor

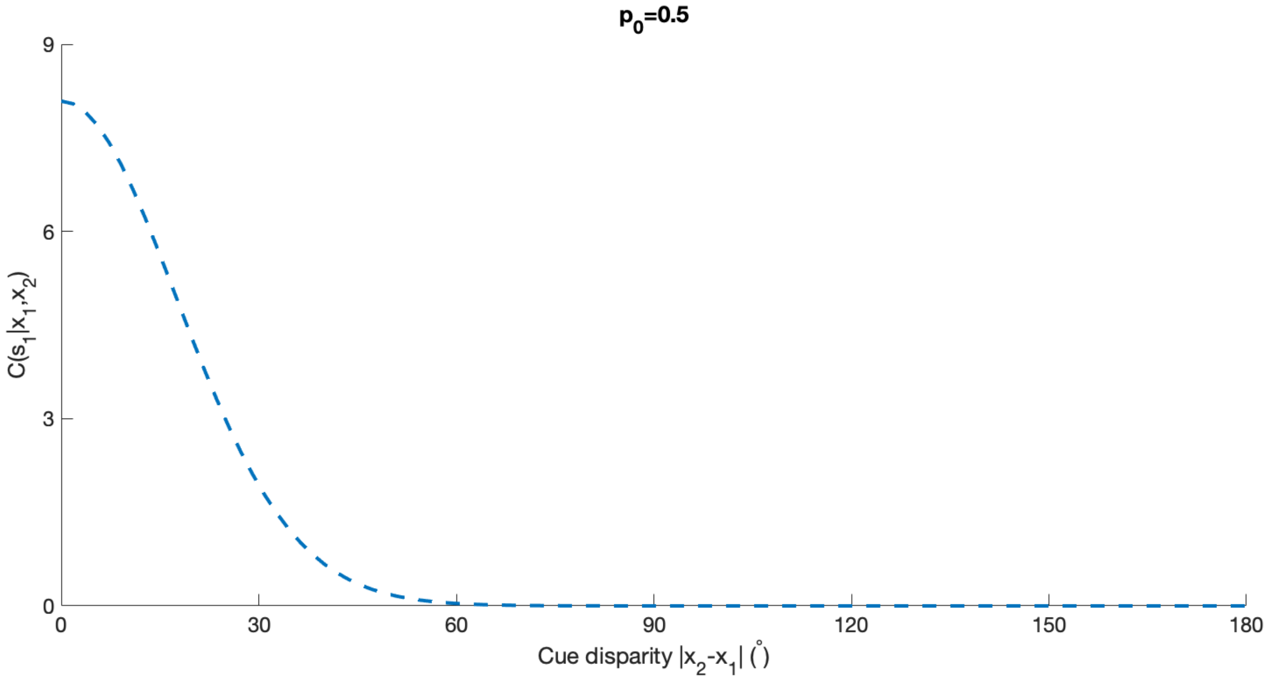
Given $p(s_1, s_2) = p_c(s_1, s_2) + p_i$, in which p_c and p_i refer to correlated prior and independent prior respectively, the marginal posterior $p(s_1|x_1, x_2)$ is calculated to be

$$\begin{aligned} p(s_1|x_1, x_2) &= \langle p_c(s_1, s_2)p(x_2|s_2)\rangle_{s_2} p(x_1|s_1) + \langle p_i p(x_2|s_2)\rangle_{s_2} p(x_1|s_1) \\ &= [p(M_1|x_1, x_2) + p(M_2|x_1, x_2)]\, p(x_1|s_1). \end{aligned} \tag{32}$$

We define the Bayes factor $C(s_1|x_1, x_2)$ for causal inference as

$$C(s_1|x_1, x_2) \triangleq \frac{p(M_1|x_1, x_2)}{p(M_2|x_1, x_2)}. \tag{33}$$

In our model, $C(s_1|x_1, x_2) \approx \frac{2\pi p_0}{1-p_0} V(s_1 - x_2, \kappa_{2s}) \approx \frac{2\pi p_0}{1-p_0} V(\Delta x, \kappa_s)$.



p_0=0.5
9
6
3
0
$C(s_1|x_1,x_2)$
0
30
60
90
120
150
180
Cue disparity $|x_2-x_1|$ (°)