THEORETICAL NEUROSCIENCE I

Lecture 15: Decoding neural populations

Prof. Jochen Braun

Otto-von-Guericke-Universität Magdeburg,
Cognitive Biology Group

Outline

1. Decoding diverse populations (recap)

2. Optimal inference: ML and MAP

3. Computing log-likelihood with neurons

4. Detecting, discriminating, and identifying with log-likelihood

1 Decoding diverse populations (recap)

In general, neuronal populations are diverse, in that each neuron prefers a different stimulus attribute (*e.g.*, a different direction of motion).

To decode the response of a diverse population, we must rely on 'prior knowledge' about the responsiveness of each individual neuron, namely, the conditional probability distribution $P(x|s)$.

As a first step, we infer from each individual neuronal response a stimulus *likelihood* distribution $L(s|x)$: the likelihood that an observed response was caused by different possible stimuli.

Assuming independence, we can combine these likelihood across the population to obtain a joint likelihood: the most likely *hypothetical* stimulus to have cause *all* observed responses is the *maximum likelihood* estimate.
Observing neural responses n_1 and n_2, we seek the **likelihoods** of alternative stimuli θ_1 and θ_2.

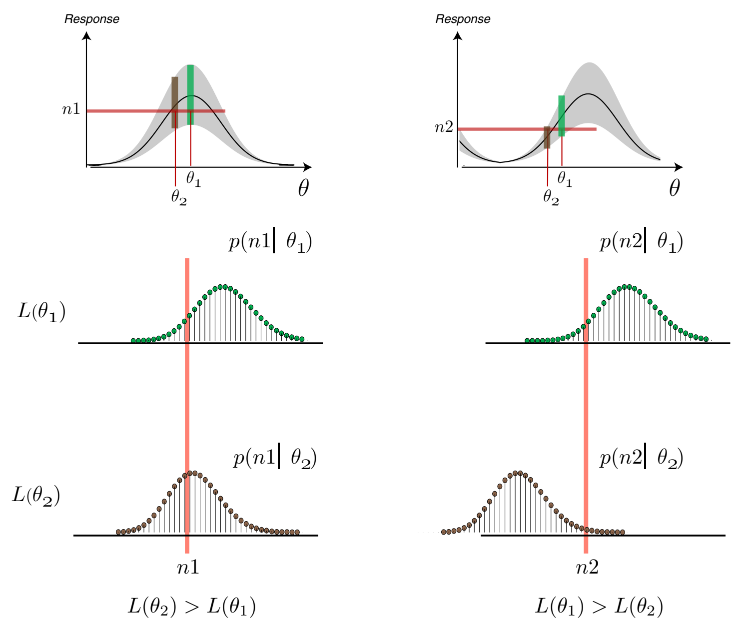


Figure 1: Likelihoods of alternative stimuli θ_1 and θ_2.

Prior knowledge: conditional probability $P[x|\theta]$

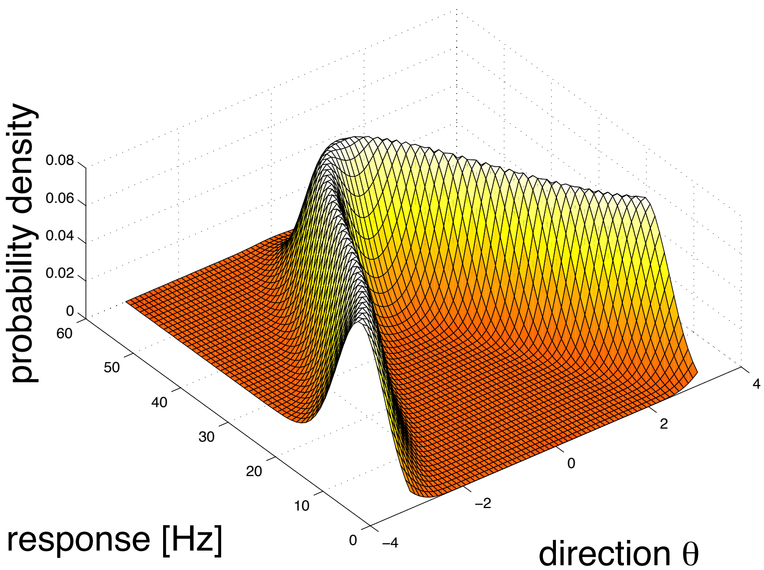


Figure 2: Conditional probability.

Joint probability $P[x, \theta] = P[x|\theta]\, P[\theta]$, with $P[\theta] = \frac{1}{2\pi}$

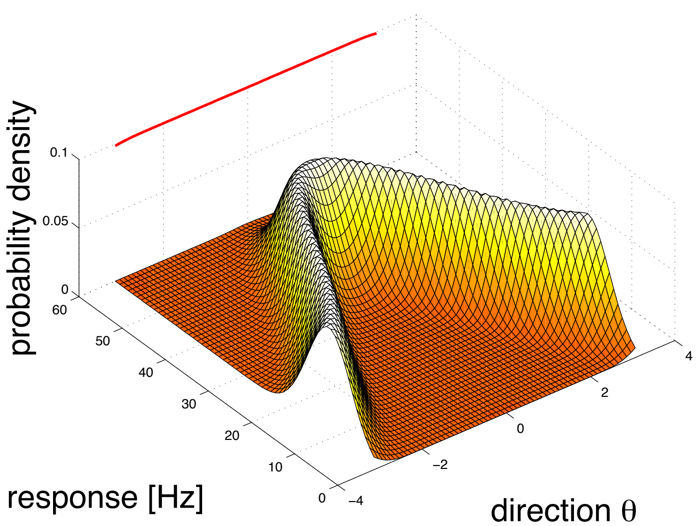


Figure 3: Joint probability.

Marginal probability $P[x] = \int P[x, \theta] \, d\theta$

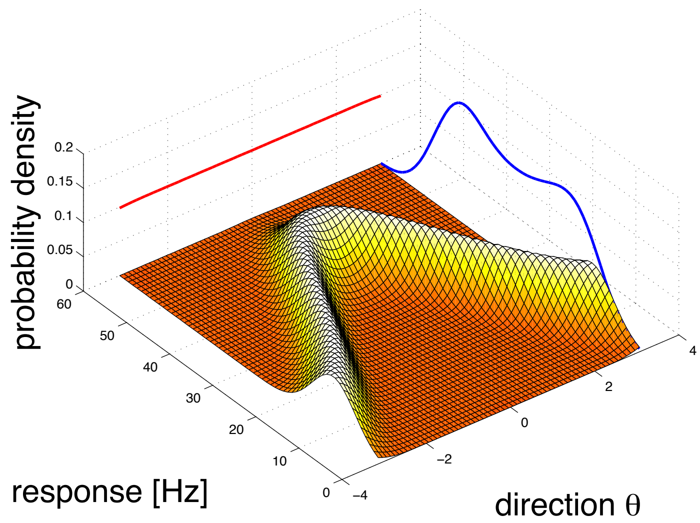


Figure 4: Marginal probability.

Stimulus likelihood $P[\theta|x]$

$$P[\theta|x] = \frac{P[x, \theta]}{P[x]} = \frac{P[x|\theta] \, P[\theta]}{P[x]}$$

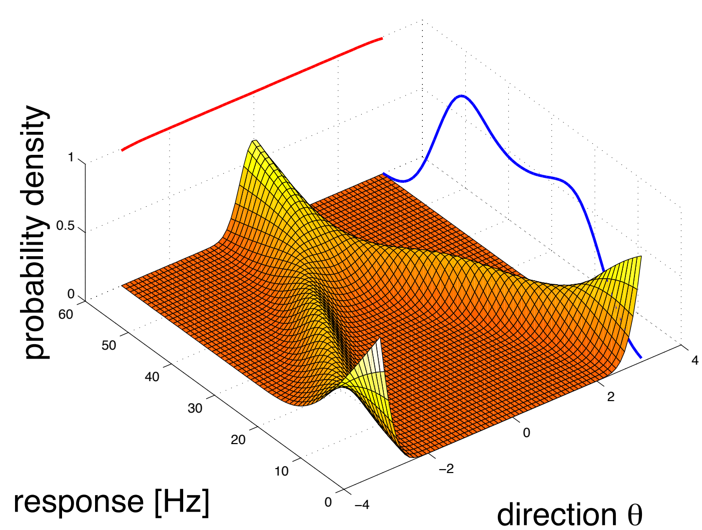


Figure 5: Stimulus likelihood.

2 Optimal inference (ML and MAP)

There are several methods for inferring a stimulus from the firing of a population of neurons. Two methods are statistically efficient (as good as possible, given the available information):

- Maximum likelihood inference (ML) maximizes $P(x|s)$, the conditional probability of an observed *response* x, given the unknown stimulus s.

- Maximum *a posteriori* (MAP) inference maximizes $L(s|x)$, the conditional likelihood of an unknown *stimulus* s, given the observed response x.

Comparison ML and MAP

Both ML and MAP require extensive *prior knowledge*, specifically, the conditional probability of responses, given any stimulus.

$$P(x|s)$$

MAP additionally requires the prior probability of different stimuli.

$$P(s)$$

Conditional stimulus *likelihood* follows from conditional response probability

$$L(s|x) = \frac{P(x|s)\,P(s)}{P(x)}$$

$$\Leftrightarrow \qquad \ln L(s|x) = \ln P(x|s) + \ln P(s) - \ln P(x)$$

Comparison ML and MAP

$$\ln L(s|x) = \ln P(x|s) + \ln P(s) - \ln P(x)$$

If $P(s) = const$,

$$\frac{\partial}{\partial s} \ln L(s|x) = \frac{\partial}{\partial s} \ln P(x|s) \qquad \textit{maximum likelihood (ML)}$$

otherwise,

$$\frac{\partial}{\partial s} \ln L(s|x) = \frac{\partial}{\partial s} \ln P(x|s) + \frac{\partial}{\partial s} \ln P(s)$$

maximum a posteriori likelihood (MAP)

Prior and posterior likelihood

When some stimuli are more probable than others, we distinguish *prior* and *posterior* stimulus probabilities/likelihoods:

Prior	$P(s)$
Posterior	$L\left(s\vert\{x_i\}\right) = \frac{P(x_{1:n}\vert s)\ P(s)}{P(x_{1:n})}$

The *prior* is the probability *before* and the *posterior* the probability (likelihood) *after* observing a population response $x_{1:n}$.

Joint probability

Assuming independence, the joint conditional probability of a population response

$$x_{1:n} \equiv \{x_1, x_2, \ldots, x_n\},$$

given stimulus s, is the product of individual conditional probabilities of x_i, given s:

$$P(x_{1:n}|s) = \prod_{i=1}^{n} P(x_i|s)$$

$$\ln P(x_{1:n}|s) = \ln \prod_{i=1}^{n} P(x_i|s) = \sum_{i=1}^{n} \ln P(x_i|s)$$

Joint posterior likelihood

The joint posterior likelihood of the observed response is

$$L(s|x_{1:n}) = \frac{P(x_{1:n}|s)\, P(s)}{P(x_{1:n})}$$

$$\begin{aligned}
\ln L(s|x_{1:n}) &= \ln P(x_{1:n}|s) + \ln P(s) - \ln P(x_{1:n}) \\
&= \sum_{i=1}^{n} \ln P_i(x_i) + \ln P(s) - \ln P(x_{1:n})
\end{aligned}$$

Maximization

We seek the value of s that maximizes $\ln L(s|x_{1:n})$

$$\begin{aligned}
0 &\stackrel{!}{=} \frac{\partial}{\partial s} \ln\, L(s|x_{1:n}) = \\
&= \frac{\partial}{\partial s} \sum_{i=1}^{n} \ln\, P(x_i|s) + \frac{\partial}{\partial s} \ln\, P(s) \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial s} \ln\, P(x_i|s) + \frac{\partial}{\partial s} \ln\, P(s)
\end{aligned}$$

This is the **MAP condition**!

Gaussian prior:

$$P(s) = \frac{1}{\sqrt{2\pi\sigma_{prior}^2}} \; \exp - \frac{(s - s_{prior})^2}{2\sigma_{prior}^2}$$

$$\ln P(s) = -\frac{(s - s_{prior})^2}{2\sigma_{prior}^2} - \frac{1}{2} \ln \sqrt{2\pi\sigma_{prior}^2}$$

$$\frac{\partial}{\partial s} \; \ln P(s) = -\frac{s - s_{prior}}{\sigma_{prior}^2}$$

Poisson variability & Gaussian tuning

$$\ln P(x_i|r_i(s)) = x_i \; \ln r_i - \log x_i! - r_i, \qquad \frac{\partial \ln P(x_i|r_i(s))}{\partial r_i} = \frac{x_i - r_i}{r_i}$$

$$r_i(s) = r_{max} \; \exp\left[-\frac{(s - s_i)^2}{2\sigma_{tuning}^2} \right], \qquad \frac{\partial r_i}{\partial s} = -\frac{s - s_i}{\sigma_{tuning}^2} \, r_i$$

$$\frac{\partial \ln P(x_i|r_i(s))}{\partial s} \;=\; \frac{\partial \ln P(x_i|r_i)}{\partial r_i} \, \frac{\partial r_i}{\partial s} = -\frac{(x_i - r_i)\,(s - s_i)}{\sigma_{tuning}^2}$$

MAP condition

$$0 \;\stackrel{!}{=}\; -\sum_{i=1}^{n} \frac{(x_i - r_i)\,(s - s_i)}{\sigma_{tuning}^2} - \frac{s - s_{prior}}{\sigma_{prior}^2}$$

Uniform and symmetric coverage

$$\sum_{i=1}^{n} r_i(s)\,(s - s_i) \approx 0$$

simplifies MAP condition to

$$0 \;=\; -\sum_{i=1}^{n} \frac{x_i\,(s - s_i)}{\sigma_{tuning}^2} - \frac{s - s_{prior}}{\sigma_{prior}^2} =$$

$$=\; -\sum_{i=1}^{n} \frac{x_i\, s}{\sigma_{tuning}^2} + \sum_{i=1}^{n} \frac{x_i\, s_i}{\sigma_{tuning}^2} - \frac{s}{\sigma_{prior}^2} + \frac{s_{prior}}{\sigma_{prior}^2}$$

$$s\left[\frac{1}{\sigma_{tuning}^2}\sum_{i=1}^{n} x_i + \frac{1}{\sigma_{prior}^2}\right] = \frac{1}{\sigma_{tuning}^2}\sum_{i=1}^{n} x_i\, s_i + \frac{1}{\sigma_{prior}^2}$$

$$s_{MAP} \;=\; \frac{\frac{1}{\sigma_{tuning}^2}\sum_{i=1}^{n} x_i\, s_i + \frac{1}{\sigma_{prior}^2}}{\frac{1}{\sigma_{tuning}^2}\sum_{i=1}^{n} x_i + \frac{1}{\sigma_{prior}^2}}$$

$$=\; \frac{\sum_{i=1}^{n} x_i\, s_i + \frac{\sigma_{tuning}^2}{\sigma_{prior}^2}\, s_{prior}}{\sum_{i=1}^{n} x_i + \frac{\sigma_{tuning}^2}{\sigma_{prior}^2}}$$

The MAP estimate s_{MAP} is the response-weighted sum of the preferred stimuli s_i and the most probable prior stimulus s_{prior}. The prior is weighted by the variance ratio.

Summary MAP inference

- The MAP estimate s_{MAP} is the response-weighted average of the preferred stimuli s_i and of the most probable prior stimulus s_{prior}

$$s_{MAP} = \frac{\sum_{i=1}^{n} x_i\, s_i + \frac{\sigma_{tuning}^2}{\sigma_{prior}^2}\, s_{prior}}{\sum_{i=1}^{n} x_i + \frac{\sigma_{tuning}^2}{\sigma_{prior}^2}}$$

- The weight of s_{prior} is the variance ratio of tuning curve and prior distribution.

- The ML estimate s_{ML} was simply the response-weighted average

$$s_{ML} = \frac{\sum_{i=1}^{n} x_i \, s_i}{\sum_{i=1}^{n} x_i}$$

3 Computing log-likelihood with neurons

The Bayesian formalism reveals how to optimally combine information from different sources (e.g., diverse neurons responding to the same stimulus).

It holds under ideal conditions: exhaustive prior knowledge, independent variability, uniform coverage.

Sensory systems are sufficiently evolved to apply Bayesian statistics!

The required computations are not complex and readily performed by neurons: weighted sums and peak localization.

Following Jazayeri & Movshon (2006), we construct a two-layer feedforward network to compute maximum likelihood.
First layer Consider a population of **first layer** neurons i with circular Gaussian tuning $f_i(\theta)$, Poisson variability, uniform and symmetric coverage, and independently variable responses x_i:

$$f_i(\theta) = r_{max}\, e^{[\kappa\, \cos(\theta - \theta_i) - \kappa]},$$

$$\ln p(x_i|\theta) \quad \approx \quad x_i \ln f_i(\theta) \quad \approx \quad x_i\, \cos(\theta - \theta_i)$$

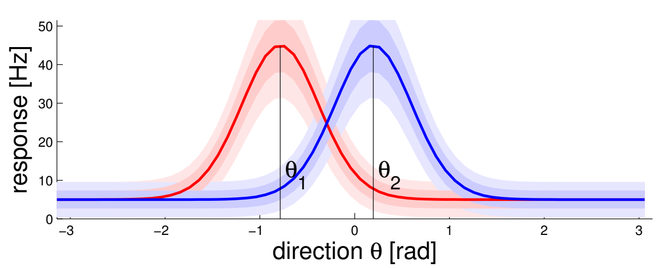


Figure 6: Gaussian tuning aproximation.

Only one important term

Applying ML inference, only one term of the joint log likelihood remains important (i.e., varies with s)

$$\begin{aligned} \ln L(\theta|x_{1:n}) &= \sum_{i=1}^{n} \ln P(x_i|\theta) = \\ &\approx \sum_{i=1}^{n} x_i \ln f_i(\theta) = \\ &\approx \sum_{i=1}^{n} x_i \cos(\theta - \theta_i) \end{aligned}$$

The ML estimate θ_{ML} is the value that maximises this sum of products: *actual* response $\times$ *log hypothetical* response.

Second layer

To compute $\ln L(\theta|x_{1:n})$, we construct a **second layer** population of neurons j with responses y_j and preferred stimuli θ_j. The activity of second layer neurons is obtained in a feedforward fashion as weighted sums of **first layer** responses x_i.

$$y_j = \sum_{i=1}^{n} x_i \, g(\theta_j - \theta_i)$$

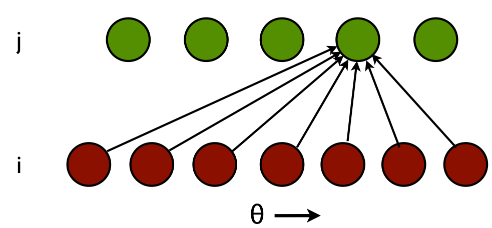


Figure 7: Interaction with a second layer of population.

Approximating log-likelihood

We choose the second-layer weights

$$g(\theta_i - \theta_j) = \ln f_i(\theta_j) = \cos(\theta_i - \theta_j)$$

such that activity j represents log-likelihood of θ_j:

$$y_j = \sum_{i=1}^{n} x_i\, g(\theta_i - \theta_j) = \sum_{i=1}^{n} x_i\, \cos(\theta_i - \theta_j) = \ln L(\theta_j | x_{1:n})$$

Each second-layer activity y_j sums (with a different set of weights) over the same first-layer activities x_i: the log-tuning of i to a hypothetical stimulus θ_j.

Thus, second-layer activities give *more* weight to first-layer activities (same) preference, $\theta_i = \theta_j$, and *less* weight to first-layer activities with other preferences, $\theta_i \neq \theta_j$.

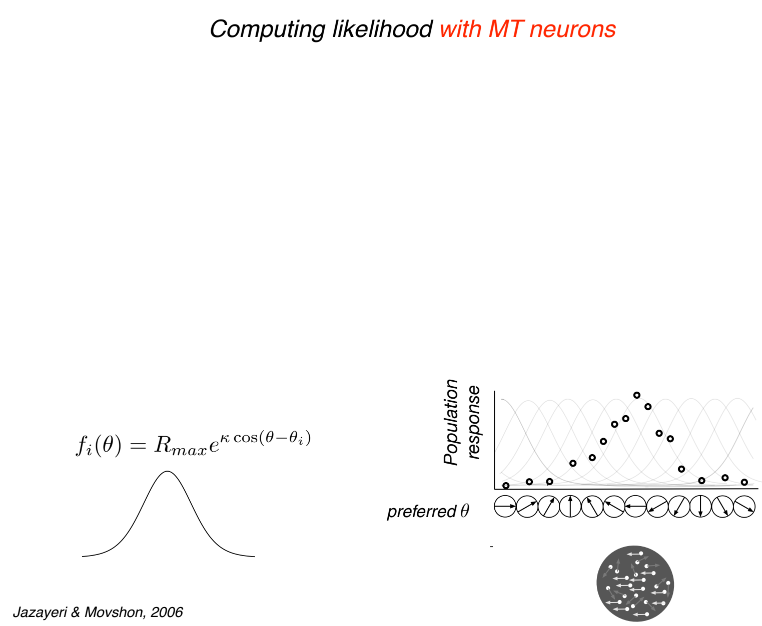


Figure 8: Approximating log-likelihood. [1]

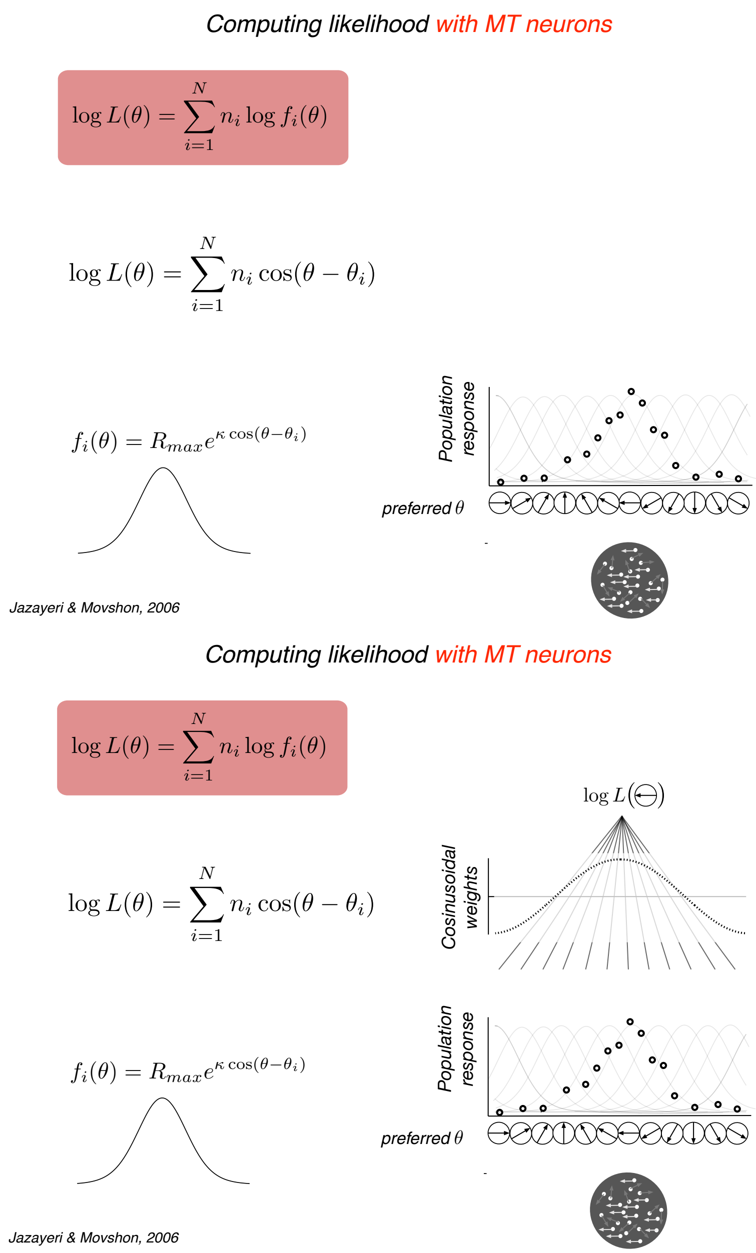
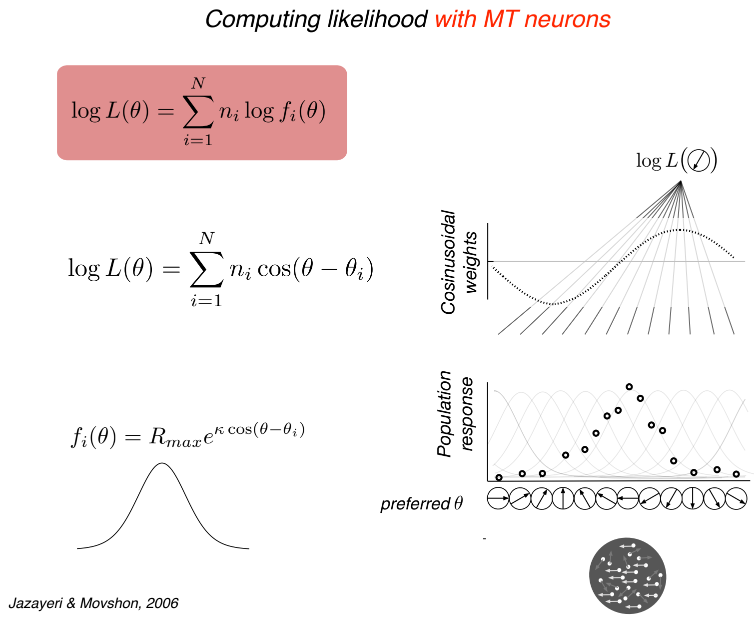


Figure 9: Approximating log-likelihood. [1]



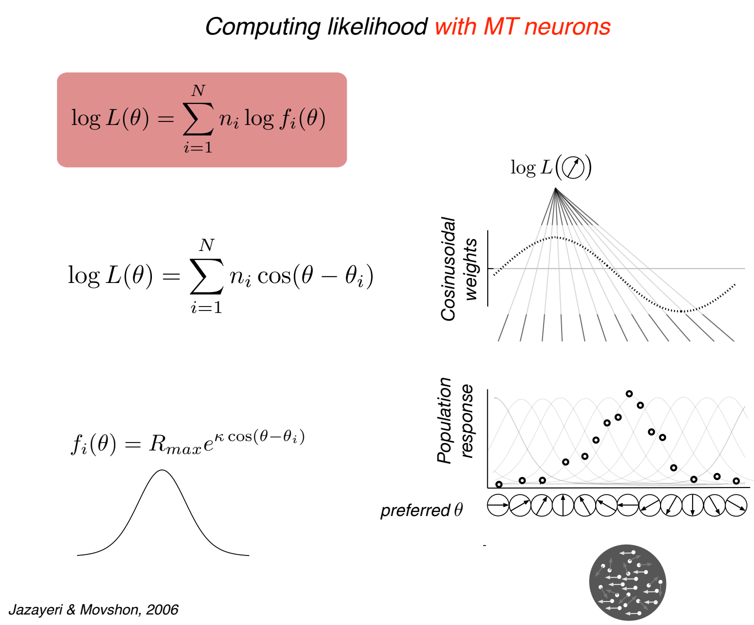


Figure 10: Approximating log-likelihood. [1]

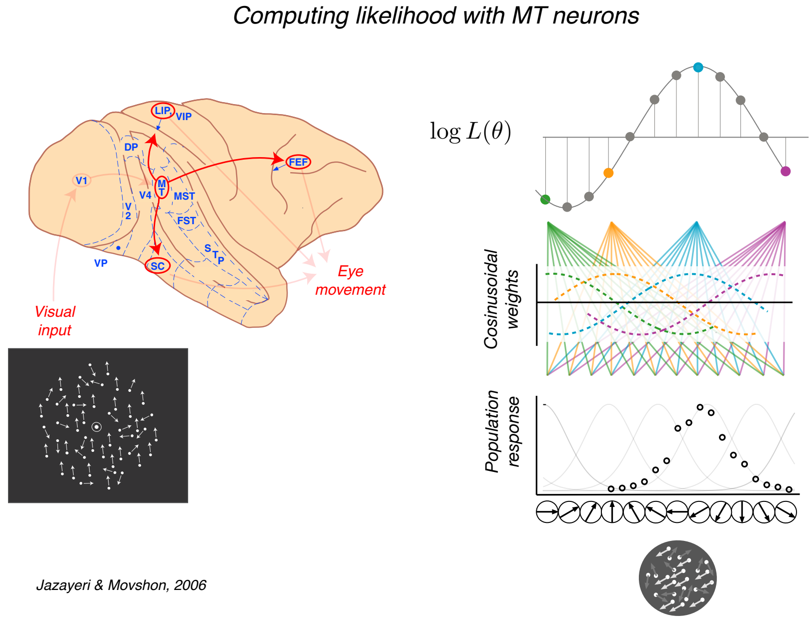


Figure 11: Approximating log-likelihood. [1]

Summary computing with neurons

- In visual cortical area V1, neurons i are tuned for directions of visual motion θ_i and their response variability is Poisson distributed.
- Jazayeri and Movshon (2006) proposed that neurons in the visual cortical area MT perform log-likelihood decoding by summing over the V1 activities x_i with suitable weights w_{ji}.

$$X_j = \sum_{i=1}^{n} w_{ji}\, x_i$$

- Specifically, they propose that an area MT neuron representing θ_j applies the weights

$$w_{ji} = \log f_i(\theta_j) \approx \cos\left(\theta_j - \theta_i\right)$$

- In this way, MT activity X_j approximates the log-likelihood that θ_j caused V1 activity $x_{1:n}$.

4 Discriminating, detecting, and identifying with log-likelihood

The decoding scheme proposed by Jazayeri & Movshon (2006) works equally well for

“discrimination tasks” (stimulus A or B?)

“detection tasks” (stimulus A absent or present?)

“identification tasks” (stimulus A, B, C, ...)

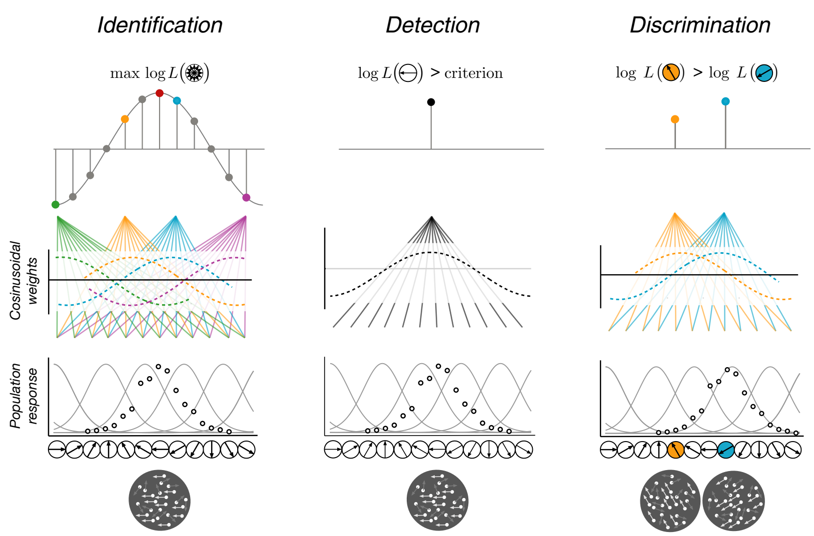


Figure 12: Left: identification. Middle: detection. Right: Discrimination. [1]

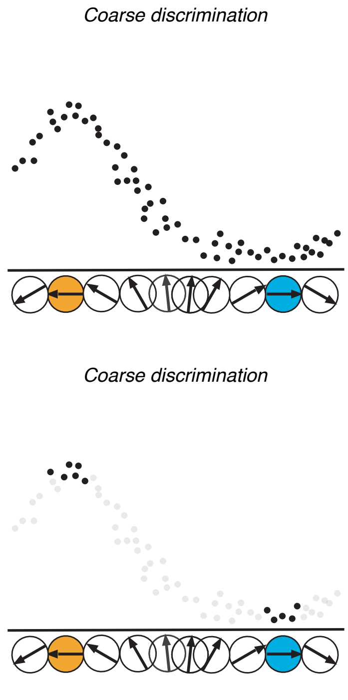


Figure 13: Informative sub-population for “coarse discrimination”. [1]

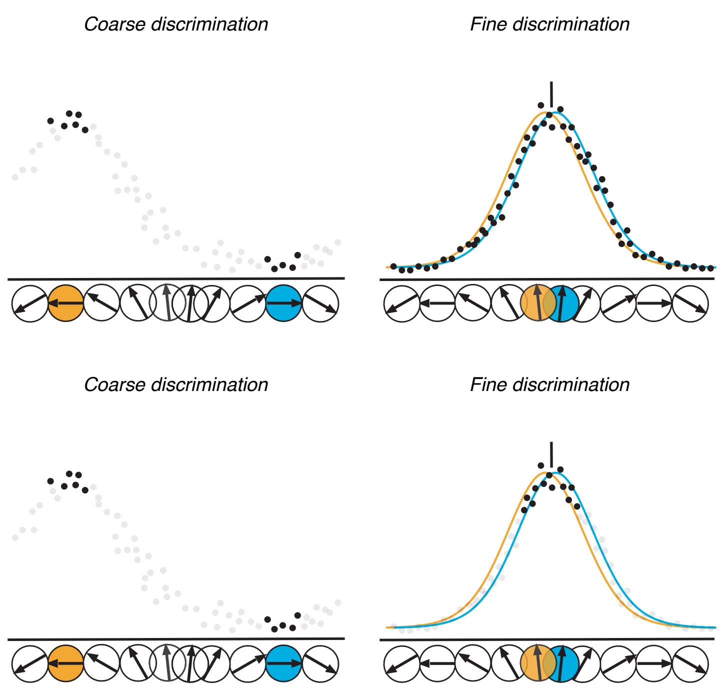


Figure 14: Informative sub-population for “fine discrimination”. [1]

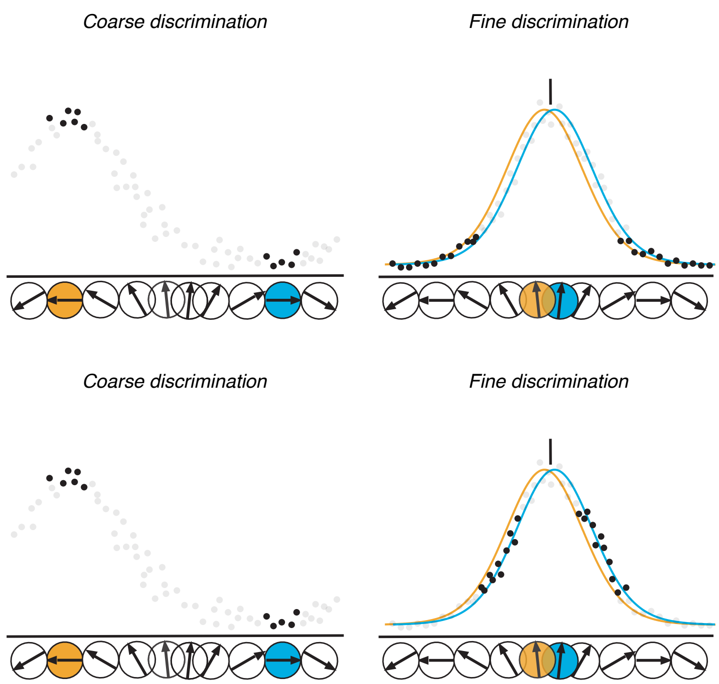


Figure 15: Informative sub-population for "fine discrimination". [1]

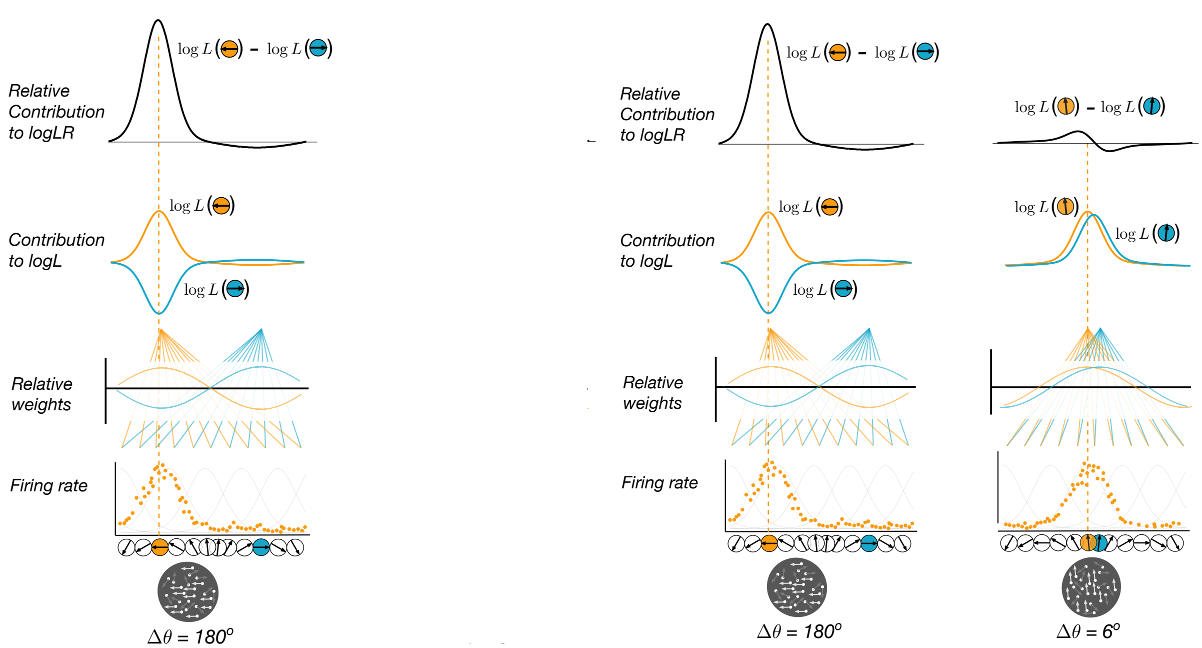


Figure 16: Contribution to log-likelihood ratio. [1]

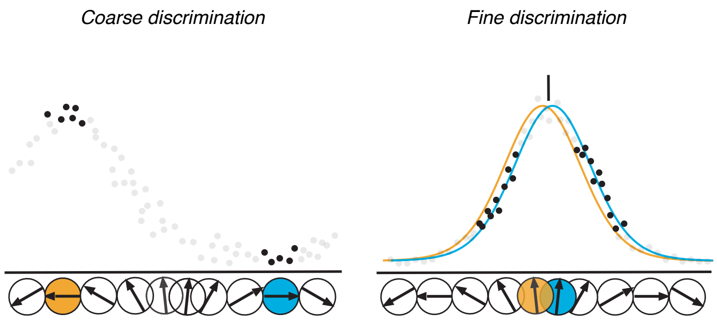


Figure 17: Informative sub-populations depend on task. [1]

Informative for fine discrimination

For fine discrimination, the most informative sub-populations are not those that respond **maximally** but, rather, those that respond **half-maximally**.

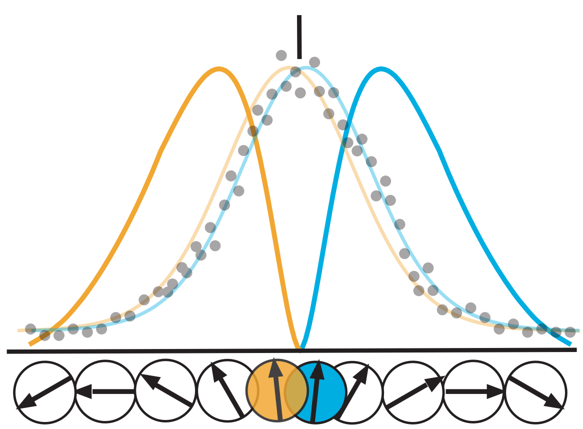


Figure 18: Informative for fine discrimination. [1]

What you should know about population decoding

- To interpret neuronal responses efficiently, we need *prior information* about the responsiveness of each neuron in the population.

- Specifically, we need to know the mean response to different stimuli (*tuning*) and the distribution of actual responses around this mean (*response variability*).

- Taken together, this provides the conditional probability $P(x|s)$ of different possible responses x, given different possible stimuli s.

- $P(x|s)$ is conditional probability of response x, given stimulus s.

- If responses vary independently, the joint probability is the product of the individual probabilities
$$P(x_{1:n}|s) = \prod_{i=1}^{n} P(x_i|s)$$

- The stimulus s maximizing the joint probability is the *maximum likelihood (ML)* estimate. Often, it is well approximated by the response-weighted average of preferred stimuli s_i:
$$\theta_{est} = \frac{\Sigma_{i=1}^{n}\ x_i\, s_i}{\Sigma_{i=1}^{n}\ x_i}$$

- $L(s|x)$ is the *posterior* likelihood of different possible stimuli s, given different observed responses x.

- $L(s|x)$ differs from $P(x|s)$ by taking into account *prior* stimulus probability $P(s)$
$$L(s|x) \propto P(x|s)\, P(s)$$

- If responses vary independently, the joint likelihood is the product of the individual likelihoods

$$L(s|x_{1:n}) = \prod_{i=1}^{n} L(s|x_i)$$

- The stimulus s maximizing the joint likelihood is the *maximum a posteriori (MAP)* estimate.

5 Bibliography

1. Dayan & Abbott (2001) Theoretical Neuroscience, MIT Press Jazayeri & Movshon (2006) Nature Neuroscience