Lecture 11

2025-11-12

Population decoding; Optimal decoding methods: Bayesian inference, Maximum a posteriori, maximum likelihood; Spike-train decoding

- Online slides: lec11-population_decoding.html
- Code: code11.ipynb
- Homework: hw11.pdf

Username: `cns`, Password: `nycu2025`

Laboratory of Computing Biological Networks

Population decoding

Benefits of multiple neurons:

- Reduction of uncertainty due to neuronal variability
- Ability to represent different stimulus attributes simultaneously

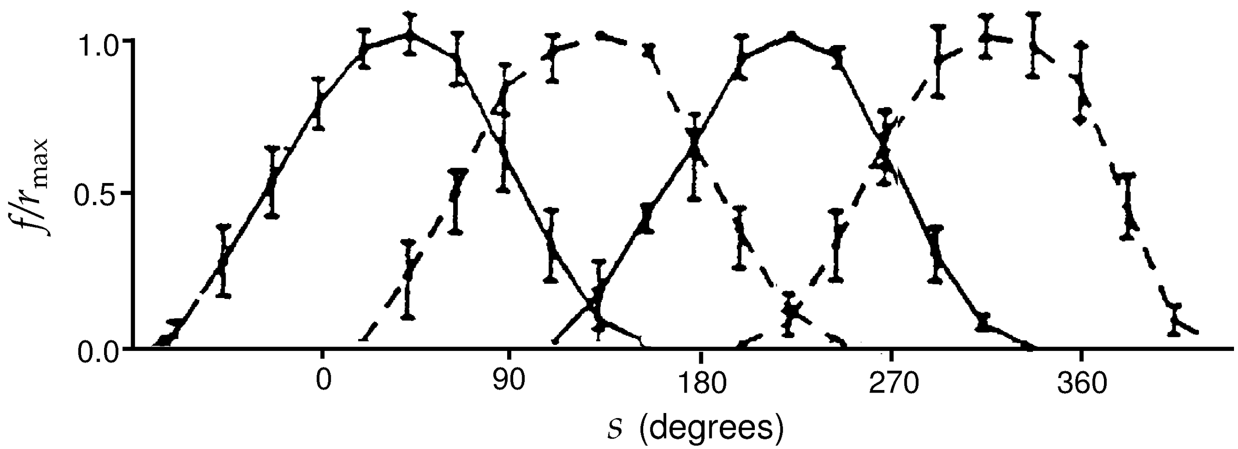
Typically: neurons have different but overlapping ranges of selectivity.

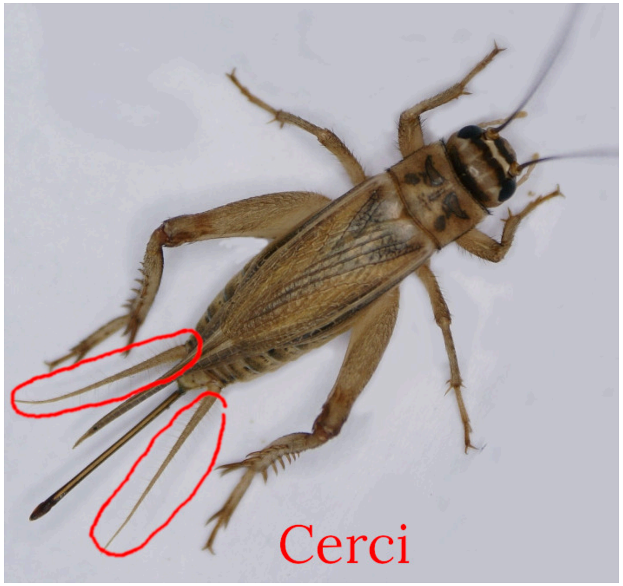
Essential difference: terms such as $p[r|s]$ is replaced by $p[\mathbf{r}|s]$.

Continue to apply: ROC analysis, likelihood ratio test, Neyman–Pearson lemma ($l(r)$ test is optimal in maximizing the power for a given size).

Discrimination → Extraction of continuous parameters

Decoding direction in cricket cercal system



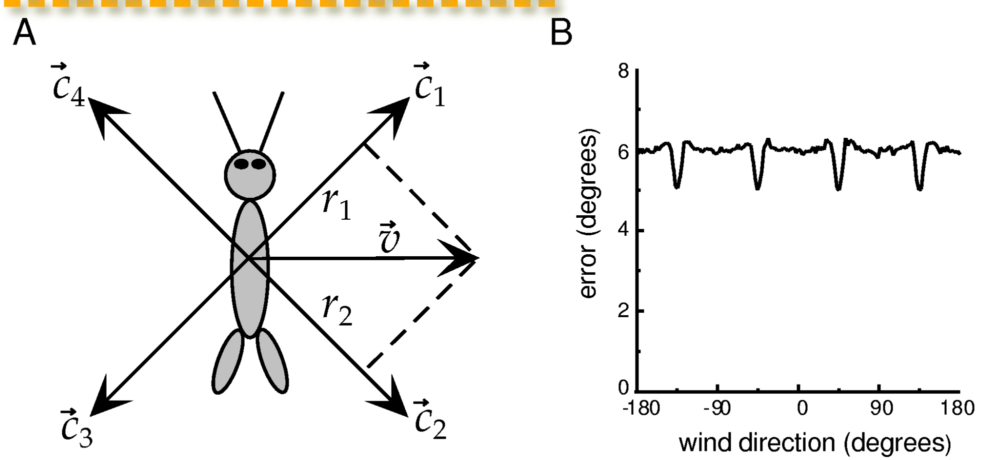


Tuning curves for the four low-velocity interneurons of the cricket cercal system, D&A Fig 3.4, adapted from Theunissen and Miller, 1991

Vector method

Sum of preferred wind vectors weighted by firing rates

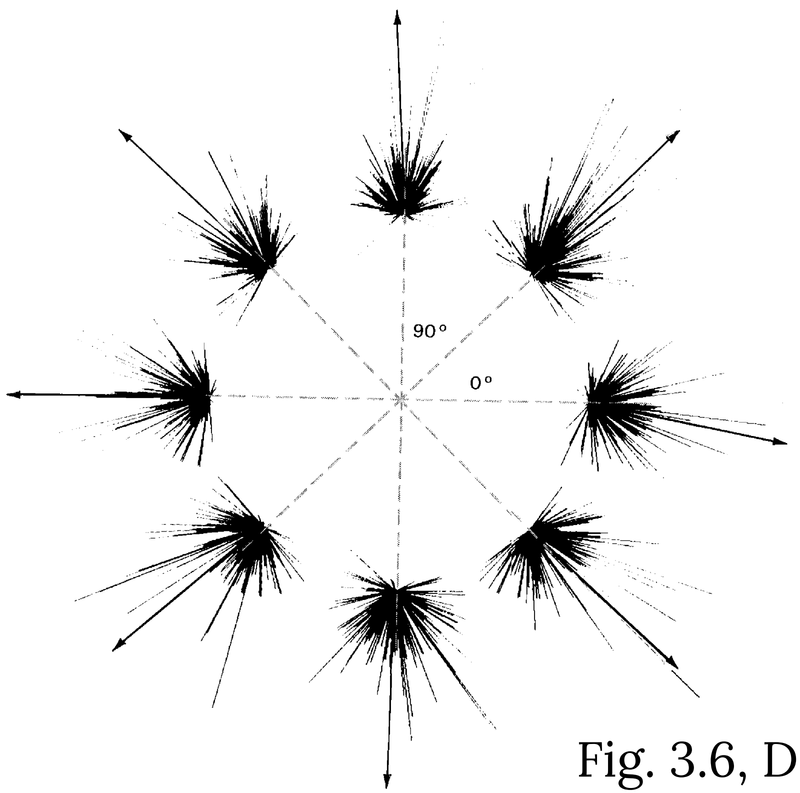
$$\vec{v}_{\text{pop}} = \sum_{a=1}^{4} \left(\frac{r}{r_{\max}} \right)_a \vec{c}_a$$



Monkey primary motor cortex (M1)

With nonzero offset rates r_0, full range:

$$\left(\frac{\langle r\rangle - r_0}{r_{\max}}\right)_a = \left(\frac{f(s) - r_0}{r_{\max}}\right)_a = \vec{v}\cdot\vec{c}_a \qquad (3.23)$$



Assume uniform prefer directions and large N

$$\vec{v}_{\text{pop}} = \sum_{a=1}^{N}\left(\frac{r - r_0}{r_{\max}}\right)_a \vec{c}_a$$

$$\Rightarrow \langle\vec{v}_{\text{pop}}\rangle = \sum_{a=1}^{N}(\vec{v}\cdot\vec{c}_a)\vec{c}_a$$

Fig. 3.6, D&A, adapted from Kandel et al, 2001

Comparing the two cases for the vector method

Cercal system

- Few neurons in Cartesian arrangement
- Rectified tuning curves sensitive to half of all angles
- Population vector ideally should give the correct wind direction

Motor cortex

- Many neurons in random uniform distribution
- Offset tuning curves sensitive to all directions
- The population vector approximates motion directly at large N limit
- Encodes additional information: e.g., initial position of arm, movement velocity and acceleration...

Optimal decoding methods

- Bayesian inference
  Finding the minimum of a loss function expressing cost of estimation errors: $s_{\text{bayes}} = \text{argmin}_s \int ds L(s', s) p[s'|\mathbf{r}]$
- Maximum a posteriori (MAP) inference
  Maximize conditional probability of the stimulus
$$s_{\text{MAP}} = \underset{s}{\text{argmax}}\, p[s|\mathbf{r}]$$
- Maximum likelihood
  Maximize the likelihood function
$$s_{\text{ML}} = \underset{s}{\text{argmax}}\, p[\mathbf{r}|s]$$
  which is the same as MAP when the prior $p[s]$ is uniform.

Bayesian loss functions

$L(s, s')$: the "cost" of reporting s' for correct s. Minimize

$$\int ds\, L(s, s') p[s|\mathbf{r}]$$

gives the optimal estimate s_{bayes}.

For $L(s, s') = (s - s')^2$, we get the mean

$$s_{\text{bayes}} = \int ds\, p[s|\mathbf{r}] s.$$

For $L(s, s') = |s - s'|$, s_{bayes} is the median of $p[s|\mathbf{r}]$.

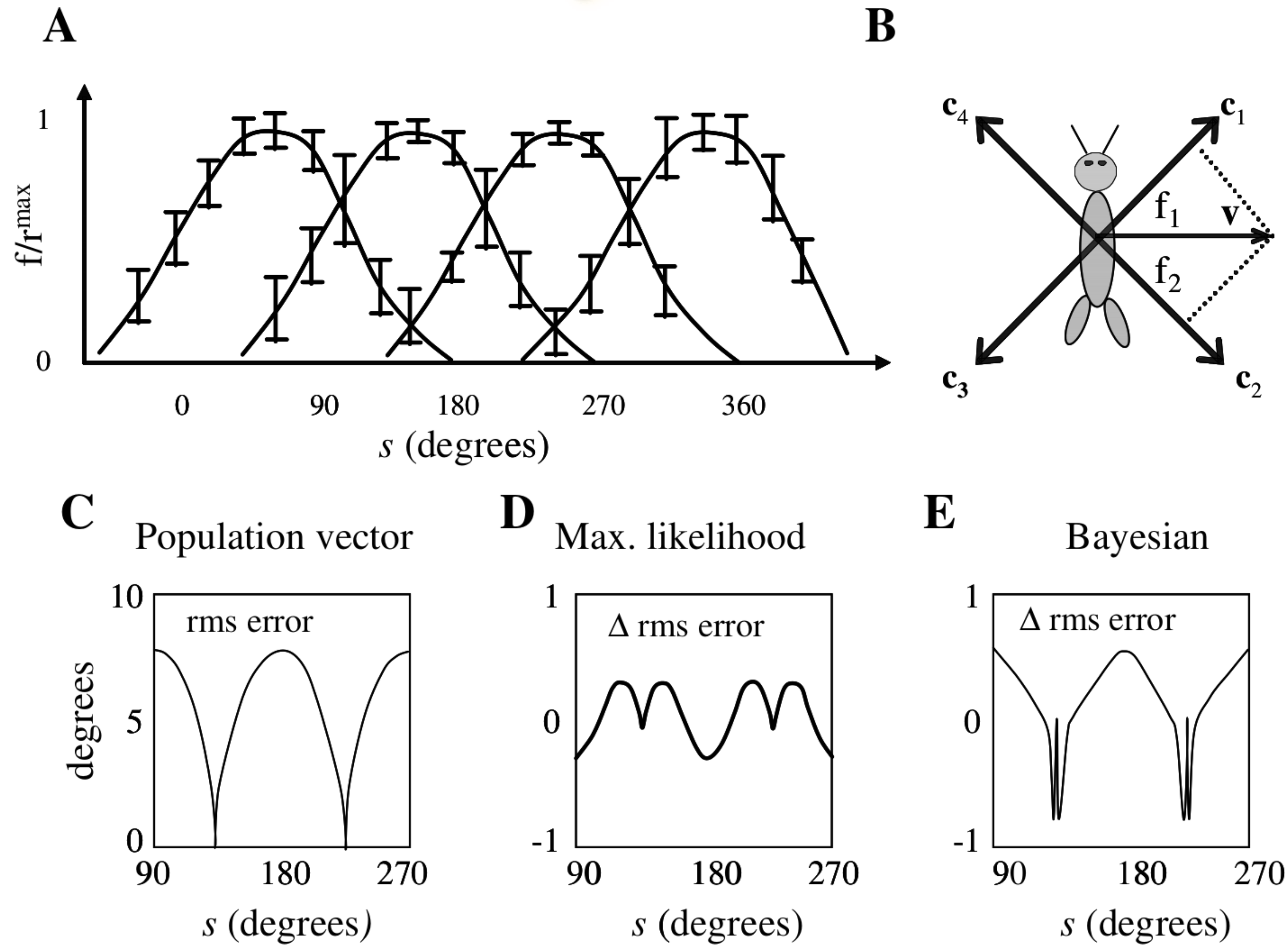
Maximum a posteriori and likelihood

For constant $p[s]$,

$$
\begin{aligned}
s_{\mathrm{MAP}} &= \operatorname*{argmax}_s p[s|\mathbf{r}] = \operatorname*{argmax}_s \frac{p[\mathbf{r}|s]\,p[s]}{p[\mathbf{r}]} \\
&= \operatorname*{argmax}_s p[\mathbf{r}|s] = s_{\mathrm{ML}}
\end{aligned}
$$

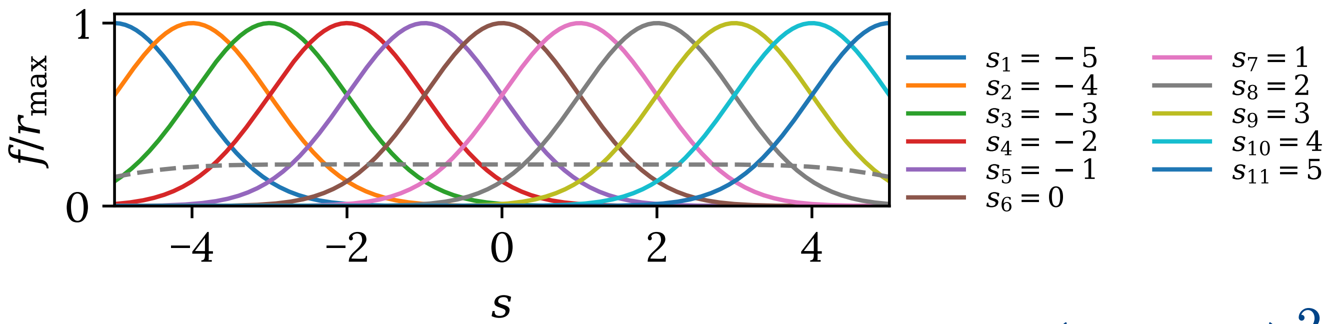
The maxima are generally obtained by taking derivatives with respective to s and setting them to zero.

Application to cercal system



[Pouget et al (2003)]

Non-directional continuous variables



Gaussian tuning curves: $f_a(s) = r_{\mathrm{max}} \exp \frac{-(s - s_a)^2}{2\sigma_a^2}$.

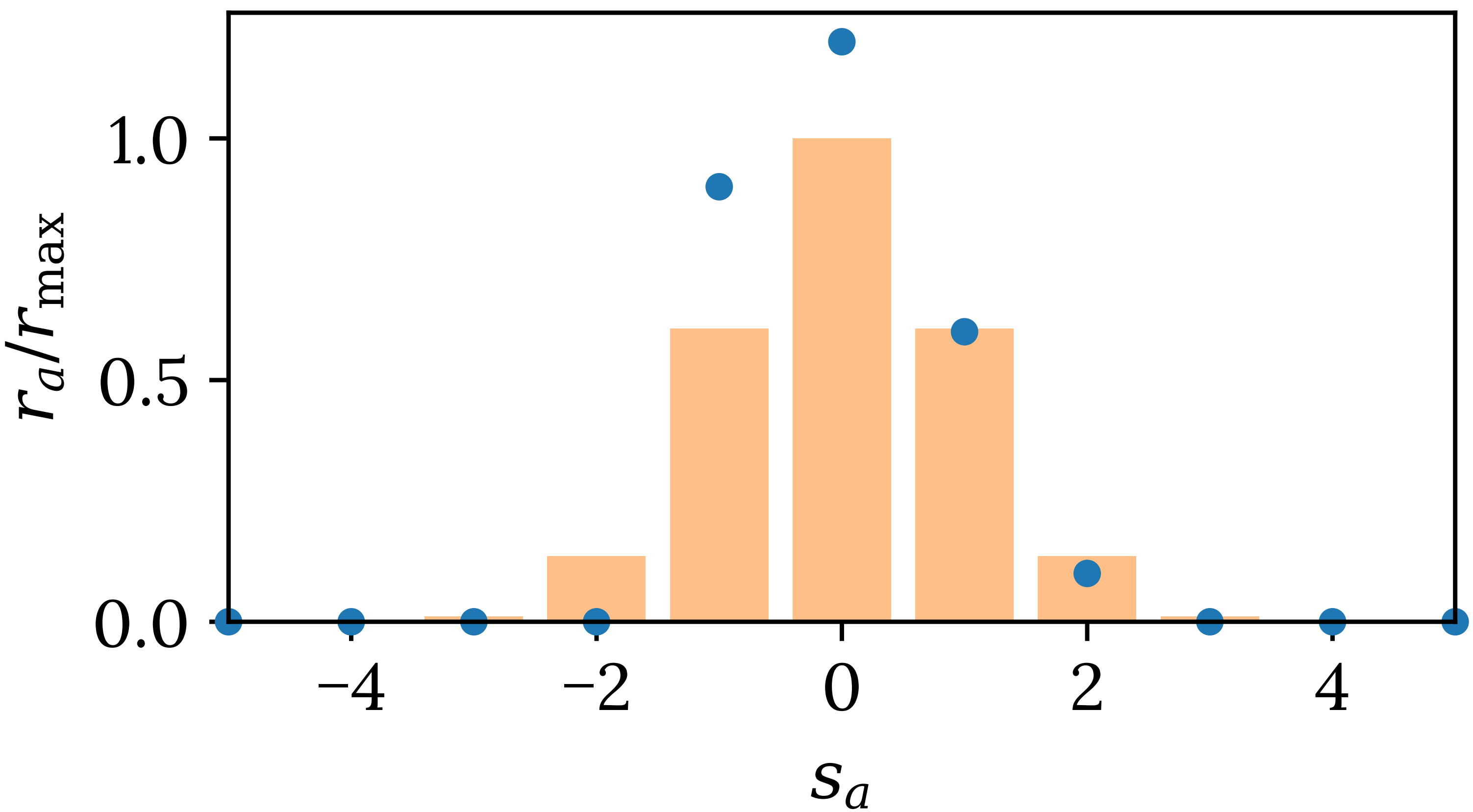
For given s, the firing rate r_a is measured by counting spikes n_a within interval T: $r_a = n_a/T$ where for large T, $\langle r_a \rangle \to f_a(s)$.

For finite T, Poisson model gives

$$P[r_a|s] = \frac{(f_a(s)T)^{r_a T}}{(r_a T)!} \exp(-f_a(s)T)$$

Example response

When $s = 0$,



Orange bars show the expected activity rates of the cells of different s_a.

Independent neuron assumption

Population firing rate probability

$$P[\mathbf{r}|s] = \prod_{a=1}^{N} \frac{(f_a(s)T)^{r_a T}}{(r_a T)!} \exp(-f_a(s)T)$$

Taking logarithm,

$$\ln P[\mathbf{r}|s] = T \sum_{a=1}^{N} r_a \ln(f_a(s)) + \ldots$$

The ellipsized is independent of s, including $f(s) = \sum_a f_a(s)$.

ML condition: $\sum_{a=1}^{N} r_a \frac{f_a'(s_{\mathrm{ML}})}{f_a(s_{\mathrm{ML}})} = 0 \quad \Rightarrow s_{\mathrm{ML}} = \frac{\sum_a r_a s_a/\sigma_a^2}{\sum_a r_a/\sigma_a^2}$

for Gaussian f_a, or $s_{\mathrm{ML}} = \frac{\sum_a r_a s_a}{\sum_a r_a}$ if all widths are the same.

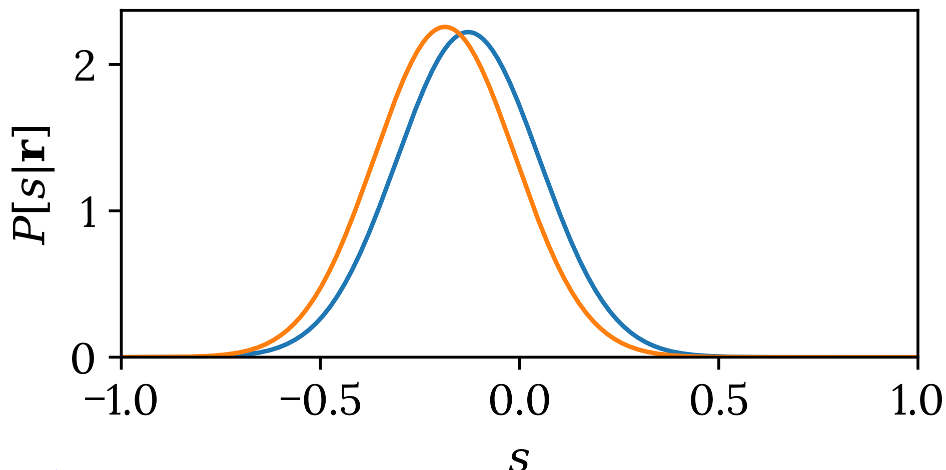
MAP condition

A posteriori

$$\ln p[s|\mathbf{r}] = T \sum_{a=1}^{N} r_a \ln\left(f_a(s)\right) + \ln p[s] + \ldots$$

$$\Rightarrow T \sum_{a=1}^{N} \frac{r_a f_a'(s_{\mathrm{MAP}})}{f_a(s_{\mathrm{MAP}})} + \frac{p'[s_{\mathrm{MAP}}]}{p[s_{\mathrm{MAP}}]} = 0$$

$$\Rightarrow s_{\mathrm{MAP}} = \frac{T \sum r_a s_a/\sigma_a^2 + s_{\mathrm{prior}}/\sigma_{\mathrm{prior}}^2}{T \sum r_a/\sigma_a^2 + 1/\sigma_{\mathrm{prior}}^2}$$



Blue: without prior;
Orange: with a Gaussian prior
$(s_{\mathrm{prior}}, \sigma_{\mathrm{prior}}^2) = (-2, 1)$

Bias and variance

Bias

Difference between average of s_{est} across trials with the given stimulus s.

$$b_{\text{est}}(s) = \langle s_{\text{est}} \rangle - s$$

Variance

How much the estimate varies about its mean value.

$$\sigma^2_{\text{est}}(s) = \left\langle \left(s_{\text{est}} - \langle s_{\text{est}} \rangle \right)^2 \right\rangle$$

Trial-average square estimation error

$$\left\langle \left(s_{\text{est}} - s \right)^2 \right\rangle = \left\langle \left(s_{\text{est}} - \langle s_{\text{est}} \rangle + b_{\text{est}}(s) \right)^2 \right\rangle = \sigma^2_{\text{est}}(s) + b^2_{\text{est}}(s)$$

Spike-train decoding

Given response spike-train, estimate the input stimulus

Linear estimate of stimulus

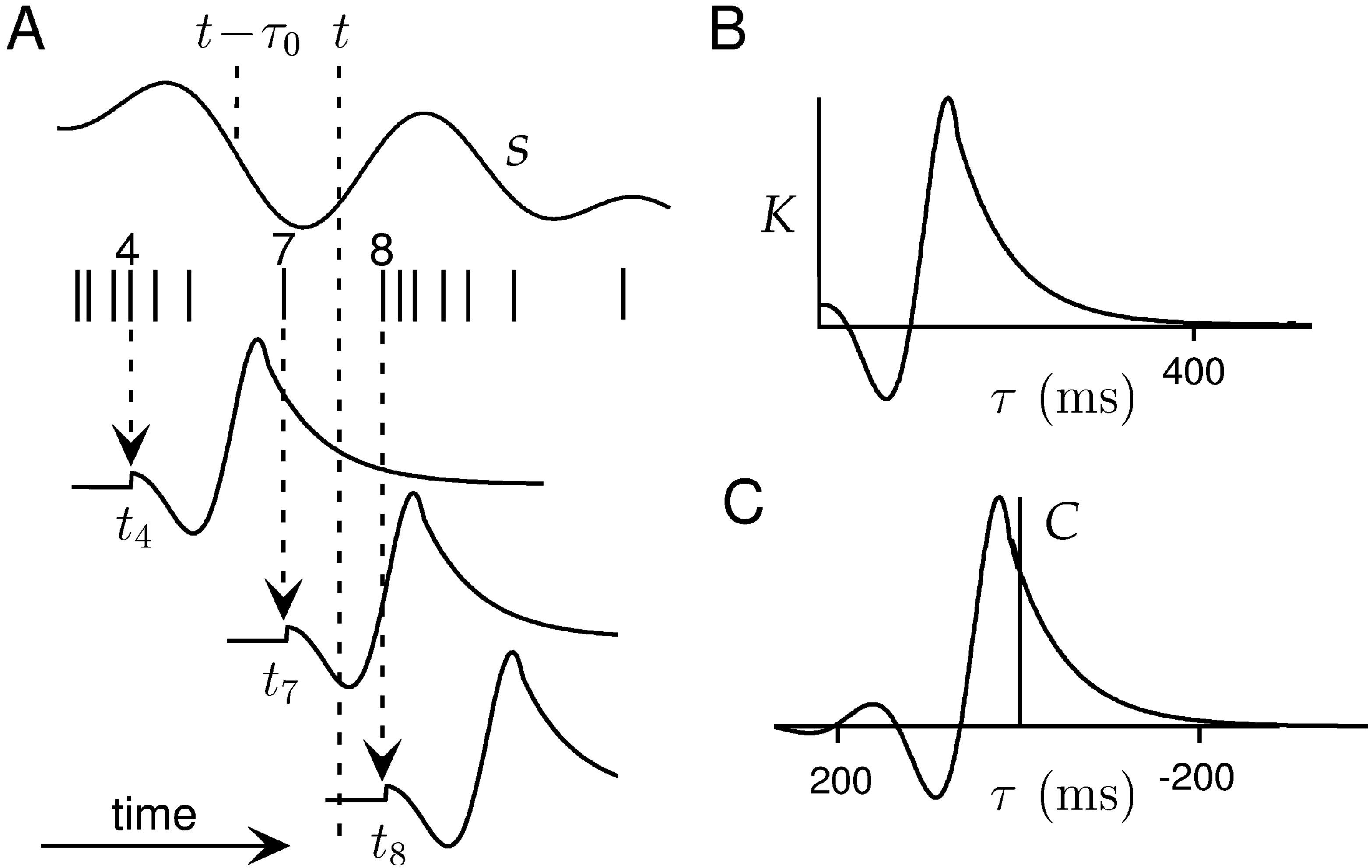
With a prediction delay τ_0, estimated stimulus:

$$s_{\text{est}}(t-\tau_0) = \sum_{i=1}^{n} K(t-t_i) - \langle r \rangle \int_{-\infty}^{\infty} d\tau K(\tau)$$

where K is the kernel for spike contributions. Last term ensures time average of s_{est} is zero. For causal kernel, $K(\tau \leq 0) = 0$. Using $\rho(t) = \sum \delta(t-t_i)$:

$$s_{\text{est}}(t-\tau_0) = \int_{-\infty}^{\infty} d\tau \left(\rho\left(t-\tau\right) - \langle r \rangle\right) K\left(\tau\right)$$

Spike-train decoding illustrated



D&A Fig. 3.13

Optimal kernel construction

Minimizing the error of prediction, $E \equiv$

$$\frac{1}{T}\int_0^T dt \left\langle \left(\int_{-\infty}^{\infty} d\tau (\rho(t-\tau) - \langle r \rangle)\, K(\tau) - s(t-\tau_0) \right)^2 \right\rangle$$

Resulting condition for K, $\partial E / \partial K(\tau) = 0$

$$\Rightarrow \int_{-\infty}^{\infty} d\tau' Q_{\rho\rho}(\tau - \tau') K(\tau') = Q_{\mathrm{r}s}(\tau - \tau_0) \qquad (3.54)$$

similar to the construction of LNP kernel with STA ...

However, unlike the stimulus in STA, we do not have control over the spike train to make it "*white*".

The auto- and cross-correlations

Autocorrelation of the spike train

$$Q_{\rho\rho}(\tau-\tau') = \frac{1}{T}\int_0^T dt\,\langle(\rho(t-\tau)-\langle r\rangle)\,(\rho(t-\tau')-\langle r\rangle)\rangle$$

Correlation between firing rate and stimulus

$$Q_{\mathrm{rs}}(\tau-\tau_0) = \langle r\rangle C(\tau_0-\tau) = \frac{1}{T}\left\langle\sum_{i=1}^{n} s(t_i+\tau-\tau_0)\right\rangle$$

where C is the spike-trigger average as discussed in Lecture 4.

For uncorrelated spike train

– tends to occur at low rate limit

We have $Q_{\rho\rho} = \langle r \rangle \delta(\tau)$ and the kernel becomes

$$K(\tau) = \frac{1}{\langle r \rangle} Q_{\mathrm{rs}}(\tau - \tau_0) = \frac{1}{\langle n \rangle} \left\langle \sum_{i=1}^{n} s(t_i + \tau - \tau_0) \right\rangle ,$$

that is, adding up stimulus around each spike time shifted by τ_0.

For causal system, spiking depends only on past stimulus. Thus, without auto-correlated stimulus, $K(\tau > \tau_0) = 0$.

For causal decoding, we can only use past spikes in prediction, that is, $K(\tau < 0) = 0$. Therefore, if there is no delay $\tau_0 = 0$, we have $K(\tau) = 0$ for all values of τ.

Acausal kernel with "colored" spike train

Solving by Fourier transformation

$$K(\tau) = \frac{1}{2\pi} \int d\omega\, \tilde{K}(\omega) \exp(-i\omega\tau).$$

Multiplying Eq. (3.54) by $\exp(i\omega\tau)$ and integrating over τ, we get

$$\int_{-\infty}^{\infty} d\tau\ \exp(i\omega\tau) \int_{-\infty}^{\infty} d\tau' Q_{\rho\rho}(\tau-\tau')K(\tau') = \int_{-\infty}^{\infty} d\tau\, Q_{\mathrm{rs}}(\tau-\tau_0).$$

$$\Rightarrow \tilde{Q}_{\rho\rho}(\omega)\tilde{K}(\omega) = \exp(i\omega\tau_0)\tilde{Q}_{\mathrm{rs}}(\omega)$$

which becomes algebraic, and we have the solution

$$\tilde{K}(\omega) = \frac{\tilde{Q}_{\mathrm{rs}}(\omega)\exp(i\omega\tau_0)}{\tilde{Q}_{\rho\rho}(\omega)}.$$

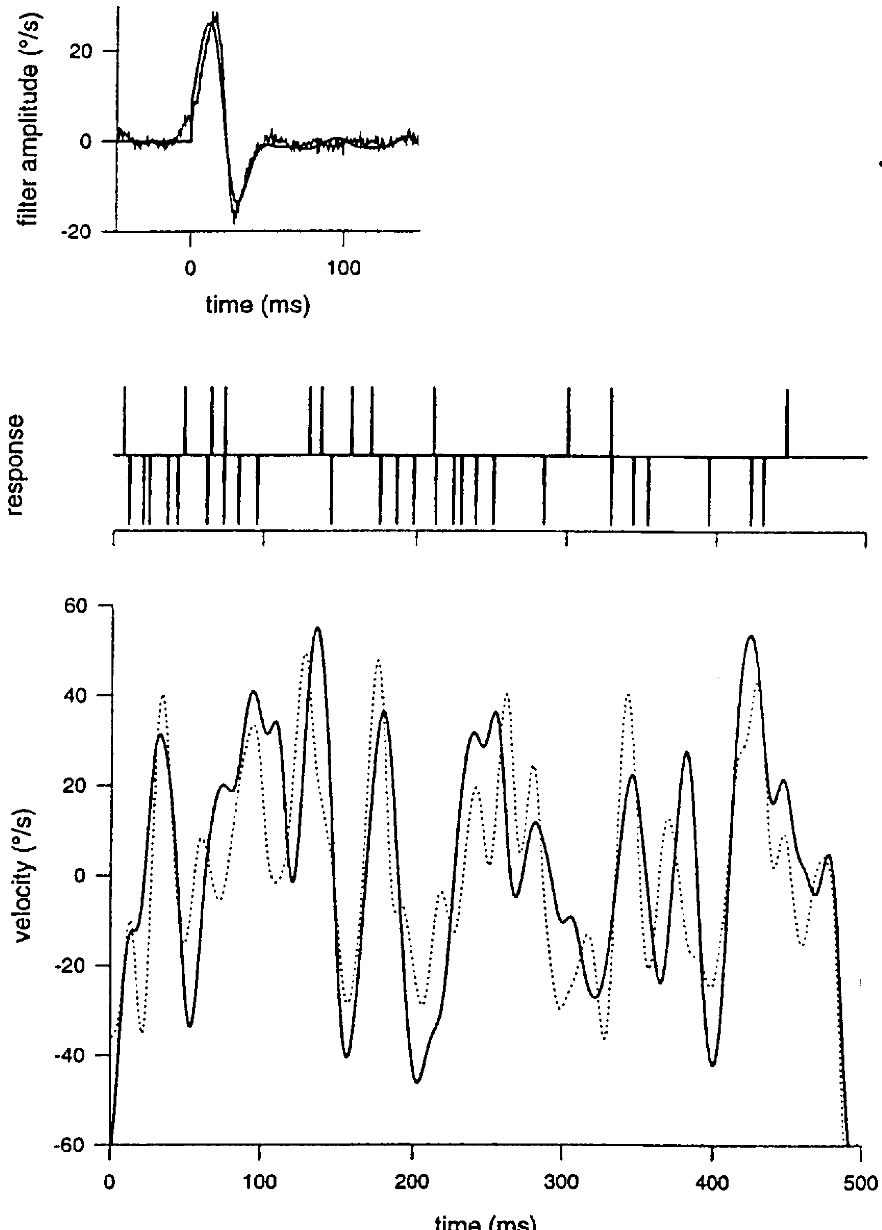
in the Fourier space. $K(\tau)$ can then be obtained by the Fourier transform.

Take care of the causality

- Truncation — After obtaining $K(\tau)$, use $\Theta(\tau)K(\tau)$, where $\Theta(\tau) = \mathrm{I}(\tau > 0)$ is the step function. → Not optimal but close
- Use causal basis functions (instead of the Fourier basis) for expansion and optimization. → Optimal
- Fixed at a conforming functional form with finite number of parameters and optimize with respect to these parameters. → Not optimal but simple
- Use techniques involving spectral factorization of $\tilde{Q}_{\rho\rho}(\omega)$. → Optimal (Wiener-Hopf filter)

* Note: All these are within the linear theory of neural processing...

Example: decoding fly H1 neuron



Upper panel: the decoding kernel. jagged curve: optimal acausal filter; the smooth curve: from causal set of basis functions.

Middle panel: typical responses of H1 to the stimuli $s(t)$ (upper) and $-s(t)$ (bottom)

Lower panel: dashed line: actual stimulus; solid line: estimated stimulus from the optimal linear reconstruction using the acausal filter