Population coding in a fast-changing world

Quentin JM Huys[1], Richard S Zemel[2], Rama Natarajan[2], Peter Dayan[1]

[1]Gatsby Computational Neuroscience Unit, University College London, UK; [2] Dept. of Computer Science, University of Toronto, Canada
{qhuys, dayan}@gatsby.ucl.ac.uk, {zemel, rama}@cs.toronto.edu

Introduction

Probabilistic population coding **in time**

Stimulus inference on timescale of spike production
is an underconstrained problem ⟶ need **prior**

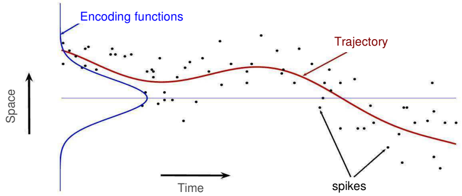
We analyse a very simple case
- Gaussian process prior over stimulus trajectory
- Bell-shaped tuning functions
- Independent Poisson noise

We present an exact expression for the posterior

$$\boxed{p(s_T|\boldsymbol{\xi}_{(0,T]}) = ?}$$

And a simple approximation

$$\boxed{\hat{p}(s_T|\boldsymbol{\xi}_{(0,T]}) = ?}$$



Posterior distribution

We find the posterior distribution over the stimulus at time T given all spikes observed so far (up to time T) by Bayes theorem

$$p(s_T|\boldsymbol{\xi}_{(0,T]}) \overset{\text{Bayes}}{\propto} p(\boldsymbol{\xi}_{(0,T]}|s_T)\, p(s_T)$$
$$= \int ds_{(0,T]}\, p(\boldsymbol{\xi}_{(0,T]}|s_{(0,T]})\, p(s_{(0,T]}, s_T)$$

Assumption: independent, identical Poisson neurones

$$p(\boldsymbol{\xi}_{(0,T]}|s_{(0,T]}) \overset{\text{Poisson}}{=} \prod_{t=0}^{T} p(\boldsymbol{\xi}_t|s_t) \overset{\text{nms}}{=} \prod_{t=0}^{T}\prod_i \phi_i(s_t)e^{-\sum_i \int^t d\tau\, \phi_i(s(\tau))} \propto \prod_{i,t}\phi_i(s(t))$$

Assumption: Gaussian process prior over **entire** stimulus (trajectory)

$p(s_{(0,T]}, s_T) = \mathcal{N}(\mathbf{m}, \mathcal{C})$ Gaussian process prior

$\mathcal{C}_{t,t'} = \lambda^{\|t-t'\|^x}$ OU process x = 1  Smooth process x = 2 

Result: Posterior distribution is a simple Gaussian

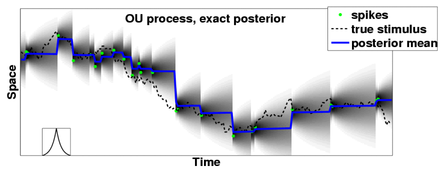
$$\boxed{\begin{aligned} p(s_T|\boldsymbol{\xi}_{(0,T]}) &= \mathcal{N}_{s_T}(\mathbf{k}\boldsymbol{\Theta}, \mathcal{C}_{TT} - \mathbf{k}\mathcal{C}_{tT}) \\ \mathbf{k} &= \mathcal{C}_{Tt}(\mathcal{C}_{tt} + \mathbf{I}\sigma^2)^{-1} \end{aligned}}$$

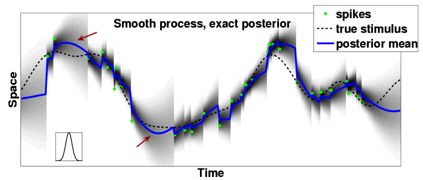
$$\mathcal{C} = \begin{pmatrix} \mathcal{C}_{tt} & \mathcal{C}_{tT} \\ \mathcal{C}_{Tt} & \mathcal{C}_{TT} \end{pmatrix}$$

θ_t = preferred position of neurone that emitted spike at time t

k_{T-t} = "weight" of spike at time t

Example: Posterior distribution for a Ornstein-Uhlenbeck and smooth prior



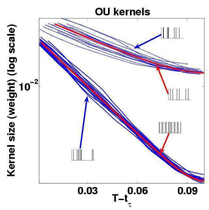


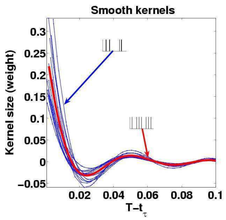
Exact kernels k

$$\mathbf{k} = \mathcal{C}_{Tt}(\mathcal{C}_{tt} + \mathbf{I}\sigma^2)^{-1}$$

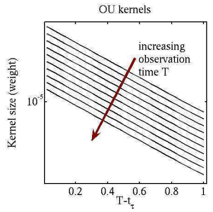
a) depends only on spike and observation times, not on spike locations
b) determines the weight of each spike
c) has a shape that is determined by the covariance of the Gaussian process prior

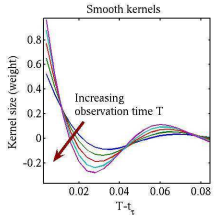
Exact and metronomic kernels



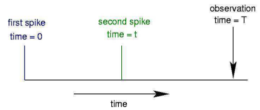


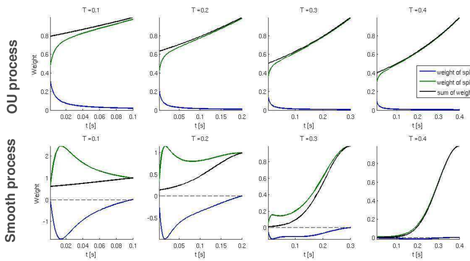
Observation time T relative to most recent spike





A simple case with two spikes elucidates the two types of dependence on tpike times



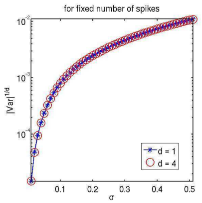


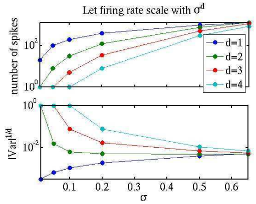
References
[1] Brown et al. (1998)
[2] Zemel, Natarajan, Huys and Dayan, Cosyne 2005
[3] Parts of this work were presented in Zemel, Huys, Natarajan and Dayan, NIPS 2005

Exact posterior variance

Dependence on σ^2

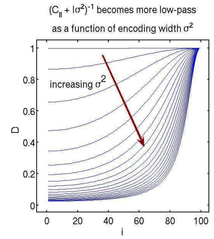
Narrower encoding tuning functions (smaller σ^2) gives narrower posterior for fixed number of spikes



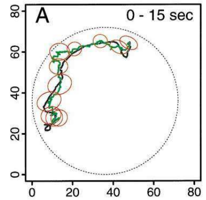


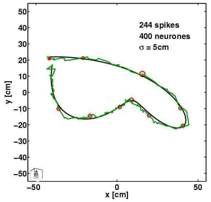
As encoding width σ^2 increases the covariance vector C_{lT} is projected onto eigenvectors of lower frequency.

$$\nu^2 = \mathcal{C}_{TT} - \mathbf{k}\mathcal{C}_{lT}$$
$$\approx \mathcal{C}_{TT} - a\mathbf{D}\mathbf{V}^T\mathcal{C}_{lT}$$



Example: Hippocampus, smooth trajectories

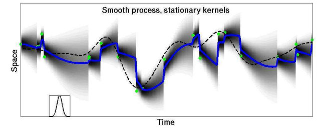




Brown et al. (1998)

Approximations

Metronomic



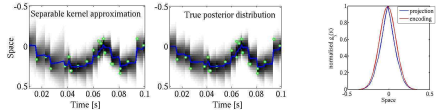
Approximate posterior distribution by treating each spike as an expert in a Product of Experts

$$\boxed{\hat{p}(s_T|\boldsymbol{\xi}_{(0,T]}) \propto \exp\left(\sum_{i,\tau} g(i,s,\tau)\xi(i,T-\tau)\right)}$$

Minimize Kullback Leibler Divergence with respect to projective fields

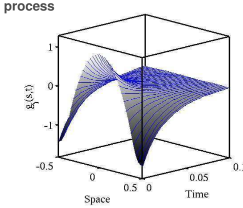
$$g(i,s,t) \leftarrow g(i,s,t) - \epsilon\frac{\partial D(p\|\hat{p})}{\partial g(i,s,t)}$$

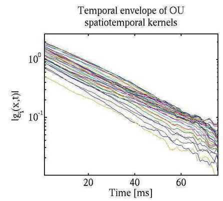
Inferring the temporal and spatial kernels separately, a good approximation is only obtained for the OU process.



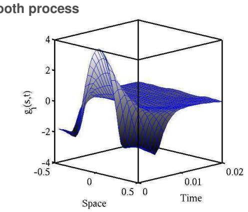
When allowing spatiotemporally nonseparable projective fields we still infer approximately separable projective fields.
A good approximation is only obtained for the OU process.

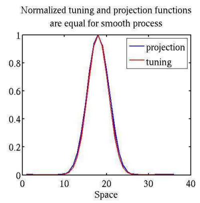
OU process





Smooth process





Conclusions

Representation of **time-varying probabilistic information** in a **population of spiking neurones**

Spike-by-spike decoding: The effect of each spike
- is described by a kernel
- depends on the prior
- depends on other spikes if the process is smoother than OU

Width of the posterior
- Narrower encoding tuning functions are always better (in the dense regime)

Approximations treating each spike as an expert in a Product of Experts setting
- Projective fields tend to be separable
- Interactions between spikes can not be captured by independent treatment of the spikes in smooth process
- Interactions between spikes do not produce spatiotemporally inseparable projective fields
  - use two-layer recurrent network [2]