Suplementary Information to: *Error-based Analysis of Optimal Encoding of Dynamic Stimuli: An Analytical Study*

Alex Susemihl
Bernstein Center for Computational Neuroscience Berlin,Technische Universität Berlin
alex.susemihl@bccn-berlin.de

Ron Meir
Department of Eletrical Engineering, Technion, Haifa
rmeir@ee.technion.ac.il

Manfred Opper
Bernstein Center for Computational Neuroscience Berlin, Technische Universität Berlin
opperm@cs.tu-berlin.de

1 Gaussian Process Inference from Point Processes

We follow the demonstration in [1]. We have

$$P(\boldsymbol{\xi}|X) = \exp\left(-\sum_m \int_{t_0}^{t_f} \lambda_m(X(t),t)dt\right) \prod_i \lambda_{m_i}(X(t_i),t_i),$$

where X represents all values of $X(t)$ for $t_0 < t < t_f$. Using that $\sum_m \lambda_m(X(t),t) = \lambda(t)$ we can write

$$P(\boldsymbol{\xi}|X) = \exp\left(-\int_{t_0}^{t_f} \lambda(t)dt - \sum_i (X(t_i) - \theta_{m_i})^2/2\alpha(t_i)^2\right)$$

We want to obtain the marginal probability

$$P(X(t_f)|\boldsymbol{\xi}) \propto P(X(t_f))P(\boldsymbol{\xi}|X(t_f)) = P(X(t_f)) \int d\mu(X_{[t_0,t_f)})P(\boldsymbol{\xi}|X_{[t_0,t_f)})P(X_{[t_0,t_f)}|X(t_f)),$$

where we abuse the notation saying that $\int d\mu(X_{[t_0,t_f)})$ is an integral over all values of $X(t)$ apart from $X(t_f)$. Since $X(t)$ is a Gaussian process with covariance function $K(t - t')$, we have $P(X(t_1), X(t_2), \ldots, X(t_n)) = \mathcal{N}(X(t_i); 0, K)$, where $K_{i,j} = K(t_i - t_j)$. We then have $P(X(t_1), \ldots, X(t_n)|X(t_f)) = \mathcal{N}(X(t_1), \ldots, X(t_n); m_i, \mathcal{K}_{i,j})$, with $m_i = X(t_f)K(t_i - t_f)/K(0)$ and

$$\mathcal{K}_{i,j} = K(t_i - t_j) - K(t_f - t_i)K(t_f - t_j)/K(0).$$

Since the integral is a convolution, the variances and means simply add up. Evaluating the integrals and using the Woodbury formula we obtain

$$P(\boldsymbol{\xi}|X(t_f)) = \mathcal{N}(\theta; m_i, \mathcal{K} + \text{Diag}(\alpha(t_i)^2)).$$

And finally, applying the matrix inversion lemma we obtain $P(X(t_f)|\boldsymbol{\xi}) = \mathcal{N}(X(t_f); \mu(t,\boldsymbol{\xi}), \sigma^2(t,\boldsymbol{\xi})$, with

$$\mu(t,\boldsymbol{\xi}) = \sum_{i,j} K(t - t_i)C_{ij}^{-1}\Theta_j, \quad \sigma^2(t,\boldsymbol{\xi}) = K(0) - \sum_{i,j} K(t - t_i)C_{ij}^{-1}K(t_j - t),^1 \quad (1)$$

---

[1]$C_{ij}(\boldsymbol{\xi}) = K(t_i - t_j) + \delta_{ij}\alpha(t_i)^2$

1

2 Differential Chapman-Kolmogorov Equation

Given a random process $X(t)$ with a transition probability $P(z, t'|x, t) = P(X(t') = z|X(t) = x)$, we define the quantities for all $\epsilon > 0$

$$W(x|z, t) = \lim_{\Delta t \to 0} P(x, t + \Delta t|z, t)/\Delta t, \forall x, z \quad \text{such that} \quad |x - z| \geq \epsilon,$$

$$A_i(z, t) = \frac{1}{\Delta t} \lim_{\Delta t \to 0} \int_{|x-z|<\epsilon} dx(x_i - z_i)P(x, t + \Delta t|z, t) + O(\epsilon), \text{ and}$$

$$B_{ij}(z, t) = \frac{1}{\Delta t} \lim_{\Delta t \to 0} \int_{|x-z|<\epsilon} dx(x_i - z_i)(x_j - z_j)P(x, t + \Delta t|z, t) + O(\epsilon).$$

The evolution of $P(z, t|y, t')$ is then given by

$$
\begin{aligned}
\frac{\partial P(z, t|y, t')}{\partial t} = \quad & -\sum_i \frac{\partial}{\partial z_i} \left[A_i(z, t)P(z, t|y, t')\right] + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_i} \left[B_{ij}(z, t)P(z, t|y, t')\right] \\
& + \int dx \left[W(z|x, t)P(x, t|y, t') - W(x|z, t)P(z, t|y, t')\right]
\end{aligned}
\tag{2}
$$

When $W(z|x, t) = 0$ this reduces to the Fokker-Planck equation and when $A_i(z, t) = B_{ij}(z, t) = 0$ we obtain the master equation. We refer the interested reader to [2, p. 47] for more details.

3 Filtering Smooth Markov Processes from Spike Trains

We consider Markov processes to obtain processes of a certain correlation structure. Consider a continuous-time version of the classic autoregressive process of order P

$$\sum a_p X^{(p)}(t) = bZ(t),$$

where $X^{(p)}(t)$ denotes the p-th derivative of the process $X(t)$. This can be cast into the form of a system of stochastic differential equations

$$dX_i = X_{i+1}dt, \quad i \in \{0, p-2\}, \quad a_P dX_{P-1} = -\sum_{i=0}^{P-1} a_i X_i dt + bdW, \tag{3}$$

this can be easily simulated via an Euler numerical integration scheme. The correlation structure of the stochastic process $X_0(t)$ can be obtained as follows. We start by calculating the Fourier transform of equation 3. We obtain

$$\sum a_p (2\pi i w)^p \tilde{X}(w) = b\tilde{Z}(w)$$

Taking a product with itself and averaging, we obtain the power spectrum of X

$$\left(\sum a_p (2\pi i w)^p\right) \left(\sum a_p (-2\pi i w')^p\right) \left\langle \tilde{X}(w)\tilde{X}(w') \right\rangle = b^2.$$

Taking $a_p = \binom{P}{p}\gamma^{P-p}$, we obtain

$$\left\langle \tilde{X}(w)\tilde{X}(w') \right\rangle = \frac{b^2}{(\gamma + 2\pi i w)^P (\gamma - 2\pi i w')^P}.$$

For this particular choice of process, we can Fourier transform the power spectrum $\left\langle \tilde{X}(w)\tilde{X}(w) \right\rangle$ to obtain the autocorrelation of $X_0(t)$. We obtain

$$\langle X_0(t)X_0(t + \tau) \rangle = \frac{\eta^2 2^{-\nu}}{\sqrt{\pi}\Gamma(\nu + 1/2)\gamma^\nu} (\gamma\tau)^\nu K_\nu(\gamma\tau),$$

where $K_\nu(t)$ is the modified Bessel function of the second kind, and $P = \nu - 1/2$. The advantage of using Markov processes is that the covariance between observations does not depend explicitly on the observations. We rewrite the stochastic equations as a multidimensional OU process Y:

$$d\vec{Y} = -\Gamma\vec{Y}dt + HdW,$$

2

with
$$\Gamma_{i,j} = -\delta_{i+1,j} + \delta_{P,i}\binom{P}{j-1}\gamma^{P-j+1}, \quad H_{i,j} = \delta_{i,P}\delta j, P\eta.$$

Note that to avoid confusion we have renamed the process to Y, so that $Y_i = X_{i-1}, Y_1 = X_0$. From standard OU process theory, we can write the evolution of the mean $\mu = \langle Y \rangle$ and covariance matrix $\Sigma_{i,j} = \langle Y_i Y_j \rangle$ as

$$\dot{\mu} = -\Gamma\mu, \quad \dot{\Sigma} = -\Gamma\Sigma - \Sigma\Gamma^T + H^2.$$

Under the assumption of dense tuning functions, which in turn implies that the probability of no spikes occurring is independent of the value of the observed value, the optimal Bayesian filtering scheme evolves according to the equations above in the absence of spikes. The posterior estimator for $X_1(t)$ is simply given by $\mu_1(t)$ with a variance associated to the estimative of $\Sigma_{1,1}(t)$. In the case of a spike, we have to update the distribution according to Bayes' rule. The likelihood of an observed spike at time t is given by $P(spike|Y_1(t)) = \phi \exp^{-(\theta-Y_1(t))^2/2\alpha^2}$. We write the precision matrix $\Lambda = \Sigma^{-1}$ as follows

$$\Lambda = \begin{pmatrix} c & b^t \\ b & A \end{pmatrix}.$$

By matrix inversion lemmas we have

$$\Sigma = \begin{pmatrix} (c - b^t A^{-1} b)^{-1} & (c - b^t A^{-1} b)^{-1} A^{-1} b^t \\ (c - b^t A^{-1} b)^{-1} b A^{-1} & A^{-1} + (c - b^t A^{-1} b)^{-1} A^{-1} b b^t A^{-1} \end{pmatrix}.$$

The multiplication by the likelihood results in a Gaussian distribution with precision matrix

$$\Lambda' = \begin{pmatrix} c + \frac{1}{\alpha^2} & b^t \\ b & A \end{pmatrix}.$$

Inverting we obtain

$$\Sigma' = \begin{pmatrix} (c + \frac{1}{\alpha^2} - b^t A^{-1} b)^{-1} & (c + \frac{1}{\alpha^2} - b^t A^{-1} b)^{-1} A^{-1} b^t \\ (c + \frac{1}{\alpha^2} - b^t A^{-1} b)^{-1} b A^{-1} & A^{-1} + (c + \frac{1}{\alpha^2} - b^t A^{-1} b)^{-1} A^{-1} b b^t A^{-1} \end{pmatrix}.$$

Rearranging and writing $k = c - b^t A^{-1} b$, we obtain

$$\Sigma' = \begin{pmatrix} (1 + \frac{1}{\alpha^2 k})^{-1} k^{-1} & (1 + \frac{1}{\alpha^2 k})^{-1} k^{-1} A^{-1} b^t \\ (1 + \frac{1}{\alpha^2 k})^{-1} k^{-1} b A^{-1} & A^{-1} + (1 + \frac{1}{\alpha^2 k})^{-1} k^{-1} A^{-1} b b^t A^{-1} \end{pmatrix}.$$

Rearranging and substituting the terms for $\Sigma_{i,j}$ we obtain

$$\Sigma' = \begin{pmatrix} \frac{\Sigma_{1,1}}{1 + \frac{\Sigma_{1,1}}{\alpha^2}} & \frac{\Sigma_{1,i}}{1 + \frac{\Sigma_{1,1}}{\alpha^2}} \\ \frac{\Sigma_{i,1}}{1 + \frac{\Sigma_{1,1}}{\alpha^2}} & \Sigma_{i,j} - \frac{\Sigma_{1,i}\Sigma_{i,j}}{\alpha^2 + \Sigma_{1,1}} \end{pmatrix}.$$

The update relations for the mean are also straightforward. We obtain

$$\mu' = \Sigma'\left(\Sigma^{-1}\mu + \frac{\Theta}{\alpha^2}\right),$$

where $\Theta = (\theta, 0, \dots)$ is a vector with the preferred stimulus of the spiking neuron θ in the first coordinate and zero elsewhere. This can also be written more summarily, writing the likelihood as $P(spike|Y_1(t)) = \phi \exp^{-(\Theta - Y(t))^t A^+ (\Theta - Y(t))/2}$, where $A_{i,j} = \delta_{1,i}\delta_{1,j}\alpha^2$ and $A^+ = \delta_{1,i}\delta_{1,j}/\alpha^2$ is the pseudoinverse of A. The relations then simplify to

$$\Sigma' = \left(\Sigma^{-1} + A^+\right)^{-1},$$

and

$$\mu' = \left(\Sigma^{-1} + A^+\right)^{-1}\left(\Sigma^{-1}\mu + A^+\Theta\right).$$

When A^+ is of the form mentioned above we can simplify it further, and obtain

$$\Sigma' = \Sigma - \Sigma_{:,1}\Sigma_{:,1}^t/(\alpha^2 + \Sigma_{1,1}),$$

in which no matrix inversion is required. Here $\Sigma_{:,1} = (\Sigma_{1,1}, \Sigma2, 1, \dots)^t$.

We can summarize this dynamics in two sets of filtering equations for the mean and variance of the estimator. We have then

$$\dot{\mu}(t) = -\Gamma\mu(t) + S(t)\left(\Sigma(t)^{-1} + A^+\right)^{-1} A^+ (\Theta_t - \mu(t))$$

and

$$\dot{\Sigma}(t) = -\Gamma\Sigma(t) - \Sigma(t)\Gamma^t + H^2 + S(t)\left(\Sigma(t)^{-1} + A^+\right)^{-1} A^+\Sigma(t),$$

where $S(t) = \sum_i \delta(t - t_i)$ is a sum of Dirac deltas over the spike times and $\Theta_t = (\theta_t, 0, \dots)$, where θ_t is the preferred stimulus of the spiking neuron.

3

4 Additional Figures
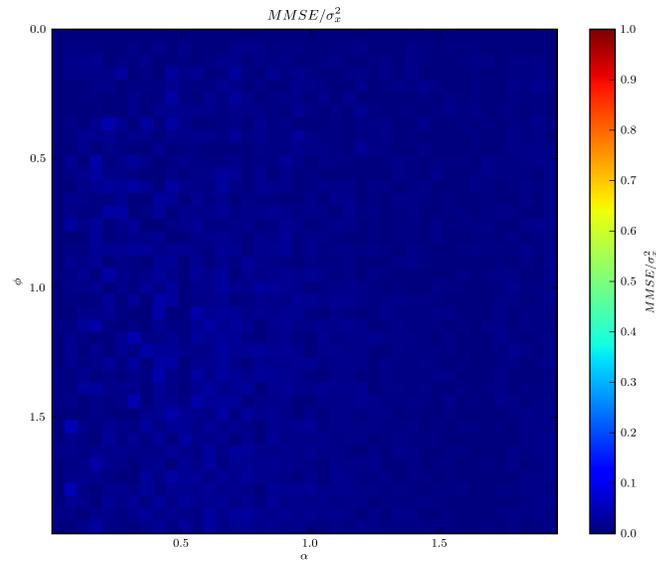
Figure 1: Relative error of the mean-field approach for the OU process.
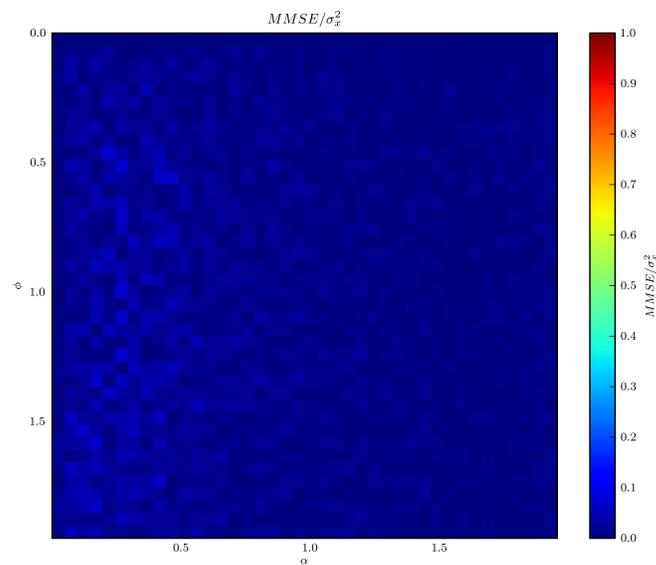
Figure 2: Relative error of the mean-field approach for the second-order Markov process.

References

[1] Quentin J. M. Huys, Richard S. Zemel, Rama Natarajan, and Peter Dayan. Fast population coding. *Neural Computation*, 19(2):404–441, 2007.

4
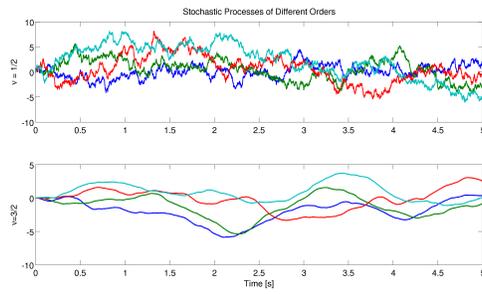
Figure 3: Stochastic processes of different orders as described above. From top to bottom we have $p = 1, 2$ corresponding to Matern kernels with $\nu = 1/2, 3/2$.

[2] C.W. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences*, volume 13 of *Springer Serier in Synergetics*. Springer, Berlin Heidelberg, fourth edition, 2009.

5