A Mixed-Filter Algorithm for Dynamically Tracking Learning from Multiple Behavioral and Neurophysiological Measures

A mixed-filter algorithm for dynamically tracking learning from multiple behavioral and neurophysiological measures

Todd P. Coleman, Marianna Yanike, Wendy Suzuki, Emery N. Brown

I. Introduction

Learning is a dynamic process generally defined as a change in behavior as a result of experience [2], [11], [26], [10], [19], [21], [23]. Understanding how processes at the molecular and neuronal levels integrate so that an organism can learn is a central question in neuroscience. Most learning experiments consist of a sequence of trials. During each trial, a subject is given a fixed amount of time to execute a task and the resulting performance is recorded. During each trial, performance can be measured with a continuous variable (i.e. reaction time) as well as a binary one (whether or not the subject executes task correctly). The spiking behavior of certain neurons can also be used to characterize learning [26],[28],[27]. Learning is usually illustrated by using the behavioral variables to show that the subject has successfully performed the previously unfamiliar task with greater reliability than would be predicted by chance. When neural activity is recorded at the same time as the behavioral measures, an important question is the extent to what neural correlates can be associated with the changes in behavior.

We have developed a state-space model to analyze binary behavioral data [26], [21], [23],[24],[22]. The model has been successfully applied in a number of learning studies [26], [14], [25], [12],[22]. Recently, we have extended this model to analyze simultaneously recorded continuous and binary measures of behavior [17], [16]. An open problem is the analysis in a state-space framework of simultaneously recorded continuous and binary performance measures along with neural spiking activity modeled as a point process.

To develop a dynamic approach to analyzing data from learning experiments in which continuous and binary and responses are simultaneously recorded along with neural spiking activity, we extend our previously developed state-space model of learning to include a lognormal probability model for the continuous measurements, a Bernoulli probability model for the binary measurements and a point process model for the neural spiking activity. We estimate the model using an approximate EM algorithm [20],[21],[23], [16] to conduct the model fitting. We illustrate our approach in the analysis of a simulated learning experiment, and an actual learning experiment, in which a monkey rapidly learns new associations within a single session.

II. A State-Space Model of Learning

We assume that learning is a dynamic process that can be analyzed with the well-known state-space framework used in engineering, statistics and computer science . The state-space model is comprised of two equations: the state equation and the observation equation. The state equation defines the temporal evolution of an unobservable process. State models with unobservable processes are also referred to as latent process or hidden Markov models [7],[6],[9],[13],[15], [20]. Here, the state equation defines an unobservable cognitive state that characterizes the subject's understanding of the task. We track the evolution of this cognitive state across the trials in the experiment. We formulate our model so that as learning occurs the state increases and when learning does not occur it decreases. The observation equation relates the observed data to the cognitive state process. The data we observe in the learning experiment are the neural spiking activity and the continuous and binary responses. Our objective is to characterize learning by estimating the cognitive state process using simultaneously all three types of data.

To develop our model we extend the work in [17], [16] and consider a learning experiment consisting of K trials in which on each trial, a continuous reaction time, neural spiking activity, and a binary response measurement of performance are recorded. Let Z_k and M_k be respectively the values of the continuous and binary measurements on trial k for $k = 1, \ldots, K$. We assume that the cognitive state model is the first-order autoregressive process:

$$X_k = \gamma + \rho X_{k-1} + V_k \tag{1}$$

where $\rho \in (0, 1)$ represents a forgetting factor, γ is a learning rate, and the V_k's are independent, zero mean, Gaussian random variables with variance σ_V^2. Let $X = [X_1, \ldots, X_K]$ be the unobserved cognitive state process for the entire experiment.

For the purpose of exposition, we assume that the continuous measurements are reaction times and that the observation model for the reaction times is given by

$$Z_k = \alpha + hX_k + W_k \tag{2}$$

where Z_k is the logarithm of the reaction time on the kth trial, and the W_k's are independent zero mean Gaussian random variables with variance σ_W^2 . We assume that $h < 0$ to insure that on average, as the cognitive state X_k increases with learning, then the reaction time decreases. We let $Z = [Z_1, \ldots, Z_K]$ be the reaction times on all K trials.

We assume that the observation model for the binary responses, the M_k's obey a Bernoulli probability model

$$P\left(M_k = m | X_k = x_k\right) \;\; = \;\; p_k^m (1 - p_k)^{1-m} \tag{3}$$

where $m = 1$ if the response is correct and 0 if the response is incorrect. We take p_k to be the probability of a correct response on trial k, defined in terms of the unobserved cognitive state process x_k as

$$p_k \;\; = \;\; \frac{\exp\left(\mu + \eta x_k\right)}{1 + \exp\left(\mu + \eta x_k\right)}. \tag{4}$$

Formulation of p_k as a logistic function of the cognitive state process (4) ensures that the probability of a correct response on each trial is constrained to lie between 0 and 1, and that as the cognitive state increases, the probability of a correct responses approaches 1.

Assume that each of the K trials lasts T seconds. Divide each trial into $J = \frac{T}{\Delta}$ bins of width Δ so that there is at most one spike per bin. Let $N_{k,j} = 1$ if there is a spike on trial k in bin j and 0 otherwise for $j = 1, \ldots, T$ and $k = 1, \ldots, K$. Let $N_k = [N_{k,1}, \ldots, N_{k,J}]$ be the spikes recorded on trial k, and $N^k = [N_1, \ldots, N_k]$ be the spikes observed from trial 1 to k. We assume that the probability of a spike on trial k in bin j may be expressed as

$$P(N_{k,j} = n_{k,j} | X^k = x^k, N^{k-1} = n^{k-1}, N_{k,1} = n_{k,1}, \ldots, N_{k,j-1} = n_{k,j-1}) = (\lambda_{k,j}\Delta)^{n_{k,j}} e^{-\lambda_{k,j}\Delta} \tag{5}$$

and thus the joint probability mass function of N_k on trial k is

$$P(N_k = n_k | X^k = x^k) = \exp\left(\sum_{j=1}^{J} \log\left(\lambda_{k,j}\right) n_{k,j} - \lambda_{k,j}\Delta\right) \tag{6}$$

where (6) follows from the likelihood of a point process [4]. We define the conditional intensity function $\lambda_{k,j}$ as

$$\log \lambda_{k,j} \;\; = \;\; \psi + g x_k + \sum_{s=1}^{S} \beta_s n_{k,j-s}. \tag{7}$$

The state model (1) provides a stochastic continuity constraint [13] so that the current cognitive state, reaction time (2), probability of a correct response (4), and the conditional intensity function (7) all depend on the prior cognitive state. In this way, the state-space model provides a simple, plausible framework for relating performance on successive trials of the experiment.

We denote all of our observations at trial k as $Y_k = (M_k, N_k, Z_k)$. Because X is unobservable, and because $\theta = (\gamma, \rho, \sigma_V^2, \alpha, h, \sigma_W^2, \mu, \eta, g, \psi)$ is a set of unknown parameters, we use the Expectation-Maximization (EM) algorithm to estimate them by maximum likelihood [21],[23],[20], [9],[16]. The EM algorithm is a well-known procedure for performing maximum likelihood estimation when there is an unobservable process or missing observations. The EM algorithm has been used to estimate state-space models from point process and binary observations with linear Gaussian state processes [5]. The current EM algorithm combines features of the ones in [18],[21],[23]. The key technical point that allows implementation of this algorithm is the combined filter algorithm in (8)-(12). Its derivation is given in Appendix A.

III. Discrete-Time Recursive Estimation Algorithms

In this section, we develop a recursive, causal estimation algorithm to estimate the state at trial k, X_k, given the observations up to and including time k, $Y^k = y^k$. Define

$$\begin{aligned} x_{k|k'} &\triangleq E[X_k | Y^{k'} = y^{k'}] \\ \sigma^2_{k|k'} &\triangleq \text{var}\left[X_k | Y^{k'} = y^{k'}\right] \end{aligned}$$

as well as $p_{k|k}$ and $\lambda_{k,j|k,j}$ by (4) and (7), respectively, with with x_k replaced by $x_{k|k}$.

In order to derive closed form expressions, we develop a Gaussian approximation to the posterior, and as such, assume that that the posterior distribution on X at time k given $Y^k = y^k$ is the Gaussian density with mean $x_{k|k}$ and variance $\sigma^2_{k|k}$. Using the Chapman-Kolmogorov equations (25) with the Gaussian approximation to the posterior density, i.e. X_k given y^k, we obtain the following recursive filter algorithm:

One-Step Prediction

$$x_{k|k-1} = \gamma + \rho x_{k-1|k-1} \tag{8}$$

One-Step Prediction Variance

$$\sigma^2_{k|k-1} = \rho^2 \sigma^2_{k-1|k-1} + \sigma^2_V \tag{9}$$

Gain Coefficient

$$C_k = \frac{\sigma^2_{k|k-1}}{h^2 \sigma^2_{k|k-1} + \sigma^2_W} \tag{10}$$

Posterior Mode

$$x_{k|k} = x_{k|k-1} + C_k \left[h\left(z_k - \alpha - h x_{k|k-1}\right) + \eta \sigma^2_W (m_k - p_{k|k})\right] + \sum_{j=1}^{J} C_k \sigma^2_W \left[g(n_{k,j} - \lambda_{k,j|k,j}\Delta)\right] \tag{11}$$

Posterior Variance

$$\sigma^2_{k|k} = \left[\frac{1}{\sigma^2_{k|k-1}} + \frac{h^2}{\sigma^2_W} + \eta^2 p_{k|k}(1 - p_{k|k}) + \sum_{j=1}^{J} g^2 \lambda_{k,j|k,j}\Delta\right]^{-1} \tag{12}$$

Details can be found in Appendix A. Because there are three observation processes, (11) has a continuous-valued innovation term, $(z_k - \alpha - h x_{k|k-1})$, a binary innovation term, $(m_k - p_{k|k})$, and a point-process innovation term, $(n_{k,j} - \lambda_{k,j|k,j}\Delta)$. As is true in the Kalman filter, the continuous-valued innovation compares the observation z_k with its one-step prediction. The binary innovation compares the binary observation m_k with $p_{k|k}$, the probability of a correct response at trial k. Finally, the point process innovation compares the $n_{k,j}$, whether or not a spike occurred in bin j on trial k, with the expected number of occurrences, $\lambda_{k,j|k,j}\Delta$. As in the Kalman filter, C_k in (10), is a time-dependent gain coefficient. At trial k, the amount by which the continuous-valued innovation term affects the update is determined by $C_k h$, the amount by which the binary innovation affects the update is determined by $C_k \eta \sigma_W^2$, and the amount by which the point process innovation for neuron j affects the update is determined by the sum of $C_k \sigma_W^2 g$. Unlike in the Kalman filter algorithm, the left and right hand sides of the posterior mode (11) and the posterior variance (12) depend on the state estimate $x_{k|k}$. That is, because $p_{k|k}$ and $\lambda_{k,j|k,j}$ depend on $x_{k|k}$ through (4) and (7). Therefore, at each step k of the algorithm, we use Newton's methods (developed in Appendix A) to compute $x_{k|k}$ in (11).

IV. An Expectation-Maximization Algorithm for Efficient Maximum Likelihood Estimation

We next define an EM algorithm [5] to compute jointly the state and model parameter estimates. To do so, we combine the recursive filter given in the previous section with the fixed interval smoothing algorithm and the covariance smoothing algorithms to efficiently evaluate the E-step.

A. E-Step

The E-step of the EM algorithm only requires the calculation of the posterior $f_{X_k|Y}(x_k|y)$. As mentioned in Section III, we use a Gaussian approximation to the posterior. Although in general this is a multi-dimensional Gaussian, we need only compute the mean and certain components of the covariance of this distribution.

1) E-step I: Nonlinear Recursive Filter: The nonlinear recursive filter is given in (8) through (12).

2) E-step II: Fixed Interval Smoothing (FIS) Algorithm: Given the sequence of posterior mode estimates $x_{k|k}$ and the variance $\sigma^2_{k|k}$, we use the fixed interval smoothing algorithm [20], [3] to compute $x_{k|K}$ and $\sigma^2_{k|K}$

$$\begin{aligned} A_k &\triangleq \rho \frac{\sigma^2_{k|k}}{\sigma^2_{k+1|k}} && (13)\\ x_{k|K} &= x_{k|k} + A_k \left(x_{k+1|K} - x_{k+1|k} \right) && (14)\\ \sigma^2_{k|K} &= \sigma^2_{k|k} + A_k^2 \left(\sigma^2_{k+1|K} - \sigma^2_{k+1|k} \right) && (15) \end{aligned}$$

for $k = K-1, \ldots, 1$ with initial conditions $x_{K|K}$ and $\sigma^2_{K|K}$ computed from the last step in (8) through (12).

3) E-step III: State-Space Covariance Algorithm: The conditional covariance, $\sigma_{k,k'|K}$, can be computed from the state-space covariance algorithm and is given for $1 \leq k \leq k' \leq K$ by

$$\sigma_{k,k'|K} = A_k \sigma_{k+1,k'|K} \tag{16}$$

Thus the covariance terms required for the E-step are

$$\begin{aligned} \tilde{W}_{k,k+1} &= \sigma_{k,k+1|K} + x_{k|K} x_{k+1|K} && (17)\\ \tilde{W}_k &= \sigma^2_{k|K} + x^2_{k|K} && (18) \end{aligned}$$

B. M-Step

The M-step requires maximization of the expected log likelihood given the observed data. Appendix B gives the computations that lead to the following approximate update equations:

$$\begin{bmatrix} \gamma \\ \rho \end{bmatrix} = \begin{bmatrix} K & \sum_{k=1}^{K} x_{k-1|K} \\ \sum_{k=1}^{K} x_{k-1|K} & \sum_{k=1}^{K} \tilde{W}_{k-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^{K} x_{k|K} \\ \sum_{k=1}^{K} \tilde{W}_{k-1,k} \end{bmatrix} \tag{19}$$

$$\begin{bmatrix} \alpha \\ h \end{bmatrix} = \begin{bmatrix} K & \sum_{k=1}^{K} x_{k|K} \\ \sum_{k=1}^{K} x_{k|K} & \sum_{k=1}^{K} \tilde{W}_{k} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^{K} z_k \\ \sum_{k=1}^{K} z_k x_{k|K} \end{bmatrix} \tag{20}$$

$$\sigma_W^2 = \frac{1}{K}\left[\sum_{k=1}^{K} (z_k - \alpha)^2 - 2(z_k - \alpha) h x_{k|K} + h^2 \tilde{W}_k\right] \tag{21}$$

$$\psi = \log\left(\frac{\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}}{\sum_{k=1}^{K}\sum_{j=1}^{J} \Delta \exp\left(g x_{k|K} + \frac{1}{2}\sigma_{k|K}^2 g^2 + \sum_{s=1}^{S} \beta_s n_{k,j-s}\right)}\right) \tag{22}$$

To solve for $\mu, \eta, \psi, g, \{\beta_s\}$, we use Newton's method techniques, described in Appendix C.

V. Algorithm Performance and Simulation

A. Application of the Methods to Simulated Data

To illustrate our analysis paradigm, we apply it first to simulated data. We simulated neural spiking activity, reaction times and binary responses for a 25-trial learning experiment during which each trial lasted 5 seconds. We discretized time into 5000 one-millisecond bins. To simulate the state process we used the parameter values $\gamma = 0.1, \rho = 0.99$, and $\sigma_V^2 = 0.03$. For the continuous-valued reaction time process, we used the parameters $\alpha = 3.69, h = -0.38$, and $\sigma_W^2 = 0.75$. For binary-valued data, we used the parameter values $\mu = -1.4170$ and $\eta = 1.75$. For the point process parameters we chose $\psi = -3.5, g = 2.0$, and $\beta = (-20, -5, 1, 3)$. The simulated data are shown in Figure 1.

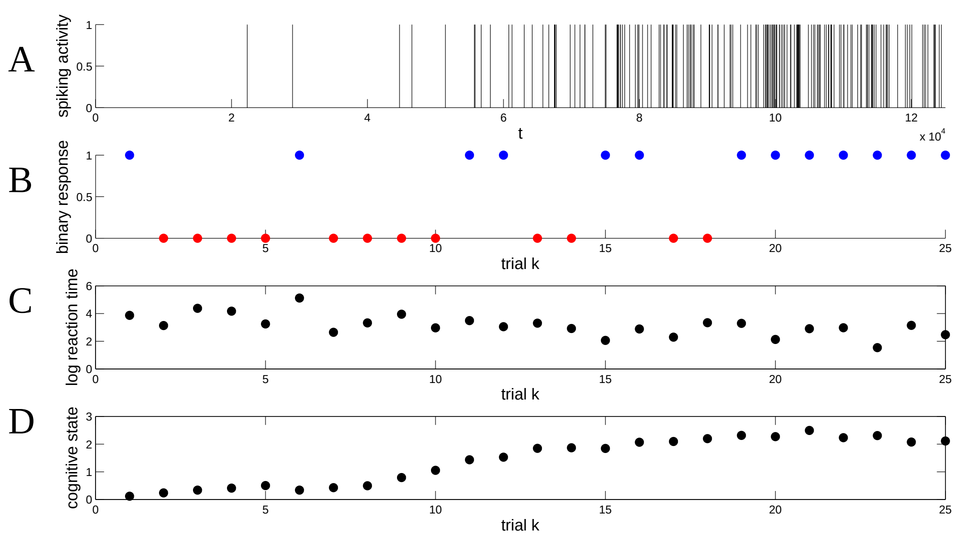

Fig. 1. Visualization of the simulated data. Panel A shows the simulated spiking activity. Panel B shows the binary responses, with blue (red) corresponding to correct (incorrect) responses. Panel C shows the log reaction times. Panel D shows the cognitive state.

The state estimates are in close agreement with the true state for all trials (Figure 2A). The Kolmogorov-Smirnov plot [4] confirms that the model describes well the point process component of the model

(Figure 2B). These results demonstrate that the mixed analysis is capable of recovering the unobserved states and the components the three observation models from simulated data.

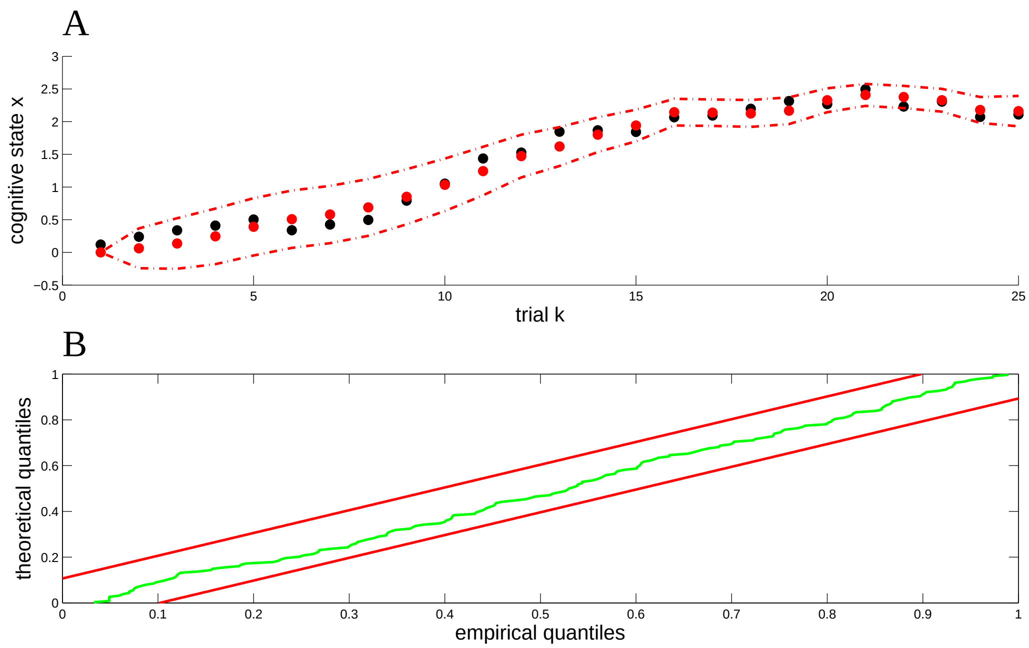

Fig. 2. A: performance of the recursive estimation procedure. The true cognitive state is given in black, while estimates are given in red. 95% Confidence intervals are given with the red dashed lines. (B): Kolmogorov-Smirnov plot confirms that the model describes well the point process component of the model.

B. Application of the Methods to Experimental Data

In this section we apply the analysis paradigm to an actual learning experiment in which neural spiking activity was recorded along with binary and continuous performance measures as a rhesus monkey executed a location-scene association task described in detail in [26]. The experiment consists of 45 trials with each trial lasting 3,300 msec. In this task, each trial started with a baseline period (0 to 400 msec) during which the monkey fixated on a cue presented on a computer screen. The animal was then presented with three identical targets (north, east, and west) superimposed on a novel visual scene (401 to 900 msec). The scene disappeared and the targets remained on the screen during a delay period (901 to 1600 msec). At the end of the delay period, the fixation point disappeared cueing the animal to make an eye-movement to one of the three targets (1,600 to 3,300 msec). For each scene, only one target was rewarded and 3 novel scenes were typically learned simultaneously. Trials of novel scenes were interspersed with trials in which three well-learned scenes were presented. The probability of a correct response occurring by chance was 0.33 because there were three locations the monkey could choose as a response. To characterize learning we reported for all trials the reaction times (time from the go-cue to the response), the correct and incorrect responses, and neural spiking activity recorded in the perirhinal cortex.

The correct and incorrect responses and neural spiking activity are shown in Figure 3A for one scene. The spiking activity on a trial is red if the behavioral response was incorrect on that trial and blue if the response was correct. The response times are shown in Figure 3B. The animal clearly showed a change in responses from all incorrect to correct around trial 23 or 24. The response time decreased from trial 1 to 45. The spiking rate of the neural firing increased with learning. To analyze the spiking activity we used one milliseconds time bins and chose the order of the autoregressive for the spiking activity equal to 10 milliseconds.

The cognitive state estimates in Figure 4A are consistent with the animal learning the task. The KS plot in Figure 4B suggests that the point process component of the model describes the neural spiking activity well. The learning curve plot of the probability of correct response, overlayed with the binary responses, is given by techniques in [16] and shown in Figure 5A. This information, as well as the decrease in the reaction time of Figure 5B, is consistent with learning. The estimated value of the parameter $\hat{g} = 0.0232$ is consistent with increasing spiking activity as the animal learned whereas the estimated coefficients

$$\hat{\beta} = (-3.0278, -2.3581, -0.4836, -0.9458, -0.1914, -0.3884, -0.7690, 0.1783, -0.4119, 0.1066)$$

are consistent with a refractory period and a relative refractory period for the neuron. The results establish the feasibility of conducting simultaneous analysis of continuous and binary behavioral data along with neural spiking activity using the mixed model.

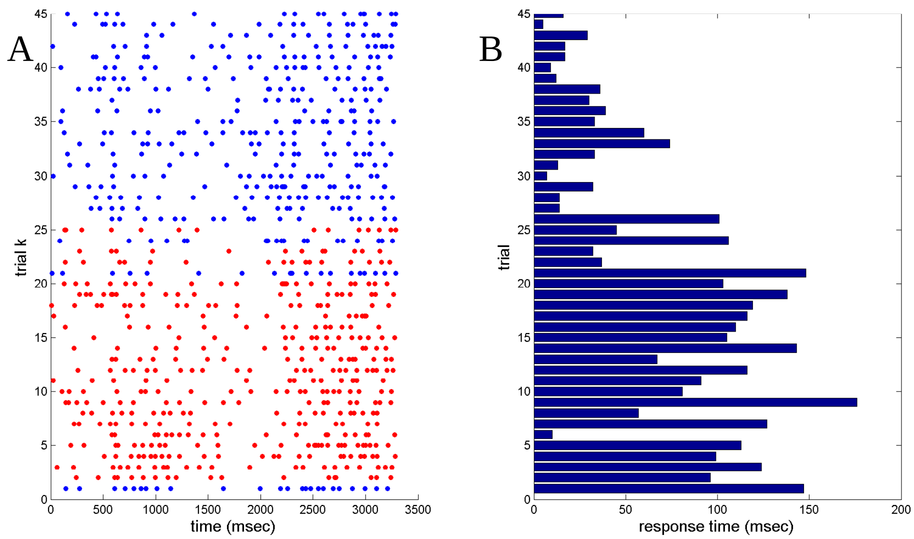

Fig. 3. A: correct/incorrect responses in blue/red rows; a spike in bin j of trial k is present if a dot appears in the associated (k, j) row and column. A change in responses from correct to incorrect is clear around trial 23 or 24.B: The response times in milliseconds on each trial. The response times on average decreased from trial 1 to 45.

VI. Discussion and Conclusion

Continuous observations, such as reaction times and run times, neural spiking activity and binary observations, such as correct/incorrect responses are frequently recorded simultaneously in behavioral learning experiments. However, the two types of performance measures and neurophysiological recordings are not analyzed simultaneously to study learning. We have introduced a state-space model in which the observation model makes use of simultaneously recorded continuous and binary measures of performance, as well as neural spiking activity to characterize learning. Using maximum likelihood implemented in the form of an EM algorithm we estimated the model from these simultaneously recorded performance measures and neural spiking activity. We illustrated the new model and algorithm in the analysis of simulated data and data from an actual learning experiment.

The computational innovation that enabled our combined model analysis is the recursive filter algorithm for mixed observation processes, i.e. continuous, point process and binary observations, the fixed-interval smoothing algorithm, and an approximate EM algorithm for combined cognitive state and model parameter estimation. Our mixed recursive filter algorithm [21] combines the well-known Kalman filter with a recently developed binary filter [17] and the point process filter [3], [1], [8]. In this way, the mixed filter makes possible simultaneous dynamic analysis of behavioral performance data and neural spiking activity.

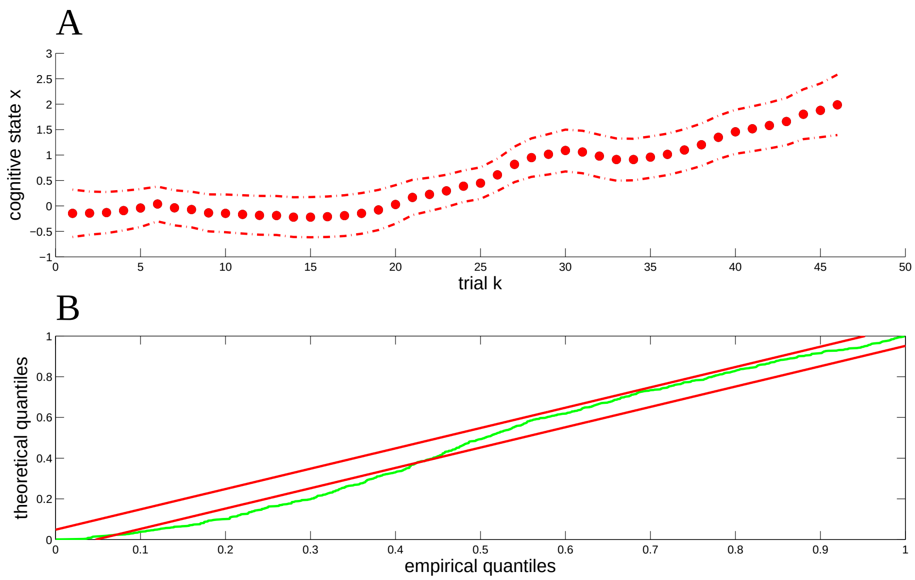

Fig. 4. Mixed modality recursive filtering results. A: the estimate and confidence interval of the cognitive state process. B: a Kolmogorov-Smirnov plot of the time-rescaled inter-spike intervals from the learned parameters.

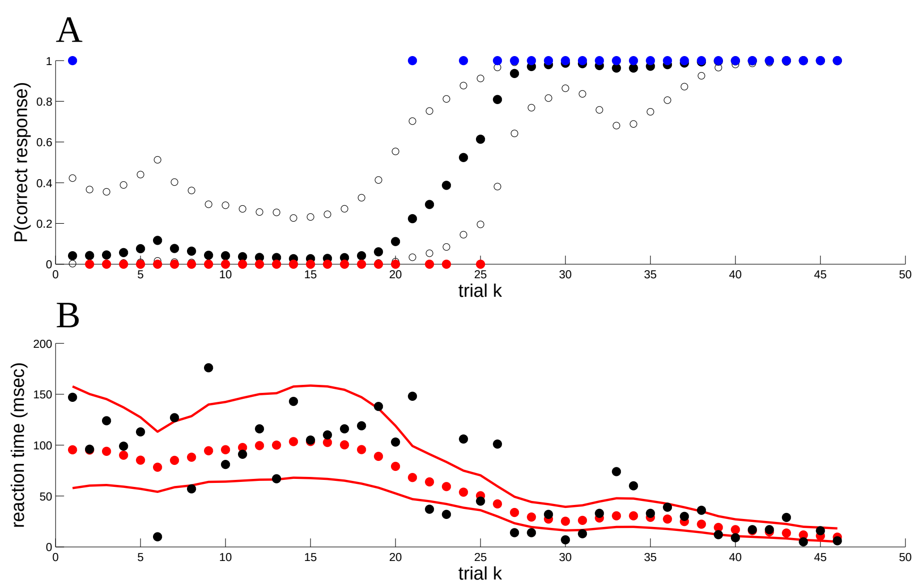

Fig. 5. A: plot of the estimated probability of correct response (black filled circles), along with 95% confidence intervals (black hollow circles), as well as the correct (blue) and incorrect (red) behavioral responses. B: plot of the estimated reaction times, along with 95% confidence intervals (red), as well as the true reaction times (black).

Several extensions of the current work are possible to more complex models of performance and neural spiking data. These model extensions could be fit by constructing the appropriate extensions of our EM algorithm. An alternative approach would be to formulate the model parameter estimation as a Bayesian question and take advantage of readily available Bayesian analysis software packages such as BUGS to conduct the model fitting [24].

The question we have studied here of simultaneously analyzing performance data and neural spiking activity offers a solution to the now ubiquitous problem of combining information dynamically from different measurement types. Possible extensions of this paradigm in neuroscience include combining information from local field potentials and ensemble neural spiking activity to devise algorithms for neural prosthetic control. Another extension of this approach is to functional neural imaging studies in which combinations of functional magnetic resonance imaging, electroencephalographic and magnetoencephalographic recordings are made simultaneously or in sequence. Again, the state-space modeling framework provides an optimal strategy for combining the information from the various sources. We will investigate these theoretical and applied problems in future investigations.

ACKNOWLEDGEMENTS

Support was provided by National Institutes of Health Grants DA-015644 to E. N. Brown and W. Suzuki; DPI0D003646, MH-59733, and MH-071847 to E. N. Brown; and by the AFOSR Complex Networks Program via Award No. FA9550-08-1-0079 to T. P. Coleman.

REFERENCES

[1] R. Barbieri, L. M. Frank, D. P. Nguyen, M. C. Quirk, V. Solo, M. A. Wilson, and E. N. Brown. Dynamic analyses of information encoding by neural ensembles. *Neural Computation*, 16(2):277–308, 2004.
[2] T.D. Barnes, Y. Kubota, D. Hu, D.Z. Jin, and A.M. Graybiel. Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, 437(7062):1158–1161, 2005.
[3] E. N. Brown, L. M. Frank, L. M. Tang, M.C. Quirk, and M. Wilson. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18:7411–7425, 1998.
[4] E. N. Brown, R. Barbieri R. E. Kass, and L. M. Frank. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 14(2):325–346, 2002.
[5] A.P. Dempster, N.M. Laird, D.B. Rubin, et al. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
[6] A. Doucet, N. De Freitas, and N. Gordon. *Sequential Monte Carlo methods in practice*. Springer Verlag, 2001.
[7] J. Durbin and S.J. Koopman. *Time series analysis by state space methods*. Oxford Univ Pr, 2001.
[8] U. T. Eden, L. M. Frank, R. Barbieri, V. Solo, and E. N. Brown. Dynamic analyses of neural encoding by point process adaptive filtering. *Neural Computation*, 16(5):971–998, 2004.
[9] L. Fahrmeir, G. Tutz, and W. Hennevogl. *Multivariate statistical modelling based on generalized linear models*. Springer Verlag, 2001.
[10] C.R. Gallistel. Learning and representation. *Learning and Memory: A Comprehensive Reference. Elsevier*, 2008.
[11] M.S. Jog, Y. Kubota, C.I. Connolly, V. Hillegaart, and A.M. Graybiel. Building neural representations of habits. *Science*, 286(5445):1745–9, 1999.
[12] M.P. Karlsson and L.M. Frank. Network dynamics underlying the formation of sparse, informative representations in the hippocampus. *Journal of Neuroscience*, 28(52):14271, 2008.
[13] G. Kitagawa and W. Gersch. *Smoothness priors analysis of time series*. Springer Verlag, 1996.
[14] J.R. Law, M.A. Flanery, S. Wirth, M. Yanike, A.C. Smith, L.M. Frank, W.A. Suzuki, E.N. Brown, and C.E.L. Stark. Functional magnetic resonance imaging activity during the gradual acquisition and expression of paired-associate memory. *Journal of Neuroscience*, 25(24):5720–5729, 2005.
[15] J.M. Mendel. Lessons in estimation theory for signal processing, communications, and control(Book). *Englewood Cliffs, NJ: Prentice Hall PTR*, 1995.
[16] M.J. Prerau, A.C. Smith, U.T. Eden, Y. Kubota, M. Yanike, W. Suzuki, A.M. Graybiel, and E.N. Brown. Characterizing learning by simultaneous analysis of continuous and binary measures of performance. *Journal of Neurophysiology*, pages 3060–72, 2009.
[17] M.J. Prerau, A.C. Smith, U.T. Eden, M. Yanike, W.A. Suzuki, and E.N. Brown. A mixed filter algorithm for cognitive state estimation from simultaneously recorded continuous and binary measures of performance. *Biological Cybernetics*, 99(1):1–14, 2008.
[18] R.H. Shumway and D.S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of time series analysis*, 3(4):253–264, 1982.
[19] S. Siegel and N. J. Castellan. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill Book Company, 1988.
[20] A. C. Smith and E. N. Brown. Estimating a state-space model from point process observations. *Neural Computation*, 15:965–991, 2003.
[21] A. C. Smith, L. M. Frank, S. Wirth, M. Yanike, D. Hu, Y. Kubota, A. M. Graybiel, W. Suzuki, and E. N. Brown. Dynamic analysis of learning in behavioral experiments. *Journal of Neuroscience*, 24(2):447–461, 2004.
[22] A. C. Smith, J. D. Scalon, S. Wirth, M. Yanike, W. A. Suzuki, and E. N. Brown. State-space algorithms for estimating spike rate functions. *Computational Intelligence and Neuroscience*, 2010, 2009.
[23] A. C. Smith, M. R. Stefani, B. Moghaddam, and E. N. Brown. Analysis and design of behavioral experiments to characterize population learning. *Journal of Neurophysiology*, 93:1776–1792, 2005.
[24] A.C. Smith, S. Wirth, W.A. Suzuki, and E.N. Brown. Bayesian analysis of interleaved learning and response bias in behavioral experiments. *Journal of Neurophysiology*, 97(3):2516–2524, 2007.
[25] Z.M. Williams and E.N. Eskandar. Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nature Neuroscience*, 9:562–568.
[26] S. Wirth, M. Yanike, L.M. Frank, A.C. Smith, E.N. Brown, and W.A. Suzuki. Single neurons in the monkey hippocampus and learning of new associations. *Science*, 300(5625):1578–1581, 2003.

[27] Chiu CC Varun S Smith AC Brown EN Suzuki Wirth S, Avsar E and WA. Trial outcome and associative learning signals in the monkey hippocampus. *Neuron*, 61(6):930–940, 2009.
[28] M. Yanike, S. Wirth, A. C. Smith, E. N. Brown, and W. A. Suzuki. Comparison of associative learning-related signals in the macaque perirhinal cortex and hippocampus. *Cerebral Cortex*, 19(5):1064–1078, 2009.

Appendix A
Details of the Recursive Filter

In this section, we provide details of the derivation of equations (8)-(12). Our objective is to construct a recursive filter to estimate the state X_k at trial k from $Y^k = \left\{Z^k, M^k, N^k\right\}$. The standard approach to deriving such a filter is to express recursively the probability density of the state given the observations. For events $\{A, B, C\}$, we have from Bayes' rule that

$$P(A|B,C) \;=\; \frac{P(A,B|C)}{P(B|C)} = \frac{P(A|C)P(B|A,C)}{P(B|C)} \tag{23}$$

Denote A according to $\{X_k = x_k\}$, B according to $\{Y_k = y_k\}$, and C according to $\left\{Y^{k-1} = y^{k-1}\right\}$. Then we have

$$\begin{aligned} f_{X_k|Y^k}\left(x_k|y^k\right) &= \frac{f_{X_k|Y^{k-1}}\left(x_k|y^{k-1}\right) f_{Y_k|X_k}\left(y_k|x_k\right)}{f_{Y_k|Y^{k-1}}\left(y_k|y^{k-1}\right)} \\ &= \frac{f_{X_k|Y^{k-1}}\left(x_k|y^{k-1}\right) P_{M_k|X_k}\left(m_k|x_k\right) P_{N_k|X_k}\left(n_k|x_k\right) f_{Z_k|X_k}\left(z_k|x_k\right)}{f_{Y_k|Y^{k-1}}\left(y_k|y^{k-1}\right)} \\ &\propto f_{X_k|Y^{k-1}}\left(x_k|y^{k-1}\right) P_{M_k|X_k}\left(m_k|x_k\right) P_{N_k|X_k}\left(n_k|x_k\right) f_{Z_k|X_k}\left(z_k|x_k\right) \end{aligned} \tag{24}$$

and the associated one-step prediction probability density or Chapman-Kolmogorov equation is

$$f_{X_k|Y^{k-1}}\left(x_k|y^{k-1}\right) \;=\; \int f_{X_{k-1}|Y^{k-1}}\left(x_{k-1}|y^{k-1}\right) f_{X_k|X_{k-1}}\left(x_k|x_{k-1}\right) dx_{k-1} \tag{25}$$

Together (24) and (25) define a recursion that can be used to compute the probability of the state given the observations.

We derive the mixed filter algorithm by computing a Gaussian approximation to the posterior density $f_{X_k|Y^k}\left(x_k|y^k\right)$ in (24). At time k, we assume the one-step prediction density (25) is the Gaussian density

$$f_{X_k|Y^{k-1}}\left(x_k|y^{k-1}\right) \sim \mathcal{N}\left(x_{k|k-1}, \sigma^2_{k|k-1}\right). \tag{26}$$

To evaluate $x_{k|k-1}$ and $\sigma^2_{k|k-1}$, we note that they follow in a straightforward manner:

$$x_{k|k-1} \;=\; \mathbb{E}\left[X_k|Y^{k-1} = y^{k-1}\right] = \gamma + \rho x_{k-1|k-1} \tag{27}$$

$$\sigma^2_{k|k-1} \;=\; \text{var}\left(X_k|Y^{k-1} = y^{k-1}\right) \tag{28}$$

$$\;=\; \text{var}\left(\gamma + \rho X_{k-1} + V_k|Y^{k-1} = y^{k-1}\right) \tag{29}$$

$$\;=\; \rho^2 \sigma^2_{k-1|k-1} + \sigma^2_V \tag{30}$$

Substituting all these equations together, then we have that the posterior density can be expressed as

$$f_{X_k|Y^k}\left(x_k|y^k\right) \;\propto \exp\left\{ -\frac{(x_k - x_{k|k-1})^2}{2\sigma^2_{k|k-1}} + m_k \log[p_k(1-p_k)^{-1}] + \log(1-p_k) -\frac{(z_k-\alpha-hx_k)^2}{2\sigma^2_W} + \sum_{j=1}^J n_{k,j}\log\left(\lambda_{k,j}\right) - \lambda_{k,j}\Delta \right\}$$

$$\Rightarrow \log f_{X_k|Y^k}\left(x_k|y^k\right) \;= C(y^k) - \frac{(x_k - x_{k|k-1})^2}{2\sigma^2_{k|k-1}} + m_k \log[p_k(1-p_k)^{-1}] + \log(1-p_k) \tag{31}$$

$$-\frac{(z_k-\alpha-hx_k)^2}{2\sigma^2_W} + \sum_{j=1}^J n_{k,j}\log\left(\lambda_{k,j}\right) - \lambda_{k,j}\Delta \tag{32}$$

Now we can compute the *maximum-a-posteriori* estimate of x_k and its associated variance estimate. To do this, we compute the first and and second derivatives of the log posterior density with respect to x_k, which are respectively

$$\begin{aligned} 0 = \frac{\partial \log f_{X_k|Y^k}\left(x_k|y^k\right)}{\partial x_k} &= -\frac{(x_k - x_{k|k-1})}{\sigma^2_{k|k-1}} + \frac{h(z_k - \alpha - hx_k)}{\sigma^2_W} + \eta(m_k - p_k) \\ &\quad + \sum_{j=1}^{J} g(n_{k,j} - \lambda_{k,j}\Delta) \\ \frac{\partial^2 \log f_{X_k|Y^k}\left(x_k|y^k\right)}{\partial x_k^2} &= -\frac{1}{\sigma^2_{k|k-1}} - \frac{h^2}{\sigma^2_W} - \eta^2 p_k(1-p_k) - \sum_{j=1}^{J} g^2\lambda_{k,j}\Delta \end{aligned}$$

where we have exploited the fact that from (4) and (7), the following properties hold:

$$\frac{\partial p_k}{\partial x_k} = \eta p_k(1-p_k) \Rightarrow \begin{cases} \frac{\partial \log(1-p_k)}{\partial x_k} = -\eta p_k, \\ \frac{\partial \log p_k}{\partial x_k} = \eta(1-p_k) \end{cases} \tag{33}$$

$$\frac{\partial \lambda_{k,j}}{\partial x_k} = g\lambda_{k,j} \tag{34}$$

$$\frac{\partial \log \lambda_{k,j}}{\partial x_k} = g \tag{35}$$

Combining all this together, using the Gaussian approximation, we arrive at (8)-(12).

Appendix B
Details of the M step update equations

In this section, we derive details of the update equations provided in Section IV-B. Note that the joint distribution on all (observed and latent) variables is given by

$$\begin{aligned} \log f_{X^K|Y^K}\left(x^K|y^K;\theta\right) &= C(y^K) + \sum_{k=1}^{K} -\frac{1}{2\sigma^2_V}(x_k - \gamma - \rho x_{k-1})^2 - \frac{1}{2\sigma^2_W}(z_k - \alpha - hx_k)^2 \\ &\quad + \sum_{k=1}^{K} m_k\left(\mu + \eta x_k\right) - \log\left(1 + \exp(\mu + \eta x_k)\right) \\ &\quad + \sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}\left[\psi + gx_k + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right] - \Delta \exp\left(\psi + gx_k + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right) \end{aligned}$$

Note that the expected log-likelihood $Q(\theta) \triangleq \mathbb{E}\left\{\log f_{X^K|Y^K}\left(x^K|Y^K;\theta\right)|Y^K = y^K\right\}$ has linear terms in $E[X_k|Y^k = y^k]$ along with quadratic terms involving $\tilde{W}_{k,j} \triangleq \mathbb{E}\left\{X_k X_j|Y^K = y^K\right\}$, except for a couple of terms, including $E[e^{gx_k}|Y^k = y^k]$. We note that if $\tilde{X} \sim \mathcal{N}\left(\mu, \sigma^2\right)$ then its moment generating function $M(t) \triangleq E[e^{\tilde{x}t}]$ is given by

$$M(t) = e^{ut + \frac{1}{2}\sigma^2 t^2}. \tag{36}$$

With this, we have

$$
\begin{aligned}
Q(\theta) \simeq C(y^K) &- \sum_{k=1}^{K} \frac{1}{2\sigma_V^2} \mathbb{E}\left[(X_k - \gamma - \rho X_{k-1})^2 | Y^K = y^K\right] - \frac{1}{2\sigma_W^2} \mathbb{E}\left[(Z_k - \alpha - hX_k)^2 | Y^k = y^k\right] \\
&+ \sum_{k=1}^{K} m_k \left(\mu + \eta x_{k|K}\right) - \mathbb{E}\left[\log\left(1 + \exp(\mu + \eta X_k)\right) | Y^k = y^k\right] \qquad (37) \\
&+ \sum_{k=1}^{K} n_{k,j} \left(\psi + g x_{k|K} + \sum_{s=1}^{S} \beta_s n_{k,j-s}\right) - \mathbb{E}\left[\exp\left(\psi + gX_k + \sum_{s=1}^{S} \beta_s n_{k,j-s}\right) \Delta | Y^K = y^K\right] \qquad (38) \\
&= C(y^K) - \frac{1}{2\sigma_V^2}\left[\sum_{k=1}^{K} \tilde{W}_k - 2\gamma x_{k|K} - 2\rho \tilde{W}_{k-1,k} + \gamma^2 + 2\gamma\rho x_{k-1|K} + \rho^2 \tilde{W}_{k-1}\right] \\
&- \frac{1}{2\sigma_W^2}\left[\sum_{k=1}^{K} (z_k - \alpha)^2 - 2(z_k - \alpha) h x_{k|K} + h^2 \tilde{W}_k\right] \\
&+ \sum_{k=1}^{K} m_k \left(\mu + \eta x_{k|K}\right) - \mathbb{E}\left[\log\left(1 + \exp(\mu + \eta x_k)\right) | Y^K = y^K\right] \qquad (39) \\
&+ \sum_{k=1}^{K} n_{k,j} \left(\psi + g x_{k|K} + \sum_{s=1}^{S} \beta_s n_{k,j-s}\right) - \Delta \exp\left(\psi + g x_{k|K} + \frac{1}{2}\sigma_{k|K}^2 g^2 + \sum_{s=1}^{S} \beta_s n_{k,j-s}\right) \qquad (40)
\end{aligned}
$$

where in going from (38) to (40), we have used the (36).

We now rely upon the Taylor series approximation around $x_{k|K}$

$$
E[\rho(X_k)|Y^K = y^K] \simeq \rho(x_{k|K}) + \frac{1}{2}\sigma_{k|K}^2 \rho''(x_{k|K})
$$

Let us now consider the conditional expectation term involves $\log\left(1 + \exp(\mu + \eta x_k)\right)$

$$
\begin{aligned}
\rho_1(x_k) &\triangleq \frac{\partial \log\left(1 + \exp(\mu + \eta x_k)\right)}{\partial \mu} = \frac{\exp(\mu + \eta x_k)}{1 + \exp(\mu + \eta x_k)} = p_k, \qquad (41) \\
\rho_2(x_k) &\triangleq \frac{\partial \log\left(1 + \exp(\mu + \eta x_k)\right)}{\partial \eta} = \frac{x_k \exp(\mu + \eta x_k)}{1 + \exp(\mu + \eta x_k)} = x_k p_k \qquad (42)
\end{aligned}
$$

Note from before that

$$
\begin{aligned}
\rho_1'(x_k) &= \eta p_k(1 - p_k) = \eta(p_k - p_k^2) \\
\Rightarrow \rho_1''(x_k) &= \eta[\eta p_k(1 - p_k) - 2p_k \eta p_k(1 - p_k)] \\
&= \eta^2 p_k(1 - p_k)(1 - 2p_k)
\end{aligned}
$$

Thus we have that

$$
\begin{aligned}
f_1(\mu, \eta) &= \frac{\partial}{\partial \mu}\left\{\sum_{k=1}^{K} m_k \left(\mu + \eta x_{k|K}\right) - \mathbb{E}\left[\log\left(1 + \exp(\mu + \eta X_k)\right) | Y^K = y^K\right]\right\} \qquad (43) \\
&= \sum_{k=1}^{K} m_k - \mathbb{E}[\rho_1(X_k)|Y^K = y^K] \qquad (44) \\
&\simeq \sum_{k=1}^{K} m_k - p_{k|K} - \frac{1}{2}\sigma_{k|K}^2 \eta^2 p_{k|K}(1 - p_{k|K})(1 - 2p_{k|K}) \qquad (45)
\end{aligned}
$$

Let us now consider $\rho_2(x_k) = x_k p_k$. Note from (42) that

$$\begin{aligned}
\rho_2'(x_k) &= [x_k \rho_1(x_k)]' \\
&= x_k \rho_1'(x_k) + \rho_1(x_k) \\
\Rightarrow \rho_2''(x_k) &= \rho_1'(x_k) + \rho_1'(x_k) + x_k \rho_1''(x_k) \\
&= 2\rho_1'(x_k) + x_k \rho_1''(x_k) \\
&= 2\eta p_k(1-p_k) + x_k \eta^2 p_k(1-p_k)(1-2p_k) \\
&= \eta p_k(1-p_k)\left[2 + x_k \eta (1-2p_k)\right]
\end{aligned}$$

Thus we have that

$$f_2(\mu,\eta) = \frac{\partial}{\partial \eta}\left\{\sum_{k=1}^{K} m_k\left(\mu + \eta x_{k|K}\right) - \mathbb{E}\left[\log\left(1+\exp(\mu+\eta x_k)\right)|Y^K = y^K\right]\right\} \tag{46}$$

$$= \sum_{k=1}^{K} m_k x_{k|K} - \mathbb{E}[\rho_2(X_k)|Y^K = y^K] \tag{47}$$

$$\simeq \sum_{k=1}^{K} m_k x_{k|K} - x_{k|K}p_{k|K} - \frac{1}{2}\sigma^2_{k|K}\eta p_{k|K}(1-p_{k|K})\left[2 + x_{k|K}\eta(1-2p_{k|K})\right] \tag{48}$$

Thus we differentiate to find a local minimum

$$0 = \frac{\partial Q}{\partial \gamma} = -\frac{1}{\sigma_V^2}\left[\sum_{k=1}^{K} -x_{k|K} + \gamma + \rho x_{k-1|K}\right]$$

$$0 = \frac{\partial Q}{\partial \rho} = -\frac{1}{\sigma_V^2}\left[\sum_{k=1}^{K} -\tilde{W}_{k-1,k} + \gamma x_{k-1|K} + \rho \tilde{W}_{k-1}\right]$$

$$0 = \frac{\partial Q}{\partial \alpha} = -\frac{1}{\sigma_W^2}\left[\sum_{k=1}^{K} \alpha - z_k + h x_{k|K}\right]$$

$$0 = \frac{\partial Q}{\partial h} = -\frac{1}{\sigma_W^2}\left[\sum_{k=1}^{K} -(z_k-\alpha)x_{k|K} + h\tilde{W}_k\right]$$

$$0 = \frac{\partial Q}{\partial \sigma_W^2} = \frac{1}{2[\sigma_W^2]^2}\left[-K\sigma_W^2 + \sum_{k=1}^{K}(z_k-\alpha)^2 - 2(z_k-\alpha)h x_{k|K} + h^2\tilde{W}_k\right]$$

$$0 = \frac{\partial Q}{\partial \mu} = \sum_{k=1}^{K} m_k - p_{k|K} - \frac{1}{2}\sigma^2_{k|K}\eta^2 p_{k|K}(1-p_{k|K})(1-2p_{k|K}) \tag{49}$$

$$0 = \frac{\partial Q}{\partial \eta} = \sum_{k=1}^{K} m_k x_{k|K} - x_{k|K}p_{k|K} - \frac{1}{2}\sigma^2_{k|K}\eta p_{k|K}(1-p_{k|K})\left[2 + x_{k|K}\eta(1-2p_{k|K})\right] \tag{50}$$

$$0 = \frac{\partial Q}{\partial \psi} = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j} - \Delta\exp\left(\psi + g x_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right) \tag{51}$$

$$0 = \frac{\partial Q}{\partial g} = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}x_{k|K} - \Delta(x_{k|K} + g\sigma^2_{k|K})\exp\left(\psi + g x_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right) \tag{52}$$

$$0 = \frac{\partial Q}{\partial \beta_s} = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}n_{k,j-s} - \Delta n_{k,j-s}\exp\left(\psi + g x_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right) \tag{53}$$

Simplifying, we get

$$\begin{bmatrix} \gamma \\ \rho \end{bmatrix} = \begin{bmatrix} K & \sum_{k=1}^{K} x_{k-1|K} \\ \sum_{k=1}^{K} x_{k-1|K} & \sum_{k=1}^{K} \tilde{W}_{k-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^{K} x_{k|K} \\ \sum_{k=1}^{K} \tilde{W}_{k-1,k} \end{bmatrix} \tag{54}$$

$$\begin{bmatrix} \alpha \\ h \end{bmatrix} = \begin{bmatrix} K & \sum_{k=1}^{K} x_{k|K} \\ \sum_{k=1}^{K} x_{k|K} & \sum_{k=1}^{K} \tilde{W}_{k} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^{K} z_k \\ \sum_{k=1}^{K} z_k x_{k|K} \end{bmatrix} \tag{55}$$

$$\sigma_W^2 = \frac{1}{K}\left[\sum_{k=1}^{K} (z_k-\alpha)^2 - 2(z_k-\alpha)hx_{k|K} + h^2\tilde{W}_k\right] \tag{56}$$

$$\psi = \log\left(\frac{\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}}{\sum_{k=1}^{K}\sum_{j=1}^{J} \Delta \exp\left(gx_{k|K} + \frac{1}{2}\sigma_{k|K}^2 g^2 + \sum_{s=1}^{S} \beta_s n_{k,j-s}\right)}\right) \tag{57}$$

Details for solving for the remaining parameters $\mu, \eta, g, \{\beta_s\}$ using Newton-like methods are given in Appendix C.

Appendix C
Newton Algorithms to Solve Fixed Point Equations

A. Newton Algorithm for the Posterior Update

We note that $x_{k|k}$ as defined in (11) is the root of the function ρ:

$$\begin{aligned} \rho(x_{k|k}) &= -x_{k|k} + x_{k|k-1} + C_k\left[h\left(z_k - \alpha - hx_{k|k-1}\right) + \eta\sigma_W^2(m_k - p_{k|k})\right] \\ &\quad + \sum_{j=1}^{J} C_k\sigma_W^2\left[g(n_{k,j} - \lambda_{k,j|k,j}\Delta)\right] \\ \Rightarrow \rho'(x_{k|k}) &= -1 - C_k\sigma_W^2\left[\eta^2 p_{k|k}(1-p_{k|k}) + \sum_{j=1}^{J} g^2\lambda_{k,j|k,j}\Delta\right] \end{aligned}$$

Either the previous state estimate, $x_{k-1|k-1}$, or the one-step prediction estimate, $x_{k|k-1}$, can provide a reliable starting guess.

B. Binary Parameters

In this section we develop derivatives of the functions f_3 and f_4 for the purpose of enabling a Newton-like algorithm to find the fixed point pertaining to (49)-(50). Define:

$$\begin{aligned} f_3 &= \sum_{k=1}^{K} m_k - p_k - \frac{1}{2}\sigma_{k|K}^2\eta^2 p_k(1-p_k)(1-2p_k) \\ f_4 &= \sum_{k=1}^{K} m_k x_{k|K} - x_{k|K}p_k - \frac{1}{2}\sigma_{k|K}^2\eta p_k(1-p_k)\left[2 + x_{k|K}\eta(1-2p_k)\right] \end{aligned}$$

We arrive at the Jacobian:

$$\frac{\partial}{\partial\mu}\{f_3\} = -\sum_{k=1}^{K} p_k(1-p_k)\left[1+\frac{1}{2}\sigma_{k|K}^2\eta^2\left[(1-2p_k)^2-2p_k(1-p_k)\right]\right]$$

$$\frac{\partial}{\partial\eta}\{f_3\} = -\sum_{k=1}^{K} p_k(1-p_k)\left[x_{k|K}+\frac{1}{2}\sigma_{k|K}^2\eta^2(1-2p_k)^2x_{k|K}+\sigma_{k|K}^2\eta[1-2p_k-\eta x_{k|K}p_k(1-p_k)]\right]$$

$$\frac{\partial}{\partial\mu}\{f_4\} = -\sum_{k=1}^{K} p_k(1-p_k)\left[x_{k|K}+\sigma_{k|K}^2(1-2p_k)+\frac{1}{2}\sigma_{k|K}^2\eta^2\left[(1-2p_k)^2-2p_k(1-p_k)\right]\right]$$

$$\frac{\partial}{\partial\eta}\{f_4\} = -\sum_{k=1}^{K} p_k(1-p_k)\Bigg[x_{k|K}^2+\sigma_{k|K}^2\left[\eta x_{k|K}(1-2p_k)+1\right] + \frac{1}{2}\sigma_{k|K}^2x_{k|K}\left[2\eta\left(1-2p_k-\eta x_{k|K}p_k(1-p_k)\right)+\eta^2(1-2p_k)^2x_{k|K}\right]\Bigg]$$

C. Spiking Parameters

1) Finding g: In this section we develop derivatives of the functions f_5 and f_6 for the purpose of enabling a Newton-like algorithm to find the fixed point pertaining to (52). From (57), note that

$$\begin{aligned}\psi &= \log\left(\frac{\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}}{\sum_{k=1}^{K}\sum_{j=1}^{J}\Delta\exp\left(gx_{k|K}+\frac{1}{2}\sigma_{k|K}^2g^2+\sum_{s=1}^{S}\beta_s n_{k,j-s}\right)}\right)\\ \text{or } \Delta\exp(\psi) &= \frac{\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}}{\sum_{k=1}^{K}\sum_{j=1}^{J}\exp\left(gx_{k|K}+\frac{1}{2}\sigma_{k|K}^2g^2+\sum_{s=1}^{S}\beta_s n_{k,j-s}\right)}\end{aligned} \tag{58}$$

Define:

$$
\begin{aligned}
f_5 = \frac{\partial Q}{\partial g} &= \sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}x_{k|K} - \Delta(x_{k|K} + \sigma^2_{k|K}g)\exp\left(\psi + gx_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right) \\
&= \left(\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}x_{k|K}\right) \\
&- \sum_{k=1}^{K}\sum_{j=1}^{J} \frac{\left(\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}\right)(x_{k|K} + \sigma^2_{k|K}g)\exp\left(gx_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right)}{\sum_{k=1}^{K}\sum_{j=1}^{J}\exp\left(gx_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right)} \\
&= \left(\sum_{k=1}^{K} n_{k,j}x_{k|K}\right) - \left(\sum_{k=1}^{K} n_{k,j}\right)\left(\frac{a(g)}{b(g)}\right) \\
\Rightarrow f_5'(g) &= -\left(\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}\right)\frac{\partial}{\partial g}\left\{\frac{a(g)}{b(g)}\right\} = -\left(\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}\right)\frac{b(g)a'(g) - a(g)b'(g)}{b(g)^2} \\
a(g) &\triangleq \sum_{k=1}^{K}\sum_{j=1}^{J}(x_{k|K} + \sigma^2_{k|K}g)\exp\left(gx_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right) \\
b(g) &\triangleq \sum_{k=1}^{K}\sum_{j=1}^{J}\exp\left(gx_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right) \\
a'(g) &= \sum_{k=1}^{K}\left[\left(x_{k|K} + \sigma^2_{k|K}g\right)^2 + \sigma^2_{k|K}\right]\exp\left(gx_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right) \\
b'(g) &= \sum_{k=1}^{K}\left(x_{k|K} + \sigma^2_{k|K}g\right)\exp\left(gx_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s=1}^{S}\beta_s n_{k,j-s}\right) = a(g) \\
\Rightarrow \frac{f_5(g)}{f_5'(g)} &= \frac{\left(\sum_{k=1}^{K} n_{k,j}x_{k|K}\right)b(g)^2 - \left(\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}\right)a(g)b(g)}{\left(\sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}\right)\left(a(g)^2 - b(g)a'(g)\right)}
\end{aligned}
$$

2) Finding β_s: In this section we develop the derivative of the functions f_6 to find the fixed point pertaining to (53). Note that the equation for $\frac{\partial Q}{\partial \beta_s} = 0$ in (53) can be expressed as

$$
\begin{aligned}
f_6(\beta_s) &= \tilde{f}_5(\phi(\beta_s)) \\
&= a_0 - a_1\phi(\beta_s), \\
\phi(\beta_s) &= \exp(\beta_s), \\
\Rightarrow f_6'(\beta_s) &= -a_1\phi'(\beta_s) = -a_1\exp(\beta_s)
\end{aligned}
$$

where

$$
\begin{aligned}
a_0 &= \sum_{k=1}^{K}\sum_{j=1}^{J} n_{k,j}n_{k,j-s}, \\
a_1 &= \sum_{k,j:n_{k,j-s}=1}\Delta\exp\left(\psi + gx_{k|K} + \frac{1}{2}\sigma^2_{k|K}g^2 + \sum_{s'\neq s}\beta_{s'}n_{k,j-s'}\right)
\end{aligned}
$$

Note that since each $a_0 \geq 0$, $a_1 > 0$, and $\phi > 0$, it follows that $f_6'(\beta_s) < 0$ and thus f is monotonically decreasing. Moreover, since $f_6(0) \geq 0$, it follows that f has a unique zero x^* and thus a unique fixed point β_s^*, when considering all other parameters $\beta_{s'}$ fixed.