

CONTINUOUS-TIME FILTERS FOR STATE ESTIMATION FROM POINT PROCESS MODELS OF NEURAL DATA

Uri T. Eden and Emery N. Brown

Boston University and Massachusetts Institute of Technology

Abstract: Neural spike trains, the primary communication signals in the brain, can be accurately modeled as point processes. For many years, significant theoretical work has been done on the construction of exact and approximate filters for state estimation from point process observations in continuous-time. We have previously developed approximate filters for state estimation from point process observations in discrete-time and applied them in the study of neural systems. Here, we present a coherent framework for deriving continuous-time filters from their discrete-counterparts. We present an accessible derivation of the well-known unnormalized conditional density equation for state evolution, construct a new continuous-time filter based on a Gaussian approximation, and propose a method for assessing the validity of the approximation following an approach by Brockett and Clark. We apply these methods to the problem of reconstructing arm reaching movements from simulated neural spiking activity from the primary motor cortex. This work makes explicit the connections between adaptive point process filters for analyzing neural spiking activity in continuous-time, and standard continuous-time filters for state estimation from continuous and point process observations.

Key words and phrases: Adaptive estimation, neural data analysis, state-space models.

1. Introduction

The primary communication signal within the brain is the action potential. Action potentials are nearly instantaneous voltage discharges that propagate rapidly across neurons and cause neurotransmitter release and signal propagation between neurons (Dayan and Abbot (2001)). The sudden stereotyped nature of these action potentials suggests that they can be accurately described under a point process modeling framework (Brillinger (1992) and Brown, Barbieri, Eden and Frank (2003)). A sequence of action potentials plotted in time is called a neural spike train. In addition to communicating with spike trains, neural systems are dynamic in that individual neurons constantly change their response properties to relevant stimuli (Brown, Nguyen, Frank, Wilson and Solo (2001), Frank, Eden, Solo, Wilson and Brown (2002), and Frank, Stanley and Brown (2004)). Furthermore, groups of neurons maintain dynamic representations of

relevant stimuli in their ensemble firing patterns (Brown, Frank, Tang, Quirk and Wilson (1998) and Barbieri, Frank, Nguyen, Quirk, Solo, Wilson and Brown (2004)). For these reasons, the development of state-space algorithms to characterize the dynamic properties of neural systems from point process observations has been a productive research area in computational neuroscience.

State-space estimation from continuous-valued observations is a well-developed area in adaptive filter theory (Haykin (1996) and Jazwinski (1970)). Under this framework, the behavior of a system is modeled with dynamic state variables that follow stochastic evolution and are observed through a continuous-valued process with stochastic disturbances. When both the state and observation models are linear and Gaussian processes, the conditional densities of the states given the observations are solutions to recursive filter equations that describe time-varying Gaussian distributions (Stengel (1994)). In discrete-time, this solution is the well-known Kalman filter. In continuous-time the recursive filter equations are stochastic partial differential equations whose solution is given by the Kalman-Bucy filter. The Kalman and Kalman-Bucy filters are equivalent in that they provide matching solutions for the mean and variance estimate of the conditional density at each point in time.

Exact continuous-time point process filtering algorithms, as well as discrete and continuous-time approximate filters for state estimation from point process observations, are well known (Elliott (1990), Kushner (1990), and Snyder and Miller (1991)). In particular, formulae for the evolution of both the unnormalized and normalized conditional density of a state given point process observations have been derived (Snyder and Miller (1991), Twum-Danso (1997) and Solo (2000)).

We have previously developed discrete-time, adaptive filter algorithms for state estimation from point process observations by computing recursively a Gaussian approximation to the conditional density and the Chapman-Kolmogorov (one-step) prediction density of the state (Brown, Frank, Tang, Quirk and Wilson (1998), Smith and Brown (2003) and Eden, Frank, Barbieri, Solo and Brown (2004)). These filters are point process analogues of the Kalman filter, recursive least-squares, and steepest descent algorithms for Gaussian observations.

The remainder of this paper is organized as follows. In Section 2, we define the continuous and discrete-time state and observation models, and in Section 3 we provide an elementary derivation of the unnormalized conditional density equation. Section 4 highlights the effect of spike observations on the conditional density. Section 5 discusses exponential representations of the solution to the conditional density equation and suggests a method for predicting the accuracy of approximate Gaussian filters. In Section 6, we derive a new continuous-time Gaussian approximate filter for state estimation from point processes. This filter is a point process analogue to the Kalman-Bucy filter. In Section 7, these

continuous-time filtering algorithms are applied to the problem of reconstructing a hand trajectory from simulated neural spiking activity from the primate primary motor cortex. In this example, we fit physiologically motivated neural models to the simulated data, and subsequently show that a Gaussian approximation to the conditional density of the state leads to accurate trajectory estimates. Using the approximate filter derived in Section 6, we reconstruct a sample trajectory from simulated ensemble spike train activity from motor cortical neurons in response to reaching movements.

2. Continuous and Discrete-Time State and Observation Models

To construct continuous-time analogues of our point process estimation algorithms, we express the stochastic models on which they are built, one for the state evolution and one for the observations process, in equivalent discrete and continuous-time forms.

The state evolution model used in the development of standard continuous-time estimation procedures, such as the Kalman-Bucy filter, takes the form of a Wiener-driven stochastic process

$$dx(t) = Ax(t)dt + Bdw(t), \quad (2.1)$$

where at any time t , $x(t)$ is a state vector of dimension n , $w(t)$ is a standard multivariate Wiener process, and A and B are constant $n \times n$ matrices defining the drift and scale parameters of the process, respectively.

To describe the discrete-time filters, we partition an observation interval, $[0, T]$, into $\{t_k : 0 \leq t_1 < \dots < t_N \leq T\}$. All pertinent model components, including the state model and observation values, are then measured or estimated at these specified times. To obtain a discrete-time expression for the state transition equation, we use the fact that the solution to (2.1) is a time-varying Gaussian process, and stochastically integrate this Wiener-driven process to compute the mean and variance of this density as a function of the discretization times. This yields the stochastic difference equation

$$x_k = F_k x_{k-1} + \varepsilon_k, \quad (2.2)$$

where $x_k = x(t_k)$, $F_k = \exp(A(t_k - t_{k-1}))$ and $\varepsilon_k = \int_{t_{k-1}}^{t_k} \exp(A(t_k - \tau))dw(\tau)$ is a zero-mean Gaussian random variable with variance $Q_k = \int_{t_{k-1}}^{t_k} \exp(A(t_k - \tau))BB'(\exp(A'(t_k - \tau)))d\tau$.

Next, we define discrete and continuous-time models for the observation process. In continuous-time we define the conditional intensity function for the i th neuron in an ensemble of C neurons as

$$\lambda^i(t|H_t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(i\text{th neuron spikes in } (t, t + \Delta t]|H_t)}{\Delta t} \quad (2.3)$$

where H_t is the history of the spiking activity of the entire ensemble up to time t . The conditional intensity function provides a complete description of the point process distributions (Daley and Vere-Jones (2003)). We can therefore formulate parametric models for these conditional intensity functions that fully describe the firing properties of each neuron in time. In general, the conditional intensity models can be a function of external or biological signals, the neuron's own previous spiking history, or the activity of other neurons within the ensemble (Truccolo, Eden, Fellows, Donoghue and Brown (2005)). By expressing the conditional intensity as a function of the state process (i.e., writing down a model of the form $\lambda^i(t|H_t) = g(t, x(t), H_t)$ for some function g), we define the probability distribution of spiking activity as a function of the state. Since by (2.3), the conditional intensity implicitly defines the conditional probability of a spike in a small time interval, it is also appropriate for describing the stochastic model for the spiking process in discrete time, in which case we write λ_k^i in place of $\lambda^i(t_k|H_{t_k})$.

The observations must also be expressed in continuous and discrete-time forms for the appropriate estimation paradigm. This is easily done by considering the counting process in continuous-time, $N^i(t) = \int_0^t dN^i(u)$, which counts the total number of spikes observed from the i th neuron in the interval $(0, t]$. Here, dN^i is the differential element of the counting process for the i th neuron. In discrete time, $\Delta N_k^i = N^i(t_k) - N^i(t_{k-1}) = \int_{t_{k-1}}^{t_k} dN^i(u)$ is the number of spikes observed in the interval $(t_{k-1}, t_k]$. We assume that if the $\Delta t_k = t_k - t_{k-1}$ elements are sufficiently small then there will be at most one spike in the interval, and the collection ΔN_k^i for all k is a sequence of zeros and ones used to express spike data in a discrete time series. The probability of neuron i spiking in $(t_{k-1}, t_k]$ is

$$\Pr(\Delta N_k^i | x_k, H_{t_k}) = (\lambda_k^i \Delta t_k)^{\Delta N_k^i} e^{-\lambda_k^i \Delta t_k} + o(\Delta t_k), \quad (2.4)$$

where $x_k = x(t_k)$ and $o(\Delta t_k)$ is a function such that $\lim_{\Delta t_k \rightarrow 0} (o(\Delta t_k)/\Delta t_k) = 0$ (Andersen (1996)).

3. A Continuous-Time Equation for the Unnormalized Density

The estimation problem deals with computing the time-varying conditional probability density of the state given the observations. One approach to tracking this probability density in continuous-time is to construct a differential equation, based on the Fokker-Planck equation, describing the evolution of the conditional density as a diffusion process (Stengel (1994)). It is sufficient to track an unnormalized version of this density in time, since it is always possible to compute its integral and renormalize at any later point.

As with our discrete-time point process filtering paradigm (Brown, Frank, Tang, Quirk and Wilson (1998) and Eden, Frank, Barbieri, Solo and Brown

(2004)), we develop a recursive expression for the conditional density given the observations process up to a given time from its value at a previous point in time by using Bayes' rule:

$$\rho(x(t + \Delta t) | \{\Delta N^i\}_{i=1}^C, H_t) = \prod_{i=1}^C \frac{\Pr(\Delta N^i | x(t + \Delta t), H_t)}{\Pr(\Delta N^i | H_t)} p(x(t + \Delta t) | H_t), \quad (3.1)$$

where $\Delta N^i = N^i(t + \Delta t) - N^i(t)$ and H_t is the complete history of spiking observations from the entire neural population up to time t . We call $p(x(t + \Delta t) | H_t)$ the one-step prediction density, $\Pr(\Delta N^i | x(t + \Delta t), H_t)$ the firing probability distribution for the i th neuron, and $\rho(x(t + \Delta t) | \{\Delta N^i\}_{i=1}^C, H_t)$ the conditional density for the state process. Here we have assumed that, within the small time interval Δt , the firing intensity of each neuron does not depend on the firing of other neurons in the ensemble within that interval. In the limit as $\Delta t \rightarrow 0$, this limitation vanishes as long as the entire process is orderly. For notational simplicity, we derive the update equation for a single neuron, and then discuss how it generalizes to spiking observations for a neural ensemble.

The denominator on the right side of (3.1) is a normalization term for the conditional density that does not depend on $x(t)$. Computing this normalization constant at each time is computationally demanding. An alternate approach is to replace this with another normalization term that facilitates the computation of an unnormalized version of the conditional density. Following Solo (2000), we use the probability of seeing the same observation under a simple homogeneous Poisson process, $\Pr(\Delta N | \text{Poisson}(1))$, as the normalization term. This renormalization serves to simplify the computation of (3.1). The resulting expression for the discrete-time evolution of the unnormalized conditional density for a single neuron is then

$$\rho(x(t + \Delta t) | \Delta N, H_t) = \frac{\Pr(\Delta N | x(t + \Delta t), H_t)}{\Pr(\Delta N | \text{Poisson}(1))} p(x(t + \Delta t) | H_t). \quad (3.2)$$

This unnormalized density can later be renormalized as needed at any point in time.

For notational convenience, we write $\rho(t)$ (or simply ρ) in place of $\rho(x(t) | \Delta N, H_t)$ and λ in place of $\lambda(t | H_t)$, from here on. The fraction on the right side of (3.2) is a likelihood ratio that is completely determined by the conditional intensity functions of the neurons in the ensemble. The only remaining term that needs to be evaluated is the one-step prediction density. One approach to computing this density is to use the Chapman-Kolmogorov equation and pass to the limit as $\Delta t \rightarrow 0$ (Solo (2000)). Here, we take an alternate approach. In the absence

of any observations, the evolution of ρ is given by the Fokker-Plank operator, defined under an Itô calculus to be

$$L(\rho) \equiv -A\rho - Ax \frac{\partial \rho}{\partial x} - \frac{1}{2} BB' \frac{\partial^2 \rho}{\partial t^2} \tag{3.3}$$

(Stengel (1994)). We use this expression to define the evolution of the conditional density from t to $t + \Delta t$ which, for small Δt , gives the one-step prediction density $p(x(t + \Delta t)|H_t) = \rho(t) + L(\rho(t))\Delta t + o(\Delta t)$.

We next compute the unnormalized conditional density in the limit as the discrete time interval approaches zero, obtaining

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \rho(t + \Delta t) &= \lim_{\Delta t \rightarrow 0} \left(\frac{(\lambda \Delta t)^{\Delta N} e^{-\lambda \Delta t}}{(\Delta t)^{\Delta N} e^{-\Delta t}} (\rho(t) + L(\rho(t))\Delta t) \right) \\ &= \lim_{\Delta t \rightarrow 0} \left((\lambda)^{\Delta N} e^{-(\lambda-1)\Delta t} (\rho(t) + L(\rho(t))\Delta t) \right). \end{aligned} \tag{3.4}$$

We examine the cases of spike times and non-spike times separately. By a non-spike time, we mean a time t for which there exists an ε such that for all $\Delta t < \varepsilon$ no spike occurs in $[t, t + \Delta t)$. In this case, we are interested in the standard definition of the derivative of the unnormalized density with respect to time,

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} (\rho(t + \Delta t) - \rho(t))\Delta t^{-1} &= \lim_{\Delta t \rightarrow 0} \left((e^{-(\lambda-1)\Delta t} - 1)\rho(t) + e^{-(\lambda-1)\Delta t} L(\rho(t))\Delta t \right) \Delta t^{-1} \\ &= \lim_{\Delta t \rightarrow 0} \left(-(\lambda - 1)\rho(t) + L(\rho(t)) \right). \end{aligned} \tag{3.5}$$

Therefore, the differential equation describing the evolution of the conditional density at times when no spike occurs is

$$\frac{\partial \rho}{\partial t} = L(\rho) - (\lambda - 1)\rho. \tag{3.6}$$

On the other hand, when a spike does occurs at time t , we are interested in the instantaneous jump in the density between the instants just before and just after the spike. We compute this as the limit of the change in density at t as $\Delta t \rightarrow 0$:

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \rho(t + \Delta t) - \rho(t) &= \lim_{\Delta t \rightarrow 0} \left(\lambda e^{-(\lambda-1)\Delta t} (\rho(t) + L(\rho(t))\Delta t) - \rho(t) \right) \\ &= (\lambda - 1)\rho(t). \end{aligned} \tag{3.7}$$

This is the jump in the density caused by a spike. Combining (3.6) and (3.7), we can write a diffusion equation for the unnormalized conditional density:

$$d\rho = (L(\rho) - (\lambda - 1)\rho)dt + (\lambda - 1)\rho dN. \tag{3.8}$$

Since adding any multiple of ρ to itself only serves to change the normalizing factor, we could alternately write an unnormalized density equation with the first

$(\lambda - 1)$ term replaced by λ . If observations from an ensemble of simultaneously recorded spiking neurons are available, and we assume that each neuron can only affect the future firing of other neurons, it is easy to show that the unnormalized conditional density equation reflects the superposition of the contributions from each neuron individually, and is

$$d\rho = \left(-A\rho - Ax \frac{d\rho}{dx} + \frac{1}{2}BB' \frac{d^2\rho}{dx^2} - \sum_{i=1}^C \lambda^i \rho \right) dt + \sum_{i=1}^C (\lambda^i - 1) \rho dN^i. \quad (3.9)$$

The solution to this density equation depends on the initial value of the distribution, on the form of the state and observation models, and on the spiking data. For initial densities with broad support over the appropriate region of the state space, the effect of the initial distribution will vanish as the observation interval and the number of observed spikes becomes large.

4. Instantaneous Density Changes

Consider the evolution of the conditional density based on the observation of a single neuron. The dN term on the right hand side of (3.9) indicates that when a spike occurs the unnormalized density undergoes a jump discontinuity at every value of the state. Specifically, if we let $\rho^- = \lim_{t \uparrow u} \rho(t)$ and $\rho^+ = \lim_{t \downarrow u} \rho(t)$ then, at a spike time u ,

$$\begin{aligned} \rho^+ &= \lim_{a \uparrow u} \lim_{b \downarrow u} \left(\int_a^b d\rho \right) + \rho^- \\ &= \lim_{a \uparrow u} \lim_{b \downarrow u} \left(\int_a^b \left(L(\rho) - \lambda(t|H_t)\rho(t) \right) dt + \left(\lambda(t|H_t) - 1 \right) \rho(t) dN(t) \right) + \rho^- \\ &= \lambda(u|H_u)\rho^-. \end{aligned} \quad (4.1)$$

The conditional density immediately following a spike is simply the product of the conditional density preceding the spike and the conditional intensity function for the spiking neuron. Therefore, if the conditional intensity is small in the region of the state space where most of the mass of the conditional density resides, and high elsewhere, then a spike can significantly and instantaneously change the overall structure of the conditional density.

5. Exponential Representations of Solutions

Based on Brockett and Clark (1980), we can construct a representation of the unnormalized conditional density solution for (3.9) as $\rho(t) = \exp\{\sum f_i G_i(x)\} \rho(0)$, where $\{f_i\}$ is a set of real-valued functions in time, and $\{G_i(x)\}$ is a set of functions of the state x . The G_i are chosen so that this exponential form encompasses the solution to the diffusion equation for any possible set of spiking observations.

In the absence of any spiking activity, the solution to the diffusion equation has an exponential representation determined by the drift operator in (3.9),

$$\left(-A - Ax \frac{d}{dx} + \frac{1}{2} \left(\frac{\partial}{\partial x} \right)' BB' \left(\frac{\partial}{\partial x} \right) - \sum_{i=1}^C \lambda^i \right), \quad (5.1)$$

and the initial state distribution. Equation (4.1) makes clear the effect of a spike from any neuron on the conditional density, namely that the function for the conditional density immediately after the spike is the product of the density just before the spike and the intensity of the spiking neuron as a function of the state. Writing this to suit an exponential representation, we obtain

$$\rho^+ = \lambda \rho^- = \exp\{\log(\lambda)\} \rho^-. \quad (5.2)$$

Therefore, to construct an exponential representation for the conditional density, we need a set of basis functions, $\{G_i(x)\}$, that include the contribution from the drift term and the initial distribution on the state, as well as $\log(\lambda^i)$ for every neuron in the ensemble. Additionally, since the drift term will exert its influence before and after each spike event, we need this set of basis functions to span all commutators of the drift and spike terms, all commutators of those operators, and so on. In other words, if we can construct a finite basis for a set of operators that includes this drift and the jump terms for each neuron and is closed under commutation, then there exists an exponential representation for the solution with a finite number of terms. This set of operators is called the Lie algebra generated by this drift, $L(\cdot) - \sum \lambda^i$, and the jump terms, $\log(\lambda^i)$, for $i = 1, \dots, C$ (Brockett (1981) and Sagle and Walde (1973)).

For the unnormalized density to have a Gaussian solution for our Gaussian state and point process observation model, the basis functions must be given by quadratic polynomials. In general, there is no functional form for the conditional intensity models so that the Lie algebra generated by this system has such a basis. Therefore, Gaussian solutions to the diffusion equation will not exist for these spiking systems. However, in many cases, Gaussian approximations to the conditional density have been used to compute accurate state estimators and characterize their uncertainty (Brown, Frank, Tang, Quirk and Wilson (1998), Barbieri, Frank, Nguyen, Quirk, Solo, Wilson and Brown (2004) and Ergun, Barbieri, Eden, Wilson and Brown (2007)). Under such a scheme, it is possible that errors related to the Gaussian approximation will be large and accumulate over time, leading to inaccurate estimators, or that these errors will remain small and their effects on the estimators limited in duration. The degree to which either of these possibilities occurs will depend on how each term in the Lie algebra affects the conditional density at each point in time.

This theory suggests a practical approach to evaluate the potential accuracy of the Gaussian approximation to the conditional density used in point process adaptive filters. To the extent that quadratic functions approximate well the effect of the drift and jump terms and their commutators, the log of the conditional density will be nearly quadratic and the Gaussian approximation should be accurate. For example, when the neural spiking models have the form $\lambda^i(t) = \exp\{a_i x(t)^2 + b_i x(t) + c_i\}$, then the $\sum \lambda^i$ in the drift operator is the only term causing the conditional distribution to deviate from a Gaussian. Twum-Danso (1997) and Twum-Danso and Brockett (2001) showed that for a regular lattice of hippocampal spatial receptive fields with identical Gaussian shapes, this term is remarkably well-approximated by a constant, and therefore an expression for an accurate Gaussian approximation to the conditional density can be constructed. More generally, if the $\sum \lambda^i$ term is well approximated by any quadratic function, then the Gaussian approximation should be accurate.

This result suggests that, for any neural decoding solution computed from an ensemble of neurons of this form, it is prudent to examine the structure of $\sum_{i=1}^c \lambda^i(t|H_i)$ over the relevant region of the state space to assess the accuracy of the Gaussian approximation. For a low-dimensional state vector, this could be accomplished by plotting the sum of intensities over the state-space (e.g. Figure 2 in the example in Section 7) to assess how well it is fit by a quadratic polynomial. For more complicated intensity models, it may be useful to expand the sum of intensities in a Taylor series and compare the magnitude of the constant, linear, and quadratic coefficients to higher order terms.

6. A Continuous-time Approximate Filter

The diffusion equation derived in Section 3 is exact in the sense that its solution at any point in time is equal to the conditional probability density up to a normalization constant. However, this differential equation cannot typically be solved analytically, and standard numerical methods can become computationally difficult. On the other hand, the discrete-time stochastic state point process filter (SSPPF) (Eden, Frank, Barbieri, Solo and Brown (2004)) uses a Gaussian approximation to the conditional density at each time step to construct a computationally simple recursion (3.1). Similar to the derivation of the Kalman-Bucy filter, we can construct a continuous-time analogue to the SSPPF by applying a limiting procedure as the time step in the discretization approaches zero.

In order to construct the SSPPF, we used the Bayesian recursion in (3.1) and applied the Chapman-Kolmogorov equation to compute the one-step prediction density, expanded the log of the posterior in a Taylor series about the one step prediction mean, $x_k = x_{k|k} \equiv E[x_k | \{\Delta N_k^i\}_{i=1}^C, H(t_k)]$, dropped terms higher than second order, and obtained a linear, discrete-time recursive Gaussian

approximation to the conditional density. Under the discrete-time state model given by (2.2) with observations process (2.3), the equations for the estimated conditional variance, $W_{k|k} = \text{var}(x_k | \{\Delta N_k^i\}_{i=1}^C, H(t_k))$, and mean are, respectively,

$$(W_{k|k})^{-1} = (F_k W_{k-1|k-1} F_k' + Q_k)^{-1} + \sum_{i=1}^C \left[\left(\frac{\partial \log \lambda^i}{\partial x_k} \right)' [\lambda^i \Delta t_k] \left(\frac{\partial \log \lambda^i}{\partial x_k} \right) - (\Delta N_k^i - \lambda^i \Delta t_k) \frac{\partial^2 \log \lambda^i}{\partial x_k \partial x_k'} \right]_{x_k = F_k x_{k-1|k-1}}, \tag{6.1}$$

$$x_{k|k} = F_k x_{k-1|k-1} + W_{k|k} \sum_{i=1}^C \left[\left(\frac{\partial \log \lambda^i}{\partial x_k} \right)' (\Delta N_k^i - \lambda^i \Delta t_k) \right]_{x_k = F_k x_{k-1|k-1}}. \tag{6.2}$$

For the derivation of a continuous-time analogue to the SSPPF, we are interested in the observation interval $(t, t + \Delta t]$, and the values of the estimated mean and variance process solely at the beginning and end of this interval, in the limit as $\Delta t \rightarrow 0$. We define \hat{x} and \hat{W} as our continuous-time expressions for the conditional mean and variance, respectively, in this limit, and write $\hat{x}^- = \lim_{\Delta t \rightarrow 0} x_{k-1|k-1}$ and $\hat{x}^+ = \lim_{\Delta t \rightarrow 0} x_{k|k}$ for the mean estimator, and $\hat{W}^- = \lim_{\Delta t \rightarrow 0} W_{k-1|k-1}$ and $\hat{W}^+ = \lim_{\Delta t \rightarrow 0} W_{k|k}$ for the variance estimator. In this limit, the discrete-time expressions in (6.1) and (6.2) become continuous-time stochastic differential equations that characterize the evolution of the mean and variance of the conditional density

$$d\hat{x} = A\hat{x}^- dt + \hat{W}^+ \sum_{j=1}^C \left[\left(\frac{\partial \log \lambda^j}{\partial x} \right)' (dN^j - \lambda^j dt) \right]_{\hat{x}^-}, \tag{6.3}$$

$$d\hat{W} = (A\hat{W}^- + \hat{W}^- A' + BB')dt - \sum_{j=1}^C \hat{W}^- \left(\frac{d^2 \lambda^j}{dx dx'} dt + S^j dN^j \right)_{\hat{x}^-} \hat{W}^- \tag{6.4}$$

where,

$$S^j = \begin{cases} \left(\hat{W}^- - \left(\frac{\partial^2 \log \lambda^j}{\partial x \partial x'} \right)^{-1} \right)^{-1} & \text{if } \left(\frac{\partial^2 \log \lambda^j}{\partial x \partial x'} \right) \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

From any initial values, this system of differential equations completely describes the evolution of the Gaussian approximation to the conditional density in continuous time. As with the diffusion equation for the unnormalized density, when the initial estimate of the variance term is sufficiently large relative to the range of values that the state takes on, the effect of the initial estimates vanishes as

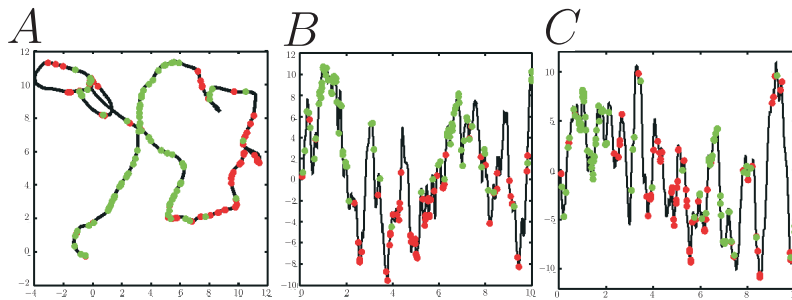


Figure 1. Sample arm movement with simulated spiking activity of two neurons, one in red and the other in green, overlaid. (A) (x, y) coordinates of movement trajectory, (B) x -velocity versus time, (C) y -velocity versus time.

the size of the observation interval and the number of spikes becomes large. The step-by-step procedure for computing these limiting equations is available as an online supplement to this article on the Statistica Sinica website.

7. An Application to Motor Decoding

We illustrate the application of these continuous-time approximate methods to the problem of estimating the firing properties of an ensemble of simulated primary motor cortical neurons that respond during reaching arm movements, and reconstructing movement trajectories from this simulated ensemble spiking activity. This problem relates to the important application of developing neural prosthetic devices for patients who have lost the ability to perform voluntary arm movements, for example through brain or spinal cord injury or neural degeneration.

Elsewhere (Truccolo, Eden, Fellows, Donoghue and Brown (2005)), we applied the SSPPF to real spike train data from the primary motor cortex of macaque monkeys to characterize the firing properties of these neurons in relation to arm movement, spiking history, and interactions among neurons. Here, we simulate neurons based on these neural models with realistic parameter values drawn from these previously characterized spiking properties. We generated smooth simulated arm movement data as an autoregressive process with parameters estimated from real arm trajectories recorded from a monkey executing continuous reaching movements. The values of these parameters, as well as the specific details of the state and observation models used in this simulation, are described in further detail in the online supplement to this article on the Statistica Sinica website. An example of one of these reaching arm movements is shown in black in Figure 1.

We simulated spiking activity from neural intensity models of the form

$$\lambda^i(t|H_t) = \exp \left\{ \alpha^i + \beta^i |v(t + 150\text{ms})| \cos(\phi(t + 150\text{ms}) - \phi_{\text{pref}}^i) + \sum_{\tau=1}^{130\text{ms}} \gamma_{\tau}^i \Delta N_{[t-\tau, t-(\tau-1)]}^i \right\}, \quad (7.1)$$

where $v(t)$ and $\phi(t)$ are the speed and direction of the arm movement, $\Delta N_{[a,b]}^i$ is the total number of spikes fired by the i th neuron in the interval $[a, b)$, and α^i , β^i , ϕ_{pref}^i , $\{\gamma_{\tau}^i\}_{\tau=1}^{130}$ are respectively the model parameters related to the baseline firing rate, velocity modulation, preferred direction, and history dependence of the neuron. Under this model, the spiking activity predicts movement variables that occur after a lag of 150 ms. This form of kinematic tuning is based on a model of Moran and Schwartz (1999). We simulated 20 neurons generated by selecting α^i , β^i , ϕ_{pref}^i , $\{\gamma_{\tau}^i\}_{\tau=1}^{130}$ randomly over a range of realistic parameter values estimated previously in Truccolo et al. (2005). The spiking activity from two of these neurons relative to an arm trajectory is shown in Figure 1. The neuron whose spiking is labeled in green tends to fire more during movements in the positive x and y -directions, while the one whose spiking is labeled in red tends to fire during movements in the negative x and y -directions. The true tuning functions for each of the 20 simulated neurons is shown in Figure 2A.

We performed an encoding analysis by estimating model parameters for the conditional intensity function for each neuron, for models of the form of (7.1), from a training set containing 20 minutes of simulated spiking data. The methods and results of this analysis, along with a goodness-of-fit analysis are presented in the online supplement to this paper on the *Statistica Sinica* website. We can ask the question whether the ensemble activity of these 20 estimated neurons contains sufficient information about the arm kinematics to reconstruct accurately the trajectory of a new arm movement solely from the observed ensemble spiking activity.

To address this question we adopt a state space model for the arm movement trajectory as in (2.1), with $x(t) = \begin{bmatrix} v(t) \cos(\phi(t)) \\ v(t) \sin(\phi(t)) \end{bmatrix}$, $A = 0$ and $B = 10 \cdot I$. Given the maximum likelihood estimates of the conditional intensity functions for the ensemble, the solution to the conditional density is given by (3.9). To determine whether a Gaussian approximation to this solution would be appropriate, we calculated the value of $\sum \lambda^i$ over the range of movement velocities. The result is shown in Figure 2A. It is clear that the depth of directional tuning for this

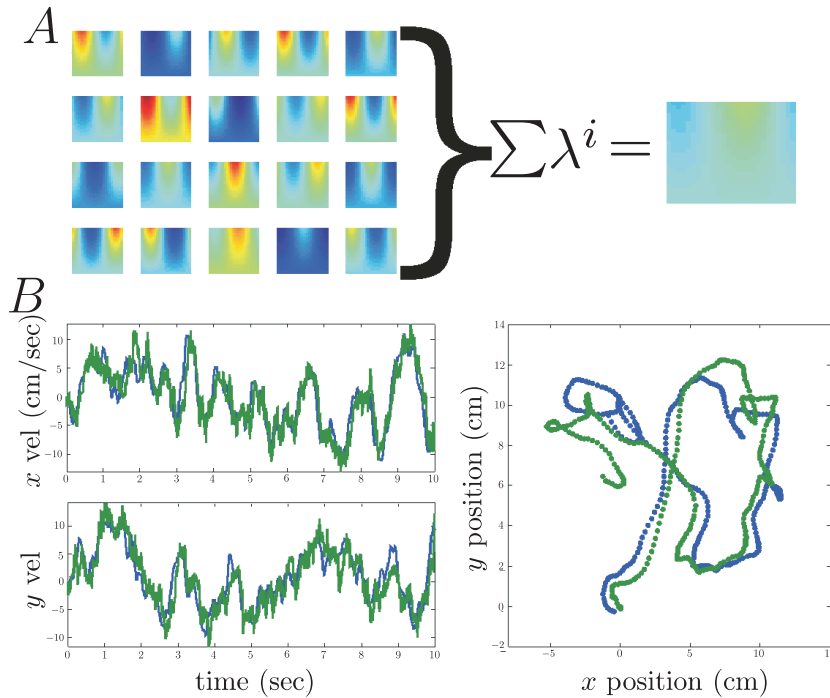


Figure 2. Decoding Analysis. (A) True receptive field properties of each neuron, and the sum term, $\sum \lambda^i$. This sum term has a tuning depth that is much smaller than any individual neuron and can be well approximated by a linear function of the movement speed, suggesting that the Gaussian approximation in the filter algorithm will be good. (B) True arm movement profile (blue) and filter estimates (green) for a single arm movement. Velocity tracking remains accurate, while position estimation suffers from integrated errors.

combined ensemble activity is much smaller than that of any individual neuron. This sum of conditional intensities can be well approximated by a function that increases linearly with speed and is independent of direction. Therefore, we expect Gaussian approximations to the conditional density to be quite accurate.

Based on these results, we applied the continuous-time Gaussian approximate filter given by (6.3) and (6.4) to these data. The differential equations were solved numerically using an Euler's method at a time step of 1 msec, for non-spike intervals, and computing the jumps directly at the spike times. The resulting estimated arm movement, illustrated in Figure 2B, is in close agreement with the actual movement. The root mean squared error for reconstructing the velocity signal was 2.56 cm/sec in the x-direction, and 2.37 cm/sec in the y-direction. These mean squared error values combine error related to the Gaussian approximation with observation error due to the fact that the spiking observations are

not perfectly informative about the state. From previous work (Ergun, Barbieri, Eden, Wilson and Brown (2007)), we conjecture that the error observed here is overwhelmingly dominated by observation error. This could be verified explicitly (at great computational expense) by comparing this filter to the mean of the solution of the exact conditional density equation.

8. Discussion

We previously used a state space framework with point process observations to track neural plasticity (Frank, Eden, Solo, Wilson and Brown (2002)), to characterize learning from changes in neural firing (Czanner, Eden, Wirth, Yanike, Suzuki and Brown (2008)), and to reconstruct behavioral and biological signals from observed spiking activity (Brown, Frank, Tang, Quirk and Wilson (1998), Barbieri, Frank, Nguyen, Quirk, Solo, Wilson and Brown (2004) and Eden, Trucolo, Fellows, Donoghue and Brown (2004)). In each of these cases, the key to performing estimation is tracking the evolution of the conditional density with time and with incoming spike observations. Whereas these previous analyses have focused on tracking this density, or a Gaussian approximation, in discrete time, the continuous-time framework offers an important perspective on point process estimation.

Computing the conditional density in continuous-time allows us to calculate expectations of any function of the state process and therefore to construct estimators for these processes at any point in time within an observation interval, not just those times selected for the discrete-time partition of this interval. Beyond that, the coefficients associated with differential equations that describe a system are generally more amenable to physical interpretation than those associated with difference equations. For example, whereas Q , the covariance of the discrete-time state equation only has meaning for a specific time partition, the scale matrix or volatility, B , of the Wiener process driving the continuous-time state equation (2.1) describes a physical property of the state process at all time scales. Furthermore, this continuous-time framework makes available a wide range of additional numerical methods for calculation of the conditional density associated with numerical methods for solving partial differential equations (Evans, Blackledge and Yardley (1999)).

From a control theoretic perspective, the differential equation for the unnormalized conditional density maps out the space of possible distributions that can be achieved. In this point-of-view, (3.9) can be seen as a control equation composed of a drift term and terms related to the spiking activity of each observed neuron. The drift term, represented by the first bracketed term on the right side of (3.9), contains combined features of each neural receptive field and determines how the conditional density evolves absent any spiking observations.

Intuitively, when the ensemble is silent, the conditional density of regions of the state space with high firing probability for any neuron decreases while that of areas with low firing probability for all neurons increases. The second term on the right side of (3.9) can be thought of as the sum of control terms related to each observed neuron. When a neuron fires, it causes the conditional density to jump discontinuously to a new density whose value at each point in the state space is equal to the product of the conditional density immediately before that spike and the intensity function of the neuron that fired. Therefore, each neuron causes the conditional density to jump so that it reflects the features of its own intensity function more strongly.

The continuous-time perspective can also serve as a guide to help us improve our discrete-time estimation methods. For example, a common numerical technique for tracking the conditional density to high precision in discrete-time is sequential Monte Carlo, also known as particle filtering (Doucet, de Freitas and Gordon (2001)). The method involves maintaining a cloud of samples that evolve in time, and empirically approximating the conditional density. The most commonly applied particle filtering algorithm is the bootstrap filter whereby, at each point in time, the empirical estimate from the previous point in time is used as the starting point for the evolution of the particles. From the above discussion about the continuous-time conditional density equation, it is clear that a single spike can cause the conditional density to jump so that most of its mass moves from one region of the state space to another. This jump will cause bootstrap particle filters to be computationally inefficient, since they require a significant number of particles in the region where the density prior to a spike was low. In Ergun, Barbieri, Eden, Wilson and Brown (2007), we describe an alternate particle filter that accounts for these jumps to improve the computational efficiency of estimation.

In Eden (2005), we made an analogy between discrete-time point process filters and the Kalman filter for systems with continuous-valued observations with Gaussian errors. Likewise, the continuous-time approximate Gaussian filter developed in Section 6 can be seen as a point process analogue of the Kalman-Bucy filter, in that we construct stochastic differential equations for the first two moments of the conditional density and use those, respectively, as estimates for the state and of the gain. The primary advantage of this approximate filter over the exact conditional density equation is that it allows us to obtain estimates of the state, and a sense of its uncertainty, using methods associated with ordinary differential equations rather than with partial differential equations, as are needed to calculate the full conditional density.

The curvature of the conditional intensity function plays a significant role in determining the evolution of the covariance or gain term in (6.4) of the

continuous-time approximate filter. For any neuron, when the curvature of its intensity function is high, the spike observation makes the conditional covariance drop toward zero. When the intensity function is log linear, spikes will not cause the conditional covariance to change at all. Therefore spikes from neurons with narrow intensity functions are much more informative than those with broader ones.

This continuous-time approximate filter in (6.3) and (6.4) is similar to one developed by Snyder and Miller (1991). They took the first two moments of the normalized density equation and replaced the expectations of the intensity at each point in time with the intensity at the mean estimate. This approximation should only be good when the intensity function is close to log linear in the state. As a consequence, Snyder's approximate filter differs significantly from our result in the magnitude of the jump in the conditional variance at spike times.

Classically, another benefit of continuous-time expressions for state space filters has been the fact that they can be easily implemented using analog electronic circuitry. The value of an estimated state can be encoded as the voltage level at a node, and the effects of the drift and observations can be simulated using circuit elements such as differentiators and integrators. Although analog computation has decreased in popularity with the rise of rapid digital computation, there are at least two promising areas of application for analog computation of neural state space estimates: the dynamic clamp and neural prosthetics. Dynamic clamping involves recording from a cell or ensemble and dynamically injecting current into each cell to mimic the effect of physiological inputs. Analog electronics are often used to rapidly compute the desired current injection as a function of the spiking observations. The continuous-time filters developed here would allow for the rapid computation of state estimates upon which to base a dynamic clamp injection paradigm. Similarly, closed-loop motor prosthetic devices must convert neural activity related to motor intent to a control signal for a cursor or robotic prosthesis in real time. Current motor prosthesis implementations digitize these signals and transmit them to a remote computer to perform estimation. The computational speed of these devices could be significantly improved through hardware implementation of neural filtering algorithms, which could occur locally at the recording site.

From a theoretical standpoint, these results provide an effective framework for addressing the problem of estimating dynamic signals from neural spike train data. By making the link between these continuous-time point process filters and our previous estimation algorithms, as well as to classical filtering algorithms whose properties are well understood, we can be sure that these methods are internally consistent, regardless of whether we choose to work in discrete or continuous time.

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant no. IIS-0643995, and the National Institutes of Health under grants no. R01-DA015644, R01-MH59733 and NSF CDI-Type II, #0836720.

References

- Andersen, P. K. (1996). *Statistical Methods Based on Counting Processes*. Springer-Verlag, New York.
- Barbieri, R., Frank, L. M., Nguyen, D. P., Quirk, M. C., Solo, V., Wilson, M. A. and Brown, E. N. (2004). Dynamic analyses of information encoding by neural ensembles. *Neural Computation*. **16**, 277-308.
- Brillinger, D. R. (1992). Nerve cell spike train data analysis: a progression of technique. *J. Amer. Statist. Assoc.* **87**, 260-271.
- Brockett, R. W. and Clark, J. M. C. (1980). The geometry of the conditional density equation. In *Analysis and Optimisation of Stochastic Systems* (Edited by O. L. R. Jacobs), Academic Press, New York.
- Brockett, R. W. (1981). Nonlinear systems and nonlinear estimation theory, In *The Mathematics of Filtering and Identification and Applications* (Edited by M. Hazewinkel and J. Willems), Reidel, Dordrecht.
- Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C. and Wilson MA. (1998). 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.
- Brown, E. N., Nguyen, D. P., Frank, L. M., Wilson, M. A. and Solo, V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *Proc. Nat. Acad. Sci.* **98**, 12261-66.
- Brown, E. N., Barbieri, R., Eden, U. T. and Frank, L. M. (2003). Likelihood methods for neural data analysis. In *Computational Neuroscience: A Comprehensive Approach* (Edited by J. Feng), 253-286. CRC, London.
- Czanner, G., Eden, U. T., Wirth, S., Yanike, M., Suzuki, W. A. and Brown, E. N. (2008). Analysis of between-trial and within-trial neural spiking dynamics. *J. Neurophys.* Submitted.
- Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, 2nd Edition*. Springer, New York.
- Dayan, P. and Abbot, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, Cambridge, MA.
- Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- Eden, U. T., Frank, L. M., Barbieri, R., Solo, V. and Brown, E. N. (2004). Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Comp.* **16**, 971-998.
- Eden, U. T., Truccolo, W., Fellows, M. R., Donoghue, J. P. and Brown, E. N. (2004). Reconstruction of Hand Movement Trajectories from a Dynamic Ensemble of Spiking Motor Cortical Neurons. Proceedings of the 26th annual international conference IEEE Engineering in Medicine and Biology Society. San Francisco, CA.
- Eden, U. T. (2005). Point process filters in the analysis of neural spiking models. PhD. Thesis in Medical Engineering/Medical Physics. Harvard/MIT Division of Health Sciences and Technology.

- Elliott, R. J. (1990). Filtering and control for point process observations. *Recent Advances in Stochastic Calculus* (Edited by J. Baras and V. Mirelli), 1-27. Progress in Automation and Information Systems, Springer-Verlag.
- Ergun, A., Barbieri, R., Eden, U. T., Wilson, M. A. and Brown, E. N. (2007). Construction of point process adaptive filter algorithms for neural systems using sequential Monte Carlo methods. *IEEE Transactions on Biomedical Engineering*. **54**, 419-428.
- Evans, G., Blackledge, J. and Yardley, P. (1999). *Numerical Methods for Partial Differential Equations*. Springer.
- Frank, L. M., Eden, U. T., Solo, V., Wilson, M. A. and Brown, E. N. (2002). Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. *Journal of Neuroscience*. **22**, 3817-30.
- Frank, L. M., Stanley, G. B. and Brown, E. N. (2004). Hippocampal plasticity across multiple days of exposure to novel environments. *Journal of Neuroscience*. **24**, 7681-89.
- Haykin, S. (1996). *Adaptive Filter Theory*. 3rd Edition. Prentice-Hall.
- Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.
- Kushner, H. J. (1990). Numerical methods for stochastic control problems in continuous time. *SIAM J. Control Optim.* **28**, 999-1048.
- Moran, D. W. and Schwartz, A. B. (1999). Motor cortical representation of speed and direction during reaching. *J. Neurophysiol.* **82**, 2676-2692.
- Sagle, A. and Walde, R. (1973). *Introduction to Lie Groups and Lie Algebras*. Academic Press, New York.
- Smith, A. C. and Brown, E. N. (2003). Estimating a state-space model from point process observations. *Neural Computation*. **15**, 965-991.
- Snyder, D. L. and Miller, M. I. (1991). *Random Point Processes in Time and Space*. Springer-Verlag. New York.
- Solo, V. (2000). Unobserved Monte Carlo method for identification of partially observed nonlinear state space systems, Part II: Counting process observations. Proc 39th IEEE Conference on Decision and Control, 3331-3336. Sydney, Australia.
- Stengel, R. F. (1994). *Optimal Control and Estimation*. Dover, Mineola, NY.
- Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P. and Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol.* **93**, 1074-1089.
- Twum-Danso, N. T. (1997). *Estimation, information and neural signals*. PhD. Thesis, Division of Engineering and Applied Sciences, Harvard.
- Twum-Danso, N. T. and Brockett, R. (2001). Trajectory estimation from place cell data. *Neural Networks*. **14**, 835-844.

Department of Mathematics and Statistics, 111 Cummington St., Boston University, Boston, MA 02215, U.S.A.

E-mail: tzvi@bu.edu

Department of Brain and Cognitive Sciences, Harvard/MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

E-mail: brown@neurostat.mgh.harvard.edu

(Received April 2007; accepted March 2008)