PERGAMON

Neural Networks 14 (2001) 835–844

Neural Networks

www.elsevier.com/locate/neunet

2001 Special issue

Trajectory estimation from place cell data

Nanayaa Twum-Danso, Roger Brockett*

Division of Engineering and Applied Sciences, Harvard University, 33 Oxford Street, Cambridge, MA 02138, USA

Accepted 8 March 2001

Abstract

We consider the problem of propagating the conditional probability density associated with the movement parameters (position, heading, velocity, etc.) of an animal, given the responses of an ensemble of place cells. While we are not the first to look at this question, ours seems to be the first treatment that incorporates a general Markov process model for the motion parameters and a general observation model postulating place cells centered in a lower dimensional 'measurement space' formed from combinations of the Markovian variables. An important part of our analysis involves the determination of a suitable set of sufficient statistics for propagating the conditional density in this context. Making use of these results we are led to approximations which greatly simplify the estimation problem and various aspects of its neuroscientific interpretation. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Place cells; Poisson processes; Conditional density; Position estimation; Measurement model

1. Introduction

The purpose of this paper is to discuss the fundamental limitations on the accuracy of place cell decoding from a communications theoretic point of view. We develop a general definition of place cell coding, derive and analyze the optimal method for extracting data from the spike trains generated by place cell coding. In doing so, we make use of an unusually effective approximation which greatly simplifies the analysis of the conditional density for the movement parameters. The potential value of this analysis for neuroscientists is that it makes quantitative predictions about performance based on a more flexible model of both the movement process and the measurement mechanism. This formalism can be easily adapted to accommodate other modulation effects, such as the use of timing relative to EEG rhythms, etc.

The idea that certain areas of the hippocampus represent the present position of the animal by means of groups of cells that are tuned to fire rapidly only when the animal is at, or near, a particular location, has proven to be quite fruitful, both as an inspiration for experimental work and as a starting point for theoretical analyses. More specifically, the problem of finding the most likely reconstruction of the trajectory describing the animal's recent movements based on a knowledge of the firing pattern of an ensemble of such place cells, has been considered by a number of authors and the literature is growing rapidly. Recent papers include a broad analysis of various decoding schemes by Zhang, Ginzburg, McNaughton and Sejnowski (1998) and a more narrowly focused statistical analysis by Brown, Frank, Tang, Quirk and Wilson (1998). These authors argue, and this is our point of view as well, that it is of interest to find the theoretical limits on the extent to which the trajectory can be determined because this permits one to make definitive statements about the adequacy of information carried by these signals for describing the motion.

Experimental evidence has been reported suggesting a role for several variations on the basic place cell model. One important variation is that the firing rate is a function of both the position and the velocity. We regard this as place cell representation in position–velocity space. Evidence for other modulation schemes have been reported, including the possibility of coding information about position within the support set of the tuning curve by the timing of the spiking relative to the waveform of the theta rhythm component of the EEG. Thus, there are good arguments for treating the subject in some generality.

Our analysis builds on work done over the last 50 years in which the critical question for estimation theory has come to center on the propagation of the conditional density. This point of view makes it clear that the problem of forming a causal estimate of a stochastic process depends strongly on the statistics of the process whose path is to be estimated. In this literature, a prominent role is played by processes that

* Corresponding author. Tel.: +1-617-495-3922; fax: +1-617-496-6404.
E-mail address: brockett@hrl.harvard.edu (R. Brockett).

PII: S0893-6080(01)00079-X

are generated from Brownian motion via linear or nonlinear stochastic differential equations. In particular, the so-called Gauss–Markov processes play an important role.

There are two ways in which our analysis differs from that discussed in the literature.

1. We work within a mathematically natural setting established for generating causal estimates of stochastic processes. In this theory one begins with a Markov model for the process to be estimated and a model for the observables that is independent of the reconstruction process.
2. We consider the possibility of estimating the state of an n-dimensional Markov process from the spike trains coming from a regular array of place cells covering an m-dimensional space. This is the type of formalism needed, for example, to estimate the position and velocity in two dimensions from place cell data responding to position alone.

The paper is organized as follows. We begin with a quick review of the relevant parts of estimation theory adapting it to suit the present problem. We then derive the evolution equation for an unnormalized form of the conditional density equation which forms the basis of our subsequent analysis. We follow this with a discussion of the question of the computational complexity associated with propagating the conditional density equation. We then discuss an approximation scheme which leads to a pair of equations for the conditional mean and the conditional variance. Finally, we study the reconstruction error as a function of the relevant parameters. Space limitations prevent us from doing complete justice to the underlying mathematical theory. The reader interested in these matters should consult the thesis of Twum-Dansio (1997) for this and for a detailed numerical study of some performance issues.

2. Markov models for movement

Beginning around 1960, Kalman and Bucy (1961) and others reformulated in an iterative form the theory initiated by Wiener and Kolomogrov relating to the use of noisy measurements and an a priori model to generate a least squares causal estimate of the current value of a stochastic process. This work made the subject both more relevant in various applied situations and more accessible to non specialists. By now, there is a huge literature and a wealth of applications. Although in its original form the theory applied only to linear systems whose measurements were corrupted by additive Gaussian noise, this work suggested extensions embracing nonlinear problems and over the years it has been extended in numerous ways. In outline form, the ingredients of this theory as it now exists are:

- a Markov model for the stochastic process that is to be estimated;
- a probabilistic model for the way in which the measurements relate to the true values of variables associated with the underlying stochastic process; and
- a propagation rule for the conditional density.

The book of Jazwinski (1970) can be consulted for a detailed account.

In this paper, we apply this methodology to study the problems involving coding and decoding relevant to problems involving place cells and their responses. Two particular types of primitive stochastic processes will play a role—the Poisson counter and Wiener's model for Brownian motion. In particular, we will make use of the idea of a Poisson counter of rate λ. These will usually be denoted by N or N_i, if several are involved. Poisson processes take on the value 0 at $t = 0$ and assume values in the set of nonnegative integers, $\{0, 1, 2, \ldots\}$. The probability that N will advance in a small time interval of length Δ is given by $\lambda\Delta + \gamma$ with γ being second order in Δ.

Definition. We will say that N is a variable rate Poisson counter of rate λ if N takes on values in the set $\{0, 1, 2, \ldots\}$ and for all t the probability $p(t, \Delta)$ associated with the event $N(t + \Delta) - N(t) = 1$ satisfies

$$\lim_{\Delta \to 0} \frac{p(t, \Delta)}{\Delta} = \lambda(t)$$

By definition, the counting rate λ is nonnegative. If λ is a constant, then such processes are said to be homogeneous; if λ depends on t, they are called inhomogeneous. They are said to be conditional if λ is allowed to depend on auxiliary random variables or random processes. The use of such models requires some care as will be touched on below.

From the point of view of minimizing mathematical technicalities, the simplest problems in this area involve estimation of the state of a finite state Markov process. Because we find it most reasonable to work with continuous time models, we consider processes that jump at a random time from one of a finite set of possible values to another according to a probabilistic rule. More concretely, let $x(t)$ take on values in the finite set $\{x_1, x_2, \ldots, x_r\}$. Let $p_i(t)$ denote the probability that $x(t) = x_i$ and suppose that the probability law is such that

$$\dot{p}(t) = A(t)p(t)$$

The x-process is then a finite state, continuous time, Markov process. The matrix A, called the intensity matrix, provides a description of how its probability law evolves in time. Of course, the description of the evolution of the probability law does not provide a description of the sample paths. Sample path descriptions are conveniently thought of as being generated from Poisson counters using differential equations in the following way. If $N(t)$ is a continuous time Poisson process, then, of course, $\mathrm{d}N(t)/\mathrm{d}t$ is zero except

at the times when N jumps. At these times, the derivative of N fails to exist in the ordinary sense, but can be interpreted as a Dirac delta function. In this sense, we can write

$$\frac{dN(t)}{dt} = \sum \delta(t - \alpha_i)$$

Considering the times α_i to be the times of initiation of action potentials, this can model neural signals. We interpret differential equations written as

$$\dot{x} = \sum_{i=1}^{m} \phi_i(x(t)) \frac{dN_i(t)}{dt}$$

in the following way. When all the Ns are constant $x(t)$ does not change. If the i-th counter jumps at time β, then $x(t)$ jumps from its value just before β to the value

$$x(\beta_+) = x(\beta_-) + \phi_i(x(\beta_-))$$

That is to say, the amount of the jump is determined by the value of $x(t)$ at the left-hand end point of the trajectory. This is the Itô interpretation of the differential equation. (See Appendix A for a discussion of alternative notations found in the literature.) Under standard assumptions, the probability that two or more counters would advance at the same time is zero and so this gives us a complete description of the sample path behavior. It is not difficult to show that any finite state, continuous time, a Markov process can be realized by an equation of this form by suitably selecting the functions ϕ_i and the rates of the Poisson counters.

Remark 1. If N is a counter of constant rate λ, then the expected value of the square of the deviation between $N(t + T) - N(t)$ and λT is given by

$$\mathcal{E}(N(t + T) - N(t) - \lambda T)^2 = \lambda T$$

Put in a more directly relevant form, the ability to estimate the rate of a Poisson counter from the sample path is limited by

$$\mathcal{E}\left(\frac{N(t + T) - N(t)}{T} - \lambda \right)^2 = \frac{\lambda}{T}$$

To reduce the variance we would like to take T to be large, but this makes it impossible to rapidly sense and respond to a change in the counting rate. This expression makes it clear that there is uncertainty about the underlying signal λ even though N is observed perfectly. It also puts in evidence the source of the timeliness–accuracy trade-off.

A second class of Markov models, useful for modeling situations in which the x process takes on values in a Euclidean space or some other manifold, are the diffusion models, to which we now turn. In this case, the sample path description depends on the Wiener process, which we usually denote by w or w_i if several are needed. The x process is generated from one or more Wiener processes by a stochastic differential equation of the form

$$\dot{x} = f(x) + \sum_{i=1}^{n} g_i(x) \dot{w}_i$$

The corresponding equation for the probability law is typically expressed in terms of a density, $\rho(t, x)$ which satisfies a second order equation of the diffusion type,

$$\frac{\partial \rho(t, x)}{\partial t} = L\rho(t, x)$$

Remark 2. The Fokker–Planck operator for the Gauss–Markov model

$$\dot{x} = Ax + B\dot{w}$$

is

$$L = -\sum \frac{\partial}{\partial x_i} a_{ij} x_j + \frac{1}{2} \sum \gamma_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$$

where γ_{ij} is the ij-th entry of the matrix BB^T. These are the so-called Gauss–Markov models. They are especially useful as prototypes because they often lead to closed form expressions and yet are general enough to generate arbitrary second order statistics.

3. Place cell models for measurement

Whereas many, or even most, aspects of the representation of information in the nervous system remains mysterious, one experimentally verified idea about the representation of geometrical data is that of place cell representation. This way of representing positions and trajectories differs markedly from what one finds in most areas of science and engineering. Usually, we think of the evolution of quantities such as the position of a particle or the temperature of an object as defining a curve in a space in which time is represented on the horizontal axis and the quantity in question is represented on the vertical axis, using some numerical scale. When the numerical representation is taken to be continuous, this gives a continuous evolution in Cartesian space. In electrical engineering, one often calls such a representation an analog representation to distinguish it from a digital representation, where typical waveforms do not show continuity with respect to the underlying phenomenon being represented. Place cell representation can also be described in mathematical terms, but the representation differs from those traditionally studied in physical science.

Let S be a convex, subset of $\mathbb{R}^n$ and let $\{s_k\}_{k=1,2,\ldots,r}$ be a set of points in S. Let $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ be a set of functions mapping S into the nonnegative real numbers. Define a function $\phi_\Lambda : S \to S$ with

$$\phi_\Lambda(x) = \sum_{k=1}^{r} \frac{\lambda_k(x)}{\sum \lambda_k(x)} s_k$$

and a second function $\psi_x : \mathbb{R}^r \to S$ with

$$\psi_x(\Lambda) = \sum_{k=1}^{r} \frac{\lambda_k(x)}{\sum \lambda_k(x)} s_k$$

Definitions. Let S and Λ be as above. Suppose that the λs have the properties

1. The functions λ_k are continuous and for each λ_k there exists a single point $s_i \in S$ where λ_k takes on its maximum value.
2. The values of each λ_k are monotone decreasing along radial lines directed away from the point where it takes on its maximum value.
3. At each point $x \in$ S at least one $\lambda_k(x)$ is nonzero.

Under these circumstances, we will say that the set Λ defines a *place cell family* for S. The points s_i will be called the *centers* and the function λ_k will be called *tuning curves*. We will say that a collection of spike trains $\{dN_k/dt\}$; $k = 1, 2, \ldots, r$ is a *place cell-spike frequency* representation of a point $x \in$ S if the counting rates of the Poisson counters N_k are given by λ_k.

Examples. Prototypical one dimensional tuning curves include the Gaussian shape

$$\lambda_k(x) = a e^{-(cx - s_k)^2/2}$$

having a center at s_k/c and a width which decreases with increasing c, and the raised cosine

$$\lambda_k(x) = b(1 + \cos(cx - s_k)); \qquad |x| \le \pi/c$$

also having a center at s_k/c and a width which decreases with increasing c.

The function ϕ_Λ assigns to the point x a convex combination of the centers s_k. If $s_1, s_2, \ldots, s_r$ are points in a vector space, the *convex hull* of these points consists of the set of all points of the form

$$x = \sum_{k=1}^{r} \alpha_k s_k; \alpha \ge 0; \qquad \sum \alpha_k = 1$$

In this context, it is useful to recall the well known result of Carathéodory stating that given a closed convex subset of n-dimensional space, and a point x in the set, there always exist $n + 1$ points on the boundary of the subset such that x is a convex combination of these points. Place cell representations are by no means constrained to use this minimal number of points and, consequently, can be much more robust and less sensitive to small changes in the weights α_k.

Early work by Georgopolis, Kettner and Schwartz (1988) considered the situation in which the place cells were explicitly related to geometric data. It has been reported subsequently by Zhang et al. (1998) that place cells can have directional dependence as well as a position dependence. For this reason, we formulate our problem in a general way, retaining the words place cell, but extending the meaning so as to permit the interpretation that 'place' may be a location anywhere in the state space of the Markov process defining the motion. For example, it might define a field in some position–velocity space with coordinates $(\xi, \dot{\xi})$. Thus, if x takes one values in $\mathbb{R}^n$ and C is an $m \times n$ matrix, we consider ensembles of place cells with tuning curves of the form

$$\lambda_k(x) = a e^{-(Cx - s_k)^T (Cx - s_k)/2}$$

Assuming a fixed lower bound on the spiking rate, one expects that it will be possible to more accurately represent the variables being coded if the place cells are more dense and if the upper limit on spiking frequencies is higher. This is most easily quantified in a situation in which the place cells are uniformly distributed. In one dimension, this means that the tuning curves of the ensemble of place cells each have the same shape, differing in that they are generated by repeatedly shifting the center of this shape by a fixed distance. For reasons that will emerge later, we characterize the geometry of the place cell distribution using two variables, C and α. The matrix C defines a linear transformation from the state space of the Markov process to an m dimensional *measurement space*, where the place cells are situated, and α is a positive number that we use to adjust the spacing between the place cells. If y is one dimensional we have

$$\lambda(y) = a e^{-(y - \alpha s)^2/2}; \qquad s = 0, \pm 1, \pm 2, \ldots$$

We can say that $1/\alpha$ determines the density of the place cells in y-space.

If x is assumed to be a random variable with a Gaussian probability density $\rho(x) = (1/\sqrt{2\pi q}) e^{-x^2/2q}$, then the expected counting rate for a cell centered at s is

$$\lambda_s = \int_{-\infty}^{\infty} a e^{-(cx - \alpha s)^2/2} \frac{1}{\sqrt{2\pi q}} e^{-x^2/2q} dx$$

$$= \frac{a}{\alpha} \frac{\sqrt{c^2 q + 1}}{\sqrt{2\pi q}} e^{-(\alpha^2 s^2)/[2(c^2 q + 1)]}$$

The total number of spikes produced by the entire ensemble in a fixed interval of time is an important parameter, often considered to be a constraint on the system. This overall counting activity is the sum of these terms over the entire lattice and is given by

$$\lambda^* = \frac{a}{\alpha} \frac{\sqrt{c^2 q + 1}}{\sqrt{2\pi q}} \sum_{s \in \mathbb{Z}} e^{-(\alpha^2 s^2)/[2(c^2 q + 1)]}$$

where $\mathbb{Z}$ denotes the set of integers. Appealing to the

inequalities

$$\sqrt{2\pi q} - \frac{1}{q} = \frac{-2}{2q} + 2\int_0^\infty \mathrm{e}^{-s^2/2q}\mathrm{d}s \le \sum_{s\in\mathbb{Z}} \mathrm{e}^{-s^2/2q}$$

$$\le \frac{2}{2q} + 2\int_0^\infty \mathrm{e}^{-s^2/2q}\mathrm{d}s = \sqrt{2\pi q} + \frac{1}{q}$$

we see that

$$\sqrt{(c^2q+1)/\alpha^2} - 1/q \le \lambda^* \le \sqrt{(c^2q+1)/\alpha^2} + 1/q$$

In a higher dimensional setting, uniformity can be modeled by assuming that there exists a place cell centered at each point in a point lattice. Let the set of points in $\mathbb{R}^m$ that have all integer components be denoted by $\mathbb{Z}^m$. We call this the *standard m-dimensional point lattice*. Let the place cells and their tuning curves be indexed by a points $s \in \mathbb{Z}^m$ according to

$$\lambda_s(y) = a\mathrm{e}^{-(y-s)^T(y-s)/2}$$

It would be unduly restrictive to limit the discussion to place cells that are separated by integers in each coordinate. To achieve more generality, we incorporate a scaling parameter α which rescales the point lattice and leads to the family

$$\lambda_s(y) = a\mathrm{e}^{-(y-\alpha s)^T(y-\alpha s)/2}$$

Theorem 1. Suppose that x is a Gaussian random variable with zero mean and variance Q. Assume that x determines the counting rate of a family of Poisson counters $\{N_s\}$ according to the rule

$$\lambda_s(x) = a\mathrm{e}^{-(1/2)(Cx-\alpha s)^T(Cx-\alpha s)}; \qquad s \in \mathbb{Z}^m$$

Then, the expected spiking rate for λ_s is

$$\bar{\lambda}_s = \frac{a\sqrt{\det Q_1}}{\alpha_m\sqrt{\det Q_0}} e^{-(\alpha^2/2)s^TQ_1^{-1}s}$$

where $Q_0 = CQC^T$ and $Q_1 = (I + (CQC^T)^{-1})^{-1}$. The expected value of the total number of spikes from the ensemble per unit time is bounded by

$$\frac{a\sqrt{\det Q_1}}{\alpha^m\sqrt{\det Q_0}}(1 + \sqrt{2\pi\lambda_{\min}(Q_1)})^m \le \lambda^*$$

$$\le \frac{a\sqrt{\det Q_1}}{\alpha^m\sqrt{\det Q_0}}(1 + \sqrt{2\pi\lambda_{\max}(Q_1)})^m$$

where $\lambda_{\min}(Q_1)$ and $\lambda_{\max}(Q_1)$ are the minimum and maximum eigenvalue of Q_1, respectively.

Proof. Let $y = Cx$ so that y is a Gaussian random variable with covariance Q_0. Then

$$\mathrm{E}\lambda^* = a\int_{\mathbb{R}^m} \mathrm{e}^{-(1/2)(y-\alpha s)^T(y-\alpha a)} \frac{1}{\sqrt{(2\pi)^m\det Q_0}} e^{-(1/2)y^T(Q_0)^{-1}y}\mathrm{d}y$$

Rearranging the exponents we get

$$\mathrm{E}\lambda^* = \frac{a}{\sqrt{(2\pi)^m\det Q_0}} \int_{\mathbb{R}^m} e^{-(1/2)(y^T(Q_1)^{-1}y - \alpha y^Ts - (1/2)\alpha^2 s^Ts)}\mathrm{d}y$$

which integrates to give the result of the theorem. The upper bound on the total ensemble spike rate is obtained using the estimate

$$\mathrm{e}^{-y^TQ^{-1}y} \le \mathrm{e}^{-y^Ty/\lambda_{\max}(Q)}$$

valid for positive definite matrices Q, together with the scalar version of this estimate invoked above.

Remark 3. If we scale C by a factor β, then the above estimates show that the expected value of the total spike rate saturates at a multiple of a/α^m as β grows large.

Corollary. Let A be a square matrix having eigenvalues with negative real parts. If the stochastic process x is generated by

$$\dot{x} - Ax + B\dot{w}$$

and if

$$Q = \int_0^\infty \mathrm{e}^{At}BB^T\mathrm{e}^{A^Tt}\mathrm{d}t$$

then when the steady state conditions are achieved, the counting rate is given by Theorem 1 with this value of Q.

We point out for future reference what is otherwise reasonably clear, namely that if the overall counting rate λ^* is to remain bounded in the limit as α goes to zero then the rate coefficient a must be scaled with α^m.

4. Reconstruction I: the conditional density equation

Following Sanger (1996) who investigated the conditional density associated with a constant process, we now consider the problem of the optimal estimation of the evolution of the process x based on place cell signals. We study this to determine how accurately a place cell representation captures the signal, not because there is any evidence that this reconstruction is needed for some biological purpose.

As we have said, in order to carry out a Bayesian analysis of the problem of recovering $x(t)$ from the past values of $N_1(t)$, $N_2(t)$, …, $N_m(t)$ it is necessary to have a probabilistic model for the stochastic process x. We provide an outline of this analysis for a simple model and refer the reader to the literature for the generalizations that are used below. Suppose that $x(t)$ is a continuous time jump process taking on values in the set $X = \{x_1, x_2, \ldots, x_n\}$. Let $p_i(t)$ be the probability that $x(t)$ takes on the value x_i and assume that

$$\dot{p} = A_p$$

for some known intensity matrix A. Assume further, that we are given m functions $\lambda_1(x), \lambda_2(x), \ldots, \lambda_m(x)$ mapping the set

X into the nonnegative real line. Let $\{N_i\}_{i=1}^m$ be m conditional Poisson counters with the rate of the i-th counter being $\lambda_i(x)$. If we observe the sample paths of the counters N_1, N_2, …, N_m we would naturally ask what is the conditional probability of x given these observations? This problem is conveniently analyzed using Bayes' rule in the form

$$p(x|N_1, N_2, ..., N_m) = \frac{p(N_1, N_2, ..., N_m|x)p(x)}{p(N_1, N_2, ..., N_m)}$$

To streamline the notation, we introduce a family of diagonal matrices $\{\Lambda_k(x)\}$ defined as

$$\Lambda_k(x) = \begin{bmatrix} \lambda_k(x_1) & 0 & 0 & \cdots & 0 \\ 0 & \lambda_k(x_2) & 0 & \cdots & 0 \\ 0 & 0 & \lambda_k(x_3) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \lambda_k(x_n) \end{bmatrix}; \qquad k = 1, 2, ..., r$$

In a small time interval h in which none of the counters advance, the Bayes' formula implies that the conditional probability evolves in accordance with

$$p(t + h) = \alpha\left(p(t) + Ap - h\sum_{k=1}^{m} \Lambda_k(x)p + \mathrm{e} \right)$$

where e is of order h^2 and α is the (unique!) normalization factor required to make the components of $p(t + h)$ sum to one. Taking the limit as h goes to zero we get

$$\dot{p}(t) = \alpha(t)\left(A - \sum_{k=1}^{m} \Lambda_k \right)p(t)$$

The difference between this evolution and the evolution of the probability of x with no observation is the presence of the negative sum. This term reflects the fact that if there are no jumps the probability shifts toward the values of x associated with slower counting rates. On the other hand, if the k-th counter advances at time t, then Bayes' rule implies that the probabilities should immediately change to reflect the fact that when the k-th counter jumps the probabilities shift in such as way as to reflect the tuning curve associated with the k-th counter. A direct application of Bayes' rule in this situation gives

$$p(t_+) = \alpha\Lambda_k(x)p(t_-)$$

Notice that the Λ term increases the probability of the values of x associated with the larger counting rates of the k-th counter. Putting these two types of evolution together we have

$$\dot{p} = \alpha(t)\left(A - \sum_{k=1}^{m} (\Lambda_k(x)) + \sum_{k=1}^{m} (\Lambda_k(x) - I)p(t)\frac{\mathrm{d}N_k}{\mathrm{d}t} \right)$$

This is to be interpreted as a stochastic differential equation in the Itô sense. If N_k advances at time β then $p(\beta_-)$ changes to $\Lambda_k p(\beta_-)$. Moreover, the normalization factor α can be dispensed with because its only effect is to rescale p. In fact, the solution of the simpler equation

$$\dot{p} = \left(A - \sum_{k=1}^{m} (\Lambda_k(x))\mathrm{d}t + \sum_{k=1}^{m} (\Lambda_k(x) - I)p(t)\frac{\mathrm{d}N_k}{\mathrm{d}t} \right)$$

is related to the version with α present only by a scaling factor. However, one can normalize the solution of the unscaled equation after it is solved and this is generally much easier, so we will work with the unscaled version. Equations of this type have been in the literature for some time, the paper of Davis and Marcus (1981) surveys many aspects of the field. More relevant for our specific problem is the important paper of Boel and Beněs (1980).

We pause to point out two aspects of this equation. First of all it is to be interpreted as an Itô equation (see Appendix A) implying that the solutions are of the form

$$p(t) = \mathrm{e}^{At_r}\mathrm{e}^{\Lambda_{k_r}(x)}\mathrm{e}^{At_{r-1}}\mathrm{e}^{\Lambda_{k_{r-1}}(x)}\cdots\mathrm{e}^{\Lambda_{k_2}(x)}\mathrm{e}^{At_1}p(0)$$

Secondly, appealing to the symmetry principles discussed in Brockett and Clark (1980) we see that the complexity associated with solving this equation depends on the dimensionally of the Lie algebra generated by matrix $A - \sum \Lambda_k$ and the logarithms of the individual Λ_k.

With this analysis in mind, we now turn to a situation more directly relevant to the place cell reconstruction problem. Suppose that x takes on values in some n-dimensional space and that its probability density, in the absence of any measurements, evolves according to the Fokker–Planck equation

$$\frac{\partial\rho(t,x)}{\partial t} = \sum_{i=1}^{d} b(x)_i \frac{\partial\rho(t,x)}{\partial x_i} + \sum_{i,j=1}^{n} c_{i,j}\frac{\partial^2\rho(t,x)}{\partial x_i\partial x_j}$$

As suggested by the application of Bayes' rule carried out above, in this situation we have an equation for the unnormalized conditional density which takes the form

$$\frac{\partial\rho(t,x)}{\partial t} = (L - \sum \lambda_k(x))\rho(t,x) + \sum_{k=1}^{r} \lambda_k(x)\rho(t,x)\frac{\mathrm{d}N_k}{\mathrm{d}t}$$

where L is the operator defined by the right-hand-side of the Fokker–Planck equation. In this case, one says that x is a diffusion process. Typically, such a process is associated with a stochastic differential equation in which the stochastic element is the increments of a Brownian motion, i.e. x is the solution of a stochastic differential equation of the form

$$\dot{x} = f(x) + G(x)\dot{w}$$

suitably interpreted. The probability density itself evolves according to the corresponding Fokker–Planck equation

$$\frac{\partial\rho(t,x)}{\partial t} = L\rho(t,x)$$

Then, the conditional density equation takes the form

$$\frac{\partial\rho(t,x)}{\partial t} = \left(L - \sum_{k=1}^{m}\lambda_k(x)\right)\rho(t,x) + \sum_{k=1}^{m}(\lambda(x) - 1)\rho(t,x)\frac{\mathrm{d}N_i}{\mathrm{d}t}$$

Ordinarily, equations of this type cannot be solved analytically and must be approached numerically. However, there are special cases in which this partial differential equation can be reduced to a finite set of ordinary (but stochastic) differential equations. In such cases, one says that there are a finite set of *sufficient statistics* for the problem. The Kalman–Bucy filter, based on a quite different set of assumptions, is probably the best known example of a situation in which there is a finite set of sufficient statistics. The corresponding equation in that setting is

$$\frac{\partial\rho(t,x)}{\partial t} = \left(L - \frac{1}{2}(cx)^2\right)\rho(t,x) + cx\rho(t,x)\frac{\mathrm{d}y}{\mathrm{d}t}$$

Example. Let x be a two-dimensional vector with components x_1, x_2 representing a point in the plane. Suppose that x performs two-dimensional Brownian motion (i.e. a random walk in the plane). Then

$$L = \frac{1}{2}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right)$$

and the conditional density equation takes the form

$$\frac{\partial\rho(t,x)}{\partial t} = \left(\frac{1}{2}\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} - \sum_{k=1}^{m}\lambda_k(x)\right)\rho(t,x) + \sum_{k=1}^{m}(\lambda(x) - 1)\rho(t,x)\frac{\mathrm{d}N_i}{\mathrm{d}t}$$

This is a special case of the more general situation in which the x process is generated by the more general n-dimensional Gauss–Markov model described by the linear equation

$$\dot{x} = Ax + B\dot{w}$$

5. Reconstruction II: Gaussian solutions

In order to understand when and how the conditional density equation can be reduced to a set of ordinary stochastic differential equations driven by the observations, we focus on the objects

$$L_0 = L - \sum_{k=1}^{m}\lambda(x) \text{ and } L_k = \ln(\lambda_k(x)) \qquad k = 1, 2, \ldots, m$$

We not only want to think of L_0 as an operator, but we also want to think of the functions $L_k = \ln(\lambda_k(x))$ as operators in that they act on ρ sending ρ into $L_i\rho$.

The theory of sufficient statistics for problems of this type tells us that it is the structure of the Lie algebra generated from the operators by the adjoining commentators $[L_k, L_l]$, their commutators, etc. that ordinarily determinates the dimensionally of the possible sets of sufficient statistics. An important example of such a finite dimensional Lie algebra is afforded by the algebra consisting of linear combinations of the elements

$$\frac{\partial^2}{\partial x_i^2}, \qquad x_j\frac{\partial}{\partial x_i}, \qquad x_ix_j, \qquad \frac{\partial}{\partial x_i}, \qquad x_i, \qquad 1$$

Notice that the terms in the Fokker–Planck operator for the Gauss–Markov model are of the form appearing in this algebra. There is no choice for the functional description of the tuning curves $\lambda_k(x)$ such that both the individual curves and their sum are a linear combination of constants, terms linear in x, and terms quadratic in x. However, as we discuss in Appendix B,

$$\lim_{\alpha\to\infty}\sum_{s\in\mathbb{Z}^d}\frac{\alpha^m}{\sqrt{(2\pi)^m}}\mathrm{e}^{-(Cx-\alpha s)^T(Cx-\alpha s)/2} = 1$$

Moreover, and this is a significant point, even for values of α as large as one this approximation is remarkably good. Using this approximation with a suitable set of 'Gaussian' tuning curves we will show that if L_0 comes from a Gauss–Markov model, then because

$$\ln(\lambda_s(x)) = \alpha(x - s)^2$$

the conditional density can be propagated in a way analogous to the Kalman–Bucy filter. In this case, the standard unnormalized conditional density equation

$$\frac{\partial\rho}{\partial x} = L\rho - \sum_{s\in\mathbb{Z}^m}\gamma\mathrm{e}^{-(Cx-\alpha s)^T(Cx-\alpha s)/2}\rho + \sum_{s\in\mathbb{Z}^d}(\gamma\mathrm{e}^{-(Cx-\alpha s)^T(Cx-\alpha s)/2} - 1)\rho\frac{\mathrm{d}N_s}{\mathrm{d}t}$$

where

$$\gamma = \frac{a\alpha^m}{\sqrt{(2\pi)^m}}$$

If we approximate the first sum by its limit as α goes to zero then

$$\frac{\partial\rho}{\partial x} = L\rho - a + \sum_{s\in\mathbb{Z}^d}(\gamma\mathrm{e}^{-(Cx-\alpha s)^T(Cx-\alpha s)/2} - 1)\rho\frac{\mathrm{d}N_s}{\mathrm{d}t}$$

However, this equation does not enforce normalization, so we can as well drop the a. Reverting to a notation that places α in evidence, we get our final form for the unnormalized conditional density.

$$\frac{\partial\rho(t,x)}{\partial x} = L\rho + \sum_{s\in\mathbb{Z}^m}\left(\frac{a\alpha^m}{\sqrt{(2\pi)^m}}\mathrm{e}^{-(Cx-\alpha s)^T(Cx-\alpha s)/2} - 1\right)\rho(t,x)\frac{\mathrm{d}N_s}{\mathrm{d}t}$$

As just discussed, the conditional density equation leads to a finite set of sufficient statistics if the sum of the tuning curves, together with the logarithms of the individual tuning curves, match certain profiles. In particular, we observed that if the logarithms are quadratic, then there is a possibility

of finding solutions of the conditional density equation that are Gaussian.

If, as this analysis suggests, we assume that there is a regular array of exponential tuning curves centered on the points of a scaled regular point lattice and if we replace the sum of the tuning curves by a constant, then there will exist solutions of the form

$$\rho(t,x) = n(t)\mathrm{e}^{-(1/2)(x-m)^T\Sigma^{-1}(x-m)}$$

Theorem 2. Suppose that x satisfies the stochastic equation

$$\dot{x} = Ax + B\dot{w}; \qquad y = Cx$$

and that $x(t)$ determines the counting rate of a family of Poisson counters $\{N_s\}$ according to the rule

$$\lambda_s(x) = a\frac{\alpha^m}{\sqrt{(2\pi)^m}}\mathrm{e}^{-(1/2)(y-\alpha s)^T(y-\alpha s)}; \qquad s \in \mathbb{Z}^m$$

Assume further that the limiting approximation yielding the unnormalized conditional density equation

$$\frac{\partial\rho(t,x)}{\partial x} = (L-a)\rho(t,x) + \left(\sum_{s\in\mathbb{Z}^m}\frac{a\alpha^m}{(2\pi)^m}\mathrm{e}^{-(1/2)(y-s)^T(y-s)}(x) - 1\right)\rho(t,x)\frac{\mathrm{d}N_s}{\mathrm{d}t}$$

is used. Then if $\hat{x}$ and Σ satisfy the equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\dot{x} = A\hat{x} + \sum_{s\in\mathbb{Z}^m}(\Sigma + C^TC)^{-1}C^T(\alpha s - C\hat{x})\frac{\mathrm{d}N_s}{\mathrm{d}t}$$

and

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}t} = A\Sigma + \Sigma A^T + BB^T + ((\Sigma^{-1} + C^TC)^1 - \Sigma)\sum_{s\in\mathbb{Z}^m}\frac{\mathrm{d}N_s}{\mathrm{d}t}$$

respectively, then the conditional density of x given the past values of the counting processes is given by

$$\rho(t,x) = \frac{1}{\sqrt{(2\pi)^n\det\Sigma}}\mathrm{e}^{-(1/2)(x-\hat{x})^T\Sigma^{-1}(x-\hat{x})}$$

Remark 4. The differential equations are to be interpreted in the Itô sense as discussed in Appendix A.

Remark 5. We observe (see Appendix B) that because

$$\epsilon\sum_{s\in\mathbb{Z}^m}\alpha s\frac{\mathrm{d}N_s}{\mathrm{d}t} = \sum_{s\in\mathbb{Z}^m}\left(\alpha s a\frac{\alpha^m}{\sqrt{(2\pi)^m}}\mathrm{e}^{-(1/2)(y-\alpha s)^T(y-\alpha s)}\right) \approx ay$$

The equation for $\hat{x}$ can be thought of as being driven by a multiple of $(y - C\hat{x})$, much as the innovations drive the Kalman–Bucy filter.

Remark 6. The error variance is a function of the overall spike rate, λ^*. The spacing variable α does not enter the variance equation. In the close spacing situation considered here, a large value of C (narrow tuning curves) makes the variance smaller because of the way in which C, enters the variance equation. However, this effect saturates as C becomes larger, with the effect beginning to diminish when CC^T is comparable in size with Σ. The overall spike rate can be nondimensionalized by comparing it with the frequency range in which the power spectrum of the x process contains the most power.

Proof. We can solve the conditional density equation most easily by treating separately the intervals on which there are no spikes and the moments at which there are spikes. The probability of two or more spikes occurring at the same is zero so these events can be ignored. Between spikes, the normalized conditional density and the Fokker–Planck equation are the same so in these intervals the mean and variance satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}x = A\hat{x}$$

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}t} = A\Sigma + \Sigma A^T + BB^T$$

At a jump, Bayes' rule implies that the Gaussian density existing prior to the jump is multiplied by the tuning curve associated with the excited spike. The solution can be found by rearranging the product of two Gaussians. To this end, observe

$$\mathrm{e}^{-(1/2)(Cx-\alpha s)^T(Cx-\alpha s)}\mathrm{e}^{-(1/2)(x-m_-)^T\Sigma^{-1}(x-m_-)}$$

$$= \mathrm{e}^{-(1/2)(x-m_+)(C^TC+\Sigma^{-1})(x-m_+)}$$

with

$$m_+ = (\Sigma^{-1} + C^TC)^{-1}(\alpha C^Ts + \Sigma^{-1}m)$$

and, thus

$$\Sigma_+ = (\Sigma_-^{-1} + C^TC)^{-1}$$

Putting these two together gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{x} = A\hat{x} + \sum_{s\in\mathbb{Z}^m}(C^TC + \Sigma_-)^{-1}(\alpha C^Ts + \Sigma^{-1}\hat{x})\frac{\mathrm{d}N_s}{\mathrm{d}t} - \hat{x}\frac{\mathrm{d}N_s}{\mathrm{d}t}$$

which can be rearranged as

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{x} = A\hat{x} + \sum_{s\in\mathbb{Z}^m}(C^TC + \Sigma_-)^{-1}C^T(\alpha s - C\hat{x})\frac{\mathrm{d}N_s}{\mathrm{d}t}$$

For the variance, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum = A\Sigma + \Sigma A^T + BB^T + \sum_{s\in\mathbb{Z}^m}((\Sigma_-^{-1} + C^TC)^{-1} - \Sigma)\frac{\mathrm{d}N_s}{\mathrm{d}t}$$

which completes the proof.

6. Comparison with additive noise model

Our goal here is to establish an approximate correspondence between the place cell model and the additive noise measurement model more widely studied in other fields. This analysis gives new insights into the nature of the approximations made above and will allow us to convert the spacing and firing rate data into an equivalent signal to noise ratio.

The update rule for the continuous time situation associated with the model

$$\dot{x} = Ax + B\dot{w}; \qquad \dot{y} = RCx + \dot{v}$$

is

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{x} = A\hat{x} + \Sigma C^T(y - C\hat{x})$$

$$\dot{\Sigma} = A\Sigma + \Sigma A^T + BB^T - \Sigma C^T R^T RC\Sigma$$

In the case of the Kalman–Bucy filter the variance is not sample path dependent, whereas in the situation described in Theorem 2, it is. More specifically, we see that in the case of rate observations, the variance drops immediately after a pulse is received and rises between pulses. If Σ^{-1} is large relative to C^TC and/or the pulses occur frequently relative to the natural time constants of the A matrix, the variance is nearly constant and can be expected to nearly equal to the solution of

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{\Sigma} = A\Sigma + \Sigma A^T + BB^T - ((\Sigma^{-1} + C^TC)^{-1} - \Sigma)\lambda^*$$

To get to this approximation we have replaced the counting term in the variance equation by its average value. This suggests that the estimation problem of Theorem 2 could be approximately modeled by an estimation of the Gauss–Markov, additive white noise type provided we can find $R = R^T$ be such that

$$\Sigma CRC^T\Sigma = (\Sigma + Q)^{-1} - \Sigma$$

with Σ being the solution to the variance equation associated with the counting process.

Lemma. Assume that C is $m \times n$ and of rank m. Suppose that Σ_0 is the positive definite solution of

$$A\Sigma + \Sigma A^T + BB^T + ((\Sigma^{-1} + C^TC)^{-1} - \Sigma)\lambda^* = 0$$

If

$$R = \lambda^*((CC^T)^{-1}C(\Sigma_0 - (\Sigma_0 + \Sigma_0 C^TC\Sigma_0)^{-1})C^T(CC^T)^{-1})$$

then Σ_0 also satisfies

$$A\Sigma + \Sigma A^T + BB^T - \Sigma C^T RC\Sigma = 0$$

Proof. Let $F = F^T = (\Sigma_0^{-1} + CC^T)^{-1} - \Sigma_0$, so that

$$I - (\Sigma_0^{-1} + CC^T)\Sigma_0 = (\Sigma_0^{-1} + CC^T)F$$

This implies that $I - I - CC^T\Sigma_0 = (\Sigma_0^{-1} + CC^T)F$ and, thus

$$-(\Sigma_0^{-1} + CC^T)^{-1}CC^T\Sigma_0 = F = (\Sigma_0^{-1} + CC^T)^{-1} - \Sigma_0$$

Transposing we get

$$\Sigma_0 CC^T = (\Sigma_0^{-1} + CC^T)^{-1} = (\Sigma_0^{-1} + CC^T)^{-1} - \Sigma_0$$

Thus, we see that the range spaces of $(\Sigma_0^{-1} + CC^T)^{-1} - \Sigma_0$ and $\Sigma_0 CRC^T\Sigma_0$ agree, as do the null spaces. Because C is of full rank, this means that we can conclude that

$$R = C^{T\#}(\Sigma_0^{-1}(\Sigma^{-1} + C^TC)^{-1}\Sigma_0^{-1} - \Sigma^{-1})C^{\#}$$

where # denotes the Moore–Penrose inverse. The final form follows from some further matrix identities.

This can be viewed as a constructive process for finding an approximately equivalent white noise measurement problem. Because the matrix $\sqrt{R}C$ is a measure of the signal to noise ratio this also gives a way to associate an approximate signal to noise ratio with the model of Theorem 2.

Acknowledgements

This work was supported in part by the US Army Research Office under grants DAAH 04 96 1 0445 and DAAG 55 97 1 0114, and the National Science Foundation under grant CCR 9980058.

Appendix A. Stochastic differential equations

Our models involve differential equations with right-hand-sides that are stochastic. One type involves differential equations with the derivatives of counting processes appearing, such as

$$\dot{x} = f(x) + g(x)\frac{\mathrm{d}N}{\mathrm{d}t}$$

The meaning that we assign to such equations was discussed in the main text. Within a time interval over which $N(t)$ is constant, the evolution of x is governed by the ordinary deterministic equation $\dot{x} = f(x)$. If N has a jump of magnitude one at $t = \beta$, then $x(t)$ moves from its limiting value as t approaches β from below to the value $x(t_-) + g(x(t_-))$. In the stochastic differential equations literature it is common to use the notation

$$\mathrm{d}x = f(x)\mathrm{d}t + g(x)\mathrm{d}N$$

for this model and to refer to the solution we have described as the solution in the sense of Itô (Brémaud, 1981). Notice that when doing analysis with Itô equations it is necessary to use a special form of the chain rule. The standard rules of calculus are not directly applicable.

The situation for differential equations with the derivative of Brownian motion appearing also require some explanation. The equations we have written in the form

$$\dot{w} = f(x) = g(x)\dot{w}$$

are more commonly written as

$$\mathrm{d}x = f(x)\mathrm{d}t + g(x)\mathrm{d}w$$

and referred to as Itô equations. It is not possible to relate the trajectory of x to the trajectory of w in the direct way used above and we must refer the reader to the literature (e.g. Jazwinski, 1970) for an explanation.

Appendix B. Some identities

There are many ways of seeing that for $t > 0$ and $\alpha > 0$

$$\lim_{\alpha\to 0} \sum_{s\in\mathbb{Z}} \frac{\alpha}{\sqrt{2\pi}} \mathrm{e}^{-(y-\alpha s)^2/2} = 1; \qquad t > 0$$

One way, having the advantage of quantifying the rate, is to use the identity from the theory of theta functions, (e.g. Whittaker & Watson, 1963, p. 476) which states

$$\sum_{k\in\mathbb{Z}} \frac{1}{\sqrt{2\pi t}} \mathrm{e}^{-(x-ka)^2/2t} = \sum_{k\in\mathbb{Z}} \mathrm{e}^{-(2\pi k/a)^2 t} \cos(2\pi kx/a); \qquad t > 0$$

Let t be positive and take the limit as a goes to zero. All the terms on the right except the one corresponding to $k = 0$ go to zero and the $k = 0$ term contributes a 1.

If we take the derivative of this identify with respect to y, we get

$$\lim_{\alpha\to 0} \sum_{s\in\mathbb{Z}} \frac{\alpha(y - \alpha s)}{\sqrt{2\pi}} \mathrm{e}^{-(y-\alpha s)^2/2} = 0$$

Splitting this up we get

$$\lim_{\alpha\to 0} \sum_{s\in\mathbb{Z}} \frac{\alpha}{\sqrt{2\pi}} \mathrm{e}^{-(y-\alpha s)^2/2} = \lim_{\alpha\to 0} \sum_{s\in\mathbb{Z}} \frac{\alpha^2 s}{\sqrt{2\pi}} \mathrm{e}^{-(y-\alpha s)^2/2} = 0$$

and so

$$y = \lim_{\alpha\to 0} \sum_{s\in\mathbb{Z}} \frac{\alpha^2 s}{\sqrt{2\pi}} \mathrm{e}^{-(y-\alpha s)^2/2}$$

Now, consider a higher dimensional version. Recall the definition of $\mathbb{Z}^m$ as the set of all m-vectors in $\mathbb{R}^m$ that have integer components. Let C be an nxm matrix. For $s \in \mathbb{Z}^m$ define λ_s to be

$$\lambda_s(x) = \frac{a}{\sqrt{(2\pi)^m}} \mathrm{e}^{-(Cx-\alpha s)^T(Cx-\alpha s)/2}$$

Lemma 2. Let x belong to $\mathbb{R}^n$ and let C be an $m \times n$ matrix of rank m. Assume that α is positive. Then

$$\lim_{\alpha\to 0} \sum_{s\in\mathbb{Z}^m} \frac{\alpha^m}{\sqrt{(2\pi)^m}} \mathrm{e}^{-(Cx-\alpha s)^T(Cx-\alpha s)/2} = 1$$

and

$$\lim_{\alpha\to 0} \sum_{s\in\mathbb{Z}^m} \frac{\alpha^m \alpha s}{\sqrt{(2\pi)^m}} \mathrm{e}^{-(Cx-\alpha s)^T(Cx-\alpha s)/2} = Cx$$

Proof. Let $y = Cx$ and let y_i be the i-th component of y. Then

$$\lim_{\alpha\to 0} \sum_{s\in\mathbb{Z}^m} \frac{\alpha^m}{\sqrt{(2\pi)^m}} \mathrm{e}^{-(y-\alpha s)^T(y-\alpha s)/2}$$

$$= \lim_{\alpha\to 0} \prod_{i=1}^{m} \sum_{s\in\mathbb{Z}} \frac{\alpha}{\sqrt{(2\pi)}} \mathrm{e}^{-(y_i - s\alpha)^2/2}$$

but this is 1 by virtue of the identity proved above. For the second part, we proceed as above, taking the derivative with respect to y.

References

Boel, R., & Beněs, V. (1980). Recursive nonlinear estimation of a diffusion acting as the rate of an observed Poisson process. *IEEE Trans. Information Theory*, *IT-26*, 561–576.

Brémaud, P. (1981). *Point processes and queues*, New York: Springer-Verlag.

Brockett, R. W., & Clark, J. M. C. (1980). The geometry of the conditional density equation. In: O. L. R. Jacobs, M. H. A. Davis, M. H. A. Dempster, C. J. Harris & P. C. Parks, *Analysis and optimization of stochastic systems*, New York: Academic Press.

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., & Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.*, *18*, 7411–7425.

Davis, M. H. A., & Marcus, S. I. (1981). An introduction to nonlinear filtering. In M. Hazewinkel & J. C. Willems, *The mathematics of filtering and identification and applications*Dordrecht: Reidel.

Georgopolis, A., Kettner, R. E., & Schwartz, A. B. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. *II. Coding of the direction of movement by neuronal population. J. Neurosci.*, *8*, 2928–2937.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*, New York: Academic Press.

Kalman, R. E., & Bucy, R. S. (1961). New results in estimation theory. *ASME J. Basic Engineering*, *83*, 95–108.

Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *J. Neurophysiol.*, *26*, 2790–2793.

Twum-Danso, N. (1997). *Estimation, information, and neural signals*. PhD thesis, Division of Engineering and Applied Sciences, Harvard University.

Whittaker, E. T., & Watson, G. N. (1963). *A course in modern analysis*, . (4th ed.) Cambridge, UK: Cambridge University Press.

Zhang, K., Ginsburg, I., McNaughton, B., & Sejnowski, T. (1998). Interpreting neuronal activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.*, *79*, 1017–1044.