INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. **13**, 393–416 (1999)

DISCRETE TIME FILTERS FOR DOUBLY STOCHASTIC POISSON PROCESSES AND OTHER EXPONENTIAL NOISE MODELS

JONATHAN H. MANTON [1,*], VIKRAM KRISHNAMURTHY [1] AND ROBERT J. ELLIOTT [2]

[1] *Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052, Australia*
[2] *Department of Mathematical Sciences, University of Alberta, Edmonton, T6G 2G1, Canada*

SUMMARY

The well-known Kalman filter is the optimal filter for a linear Gaussian state-space model. Furthermore, the Kalman filter is one of the few known finite-dimensional filters. In search of other discrete-time finite-dimensional filters, this paper derives filters for general linear exponential state-space models, of which the Kalman filter is a special case. One particularly interesting model for which a finite-dimensional filter is found to exist is a doubly stochastic discrete-time Poisson process whose rate evolves as the square of the state of a linear Gaussian dynamical system. Such a model has wide applications in communications systems and queueing theory. Another filter, also with applications in communications systems, is derived for estimating the arrival times of a Poisson process based on negative exponentially delayed observations. Copyright © 1999 John Wiley & Sons, Ltd.

Key words: finite-dimensional filters; doubly stochastic Poisson models; linear Gaussian systems; exponential families

1. INTRODUCTION

Doubly stochastic Poisson processes were first introduced in 1955 by Cox.[1] A doubly stochastic Poisson process is, loosely speaking, a Poisson process whose rate is modulated by a second stochastic process.

Doubly stochastic Poisson processes are widely used in modelling communication systems.[2, 3] At the optical level, photons of light strike a photodetector with a Poisson distribution. The Poisson

This paper was recommended for publication by editor J. J. Shynk

* Correspondence to: Jonathan H. Manton, Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052, Australia. E-mail: jon@ee.mu.oz.au

Contract/grant sponsor: Australian Telecommunication and Engineering Research Board
Contract/grant sponsor: Cooperative Research Centre for Sensor Signal and Information Processing
Contract/grant sponsor: Australian Research Council
Contract/grant sponsor: NSERC
Contract/grant number: A7694

CCC 0890–6327/99/050393–24$17.50

Received 18 September 1997
Revised 18 December 1998
Accepted 19 February 1999

rate corresponds to the intensity of the light. Traffic arrival and departure in computer networks[4] are also amenable to modelling by doubly stochastic Poisson processes.

This paper primarily considers a discrete-time doubly stochastic Poisson process. The rate is determined by the square of the state of a discrete-time linear dynamical Gaussian system. The main result of this paper is Theorem 4, which presents a finite-dimensional filter for the square of the underlying state given the Poisson events. This result is interesting since there are very few known finite-dimensional filters. (This paper defines a finite-dimensional filter to be one where the filtered density can be completely characterized by a finite number of sufficient statistics. If furthermore the number of sufficient statistics is a constant, independent of the number of observations, the filter is said to be strictly finite-dimensional.)

The discrete-time case considered in this paper is worthy of consideration not only in its own right, but also as an approximation to the continuous-time case. Moreover, the continuous-valued Poisson rate can be used as an approximation in the case where the Poisson rate is discrete (for example a Markov chain). If the number of discrete states is very large, it is attractive to approximate discrete states by continuous-valued states.

Also presented is a sub-optimal filter (Section 6) which is based on approximating the filtered density by an Edgeworth series. Such an approximation is quite natural in this context. The sub-optimal filter is computationally inexpensive and performs satisfactorily in simulations.

1.1. Applications

Recent applications of doubly stochastic Poisson processes include the following. In Reference 5 a doubly stochastic Poisson process is used to model (and thus filter) the image of a faint object that appears to be moving in a random walk when viewed with an imager having a finite point-spread function, such as when viewing a planet through a telescope. Doubly stochastic Poisson processes have also been applied to modelling the surface reflectivity of SAR images,[6] modelling a quantum-limited optical DPSK receiver[7] and modelling network traffic.[8, 9]

The papers[10, 11] consider the estimation of Markov-modulated Poisson processes. As the number of discrete states in the Markov model increases, it is possible to approximate discrete states by continuous states, thus arriving at the models considered in the present paper.

The doubly stochastic Poisson process derived in this paper is currently being used to model rainfall data. Other applications of Poisson processes to meteorology include References 12–14.

In summary, the present paper extends research in this field by showing how to construct the optimal filter for a certain class of doubly stochastic Poisson processes. Moreover, the optimal filter is finite-dimensional. This allows the filter to be easily implemented in practice.

1.2. Related work—Poisson processes

A brief summary of other work in the literature relating to filtering of Poisson processes is now given. Continuous-time filtering results for a Poisson process whose rate evolves according to either a finite state or a diffusion Markov process have been widely studied, see References 2 and 15 for example. In Reference 15 stochastic differential equations are derived for the characteristic function of the filtered density. In general, going from this equation to an explicit formula for the filtered estimate is analytically intractible. The solution requires numerical integration, which can be computationally infeasible. An approximation to these differential equations is also derived in Reference 15 which leads to a sub-optimal filter.

More recently, continuous-time filters for doubly stochastic Poisson processes whose rate evolves according to a positive function of the state of a continuous-time linear Gaussian system are derived in Reference 16. It should be noted that the results of the continuous-time case in Reference 16 cannot be used to derive the discrete time results presented in the present paper.

Some filtering results for discrete time point processes are presented in Reference 17. These results are used to estimate the state of a random time-division multiple access computer network.

1.3. Related work—Finite-dimensional filters

The definition of a finite-dimensional filter, as given for example in Reference 18, is a filter whose filtered density always belongs to a parametrized set of density functions. The parametrized set is indexed by some subset of $\mathbb{R}^N$ — the smallest such N being referred to as the dimension of the filter. The present paper renames such a filter as a *strictly* finite-dimensional filter, to distinguish it from the case when the dimension, although finite, is allowed to increase as the number of observations increases.

The key result in the literature is that, under regularity conditions, a strictly finite-dimensional filter exists for a partially observable Markov process if and only if each of the conditional distributions involved form an exponential family of distributions.[18] Here, the "involved conditional distributions" include the filtered density too. This result was proved earlier in Reference 19 but under stronger regularity conditions.

The necessity of exponential families of distributions should come as no surprise. In Reference 20 (and later in Reference 21 under milder assumptions) it was shown that a sequence of i.i.d. random variables possesses a finite set of sufficient statistics if and only if the distribution of the random variables belongs to an exponential family.

1.4. Outline of approach

While the main contribution of this paper is the finite-dimensional filter for the doubly stochastic Poisson process, it is more illuminating to derive finite-dimensional filters for more general models and then consider the doubly stochastic Poisson process as a special case. In particular, the secondary aim of this paper is to show how classes of finite-dimensional filters may be found.

The Kalman filter is the optimal filter for the linear Gaussian state-space model:

$$x_{k+1} = A_{k+1}x_k + w_{k+1}, \qquad w_k \sim \text{i.i.d. } \mathrm{N}(0, Q_k) \tag{1}$$

$$z_k = C_k x_k + e_k, \qquad e_k \sim \text{i.i.d. } \mathrm{N}(0, \Gamma_k) \tag{2}$$

where $x_k \in \mathbb{R}^p$ is the state vector and $z_k \in \mathbb{R}^n$ the observation vector. The matrices $A_{k+1} \in \mathbb{R}^{p\times p}$ and $C_k \in \mathbb{R}^{n\times p}$ are deterministic. The state-space noise $w_k \in \mathbb{R}^p$ and the observation noise $e_k \in \mathbb{R}^n$ are zero-mean independent Gaussian random vectors with covariance matrices $Q_k \in \mathbb{R}^{p\times p}$ and $\Gamma_k \in \mathbb{R}^{n\times n}$ respectively.

A Gaussian distribution is a special example of an exponential family. Exponential families have a very special geometric significance.[22] A parameterized exponential family forms an affine space. (The converse is also true; if a parameterized family of distributions forms an affine space, it must be an exponential family.)

It is natural to ask if the various properties of the Kalman filter are specifically due to the Gaussian noise distribution, or instead are due to the affine nature of the Gaussian distribution. Therefore, this paper considers state-space models with exponential noise distributions. Further motivation for examining exponential noise models is based on the prominence of exponential families of distributions involved in a strictly finite-dimensional filter.[18]

The precise exponential noise models we will consider are detailed in Section 2. The evolution of the (unnormalized) filtered density of the exponential noise models is derived in Section 3. When applied to a doubly stochastic Poisson process, the filtered density can be computed analytically. Due to the practical importance of doubly stochastic Poisson processes, Section 4 is devoted to the derivation and application of the finite-dimensional filtering equations for such a process. Section 5 gives sufficient conditions for the filtered density equations of Section 3 to lead to finite-dimensional filters. Included in Section 5 is a finite-dimensional filter for the filtering of Poisson arrival times given only 'randomly delayed' observations. Section 6 introduces a general method of obtaining sub-optimal filters from the optimal filters of Section 5.

Notation. $\sigma\{x, y, \ldots\}$ denotes the smallest σ-algebra such that each of the random variables $x, y, \ldots$ are Borel-measurable. If ϕ_k is a vector, then its ith element is denoted by $(\phi_k)_i$. $\binom{y}{x}$ is the binomial symbol. $\wedge$ denotes minimum, i.e. $x \wedge y = \min\{x, y\}$.

2. SIGNAL MODEL

After introducing exponential families of distributions, the exponential noise models studied in this paper are expounded.

1. *Exponential family*: Following [Reference 23, Section 1.4], a family $\{P_\theta\}$ of distributions is said to form an s-parameter exponential family if the distributions P_θ have densities of the form

$$p_\theta(z) = \exp\left[\sum_{i=1}^{s} \theta_i T_i(z) - R(\theta_1, \ldots, \theta_s)\right] h(z), \quad z \in \mathbb{R}^n \tag{3}$$

with respect to some carrier measure μ. Here, $\theta_i \in \mathbb{R}$ are the canonical parameters, while $T_i : \mathbb{R}^n \to \mathbb{R}$, $R : \mathbb{R}^s \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$ are Borel-measurable functions. In this paper, the carrier measure μ is the Lebesgue measure for a continuous distribution, or the counting measure for a discrete distribution. Some common one and two parameter exponential families include the Gaussian, Gamma, Chi-square and Beta continuous distributions and the Binomial, Poisson and Negative Binomial discrete distributions. Two examples are now given.

2. *Gaussian*: The Gaussian density $\mathrm{N}(\mu, \sigma^2)$

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right], \quad z \in \mathbb{R} \tag{4}$$

can be written as (cf., (3)) $p(z) = \exp[\theta_1 T_1(z) + \theta_2 T_2(z) - R(\theta_1, \theta_2)] h(z)$ where

$$\theta_1 = -\frac{1}{2\sigma^2}, \ \theta_2 = \frac{\mu}{\sigma^2}, \ T_1(z) = z^2, \ T_2(z) = z, \ R(\theta_1, \theta_2) = -\frac{\theta_2^2}{4\theta_1} - \frac{1}{2}\log(-\theta_1), \ h(z) = \frac{1}{\sqrt{\pi}}. \tag{5}$$

Notice that the canonical parameters (θ_1, θ_2) are not the 'usual' parameters (μ, σ^2).

3. *Poisson*: The Poisson density with rate $\lambda \in \mathbb{R}$, $\lambda > 0$ is

$$p(z) = \frac{\lambda^z}{z!} \exp[-\lambda], \quad z = 0, 1, 2, \ldots \tag{6}$$

This is equivalent to (3) with $\theta_1 = \log \lambda$, $T_1(z) = z$, $R(\theta_1) = \exp[\theta_1]$, $h(z) = 1/z!$.

The signal model examined in this paper is now stated.

Signal model: For clarity of presentation both the state vector and the observation vector are assumed to be one-dimensional, i.e., scalars. The exponential noise model studied in this paper is a generalisation of the linear Gaussian state-space model (1), (2). The underlying probability space is $(\Omega, \mathscr{F}, \mathscr{P})$. The event space $\Omega = (\mathbb{R} \times \mathbb{R})^\infty$ contains elements of the form $\{x_0, z_0, x_1, z_1, \ldots\} \in \Omega$. (The random variables $x_k \in \mathbb{R}$, $k = 0, 1, \ldots$ are referred to as the state, while the $z_k \in \mathbb{R}$, $k = 0, 1, \ldots$ are the observations.) The σ-algebra $\mathscr{F}$ is $\mathscr{F} = \sigma\{x_0, z_0, x_1, z_1, \ldots\}$. The probability measure $\mathscr{P}$ is the measure† which gives the random variable w_k defined by (7) the density (8) and the random variable z_k (defined above) the conditional density (9). In particular, the state equation

$$x_{k+1} = A_{k+1} x_k + w_{k+1}, \quad x_0 = w_0 \tag{7}$$

(where $A_{k+1} \in \mathbb{R}$ is a known scalar) defines the random variable $w_{k+1} : \Omega \to \mathbb{R}$. The density of w_k, denoted by $\Psi^w(w_k)$, belongs to a q-parameter independent continuous exponential family, i.e.

$$\Psi^{(w)}(w_k; \phi_k) = \exp\left[\sum_{i=1}^{q} (\phi_k)_i \; T_i^{(w)}(w_k) - R^w(\phi_k)\right] h^{(w)}(w_k) \tag{8}$$

where $\phi_k \in \mathbb{R}^q$ is a known parameter and $T_i^{(w)}$, $R^{(w)}$ and $h^{(w)}$ are Borel-measurable functions (see (3)). (Since w_k is continuous, the carrier measure is the Lebesgue measure.) The observation z_k has an s-parameter independent exponential density

$$\Psi^{(z)}(z_k; \boldsymbol{\theta}_k) = \exp\left[\sum_{i=1}^{s} (\boldsymbol{\theta}_k)_i T_i^{(z)}(z_k) - R^{(z)}(\boldsymbol{\theta}_k)\right] h^{(z)}(z_k) \tag{9}$$

conditioned on the parameter $\boldsymbol{\theta}_k \in \mathbb{R}^s$ given by

$$\boldsymbol{\theta}_k = r_k(x_k) \tag{10}$$

Here, $r_k : \mathbb{R} \to \mathbb{R}^s$, $k = 0, 1, \ldots,$ are Borel-measurable functions. Note that the carrier measure for the observation density (9) is the Lebesgue measure if z_k is a continuous random variable, or the counting measure if z_k is a discrete random variable.

Definition

Let $\mathscr{Z}_k$ denote the observation history and $\mathscr{G}_k$ the complete history, i.e.

$$\mathscr{Z}_k = \sigma\{z_0, \ldots, z_k\} \tag{11}$$

$$\mathscr{G}_k = \sigma\{x_0, z_0, \ldots, x_k, z_k\} \tag{12}$$

† Existence and uniqueness up to sets of Lebesgue-measure zero follow from Kolmogorov's theorem.

Expectation with respect to the measure $\mathscr{P}$ is denoted by $\mathbf{E}[\cdot]$. In Section 3 a new measure $\bar{\mathscr{P}}$ is introduced. Expectation with respect to $\bar{\mathscr{P}}$ is denoted by $\bar{\mathbf{E}}[\cdot]$.

Aim: Given the observations $\{z_0, \ldots, z_k\}$ the aim is to estimate (some measurable function $g(x_k)$ of) the state x_k. More precisely, the aim is to calculate the optimal (minimum mean-square error) estimator $\mathbf{E}[g(x_k) \mid \mathscr{Z}_k]$.

3. DERIVATION OF UNNORMALIZED FILTERED DENSITY

In this section the unnormalized filtered density for the state x_k given the observations $\{z_0, \ldots, z_k\}$ is derived. Standard techniques based on the reference probability method[24] are used.

On the probability space $(\Omega, \mathscr{F}, \mathscr{P})$ (defined in Section 2) define the new probability measure $\bar{\mathscr{P}}$ by

$$\left.\frac{\mathrm{d}\mathscr{P}}{\mathrm{d}\bar{\mathscr{P}}}\right|_{\mathscr{G}_k} = \Lambda_k = \prod_{i=0}^{k} \lambda_i \tag{13}$$

$$\lambda_k = \exp\left[\sum_{i=1}^{q} (\phi_k)_i T_i^{(w)}(x_k - A_k x_{k-1}) - \sum_{i=1}^{q} (\bar{\phi})_i T_i^{(w)}(x_k) - R^{(w)}(\phi_k) + R^{(w)}(\bar{\phi})\right]$$
$$\times \exp\left[\sum_{i=1}^{s} (\boldsymbol{\theta}_k - \bar{\boldsymbol{\theta}})_i T_i^{(z)}(z_k) - R^{(z)}(\boldsymbol{\theta}_k) + R^{(z)}(\bar{\boldsymbol{\theta}})\right] \frac{h^{(w)}(x_k - A_k x_{k-1})}{h^{(w)}(x_k)}, \quad k = 1, 2, \ldots \tag{14}$$

$$\lambda_0 = \exp\left[\sum_{i=1}^{q} (\phi_0 - \bar{\phi})_i T_i^{(w)}(x_0) - R^{(w)}(\phi_0) + R^{(w)}(\bar{\phi})\right]$$
$$\times \exp\left[\sum_{i=1}^{s} (\boldsymbol{\theta}_0 - \bar{\boldsymbol{\theta}})_i T_i^{(z)}(z_0) - R^{(z)}(\boldsymbol{\theta}_0) + R^{(z)}(\bar{\boldsymbol{\theta}})\right] \tag{15}$$

where $\bar{\boldsymbol{\theta}} \in \mathbb{R}^s$ and $\bar{\phi} \in \mathbb{R}^q$ are constant vectors and $\boldsymbol{\theta}_k$ is defined in (10). Then Girsanov's theorem[24] implies that under $\bar{\mathscr{P}}$, x_k and z_k are i.i.d.; x_k has density $\Psi^{(w)}(x_k; \bar{\phi})$ where $\Psi^{(w)}$ is defined in (8); and z_k has density $\Psi^{(z)}(z_k; \bar{\boldsymbol{\theta}})$ where $\Psi^{(z)}$ is defined in (9).

Remark

The above holds irrespective of whether z_k is a continuous or a discrete random variable. $\mathscr{P}$ denotes the real-world probability, whereas working under $\bar{\mathscr{P}}$ is very convenient. The filtered density is derived under $\bar{\mathscr{P}}$ and then mapped back to $\mathscr{P}$. This mapping is done using an abstract version of Bayes' theorem[24] as follows. For any Borel-measurable test function g,

$$\mathbf{E}[g(x_k) \mid \mathscr{Z}_k] = \frac{\bar{\mathbf{E}}[\Lambda_k g(x_k) \mid \mathscr{Z}_k]}{\bar{\mathbf{E}}[\Lambda_k \mid \mathscr{Z}_k]} \tag{16}$$

The unnormalized filtered density at time k, denoted $q_k(x)$, is now formally defined.

Definition

Define the unnormalized filtered density $q_k(x)$, for x_k given $\{z_0, \ldots, z_k\}$ implicitly by

$$\bar{\mathbf{E}}[\Lambda_k g(x_k) \mid \mathscr{Z}_k] = \int_{\mathbb{R}} g(x) q_k(x) \, \mathrm{d}x \tag{17}$$

for any Borel-measurable test function g.

The unnormalized filtered density for the signal model is stated in Lemma 3.

Lemma 1 (*Unnormalized filtered density*)

The unnormalized filtered density q_k defined in (17) for the exponential noise model defined in Section 2 is recursively given by

$$q_{k+1}(y) = K_{k+1} \exp\left[\sum_{i=1}^{s} (r_{k+1}(y))_i T_i^{(z)}(z_{k+1}) - R^{(z)}(r_{k+1}(y))\right] \tag{18}$$

$$\times \int_{\mathbb{R}} \exp\left[\sum_{i=1}^{q} (\phi_{k+1})_i T_i^{(w)}(y - A_{k+1}x)\right] h^{(w)}(y - A_{k+1}x) q_k(x) \, \mathrm{d}x, \quad k = 1, 2, \ldots$$

$$q_0(y) = K_0 \exp\left[\sum_{i=1}^{s} (r_0(y))_i T_i^{(z)}(z_0) - R^{(z)}(r_0(y)) + \sum_{i=1}^{q} (\phi_0)_i T_i^{(w)}(y)\right] h^{(w)}(y) \tag{19}$$

where $K_k \in \mathbb{R}$, $k = 0, 1, \ldots$ are independent of y.

Proof: Define the function $\lambda_k(x, y)$ to be the right-hand side of (14) with x_{k-1} replaced by x and x_k by y (i.e. $\lambda_k \equiv \lambda_k(x_{k-1}, x_k)$). Similarly $\lambda_0(x)$ denotes the right-hand side of (15) with x_0 replaced by x. The recursive update of the unnormalized filtered density is derived as follows:

$$\bar{\mathbf{E}}[\Lambda_{k+1} g(x_{k+1}) \mid \mathscr{Z}_{k+1}] = \bar{\mathbf{E}}\left[\Lambda_k \int_{\mathbb{R}} \lambda_{k+1}(x_k, y) \, g(y) \, \Psi^{(w)}(y; \bar{\phi}) \, \mathrm{d}y \mid \mathscr{Z}_{k+1}\right] \tag{20}$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \lambda_{k+1}(x, y) \, g(y) \, \Psi^{(w)}(y; \bar{\phi}) \, \mathrm{d}y \right) q_k(x) \, \mathrm{d}x \tag{21}$$

$$= \int_{\mathbb{R}} g(y) \left(\int_{\mathbb{R}} \lambda_{k+1}(x, y) \, \Psi^{(w)}(y; \bar{\phi}) \, q_k(x) \, \mathrm{d}x \right) \mathrm{d}y \tag{22}$$

Here, (21) follows from (20) by (17) and the fact that, under $\bar{\mathscr{P}}, z_{k+1}$ is independent of $\{z_0, \ldots, z_k, x_k\}$. Comparing (22) with (17) shows $q_{k+1}(y) = \int_{\mathbb{R}} \lambda_{k+1}(x, y) \, \Psi^{(w)}(y; \bar{\phi}) \, q_k(x) \, \mathrm{d}x$, and (18) follows by expanding the integrand.

The initial $(k = 0)$ unnormalized filtered density is derived thus:

$$\bar{E}[\Lambda_0 g(x_o) \mid \mathscr{Z}_0] = \int_{\mathbb{R}} g(x) (\lambda_0(x) \, \Psi^{(w)}(x; \bar{\phi})) \, \mathrm{d}x \tag{23}$$

Comparing with (17) shows $q_0(x) = \lambda_0(x) \, \Psi^{(w)}(x; \bar{\phi})$, and (19) follows upon expansion. □

4. DISCRETE-TIME FILTER FOR A DOUBLY STOCHASTIC POISSON PROCESS

This section applies the unnormalized filtered density derived in Section 3 to a doubly stochastic Poisson process. The reasons for the specific form of the doubly stochastic Poisson process chosen, along with generalizations of the model, are given in further detail in Section 5.

Signal model (*Doubly stochastic Poisson process*)

The underlying process $x_k \in \mathbb{R}$ is a scalar linear dynamical system of the form

$$x_{k+1} = A_{k+1}x_k + w_{k+1}, \quad w_k \sim \text{i.i.d.}\,\mathrm{N}(0,\sigma_k^2) \tag{24}$$

$$x_0 = w_0, \quad w_0 \sim \text{i.i.d.}\,\mathrm{N}(0,\sigma_0^2) \tag{25}$$

The noise $w_k \in \mathbb{R}$ is a scalar-independent Gaussian sequence with zero mean and variance σ_k^2. The parameters $A_k \in \mathbb{R}$ and $\sigma_k^2 \in \mathbb{R}$ are deterministic and assumed to be known. The process x_k is not observed directly. Instead, it is used to modulate a Poisson process with rate $(c_k x_k)^2$, where $c_k \in \mathbb{R}$ is a known deterministic parameter. The choice of $(c_k x_k)^2$ ensures that the rate is non-negative and that the filtered density has the simple form of a polynomial times a Gaussian (see Section 5). The observations $\{z_0,\ldots,z_k\}$ thus have the following independent Poisson density:

$$\Psi^{(z)}(z_k;x_k) = \frac{(c_k x_k)^{2z_k}}{z_k!}\exp[-(c_k x_k)^2] \tag{26}$$

Aim: Given the observations $\{z_0,\ldots,z_k\}$ the aim is to derive an optimal recursive filter for the rate of the Poisson process, i.e., the aim is to compute $\mathbf{E}[x_k^2 \mid \mathscr{Z}_k]$. Note that $\mathbf{E}[x_k \mid \mathscr{Z}_k] = 0$ because x_0 is distributed symmetrically about the origin and the observations give no information about the sign of x_k, only its magnitude squared.

The optimal recursive filter is presented in Theorem 1 below.

Theorem 1 (*Optimal finite-dimensional filter*)

At time k, the unnormalized filtered density $q_k(x)$ defined in (17) for the doubly stochastic Poisson process defined above can be expressed as

$$q_k(x) = \sum_{t=0}^{L_k} P_k(t)\,x^t \exp\left[-\frac{x^2}{2\Omega_k}\right] \tag{27}$$

Thus, $q_k(x)$ is completely characterized by the $L_k + 1$ sufficient statistics $P_k(t) \in \mathbb{R}$, $t = 0,\ldots,L_k$, and the parameter $\Omega_k \in \mathbb{R}$. These statistics are recursively computed by

$$L_{k+1} = L_k + 2z_{k+1}, \qquad L_o = 2z_0 \tag{28}$$

$$\Omega_{k+1}^{-1} = 2c_{k+1}^2 + (\sigma_{k+1}^2 + \Omega_k A_{k+1}^2)^{-1}, \qquad \Omega_0^{-1} = 2c_0^2 + \sigma_0^{-2} \tag{29}$$

$$P_{k+1}(t + 2z_{k+1}) = \sum_{\alpha=t}^{L_k} Q_\alpha(\alpha - t)P_k(\alpha), \qquad P_0(2z_0) = 1 \tag{30}$$

where t ranges from 0 to L_k in (30), while $P_k(m)=0$ for $m<2z_k$. $Q_\alpha(m)$ in (30) is zero for m odd, and for m even, is defined recursively:

$$Q_\alpha(m)=\begin{cases} A_{k+1}^{\alpha}(\sigma_{k+1}^2\Omega_k^{-1}+A_{k+1}^2)^{-\alpha}, & m=0 \\ Q_\alpha(m-2)\,\frac{(\alpha-m+2)(\alpha-m+1)}{m}\,\sigma_{k+1}^2\,(1+\frac{\sigma_{k+1}^2}{A_{k+1}^2}\Omega_k^{-1}), & m>0 \text{ and even} \end{cases} \tag{31}$$

Furthermore, the filtered estimate of the Poisson rate is given by

$$\mathbf{E}[x_k^2 \mid \mathscr{Y}_k]=\frac{\sum_{t=0}^{L_k} P_k(t)\,S_k(t+2)}{\sum_{t=0}^{L_k} P_k(t)\,S_k(t)} \tag{32}$$

where $S_k(t)$ is zero for t odd, and for t even, is defined recursively:

$$S_k(t)=\begin{cases} 1, & t=0 \\ (t-1)\,\Omega_k\,S_k(t-2), & t>0 \text{ and even} \end{cases} \tag{33}$$

Proof: The doubly stochastic Poisson process (24)–(26) is a special case of the exponential noise model (7)–(10), where

$$\phi_k=\frac{-1}{2\sigma_k^2},\qquad T_1^{(w)}(x)=x^2,\quad R^{(w)}(\phi)=-\frac{1}{2}\log(-\phi),\quad h^{(w)}(x)=\frac{1}{\sqrt{\pi}} \tag{34}$$

$$r_k(x_k)=\log(c_kx_k)^2,\quad T_1^{(z)}(z)=z,\qquad R^{(z)}(\theta)=\exp[\theta],\qquad h^{(z)}(z)=\frac{1}{z!} \tag{35}$$

Note that $\boldsymbol{\theta}_k,\phi_k\in\mathbb{R}$ are scalars (i.e., $q=1$ and $s=1$ in (8) and (9) respectively).

The unnormalized filtered density is given by substituting (34)–(35) into (18) and (19). In particular, from (19),

$$q_0(y)=\frac{1}{\sqrt{\pi}}K_0\,c_0^{2z_0}\,y^{2z_0}\exp\left[-\left(c_0^2+\frac{1}{2\sigma_0^2}\right)y^2\right] \tag{36}$$

This is in the form (27) by defining $L_0=2z_0$, $P_0(2z_0)=1$ and $\Omega_0^{-1}=2c_0^2+\sigma_0^{-2}$ where the multiplicative constant $(1/\sqrt{\pi})K_0\,c_0^{2z_0}$ has been omitted since only the *unnormalized* density is of importance.

Using (18) it is now shown that if

$$q_k(x)=x^{\alpha}\exp\left[-\frac{x^2}{2\Omega_k}\right] \tag{37}$$

then

$$q_{k+1}(y)=\frac{1}{\sqrt{\pi}}K_{k+1}\,c_{k+1}^{2z_{k+1}}y^{2z_{k+1}}\left(\sqrt{2\pi}\sum_{m=0}^{\alpha}Q_\alpha(m)\,y^{\alpha-m}\right)\exp\left[-\frac{y^2}{2\Omega_{k+1}}\right] \tag{38}$$

where Ω_{k+1} and Q_α are defined in (29) and (31), respectively. Substituting (34)–(35) and (37) into (18) gives, after rearrangement

$$q_{k+1}(y) = \frac{1}{\sqrt{\pi}} K_{k+1} c_{k+1}^{2z_{k+1}} y^{2z_{k+1}} \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_{k+1}^2 + A_{k+1}^2\Omega_k} + 2c_{k+1}^2\right) y^2\right]$$
$$\times \int_{\mathbb{R}} x^\alpha \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x \tag{39}$$

where $\mu = (A_{k+1}/(\sigma_{k+1}^2\Omega_k^{-1} + A_{k+1}^2))y$ and $\sigma^2 = (1/\Omega_k + (A_{k+1}^2/\sigma_{k+1}^2))^{-1}$. To show (39) is identical to (38) requires proving

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^\alpha \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x = \sum_{m=0}^{\alpha} Q_\alpha(m) y^{\alpha-m} \tag{40}$$

Indeed,

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^\alpha \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\sigma x + \mu)^\alpha \exp\left[-\frac{1}{2}x^2\right] \mathrm{d}x \tag{41}$$

$$= \sum_{m=0}^{\alpha} \binom{\alpha}{m} \mu^{\alpha-m}\sigma^m \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^m \exp\left[-\frac{1}{2}x^2\right] \mathrm{d}x \tag{42}$$

Comparing (40) with (42) shows that

$$Q_\alpha(m) = \binom{\alpha}{m} \left(\frac{\mu}{y}\right)^{\alpha-m} \sigma^m \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^m \exp\left[-\frac{1}{2}x^2\right] \mathrm{d}x \tag{43}$$

The odd moments of a Gaussian distribution are zero, while the even moments are given by

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^m \exp\left[-\frac{1}{2}x^2\right] \mathrm{d}x = \begin{cases} 1, & m = 0 \\ 1 \cdot 3 \cdots (m-1), & m > 0 \text{ and even} \end{cases} \tag{44}$$

Therefore,

$$\frac{Q_\alpha(m)}{Q_\alpha(m-2)} = \frac{(\alpha - m + 2)(\alpha - m + 1)}{m} \left(\frac{\mu}{y}\right)^{-2} \sigma^2 \tag{45}$$

which agrees with (31).

Therefore, if $q_k(x)$ is given by (27), then from (38) it follows that

$$q_{k+1}(y) = \sqrt{2}\,K_{k+1} c_{k+1}^{2z_{k+1}} y^{2z_{k+1}} \sum_{\alpha=0}^{L_k} P_k(\alpha) \sum_{m=0}^{\alpha} Q_\alpha(m) y^{\alpha-m} \exp\left[-\frac{y^2}{2\Omega_{k+1}}\right] \tag{46}$$

from which (30) follows straightforwardly.

Lastly, the filtered estimate (32) is derived. Combining (16), (17) and (27) gives

$$\mathbf{E}[x_k^2 \mid \mathscr{Y}_k] = \frac{\int_{\mathbb{R}} x^2 \sum_{t=0}^{L_k} P_k(t) x^t \exp[-\frac{x^2}{2\Omega_k}]\,\mathrm{d}x}{\int_{\mathbb{R}} \sum_{t=0}^{L_k} P_k(t) x^t \exp[-\frac{x^2}{2\Omega_k}]\,\mathrm{d}x} = \frac{\sum_{t=0}^{L_k} P_k(t) S_k(t+2)}{\sum_{t=0}^{L_k} P_k(t) S_k(t)} \tag{47}$$

where $S_k(t) = (1/\sqrt{2\pi\Omega_k}) \int_{\mathbb{R}} x^t \exp[-x^2/2\Omega_k]\,\mathrm{d}x$. Applying integration by parts shows that $S_k(t)$ can be calculated by (33). □

Implementation details and computational complexity

1. *Computational complexity*: The computational complexity of filtering a block of observations $\{z_0, \ldots, z_k\}$ is $\mathrm{O}(kN^2)$ where $N = z_0 + \cdots + z_k$ and k denotes time.
2. The parameter Ω_k in (29) is independent of the data and can be computed off-line.
3. Only the coefficients of the even powers of x in the density function (27) need be calculated (cf. (30)), since $P_k(t)$ is zero when t is odd.
4. When implementing the filter, it is necessary to scale $P_k(t)$ to avoid numerical underflow or overflow.

5. OTHER FINITE-DIMENSIONAL FILTERS

This section gives sufficient conditions for a finite-dimensional filter to exist. The key idea is to investigate filtered densities which satisfy a closedness condition, defined in Section 5.1. Two forms of the filtered density which satisfy the closedness condition are subsequently considered. The first form is a polynomial times a Gaussian, considered in Section 5.2. Section 5.3 considers a filtered density having the form of a Gaussian mixture (i.e. a weighted sum of Gaussian distributions). Both Sections 5.2 and 5.3 give conditions (stated as constraints on the functions $T_i^{(w)}$, $T_i^{(z)}$, $R^{(w)}$, $R^{(z)}$, $h^{(w)}$, $h^{(z)}$ and r_k defined in Section 2) which ensure the existence of a finite-dimensional filter.

Given a specific filtering problem, it is straightforward to check if the sufficient conditions in Sections 5.2 or 5.3 are satisfied. If they are, then an explicit filtering algorithm can be readily derived from the unnormalized filtered density (18), (19) in an equivalent way to the doubly stochastic Poisson filter derivation in Section 4.

Lastly, Section 5.4 presents a finite-dimensional filter for the arrival times of a Poisson process. It shows how one-sided noise distributions (i.e. noise which is nonnegative) can also lead to finite-dimensional filters.

5.1. Closed classes of filtered densities

The exponential noise model defined in Section 2 contains the following arbitrary functions: $T_i^{(w)}$, $T_i^{(z)}$, $R^{(w)}$, $R^{(z)}, h^{(w)}$, $h^{(z)}$ and r_k. The form of these functions determines whether or not the integrals in (18) and (19) can be evaluated analytically. Lemma 2, combined with Definition 1 below, gives sufficient and easily verifiable conditions for a finite-dimensional filter to exist.

Definition 1 (*Closedness of filtered density*)

Let $\mathscr{C}$ denote a class of unnormalized probability densities. Define $\mathscr{C}$ to be *closed* if it satisfies the following constraints:

1. If $P(x) \in \mathscr{C}$ and $\lambda \in \mathbb{R}^+$ then $\lambda P(x) \in \mathscr{C}$.
2. If $P_1(x), P_2(x) \in \mathscr{C}$ then $P_1(x)P_2(x) \in \mathscr{C}$.
3. If $P_1(x), P_2(x) \in \mathscr{C}$ and $a \in \mathbb{R}$ then $P(y) = \int_{\mathbb{R}} P_1(y - ax)P_2(x)\,\mathrm{d}x$ exists, can be computed analytically, and furthermore $P(y) \in \mathscr{C}$.

The next lemma shows the relevance of classes $\mathscr{C}$ which are closed.

Lemma 2

The (unnormalized) filtered density $q_k(x)$, defined in Lemma 1, will always be an element of $\mathscr{C}$ if:

1. $\mathscr{C}$ is closed (see Definition 1).
2. For all $k \geqslant 0, \phi \in \mathbb{R}$ and $i = 1, \ldots, s$,

$$\exp[\phi\,(r_k(y))_i] \in \mathscr{C}, \quad \exp[-R^{(z)}(r_k(y))] \in \mathscr{C}, \quad \exp[\phi\, T_i^{(w)}(y)] \in \mathscr{C}, \quad h^{(w)}(y) \in \mathscr{C}. \tag{48}$$

Proof: Take $\mathscr{C}$ to be closed. Then the conditions (48) ensure that (18) has the form $q_{k+1}(y) = P_1(y)\int_{\mathbb{R}} P_2(y - A_{k+1}x)q_k(x)\,\mathrm{d}x$, where $P_1, P_2 \in \mathscr{C}$. Therefore, if $q_k(x) \in \mathscr{C}$, then $q_{k+1}(x) \in \mathscr{C}$ too. The conditions (48) are sufficient to ensure $q_0(x) \in \mathscr{C}$ (see (19)). □

5.2. Polynomial times Gaussian filtered density

The filtered density for the linear Gaussian state-space model (1), (2) is well-known to be a Gaussian distribution. (It is easily verified that the family of (unnormalized) Gaussian distributions is *closed* in the sense of Definition 1.) A more general class of density functions‡ is defined by

$$\mathscr{C} = \left\{ q(x) : q(x) = \sum_{j=0}^{N} a_j x^j \exp[bx^2 + cx], \;\; a_j, b, c \in \mathbb{R}, \;\; b < 0 \right\} \tag{49}$$

Such a class is referred to as the class of polynomials times a Gaussian (PTG). It is readily verified that this class is closed (see Definition 1). Therefore, Lemma 2 shows that, under certain conditions on $T_i^{(w)}$, $R^{(z)}$, $h^{(w)}$ and r_k, the filtered density will always belong to the class of PTGs. These conditions are stated explicitly in the following lemma.

Lemma 3 (*Sufficient conditions*)

If the restrictions below are imposed on the signal model defined in Section 2, the filtered density will belong to the class $\mathscr{C}$ defined in (49), and is said to have the form of a PTG, a

‡ The class also includes functions which are not density functions. However, this causes no problems, since the filtering equations (18), (19) themselves ensure that the filtered density is an unnormalized density function (i.e. non-negative and integrable)

polynomial times a Gaussian. (In the table below, a — denotes no restriction. $R^{(z)} \circ r_k$ denotes the composite function, i.e. $R^{(z)}(r_k(x))$.)

$T_i^{(w)}$	quadratic	$T_i^{(z)}$	—
$R^{(w)}$	—	$R^{(z)}$	see $R^{(z)} \circ r_k$
$h^{(w)}$	polynomial $\times$ exp(quadratic)	$h^{(z)}$	—
$(r_k)_i$	quadratic + log(polynomial)	$R^{(z)} \circ r_k$	quadratic + log(polynomial)

Proof: Follows immediately from Lemma 2 and the fact that the class of PTGs is closed. □

Remark

Although $T_i^{(z)}$ and $h^{(z)}$ are marked as arbitrary, indirectly they are restricted by the restrictions on $R^{(z)}$. Specifically, because $\int_{\mathbb{R}} \Psi^{(z)}(z;\theta)\,\mathrm{d}z = 1$, $R^{(z)}$ is given by§

$$R^{(z)}(\theta) = \log \int_{\mathbb{R}} \exp\left[\sum_{i=1}^{s} \theta_i T_i^{(z)}(z)\right] h^{(z)}(z)\,\mathrm{d}z \tag{50}$$

Models which satisfy the sufficient conditions in Lemma 3, yet have not been introduced into the literature previously, include the following.

Doubly stochastic Poisson process: The doubly stochastic Poisson process has a finite-dimensional filter, as derived in Section 4.

Doubly stochastic binomial process: The canonical parameter (see Section 2) of a binomial distribution having success probability p is $\log p/(1-p)$. Rather than Poisson observations with rate $(c_k x_k)^2$ (see Section 4), a finite-dimensional filter also exists for Binomial observations with canonical parameter $\boldsymbol{\theta}_k = \log p/(1-p) = \log x_k^2$.

There are myriad other noise densities which satisfy Lemma 3 yet do not have a name. Furthermore, models which do not meet the criteria directly can be approximated by models which do. For example, an Edgeworth series (a polynomial times a Gaussian) can be used to approximate the state-space noise density (8). Because most distributions can be approximated by an Edgeworth series, $\Psi^{(w)}$ can be quite general.

5.3. *Gaussian mixture filtered density*

Another interesting form the filtered density can take is that of a weighted sum of Gaussian densities, referred to as a Gaussian mixture. In Reference 25, all probability densities are approximated by Gaussian mixtures, thus forcing the filtered density to also have the same form. In the same way, Lemma 3 used the recursive update equations (18) and (19) to determine sufficient conditions for the filtered density to be a PTG (polynomial times a Gaussian), conditions

§ For convenience, z_k is assumed to be continuous. If it is a discrete random variable, the integral can be replaced by a suitable summation.

on $T_i^{(w)}$, $T_i^{(z)}$, $R^{(w)}$, $R^{(z)}$, $h^{(w)}$, $h^{(z)}$ and r_k can likewise be given to ensure the filtered density be a Gaussian mixture.

The class of Gaussian mixtures is formally defined by

$$\mathscr{C} = \left\{ q(x) : q(x) = \sum_{j=1}^{n} a_j \exp[b_j x^2 + c_j x], \; a_j, b_j, c_j \in \mathbb{R}, \; a_j > 0, \; b_j < 0 \right\} \tag{51}$$

The class of Gaussian mixtures is closed (see Definition 1). The next lemma gives conditions for the filtered density to be a Gaussian mixture.

Lemma 4 (*Sufficient conditions*)

If the restrictions below are imposed on the signal model defined in Section 2, the filtered density will belong to the class $\mathscr{C}$ defined in (51), and is said to have the form of a Gaussian mixture. (In the table below, a — denotes no restriction. $R^{(z)} \circ r_k$ denotes the composite function, i.e. $R^{(z)}(r_k(x))$.)

$T_i^{(w)}$	quadratic	$T_i^{(z)}$	—
$R^{(w)}$	—	$R^{(z)}$	see $R^{(z)} \circ r_k$
$h^{(w)}$	Gaussian mixture	$h^{(z)}$	—
$(r_k)_i$	quadratic	$R^{(z)} \circ r_k$	quadratic

Proof: Follows immediately from Lemma 2 and the fact that the class of Gaussian mixtures is closed. □

Remark

Although $T_i^{(z)}$ and $h^{(z)}$ are marked as arbitrary, indirectly they are restricted by the restrictions on $R^{(z)}$. See (50).

5.4. *Poisson arrival time filter*

Up until now, only filtered densities with $\exp[-x^2]$ terms have been considered in the present paper. The Poisson arrival time filter presented in this section arises from the consideration of filtered densities with $\exp[-x]$ terms.

After stating the signal model, some intended applications of the Poisson arrival time filter are discussed. The general form of the filtered density is then given. Lastly, an algorithm for the resulting finite-dimensional filter is presented.

Signal model (*Poisson arrival times*)

The underlying process $x_k \in \mathbb{R}$ is a scalar linear dynamical system of the form

$$x_{k+1} = A_{k+1} x_k + w_{k+1}, \quad x_0 = w_0, \quad A_{k+1} > 0 \tag{52}$$

$$\Psi^{(w)}(w_k) = \frac{1}{a_k}\exp\left[-\frac{w_k}{a_k}\right], \quad w_k > 0, \quad k = 0, 1, \ldots \tag{53}$$

The noise $w_k \in \mathbb{R}$, $w_k > 0$ is a scalar-independent negative exponentially distributed sequence with parameter $a_k \in \mathbb{R}$, $a_k > 0$. The parameters $A_k \in \mathbb{R}$, $A_k > 0$ and $a_k > 0$ are deterministic and assumed to be known. The process x_k is not observed directly. Instead, a 'delayed' version z_k is observed. The delay is another independent negative exponentially distributed random variable, with (deterministic and known) parameter $b_k \in \mathbb{R}$, $b_k > 0$. The observations $\{z_0, \ldots, z_k\}$ thus have the following independent density:

$$\Psi^{(z)}(z_k; x_k) = \frac{1}{b_k}\exp\left[-\frac{z_k - x_k}{b_k}\right], \quad z_k \geqslant x_k. \tag{54}$$

Aim: Given the observations $\{z_0, \ldots, z_k\}$ the aim is to derive an optimal recursive filter for the arrival times x_k, i.e., the aim is to compute $\mathbf{E}[x_k \mid \mathscr{Z}_k]$.

Remark

1. If $A_k = 1$, then x_k models the arrival time of the kth Poisson event. Furthermore, if there is a delay between the arrival of each Poisson event and its observation, then z_k models the observation time of the kth Poisson event. Based only on the observation times, the aim is to determine the actual arrival times.
2. The requirement $A_k > 0$ in (52) ensures that the state x_k in (52) is always non-negative. This in turn ensures that the filtered density $q_k(x)$ has one-sided support $[0, \infty]$ for all k.
3. Because the support of $\Psi^{(z)}$ in (54) depends on x_k, $\Psi^{(z)}$ is *not* a family of exponential distributions, i.e.$\Psi^{(z)}$ cannot be put into the form (9).

Practical applications: Let the state x_k in (54) be the time that the kth event occurs. Furthermore, as soon as an event occurs, a message is dispatched. Let the observation z_k in (54) be the time that the message arrives at its destination. Therefore, the observation time z_k is the arrival time x_k plus the transit time of the message.

A simple example is the case of a sensor connected to the interrupt line of a CPU. When an event is detected by the sensor, an interrupt will be generated. However, there will be a delay between the interrupt being triggered and the CPU actually handling the interrupt. (The delay will be caused by the CPU servicing higher-priority interrupts, for example.) Under certain circumstances, it is reasonable to model this delay by a negative exponentially distributed random variable.

A more complex example of a negative exponentially distributed delay is in ATM networks.[26] Assume there are a number of virtual circuits (VC) carrying voice calls. Each voice call can be characterized by an on–off two-state Markov chain. When it is in the ON state, it can be assumed that cells are generated according to a Poisson process. When such calls are multiplexed together and fed into an ATM node, it has been found[27] that the queue length has a negative exponential distribution. Since ATM cells are constant in length, the time spent in the queue for each cell is proportional to the queue length and hence will also have a negative exponential distribution.

Therefore, if a cell is sent whenever some event occurs, the destination ATM node can be expected to receive that cell delayed in time by a negative exponentially distributed amount. If the events correspond to arrival times of a Poisson process, then the Poisson arrival time filter can be used to estimate the exact time the events occurred based on the arrival time at the destination ATM node.

Finite-dimensional filter: For simplicity of derivation (see Remark 2 below the proof) only the case $a_k \geqslant b_k$ in the model (52)–(54) is considered. This criterion is a realistic assumption since it implies that the measurement delay is on average smaller than the underlying process delay.

The step function $u(\cdot)$ defined by

$$u(x) = \begin{cases} 1, & x \geqslant 0 \\ 0, & x < 0 \end{cases} \tag{55}$$

is used in the optimal Poisson arrive time filter derived in Theorem 2.

Theorem 2 (Optimal Poisson arrival time filter)

At time k, the unnormalized filtered density $q_k(x)$ defined in (17) or the Poisson arrival time process defined above (52)–(54) can be expressed as

$$q_k(x) = \sum_{i=0}^{k} \left(\sum_{j=0}^{k-i} \alpha_{ij}^{(k)} \exp[c_j^{(k)} x] \right) u(\beta_i^{(k)} - x), \quad x \geqslant 0 \tag{56}$$

Thus, $q_k(x)$ is completely determined by the $\alpha_{ij}^{(k)} \in \mathbb{R}$, $i = 0, \ldots, k$, $j = 0, \ldots, k - i$; $c_j^{(k)} \in \mathbb{R}$, $j = 0, \ldots, k$ and $\beta_i^{(k)} \in \mathbb{R}$, $i = 0, \ldots, k$. These statistics are initialized to

$$\beta_0^{(0)} = z_0, \quad c_0^{(0)} = \frac{1}{b_0} - \frac{1}{a_0}, \quad \alpha_{0,0}^{(0)} = 1 \tag{57}$$

and are recursively updated by

$$\begin{aligned}
\beta_{k+1}^{(k+1)} &= z_{k+1} \quad \beta_i^{(k+1)} = A_{k+1} \beta_i^{(k)} \wedge z_{k+1}, \quad i = 0, \ldots, k \\
c_0^{(k+1)} &= \frac{1}{b_{k+1}} - \frac{1}{a_{k+1}} \quad c_{j+1}^{(k+1)} = \frac{1}{b_{k+1}} + \frac{1}{A_{k+1}} c_j^{(k)}, \; j = 0, \ldots, k \\
\alpha_{i,j+1}^{(k+1)} &= \alpha_{ij}^{(k)} \left(\frac{A_{k+1}}{a_{k+1}} + c_j^{(k)} \right)^{-1}, \quad i = 0, \ldots, k; \; j = 0, \ldots, k - i \\
\alpha_{i,0}^{(k+1)} &= -\sum_{j=0}^{k-i} \alpha_{ij}^{(k)} \left(\frac{A_{k+1}}{a_{k+1}} + c_j^{(k)} \right)^{-1} \exp\left[\left(\frac{A_{k+1}}{a_{k+1}} + c_j^{(k)} \right) \beta_i^{(k)} \right], \quad i = 0, \ldots, k \\
\alpha_{k+1,0}^{(k+1)} &= \sum_{i=0}^{k} \sum_{j=0}^{k-i} \alpha_{ij}^{(k)} \left(\frac{A_{k+1}}{a_{k+1}} + c_j^{(k)} \right)^{-1} \left(\exp\left[\left(\frac{A_{k+1}}{a_{k+1}} + c_j^{(k)} \right) \beta_i^{(k)} \right] - 1 \right)
\end{aligned} \tag{58}$$

Furthermore, the filtered estimate of the Poisson arrival time is given by

$$\mathbf{E}[x_k \mid \mathscr{Y}_k] = \frac{\sum_{i=0}^{k} \sum_{j=0}^{k-i} \alpha_{ij}^{(k)} S_1(\beta_i^{(k)}, c_j^{(k)})}{\sum_{i=0}^{k} \sum_{j=0}^{k-i} \alpha_{ij}^{(k)} S_0(\beta_i^{(k)}, c_j^{(k)})} \tag{59}$$

where

$$S_0(\beta_i^{(k)}, c_j^{(k)}) = \begin{cases} \beta_i^{(k)}, & c_j^{(k)} = 0 \\ \frac{\exp[c_j^{(k)}\beta_i^{(k)}]-1}{c_j^{(k)}}, & c_j^{(k)} \neq 0 \end{cases}$$

$$S_1(\beta_i^{(k)}, c_j^{(k)}) = \begin{cases} \frac{(\beta_i^{(k)})^2}{2}, & c_j^{(k)} = 0 \\ \frac{1+(c_i^{(k)}\beta_i^{(k)}-1)\exp[c_j^{(k)}\beta_i^{(k)}]}{(c_j^{(k)})^2}, & c_j^{(k)} \neq 0 \end{cases} \tag{60}$$

Proof: Due to the parameter-dependent support of the observation noise density $\Psi^{(z)}$, the signal model does not belong to the class of signal models defined in Section 2. However, the derivation of the unnormalized filtered density (Section 3) is applicable to any partially observable Markov process. Define

$$\lambda_0(x_0) = \frac{(1/a_0)\exp[-x_0/a_0]\,(1/b_0)\exp[-(z_0-x_0)/b_0]}{\bar{p}(x_0)\,\bar{p}(z_0)} u(z_0 - x_0) \tag{61}$$

$$\lambda_k(x_{k-1}, x_k) = \frac{(1/a_k)\exp[-(x_k - A_k x_{k-1})/a_k]\,(1/b_k)\exp[-(z_k - x_k)/b_k]}{\bar{p}(x_k)\,\bar{p}(z_k)} \times u(x_k - A_k x_{k-1})\; u(z_k - x_k) \tag{62}$$

where $\bar{p}(x) = \exp[-x]$ and $u(\cdot)$ is defined in (55). Then, as shown in the proof of Lemma 1, the unnormalized filtered density q_k is given recursively by

$$q_0(x) = \lambda_0(x)\,\bar{p}(x), \quad x \geqslant 0 \tag{63}$$

$$q_{k+1}(y) = \int_0^\infty \lambda_{k+1}(x, y)\,\bar{p}(y) q_k(x)\,\mathrm{d}x, \quad y \geqslant 0 \tag{64}$$

Substituting (61) into (63) gives

$$q_0(x) = K_0 \;\exp\left[\left(\frac{1}{b_0} - \frac{1}{a_0}\right) x\right] u(z_0 - x_0), \quad x \geqslant 0 \tag{65}$$

where $K_0 \in \mathbb{R}$ does not depend on x. This is equivalent to (56) when (57) holds. To verify (58), it is necessary to show that if $q_k(x) = \exp[c_j^{(k)} x] u(\beta_i^{(k)} - x)$, then

$$q_{k+1}(y) = \left\{\frac{1}{\alpha}\exp\left[\left(\frac{\alpha}{A_{k+1}} + c_0^{(k+1)}\right) y\right] - \frac{1}{\alpha}\exp[\alpha\beta_i^{(k)}]\exp[c_0^{(k+1)} y]\right\} u(A_{k+1}\beta_i^{(k)} - y) \times u(z_{k+1} - y) + \left\{\frac{1}{\alpha}(\exp[\alpha\beta_i^{(k)}] - 1)\exp[c_0^{(k+1)} y]\right\} u(z_{k+1} - y) \tag{66}$$

$$= \left\{\frac{1}{\alpha}\exp[c_{j+1}^{(k+1)} y] - \frac{1}{\alpha}\exp[\alpha\beta_i^{(k)}]\exp[c_0^{(k+1)} y]\right\} u(\beta_i^{(k+1)} - y) \tag{67}$$

$$+ \left\{\frac{1}{\alpha}(\exp[\alpha\beta_i^{(k)}] - 1)\exp[c_0^{(k+1)} y]\right\} u(z_{k+1} - y)$$

where

$$\alpha = \frac{A_{k+1}}{a_{k+1}} + c_j^{(k)}, \qquad c_0^{(k+1)} = \frac{1}{b_{k+1}} - \frac{1}{a_{k+1}} \tag{68}$$

$$\beta_i^{(k+1)} = A_{k+1}\beta_i^{(k)} \wedge z_{k+1}, \qquad c_{j+1}^{(k+1)} = \frac{\alpha}{A_{k+1}} + c_0^{(k+1)} = \frac{1}{b_{k+1}} + \frac{1}{A_{k+1}} c_j^{(k)} \tag{69}$$

Substituting $q_k(x) = \exp[c_j^{(k)} x] u(\beta_i^{(K)} - x)$ into (64) gives

$$q_{k+1}(y) = \exp[c_0^{(k+1)} y] \int_0^{\beta_i^{(k)} \wedge y/A_{k+1}} \exp[\alpha x]\, \mathrm{d}x u(z_{k+1} - y) \tag{70}$$

where α and $c_0^{(k+1)}$ are defined in (68). Equation (66) follows upon evaluating the integral in (70) assuming α in (68) is non-zero (see Remark 2 below).

Finally, by combining (16), (17) and (56), the filtered arrival time can be computed:

$$\mathbf{E}[x_k \mid \mathscr{Z}_k] = \frac{\int_0^\infty z q_k(x) \mathrm{d}x}{\int_0^\infty q_k(x) \mathrm{d}x} \tag{71}$$

$$= \frac{\sum_{i=1}^{k} \sum_{j=0}^{k-i} \alpha_{ij}^{(k)} S_1(\beta_i^{(k)}, c_j^{(k)})}{\sum_{i=0}^{k} \sum_{j=0}^{k-i} \alpha_{ij}^{(k)} S_0(\beta_i^{(k)}, c_j^{(k)})} \tag{72}$$

where

$$S_0(b, c) = \int_0^\infty \exp[cx]\, \mathrm{d}x, \qquad S_1(b, c) = \int_0^b x \exp[cx] \mathrm{d}x \tag{73}$$

Evaluating the above integrals gives (60).

Remark

1. *Computational complexity*: The computational complexity involved in filtering a block of observations $\{z_0, \ldots, z_k\}$ is $O(k^3)$. The amount of storage required is $O(k^2)$.
2. Division by zero will occur if $\alpha = (A_{k+1}/a_{k+1}) + c_j^{(k)} = 0$ in (66) above. Note that $a_k \geqslant b_k$ ensures that $(A_{k+1}/a_{k+1}) + c_j^{(k)} > 0,\ k = 0, 1, \ldots$. It can be shown that $\alpha = 0$ signifies that the filtered density q_{k+1} contains a polynomial term. In general, the filtered density $q_k(x)$ (17) for the Possion arrival time model (52)–(54) is always of the form

$$q_k(x) = \sum_i \left(\sum_j f_{ij}(x) \mathrm{e}^{c_{ij} x} \right) u(\beta_i - x) \tag{74}$$

   where the f_{ij} are polynomials. The two sums will always have finite (but in general growing with time k) limits.
3. Since $\beta_i^{(k+1)} = A_{k+1} \beta_i^{(k)} \wedge z_{k+1}$, it is posible for some of the $\beta_i^{(k)}$ to be equal. If this occurs, the corresponding $\alpha_{ij}^{(k)}$ can be merged into one term.

6. SUB-OPTIMAL FILTERS

This section shows how to construct a sub-optimal filter based on any of the optimal filters in Sections 4 and 5.2. The resulting sub-optimal filter has a significantly lower computational complexity compared to the original filter. After outlining the key idea, namely to approximate the filtered density by an Edgeworth series, the sub-optimal filter for the doubly stochastic Poisson model (Section 4) is stated.

The optimal filters in Sections 4 and 5.2 have a filtered density $q_k(x)$ (17) in the form of a polynomial times a Gaussian. If the polynomial has many non-zero coefficients, the update (30) requires many multiplications. Therefore, the computational complexity can be significantly reduced if at each instant, the filtered density is approximated by a low-order polynomial times a Gaussian. Such an approximation can be achieved by using an Edgeworth series expansion.[28] The Edgeworth series expansion has the property that the lth-order expansion has the same l cumulants as the original density.

To demonstrate how to construct a sub-optimal filter, the corresponding sub-optimal filter for the doubly stochastic Poisson process filter (Theorem 1) is now derived. For convenience, a fourth-order Edgeworth series expansion is used. Since the odd cumulants of the filtered density $q_k(x)$ (27) are zero, the choice of a fourth degree polynomial matches both the variance and the fourth-order cumulant, allowing for departure from normality.

Let $p_k(x)$ be the (unnormalized) density we wish to approximate. $p_k(x)$ is assumed to have its odd cumulants identically zero.

Lemma 5 (Edgeworth series approximation)

Given an unnormalised density function $p_k(x)$ with odd cumulants identically zero, the density $\bar{q}_k(x) = K[P_0 + P_2x^2 + P_4x^4]\exp[-\frac{1}{2}x^2/V]$ will have the same first four cumulants provided:

$$P_0 = m_2^2(3\rho_4 + 24), \quad P_2 = -6m_2\rho_4, \quad P_4 = \rho_4, \quad V = m_2 \tag{75}$$

where

$$m_2 = \frac{\int_{\mathbb{R}} x^2 p_k(x)\, \mathrm{d}x}{\int_{\mathbb{R}} p_k(x)\, \mathrm{d}x}, \quad m_4 = \frac{\int_{\mathbb{R}} x^4 p_k(x)\, \mathrm{d}x}{\int_{\mathbb{R}} p_k(x)\, \mathrm{d}x}, \quad \rho_4 = \frac{m_4}{m_2^2} - 3 \tag{76}$$

and K is the normalizing constant to make $\bar{q}_k(x)$ integrate to one.

Proof: Since $p_k(x)$ has its odd cumulants identically zero, we need only consider even cumulants. The standardized density $[1 + (\rho_4/4!)(z^4 - 6z^2 + 3)]\exp[-\frac{1}{2}z^2]$ has mean zero, variance one and fourth-order culumant ρ_4. The result now follows easily.

Due to the Edgeworth series expansion matching cumulants, using either $p_k(x)$ or its approximation $\bar{q}_k(x)$ as the filtered density in (32) gives the same estimate of x_k^2.

The following algorithm summarizes the implementation of the sub-optimal filter for the doubly stochastic Poisson process defined in Section 4.

Algorithm 1 (*Sub-optimal filter*)

Let $\bar{q}_k(x)$, $k=0,1,\ldots$, denote the unnormalized sub-optimal filtered density. The sub-optimal filter, at time $k=0$, is initialized by defining $p_0(x)=q_0(x)$ where $q_0(x)$ is defined in (36). For each time instant $k=0,1,\ldots$ the sub-optimal filtered density is updated as follows.

1. Use Lemma 5 to calculate m_2, m_4, ρ_4 and $\bar{q}_k(x)$.
2. Output m_2 as the filtered estimate for x_k^2.
3. Substitute $\bar{q}_k(x)$ for $q_k(x)$ and $p_{k+1}(x)$ for $q_{k+1}(x)$ in (27), Theorem 1, to compute $p_{k+1}(x)$.
4. Set $k=k+1$.

7. NUMERICAL EXAMPLE

This section presents computer simulations to illustrate the performance of the proposed optimal and sub-optimal filters for the doubly stochastic Poisson process. The doubly stochastic Poisson process model (24)–(26) with parameters $A_k=0{\cdot}95, c_k=0{\cdot}6$ and $\sigma_k^2=\frac{1}{2}$ was used to generate 20 Poisson distributed observations of the linear Gaussian dynamical system. Both the optimal filter (Theorem 1 of Section 4) and the sub-optimal filter (Algorithm 1 of Section 6) were used to estimate the state of the system given these observations. Figure 1 shows, for each time instant k, the true value of the rate $(c_k x_k)^2$, the integer-valued observation z_k, and the optimal filtered estimate $\mathbf{E}[(c_k x_k)^2 \mid \mathscr{Y}_k]$. Figure 2 shows the sub-optimal filtered estimate. The optimal filtered rate was 2·6 dB better than the estimator $(\widehat{c_k x_k})^2=z_k$. The sub-optimal filtered rate was 2·2 dB better, only 0·4 dB worse than the optimal filter. Figure 3 shows the optimal and the sub-optimal filtered rates in the one graph for comparison. Note that in this example, the sub-optimal filter performs comparably to the optimal filter.

The same doubly stochastic Poisson process model was then used to generate 100 observations. The sub-optimal filtered rate is shown in Figure 4. The improvement over using the observations alone (i.e., $(\widehat{c_k x_k})^2=z_k$) was 0·7 dB. Note that around $k=45$ to 50, due to two large ($z_k=10$) observations, the filtered rate differed significantly to the true rate in this region. Overall though, the sub-optimal filtered rate followed the true rate reasonably well.

Table I shows the improvement in the mean-square error (MSE) of the optimal and sub-optimal doubly stochastic Poisson process filters. The improvement is relative to the simple estimate of the rate based on the observations alone, namely $(\widehat{c_k x_k})^2=z_k$. The precise model used was the doubly stochastic Poisson process model (24)–(26) with parameters $\sigma_k^2=\frac{1}{2}$, and $A_k=0{\cdot}1, 0{\cdot}5, 0{\cdot}8, 0{\cdot}95$, $c_k=0{\cdot}25, 0{\cdot}5, 0{\cdot}75$ as indicated in the table. For each trial, eight observations were generated, and the filtered rate calculated for each of the eight time instants. The MSE was calculated by performing 250 trials, and averaging the squared-errors resulting from each trial. There are three general observations that can be made based on Table I:

1. As $A_k \to 0$ and $c_k \to 0$, the sub-optimal filter performs almost identically to the optimal filter.
2. The improvement in MSE increases as $c_k \to 0$.
3. The improvement in MSE increases initially but then decreases as A_k decreases towards zero, (see for example the last column of Table I)

These observations can be explained as follows. In an intuitive sense, the total information available about the true rate consists of two parts, the observations z_k, and the correlations between the x_k caused by the model (7). The more information that is contained in the correlations relative to the actual observations, the better the optimal filter will perform compared to the simple estimate

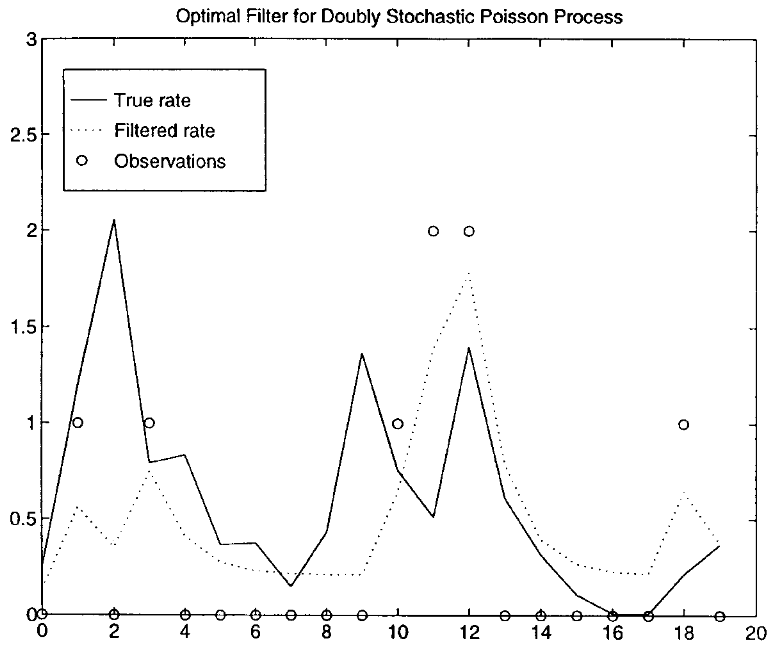

Figure 1. Simulation example of the optimal filter applied to a doubly stochastic Poisson process

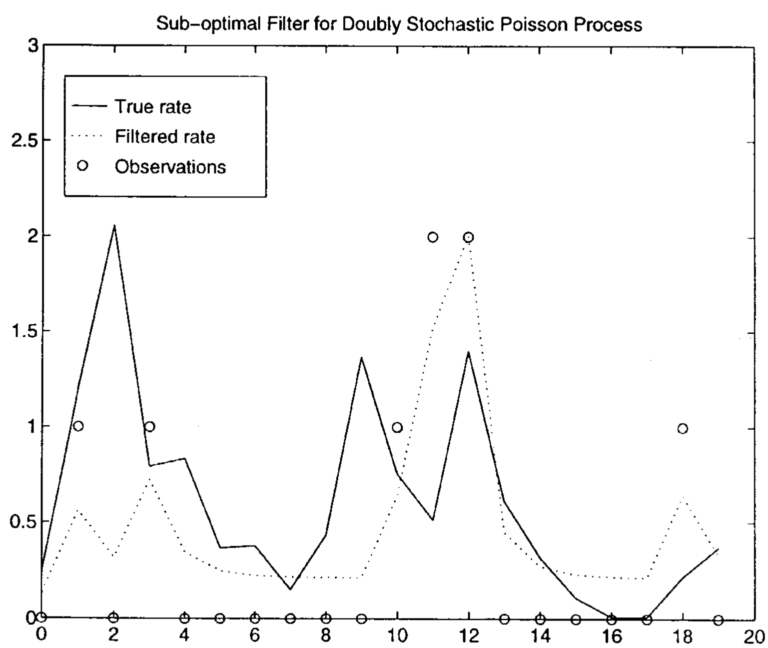

Figure 2. Simulation example of the sub-optimal filter applied to a doubly stochastic Poisson process

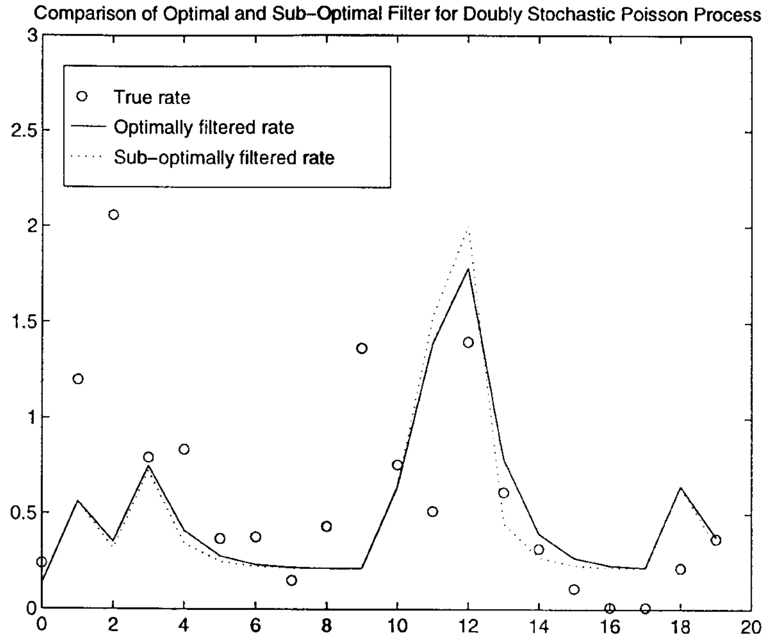

Figure 3. Comparison of the optimal and sub-optimal filters applied to a doubly stochastic Poisson process

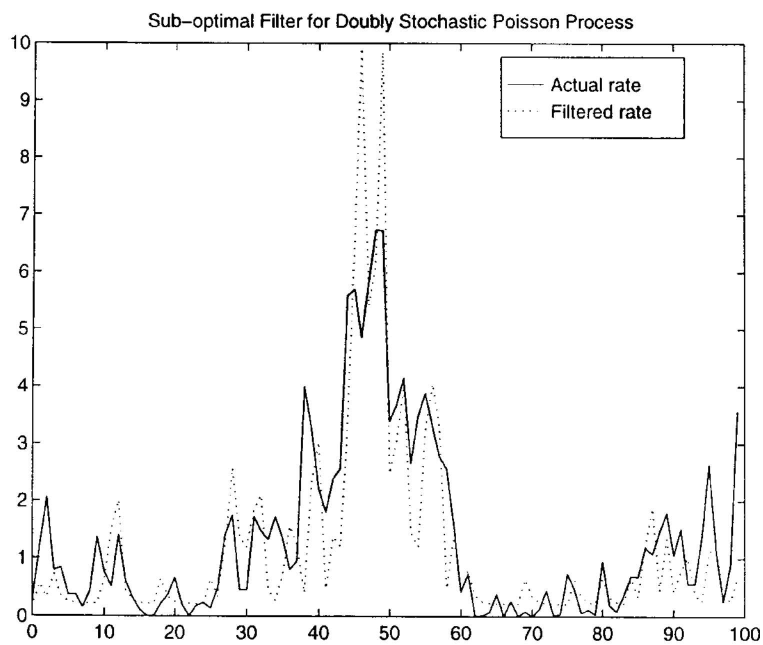

Figure 4. Simulation example of the sub-optimal filter applied to a doubly stochastic Poisson process over a longer time frame

Table I. The improvement (dB) obtained by applying the optimal and sub-optimal filters to the doubly stochastic Poisson process with parameters $\sigma_k^2 = \frac{1}{2}$ and A_k, c_k

A_k	Filter	$c_k = 0{\cdot}25$	$c_k = 0{\cdot}5$	$c_k = 0{\cdot}75$
0·1	Optimal	13·2	6·84	3·90
	Sub-optimal	13·2	6·84	3·90
0·5	Optimal	11·6	7·02	4·43
	Sub-optimal	11·6	7·01	4·43
0·8	Optimal	8·90	5·04	2·91
	Sub-optimal	8·90	4·82	2·60
0·95	Optimal	7·56	4·31	2·76
	Sub-optimal	7·41	3·23	1·82

based on the observations alone. Therefore, the improvement in MSE will increase if either the amount of information present in the observations z_k decreases, or if the correlation between the x_k increases.

The information contained in the observations z_k decreases as the rate dies off to zero. This in turn is caused by $c_k \to 0$, or similarly, by the average value of x_k^2 being small. The latter is caused to some extent by A_k being small. Therefore, for large A_k, as A_k decreases, it appears from Table I that the improvement in MSE increases, which is attributed to the average value of x_k^2 decreasing and thus removing information from the observations. However, reducing A_k also reduces the correlation between the x_k. There comes a point beyond which decreasing A_k causes the improvement in MSE to decrease.

Lastly, the sub-optimal filter is expected to perform comparably to the optimal filter when the amount of information which the optimal filter can exploit decreases. This is the case either when the correlation between the x_k is small (i.e. $A_k \to 0$), or when the observations contain little information (i.e. $c_k \to 0$).

ACKNOWLEDGEMENTS

Jonathan Manton acknowledges the support of the Australian Telecommunication and Engineering Research Board and Cooperative Research Centre for Sensor Signal and Information Processing. Vikram Krishnamurthy acknowledges the support of an Australian Research Council large grant and the Coorperative Research Centre for Sensor Signal and Information Processing. Robert Elliott acknowledges NSERC for its support under grant A7964.

REFERENCES

1. Cox, D. R., 'Some statistical methods connected with series of events', *J. Roy. Statist. Soc. Ser. B*, **17**, 129–164 (1955).
2. Brémaud, P. *Point Processes and Queues, Martingale Dynamics*, Springer, Berlin, 1981.
3. Snyder, D. L. and M. I. Miller, *Random Point Processes in Time and Space*, Springer, Berlin, 1991.
4. Ross, K. W., *Multiservice Loss Models for Broadband Telecommunication Networks*, Springer, Berlin, 1995.
5. Snyder, D. L. and T. J. Schulz, 'High resolution imaging at low light-levels through weak turbulence', *J. Opt. Soc. Amer.* A (1990).
6. Daba, J. S. and M. R. Bell, 'Estimation of the surface reflectivity of SAR images based on a marked Poisson point process model', in *Int. Symp. on Signals, Systems and Electronics*, URSI, 1995, pp. 183–186.

7. Brady, D. and S. Verdú, 'Performance analysis of an asymptotically quantum-limited optical DPSK receiver', *IEEE Trans. Commun.*, **37**(1), 46–51 (1989).
8. Slimane, S. B. and T. Le-Ngoc, 'A doubly stochastic Poisson model for self-similar traffic', in *Int. Conf. on Communications*, IEEE, New York, 1995, pp. 456–460.
9. Subramanian, S. N. and T. Le-Ngoc, 'Traffic modeling in a multi-media environment', in *Canadian Conf. on Electrical and Computer Engineering*, IEEE, New York, 1995, pp. 838–841.
10. Lindgren, G. and U. Holst, 'Recursive estimation of parameters in Markov-modulated Poisson processes', *IEEE Trans. Commun.*, **43**(11), 2812–2820 (1995).
11. Rydén, T., 'Parameter estimation for Markov-modulated Poisson processes', *Stochastic Models*, **10**, 795–829 (1994).
12. Larsen, M., 'Point process models for weather radar images', in *Int. Geoscience and Remote Sensing Symp.*, IEEE, New York, 1994, pp. 25–28.
13. Le Cam, L., 'A stochastic description of precipitation', in *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, Vol. 3, 1961, pp. 165–186.
14. Phelan, M. J., 'Point processes and inference for rainfall fields', in Basava and Taylor (eds.), *Selected Proceedings of the Sheffield Symposium on Applied Probability*, Institute of Mathematical Statistics, 1991, pp. 127–148.
15. Snyder, D. L., 'Filtering and detection for doubly stochastic poisson processes', *IEEE Trans. Inform. Theory*, **18**(1), 91–102 (1972).
16. Boel, R. K. and V. E. Beneš, 'Recursive nonlinear estimation of a diffusion acting as the rate of an observed poisson process', *IEEE Trans. Inform. Theory*, **26**(5), 561–575 (1980).
17. Segall, A., 'Recursive estimation from discrete-time point processes', *IEEE Trans. Inform. Theory*, **22**(4), 422–431 (1976).
18. Sawitzki, G., 'Finite dimensional filter systems in discrete time', *Stochastics*, **5**, 107–114 (1981).
19. Bather, J. A., 'Invariant conditional distributions', *Ann. Math. Statist.*, **36**, 829–846 (1965).
20. Koopman, B. O., 'On distributions admitting a sufficient statistic', *Trans. Amer. Math. Soc.*, **39**, 399–409 (1936).
21. Brown, L., 'Sufficient statistics in the case of independent random variables', *Ann. Math. Statist.*, **35**, 1456–1474 (1964).
22. Murray, M. K. and J. W. Rice, *Differential Geometry and Statistics*, Monographs on Statistics and Applied Probability, Vol. 48, Chapman & Hall, London, 1993.
23. Lehmann, E. L., *Theory of Point Estimation*, Wiley, New York, 1983.
24. Elliott, R. J., L. Aggoun and J. B. Moore, *Hidden Markov Models*: *Estimation and Control*, Springer, Berlin, 1995.
25. Sorenson, H. W. and D. L. Alspach, 'Recursive Bayesian estimation using Gaussian sums', *Automatica*, **7**, 465–479 (1971).
26. De Pryker, M., *Asynchronous Transfer Mode*, 3rd edn, Prentice-Hall, Englewood cliffs, NJ, 1995.
27. Anick, D., D. Mitra and M. Sonhdi, 'Stochastic theory of a data handling system with multiple resources', *Bell Systems Tech. J.*, **61**(8), 1871–1894 (1982).
28. McCullagh, P., *Tensor Methods in Statistics*, Chapman & Hall, London, 1987.
29. Basava, I. W. and R. L. Taylor (eds.), *Selected Proceedings of the Sheffield Symposium on Applied Probability*, Institute of Mathematical Statistics, 1991.

