Dynamics and Control, 7, 1–14 (1997)

Computational Suboptimal Filter for a Class of Wiener-Poisson Driven Stochastic Processes

RAGHIB M. ABU-SARIS saris@math.birzeit.edu
Mathematics Department, Birzeit University, P. O. Box 14, Birzeit, West Bank, via Israel

FLOYD B. HANSON hanson@math.uic.edu
Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 S. Morgan, MC 249, Chicago, IL 60607-7045, USA

Received January 23, 1996; Revised July 17, 1996

Editor: N. U. Ahmed

Abstract. The minimum mean square estimate (MMSE) for a stochastic process driven simultaneously by Wiener and Poisson processes is characterized by an infinite number of stochastic differential equations (even in the simplest linear case), and so is not practically implementable. In this article, a practical approximation to the solution is developed in terms of a computationally suboptimal filter for the estimation problem. Basically, it detects and estimates the Poisson driving process using a Maximum A Posteriori (MAP) criterion, and then reconstructs the entire system state using MMSE applied to a system approximating the original one.

Keywords: Filtering, partial observations, suboptimal filter, jump diffusions, computational filtering method

1. Introduction

The assumption that the disturbing noise for a dynamical system being Gaussian is not satisfactory in many realistic applications. In particular when the system is exposed to sudden, infrequent, highly localized changes or disturbances that occur in a short period of time such as earthquakes, large random weather fluctuations, or occasional mass mortalities [6, 9]. Therefore, the dynamical system under study is modeled as a stochastic differential equation (SDE) driven by a combination of Gaussian and marked Poisson processes. An application of jump-diffusion processes is target tracking (cf., Miller et al. [18]). The Gaussian-driven part represents continuous fluctuations or diffusion in the system [7, 10]. On the contrary, Poisson-driven noise models have random discontinuities in the underlying state dynamics [7, 23, 9]. The exact solution, if obtainable, depends on the driving processes which are random in nature and, in addition, the "best" estimate based upon partial observations, that are themselves contaminated by noise. Further, practical implementation, if at all possible, is complicated when there are jumps in the observations as well as in the state dynamics. In this research, a computationally, but suboptimal, filter will be developed that is practically implementable.

Many approximate filters were developed for nonlinear systems, almost all for Gauss-driven noise only. Developments have relied on three main approaches. The first is based upon Taylor's series expansion of system nonlinearities around a nominal trajectory. Performing the expansion only up to the first order terms results in the *linearized Kalman*

Filter [10]. This approach can further be improved by linearizing, again up to first order, about the most recent estimate. By doing so, the well-known *Extended Kalman Filter (EKF)* is obtained [10, 17, 21]. The *truncated minimum-variance, modified minimum-variance, and Gaussian* filters are obtained when the expansion is carried out to second order terms [10, 22, 16, 17]. In the second approach, the Conditional Probability Density Function (CDF) is parametrized via a finite and small set of parameters. This parametrization immediately suggests an *orthogonal expansion* (cf., Jazwinski [10] and Kwakernaak [15]). Emara-Shabaik [5], and Ahmed and Radaideh [2] considered the nonlinear filtering problem from a different approach, by approximating the problem and solving the approximate model. Emara-Shabaik [5] utilizes a nearly linear approximate model of a practical filter, while Ahmed and Radaideh [2] use a nearly deterministic approximate model in their *Modified Extended Kalman Filter*. Au, Haddad and Poor [3] developed an approximate suboptimal filter that is a combination of detection and estimation for a scalar, linear, and Gauss-Poisson driven system observed via Gauss-driven observations. Kushner and Dupuis [14] discuss the treatment of the nonlinear filter for the jump diffusion state using the Markov chain approximation.

The aim here is to formally develop an approximate suboptimal filter for system and observations that are both driven by Weiner and Poisson processes simultaneously. The focus here in on the application of computational methods. In Section 2, the problem formulation is presented. In Section 3, the suboptimal filter was derived by combining the two schemes due to Au, Haddad and Poor [3] and to Emara-Shabaik [5]. In Section 4, several numerical experiments with our approximate suboptimal filter are discussed. Conclusions are given in Section 5.

2. Problem Formulation

The dynamical system under study is described by the $n \times 1$ state vector $\mathbf{X}(t)$ and observed continuously via the $m \times 1$ observation vector $\mathbf{Z}(t)$. Both the state and the observations are governed by systems of stochastic differential equations:

$$d\mathbf{X}(t) = [\mathbf{A}_1(t)\mathbf{X}(t) + \mathbf{B}_1(\mathbf{X},t)]\,dt + \mathbf{G}_1(t)d\mathbf{W}_1(t) + \int_{\mathcal{U}_1} \mathbf{H}_1(t,\mathbf{U})\mathcal{P}_1(dt,d\mathbf{U}), \quad (1)$$

$$d\mathbf{Z}(t) = [\mathbf{A}_2(t)\mathbf{X}(t) + \mathbf{B}_2(\mathbf{X},t)]\,dt + \mathbf{G}_2(t)d\mathbf{W}_2(t) + \int_{\mathcal{U}_2} \mathbf{H}_2(t,\mathbf{U})\mathcal{P}_2(dt,d\mathbf{U}), \quad (2)$$

where the state coefficient set $\{\mathbf{A}_1,\ \mathbf{B}_1,\ \mathbf{G}_1,\ \mathbf{H}_1\}$ has dimensions $\{n \times n,\ n \times 1,\ n \times r_1,\ n \times 1\}$ and the observer coefficient set $\{\mathbf{A}_2,\ \mathbf{B}_2,\ \mathbf{G}_2,\ \mathbf{H}_2\}$ has dimensions $\{m \times n,\ m \times 1,\ m \times r_2,\ m \times 1\}$, respectively. The $r_i \times 1$ zero-mean Wiener process $\mathbf{W}_i(t)$ and space-time Poisson process $\mathcal{P}_i(t,\mathbf{u})$ are assumed to be mutually independent. Thus

$$\begin{aligned} \mathrm{E}[d\mathbf{W}_i(t)] &= \mathbf{0}, \\ \mathrm{Covar}[d\mathbf{W}_i(t), d\mathbf{W}_j^T(t)] &= \delta_{i,j}\mathbf{I}_{r_i}dt; \end{aligned}$$

$$\mathrm{E}\left[\int_{\mathcal{U}_i} \mathbf{H}_i(\mathbf{X},t,\mathbf{U})\mathcal{P}_i(dt,d\mathbf{U}) \mid \mathbf{X}(t) = \mathbf{x}\right] = \lambda_i dt \int_{\mathcal{U}_i} \mathbf{H}_i(\mathbf{x},t,\mathbf{u})\phi_i(\mathbf{u})d\mathbf{u}; \quad (3)$$

$$\text{Covar}\left[\int_{\mathcal{U}_i} \mathbf{H}_i \mathcal{P}_i(dt, d\mathbf{U}), \int_{\mathcal{U}_j} \mathbf{H}_j^T \mathcal{P}_j(dt, d\mathbf{U}) | \ \mathbf{X}(t) = \mathbf{x}\right] \\ = \delta_{i,j} \lambda_i dt \int_{\mathcal{U}_i} \mathbf{H}_i \mathbf{H}_i^T \phi_i(\mathbf{u}) d\mathbf{u}; \tag{4}$$

$$\text{Covar}\left[d\mathbf{W}_i(t), \mathcal{P}_j(dt, d\mathbf{U})\right] = \mathbf{0},$$

for $i, j = 1, 2$, where λ_i the Poisson intensity and $\phi_i(\mathbf{u})$ is the probability density function of marks (Poisson amplitudes) $\mathbf{U} = \mathbf{u}$ in the mark space $\mathcal{U}_i$ corresponding to the i^{th} Poisson process.

The objective is to estimate the system state vector, $\mathbf{X}(t)$ based upon a prior set of observations on $(t_0, t]$,

$$\mathcal{Z}_t = \{\mathbf{Z}(s); t_0 < s \leq t\}.$$

It is a known [10] that the conditional mean, $\mathrm{E}[\mathbf{X}(t)|\mathcal{Z}_t]$, minimizes the mean square error, $\mathrm{E}[(\mathbf{X}(t) - \widehat{\mathbf{X}}(t))^T(\mathbf{X}(t) - \widehat{\mathbf{X}}(t))|\mathcal{Z}_t]$. Abu-Saris [1] has shown that the conditional mean for this model is characterized by an infinite number of stochastic differential equations even for the simplest scalar linear case, like the nonlinear case and unlike the original Kalman filter [11, 12]. Hence, approximation and estimation are needed for realistic, physical implementation. The essential part of current research begins here.

3. Suboptimal Filter

In addition to existence and uniqueness [7], the Lipschitz continuity of the nonlinear terms in $\mathbf{x}$, and the dominance of the linear terms [5] is assumed, i.e.,

$$|B_{i,j}(\mathbf{x}, t)| < \theta \left| \sum_{k=1}^{n} A_{i,j,k}(t) x_k \right|, \tag{5}$$

for some $0 < \theta < 1$, for $i = 1$ to 2, and for all $j = 1$ to n. The assumption that linear terms dominate implies that the nonlinear terms, B_1 and B_2, can be approximated using a *good guess* $\mathbf{X} = \mathbf{X}^*$, such that such that the nearly linear conditions in (5) are satisfied. The question of the theoretical convergence of $\mathbf{X}^*(t)$ to $\mathbf{X}(t)$ is beyond the scope of the present study, but linear convergence would be anticipated provided the initial iterate $\mathbf{X}^*(t)$ is sufficiently close to the state $\mathbf{X}(t)$. Examples of numerical convergence for the nearly linear approximate model with Gaussian noise only is given by Emara-Shabaik [5] for several cases of a stochastically perturbed van der Pol oscillator. Related numerical corroboration is given by Ahmed and Radaideh [2] for the nearly deterministic approximate model with Gaussian noise only, illustrated by several quite different examples. Grewal and Andrews [8] offer a long list of convergence and divergence difficulties in Kalman filter implementations.

Thus, the system and observations equations (1, 2) can be approximated by $\{\widetilde{\mathbf{X}}(t), \widetilde{\mathbf{Z}}(t)\}$ using the following equations:

$$d\widetilde{\mathbf{X}}(t) = \left[\mathbf{A}_1(t)\widetilde{\mathbf{X}}(t) + \mathbf{B}_1(\mathbf{X}^*, t)\right] dt + \mathbf{G}_1(t)d\mathbf{W}_1(t) + \int_{\mathcal{U}_1} \mathbf{H}_1(t, \mathbf{U})\mathcal{P}_1(dt, d\mathbf{U}), \tag{6}$$

$$d\widetilde{\mathbf{Z}}(t) = \left[\mathbf{A}_2(t)\widetilde{\mathbf{X}}(t) + \mathbf{B}_2(\mathbf{X}^*, t)\right] dt + \mathbf{G}_2(t)d\mathbf{W}_2(t) + \int_{\mathcal{U}_2} \mathbf{H}_2(t, \mathbf{U})\mathcal{P}_2(dt, d\mathbf{U}). \tag{7}$$

Integration of the dynamical equation (6) leads to

$$\widetilde{\mathbf{X}}(t) = \boldsymbol{\Phi}(t, t_0)\widetilde{\mathbf{X}}(t_0) + \int_{t_0}^{t} \boldsymbol{\Phi}(t, s)\mathbf{B}_1(\mathbf{X}^*, s)ds + \int_{t_0}^{t} \boldsymbol{\Phi}(t, s)\mathbf{G}_1(s)d\mathbf{W}_1(s) + \int_{t_0}^{t} \int_{\mathcal{U}_1} \boldsymbol{\Phi}(t, s)\mathbf{H}_1(s, \mathbf{U})\mathcal{P}_1(ds, d\mathbf{U}), \tag{8}$$

where $\boldsymbol{\Phi}(t, s)$ is the linear state drift transition matrix, i.e., satisfying

$$\frac{\partial \boldsymbol{\Phi}}{\partial t}(t, s) = \mathbf{A}_1 \cdot \boldsymbol{\Phi}(t, s);\ \boldsymbol{\Phi}(s, s) = \mathbf{I}_n.$$

As can be seen from (8), the state approximation $\widetilde{\mathbf{X}}(t)$ is a sum of two parts. The first part, consisting of the first three terms of (8), is a solution to a jump-free and Gaussian-driven stochastic differential equation, while the second part and last term is solely dependent on the Poisson-driven part. Let

$$\widetilde{\mathbf{M}}(t) = \mathrm{E}[\widetilde{\mathbf{X}}(t) \mid \widetilde{\mathcal{Z}}_t, \boldsymbol{\eta}_1(t)]$$

be the mean value of $\widetilde{\mathbf{X}}(t)$ conditioned on prior observations $\widetilde{\mathcal{Z}}_t$ as usual, but also on the system Poisson component defined by

$$\boldsymbol{\eta}_1(t) = \int_{t_0}^{t} \int_{\mathcal{U}_1} \mathbf{H}_1(t, \mathbf{U})\mathcal{P}_1(ds, d\mathbf{U}); \qquad \boldsymbol{\eta}_1(t_0) = \mathbf{0}.$$

The conditional mean $\widetilde{\mathbf{M}}(t)$ can be decomposed into

$$\widetilde{\mathbf{M}}(t) = \widetilde{\mathbf{M}}_0(t) + \widetilde{\mathbf{M}}_1(t),$$

where the second part of the decomposition is defined as

$$\widetilde{\mathbf{M}}_1(t) = \int_{t_0}^{t} \int_{\mathcal{U}_1} \boldsymbol{\Phi}(t, s)\mathbf{H}_1(s, \mathbf{U})\mathcal{P}_1(ds, d\mathbf{U}) = \int_{t_0}^{t} \boldsymbol{\Phi}(t, s)\, d\boldsymbol{\eta}_1(s), \tag{9}$$

explicitly depending on the additional system Poisson conditioning. The equation for the remaining part of the conditional mean follows from the moment form of Kushner's filter equation (see [10] and [1]) modified for fixed $\mathbf{X}^*$ and additional conditioning on the current state jump cumulative amplitude $\boldsymbol{\eta}_1(t)$,

$$d\widetilde{\mathbf{M}}_0(t) = \left[\mathbf{A}_1\widetilde{\mathbf{M}}_0 + \mathbf{B}_1(\mathbf{X}^*, t)\right] dt + \widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\left\{d\widetilde{\mathbf{Z}}(t) - \mathbf{A}_2(\widetilde{\mathbf{M}}_0 + \widetilde{\mathbf{M}}_1)dt - \mathbf{B}_2(\mathbf{X}^*, t)dt - \lambda_2 \int_{\mathcal{U}_2} \mathbf{H}_2\phi_2(\mathbf{u})d\mathbf{u}dt\right\}, \tag{10}$$

The filter moment equation for the conditional covariance which is defined as

$$\widetilde{\mathbf{P}}(t) = \mathrm{E}[(\widetilde{\mathbf{X}}(t) - \widetilde{\mathbf{M}}(t)) \cdot (\widetilde{\mathbf{X}}(t) - \widetilde{\mathbf{M}}(t))^T \mid \widetilde{\mathcal{Z}}_t, \boldsymbol{\eta}_1(t)],$$

is given by

$$d\widetilde{\mathbf{P}}(t) = \left[\mathbf{A}_1\widetilde{\mathbf{P}} + \widetilde{\mathbf{P}}\mathbf{A}_1^T + \mathbf{G}_1\mathbf{G}_1^T - \widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{A}_2\widetilde{\mathbf{P}}\right] dt. \tag{11}$$

In the filter moment equations (10, 11),

$$\mathbf{R} = \mathbf{R}(t) = \mathbf{G}_2(t)\mathbf{G}_2^T(t)$$

is the covariance of the observer Gaussian term $\mathbf{G}_2(t)d\mathbf{W}_2(t)$ upon dividing the result by dt. Note that $\widetilde{\mathbf{X}}(t)$ is Gaussian for given $\boldsymbol{\eta}_1(t)$ so that its third centered moment is $\mathbf{0}$ and consequently the innovative term containing $d\widetilde{\mathbf{Z}}(t)$ in (10) does not appear in the condition covariance equation (11).

The filter moment equation (10) can be formally integrated to yield

$$\begin{aligned}\widetilde{\mathbf{M}}_0(t) \;=\; & \boldsymbol{\Psi}(t,t_0)\widetilde{\mathbf{M}}_0(t_0) + \int_{t_0}^{t} \boldsymbol{\Psi}(t,s)\mathbf{B}_1(\mathbf{X}^*,s)ds \\ & + \int_{t_0}^{t} \boldsymbol{\Psi}(t,s)\widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\Big\{d\widetilde{\mathbf{Z}}(s) - \mathbf{A}_2\widetilde{\mathbf{M}}_1 ds - \mathbf{B}_2(\mathbf{X}^*,s)ds \\ & - \lambda_2 \int_{\mathcal{U}_2} \mathbf{H}_2\phi_2(\mathbf{u})d\mathbf{u}ds\Big\},\end{aligned} \tag{12}$$

where $\boldsymbol{\Psi}(t,s)$ is the $\widetilde{\mathbf{M}}_0(t)$ linear component transition matrix, i.e.,

$$\frac{\partial \boldsymbol{\Psi}}{\partial t}(t,s) = \left[\mathbf{A}_1 - \widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{A}_2\right] \cdot \boldsymbol{\Psi}(t,s); \;\; \boldsymbol{\Psi}(s,s) = \mathbf{I}_n.$$

Further isolation of $\boldsymbol{\eta}_1$ terms can by attained by introducing the refined decomposition:

$$\widetilde{\mathbf{M}}_0(t) = \widetilde{\mathbf{M}}_2(t) + \widetilde{\mathbf{M}}_3(t),$$

so now

$$\widetilde{\mathbf{M}}(t) = \widetilde{\mathbf{M}}_1(t) + \widetilde{\mathbf{M}}_2(t) + \widetilde{\mathbf{M}}_3(t), \tag{13}$$

where

$$\begin{aligned}d\widetilde{\mathbf{M}}_2(t) \;=\; & \left[\mathbf{A}_1\widetilde{\mathbf{M}}_2(t) + \mathbf{B}_1(\mathbf{X}^*,t)\right] dt \\ & + \widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\Big\{d\widetilde{\mathbf{Z}}(t) - \mathbf{A}_2\widetilde{\mathbf{M}}_2(t)dt - \mathbf{B}_2(\mathbf{X}^*,t)dt \\ & - \lambda_2 \int_{\mathcal{U}_2} \mathbf{H}_2\phi_2(\mathbf{u})d\mathbf{u}dt\Big\} \\ =\; & \left[\mathbf{A}_1 - \widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{A}_2\right]\widetilde{\mathbf{M}}_2(t)dt + \left[\mathbf{B}_1(\mathbf{X}^*,t) - \widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{B}_2(\mathbf{X}^*,t)\right] dt \\ & + \widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\left\{d\widetilde{\mathbf{Z}}(t) - \lambda_2 \int_{\mathcal{U}_2} \mathbf{H}_2\phi_2(\mathbf{u})d\mathbf{u}dt\right\},\end{aligned} \tag{14}$$

and

$$\widetilde{\mathbf{M}}_3(t) = -\int_{t_0}^{t} \boldsymbol{\Psi}(t,s)\widetilde{\mathbf{P}}(s)\mathbf{A}_2^T(s)\mathbf{R}^{-1}(s)\mathbf{A}_2(s)\widetilde{\mathbf{M}}_1(s)ds. \tag{15}$$

Further collecting $\widetilde{\mathbf{M}}$ terms depending on $\boldsymbol{\eta}_1$, let

$$\widetilde{\mathbf{M}}_4(t) = \widetilde{\mathbf{M}}_1(t) + \widetilde{\mathbf{M}}_3(t), \tag{16}$$

so from (9) and (15),

$$\begin{aligned}
d\widetilde{\mathbf{M}}_4(t) &= d\widetilde{\mathbf{M}}_1(t) + d\widetilde{\mathbf{M}}_3(t) \\
&= \left\{\mathbf{A}_1\widetilde{\mathbf{M}}_1(t)dt + \int_{\mathcal{U}_1} \mathbf{H}_1(t,\mathbf{U})\mathcal{P}_1(dt,d\mathbf{U})\right\} \\
&\qquad + \left\{\left(\mathbf{A}_1 - \widetilde{\mathbf{P}}(t)\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{A}_2\right)\widetilde{\mathbf{M}}_3(t)dt - \widetilde{\mathbf{P}}(t)\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{A}_2\widetilde{\mathbf{M}}_1(t)dt\right\} \\
&= \left(\mathbf{A}_1 - \widetilde{\mathbf{P}}(t)\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{A}_2\right)\widetilde{\mathbf{M}}_4(t)dt + \int_{\mathcal{U}_1} \mathbf{H}_1(t,\mathbf{U})\mathcal{P}_1(dt,d\mathbf{U}).
\end{aligned}$$

Thus,

$$\widetilde{\mathbf{M}}_4(t) = \int_{t_0}^{t}\int_{\mathcal{U}_1} \boldsymbol{\Psi}(t,s)\mathbf{H}_1(s,\mathbf{U})\mathcal{P}_1(ds,d\mathbf{U}) = \int_{t_0}^{t} \boldsymbol{\Psi}(t,s)\, d\boldsymbol{\eta}_1(s). \tag{17}$$

Observe that now

$$\widetilde{\mathbf{M}}(t) = \widetilde{\mathbf{M}}_2(t) + \widetilde{\mathbf{M}}_4(t), \tag{18}$$

combining (13) and (16), where $\widetilde{\mathbf{M}}_2$ is given by equation (14), and $\widetilde{\mathbf{M}}_4$ is given by equations (17, 19).

The jump component of the condition mean $\widetilde{\mathbf{M}}_4(t)$ can also be given a discrete representation with respect to the jumps of the space-time Poisson process, so (17) becomes

$$\widetilde{\mathbf{M}}_4(t) = \sum_{j}^{N_1(t)} \boldsymbol{\Psi}(t,\tau_{1j})\mathbf{H}_1(\tau_{1j},\mathbf{U}_{1j}), \tag{19}$$

where N_1, τ_{1j} and $\mathbf{U}_{1j}$ are the number of jumps over the interval $[t_0,t)$, the jump time and mark vector of the system Poisson process, respectively. However, the jump times and marks have to be determined for the filter being approximated. To estimate the jump times and marks, the Maximum A Posteriori Estimation technique will be used (cf., Snyder and Miller [23]). In other words, one needs to maximize the likelihood function of $\{\boldsymbol{\eta}_1(s);\ t_0 < s \le t\}$ conditioned on the prior observations $\widetilde{\mathcal{Z}}_t = \{\widetilde{\mathbf{Z}}(s);\ t_0 < s \le t\}$.

The system Poisson input process, $\boldsymbol{\eta}_1(t)$ can be specified over $(t_0,t]$ by giving the number of jumps, $N_1(t)$, the set of ordered jump times, $\widetilde{\mathcal{T}}_1(t) = \{\tau_{1j};\ j = 1,\cdots,N_1(t)\}$ and the corresponding set of ordered mark vectors, $\widetilde{\mathcal{U}}_1(t) = \{\mathbf{u}_{1j};\ j = 1,\cdots,N_1(t)\}$. Therefore a likelihood function of $\boldsymbol{\eta}_1(t)$ conditioned on the prior observations $\widetilde{\mathcal{Z}}_t$ can be taken to be

$$\mathcal{L}(\boldsymbol{\eta}_1 \mid \widetilde{\mathcal{Z}}_t) = p\left(N_1(t), \widetilde{\mathcal{T}}_1(t), \widetilde{\mathcal{U}}_1(t) \mid \widetilde{\mathbf{Z}}(t) = \mathbf{z}\right),$$

since the system Poisson process, $\boldsymbol{\eta}_1(t)$ is white. By *Bayes' rule* relating mixed (i.e., continuous and discrete) conditional probabilities and since the mark vectors are stochastically independent of the jump times, the likelihood function of $\boldsymbol{\eta}_1(t)$ conditioned on the prior observations $\widetilde{\mathcal{Z}}_t$ is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\eta}_1 \mid \widetilde{\mathcal{Z}}_t) \;=\; & p\left(\widetilde{\mathbf{Z}}(t) \mid N_1(t) = n_1, \widetilde{\mathcal{T}}_1(t), \widetilde{\mathcal{U}}_1(t)\right) \cdot p\left(\widetilde{\mathcal{T}}_1(t) \mid N_1(t) = n_1\right) \\ & \cdot p\left(\widetilde{\mathcal{U}}_1(t) \mid N_1(t) = n_1\right) \cdot \mathrm{Prob}\left[N_1(t) = n_1\right] / p\left(\mathbf{z}, t\right). \quad (20)\end{aligned}$$

The observation probability density function $p(\mathbf{z}, t)$ is a function of realizations of observations and does not depend explicitly on $N_1(t)$, $\widetilde{\mathcal{T}}_1(t)$ and $\widetilde{\mathcal{U}}_1(t)$. Therefore, the denominator $p(\mathbf{z}, t)$ in (19) can be omitted and the likelihood function of $\boldsymbol{\eta}_1(t)$ conditioned on the prior observations can be replaced by:

$$\begin{aligned}\widetilde{\mathcal{L}}(\boldsymbol{\eta}_1 \mid \widetilde{\mathcal{Z}}_t) \;=\; & p\left(\widetilde{\mathbf{Z}}(t) \mid N_1(t) = n_1, \widetilde{\mathcal{T}}_1(t), \widetilde{\mathcal{U}}_1(t)\right) \cdot p\left(\widetilde{\mathcal{T}}_1(t) \mid N_1(t) = n_1\right) \\ & \cdot p\left(\widetilde{\mathcal{U}}_1(t) \mid N_1(t) = n_1\right) \cdot \mathrm{Prob}\left[N_1(t) = n_1\right].\end{aligned}$$

Further, by a simple application of Bayes' rule,

$$p\left(\widetilde{\mathcal{T}}_1(t) \mid N_1(t) = n_1\right) = \mathrm{Prob}\left[N_1(t) = n_1 \mid \widetilde{\mathcal{T}}_1(t)\right] p\left(\widetilde{\mathcal{T}}_1(t)\right) \mathrm{Prob}\left[N_1(t) = n_1\right],$$

so that

$$\begin{aligned}\widetilde{\mathcal{L}}(\boldsymbol{\eta}_1 \mid \widetilde{\mathcal{Z}}_t) \;=\; & p\left(\widetilde{\mathbf{Z}}(t) \mid N_1(t) = n_1, \widetilde{\mathcal{T}}_1(t), \widetilde{\mathcal{U}}_1(t)\right) p\left(\widetilde{\mathcal{U}}_1(t) \mid N_1(t) = n_1\right) \\ & \cdot \mathrm{Prob}\left[N_1(t) = n_1 \mid \widetilde{\mathcal{T}}_1(t)\right] p\left(\widetilde{\mathcal{T}}_1(t)\right).\end{aligned}$$

Observe that, when given the jump times $\widetilde{\mathcal{T}}_1(t)$, the

$$\mathrm{Prob}\left[N_1(t) = n_1 \mid \widetilde{\mathcal{T}}_1(t)\right] = \mathrm{Prob}\left[N_1(t) - N_1(\tau_{1,n_1}) = 0\right] = e^{-\lambda_1(t - \tau_{1,n_1})}.$$

Also, the density of the jump times according to Snyder and Miller [23] is given by

$$p(\widetilde{\mathcal{T}}_1(t)) = \lambda_1^{n_1} e^{-\lambda_1(\tau_{1,n_1} - t_0)} \mathcal{H}(t - \tau_{1,n_1}) \prod_{j=1}^{n_1} \mathcal{H}(\tau_{1,j} - \tau_{1,j-1}),$$

relying on the jump time ordering with $\tau_{1,0} = t_0$, where $\mathcal{H}(t)$ denotes the Heaviside step function. Moreover, since, by assumption, the mark vectors are mutually independent and identically distributed,

$$p\left(\widetilde{\mathcal{U}}_1(t) \mid N_1(t) = n_1\right) = \prod_{j=1}^{n_1} \phi_1(\mathbf{u}_{1j}).$$

Finally,

$$\widetilde{\mathcal{L}}(\boldsymbol{\eta}_1 \mid \widetilde{\mathcal{Z}}_t) \;=\; \lambda_1^{n_1} e^{-\lambda_1(t-t_0)} \left[\prod_{j=1}^{n_1} \phi_1(\mathbf{u}_{1j})\right] p\left(\widetilde{\mathbf{Z}}(t) \mid N_1(t) = n_1, \widetilde{\mathcal{T}}_1(t), \widetilde{\mathcal{U}}_1(t)\right),$$

where the domain defining Heaviside factors have been dropped assuming proper jump time ordering.

The likelihood of $\widetilde{\mathbf{Z}}(t)$ conditioned on both $\{\boldsymbol{\eta}_1(s);\ t_0 < s \le t\}$, and

$$\left\{\boldsymbol{\eta}_2(s) = \int_{t_0}^{t}\int_{\mathcal{U}_2} \mathbf{H}_2(t,\mathbf{U})\mathcal{P}_2(ds, d\mathbf{U});\ t_0 < s \le t\right\}$$

(see Au, Haddad and Poor [3]) is proportional to

$$\begin{aligned}\mathcal{J}(\widetilde{\mathbf{Z}} \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2) \;=\; \exp\Bigg(\int_{t_0}^{t} \Bigg\{ & \widetilde{\mathbf{M}}^T \mathbf{A}_2^T \mathbf{R}^{-1} \Bigg[d\widetilde{\mathbf{Z}}(s) - \mathbf{B}_2(\mathbf{X}^*, s)ds \\ & -\frac{1}{2}\mathbf{A}_2\widetilde{\mathbf{M}}(s)ds - \int_{\mathcal{U}_2} \mathbf{H}_2\mathcal{P}_2\Bigg]\Bigg\}\Bigg).\end{aligned}$$

Applying the expectation with respect to $\boldsymbol{\eta}_2(t)$ (see Hanson and Ryan [9]), one gets

$$\begin{aligned}\overline{\mathcal{J}}(\widetilde{\mathbf{Z}} \mid \boldsymbol{\eta}_1) \;=\; & \exp\left(\int_{t_0}^{t} \widetilde{\mathbf{M}}^T \mathbf{A}_2^T \mathbf{R}^{-1}\left[d\widetilde{\mathbf{Z}}(s) - \mathbf{B}_2(\mathbf{X}^*, s)ds - \frac{1}{2}\mathbf{A}_2\widetilde{\mathbf{M}}(s)ds\right]\right) \\ & \cdot \exp\left(\lambda_2\left[\int_{t_0}^{t}\int_{\mathcal{U}_2} \exp\left(-\widetilde{\mathbf{M}}^T\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{H}_2\right)\phi_2(\mathbf{u})d\mathbf{u}ds - (t-t_0)\right]\right).\end{aligned}$$

Therefore, the log-likelihood function can be taken to be

$$\begin{aligned}\widetilde{\mathcal{J}}(\widetilde{\mathbf{Z}} \mid \boldsymbol{\eta}_1) \;=\; & \int_{t_0}^{t} \widetilde{\mathbf{M}}^T \mathbf{A}_2^T \mathbf{R}^{-1}\left[d\widetilde{\mathbf{Z}}(s) - \mathbf{B}_2(\mathbf{X}^*, s)ds - \frac{1}{2}\mathbf{A}_2\widetilde{\mathbf{M}}(s)ds\right] \\ & +\lambda_2 \int_{t_0}^{t}\int_{\mathcal{U}_2} \exp\left(-\widetilde{\mathbf{M}}^T\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{H}_2\right)\phi_2(\mathbf{u})d\mathbf{u}ds - (t-t_0).\end{aligned}$$

Note that in $\widetilde{\mathbf{M}}(t)$ of (18), $\widetilde{\mathbf{M}}_2(t)$ of (14) does not depend upon the jump parameter sets, $\widetilde{\mathcal{T}}_1(t)$ and $\widetilde{\mathcal{U}}_1(t)$, since it depends on $\mathrm{E}[d\boldsymbol{\eta}_2(t)]$ rather than $d\boldsymbol{\eta}_1(t)$, but the dependence on $d\boldsymbol{\eta}_1(t)$ is only in $\widetilde{\mathbf{M}}_4(t)$ of (17).

Since the nonlinearities are Lipschitz continuous in $\mathbf{X}$, then as $\mathbf{X}^*(t)$ approaches $\widetilde{\mathbf{X}}(t)$, the approximate description, approaches the original system. Finally, choosing $\mathbf{X}^*(t)$ to be $\widetilde{\mathbf{M}}_2^{(1)}(t)$ from (14), the suboptimal filter is given by equation (18), where $\widetilde{\mathbf{M}}_4$ is given by (17, 19), and reusing (14) to obtain another iteration,

$$\begin{aligned}d\widetilde{\mathbf{M}}_2^{(2)}(t) \;=\; & \left[\mathbf{A}_1 - \widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{A}_2\right]\widetilde{\mathbf{M}}_2^{(2)}dt \\ & + \left[\mathbf{B}_1(\widetilde{\mathbf{M}}_2^{(1)}, t) - \widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1}\mathbf{B}_1(\widetilde{\mathbf{M}}_2^{(1)}, t)\right]dt \\ & +\widetilde{\mathbf{P}}\mathbf{A}_2^T\mathbf{R}^{-1} \cdot \left\{d\widetilde{\mathbf{Z}}(t) - \lambda_2\int_{\mathcal{U}_2}\mathbf{H}_2\phi_2(\mathbf{u})d\mathbf{u}dt\right\},\end{aligned} \tag{21}$$

with $\widetilde{\mathbf{P}}$ satisfying (11).

The jump parameter sets $\widetilde{\mathcal{T}}_1(t)$ and $\widetilde{\mathcal{U}}_1(t)$ are determined by maximizing

$$\begin{aligned}\widetilde{\mathcal{J}}^{(2)}(\widetilde{\mathbf{Z}} \mid \boldsymbol{\eta}_1) = \textstyle\int_{t_0}^{t} \widetilde{\mathbf{M}}^T \mathbf{A}_2^T \mathbf{R}^{-1} \left[d\widetilde{\mathbf{Z}}(s) - \mathbf{B}_2(\widetilde{\mathbf{M}}_2^{(2)}, s)ds - \frac{1}{2}\mathbf{A}_2 \widetilde{\mathbf{M}}^{(2)}(s)ds \right] \\ + \lambda_2 \textstyle\int_{t_0}^{t} \int_{\mathcal{U}_2} \exp\left(-\widetilde{\mathbf{M}}^T \mathbf{A}_2^T \mathbf{R}^{-1} \mathbf{H}_2 \right) \phi_2(\mathbf{u}) d\mathbf{u} ds - (t - t_0).\end{aligned} \tag{22}$$

4. Numerical Experiments

Partitioning the interval, $(t_0, t]$ into subintervals of length Δ such that:

$$\lambda_1 \Delta \ll 1, \tag{23}$$

so that the probability of more than one jump is negligible in a Δ-interval. Further, let $\hat{N}_1,\ \hat{\tau}_1$, and $\hat{u}_1$ denote the MAP estimate of the one jump parameters $N_1,\ \tau_1$, and u_1, respectively. Then

$$\begin{aligned}\widehat{\mathcal{J}}[\hat{N}_1, \hat{\tau}_1, \hat{u}_1] &= \max_{N_1, \tau_1, u_1} \left[\widehat{\mathcal{J}}[N_1, \tau_1, u_1] \right] \\ &= \max_{N_1} \left[\max_{\tau_1, u_1} \left[\widehat{\mathcal{J}}[N_1, \tau_1, u_1] \right] \right] \\ &= \max_{N_1} \left[\widehat{\mathcal{J}}[N_1, \hat{\tau}_1, \hat{u}_1] \right].\end{aligned}$$

This means one can assume a value for N_1 and maximize with respect to τ_1 and u_1, since τ_1 and u_1 are pre-conditioned on the existence of jumps. With high probability, the number of jumps in each subinterval, $[t_n, t_n + \Delta)$ can take one of only two values 0 or 1 for sufficiently small Δ. Therefore, J needs to be maximized only for one set of τ_1 and u_1 on each subinterval.

The testing data are generated by a Runge-Kutta scheme for Itô SDEs due to N. Newton [19], modified by Kloeden and Platen [13], and modified again here for Poisson noise. The Wiener noise is generated by means of *Box-Muller transformation method*, whereas the Poisson noise is generated via a *rejection method* (see Press et al. [20]).

The likelihood function $\widehat{\mathcal{J}}$ is maximized by a form of the *conjugate-direction method* of Powell with modifications due to Brent combined with *golden section search* [20]. This modified form avoids the computation of the derivatives.

Four single-state and single-observation dimensional examples were used to test the suboptimal filter. The first illustrate the linear case with stationary coefficients. The second also deals with linear case, but the coefficients are nonstationary. The third and the fourth illustrate the nonlinear, nonstationary case.

- Example 1 ($\lambda_1 = 1.5;\ \ \lambda_2 = 3.0$):

$$\begin{aligned}dX(t) &= -4X(t)dt + 0.05dW_1(t) + 0.5dP_1;\ X(0) = 0, \\ dZ(t) &= X(t)dt + dW_2(t) + dP_2(t);\ Z(0) = 0.\end{aligned}$$

- Example 2 ($\lambda_1 = 0.1;\ \ \lambda_2 = 2.5$):

$$\begin{aligned} dX(t) &= e^{-t}X(t)dt + t^2 dW_1(t) + t/2dP_1;\ X(0) = 0, \\ Z(t) &= tX(t)dt + (5+t^2)dW_2(t) + (1+t)dP_2(t);\ Z(0) = 0. \end{aligned}$$

- Example 3 ($\lambda_1 = 6.5;\ \ \lambda_2 = 5.5$):

$$\begin{aligned} dX(t) &= (X + t\sin(X))dt + tdW_1(t) + t^2 dP_1(t);\ X(0) = 0, \\ dZ(t) &= tXdt + dW_2(t) + dP_2(t);\ Z(0) = 0. \end{aligned}$$

- Example 4 ($\lambda_1 = 3.4;\ \ \lambda_2 = 1.2$):

$$\begin{aligned} dX(t) &= X(1 + X\cos(X))dt + dP_1(t);\ X(0) = 0, \\ dZ(t) &= X\left(1 + \frac{X}{\exp(t) + X}\right)dt + dW_2(t) + dP_2(t);\ Z(0) = 0. \end{aligned}$$

The suboptimal filter was solved numerically on an IBM ES/3090VF 300J main frame using an asymptotically consistent, stochastic Runge-Kutta method originally due to N. Newton [19]. For random numbers generated on the computer, the performance of the filter was extremely good, provided the jump times are known or predicted correctly. First, the jumps are detected and then estimated using the maximum a posteriori technique. After that the state vector is estimated using the optimal filter equations conditioned on the system Poisson process.

The simulated state process $X(t)$, the suboptimal filter M_a with the jump times *assumed known*, and another suboptimal filter M_b with jump times *estimated* are compared in Fig. 1 for the linear, stationary coefficient Example 1. All results for $X(t)$, M_a, and M_b were obtained by stochastic Runge-Kutta methods. Observe that the estimates M_a and M_b are very close to the state $X(t)$ simulations. Also there are few discrepancies in jump times of M_a and M_b.

In Fig. 2, the corresponding results for the linear, nonstationary coefficient Example 2 are somewhat different from the prior example, in that the estimate M_a with given jump times is not distinguishable from the state simulations X, but the estimate M_b, with estimated jump times, exhibits an essentially zero response. In this linear, nonstationary coefficient example, the detection part is very weak.

In Fig. 3, the corresponding results for nonlinear, nonstationary coefficient Example 3 again exhibit good correspondence between the state computation $X(t)$ and the state estimates M_a and M_b, except for a slow growth in overestimation for long times. Both suboptimal estimates work well in this case, independent of whether the jump times are prescribed or estimated.

Finally in Fig. 4, the results for the other nonlinear, nonstationary coefficient Example 4 exhibit good agreement between the state computation $X(t)$ and the state estimates M_a and M_b only for short times, while both suboptimal filters predict a similar underestimate of the state for longer times. This example shows that our suboptimal estimation method relies on the importance of dominance of the linear terms. Thus both suboptimal estimates fail to track the state for long times. The nearly quadratic nonlinearity of Example 4 is more dominant than the nearly linear nonlinearity of Example 3, so that the tracking failure of the estimates for Example 4 are greater.

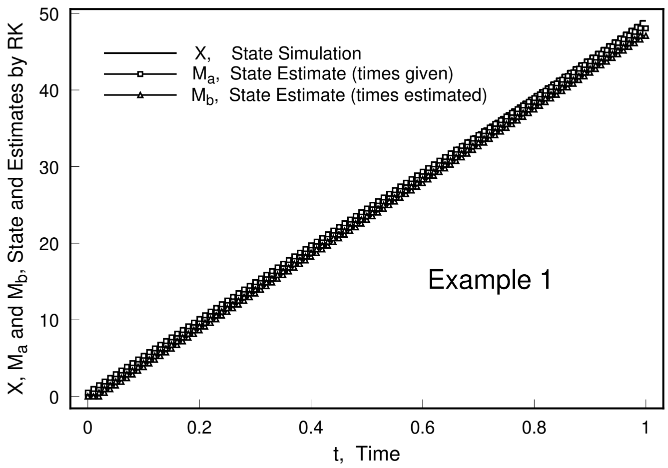

Figure 1. Comparison of state simulation and computational state estimates with jump times given and estimated, respectively, for Example 1, the linear case with stationary coefficients.

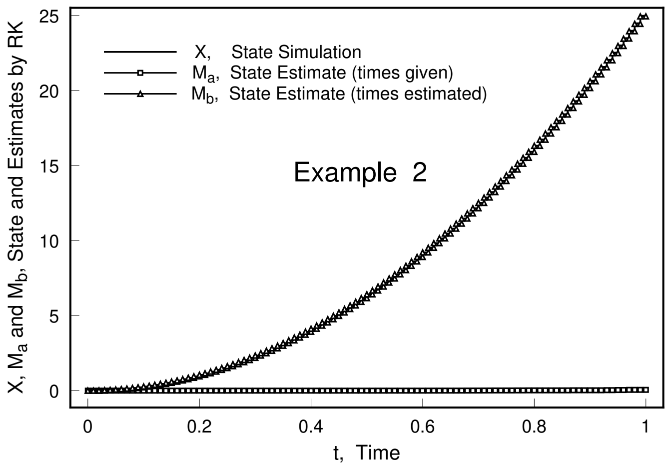

Figure 2. Comparison of state simulation and computational state estimates with jump times given and estimated, respectively, for Example 2, also a linear case but with nonstationary coefficients.

5. Summary and Conclusions

The problem of state estimation observed continuously in an uncertain jump environment is the main theme of this research. In many real problems, the state can not be observed directly and both the system and observations processes are governed by nonlinear SDEs.

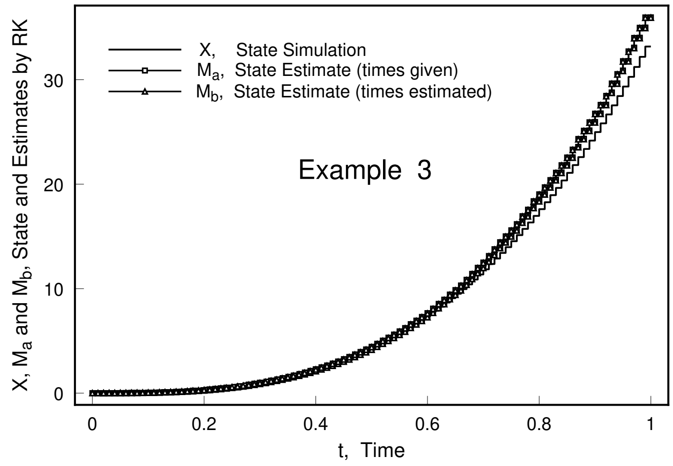

Figure 3. Comparison of state simulation and computational state estimates with jump times given and estimated, respectively, for Example 3, a nonlinear case with nonstationary coefficients.

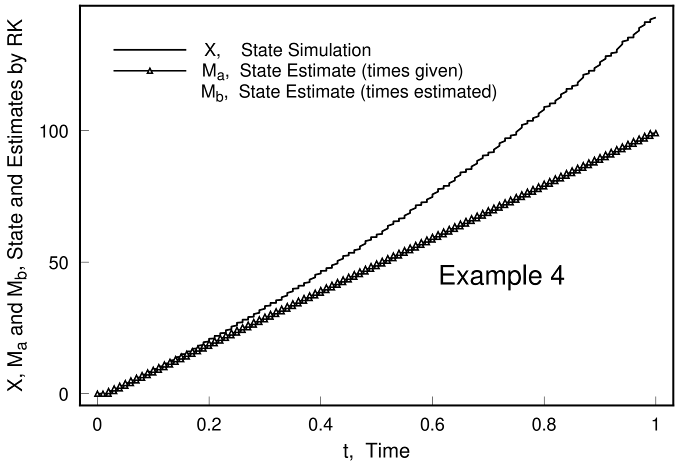

Figure 4. Comparison of state simulation and computational state estimates with jump times given and estimated, respectively, for Example 4, another nonlinear case with nonstationary coefficients.

Earlier contributors solved the problem in the case of Gaussian-driven or Gaussian-Poisson driven systems with observations contaminated by Gaussian noise only. Here we solved the case when both the system and observations are driven by Wiener-Poisson processes. The MMSE optimal filter is not closed in the sense that it is the solution of a set of an infinite number of equations, necessitating approximate techniques.

The need for a computationally implementable, approximate filter in real world applications was the motivation for the present work. The equations satisfied by the optimal filter were used and derived formally a computational but suboptimal filter conditioned on the partially observed system Poisson process. It is a combination of detection and estimation. First, the jumps are detected and then estimated using the MAP technique. After that the state vector is estimated using the optimal filter equations conditioned on the system Poisson process applied to an approximation of the system and observation SDEs.

However, the false prediction of jump times affected the performance heavily. In addition, some discrepancies in the estimates arise when the nonlinear part significantly dominates, rather than the linear part dominating. This leaves the question for a better prediction technique open. Otherwise, the current method works quite well.

Acknowledgments

We thank an anonymous reviewer for constructive and helpful criticisms. This research was supported in part by the National Science Foundation under Computational Mathematics Program Grants NSF DMS 93-01107 and 96-26692, and in part by the Birzeit University under grant BZU 235-17-99-9.

References

1. R. M. Abu-Saris, "Filtering of a Wiener-Poisson driven stochastic process," Ph.D. Thesis, University of Illinois, Chicago, 1994.
2. N. U. Ahmed and S. M. Radaideh, "Modified Extended Kalman Filtering," *IEEE Trans. Automat. Contr.*, vol. 39 (6), June 1994, pp. 1322–1326.
3. S. P. Au, A. H. Haddad and H. V. Poor, "A state estimation algorithm for linear systems driven simultaneously by Wiener and Poisson processes," *IEEE Trans. Automat. Contr.*, vol. 27 (3), June 1982, pp. 617–626.
4. B. A. Cipra, "Engineers look to Kalman filtering for guidance," *SIAM News*, vol. 26, August 1993, pp. 8–9.
5. H. E. Emara-Shabaik, "A new approach for developing practical filters for nonlinear systems," *Control and Dynamic Systems: Advances in Theory and Applications*, vol. 21, 1984, pp. 297–335.
6. J. J. Florentin, "Optimal control of systems with generalized Poisson inputs," *J. Basic Engr. Tran. ASME*, vol. 38, June 1963, pp. 217–221.
7. I. I. Gihman, and A. V. Skorohod, *Stochastic Differential Equations*, Springer-Verlag:New York, NY, 1972.
8. M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice*, Prentice-Hall:Englewood Cliffs, NJ, 1993.
9. F. B. Hanson and D. Ryan, "Mean and quasi-deterministic equivalence for linear stochastic dynamics," *Math. Biosc.*, vol. 93, 1989, pp. 1–14.
10. A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press:New York, NY, 1970.
11. R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Engr. Trans. ASME*, ser. D, vol. 35, March 1960, pp. 35–45.
12. R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *J. Basic Engr. Trans. ASME*, ser. D, vol. 36, March 1961, pp. 95–108.
13. P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag:New York, NY, 1992.
14. H. J. Kushner and P. G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag:New York, NY, 1992.

15. H. Kwakernaak, “Filtering for systems excited by Poisson white noise,” *Control Theory, Numerical Methods and Computer Systems Modeling*, vol. 107, 1975, pp. 468–492.
16. D. F. Liang, “Exact and approximate state estimation techniques for nonlinear dynamic systems,” *Control and Dynamic Systems: Advances in Theory and Applications*, vol. 19, 1983, pp. 1–80.
17. D. F. Liang, “Comparisons of nonlinear recursive filters for systems with nonnegligible nonlinearities,” *Control and Dynamic Systems: Advances in Theory and Applications*, vol. 20, 1983, pp. 341–401.
18. M. I. Miller, R. Teichman, A. Srivastava, J. A. O’Sullivan and D. Snyder, “Jump-diffusion processes for automated tracking-target recognition,” Electronic System Signals Research Laboratory, Washington University, St. Louis, MO, Research Monograph ESSRL-93-6, 1993.
19. N. J. Newton, “Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations,” *SIAM J. Appl. Math.*, vol. 51 (2), April 1991, pp. 542–567.
20. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in FORTRAN: the Art of Scientific Computing*, Cambridge University Press:Cambridge, UK, 1992.
21. G. T. Schmidt, “Linear and nonlinear filtering techniques,” *Control and Dynamic Systems: Advances in Theory and Applications*, vol. 12, 1976, pp. 63–98.
22. L. Schwartz, “Approximate continuous non-linear minimal-variance filtering,” *Advances in Control Systems: Theory and Applications*, vol. 7, 1969, pp. 31–72.
23. D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, 2nd edition, Springer-Verlag:New York, NY, 1991.