

第一页为封面页

参赛学生姓名：Lillian Sun

中学：武汉爱莎文华高级中学

省份：湖北

国家/地区：中华人民共和国

指导老师姓名：刘显明

指导老师单位：华中科技大学

论文题目：Nonlinear filtering for stochastic
dynamical systems with memory: A new
framework

Nonlinear filtering for stochastic dynamical systems with memory: A new framework

Lillian Sun

ISA Wenhua Wuhan School, Wuhan 430119, Hubei, China

September 15, 2026

Abstract: Existing nonlinear filtering algorithms have been developed primarily for Markov systems. They rely critically on solving the Duncan–Mortensen–Zakai (DMZ) equation, which governs the dynamical evolution of the conditional density of the state given the observation history. For stochastic dynamical systems with memory, however, no governing equation analogous to the DMZ equation exists for the conditional density. Consequently, existing algorithms cannot be applied to such systems, and estimating the conditional density of the state from observation-history data remains a challenging open problem. We propose a new nonlinear filtering framework for stochastic dynamical systems with memory. In particular, we establish a new algorithm for computing the conditional density of the state given the observation history, from which the optimal state estimate can be extracted. Numerical examples are provided to illustrate the proposed algorithm and verify its performance.

Keywords: Nonlinear filtering; Stochastic dynamical systems; state estimation; conditional probability density; non-Markovian; Memory

Contents

1	Introduction	3
2	Derivation Of the Algorithm	4
2.1	Preparation	5
2.2	Derivation	6
3	Examples	10
4	Concluding remarks	13

1 Introduction

We consider the following class of stochastic systems with memory [1],

$$\begin{cases} dX(t) = f(X(t), X(t - \tau_1), \dots, X(t - \tau_n)) dt + g(X(t), X(t - \tau_1), \dots, X(t - \tau_n)) dB(t), \\ X(s) = \gamma(s) \quad \text{for } s \in [-\tau, 0], \\ dY(t) = h(X(t), X(t - \tau_1), \dots, X(t - \tau_n)) dt + r(X(t), X(t - \tau_1), \dots, X(t - \tau_n)) dV(t), \\ Y(0) = y_0, \end{cases} \quad (1.1)$$

where $X(t)$ is a $\mathbb{R}^{d_1}$ -valued stochastic process denoting the system state, $Y(t)$ is a $\mathbb{R}^{d_2}$ -valued stochastic process representing the observations. The constants τ_i ($i = 1, 2, \dots, n$) are time delays. The functions f and g are the drift and diffusion coefficients of the state equation, respectively, defined on $\mathbb{R}^{(n+1)d_1}$. h and r are the corresponding coefficients for the observation equation, defined on $\mathbb{R}^{(n+1)d_2}$.

The objective of this paper is to develop a nonlinear filtering algorithm for (1.1). More specifically, we aim to develop a recursive algorithm for computing the conditional density of the state $X(t)$ given the observation history $Y(s) = y(s)$, $0 \leq s \leq t$, from which the optimal state estimate can be extracted.

Existing nonlinear filtering algorithms [2, 3] are mainly developed for stochastic systems without memory,

$$\begin{cases} dX(t) = f(X(t)) dt + g(X(t)) dB(t), \\ dY(t) = h(X(t)) dt + r(X(t)) dV(t), \end{cases} \quad (1.2)$$

which is a special case of (1.1) with no delays. Various techniques exist for estimating the conditional density of $X(t)$ given the observation history up to time t for (1.2). Most of these techniques, including the Yau–Yau algorithm [3, 4], involve solving the Duncan–Mortensen–Zakai (DMZ) equation, which governs the time evolution of the conditional density for the delay-free case.

However, these existing techniques are not directly applicable to (1.1), as no governing equation analogous to the DMZ equation exists to describe the conditional density for systems with memory. To the best of our knowledge, no method for estimating the conditional density has yet been developed, and this remains a challenging open problem.

Throughout the paper, unless we state otherwise, we assume that (i) f , g , h and r are Lipschitz continuous and (ii) g and r satisfy the strong elliptic condition, i.e., there exists a constant $\epsilon > 0$ such that $gg^T \geq \epsilon I$ and $rr^T \geq \epsilon I$, where I is the identity matrix. Note that under assumptions (i) and (ii), (1.1) has a unique, strong solution, and the probability density for the solution exists and is continuous, see [5, 6], among others, for a more detailed discussion.

In this paper, we propose a recursive algorithm to compute the conditional density, from which the optimal state estimate is derived. The derivation is presented for the general case with delays in both the state and observation equations; nevertheless, the algorithm naturally applies to simpler settings—when a particular delay is absent, the corresponding functions simply omit that argument, while the recursive structure remains unchanged. We also note that assumptions (i) and (ii) can be relaxed, though such a generalization is not pursued here.

We shall establish a new nonlinear filtering framework for stochastic systems with memory, as given in (1.1), by deriving a recursive algorithm for the conditional density of the state given the observation history. The rest of this paper is structured as follows. Section 2 provides a detailed derivation of the recursive algorithm. Section 3 presents numerical examples that demonstrate the application of the proposed method and validate its performance. Finally, Section 4 offers concluding remarks.

2 Derivation Of the Algorithm

For brevity, we restrict our attention to the case of (1.1) with $d_1 = d_2 = 1$ and $n = 2$, i.e., we consider a system with two distinct delays τ_1 and τ_2 , a scalar state $X(t)$, and a scalar observation $Y(t)$. The derivation for more general n , d_1 and d_2 is essentially the same and is therefore omitted.

2.1 Preparation

Taking $n = 2$, $d_1 = d_2 = 1$, (1.1) becomes

$$\begin{cases} dX(t) = f(X(t), X(t - \tau_1), X(t - \tau_2)) dt + g(X(t), X(t - \tau_1), X(t - \tau_2)) dB(t), \\ X(s) = \gamma(s) \quad \text{for } s \in [-\tau, 0], \\ dY(t) = h(X(t), X(t - \tau_1), X(t - \tau_2)) dt + r(X(t), X(t - \tau_1), X(t - \tau_2)) dV(t), \\ Y(0) = y_0, \end{cases} \quad (2.1)$$

where $X(t)$ and $Y(t)$ are scalar-valued stochastic processes, and $\tau = \max(\tau_1, \tau_2)$.

Let t_j , $j = 0, 1, \dots, N$ be $N + 1$ discrete time points on $[0, T]$ with $t_j = j\Delta t$, $\Delta t = \frac{T}{N}$.

Applying the Euler-Maruyama scheme to (2.1) yields the discrete-time approximation:

$$\begin{cases} \hat{X}(t_{k+1}) = \hat{X}(t_k) + f(\hat{X}(t_k), \hat{X}(t_{k-m_1}), \hat{X}(t_{k-m_2})) \Delta t + g(\hat{X}(t_k), \hat{X}(t_{k-m_1}), \hat{X}(t_{k-m_2})) \Delta B_k, \\ \text{for } k = 0, 1, 2, \dots, N-1, \\ X(-t_k) = \gamma(-t_k) \quad \text{for } k = 0, 1, \dots, m, \\ \hat{Y}(t_{k+1}) = \hat{Y}(t_k) + h(\hat{X}(t_k), \hat{X}(t_{k-m_1}), \hat{X}(t_{k-m_2})) \Delta t + r(\hat{X}(t_k), \hat{X}(t_{k-m_1}), \hat{X}(t_{k-m_2})) \Delta V_k, \\ \text{for } k = 0, 1, 2, \dots, N-1, \\ \hat{Y}(0) = y_0, \end{cases} \quad (2.2)$$

where $m_1 = \lfloor \frac{\tau_1}{\Delta t} \rfloor$, $m_2 = \lfloor \frac{\tau_2}{\Delta t} \rfloor$, $m = \max(m_1, m_2)$, $\Delta B_k = B(t_{k+1}) - B(t_k)$, $\Delta V_k = V(t_{k+1}) - V(t_k)$. Here $\lfloor \cdot \rfloor$ denotes the floor function.

Define $\hat{X}(t)$, $\hat{Y}(t)$ as the linear interpolation of $\hat{X}(t_k)$, $\hat{Y}(t_k)$ respectively ($k = 0, 1, \dots, N$), i.e.,

$$\begin{cases} \hat{X}(t) = \sum_{i=0}^{N-1} \left\{ \left(\hat{X}(t_i) \frac{t_{i+1}-t}{\Delta t} + \hat{X}(t_{i+1}) \frac{t-t_i}{\Delta t} \right) \mathbf{1}_{(t_i, t_{i+1}]}(t) \right\}, \\ \hat{Y}(t) = \sum_{i=0}^{N-1} \left\{ \left(\hat{Y}(t_i) \frac{t_{i+1}-t}{\Delta t} + \hat{Y}(t_{i+1}) \frac{t-t_i}{\Delta t} \right) \mathbf{1}_{(t_i, t_{i+1}]}(t) \right\}. \end{cases} \quad (2.3)$$

Denote $\rho(x, y, t)$ as the joint probability density of $X(t)$ and $Y(t)$ and $\hat{\rho}(x, y, t)$ as the joint probability density of $\hat{X}(t)$ and $\hat{Y}(t)$. Both the weak and strong convergence of $(\hat{X}(t), \hat{Y}(t))$ to the original solution $(X(t), Y(t))$ has been established, see [7, 8, 9], among others, for

more details. Under the condition that the probability density exists and is continuous, either weak or strong convergence as mentioned above guarantees the convergence of the probability density, i.e., $\forall x, y \in \mathbb{R}$ and $\forall t \in (0, T]$,

$$|\rho(x, y, t) - \hat{\rho}(x, y, t)| \rightarrow 0 \quad \text{as} \quad \Delta t \rightarrow 0. \quad (2.4)$$

2.2 Derivation

There is more than one way to define the optimal state estimate given the observation history for system (2.1). In this paper, we take the conditional expectation of the state $X(t)$ given the observation history $Y(s) = y(s)$, $0 \leq s \leq t$, as the optimal estimate. Expression (2.4) implies that for a sufficient small Δt , the conditional expectation of $X(t)$ given $Y(s) = y(s)$, $0 \leq s \leq t$, can be estimated by the conditional expectation of $\hat{X}(t)$ given $\hat{Y}(s) = y(s)$, $0 \leq s \leq t$. Since $\hat{X}(t)$ and $\hat{Y}(t)$ are linear interpolations of $\hat{X}(t_k)$ and $\hat{Y}(t_k)$, respectively, the conditional expectation of $\hat{X}(t)$ given $\hat{Y}(s) = y(s)$, $0 \leq s \leq t$, can be approximated by that of $\hat{X}(t)$ given $\hat{Y}(t_i) = y(t_i)$, $i = 0, 1, \dots, k$, where $k = \lfloor \frac{t}{\Delta t} \rfloor$. For brevity, from now on we write y_i for $y(t_i)$.

Therefore, the optimal estimate of $X(T)$ given observation history $Y(s) = y(s)$, $0 \leq s \leq T$ can be expressed as

$$\begin{aligned} E(X(T)|Y(s) = y(s), 0 \leq s \leq T) &\approx E(\hat{X}(T)|\hat{Y}(s) = y(s), 0 \leq s \leq T) \\ &= E(\hat{X}(t_N)|y_0, t_0; \dots; y_N, t_N) \\ &= \int_{\mathbb{R}} x_k \hat{p}(x_N, t_N|y_0, t_0; \dots; y_N, t_N) dx_N, \end{aligned} \quad (2.5)$$

where $\hat{p}(x_k, t_k|y_0, t_0; \dots; y_k, t_k)$ denotes the conditional probability density of $\hat{X}(t_k)$ at $\hat{X}(t_k) = x_k$ given observation history $\hat{Y}(t_i) = y_i$, $i = 0, 1, \dots, k$.

It is straightforward to check that

$$\hat{p}(x_k, t_k|y_0, t_0; \dots; y_k, t_k) = \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \hat{p}(x_1, t_1; \dots; x_k, t_k|y_0, t_0; \dots; y_k, t_k) dx_1 \dots dx_{k-1}, \quad (2.6)$$

and

$$\begin{aligned}
& \hat{p}(x_1, t_1; \dots; x_k, t_k | y_0, t_0; \dots; y_k, t_k) \\
&= \frac{\hat{p}(x_k, t_k; y_k, t_k | x_1, t_1; \dots; x_{k-1}, t_{k-1}; y_0, t_0; \dots, y_{k-1}, t_{k-1})}{\hat{p}(y_k, t_k | y_0, t_0; \dots; y_{k-1}, t_{k-1})} \\
& \times \hat{p}(x_1, t_1; \dots; x_{k-1}, t_{k-1} | y_0, t_0; \dots; y_{k-1}, t_{k-1}). \tag{2.7}
\end{aligned}$$

The numerator of the fraction on the right hand side of (2.7) represents the conditional joint density of $(\hat{X}(t_k), \hat{Y}(t_k))$ at $\hat{X}(t_k) = x_k$ and $\hat{Y}(t_k) = y_k$ given the state history $\hat{X}(t_i) = x_i, i = 1, 2, \dots, N$ and the observation history $\hat{Y}(t_i) = y_i, i = 1, 2, \dots, k$. By (2.2), this conditional joint density can be expressed explicitly as

$$\begin{aligned}
& \hat{p}(x_k, t_k; y_k, t_k | x_1, t_1; \dots; x_{k-1}, t_{k-1}; y_0, t_0; \dots; y_{k-1}, t_{k-1}) \\
&= \hat{p}(x_k, t_k; y_k, t_k | x_{k-1-m_2}, t_{k-1-m_2}; x_{k-1-m_1}, t_{k-1-m_1}; x_{k-1}, t_{k-1}; y_{k-1}, t_{k-1}) \\
&= \frac{1}{\sqrt{2\pi\Delta t} g^2(x_{k-1}, x_{k-1-m_1}, x_{k-1-m_2}) r^2(x_{k-1}, x_{k-1-m_1}, x_{k-1-m_2})} \\
& \times \exp \left\{ -\frac{(x_k - x_{k-1} - f(x_{k-1}, x_{k-1-m_1}, x_{k-1-m_2})\Delta t)^2}{2g^2(x_{k-1}, x_{k-1-m_1}, x_{k-1-m_2})\Delta t} \right\} \\
& \times \exp \left\{ -\frac{(y_k - y_{k-1} - r(x_{k-1}, x_{k-1-m_1}, x_{k-1-m_2})\Delta t)^2}{2r^2(x_{k-1}, x_{k-1-m_1}, x_{k-1-m_2})\Delta t} \right\}. \tag{2.8}
\end{aligned}$$

Define

$$I_k = \begin{cases} \hat{p}(x_1, t_1; \dots; x_k, t_k | y_0, t_0; \dots; y_k, t_k) & \text{for } 1 \leq k \leq m+1, \\ \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \hat{p}(x_1, t_1; \dots; x_k, t_k | y_0, t_0; \dots; y_k, t_k) dx_1 \dots dx_{k-1-m} & \text{for } k \geq m+2. \end{cases} \tag{2.9}$$

Note that I_k is itself a conditional joint density. In fact, for $1 \leq k \leq m+1$, I_k is the conditional joint density of $(\hat{X}(t_1), \dots, \hat{X}(t_k))$ given observation history $\hat{Y}(t_i) = y_i, i = 1, 2, \dots, k$. For $k \geq m+2$, I_k is the conditional joint density of $(\hat{X}(t_{k-m}), \dots, \hat{X}(t_k))$ given the same observation history.

Substituting (2.9) into (2.6), we get

$$\hat{p}(x_k, t_k | y_0, t_0; \dots; y_k, t_k) = \begin{cases} \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} I_k dx_1 \dots dx_{k-1} & \text{for } 1 \leq k \leq m+1, \\ \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} I_k dx_{k-m} \dots dx_{k-1} & \text{for } k \geq m+2. \end{cases} \tag{2.10}$$

It follows from (2.7) and (2.9) that I_k can be expressed recursively as

$$I_k = \begin{cases} \hat{p}(x_1, t_1 | y_0, t_0; y_1, t_1) = \frac{\hat{p}(x_1, t_1; y_1, t_1 | y_0, t_0)}{\hat{p}(y_1, t_1 | y_0, t_0)} & \text{for } k = 1, \\ \frac{\hat{p}(x_k, t_k; y_k, t_k | x_{k-1-m_2}, t_{k-1-m_2}; x_{k-1-m_1}, t_{k-1-m_1}; x_{k-1}, t_{k-1}; y_{k-1}, t_{k-1})}{\hat{p}(y_k, t_k | y_0, t_0; \dots; y_{k-1}, t_{k-1})} I_{k-1} & \text{for } 1 \leq k \leq m+1, \\ \int_{\mathbb{R}} \frac{\hat{p}(x_k, t_k; y_k, t_k | x_{k-1-m_2}, t_{k-1-m_2}; x_{k-1-m_1}, t_{k-1-m_1}; x_{k-1}, t_{k-1}; y_{k-1}, t_{k-1})}{\hat{p}(y_k, t_k | y_0, t_0; \dots; y_{k-1}, t_{k-1})} I_{k-1} dx_{k-1-m} & \text{for } k \geq m+2, \end{cases} \quad (2.11)$$

For each fraction appearing on the right-hand side of the above equation, the numerator is a joint transition density that we have already computed in (2.8). On the other hand, the denominator, which is constant because $y_0, y_1, \dots, y_k$ are given observations, is not as easy to compute directly. Fortunately, we do not need to compute it. To see why, recall that we noted earlier that I_k is a conditional joint density. The denominator is merely a normalization constant whose only purpose is to normalize. Hence, we can ignore the denominator, as long as we remember to normalize the density eventually.

From (2.5), we see that computing the conditional expectation of $X(t)$ requires only the conditional probability density of $\hat{X}(t)$. From (2.10), computing this conditional probability density requires only I_k . From (2.11), I_k can be computed recursively.

Now we are ready to formally present the proposed algorithm to compute the conditional expectation of $X(t)$. For numerical evaluation, we replace the integration domain $\mathbb{R}$ with the interval $[-D, D]$, where D is a sufficiently large constant. Furthermore, we select M space points $z_0, z_1, \dots, z_{M-1}$ from $[-D, D]$ to evaluate our integrals by Euler scheme, where $z_i = -D + i\Delta x$ for $i = 0, 1, 2, \dots, M-1$ and $\Delta x = \frac{2D}{M}$. Applying these changes to (2.5), (2.10), (2.11) we get

$$\begin{aligned} E(X(T) | Y(s), 0 \leq s \leq T) &\approx E(\hat{X}(T) | \hat{Y}(s), 0 \leq s \leq T) \\ &= E(\hat{X}(t_N) | y_0, t_0; \dots; y_N, t_N) \\ &= \Delta x \times \sum_{i_N=0}^{M-1} z_{i_N} \hat{p}(z_{i_N}, t_N | y_0, t_0; \dots; y_N, t_N), \end{aligned} \quad (2.12)$$

$$\hat{p}(x_k, t_k | y_0, t_0; \dots; y_k, t_k) = \begin{cases} \Delta x^{k-1} \times \sum_{i_{k-1}=0}^{M-1} \sum_{i_{k-2}=0}^{M-1} \dots \sum_{i_1=0}^{M-1} I_k(z_{i_1}, \dots, z_{i_{k-1}}, x_k) & \text{for } 1 \leq k \leq m+1, \\ \Delta x^{k-1} \times \sum_{i_{k-1}=0}^{M-1} \sum_{i_{k-2}=0}^{M-1} \dots \sum_{i_{k-m}=0}^{M-1} I_k(z_{i_{k-m}}, \dots, z_{i_{k-1}}, x_k) & \text{for } k \geq m+2, \end{cases} \quad (2.13)$$

$$I_k = \begin{cases} \frac{\hat{p}(x_1, t_1; y_1, t_1 | y_0, t_0)}{\hat{p}(y_1, t_1 | y_0, t_0)} & \text{for } k = 1, \\ \frac{\hat{p}(x_k, t_k; y_k, t_k | x_{k-1-m_2}, t_{k-1-m_2}; x_{k-1-m_1}, t_{k-1-m_1}; x_{k-1}, t_{k-1}; y_{k-1}, t_{k-1})}{\hat{p}(y_k, t_k | y_0, t_0; \dots; y_{k-1}, t_{k-1})} I_{k-1}(z_1, \dots, z_{i_{k-2}}, x_{k-1}) & \text{for } 2 \leq k \leq m+1, \\ \Delta x \times \sum_{i_{k-1-m}=0}^{M-1} \frac{\hat{p}(x_k, t_k; y_k, t_k | x_{k-1-m_2}, t_{k-1-m_2}; x_{k-1-m_1}, t_{k-1-m_1}; x_{k-1}, t_{k-1}; y_{k-1}, t_{k-1})}{\hat{p}(y_k, t_k | y_0, t_0; \dots; y_{k-1}, t_{k-1})} I_{k-1}(z_{i_{k-m}}, \dots, z_{i_{k-1}}, x_k) & \text{for } k \geq m+2. \end{cases} \quad (2.14)$$

In the above equations, we have included the arguments of I_k where necessary to avoid confusion.

Combining (2.8), (2.14), (2.13) and (2.12), we get the proposed algorithm to compute the conditional expectation of $X(T)$ given the observation history $Y(s)$, $0 \leq s \leq T$. Algorithm summarizes the complete procedure to approximate the conditional expectation of the state $X(t)$ given observation history $Y(s) = y(s)$, $0 \leq s \leq t$, for $t \in (0, T]$

Algorithm: For the system (2.1), approximate the conditional expectation of the state $X(t)$ given observation history $Y(s) = y(s)$, $0 \leq s \leq t$, by computing that of $\hat{X}(t_k)$ given observation history $Y(t_i) = y_i$, $i = 1 \dots, k$ for $k = 1, 2, \dots, N$. Here, $0 = t_0 < t_1 < \dots < t_N = T$ with $t_i = i\Delta t$ and $\Delta t = \frac{T}{N}$.

Input: $f, g, h, r, T, D, \Delta t, \Delta x$, the initial conditions of $\hat{X}(t_k)$ and $\hat{Y}(t_k)$, and $\{y_1, y_2, \dots, y_N\}$. Here, f, g, h, r are the functions in (2.1), T is the time, D is the space domain, Δt is the time step and Δx is the space step. The sequence $\{y_1, y_2, \dots, y_N\}$ represents the observations collected at the sampling times.

Output: $E(\hat{X}(t_k)|y_0, t_0; \dots; y_k, t_k)$ for $k = 1, 2, \dots, N$

```

1 Set  $N = \frac{T}{\Delta t}$  and  $M = \frac{2D}{\Delta x}$ 
2 for  $k = 0$  to  $N$  do
3   | set  $t_k = k\Delta t$ 
4 for  $k = 0$  to  $M - 1$  do
5   | set  $z_k = -D + k\Delta x$ 
6 for  $k = 0$  to  $N$  do
7   | compute  $I_k$  by using (2.8), (2.14)
8   | compute  $\hat{p}(x_k, t_k|y_0, t_0; \dots; y_k, t_k)$  from  $I_k$  using (2.13)
9   | get  $E(\hat{X}(t_k)|y_0, t_0; \dots; y_k, t_k)$  from  $\hat{p}(x_k, t_k|y_0, t_0; \dots; y_k, t_k)$  using (2.12)
10 Output  $E(\hat{X}(t_k)|y_0, t_0; \dots; y_k, t_k)$  for  $k = 1, 2, \dots, N$ .
```

3 Examples

In this section, two numerical examples are presented to illustrate the proposed algorithm and verify its performance.

Example 1. Consider the following system,

$$\begin{cases} dX(t) = -0.1 \sin(X(t - \tau)) dt + dB(t), \\ X(s) = 0 \quad \text{for } s \in [-\tau, 0], \\ dY(t) = 100(X(t) + \sin(X(t))) + dV(t), \\ Y(0) = 0, \end{cases} \quad (3.1)$$

where $B(t)$ and $V(t)$ are standard Brownian motions independent of each other.

Take $\tau = \frac{1}{20}$, total time $T = 4$, time step $\Delta t = \frac{1}{80}$, space domain $[-2, 2]$, and space step $\Delta x = \frac{1}{8}$. By using the proposed algorithm, we get the state estimate, as shown in Figure 1. It can be seen from the figure that the estimated path of the state tracks the true path very well.

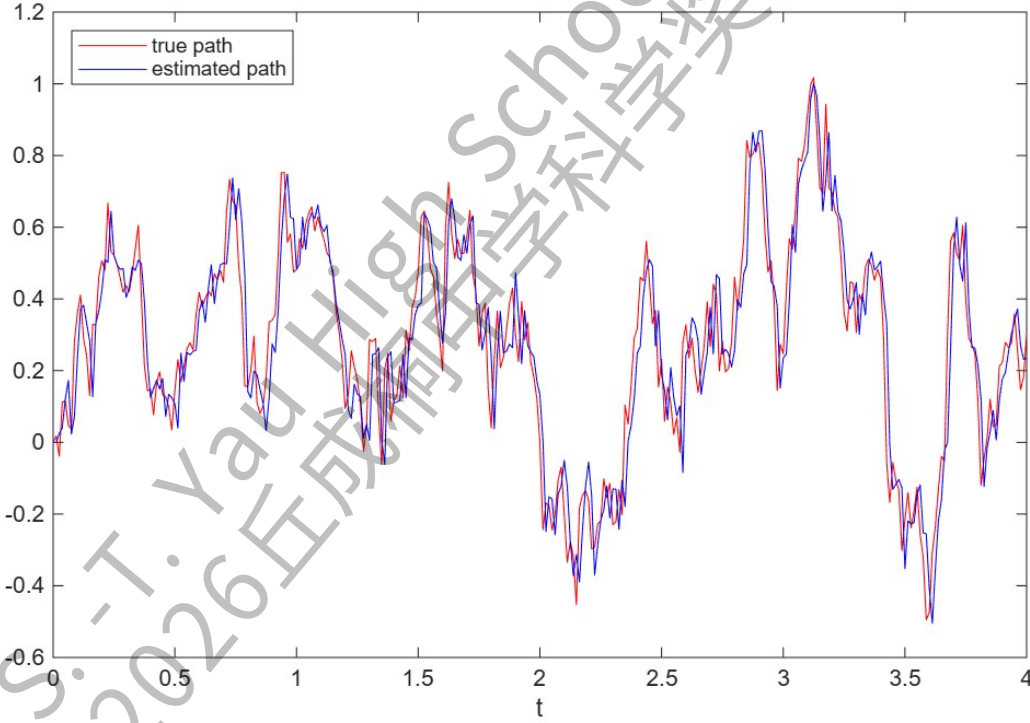


Figure 1: True and estimated paths for $X(t)$ in (3.1)

Example 2. Consider the following two dimensional stochastic system:

$$\left\{ \begin{array}{l} \begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} 8X_2(t) - 32(X_1(t - \tau))^3 \\ 5X_1(t) - 25(X_2(t - \tau))^3 \end{bmatrix} dt + \begin{bmatrix} dB_1(t) \\ dB_2(t) \end{bmatrix}, \\ X_1(s) = X_2(s) = 0 \quad \text{for } s \in [-\tau, 0], \\ \begin{bmatrix} dY_1(t) \\ dY_2(t) \end{bmatrix} = 100 \begin{bmatrix} X_1(t) + \sin(X_1(t)) \\ X_2(t) + \sin(X_2(t)) \end{bmatrix} dt + \begin{bmatrix} dV_1(t) \\ dV_2(t) \end{bmatrix}, \\ Y_1(0) = Y_2(0) = 0. \end{array} \right. \quad (3.2)$$

The proposed algorithm is applied to compute the conditional expectation of the state given the observation history. The result is shown in Figure 2. In the computation, we take $\tau = \frac{1}{100}$, total time $T = 1$, time step $\Delta t = \frac{1}{100}$, space domain $[-\frac{3}{2}, \frac{3}{2}]$, and space step $\Delta = \frac{1}{10}$. It can be seen from the figure that the estimated path of the state tracks the true path closely.

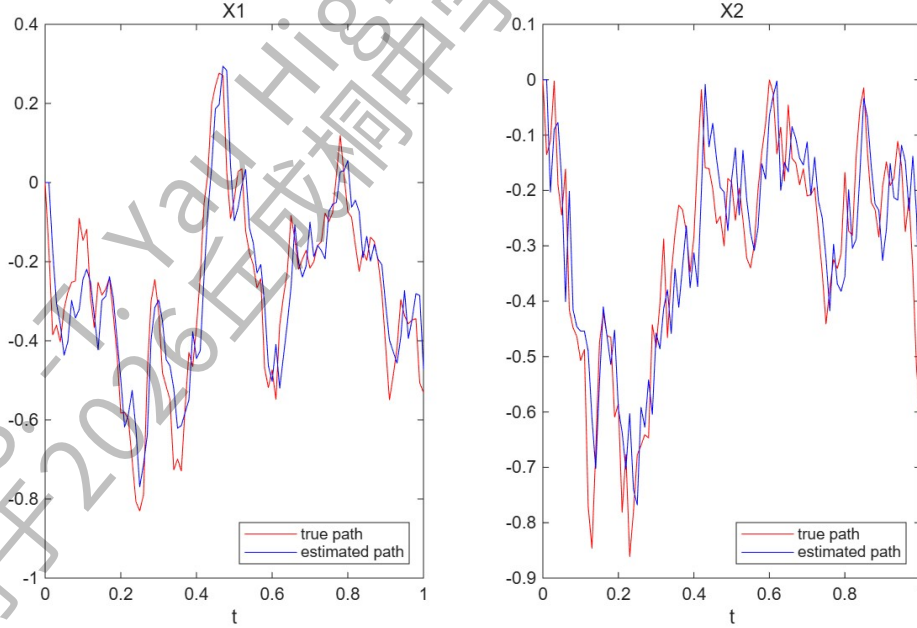


Figure 2: True and estimated paths for $X_1(t)$ (left) and $X_2(t)$ (right) in (3.2)

4 Concluding remarks

Existing nonlinear filtering algorithms rely heavily on solving the DMZ equation. The challenge of nonlinear filtering for stochastic dynamical systems with memory lies in the fact that no governing equation analogous to the DMZ equation exists to describe the conditional density of the state given the observation history.

This paper establishes a new nonlinear filtering algorithm for stochastic dynamical systems with memory modeled by (1.1). The algorithm is expressed by (2.8), (2.14), (2.13), (2.12), and is summarized at the end of Section 2. Although the algorithm is derived for systems with discrete time delays only, it is readily applicable to more general models with continuous time delays, such as stochastic functional differential equations.

References

- [1] X. Mao, *Stochastic Differential Equations and Applications*. Horwood series in mathematics & applications, Elsevier Science, 2008.
- [2] J. Xiong, *An Introduction to Stochastic Filtering Theory*. Oxford Graduate Texts in Mathematics, OUP Oxford, 2008.
- [3] S. S.-T. Yau, X. Chen, X. Jiao, J. Kang, S. Zeju, and Y. Tao, *Principles of Nonlinear Filtering Theory*. Springer Cham, 2024.
- [4] Z. Sun, J. Kang, X. Chen, and S.-T. Yau, “A tutorial on yau-yau nonlinear filtering algorithm: from model-driven to data-driven,” *Communications in Information and Systems*, vol. 26, pp. 119–149, 01 2026.
- [5] A. Ohashi, F. Russo, and E. Shamarova, “Smoothness of densities for path-dependent sdes under hörmander’s condition,” *Journal of Functional Analysis*, vol. 281, no. 11, p. 109225, 2021.
- [6] R. Chhaibi and I. Ekren, “A Hörmander condition for delayed stochastic differential equations,” *Annales Henri Lebesgue*, vol. 3, pp. 1023–1048, 2020.
- [7] E. Buckwar, R. Kuske, S.-E. A. Mohammed, and T. Shardlow, “Weak convergence of the euler scheme for stochastic differential delay equations,” *Lms Journal of Computation and Mathematics*, vol. 11, pp. 60–99, 2008.
- [8] C. T. H. Baker and E. Buckwar, “Numerical analysis of explicit one-step methods for stochastic delay differential equations,” *LMS Journal of Computation and Mathematics*, vol. 3, p. 315–335, 2000.
- [9] X. Mao, “Numerical solutions of stochastic functional differential equations,” *LMS Journal of Computation and Mathematics*, vol. 6, pp. 141–161, 2003.

Acknowledgments

I wish to express my sincere gratitude to Professor Xianming Liu of Huazhong University of Science and Technology, who served as my advisor on this project voluntarily and without compensation. Professor Liu suggested the research topic, recommended textbooks and other reading materials, and offered valuable advice and suggestions. I would also like to thank my high school for supporting my pursuits and my family for their encouragement. I am also grateful to the many researchers whose prior work laid the foundation for this project.

I first became acquainted with Professor Xianming Liu when I audited his course "Introduction to Probability and Statistics" at Huazhong University of Science and Technology in 2024. At the end of the course, I inquired whether he would be willing to mentor me on a research project. Knowing that I already had an understanding of calculus, linear algebra, and numerical analysis, Professor Liu suggested several topics for investigation, including one on nonlinear filtering for stochastic dynamical systems with memory. After some reading, I found that although stochastic dynamical systems with memory are more appropriate models in many applications, most existing nonlinear filtering algorithms are designed for systems without memory, and no nonlinear filtering algorithm currently exists for these types of models in general. I chose to pursue the nonlinear filtering topic because not only was it mathematically challenging but it also held potential for practical applications. This was precisely the kind of research I enjoy most.

Since I was unfamiliar with a lot of the advanced topics, Professor Xianming Liu recommended textbooks on stochastic differential equations, such as *An Introduction to Stochastic Differential Equations* by Lawrence C. Evans, along with reading materials on topics such as Kalman filtering, the Fokker-Planck equation, and the DMZ equation. The main challenge I faced was trying to compute the optimal state estimate in a new way that would also work for stochastic systems with memory. Trying to compute the conditional density directly without using a governing equation, I first used the Euler-Maruyama scheme to discretize the system. However, to do that, I needed to identify a structure from which a computable algorithm could be derived. After many

days without progress, I finally found a clever approach to derive a recursive structure from the expression for the conditional probability density. That was the key to the research project.

After the overall structure of the algorithm was confirmed and finalized, Professor Liu suggested several examples on which to test it. I did all the programming needed for the numerical procedure and collected the results. With Professor Liu's advice, a selection of examples was chosen for the paper.

At this point, all the material required for writing the paper had been prepared, so I started working on the first draft. Writing the first draft was a challenge, as I was not yet experienced in academic writing. The hardest part was learning to explain my ideas clearly and accessibly so that readers unfamiliar with the topic could understand them. During the editing stage, Professor Liu patiently proofread my drafts countless times, providing invaluable feedback and suggestions.