Received: 17 May 2021 | Revised: 4 July 2022 | Accepted: 1 October 2022

DOI: 10.1002/asjc.2970

RESEARCH ARTICLE

WILEY

Nonlinear Bayesian filtering via holonomic gradient method with quasi moment generating function

Tomoyuki Iori[1] | **Toshiyuki Ohtsuka**[2]

[1]Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

[2]Graduate School of Informatics, Kyoto University, Kyoto, Japan

Correspondence
Tomoyuki Iori, Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan.
Email: t-iori@ist.osaka-u.ac.jp

Funding information
JSPS KAKENHI, Grant/Award Number: JP15H02257, JP18J22093 and JP21K21285

Summary

A symbolic-numeric method is proposed for addressing the Bayesian filtering problems of a class of discrete-time nonlinear stochastic systems. We first approximate the posterior probability density function to be Gaussian. The update law of the mean and variance is formulated as the evaluation of several integrals depending on certain parameters. Unlike existing methods, such as the extended Kalman filter (EKF), unscented Kalman filter (UKF), and particle filter (PF), this formulation considers the nonlinearity of system dynamics exactly. To evaluate the integrals efficiently, we introduce an integral transform motivated by the moment generating function (MGF), which we call a quasi MGF. Furthermore, the quasi MGF is compatible with the Fourier transform of differential operators. We utilize this compatibility to decrease the number of computations of Gröbner bases in the noncommutative rings of differential operators, which reduces the offline computational time. A numerical example is presented to show the efficiency of the proposed method compared to that of other existing methods such as the EKF, UKF, and PF.

KEYWORDS

Bayesian filtering, symbolic-numeric method, theory of D-modules

1 | INTRODUCTION

Although the state of a dynamical system is important for monitoring and controlling using feedback, it is rarely observed to be the output of the system. The optimal filtering theory was developed to compute optimal estimates based on the history of observed outputs. The Kalman filter (KF) [1, 2] provides an explicit form of the optimal filter under the criterion of the minimum mean square error for linear systems with Gaussian noise. Following the success of the KF, the problem settings in optimal filtering have become diverse to an extent that they include other optimality criteria, nonlinear systems, and non-Gaussian noise [3–8].

For cases with non-Gaussian noise, Duong et al. [6] introduced an additive Laplace-distributed noise to model the impulsive behavior observed in sonar, radar, and air turbulent noise[9, 10]. Using an integration formula for Laplace distributions, they analytically derived the recursive propagation and update laws of the probability density function (PDF) of the state conditionally on the output history. The analytical form of the conditional PDF can then be used to compute its mean and variance or perform the maximum a posteriori estimation. Yin et al. [7] proposed a minimum entropy filter based on a radial basis function neural network (RBFNN) for nonlinear systems with non-Gaussian noise. This filter is based on

the data-driven minimum entropy filter, which minimizes the Rényi entropy of the output as a measurement of the randomness of the estimation. The RBFNN was introduced to model the dynamics of the entropy while reducing the computational cost, so that the filter can be implemented in real time.

Meanwhile, for cases of nonlinear systems, Chen and Hu [5] proposed a new Kalman-like filter that does not require any Gaussian assumptions but only the noise bounds. An upper bound of the estimation error is formulated as a function of the filter gain based on the noise bounds and Taylor series approximation of a nonlinear system. The optimal filter gain is obtained by minimizing the upper bound, and several numerical examples show that the Kalman-like filter with the optimal gain outperforms the other nonlinear KFs under the assumption of unknown noise variances. Li et al. [8] proposed a filtering algorithm that exploits the manifold structure underlying nonlinear system dynamics. For the spatial pose estimation via dual quaternions, they modeled the distributions on the manifold of unit dual quaternions by particles. The proposal distribution was then updated particle-wise with an unscented KF-like algorithm that utilizes the notion of locally augmented tangent space.

In Bayesian filtering, the nonlinearity of the system dynamics and the non-Gaussian noise can be considered in principle. However, from the implementation perspective, it is extremely difficult to consider them because of the following two main issues: Representing non-Gaussian distribution owing to the nonlinearity and non-Gaussian noise, and computing the integrations accompanied by the marginalization of the PDFs. One of the most popular approaches to deal with these issues is the nonlinear extension of the KF, such as the extended KF (EKF) [11], unscented KF (UKF) [12], cubature KF (CKF) [13], and Gauss-Hermite KF (GHKF) [14]. In these methods, the posterior PDF of the state of a dynamical system is assumed or approximated as a Gaussian PDF, so that it can be characterized by a finite number of parameters, namely, the mean and variance. This approximation allows us to use the Gauss quadrature rules to perform the integrations rapidly. However, the Gauss quadrature rules are approximations of the integrals; thus, they cannot capture the nonlinearity of a system exactly. Another popular approach is the Monte-Carlo scheme, such as the particle filter (PF) [15–17]. This approach has superior capabilities for describing the non-Gaussian distributions and considering the nonlinearity. These advantages can be attained if the number of particles is large enough. However, the high computational cost due to the large number of particles is problematic when the sampling intervals are tightly constrained.

To extend the boundary of the trade-off between the reduction of computational cost and the exact consideration of nonlinearity, in our previous work [18], we used an offline computation before performing the state estimation. This makes the online computation more efficient. In our previous work [18], we introduced a symbolic-numeric method called the holonomic gradient method (HGM) [19] to exactly consider the nonlinearity when updating the mean and variance of a Gaussian PDF approximating the posterior PDF. The integrations that are accompanied by the marginalization are handled offline in terms of differential operators, and then, the one-step update process in Bayesian filtering is reduced to a finite number of initial value problems (IVPs), which can be efficiently solved using numerical integration methods such as the Runge–Kutta method. Moreover, an integral transform was introduced in the preliminary result [20] to completely utilize the results of the HGM and reduce the number of IVPs that are solved online. In this paper, we refer to the integral transform as the *quasi moment generating function (MGF)* because its definition is similar to that of the MGF.

Integral transforms are useful tools for computing PDFs and have been used for filtering. Idan and Speyer [21] used an integral transform called the characteristic function in the state estimation for linear scalar systems with Cauchy noise. By utilizing the special structure of the characteristic function, they derived a state estimator by recursively propagating in time the characteristic function of the unnormalized conditional PDF of the state. From the propagated characteristic function, the posterior mean and variance can be derived explicitly. This characteristic-function approach has been extended to the case of multivariate linear systems with Cauchy noise [22]. Zhou et al. [23] used an integral transform called the cumulant generating function for the state estimation of Markov jump linear systems. By considering the higher-order moments of the state PDF, a discrete Markov jump linear system is transformed into a deterministic system, to which the Kalman filter can be applied.

In this study, we further explore the quasi MGF introduced in the preliminary work [20]. The quasi MGF is defined as the integral of the posterior PDF multiplied by a kernel. In the previous method [18], this integration is performed offline in terms of differential operators and includes the computations of several Gröbner bases. They require a high computational cost that is unacceptable even though they are performed offline. To avoid the computation of Gröbner bases, we introduce the *Fourier transform of differential operators* [24]. The definition of the quasi MGF is compatible with the Fourier transform of differential operators; that is, if a PDF is annihilated by a differential operator, its inverse Fourier transform annihilates the quasi MGF of the PDF. By making use

of this special compatibility, several computations of Gröbner bases can be replaced with simple substitutions of equations. As a result, the offline computational cost can be drastically reduced compared to that of our previous method [18].

The rest of this paper is organized as follows. First, Section 2 formulates the problems considered in this study and defines several integrals that have to be computed in the one-step estimation process. Next, Section 3 introduces a symbolic-numeric method called the HGM, which is used in the proposed method to evaluate the integrals efficiently. In Section 4, we introduce the notion of the quasi MGF to compute the mean and variance of the posterior PDF efficiently. The contents in this section are the results obtained in the preliminary work [20] and are presented for the completeness of this paper. In Section 5, we introduce the compatibility between the quasi MGF and the Fourier transform of differential operators. This compatibility is used to reduce the offline computational cost or specifically the number of computations of Gröbner bases in the rings of differential operators. Section 6 demonstrates the efficiency of the proposed method both in offline and online computations, and Section 7 concludes this paper. In Appendix A, the notion of holonomic functions is introduced, which is the basis of all the arguments in this paper. Finally, for the completeness of this paper, Appendix B briefly introduces the multiplication and integration in terms of differential operators, which are used in the proposed method.

Notations

For the field of real numbers $\mathbf{R}$ and a vector of indeterminates $X = [X_1 \ \dots \ X_n]^\top$, $\mathbf{R}[X]$ and $\mathbf{R}(X)$ denote the ring of polynomials and the field of rational functions in the components of X over $\mathbf{R}$, respectively. $\partial_X := [\partial_{X_1} \ \dots \ \partial_{X_n}]^\top$ denotes a vector of differential operators, where $\partial_{X_i} = \partial/\partial X_i$. We abbreviate ∂_{X_i} by ∂_i if X is clearly specified in accordance with the context. For a multi-index vector $d = [d_1 \ \dots \ d_n]^\top \in \mathbf{Z}^n_{\geq 0}$, X^d denotes a monomial $X_1^{d_1} X_2^{d_2} \ \dots \ X_n^{d_n}$, that is, the product of n monomials $X_i^{d_i} (i = 1, \dots, n)$. Similarly, ∂^d denotes a differential operator $\partial_1^{d_1} \partial_2^{d_2} \ \dots \ \partial_n^{d_n}$, that is, the composition of n differential operators $\partial_i^{d_i} (i = 1, \dots, n)$. The symbols $\mathcal{D}_n := \mathbf{R}[X]\langle\partial\rangle$ and $\mathcal{R}_n := \mathbf{R}(X)\langle\partial\rangle$ denote the non-commutative rings of the differential operators with coefficients in the polynomials and rational functions, respectively. The non-commutative ring $\mathcal{D}_n$ is also called the *n-dimensional Weyl algebra*. The subscript n of $\mathcal{D}_n$ or $\mathcal{R}_n$ is omitted if it is clear from the context. We denote the action of an element $l \in \mathcal{D}_n (\text{or} \, \mathcal{R}_n)$ on a sufficiently smooth function $\alpha(X) = \alpha(X_1, \dots, X_n)$ by $l \bullet \alpha(X)$; for instance, $\partial_i \bullet \alpha(X) = \partial\alpha/\partial X_i(X)$. The left ideal generated by a finite set of differential operators $\{l_1, \dots, l_s\} \subset \mathcal{D}_n (\text{or} \, \mathcal{R}_n)$ is defined as a set of differential operators $\langle l_1, \dots, l_n\rangle := \{a_1 \cdot l_1 + \dots + a_s \cdot l_s | a_1, \dots, a_s \in \mathcal{D}_n (\text{or} \in \mathcal{R}_n, \text{respectively})\}$. We omit the adjective "left" and simply call them ideals because all the ideals in this study are left ideals. For an ideal $I = \langle l_1, \dots, l_s\rangle$, the set $\{l_1, \dots l_s\}$ is called a basis of I. We say that a differential operator $l \in \mathcal{D} (\text{or} \mathcal{R})$ *annihilates* a function α if $l \bullet \alpha = 0$. Similarly, we say that an ideal $I \subset \mathcal{D} (\text{or} \, \mathcal{R})$ *annihilates* α if $l \bullet \alpha = 0$ for all $l \in I$. The set of all positive definite $n \times n$ matrices is denoted by $\text{PD}(n)$. The half-vectorization $\text{vech}(A)$ of a matrix $A = \{a_{ij}\} \in \text{PD}(n)$ is defined as an $n(n+1)/2$-dimensional vector $[a_{11} \ \dots \ a_{1n} a_{21} \ \dots \ a_{2n} \ \dots \ a_{nn}]^\top$. A smooth function $\alpha(X)$ is said to be *rapidly decreasing* if $\lim_{\|X\| \to \infty} X^a \partial^b \alpha(X) \to 0$ for all $a, b \in \mathbf{Z}^n_{\geq 0}$. For a Gaussian random vector X with mean μ and variance Σ, its PDF is denoted by $\mathcal{N}(X | \mu, \Sigma)$.

2 | PROBLEM SETTING

In this study, we consider the Bayesian filtering problems of discrete-time nonlinear systems with stochastic noise, where the posterior PDF of the state is recursively updated using the observed output. The update law of the posterior PDF is described as follows [25, 26]:

$$p(x_k \mid y_{[0:k]}) = \frac{p(y_k, x_k \mid y_{[0:k-1]})}{p(y_k \mid y_{[0:k-1]})}, \tag{1}$$

where $x_k \in \mathbf{R}^n$ and $y_k \in \mathbf{R}^r$ denote the state and output at time step k and $p(y_k, x_k | y_{[0:k-1]})$ is the joint PDF defined by

$$p(y_k, x_k \mid y_{[0:k-1]}) := p(y_k \mid x_k) \int_{\mathbf{R}^n} p(x_k \mid x_{k-1}) p(x_{k-1} \mid y_{[0:k-1]}) dx_{k-1},$$

and $p(y_k \mid y_{[0:k-1]})$ is the output PDF defined by

$$p(y_k \mid y_{[0:k-1]}) := \int_{\mathbf{R}^n} p(y_k, x_k \mid y_{[0:k-1]}) dx_k.$$

For simplicity of notation, we focus on the one-step update of the posterior PDF and omit the past outputs $y_{[0:k-1]}$ and the subscript k. The appearance of (1) can then be simplified to

$$p(x \mid y) = \frac{p(y,x)}{p(y)} = \frac{p(y \mid x) \int_{\mathbf{R}^n} p(x \mid x^-) p(x^-) dx^-}{\int_{\mathbf{R}^n} p(y \mid x) \left[\int_{\mathbf{R}^n} p(x \mid x^-) p(x^-) dx^-\right] dx}, \tag{2}$$

where x and y denote the current state and output, respectively, and x^- denotes the previous state.

We assume that the system dynamics and observation process are described by the following state and output equations:

$$x = f(x^-, u) + w, \tag{3}$$

$$y = h(x) + v, \tag{4}$$

where $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n$ and $h : \mathbf{R}^n \to \mathbf{R}^r$ are given nonlinear functions, $u \in \mathbf{R}^m$ is a given input, and w and v denote the system and observation noises, respectively. They are assumed to be independent and identically distributed with the PDFs $p_w(w)$ and $p_v(v)$, respectively. To express the conditional PDFs $p(x|x^-)$ and $p(y|x)$ in (2) in terms of f, h, p_w, and p_v, we use the rule of transformation of PDFs that is summarized as the following lemma[11].

Lemma 1. *Let X and Y be random n-dimensional vectors with the PDFs $p_X(X)$ and $p_Y(Y)$, respectively. Suppose there exists a relation $Y = g(X)$, its inverse g^{-1} exists, and both g and g^{-1} are continuously differentiable. Then,*

$$p_X(X) = p_Y(g(X)) \left| \det \left(\frac{\partial g}{\partial X} \right) \right|,$$

where $|\det(\partial g/\partial X)| > 0$ is the absolute value of the Jacobian determinant.

By using the above lemma, the conditional PDFs $p(x|x^-)$ and $p(y|x)$ in (2) can be written as follows:

$$\begin{aligned} p(x \mid x^-, u) &= p_w(x - f(x^-, u)) \left| \det \left(\frac{\partial w}{\partial x} \right) \right| \\ &= p_w(x - f(x^-, u)), \end{aligned} \tag{5}$$

$$\begin{aligned} p(y \mid x) &= p_v(y - h(x)) \left| \det \left(\frac{\partial v}{\partial y} \right) \right| \\ &= p_v(y - h(x)), \end{aligned} \tag{6}$$

where we use the facts that both derivatives $\partial w/\partial x$ and $\partial v/\partial y$ are identity matrices under the relations (3) and (4). Moreover, the former PDF is conditional not only on x^- but also on u because of (3). By substituting (5) and (6) into (2), the update law can be rewritten as

$$p(x \mid y, u) = \frac{p_{\text{joint}}(y, x \mid u)}{\int_{\mathbf{R}^n} p_{\text{joint}}(y, x \mid u) dx}, \tag{7}$$

where $p_{\text{joint}}(y, x \mid u)$ is defined using f, h, p_w, and p_v as

$$p_{\text{joint}}(y, x \mid u) := p_v(y - h(x)) \int_{\mathbf{R}^n} p_w(x - f(x^-, u)) p(x^-) dx^-. \tag{8}$$

The update law (7) with (8) can be regarded as a functional that maps the previous posterior PDF $p(x^-)$ to the current $p(x|y, u)$, depending on the parameters y and u. In general, the evaluation of this functional requires high computational cost and is unacceptable for filtering. To overcome this issue, the posterior PDFs are usually approximated by the PDFs that can be characterized by a finite number of parameters. For example, in the nonlinear KFs such as the EKF and UKF, the posterior PDFs are approximated as Gaussian PDFs [13, 27]. In this paper, we also approximate the posterior PDFs $p(x^-)$ and $p(x|y, u)$ as Gaussian, but the nonlinearity of the system, namely, f and h, are considered exactly in contrast to the nonlinear KFs.

First, suppose that a Gaussian approximation of $p(x^-)$, denoted by $\mathcal{N}(x^-|\mu^-, \Sigma^-)$, is provided by specifying the values of its mean μ^- and variance Σ^-. By replacing $p(x^-)$ in (8) with $\mathcal{N}(x^-|\mu^-, \Sigma^-)$, we can approximate $p_{\text{joint}}(y, x|u)$ as

$$\begin{aligned} p_{\text{joint}}(y, x \mid u) \approx \tilde{p}_{\text{joint}}(y, x \mid u, \mu^-, \Sigma^-) &:= p_v(y - h(x)) \\ &\int_{\mathbf{R}^n} p_w(x - f(x^-, u)) \mathcal{N}(x^- \mid \mu^-, \Sigma^-) dx^-, \end{aligned} \tag{9}$$

and the current posterior PDF can also be approximated by substituting (9) into (7), that is,

$$p(x \mid y, u) \approx \tilde{p}(x \mid y, u, \mu^-, \Sigma^-) := \frac{\tilde{p}_{\text{joint}}(y, x \mid u, \mu^-, \Sigma^-)}{\int_{\mathbf{R}^n} \tilde{p}_{\text{joint}}(y, x \mid u, \mu^-, \Sigma^-) dx}.$$

Note that in the nonlinear cases, the approximation $\tilde{p}_{\text{joint}}$ is *not* Gaussian because (5) and (6) are non-Gaussian, and the same is the case with $\tilde{p}(x|y, u, \mu^-, \Sigma^-)$. Therefore, we define an approximation of the current posterior PDF to be a Gaussian PDF $\mathcal{N}(x|\mu, \Sigma)$ with the same mean μ and variance Σ as $\tilde{p}(x|y, u, \mu^-, \Sigma^-)$, that is,

$$\begin{aligned} \mu(y, u, \mu^-, \Sigma^-) &:= \int_{\mathbf{R}^n} x \cdot \tilde{p}(x \mid y, u, \mu^-, \Sigma^-) dx \\ &= \frac{\int_{\mathbf{R}^n} x \cdot \tilde{p}_{\text{joint}}(y, x \mid u, \mu^-, \Sigma^-) dx}{\int_{\mathbf{R}^n} \tilde{p}_{\text{joint}}(y, x \mid u, \mu^-, \Sigma^-) dx}, \\ \Sigma(y, u, \mu^-, \Sigma^-) &:= \int_{\mathbf{R}^n} xx^\top \cdot \tilde{p}(x \mid y, u, \mu^-, \Sigma^-) dx - \mu\mu^\top \\ &= \frac{\int_{\mathbf{R}^n} xx^\top \cdot \tilde{p}_{\text{joint}}(y, x \mid u, \mu^-, \Sigma^-) dx}{\int_{\mathbf{R}^n} \tilde{p}_{\text{joint}}(y, x \mid u, \mu^-, \Sigma^-) dx} - \mu\mu^\top. \end{aligned}$$

For simplicity of notation, we introduce an integral Φ that depends on the parameters y, u, μ^-, and Σ^- as follows.

$$\Phi[g(x)](y,u,\mu^-,\Sigma^-) := \int_{\mathbf{R}^n} g(x)\cdot \tilde{p}_{\text{joint}}(y,x \mid u,\mu^-,\Sigma^-)dx,$$

where $g(x)$ is a scalar-, vector-, and matrix-valued function of x. Using Φ, μ and Σ can be simply rewritten as

$$\begin{aligned}\mu(y,u,\mu^-,\Sigma^-) &= \frac{\Phi[x](y,u,\mu^-,\Sigma^-)}{\Phi[1](y,u,\mu^-,\Sigma^-)},\\ \Sigma(y,u,\mu^-,\Sigma^-) &= \frac{\Phi[xx^\top](y,u,\mu^-,\Sigma^-)}{\Phi[1](y,u,\mu^-,\Sigma^-)} - \mu\mu^T.\end{aligned} \tag{10}$$

Now, the one-step update of the mean and variance of the Gaussian approximation $\mathcal{N}(x|\mu,\Sigma)$ is reduced to the evaluation of the functions that are defined by (10).

Remark 1. Although the Gaussian approximation of the posterior PDF appears in the nonlinear KFs [13, 27], this assumption is no longer applicable when the noise has a heavy-tailed distribution, such as the Cauchy noise. However, we would like to mention that the proposed method can be applicable to a certain class of problems with such a heavy-tailed distribution if we approximate the posterior PDF by a heavy-tailed distribution. More specifically, to achieve this, we require the heavy-tailed distribution used for the approximation to be (i) holonomic, (ii) characterized by a finite number of parameters, (iii) and such that the parameters can be explicitly expressed as functions of the previous estimate, observed output, and known input. Under these conditions, the proposed method can be performed with a slight modification even with heavy-tailed noise.

To update the mean and variance using (10), the three integrals $\Phi[1]$, $\Phi[x]$, and $\Phi[xx^\top]$ must be evaluated. The nonlinearity of the system dynamics is exactly considered in these integrals because f and h of (3) and (4) are included in (9) without any approximations. The nonlinearity makes it extremely hard to express the integrals in terms of elementary functions. Therefore, we have to rely on the numerical computation in one way or another. However, the numerical evaluation of the integrals is too computationally demanding for application in state estimation problems with short sampling intervals. In our previous work [18], we efficiently evaluated the integrals using the HGM, which will be introduced in the next section.

Finally, the following technical assumptions are made. The notion of holonomic functions and their remarkable properties are summarized in Appendix A.

Assumption 1. The conditional PDFs (5) and (6) are holonomic functions.

Assumption 2. We assume that the nonlinear functions f and h and the PDFs p_w and p_v are such that a function defined by them:

$$\exp(\xi^\top x)\tilde{p}_{\text{joint}}(y,x|u,\mu^-,\Sigma^-)$$

is rapidly decreasing with respect to x for any $y \in \mathbf{R}^r$, $u \in \mathbf{R}^m$, $\mu^- \in \mathbf{R}^n$, $\Sigma^- \in \text{PD}(n)$, and $\xi \in \Xi$, where Ξ is a compact subset of $\mathbf{R}^n$ including the origin.

Remark 2. Assumption 2 is made to guarantee the existence of the quasi MGF, though it is not clear which kind of problem satisfies the assumption. A simple example satisfying the assumption is the linear case with Gaussian noise. In this case, the approximated PDF $\tilde{p}_{\text{joint}}$ is given as Gaussian, which is identical to the true PDF p_{joint}. It is easy to see that each derivative of the Gaussian PDF is proportional to $\exp(-\|x\|^2)$ up to a polynomial factor in the components of x. Even if multiplied by $\exp(\xi^\top x)$, the product is proportional to $\exp(\|x\| - \|x\|^2)$ up to the polynomial factor, which still converges to 0 as $\|x\| \to \infty$. As another case that satisfies the assumption, we can consider a problem where p_v is proportional to $\exp(-\|v\|)$ up to a polynomial factor in the components of v and $h(x)$ is a polynomial of total degree $d > 2$. In this case, the PDF $p(y|x) = p_v(y - h(x))$ in (9) is proportional to $\exp(-\|x\|^d)$ up to a polynomial factor in the components of x, which makes $\tilde{p}_{\text{joint}}$ rapidly decreasing.

3 | HOLONOMIC GRADIENT METHOD

The HGM is a symbolic-numeric method that can efficiently evaluate a holonomic function (Definition 3) by using the differential operators that annihilate the holonomic function. In particular, the HGM is suitable for a holonomic function such that i) its explicit expression is described as an integral of nonlinear functions that depends on several parameters (such as the integrals $\Phi[1]$, $\Phi[x]$, and $\Phi[xx^\top]$), and ii) its evaluation based on the direct computation of the explicit expression is time-consuming; hence, it is unavailable.

Consider the evaluation of the holonomic function $\alpha(X)$ whose explicit expression is given as an integral that depends on a parameter X. For a specific point $\hat{X}$ where we want to evaluate $\alpha(X)$, the HGM is performed in the following three steps: (i) compute the Pfaffian system (A3) from the explicit expression of $\alpha(X)$; (ii) compute the value of Q in (A3) at an initial point X_{init}, which is used as an initial vector of an IVP solved in the next step; and (iii) integrate

the Pfaffian system along an integration path from X_{init} to $\hat{X}$. The following briefly explains each step.

(1) Computation of the Pfaffian system
As mentioned in the last paragraph of Appendix A, the Pfaffian system for $\alpha(X)$, or more specifically, the coefficient matrices $A_{X_i}(i = 1, \dots, n)$ can be computed from a basis $\mathcal{B}$ of a zero-dimensional ideal (Definition 2) that annihilates $\alpha(X)$. If the prescribed expression of $\alpha(X)$ is complicated,, it is extremely difficult to find $\mathcal{B}$. The multiplication and integration in terms of differential operators (Appendix B) are the key tools for finding this basis; if the expression of $\alpha(X)$ is given as an integral of the product of relatively simple holonomic functions, we can compute $\mathcal{B}$ from the set of every basis of a zero-dimensional ideal that annihilates each holonomic function. This computation requires the notion of the *holonomic ideals* in $\mathcal{D}$, which are counterparts of zero-dimensional ideals in $\mathcal{R}$, and is introduced in Section 5.

(2) Computation of the initial vector
In step 1, we can compute a finite set of differential operators $\{\partial^{d_1}, \dots, \partial^{d_{q-1}}\}$ in Lemma 6 from $\mathcal{B}$. By letting them act on the prescribed expression of $\alpha(X)$, we obtain an explicit expression for a vector-valued function $Q(X) := [\alpha\, \partial^{d_1}\alpha \cdots \partial^{d_{q-1}}\alpha]^\top$. Note that if the prescribed expression of $\alpha(X)$ is an integral depending on the parameter X, the expression of every component of $Q(X)$ is also described as an integral depending on the parameter X, which implies that the evaluation of $Q(X)$ is also time-consuming. However, as long as it is performed offline, $Q(X)$ can be evaluated numerically at a fixed point X_{init} using the obtained explicit expression, if the computational time is sufficient.

(3) Integration of the Pfaffian system along an integration path
When the evaluation point $\hat{X}$ is specified, we can define an integration path $X : [0,1] \ni s \mapsto X(s) \in \mathbf{R}^n$ such that $X(0) = X_{\text{init}}$ and $X(1) = \hat{X}$. For example, the line segment $X(s) := s\hat{X} + (1-s)X_{\text{init}}$ is available. The vector $Q(\hat{X})$, whose first component is $\alpha(\hat{X})$, can then be computed by solving the following IVP:

$$\begin{aligned} \frac{dQ}{ds} &= \frac{\partial Q}{\partial X}\frac{dX}{ds} = \sum_{i=1}^{n} A_{X_i}(X(s))Q(X(s))\frac{dX_i}{ds}, \\ Q(X(0)) &= Q(X_{\text{init}}). \end{aligned} \tag{11}$$

This IVP can be solved using numerical integration methods such as the Runge–Kutta method (RK4) or the Adams–Bashforth–Moulton predictor-corrector method (ABM4).

Note that some of the denominators of the components of $A_{X_i}(i = 1, \dots, n)$ can be zero at a certain point $X(\tilde{s})(\tilde{s} \in [0,1])$. The numerical integration may fail if such a number $\tilde{s}$ exists. The zero set of the least common multiple of all the denominators in $A_{X_i}(i = 1, \dots, n)$ is called the *singular locus* of the Pfaffian system. The initial point X_{init} and the integration path should be chosen such that the integration path does not intersect with the singular locus; that is, $\tilde{s}$ does not exist.

Here, we note that the evaluation point $\hat{X}$ does not appear until step 3. Therefore, if the specific value of $\hat{X}$ is only available online, we can perform steps 1 and 2 offline without the value of $\hat{X}$. We can use this feature of the HGM in the application of the state estimation and reduce the online computational cost [18]. However, the computational cost of Gröbner bases required in step 1 is extremely high and is not acceptable even though it is performed offline.

4 | EVALUATION OF MEAN AND VARIANCE BY QUASI MGF

This section introduces the preliminary results[20] for the completeness of this paper. Hereafter, $z \in \mathbf{R}^N$ denotes a vector that consists of the independent components of $y \in \mathbf{R}^r$, $u \in \mathbf{R}^m$, $\mu^- \in \mathbf{R}^n$, and $\Sigma^- \in \mathrm{PD}(n)$, where $N = r + m + n + n(n+1)/2$.

In our previous work[18], the HGM was utilized to evaluate integrals $\Phi[1]$, $\Phi[x_i](i = 1, \dots, n)$, and $\Phi[x_i x_j](i, j \in \{1, \dots, n\})$ to compute the estimates (10). In the HGM, an IVP needs to be solved to evaluate a scalar-valued function, which implies that, for example, the number of IVPs that are solved online for evaluating the matrix $\Phi[xx^\top]$ increases in the order of n^2. This may be unacceptable when n increases. To address this issue, we introduce an integral transform of the approximation of the joint PDF $\tilde{p}_{\text{joint}}$, which we call the *quasi MGF* for $\tilde{p}_{\text{joint}}$. The definition of the quasi MGF is similar to that of the MGF; therefore, the integrals $\Phi[x_i]$ and $\Phi[x_i x_j]$ can be obtained as partial derivatives of the quasi MGF. By making use of the finiteness of partial derivatives (Lemma 6), these partial derivatives can be evaluated by solving an IVP just once, which reduces the online computational cost.

The quasi MGF $\mathcal{T}\left[\tilde{p}_{\text{joint}}\right]$ is defined as the following integral transform of $\tilde{p}_{\text{joint}}$:

$$\begin{aligned} \mathcal{T}\left[\tilde{p}_{\text{joint}}\right](\xi, z) &:= \Phi\left[\exp(\xi^\top x)\right](z) \\ &= \int_{\mathbf{R}^n} \exp(\xi^T x)\tilde{p}_{\text{joint}}(y, x|u, \mu^-, \Sigma^-)dx. \end{aligned} \tag{12}$$

Under Assumption 2, the integrand of (12) is bounded and smooth for all $\xi \in \Xi$. Although this definition is similar to that of the MGF, it is different because the integration is considered only over x, instead of y and x. The integrals that appear in (10) are then derived from the partial derivatives of the quasi MGF as follows.

$$\begin{aligned}
\mathcal{T}[\tilde{p}_{\text{joint}}](0,z) &= \Phi[1](z),\\
\partial_{\xi_i}\mathcal{T}[\tilde{p}_{\text{joint}}](0,z) &= \int_{\mathbf{R}^n} \left[\partial_{\xi_i} \exp(\xi^\top x)\tilde{p}_{\text{joint}}(y,x|u,\mu^-,\Sigma^-)\right]\Big|_{\xi=0} dx\\
&= \int_{\mathbf{R}^n} x_i \tilde{p}_{\text{joint}}(y,x|u,\mu^-,\Sigma^-)dx = \Phi[x_i](z),\\
\partial_{\xi_i}\partial_{\xi_j}\mathcal{T}[\tilde{p}_{\text{joint}}](0,z) &= \int_{\mathbf{R}^n} \left[\partial_{\xi_i}\partial_{\xi_j} \exp(\xi^\top x)\tilde{p}_{\text{joint}}(y,x|u,\mu^-,\Sigma^-)\right]\Big|_{\xi=0} dx\\
&= \int_{\mathbf{R}^n} x_i x_j \tilde{p}_{\text{joint}}(y,x|u,\mu^-,\Sigma^-)dx = \Phi[x_i x_j](z).
\end{aligned}$$

From the closure property (Lemma 5) and Assumption 1, $\tilde{p}_{\text{joint}}$ is a holonomic function. Therefore, by considering the closure property again, $\mathcal{T}[\tilde{p}_{\text{joint}}]$ is also a holonomic function because the kernel $\exp(\xi^\top x)$ is also holonomic. Lemma 6 guarantees the existence of a finite set $\{\partial^{d_1}, \dots, \partial^{d_{q-1}}\}$ such that the q-dimensional vector-valued function

$$Q(\xi,z) := \left[\mathcal{T}[\tilde{p}_{\text{joint}}] \quad \partial^{d_0}\mathcal{T}[\tilde{p}_{\text{joint}}] \quad \dots \quad \partial^{d_{q-1}}\mathcal{T}[\tilde{p}_{\text{joint}}]\right]^\top$$

satisfies the Pfaffian system (A3). Using the Pfaffian system, we can derive $\mathcal{T}[\tilde{p}_{\text{joint}}]$ and its derivatives as a linear combination of components of $Q(\xi,z)$ with coefficients in $\mathbf{R}(\xi,z)$.

First, $\mathcal{T}[\tilde{p}_{\text{joint}}](\xi,z)$ is the first component of $Q(\xi,z)$; thus, it can be written as

$$\mathcal{T}[\tilde{p}_{\text{joint}}](\xi,z) = C^{(0)}Q(\xi,z) \tag{13}$$

using a constant coefficient vector $C^{(0)} = [1\,0\,\dots\,0] \in \mathbf{R}^{1\times q}$. Next, for $i = 1, \dots, n$, the first derivative $\partial_{\xi_i}\mathcal{T}[\tilde{p}_{\text{joint}}](\xi,z)$ appears in the first component of $\partial_{\xi_i}Q$, namely, the left-hand side of (A3). Hence, the following equality holds.

$$\partial_{\xi_i}\mathcal{T}[\tilde{p}_{\text{joint}}](\xi,z) = C_i^{(1)}(\xi,z)Q(\xi,z)\ (i = 1, \dots, n), \tag{14}$$

where the coefficient vector $C_i^{(1)}(\xi,z) \in \mathbf{R}(\xi,z)^{1\times q}$ is the first row vector of the rational matrix-valued function $A_{\xi_i}(\xi,z) \in \mathbf{R}(\xi,z)^{q\times q}$ in (A3). Finally, the second derivatives $\partial_{\xi_i}\partial_{\xi_j}\mathcal{T}[\tilde{p}_{\text{joint}}](\xi,z)(i,j = 1, \dots, n)$ are obtained by differentiating both sides of (A3), that is,

$$\begin{aligned}
\partial_{\xi_i}\bullet\partial_{\xi_j}Q &= \partial_{\xi_i}\bullet(A_{\xi_j}Q)\\
&= \partial_{\xi_i}A_{\xi_j}Q + A_{\xi_j}\partial_{\xi_i}Q\\
&= (\partial_{\xi_i}A_{\xi_j} + A_{\xi_j}A_{\xi_i})Q,
\end{aligned}$$

where $\partial_{\xi_i}A_{\xi_j} + A_{\xi_j}A_{\xi_i} \in \mathbf{R}(\xi,z)^{q\times q}$. Hence, the second derivatives can be obtained as follows:

$$\partial_{\xi_i}\partial_{\xi_j}\mathcal{T}[\tilde{p}_{\text{joint}}](\xi,z) = C_{ij}^{(2)}(\xi,z)Q(\xi,z), \tag{15}$$

where the coefficient vector $C_{ij}^{(2)}(\xi,z) \in \mathbf{R}(\xi,z)^{1\times q}$ is the first row vector of $\partial_{\xi_j}A_{\xi_i} + A_{\xi_i}A_{\xi_j}$.

The coefficient matrices $A_{\xi_i}(i = 1, \dots, n)$ and the coefficient vectors $C_i^{(1)}(\xi,z)(i = 1, \dots, n)$ and $C_{ij}^{(2)}(\xi,z)(i,j = 1, \dots, n)$ can be computed symbolically and offline. Hence, the values of the integrals $\Phi[1](z)$, $\Phi[x](z)$, and $\Phi[xx^\top](z)$ at $z = \hat{z}$ can be computed as follows:

$$\begin{aligned}
\Phi[1](\hat{z}) &= \mathcal{T}[\tilde{p}_{\text{joint}}](0,\hat{z}) = C^{(0)}Q(0,\hat{z}),\\
\Phi[x_i](\hat{z}) &= \partial_{\xi_i}\mathcal{T}[\tilde{p}_{\text{joint}}](0,\hat{z}) = C_i^{(1)}(0,\hat{z})Q(0,\hat{z}),\\
\Phi[x_i x_j](\hat{z}) &= \partial_{\xi_i}\partial_{\xi_j}\mathcal{T}[\tilde{p}_{\text{joint}}](0,\hat{z}) = C_{ij}^{(2)}(0,\hat{z})Q(0,\hat{z}).
\end{aligned}$$

Finally, the online computation of the proposed method can be summarized as Algorithm 1.

Algorithm 1 State estimation using HGM via Quasi MGF

Input: Coefficient vectors: $C^{(0)}$, $C_i^{(1)}(\xi,z)$ $(i = 1,\dots,n)$, and $C_{ij}^{(2)}(\xi,z)$ $(i,j \in \{1,\dots,n\})$, coefficient matrices A_λ $(\lambda \in \{\xi_1,\dots,\xi_n,z_1,\dots,z_N\})$ of Pfaffian system for $\mathcal{T}[\tilde{p}_{\text{joint}}]$, initial point z_{init} and corresponding initial vector $Q(0,z_{\text{init}})$, given data $\hat{z}$ consisting of given input u, observed output y, and previous estimates μ^- and Σ^-

Output: Failure or estimates of mean μ and variance Σ of current posterior PDF

1: Define $z(s)$ such that $z(0) = z_{\text{init}}$ and $z(1) = \hat{z}$
2: **if** $\exists \tilde{s} \in [0,1]$ such that $z(\tilde{s})$ is included in singular locus of Pfaffian system **then**
3: **return** algorithm has failed
4: **end if**
5: Integrate ODE (11) from $s = 0$ to 1 with the initial vector $Q(0,z_{\text{init}})$ by numerical integration and obtain $Q(0,\hat{z})$
6: Set $\Phi[1](\hat{z}) \leftarrow C^{(0)}Q(0,\hat{z})$, $\Phi[x_i](\hat{z}) \leftarrow C_i^{(1)}(0,\hat{z})Q(0,\hat{z})$, and $\Phi[x_i x_j](\hat{z}) \leftarrow C_{ij}^{(2)}(0,\hat{z})Q(0,\hat{z})$
7: **return** $\mu = \Phi[x](\hat{z})/\Phi[1](\hat{z})$ and $\Sigma = \Phi[xx^\top](\hat{z})/\Phi[1](\hat{z}) - \mu\mu^\top$

5 | COMPUTATION OF PFAFFIAN SYSTEM FOR QUASI MGF

In the preliminary work [20], only the online computational time, which directly affects the efficiency of estimation methods, is investigated and compared with that of the other existing methods. However, the offline computational cost is also important for the practical applications. For example, the tuning of the estimator may suffer from the huge offline computational time. In practical situations, the PDFs p_w and p_v in (9) are often unknown and need to be tuned. During this tuning, we have to construct many estimators with various candidate PDFs, which requires the offline part of the proposed method to be executed many times. If the offline part requires much computational time, this process cannot be acceptable even if it can be performed offline. In this section, we focus on the offline computational cost of the proposed method.

As a part of the input of Algorithm 1, the Pfaffian system for the quasi MGF $\mathcal{T}[\tilde{p}_{\text{joint}}]$ is required. It can be computed from a basis of the zero-dimensional ideal that annihilates $\mathcal{T}[\tilde{p}_{\text{joint}}]$. Furthermore, such a basis can be computed from the bases of holonomic ideals I^-, I^t, and I^o that annihilate $\mathcal{N}(x^-|\mu^-, \Sigma^-)$, $p(x|x^-, u^-)$, and $p(y|x)$, respectively (see previous studies [20, 29] for details). This computation is performed algorithmically using the multiplication and integration in terms of differential operators (Lemmas 7 and 8). The multiplication and integration in terms of differential operators require several computations of Gröbner bases of ideals in $\mathcal{D}$. This incurs a high computational cost, which may be unacceptable even when the computations are performed offline.

This section first introduces the *Fourier transform of differential operators* to partially avoid computing Gröbner bases. To this end, we need to focus on holonomic ideals in $\mathcal{D}$ rather than zero-dimensional ideals in $\mathcal{R}$. Although we skip the definition of holonomic ideals owing to space limitations, it is characterized by the dimension of ideals and modules (see previous works [28, 29] for details). Zero-dimensional ideals in $\mathcal{R}$ and holonomic ideals in $\mathcal{D}$ are related to each other by the following lemmas.

Lemma 2. [29] *For a zero-dimensional ideal $J \subset \mathcal{R}$, the intersection $J \cap \mathcal{D}$ is a holonomic ideal in $\mathcal{D}$.*

Lemma 3. [29] *Let I be a holonomic ideal in $\mathcal{D}$ and let $\mathcal{R}I$ represent the following set of differential operators.*

$$\mathcal{R}I := \{a_1 l_1 + \cdots + a_s l_s | a_1, \ldots, a_s \in \mathcal{R}; l_1, \ldots, l_s \in I; s \in \mathbf{Z}_{\geq 0}\}.$$

Then, $\mathcal{R}I$ is a zero-dimensional ideal in $\mathcal{R}$.

Lemma 3 implies that if a holonomic ideal $I \subset \mathcal{D}$ is generated by a finite number of differential operators $l_1, \ldots, l_s \in \mathcal{D}$, the zero-dimensional ideal $\mathcal{R}I$ is also generated by $l_1, \ldots, l_s$ as elements of $\mathcal{R}$. Hence, from the perspective of symbolic computation, computing a basis of I readily implies computing a basis of $\mathcal{R}I$.

The Fourier transform in terms of differential operators is defined as follows.

Definition 1. [29] For $X = [X_1 \ \ldots \ X_n]^\top$, $Y = [Y_1 \ \ldots \ Y_m]^\top$, and $Z = [Z_1 \ \ldots \ Z_n]^\top$, the *Fourier transform* of $\mathbf{R}[X, Y]\langle \partial_X, \partial_Y \rangle$ with respect to X is the ring isomorphism

$$\mathcal{F} : \mathbf{R}[X, Y]\langle \partial_X, \partial_Y \rangle \to \mathbf{R}[Z, Y]\langle \partial_Z, \partial_Y \rangle$$

defined by

$$\mathcal{F}(X_i) = -\partial_{Z_i}, \ \mathcal{F}(\partial_{X_i}) = Z_i \qquad (i = 1, \ldots, n). \tag{16}$$

The Fourier transform $\mathcal{F}$ is an automorphism of $\mathcal{D}_{n+m}$ with the inverse defined by

$$\mathcal{F}^{-1}(Z_i) = \partial_{X_i}, \ \mathcal{F}^{-1}(\partial_{Z_i}) = -X_i \qquad (i = 1, \ldots, n),$$

because both $\mathbf{R}[X, Y]\langle \partial_X, \partial_Y \rangle$ and $\mathbf{R}[Z, Y]\langle \partial_Z, \partial_Y \rangle$ are isomorphic to $\mathcal{D}_{n+m}$.

Using the Fourier transform, we can readily compute a holonomic ideal for $\mathcal{T}[\tilde{p}_{\text{joint}}]$ from that for $\tilde{p}_{\text{joint}}$, without computing any Gröbner bases. The following theorem describes the compatibility between the quasi MGF and the Fourier transform of differential operators, which is the key in the offline part of the proposed method.

Theorem 1. *Let $\mathcal{D}$ be $\mathbf{R}[x, z]\langle \partial_x, \partial_z \rangle$, $\mathcal{D}'$ be $\mathbf{R}[\xi, z]\langle \partial_\xi, \partial_z \rangle$, and $\mathcal{F}^{-1} : \mathcal{D} \to \mathcal{D}'$ be the inverse of the Fourier transform with respect to x. For a holonomic ideal $I = \langle l_1, \ldots, l_b \rangle \subset \mathcal{D}$ that annihilates $\tilde{p}_{joint}(x, z)$, its inverse Fourier transform is defined as*

$$\mathcal{F}^{-1}(I) := \left\langle \mathcal{F}^{-1}(l_1), \ldots, \mathcal{F}^{-1}(l_b) \right\rangle \subset \mathcal{D}'.$$

Then, $\mathcal{F}^{-1}(I)$ is a holonomic ideal that annihilates $\mathcal{T}[\tilde{p}_{joint}](\xi, z)$.

Proof. Whether an ideal $I \subset \mathcal{D}$ is holonomic is determined by the dimension of the left $\mathcal{D}$-module $\mathcal{D}/I$[29]. In addition, it is known that the dimension of a left $\mathcal{D}$-module is preserved under any automorphism of $\mathcal{D}$[24]. Hence, the dimension of the left $\mathcal{D}$-module $\mathcal{D}/\mathcal{F}^{-1}(I)$ is the same as $\mathcal{D}/I$, which implies $\mathcal{F}^{-1}(I)$ is holonomic if and only if I is holonomic.

It needs to be demonstrated that $\mathcal{F}^{-1}(I)$ annihilates $\mathcal{T}[\tilde{p}_{\text{joint}}]$. To this end, we show that

$$\mathcal{T}[l \bullet \tilde{p}_{\text{joint}}] = \mathcal{F}^{-1}(l) \bullet \mathcal{T}[\tilde{p}_{\text{joint}}]$$

for any element $l \in \mathcal{D}$. If this is true, we obtain

$$\mathcal{F}^{-1}(l)\bullet\mathcal{T}[\tilde{p}_{\text{joint}}] = \mathcal{T}[l\bullet\tilde{p}_{\text{joint}}] = \mathcal{T}[0] = 0\ (\forall l \in I),$$

which implies that $\mathcal{F}^{-1}(I)$ annihilates $\mathcal{T}[\tilde{p}_{\text{joint}}]$. Since $z_1, \dots, z_N$ and $\partial_{z_1}, \dots, \partial_{z_N}$ have no relevance with $\mathcal{F}$, it is sufficient to show the following equalities.

$$\mathcal{T}[x_i\bullet\tilde{p}_{\text{joint}}] = \mathcal{F}^{-1}(x_i)\bullet\mathcal{T}[\tilde{p}_{\text{joint}}], \tag{17}$$

$$\mathcal{T}[\partial_{x_i}\bullet\tilde{p}_{\text{joint}}] = \mathcal{F}^{-1}(\partial_{x_i})\bullet\mathcal{T}[\tilde{p}_{\text{joint}}]. \tag{18}$$

The former one (17) can be obtained from (12) and (16) as follows.

$$\begin{aligned}\mathcal{T}[x_i\bullet\tilde{p}_{\text{joint}}](\xi,z) &= \int_{\mathbf{R}^n} \exp(\xi^\top x)\left\{x_i\tilde{p}_{\text{joint}}(x,z)\right\}dx \\ &= \int_{\mathbf{R}^n} \left\{\partial_{\xi_i}\bullet\exp(\xi^\top x)\right\}\tilde{p}_{\text{joint}}(x,z)dx \\ &= \mathcal{F}^{-1}(x_i)\bullet\mathcal{T}[\tilde{p}_{\text{joint}}](\xi,z).\end{aligned}$$

Meanwhile, for (18), we first show the following equality that is obtained by integrating by parts:

$$\begin{aligned}&\int_{-\infty}^{\infty} \exp(\xi^\top x)\left\{\partial_{x_i}\bullet\tilde{p}_{\text{joint}}(x,z)\right\}dx_i \\ &= \left[\exp(\xi^\top x)\tilde{p}_{\text{joint}}(x,z)\right]_{x_i=-\infty}^{x_i=\infty} \\ &\quad - \int_{-\infty}^{\infty} \left\{\partial_{x_i}\bullet\exp(\xi^\top x)\right\}\tilde{p}_{\text{joint}}(x,z)dx_i \\ &= -\xi_i\int_{-\infty}^{\infty} \exp(\xi^\top x)\tilde{p}_{\text{joint}}(x,z)dx_i,\end{aligned}$$

where the second equality follows from Assumption 2. Using this equality, (18) is proven as follows:

$$\begin{aligned}\mathcal{T}[\partial_{x_i}\bullet\tilde{p}_{\text{joint}}](\xi,z) &= \int_{\mathbf{R}^{n-1}} \left\{\int_{-\infty}^{\infty} \exp(\xi^\top x)\left\{\partial_{x_i}\bullet\tilde{p}_{\text{joint}}(x,z)\right\}dx_i\right\}dx_{\hat{i}} \\ &= \int_{\mathbf{R}^{n-1}} \left\{-\xi_i\int_{-\infty}^{\infty} \exp(\xi^\top x)\tilde{p}_{\text{joint}}(x,z)dx_i\right\}dx_{\hat{i}} \\ &= \mathcal{F}^{-1}(\partial_{x_i})\bullet\mathcal{T}[\tilde{p}_{\text{joint}}](\xi,z),\end{aligned}$$

where $x_{\hat{i}} = [x_1\ \dots\ x_{i-1}\,x_{i+1}\ \dots\ x_n]^T$. This completes the proof. □

Theorem 1 shows that a basis of $\mathcal{F}^{-1}(I)$ can be derived from a basis of I by performing simple substitutions, $x_i \leftarrow \partial_{\xi_i}$ and $\partial_{x_i} \leftarrow -\xi_i$ for $i = 1, \dots, n$. As shown in the example illustrated in Section 6, this leads to a significant reduction in the offline computational cost. Finally, using Theorem 1, the Pfaffian system for the quasi MGF $\mathcal{T}[\tilde{p}_{\text{joint}}]$ is obtained from the holonomic ideals I^-, I^t, and I^o by Algorithm 2. In line 2 of Algorithm 2, the multiplication and integration in terms of differential operators are avoided by utilizing the compatibility of the quasi MGF and the Fourier transform as summarized in Theorem 1.

Algorithm 2 Offline computation of $C_i^{(1)}(\xi,z)$, $C_{ij}^{(2)}(\xi,z)$, and A_λ

Input: Bases of holonomic ideals $I^- \subset \mathbf{R}[x^-,\mu^-,\text{vech}(\Sigma^-)]\langle\partial_{x^-},\partial_{\mu^-},\partial_{\text{vech}(\Sigma^-)}\rangle$ that annihilates $\mathcal{N}(x^- \mid \mu^-,\Sigma^-)$, $I^t \subset \mathbf{R}[x,x^-,u^-]\langle\partial_x,\partial_{x^-},\partial_{u^-}\rangle$ that annihilates $p(x \mid x^-,u^-)$, and $I^o \subset \mathbf{R}[y,x]\langle\partial_y,\partial_x\rangle$ that annihilates $p(y \mid x)$

Output: Coefficient vectors: $C_i^{(1)}(\xi,z)$ $(i = 1,\dots,n)$, and $C_{ij}^{(2)}(\xi,z)$ $(i,j \in \{1,\dots,n\})$, coefficient matrices A_λ $(\lambda \in \{\xi_1,\dots,\xi_n,z_1,\dots,z_N\})$ of Pfaffian system for $\mathcal{T}[\tilde{p}_{\text{joint}}]$

1: Compute basis of holonomic ideal $I^{\text{joint}} \subset \mathbf{R}[x,z]\langle\partial_x,\partial_z\rangle$ that annihilates $\tilde{p}_{\text{joint}}(x,z)$ from bases of I^-, I^t, and I^o using Lemmas 7 and 8

2: Obtain basis of $\mathcal{F}^{-1}(I^{\text{joint}}) \subset \mathbf{R}[\xi,z]\langle\partial_\xi,\partial_z\rangle$ from basis of I^{joint} with substitutions $x_i \leftarrow \partial_{\xi_i}$ and $\partial_{x_i} \leftarrow -\xi_i$ for $i = 1,\dots,n$

3: Compute finite set $\{\partial^{d_1},\dots,\partial^{d_{q-1}}\} \subset \mathbf{R}[\xi,z]\langle\partial_\xi,\partial_z\rangle$ in Lemma 6 from basis of $\mathcal{F}^{-1}(I^{\text{joint}})$

4: Compute coefficient matrices $A_\lambda \in \mathbf{R}(\xi,z)^{q\times q}$ for $\lambda \in \{\xi_1,\dots,\xi_n,z_1,\dots,z_N\}$ from $\{\partial^{d_1},\dots,\partial^{d_{q-1}}\}$ and basis of $\mathcal{F}^{-1}(I^{\text{joint}})$

5: Let $C_i^{(1)} \in \mathbf{R}(\xi,z)^{1\times q}$ be first row vector of A_{ξ_i} for $i = 1,\dots,n$

6: Let $C_{ij}^{(2)} \in \mathbf{R}(\xi,z)^{1\times q}$ be first row vector of $\partial_{\xi_j}A_{\xi_i} + A_{\xi_i}A_{\xi_j}$ for $i,j \in \{1,\dots,n\}$

7: **return** $C_i^{(1)}$, $C_{ij}^{(2)}$, and A_λ

In the rest of this section, we discuss how to obtain the input of Algorithm 2, namely, the holonomic ideals I^-, I^t, and I^o, from the PDFs $\mathcal{N}(x^-|\mu^-,\Sigma^-)$, $p(x|x^-,u^-)$, and $p(y|x)$, respectively. If the PDFs are complex nonlinear functions, we can decompose the problem of finding a proper holonomic ideal to problems of finding more simple holonomic ideals. For example, by expressing a complex nonlinear function $\alpha(X)$ as the sum or product of several simple nonlinear functions $\alpha_1(X), \dots, \alpha_p(X)$, we can construct a holonomic ideal annihilating $\alpha(X)$ from a set of holonomic ideals annihilating $\alpha_1(X), \dots, \alpha_p(X)$ (for details, see Oaku et al. [30] and Lemma 7 in the appendix of this paper).

However, even for a simple nonlinear function $\alpha_i(X)$, it is sometimes difficult to find a holonomic ideal $I \subset \mathbf{R}[X]\langle\partial\rangle$ annihilating $\alpha_i(X)$. In such cases, instead of I, we can find a zero-dimensional ideal $J \subset \mathbf{R}(X)\langle\partial\rangle$ annihilating $\alpha_i(X)$, which is easier to compute. Once a basis of J is obtained, by clearing the denominators of the basis, we can often obtain a basis of I. If not, we can algorithmically compute a basis

of $I = J \cap \mathbf{R}[X]\langle\partial\rangle$ defined in Lemma 2 (see Tsai [31] for details of the algorithm).

There are many classes of simple nonlinear functions such that, for any function in the class, we can manually compute a zero-dimensional ideal annihilating it. Some of these classes are introduced in the following examples. Consequently, we can algorithmically compute a proper zero-dimensional ideal as well as holonomic ideal for any nonlinear function that can be expressed as the sum and product of the simple nonlinear functions.

Example 1. For any rational function $g \in \mathbf{R}(X)$ of $X = [X_1 X_2 \ \dots \ X_n]^\top$, we can easily compute a zero-dimensional ideal annihilating it. Indeed, as the derivatives $\partial_{X_i} g(X)$ are rational functions, $g(X)$ is annihilated by differential operators $\partial_{X_i} - (\partial_{X_i} g(X))/g(X) (i = 1, \dots, n)$, which constitute a basis of such a zero-dimensional ideal from Lemma 4.

Example 2. The second class is the exponential of rational functions $\exp(g(X))$. It is easy to verify that the differential operators $\partial_{X_i} - (\partial_{X_i} g(X)) (i = 1, \dots, n)$ annihilate $\exp(g(X))$ and again constitute a basis of the annihilating zero-dimensional ideal from Lemma 4.

Example 3. Finally, the composition of trigonometric functions and rational functions such as $\cos(g(X))$ also belongs to the class of simple nonlinear functions for which we can compute a proper holonomic ideal. The first and second derivatives of $\cos(g(X))$ with respect to X_i are easily computed as

$$\partial_i \bullet \cos(g(X)) = -\partial_i g(X) \sin(g(X))$$
$$\partial_i^2 \bullet \cos(g(X)) = -\partial_i^2 g(X) \sin(g(X)) - (\partial_i g(X))^2 \cos(g(X)).$$

By eliminating $\sin(g(X))$ from these equations, we obtain the differential operator $\partial_i^2 - \{\partial_i^2 g(X)/\partial_i g(X)\}\partial_i + (\partial_i g(X))^2$ for each $i = 1, \dots, n$. These n differential operators again constitute a basis of a zero-dimensional ideal annihilating $\cos(g(X))$.

6 | NUMERICAL EXAMPLE

This section provides a numerical example to demonstrate the efficiency of the proposed method. We use `Risa/Asir` [32] for the symbolic computation, `Maple` for the offline numerical computation, and `Python` for the online numerical computation on a PC (Intel(R) Core(TM) i7-1065G7 CPU @ 1.30 GHz; RAM: 16 GB).

Consider the following nonlinear stochastic system:

$$x = \frac{4x^-}{1 + (x^-)^2} + u + w, \tag{19}$$

$$y = x + v, \tag{20}$$

where the PDFs of w and v are given as $\mathcal{N}(w|0, 1)$ and $\mathcal{N}(v|0, 9)$, respectively, and u is given as a function $\cos(0.6k)$ of the time step k. To apply Algorithm 1 to this nonlinear stochastic system, several symbolic objects have to be computed beforehand using Algorithm 2.

The inputs of Algorithm 2, namely, the bases of holonomic ideals I^-, I^t, and I^o are obtained as follows. First, the one-dimensional Gaussian

$$\mathcal{N}(x^-|\mu^-, \Sigma^-) = \frac{1}{\sqrt{2\pi\Sigma^-}} \exp\left\{-\frac{(x^- - \mu^-)^2}{2\Sigma^-}\right\}$$

is annihilated by three differential operators:

$$\partial_{x^-} + \partial_{\mu^-}, \ \partial_{\Sigma^-} - \frac{1}{2}\partial_{x^-}^2, \text{ and } \Sigma^- \partial_{x^-} + x^- - \mu^-. \tag{21}$$

It can be verified that these differential operators generate a holonomic ideal that annihilates $\tilde{p}_{\text{joint}}(\xi, z)$ (see[29] for an algorithm to perform this verification). Hence, (21) can be regarded as a basis of I^- in Algorithm 2. From (5), $p(x|x^-, u)$ is obtained by substituting (19) into the PDF of $\mathcal{N}(0, 1)$ as

$$p(x|x^-, u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left\{x - \frac{4x^-}{1 + (x^-)^2} - u\right\}^2\right].$$

This PDF is annihilated by the following three differential operators:

$$\partial_x + x, \ \partial_u - u, \text{ and } \frac{\left\{(x^-)^2 + 1\right\}^2}{4(x^-)^2 - 4}\partial_{x^-} + x - \frac{4x^-}{1 + (x^-)^2} - u. \tag{22}$$

After canceling the denominators, the set of differential operators becomes a basis of a holonomic ideal for $p(x|x^-, u)$. Hence, it can be regarded as a basis of I^t. Finally, $p(y|x)$ is also obtained from (6) by substituting (20) into the PDF of $\mathcal{N}(0, 9)$:

$$p(y|x) = \frac{1}{3\sqrt{2\pi}} \exp\left\{-\frac{1}{18}(y - x)^2\right\}.$$

This PDF is annihilated by

$$\partial_y + y \text{ and } \partial_x - x, \tag{23}$$

which constitute a basis of I^o, that is, a holonomic ideal that annihilates $p(y|x)$.

Hereafter, $z = [z_1 \ \dots \ z_4]^\top$ denotes the vector $[y\, u\, \mu^- \Sigma^-]^\top$. From the bases (21)–(23), a basis of a holonomic ideal that annihilates $\tilde{p}_{\text{joint}}(x, z)$ can be obtained by performing the multiplication and integration in terms of differential operators. In this example, the obtained basis of the holonomic ideal consists of eleven differential operators:

$$\partial_{z_1} + \partial_x + \partial_{z_2}, -9\partial_{z_1} + x - z_1, \ \ldots \tag{24}$$

where the remaining elements of the basis are omitted due to space limitations. By using the Fourier transform with respect to x, we obtained a basis of a holonomic ideal that annihilates $\mathcal{T}[\tilde{p}_{\text{joint}}](\xi, z)$, which is denoted by $\mathcal{B}$ hereafter. For example, the differential operators that appears in (24) are mapped to the following differential operators by the Fourier transform.

$$\partial_{z_1} - \xi + \partial_{z_2}, -9\partial_{z_1} + \partial_\xi - z_1 \in \mathcal{B}. \tag{25}$$

From $\mathcal{B}$, we compute the Pfaffian system (A3) for $\mathcal{T}[\tilde{p}_{\text{joint}}]$. Lemma 3 implies that the computed basis $\mathcal{B} \subset \mathcal{D} = \mathbf{R}[\xi, z]\langle\partial_\xi, \partial_z\rangle$ can be regarded as a basis of the zero-dimensional ideal in $\mathcal{R} = \mathbf{R}(\xi, z)\langle\partial_\xi, \partial_z\rangle$ that annihilates $\mathcal{T}[\tilde{p}_{\text{joint}}]$. Therefore, we can compute the finite set of differential operators in Lemma 6 from $\mathcal{B}$. In this example, the finite set was obtained as $\left\{\partial_{z_1}, \partial_{z_3}, \partial_{z_4}, \partial_{z_1}^2, \partial_{z_4}\partial_{z_1}, \partial_{z_4}^2\right\}$. Hence, Q becomes a seven-dimensional vector-valued function that consists of the quasi MGF and its derivatives, that is,

$$Q = \begin{bmatrix} 1 & \partial_{z_1} & \partial_{z_3} & \partial_{z_4} & \partial_{z_1}^2 & \partial_{z_4}\partial_{z_1} & \partial_{z_4}^2 \end{bmatrix}^\top \bullet \mathcal{T}[\tilde{p}_{\text{joint}}], \tag{26}$$

where $\bullet$ represents the actions of all the components on $\mathcal{T}[\tilde{p}_{\text{joint}}]$. Note that (26) consists of integrals over x and x^- because of the definitions (12) and (9), and hence, its evaluation is time-consuming.

Meanwhile, all the entries of the coefficient matrices $A_\lambda \in \mathbf{R}(\xi, z)^{7\times 7} (\lambda \in \{\xi, z_1, z_2, z_3, z_4\})$ were explicitly computed. Due to space limitations, we only present an overview of the coefficient matrix A_ξ as follows:

$$\begin{bmatrix} z_2 + \frac{4}{5}z_3 & 0 & \frac{16}{25}z_4 + 1 & 0 & 0 & 0 & 0 \\ 0 & z_2 + \frac{4}{5}z_3 & 0 & 0 & 0 & \frac{16}{25}z_4 + 1 & 0 \\ 1 & 0 & z_2 + \frac{4}{5}z_3 & 2z_4 + \frac{25}{8} & 0 & 0 & 0 \\ * & * & * & * & 0 & -\frac{9z_4}{z_3} & 0 \\ * & * & * & * & * & \frac{3z_4^2}{25z_3^2} & \frac{3z_4^3}{25z_3} \\ * & * & * & * & * & * & \frac{3}{5}z_4 \\ * & * & * & * & -\frac{6}{z_3z_4^3} & * & * \end{bmatrix},$$

where the entries $*$ are complicated rational functions of ξ and z, and their details are omitted. The coefficient vectors in (14) and (15) were computed from A_ξ as follows:

$$C^{(1)}(\xi, z) = [z_1 \ \ 9 \ \ 0 \ \ 0 \ \ 0 \ \ 0 \ \ 0], \tag{27}$$

$$C^{(2)}(\xi, z) = [z_1^2 + 9 \ \ 18z_1 \ \ 0 \ \ 0 \ \ 81 \ \ 0 \ \ 0]. \tag{28}$$

In this example, the symbolic computation that has been described above took just a few seconds. In contrast, our previous method [18] was unable to finish its offline part within 24 h for the same example. This indicates that the quasi MGF with the Fourier transform of differential operators contributes to overcoming the difficulty of our previous method in the offline computational cost, which allows the proposed method to be applied to a larger class of problems than the previous method.

At the end of the offline part, we have to choose the initial points z_{init} and compute the corresponding initial vectors $Q(0, z_{\text{init}})$. The choice of the initial points depends on the singular locus of the Pfaffian system because the integration path must not intersect; otherwise, Algorithm 1 fails (line 3). In this example, the least common multiple of all the denominators in the coefficient matrices is z_3z_4; thus, the singular locus of the Pfaffian system is obtained as

$$\mathcal{S} := \left\{(\xi, z) \in \mathbf{R}^5 | z_3z_4 = 0\right\}. \tag{29}$$

Since $z_4 = \Sigma^-$ is always positive, only the equality $z_3 = 0$ should be considered. More specifically, it is necessary to choose at least two initial points, one satisfying $z_3 > 0$ and the other $z_3 < 0$. If we have only one, there is no integration path to the other side to which the initial point does not belong. For this example, we chose the following 16 initial points:

$$\left\{[z_1 \ \ z_2 \ \ z_3 \ \ z_4]^\top | z_1 = \pm 4, z_2 = \pm 0.1, z_3 = \pm 2, z_4 \in \{2, 4\}\right\}. \tag{30}$$

The corresponding initial vectors can be computed numerically from the definition (26). We used `Maple` to compute the initial vectors with sufficient accuracy.

In the online computation, the data z that consists of u, y, μ^-, and Σ^- is given at each sampling time. To avoid the failure of the integration in Algorithm 1, an initial point and integration path should be appropriately selected from the prescribed set (30). For this example, we fixed the integration path to a line segment from an initial point z_{init} to a given data $\hat{z}$, that is, the integration path $z(s)$ is defined by

$$z(s) := s\hat{z} + (1 - s)z_{\text{init}}. \tag{31}$$

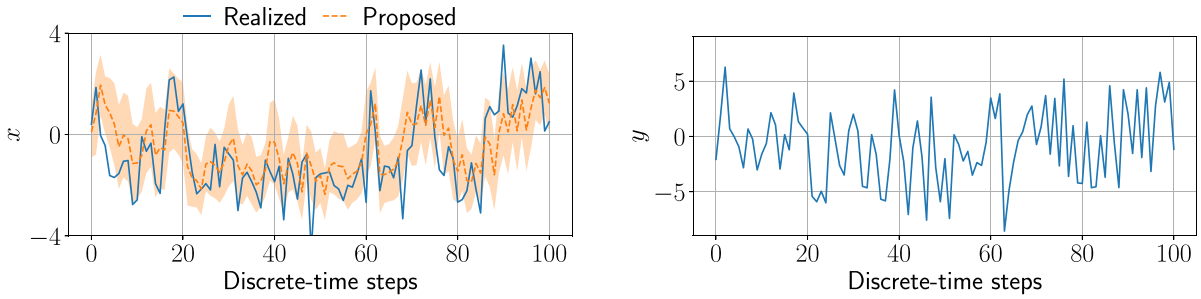
The initial point was selected for the data in such a way that the line segment did not intersect the singular locus (29) and was the shortest among the prescribed set (30). The numerical integration along the integration path is performed using ABM4 wherein the first three steps are initialized using RK4.

The proposed method was performed under the aforementioned settings. Figure1 shows a realization of the state and output trajectories of system (19)–(20) and the estimated trajectory by Algorithm 1. In Figure 1a, the dashed line (trajectory of estimated states) approximately

resembles the solid line (trajectory of true states) regardless of the existence of relatively large noise observed from the output trajectory in Figure 1b. When the estimated state deviates from the true state, for example, around the time step $k = 40$ and $k = 55$, it is also observed that the filled area is wide. Specifically, even when the estimated posterior mean is accurate, it may largely deviate from the true state because of the large true posterior variance. In this case, the estimated variance, which is depicted as the range of the filled area, should also be large. The validity of the estimated variance will also be tested in the evaluation of negative log-likelihood (NLL) later.

For comparison, we implemented the EKF, UKF, and PF. The number of particles in the PF was set to (i) 20 for its computational time to be comparable to that of the proposed method and (ii) 80 for its accuracy to be almost the same as that of the proposed method. All the methods, including the proposed method, were performed for 300 realizations.

We used the root-mean-square-error (RMSE) and averaged NLL [33] as the performance indices. The NLL is a performance index that can evaluate not only the mean but also the variance. Figure 2 shows the RMSE of each method for all realizations at each time step. The proposed method and the PF with 80 particles show similar RMSEs, which is the reason for the number of particles being set to 80. These two methods yield smaller RMSEs than the other methods, and this trend becomes clearer when the performance index is the averaged NLL as shown in Figure 3. This indicates that the proposed method can



(a) Trajectories of certain realization (solid) and its estimate (dashed) with $\pm\sqrt{\Sigma}$ region (filled area)

(b) Trajectory of observed output

FIGURE 1 Certain realization and corresponding estimate of (19) and (20)

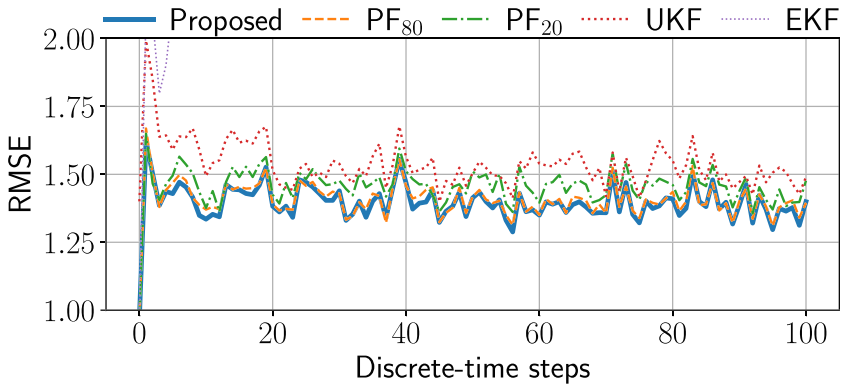


FIGURE 2 Comparison of RMSE for proposed method (solid), PFs of 80 and 20 particles (dashed and dash-dotted, respectively), UKF (dotted, thick), and EKF (dotted, thin, greater than 2.0 after the time step $k = 4$)

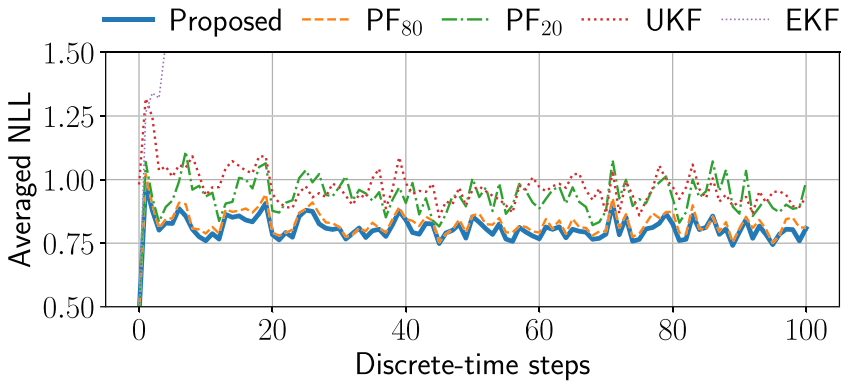


FIGURE 3 Comparison of NLL for proposed method (solid), PFs of 80 and 20 particles (dashed and dash-dotted, respectively), UKF (dotted, thick), and EKF (dotted, thin, greater than 1.5 after the time step $k = 4$)

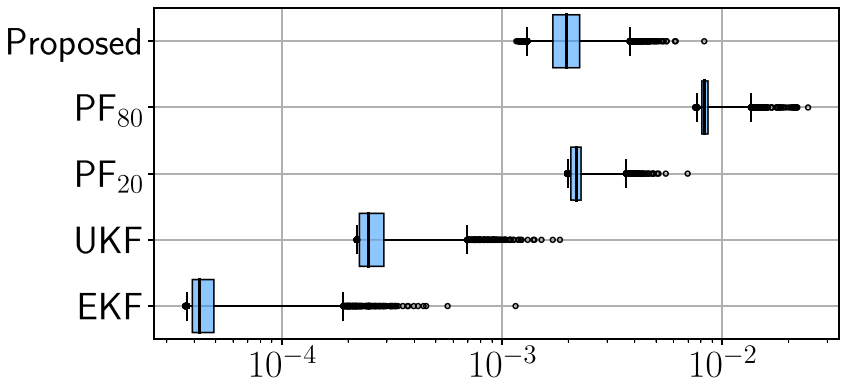


FIGURE 4 Boxplots of computational time for one-step estimations of all methods

provide a Gaussian approximation of the posterior PDF as accurately as the PF with 80 particles.

Furthermore, the computational time for the one-step estimations of all the methods are summarized as boxplots in Figure 4. It can be observed that the proposed method is faster than the PF with 80 particles. The computational time of the PF with 20 particles is, as arranged, similar to that of the proposed method, while the nonlinear Kalman filters are faster than the proposed method. However, as shown in Figure 3, these methods failed to estimate the posterior PDF as accurately as the proposed method. Consequently, the proposed method outperforms the other methods from the viewpoint of the tradeoff between computational time and accuracy. The efficiency of the proposed method results from the fact that most complicated parts of the nonlinear Bayesian filtering are performed offline, which allows us to obtain accurate estimates with relatively small online computational burdens.

7 | CONCLUSION

In this study, a symbolic-numeric Bayesian filtering algorithm for a class of nonlinear systems with stochastic noise is proposed. The mean and variance of the Gaussian approximation of the posterior PDF are updated while exactly considering the nonlinearity by using the holonomic gradient method. We introduced the notion of the quasi MGF to reduce the number of offline computations of Gröbner bases and the number of initial value problems solved online in the proposed estimation algorithm. In the numerical example, we demonstrated that the proposed method could reduce the offline computational time compared to our previous method. In addition, it is demonstrated that the proposed method is more efficient than other existing methods in terms of the trade-off between accuracy and computational cost.

The main contribution of the paper lies in the efficient evaluation of marginal mean and variance such as (10). Hence, the theoretical results of the paper may also be applied to, for example, stochastic model predictive control of uncertain discrete-time nonlinear systems, where evaluating the expectation of a cost function and predicting the state distribution of an uncertain nonlinear system are key challenges [34, 35].

It is known that ODE (11) may be stiff and difficult to solve numerically [36]. Therefore, for future work, the choice of the initial point and the integration path in the third step of the HGM can be investigated to obtain a stable ODE. This would be achieved particularly by making use of the problem-specific structure such as the quasi MGF. In another direction, we plan to use more general types of PDFs for the approximation of the posterior PDF rather than Gaussian PDFs to achieve higher accuracy.

ACKNOWLEDGEMENTS

This work was partly supported by JSPS KAKENHI Grant Numbers JP21K21285, JP18J22093, and JP15H02257.

AUTHOR CONTRIBUTIONS

Tomoyuki Iori: Conceptualization, formal analysis, methodology. **Toshiyuki Ohtsuka:** Conceptualization, methodology, supervision.

CONFLICTS OF INTEREST

The authors declare no potential conflict of interests.

ORCID

Tomoyuki Iori https://orcid.org/0000-0002-5921-8982

REFERENCES

1. R. E. Kalman, *A new approach to linear filtering and prediction problems*, ASME J. Basic Eng. **82** (1960), 35–45.
2. R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory*, ASME J. Basic Eng. **83** (1961), no. 1, 95–108.
3. I. Rusnak, *Maximum likelihood optimal estimator of non-autonomous nonlinear dynamic systems*, Proceedings of European Control Conference, 2015, pp. 909–914.

4. X. Luo, Y. Jiao, W. L. Chiou, and S. S. T. Yau, *A novel suboptimal method for solving polynomial filtering problems*, Automatica **62** (2015), 26–31.
5. B. Chen and G. Hu, *Nonlinear state estimation under bounded noises*, Automatica **98** (2018), 159–168.
6. N. Duong et al., *Laplace estimator for linear scalar systems*, Proceedings of the IEEE Conference on Decision and Control, 2018, pp. 2283–2290.
7. X. Yin, Q. Zhang, H. Wang, and Z. Ding, *RBFNN-based minimum entropy filtering for a class of stochastic nonlinear systems*, IEEE Trans. Autom. Control **65** (2020), no. 1, 376–381.
8. K. Li, F. Pfaff, and U. D. Hanebeck, *Unscented dual quaternion particle filter for SE(3) estimation*, IEEE Control Syst. Lett. **5** (2021), no. 2, 647–652.
9. E. E. Kuruoğlu, W. J. Fitzgerald, and P. J. Rayner, *Near optimal detection of signals in impulsive noise modeled with a symmetric α-stable distribution*, IEEE Commun. Lett. **2** (1998), no. 10, 282–284.
10. P. Reeves, *A non-Gaussian turbulence simulation*, 1969. AFFDL-TR-69-67, Air Force Flight Dynamics Laboratory.
11. A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, San Diego, 1970.
12. S. J. Julier and J. K. Uhlmann, *New extension of the Kalman filter to nonlinear systems*, SPIE, 1997, pp. 3068.
13. I. Arasaratnam and S. Haykin, *Cubature kalman filters*, IEEE Trans. Autom. Control **54** (2009), no. 6, 1254–1269.
14. K. Ito and K. Xiong, *Gaussian filters for nonlinear filtering problems*, IEEE Trans. Autom. Control **45** (2000), no. 5, 910–927.
15. G. Kitagawa, *Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space model*, Institute of Statistical Mathematics Research Memorandum, 1993, pp. 462.
16. G. Kitagawa, *Monte Carlo filter and smoother for non-Gaussian nonlinear state space models*, J. Comput. Graph. Stat. **5** (1996), no. 1, 1–25.
17. N. J. Gordon, D. J. Salmond, and A. F. M. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proc. F - Radar Signal Proc. **140** (1993), 107–113.
18. T. Iori and T. Ohtsuka, *Symbolic-numeric computation of posterior mean and variance for a class of discrete-time nonlinear stochastic systems*, Proceedings of the IEEE Conference on Decision and Control, 2020, pp. 4814–4821.
19. H. Nakayama et al., *Holonomic gradient descent and its application to the Fisher-Bingham integral*, Adv. Appl. Math. **47** (2011), no. 3, 639–658.
20. T. Iori and T. Ohtsuka, *Bayesian filtering for nonlinear stochastic systems using holonomic gradient method with integral transform*, Proceedings of the IEEE Conference on Decision and Control, 2021, pp. 5904–5910.
21. M. Idan and J. L. Speyer, *State estimation for linear scalar dynamic systems with additive Cauchy noises: Characteristic function approach*, SIAM J. Control. Optim. **50** (2012), no. 4, 1971–1994.
22. M. Idan and J. L. Speyer, *Multivariate Cauchy estimator with scalar measurement and process noises*, SIAM J. Control. Optim. **52** (2014), no. 2, 1108–1141.
23. Z. Zhou, X. Luan, and F. Liu, *High-order moment recursive state estimation of Markov jump linear systems*, IEEE Access **6** (2018), 70788–70793.
24. S. C. Coutinho, *A Primer of Algebraic D-Modules*, Cambridge University Press, Cambridge, 1995.
25. J. A. E. Bryson and Y.-C. Ho, *Applied Optimal Control*, 1st edn., Vol. **482**, John Wiley & Sons, Hoboken, 1975.
26. G. Goodwin, M. M. Seron, and J. A. de Doná, *Constrained Control and Estimation: An Optimisation Approach, Communications and Control Engineering*, Springer-Verlag, London, 2006.
27. H. Fang, N. Tian, Y. Wang, M. Zhou, and M. A. Haile, *Nonlinear Bayesian estimation: From Kalman filtering to a broader horizon*, IEEE/CAA J. Autom. Sinica **5** (2018), no. 2, 401–417.
28. M. Saito, B. Sturmfels, and N. Takayama, *Gröbner Deformations of Hypergeometric Differential Equations*, Springer-Verlag, Berlin Heidelberg, 2000.
29. T. Hibi Ed., *Gröbner Bases: Statistics and Software Systems*, 1st ed. *Edited by* T. Hibi, Vol. **474**, Springer Japan, Tokyo, 2013.
30. T. Oaku, Y. Shiraki, and N. Takayama, *Algebraic algorithms for D-Modules and numerical analysis*, Comput. Math. (Proc. ASCM 2003 **10** (2003), 23–39.
31. H. Tsai, Algorithms for associated primes, weyl closure, and local cohomology of D-modules, *Local Cohomology and its Applications*, CRC Press, 2001, pp. 169–194.
32. M. Noro, N. Takayama, H. Nakayama, K. Nishiyama, and K. Ohara, *Risa/Asir: A computer algebra system*, 2020. https://www.math.kobe‐u.ac.jp/Asir/asir.html
33. M. P. Deisenroth, *Efficient Reinforcement Learning Using Gaussian Processes*, Vol. **205**, KIT Scientific Publishing, Karlsruhe, 2010.
34. M. F. Huber and U. D. Hanebeck, Stochastic nonlinear model predictive control based on Gaussian mixture approcimations, *Informatics in Control, Automation and Robotics*, Springer Berlin, Heidelberg, 2009, pp. 239–252.
35. A. Mesbah, *Stochastic model predictive control: An overview and perspectives for future reasearch*, IEEE Control Systems Magazine, Vol. **36**, 2016, pp. 30–44.
36. A. Kume and T. Sei, *On the exact maximum likelihood inference of Fisher–Bingham distributions using an adjusted holonomic gradient method*, Stat. Comput. **28** (2018), no. 4, 835–847.
37. M. Kauers, The holonomic toolkit, *title=Computer Algebra in Quantum Field Theory, Integration, Summation, and Special Functions, publisher=Springer-Verlag, Wien*, pp. 119–144, 2013.

AUTHOR BIOGRAPHIES

Tomoyuki Iori received the BS degree in Engineering from Kyoto University, Kyoto, Japan, in 2016 and MS and PhD degrees in Informatics from Kyoto University in 2018 and 2021, respectively. He was a JSPS research fellow (DC1) from 2018 to 2021. He is currently an Assistant Professor at the Graduate School of Information Science and Technology, Osaka University, Osaka, Japan. His research interests include nonlinear control theory and sym-

bolic computation with algebraic techniques. He is a member of SICE, ISCIE, ORSJ, and IEEE.

Toshiyuki Ohtsuka received his BEng, MEng, and DEng from Tokyo Metropolitan Institute of Technology, Japan, in 1990, 1992, and 1995, respectively. From 1995 to 1999, he worked as an Assistant Professor at the University of Tsukuba. In 1999, he joined Osaka University and was a Professor at the Graduate School of Engineering Science from 2007 to 2013. In 2013, he joined Kyoto University as a Professor at the Graduate School of Informatics. His research interests include nonlinear control theory and real-time optimization with applications to aerospace engineering and mechanical engineering. He is a member of SICE, ISCIE, and JSASS and a senior member of IEEE and AIAA.

How to cite this article: Iori T, Ohtsuka T. *Nonlinear Bayesian filtering via holonomic gradient method with quasi moment generating function*, Asian J Control **25** (2023), 2655–2670. https://doi.org/10.1002/asjc.2970

APPENDIX A: HOLONOMIC FUNCTIONS

A holonomic function can be defined by a set of differential operators referred to as a zero-dimensional ideal in $\mathcal{R}$.

Definition 2. [29] Let I be an ideal in $\mathcal{R} = \mathbf{R}(X)\langle\partial\rangle$. The ideal I is zero-dimensional if the quotient ring $\mathcal{R}/I$ is a finite-dimensional vector space over $\mathbf{R}(X)$.

Definition 3. [29] An analytic function is called a holonomic function if there exists a zero-dimensional ideal that annihilates the function.

Remark 3. Most types of nonlinear functions that appear in the problems of systems theory are holonomic functions. For example, polynomials, rational functions, exponentials, logarithms, trigonometric functions, and their sums, products, and compositions under the mild assumptions[37] are holonomic.

The following lemma is useful to verify if a given function is holonomic.

Lemma 4. [29] *An ideal* $I \subset \mathcal{R}_n = \mathbf{R}(X)\langle\partial\rangle$ *is zero-dimensional if and only if* $I \cap \mathbf{R}(X)\langle\partial_i\rangle \neq \{0\}$ $(i = 1, \dots, n)$.

Example 4. Consider the following nonlinear function.

$$\beta(X) = \exp(\alpha(X)),$$

where $X = [X_1 \ \dots \ X_n]^\top$ and $\alpha(X) \in \mathbf{R}(X)$. It is easy to verify that $\beta(X)$ satisfies the following PDEs.

$$\begin{aligned}\partial_i \bullet \beta(X) - \partial_i\alpha(X)\beta(X) &= (\partial_i - \partial_i\alpha(X)) \bullet \beta(X) \\ &= 0 \ (i = 1 \ \dots, n).\end{aligned} \tag{A1}$$

Since every $\partial_i\alpha(X)$ is a rational function, every differential operator $\partial_i - \partial_i\alpha(X)$ is an element of $\mathbf{R}(X)\langle\partial_i\rangle$. Hence, Lemma 4 ensures that an ideal $\langle\partial_1 - \partial_1\alpha(X), \dots, \partial_n - \partial_n\alpha(X)\rangle$ is zero-dimensional; thus, the function $\beta(X)$, which is annihilated by this zero-dimensional ideal, is a holonomic function.

Holonomic functions have two remarkable properties: the closure property and finiteness of the partial derivatives.

Lemma 5. [30] *For holonomic functions* $\alpha(X)$ *and* $\beta(X)$ *in* $X = [X_1 \ \dots \ X_n]^\top$, *the following hold:*

1. *The product* $\alpha \cdot \beta$ *is also a holonomic function.*
2. *Assume that* $\alpha = \alpha(X_1, \dots, X_n)$ *is infinitely differentiable on* $\mathbf{R}^n$ *and rapidly decreasing with respect to* X_n *for any* $[X_1 \ \dots \ X_{n-1}]^\top \in \mathbf{R}^{n-1}$. *Then, the integral*

$$\int_{-\infty}^{\infty} \alpha(X) dX_n \tag{A2}$$

is also a holonomic function of $[X_1 \ \dots \ X_{n-1}]^\top$.

By virtue of the closure property, even when a given function is described as a product of two nonlinear functions, we can determine if the product is holonomic by checking whether each of the two nonlinear functions is holonomic; the same is true for an integral of a nonlinear function. Moreover, this property is applicable recursively. The integral of the product, the product of the integrals, and more complicated expressions can be defined by integrations and multiplications. These can be verified to be holonomic by checking whether all the nonlinear functions that are used for their construction are holonomic. It should be noted that if we have bases of zero-dimensional ideals for all the nonlinear functions, a basis of a zero-dimensional ideal for the complicated expression can be computed symbolically by using the

integrations and multiplications in terms of differential operators (see Appendix B for a brief introduction and[30] for details).

Example 5. The function

$$\alpha(X_1) = \int_{\mathbf{R}} \exp(-X_1^2 - X_2^2)\cos(X_2)dX_2$$

is holonomic because it is defined as an integral of product of the holonomic functions $\exp(-X_1^2 - X_2^2)$ and $\cos(X_2)$, and the integrand is rapidly decreasing with respect to X_2.

Meanwhile, the finiteness of the partial derivatives is stated by the following lemma.

Lemma 6. [29] *For a holonomic function $\alpha(X)$ of n variables $X = [X_1 \ \dots \ X_n]^\top$, there exist a finite number of its partial derivatives $\partial^{d_1}\alpha(X), \dots, \partial^{d_{q-1}}\alpha(X)$ such that the q-dimensional vector-valued function $Q(X) := [\alpha(X)\partial^{d_1}\alpha(X) \ \dots \ \partial^{d_{q-1}}\alpha(X)]^\top$ satisfies the following PDEs:*

$$\partial_{X_i} Q(X) = A_{X_i}(X)Q(X) \qquad (i = 1, \dots, n), \qquad \text{(A3)}$$

where each $A_{X_i}(X) \in \mathbf{R}(X)^{q\times q}$ is a matrix-valued function in which all the components are rational functions of X.

The set of PDEs (A3) is called the *Pfaffian system* for the holonomic function $\alpha(X)$. Lemma 6 shows that every first-order partial derivative of $Q(X)$ can be expressed as a vector that comprises linear combinations of the components of $Q(X)$ with coefficients in $\mathbf{R}(X)$. Moreover, by differentiating both sides of (A3), it can be readily shown that all the higher order partial derivatives of $Q(X)$ can also be expressed by such linear combinations. Every partial derivative of $\alpha(X)$ can be expressed as such a linear combination because $\alpha(X)$ is the first component of $Q(X)$. In other words, all the partial derivatives of $\alpha(X)$ are elements of a finite-dimensional vector space spanned by the components of $Q(X)$ over $\mathbf{R}(X)$.

Let I be a zero-dimensional ideal annihilating a holonomic function $\alpha(X)$. If we have a basis of I, we can find the differential operators $\partial^{d_1}, \dots, \partial^{d_{q-1}}$ in Lemma 6 and symbolically compute the Pfaffian system, that is, the matrices $A_{X_i}(X)(i = 1, \dots, n)$ in (A3)[29]. The dimension of $Q(X)$, namely, the number q, is called the *holonomic rank* of I and does not depend on the basis used to compute the Pfaffian system.

APPENDIX B: MULTIPLICATIONS AND INTEGRATIONS IN TERMS OF DIFFERENTIAL OPERATORS

As mentioned in Appendix A, the notion of holonomic functions is closely related to differential operators. By making use of this relationship, many manipulations of holonomic functions can be interpreted as manipulations of holonomic ideals[30]. For the completeness of this paper, we summarize the algorithms of holonomic ideals corresponding to the multiplication and integration of holonomic functions, which we call *the multiplication and integration in terms of differential operators* with a slight abuse of terminology.

Lemma 7. *(Multiplication) Let $\alpha(X)$ and $\beta(X)$ be holonomic functions, which are solutions of holonomic ideals I_α and I_β, respectively. By replacing the indeterminates of β and I_β with Y, we can define an ideal*

$$I_{\alpha\otimes\beta} := \{a_1 l_\alpha + a_2 l_\beta | l_\alpha \in I_\alpha; l_\beta \in I_\beta; a_1, a_2 \in \mathcal{D}_{2n}\}$$
$$\subset \mathcal{D}_{2n} = \mathbf{R}[X, Y]\langle\partial_X, \partial_Y\rangle.$$

This ideal can be regarded as an ideal of $\mathbf{R}[X, Z]\langle\partial_X, \partial_Z\rangle$ by applying a coordinate transformation $Z = X - Y$. Then, the product $\alpha(X)\beta(X)$ is a solution of the restriction of $I_{\alpha\otimes\beta}$ with respect to Z.

Lemma 8. *(Integration) Let $\alpha(X)$ be a holonomic function of variables $X = [X_1 \ \dots \ X_n]^\top$, which is a solution of a holonomic ideal $I \subset \mathcal{D}_n = \mathbf{R}[X]\langle\partial_X\rangle$. Moreover, assume that α is infinitely differentiable on $\mathbf{R}^n$ and rapidly decreasing with respect to X_n. Let $\mathcal{D}_{n-1}$ be $\mathbf{R}[X_1, \dots, X_{n-1}]\langle\partial_1, \dots, \partial_{n-1}\rangle$. Then, integral (A2) is a solution of the holonomic ideal $(I + \partial_n\mathcal{D}_n) \cap \mathcal{D}_{n-1}$.*

The computations of holonomic ideals that correspond to the multiplication and integration are reduced to the computation of a restriction of the holonomic ideals[28]. In addition, their computation algorithms are given in[30]. Therefore, for any function that is defined by the integrations and multiplications such as (9), a basis of a holonomic ideal that annihilates the function can always be computed algorithmically by applying Lemmas 7 and 8 recursively.