Recursive Elimination Method in Moving Horizon Estimation for a Class of Nonlinear Systems and Non-Gaussian Noise

Tomoyuki Iori * and Toshiyuki Ohtsuka *

Abstract: This paper proposes a recursive elimination method for optimal filtering problems of a class of discrete-time nonlinear systems with non-Gaussian noise. By this method, most of the computations to solve an optimal filtering problem can be carried out off-line by using symbolic computation based on the results from algebraic geometry. This property is suitable for moving horizon estimation, where a certain optimal filtering problem must be solved for different measurement sequences in each sampling interval. A numerical example is provided to compare the proposed method with other state estimation methods including the particle filter, and the efficiency of the proposed method is shown.

Key Words: nonlinear estimation, non-Gaussian distribution, moving horizon estimation, commutative algebra.

1. Introduction

After the great success of the Kalman filter (KF) [1],[2] in both theoretical and practical aspects, the study of optimal filtering has been developed to consider the problem setting of nonlinear systems with non-Gaussian noise, which is one of the most important problem settings in control engineering [3]–[5].

The most popular result in the optimal filtering theory for nonlinear systems is the extended Kalman filter (EKF) [6], which uses linear approximations to extend the concept of the KF to nonlinear cases. Its algorithm is simple and analogous to that of the KF, which not only makes it easy for us to imagine how the EKF works but also leads to a small computational cost. However, the use of linear approximations in its algorithm limits the accuracy of the EKF so that it can have surprisingly bad performance in some practical applications with nonlinear systems and non-Gaussian noise [7]. Another extension of the KF is the unscented Kalman filter (UKF) [8]. This filtering method uses the unscented transformation to approximate the distribution distorted by nonlinear systems with the Gaussian distribution whose mean and covariance are the same as those of the distorted one. Therefore, the UKF often provides more accurate estimates than the EKF (e.g., see [9]). However, both EKF and UKF require the assumption that the system and measurement noises are Gaussian.

In some applications, such an assumption does not hold. Chen and Hu [10] have proposed a new Kalman-like filtering method for nonlinear systems, which does not require any statistical information of noise but only its bounds. They constructed an upper bound of estimation error by using a Taylor series approximation and computed the gain of the filter that minimizes the upper bound. In some numerical examples, this new filtering method outperformed the EKF, UKF, and cubature Kalman Filter (CKF) with inaccurate covariances of system and measurement noises. It is often difficult to certify that the noises are Gaussian, and this method is promising for such cases. However, the bounds of noise are sometimes unavailable when, for example, the noise has an impulsive character that produces deviations of high amplitude with short durations more often than the Gaussian distribution does [11]. It is known that radar and sonar noise [12] and disturbances caused by air turbulence [13] have the impulsive character, which has motivated several researchers to model such impulsive noise by using non-Gaussian distributions [11],[12],[14],[15]. With this in mind, it is still important to study optimal filtering methods for nonlinear systems with non-Gaussian noise even if we assume that their statistical properties or distributions are known.

As a popular approach for nonlinear filtering with non-Gaussian noise, the particle filter (PF) was first proposed by Kitagawa [16],[17] and Gordon et al. [18]. This filter uses the Monte Carlo method to approximate the posterior distribution of the state, and thus Gaussian assumptions for noise distributions are not required. Although the PF with enough particles yields more accurate estimates than those derived from the EKF and UKF, it usually has a high computational cost, which can reach an unacceptable level for real-time purposes.

Another promising approach is called moving horizon estimation (MHE). In MHE, an estimation of the finite length of state trajectory is performed at every sampling time by using the same length of measurement sequence, and the length of the estimated state trajectory is called the horizon. While the EKF, UKF, and other Kalman-like filters estimate the current state by using the fixed previous estimate, MHE reoptimizes the past estimates on the horizon and estimates the current state by using those reoptimized past estimates. Indeed, it has been reported that this reoptimization property improves the current state estimate in some examples [19]. The estimation problem at each sampling time is formulated as a nonlinear optimization problem, which is usually computationally demanding to solve while the sampling interval is limited. Therefore, it is important to reduce its computational time.

In most of the efficient algorithms for MHE [20]–[23], the objective function minimized in each optimization problem is assumed to be quadratic, which is equivalent to the Gaussian...
assumption for noise distributions. For non-Gaussian cases, several methods have been developed to approximate the distributions of non-Gaussian noise by using simple distributions such as Gaussian or uniform distributions [24]. Monin [3] proposed an MHE algorithm for non-Gaussian cases by approximating the non-Gaussian distributions with the Gaussian mixture defined by the Max operator. Although these approximating methods can deal with a wide range of non-Gaussian noise, the computational complexity grows combinatorially with increasing the number of simple distributions used in the approximation. Hence, it gets more difficult to reduce the computational time for solving each optimization problem when we consider non-Gaussian cases.

To tackle such difficulties of MHE with nonlinear systems and non-Gaussian noise, note that the MHE procedure is analogous to model predictive control (MPC) in nonlinear control theory. Indeed, if we choose the joint maximum a posteriori (MAP) estimation as the optimality criterion, the derived optimization problem has a structure similar to that of an optimal control problem with a fixed horizon [24]–[26]. Therefore, it is worth exploring algorithms for MHE that make use of this analogy with MPC. For optimal control problems, Ohtsuka [27] proposed a recursive elimination method, where the variables other than the initial state and input are eliminated by using the concept of elimination ideals in commutative algebra. This method yields an implicit function representation of the initial optimal input as the function of the initial state. This representation is especially useful in MPC (e.g., see the example in [28]) because only the initial optimal input is needed in MPC. The variable elimination in this method is computed by symbolic computation, which can be performed off-line. Although several algebraic assumptions are needed to apply this method, the off-line variable elimination can reduce the time for the on-line numerical computations to solve the optimization problem at each time step.

Now, the analogy between MHE and MPC has led us to propose a recursive elimination method for optimal filtering problems. In the joint MAP estimation, the state trajectory on the horizon is computed to maximize the joint probability density of all the states conditionally on the measurements. However, in the MHE procedure, only the estimate of the current state is used as a solution for the filtering problem at each time step, which means there is no need to compute the other past estimates. Hence, by using the off-line variable elimination technique in the joint MAP estimation, we can eliminate the past state variables from the filtering problem to reduce the on-line computational time. As a preliminary result, we have proposed a recursive elimination method for optimal filtering problems of polynomial systems with a certain class of noise [29]. Compared to this previous work, in this paper, we deal with a wider class of problems where we can use rational functions (instead of polynomials) to describe systems or distributions of noise. This extension allows us to deal with, for example, the Cauchy distribution, which can represent an impulsive noise [14] and cannot be handled by the preliminary result [29].

The rest of this paper is organized as follows. In Section 2, we formulate the problems dealt with in this paper, derive the stationary conditions, and convert them into algebraic equations by introducing additional variables. Section 3 is devoted to introducing a recursive elimination method for optimal filtering problems. In Section 4, an MHE algorithm that uses the implicit function representation derived from the recursive elimination is proposed. In Section 5, a numerical example is provided to compare the proposed MHE algorithm with other state estimation methods and show the efficiency of the proposed method. Section 6 concludes this paper.

**Notations**

The subscript $k \in \mathbb{Z}_{\geq 0}$ denotes the time step of discrete-time systems. For vector variables $X_k \in \mathbb{R}^n$ depending on the discrete time $k$, the sequence $\{X_k, \ldots, X_0\} \in \mathbb{R}^n$ is abbreviated as $\{X_{[k+1]}\}$. For the field of real numbers $\mathbb{R}$ and vectors $X = [X_1 \ldots X_n]^T$ and $Y = [Y_1 \ldots Y_m]^T$, $[X, Y]$ denotes the ring of polynomials in the components of $X$ and $Y$ over $\mathbb{R}$. An ideal $I \subset \mathbb{R}[X]$ generated by the set of polynomials $\{f_1, \ldots, f_s\} \subset \mathbb{R}[X]$ is defined as $\langle f_1, \ldots, f_s \rangle := \{a_1 f_1 + \cdots + a_s f_s \mid a_1, \ldots, a_s \in \mathbb{R}[X]\}$, and the set $\{f_1, \ldots, f_s\}$ is called generators of the ideal $I$. For an ideal $I \subset \mathbb{R}[X]$, $\mathcal{V}(I) \subset \mathbb{R}^n$ is a set of elements in $\mathbb{R}^n$ where all polynomials in $I$ vanish; it is called the algebraic set defined by $I$. If an ideal $I$ is generated by $\{f_1, \ldots, f_s\}$, $\mathcal{V}(I)$ equals the set of elements where all the generators $f_1, \ldots, f_s$ vanish [30].

2. Problem Formulation

Consider the following joint MAP estimation problem with a finite horizon $N \in \mathbb{Z}_{\geq 0}$, where the joint probability density function (PDF) $p(x_{0:N} | y_{0:N})$ of the state trajectory $x_{0:N} \subset \mathbb{R}^n$ conditionally on the measurements $y_{0:N} \subset \mathbb{R}^m$ is maximized with respect to a given set of inputs $u_{0:(N-1)} \subset \mathbb{R}^m$:

$$
\begin{align*}
x^*_{0:N} & := \arg \max_{x_{0:N}} p(x_{0:N} | y_{0:N}) \quad (1) \\
\text{s. t.} \quad & x_k = f_{k-1}(x_{k-1}, u_{k-1}) + w_k, \quad (2) \\
& y_k = h_k(x_k) + v_k, \quad (3) \\
& w_k \sim p_w(w_k), \quad (4) \\
& v_k \sim p_v(v_k), \quad (5) \\
& x_0 \sim p_0(x_0, \mu), \quad (6)
\end{align*}
$$

where $w_k$ and $v_k$ are i.i.d. noise vectors of system dynamics and observation, respectively, and $\mu \in \mathbb{R}^n$ is a vector of parameters that describes the time evolution of the PDF of the initial state on the horizon; in the MHE context, this PDF corresponds to the so-called arrival cost [20]. In this problem, although functions $f_k, h_k, p_w, p_v$, and $p_0$ are assumed to be given, $p_0$ is what we have to choose. Typically, $p_0$ is chosen as Gaussian with the mean of the past estimate, which corresponds to the quadratic arrival cost [20],[22]. In this case, $\mu$ is a parameter denoting the past estimate. For a more reasonable design method, we can take into account the state equation (2) and the PDF of system noise (4); $p_0$ can be defined as

$$
p_0(x_0, \mu) := p_w(x_0 - f_0(\mu_1, \mu_2)),
$$

where $\mu = [\mu^T_1 \mu^T_2]^T \in \mathbb{R}^{n+m}$ is the parameter representing the state and input at $N + 1$ real-time steps ago. This would be a better choice than the typical one if a Gaussian distribution cannot approximate the PDF of system noise $p_w$.

To solve the joint MAP estimation problem (1)–(6), we first reformulate this problem as an unconstrained nonlinear optimization problem. According to Bayes’ rule, the conditional joint PDF in (1) can be written as
p(x_{0:N} \mid y_{0:N}) = \frac{p(x_{0:N}, y_{0:N})}{p(y_{0:N})}.

The denominator on the right-hand side is the joint probability of the output trajectory $y_{0:N}$ marginalized over all state trajectories and thus is independent of $x_{0:N}$. Therefore, the maximization (1) is equivalent to another maximization:

$$
x^*_{0:N} = \arg \max_{x_{0:N}} p(x_{0:N}, y_{0:N}).
$$

The state trajectory $x_{0:N}$ is Markovian due to the constraint of the state equation (2), and similarly, the output $y_k$ at a time step $k$ only depends on the state $x_k$ at the same time step due to the observation equation (3). Therefore, the joint PDF in (7) can be rewritten as the product of the transition PDFs $p(x_k \mid x_{k-1})$, observation PDFs $p(y_k \mid x_k)$, and PDF of the initial state $p_0(x_0; \mu)$ [24]:

$$
p(x_{0:N}, y_{0:N}) = p(y_{0} \mid x_{0})p_0(x_0; \mu) \prod_{k=1}^{N} p(y_k \mid x_k)p(x_k \mid x_{k-1}).
$$

Hence, by taking the logarithms on both sides, maximization (7) is reformulated as minimization of the sum of logarithms:

$$
- \ln p(x_{0:N}, y_{0:N}) = - \{\ln p_0(x_0; \mu) + \ln p(y_0 \mid x_0)\}
- \sum_{k=1}^{N} [\ln p(x_k \mid x_{k-1}) + \ln p(y_k \mid x_k)].
$$

(9)

From the rule of transformation of PDFs, the conditional PDFs in (9) are rewritten as

$$
p(x_k \mid x_{k-1}) = p_{\nu}(w_{k-1}) \left| \frac{\partial x_k}{\partial w_{k-1}} \right|^{-1} = p_{\nu}(w_{k-1}),
$$

and similarly,

$$
p(y_k \mid x_k) = p_{\nu}(v_k) \left| \frac{\partial y_k}{\partial v_k} \right|^{-1} = p_{\nu}(v_k),
$$

(11)

where the derivatives $\partial x_k / \partial w_{k-1}$ and $\partial y_k / \partial x_k$ are both identity matrices owing to (2) and (3), and thus the absolute values of their determinants are 1. In the end, by substituting the state and observation equations into (10) and (11), respectively, and then substituting these equations into the objective function (9), the joint MAP estimation problem is reformulated as the following unconstrained nonlinear optimization problem:

$$
\min_{p_0(x_0; \mu)} \left\{ \ln p_0(x_0; \mu) + \ln p_{\nu}(y_0 - h_0(x_0)) \right\}
- \sum_{k=1}^{N} [\ln p_{\nu}(x_k - f_k(x_{k-1}, u_{k-1})) + \ln p_{\nu}(y_k - h_k(x_k))].
$$

(12)

By differentiating the objective function with respect to the state trajectory, the stationary conditions for the minimization problem (12) can be obtained as follows:

$$
- \frac{1}{p_0} \frac{\partial p_0}{\partial x_0} + \frac{1}{p_{\nu}} \frac{\partial p_{\nu}}{\partial v_0} \left|_{y_0=y_0-h_0} \right| \frac{\partial h_0}{\partial x_0} + \frac{1}{p_{\nu}} \frac{\partial p_{\nu}}{\partial w_0} \left|_{y_0=y_0-h_0} \right| \frac{\partial f_k}{\partial x_0} = 0,
$$

(13)

$$
- \frac{1}{p_{\nu}} \left. \frac{\partial p_{\nu}}{\partial w_{k-1}} \right|_{y_k=y_k-h_k} + \frac{1}{p_{\nu}} \frac{\partial p_{\nu}}{\partial v_k} \left|_{y_k=y_k-h_k} \right| \frac{\partial h_k}{\partial x_k} + \frac{1}{p_{\nu}} \frac{\partial p_{\nu}}{\partial w_k} \left|_{y_k=y_k-h_k} \right| \frac{\partial f_k}{\partial x_k} = 0,
$$

(14)

$$
- \frac{1}{p_{\nu}} \frac{\partial p_{\nu}}{\partial w_N} \left|_{y_N=y_N-h_N} \right| \frac{\partial h_N}{\partial x_N} = 0,
$$

(15)

where the arguments of all functions are omitted for simplicity. We assume that the unconstrained problem (12) has at least one global minimizer $x^*_{0:N} \in \mathbb{R}^{n(N+1)}$ with a finite minimum value of the cost function, and thus (13)–(15) are necessary conditions for optimality. Note that when the parameter $\mu$ and the histories of inputs $u_{0:N-1}$ and measurements $y_{0:N}$ are given, (13)–(15) can be viewed as a two-point boundary value problem for the sequence of tuples $(x_k, x_{k-1}) (k = 1, \ldots, N)$ because (13) and (15) can be regarded as the initial and the terminal conditions, respectively.

Now, we assume that functions $f_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ and $h_k : \mathbb{R}^p \to \mathbb{R}^r$ in the state and observation equations (2) and (3), respectively, are vectors of rational functions. In addition, we also assume that the derivatives of logarithms

$$
\frac{\partial}{\partial w_k} [\log p_\nu(w_k)] = \frac{1}{p_\nu(w_k)} \frac{\partial p_\nu(w_k)}{\partial w_k},
\frac{\partial}{\partial v_k} [\log p_\nu(v_k)] = \frac{1}{p_\nu(v_k)} \frac{\partial p_\nu(v_k)}{\partial v_k},
\frac{\partial}{\partial x_0} [\log p_0(x_0, \mu)] = \frac{1}{p_0(x_0, \mu)} \frac{\partial p_0(x_0, \mu)}{\partial x_0}
$$

are described by rational functions. Under these assumptions, all the left-hand sides of (13) and (15) consist of rational functions.

**Remark 1** The assumption that the derivatives (16) consist of rational functions is not so restrictive. For example, if $p_\nu$ is described as an exponential of a rational function $g$,

$$
p_\nu(w_k) = \exp(g(w_k)),
$$

then its logarithm $g$ and its derivative of the logarithm $\partial g / \partial w_k$ are both rational functions. Moreover, even if $p_\nu$ itself is a rational function $g$, its derivative of the logarithm $\partial (\log g) / \partial w_k = (\partial g / \partial w_k) / g$ is a rational function though its logarithm $\log g$ is not.

**Remark 2** Although the unconstrained problem (12) is considered in the paper, equality and inequality constraints can be taken into account if they consist of rational functions. In this case, the stationary conditions (13)–(15) are replaced by the Karush-Kuhn-Tucker (KKT) conditions, and thus we can apply the following discussions and algorithms by using some minor modifications to deal with the Lagrange multipliers associated with the additional constraints. Note, however, that a global minimizer of the constrained optimization problem may no longer be interpreted as a MAP estimation, that is, a maximizer of the conditional joint PDF (1). This is because, for example, a state constraint can destroy the independence of the noise sequence $[w_k]_{k=1}^m$, which we assume throughout the paper.
(see [31] for details). As an example of constraints that are compatible with the MAP estimation interpretation, inequality constraints independent of the state such as box constraints \( w_{i < \omega_i < \omega_i} (i = 1, \ldots, n) \) can be considered for either or both \( w \) and \( v \).

To apply the algebraic geometry techniques used in [27],[28], we next convert these equations into algebraic equations. This conversion can be performed for each time step \( k \) by canceling the denominators of the \( n \) equations at the time step; first, let us select \((3)\), which consists of an \( n \)-dimensional vector-valued rational function and can be rewritten as follows:

\[
D_0^{-1} (x_0, x_1, \mu, u_0, y_0) n_0 (x_0, x_1, \mu, u_0, y_0) = 0, \tag{17}
\]

where \( n_0 \) is a vector of polynomials whose \( i \)-th component \( n_{0,i} \) is the numerator of the left-hand side of the \( i \)-th equation in \((3)\) and \( D_0 \) is a diagonal matrix-valued function whose \( i \)-th diagonal element is the denominator polynomial corresponding to \( n_{0,i} \). It is obvious that the solution of \((3)\) satisfies a set of \( n \) algebraic equations

\[
n_0 (x_0, x_1, \mu, u_0, y_0) = 0. \tag{18}
\]

To prevent all denominators from vanishing, consider an algebraic equation

\[
1 - d_0 \det D_0 (x_0, x_1, \mu, u_0, y_0) = 0, \tag{19}
\]

where \( d_0 \in \mathbb{R} \) is an additional scalar variable, and \( \det D_0 \) denotes the product of all the diagonal polynomials of \( D_0 \). Note that if \((19)\) has a solution, the value of \( \det D_0 \) is given as \( \det D_0 = 1/d_0 \neq 0 \), which means any denominator included in \((3)\) does not vanish at the solution; in other words, all the denominators do not vanish as long as \( (19) \) holds. Consequently, \((3)\) is equivalent to \((n+1)\) algebraic equations, namely, \((18)\) and \((19)\). Note that if \( \det D_0 \) is strictly positive or negative on its domain, \((19)\) always has a solution and can be omitted.

The process of the conversion from \((3)\) to the algebraic equations \((18)\) and \((19)\) can also be performed for \((14)\) and \((15)\) in the same way. In the following discussion, we denote a vector of polynomials \([n_i^T 1 - d_0 \det D_0]^T \) by \( F_0 \) and derive the vectors of polynomials \( F_k (k = 1, \ldots, N - 1) \) and \( F_N \) from \((14)\) and \((15)\), respectively, in the same way. In the end, all equations \((13)-(15)\) can be equivalently converted into the following equations:

\[
F_0 (x_0, x_1; \mu, u_0, y_0, d_0) = 0, \tag{20}
\]

\[
F_k (x_{[k-1,k+1]}; u_{k-1}, u_k, y_k, d_k) = 0 \quad (k = 1, \ldots, N - 1), \tag{21}
\]

\[
F_N (x_{[N-1]}, x_N; u_{N-1}, y_N, d_N) = 0. \tag{22}
\]

Equations \((20)-(22)\) are still necessary optimality conditions because they are equivalent to the original stationary conditions \((13)-(15)\). By solving these equations for given parameter \( \mu \) and histories of inputs \( u_{[0,N-1]} \) and outputs \( y_{[0,N]} \), we can obtain the candidates of the optimal solution of the MAP estimation problem \((1)-(6)\). However, it is still difficult and time-consuming to solve \((20)-(22)\) because they include all the state variables on the horizon and are coupled by these variables. Recall that, in the MHE procedure, just the terminal estimate \( x_N^* \) is used as the estimate of the current state, which means we do not need to compute the past estimates \( x_{[0,N-1]}^* \) on-line. By using algebraic geometry techniques, we can eliminate the past estimates from \((20)-(22)\) off-line without any approximation and obtain equations only in the terminal state \( x_N \) and other parameters \( \mu, u_{[0,N-1]} \), and \( y_{[0,N]} \), which leads to a reduction of the computational cost.

3. Recursive Elimination

If the state and observation equations \((2)-(3)\) are linear and the PDFs of \( w_k \) and \( v_k \) are Gaussian, \((20)-(22)\) (which are exactly the same as \((13)-(15)\) in linear cases) are explicitly solvable and have a unique solution, which corresponds to the KF [32]. On the other hand, in nonlinear cases, these equations are no longer explicitly solvable, and we have to rely on numerical methods such as Newton's method to solve them. Generally speaking, nonlinear equations become harder to solve as the number of variables increases, which means decreasing the number of variables can reduce the computational cost for solving the equations; indeed, in the numerical examples introduced in [27],[33], computational costs are reduced by decreasing the numbers of variables. Elimination theory [30] in algebraic geometry is one of the most powerful tools to decrease the number of variables, and it can eliminate some of the variables from nonlinear equations symbolically. It is shown by several numerical examples in [27],[28] that this off-line variable elimination leads to a reduction of the on-line computational cost.

In MHE, as mentioned before, we only need to compute the terminal estimate \( x_N^* \), and the other estimates \( x_{[0,N-1]}^* \) are introduced just for connecting \( x_N^* \) to the past measurements \( y_{[0,N-1]} \). Therefore, by eliminating the estimates \( x_{[0,N-1]}^* \) from \((20)-(22)\), we can reduce the on-line computational cost to solve the MAP estimation problem \((1)\). In this section, we introduce the recursive elimination method for \((20)-(22)\). Although it is similar to the method proposed in [29], the main difference is the existence of the additional variables \( d_i \).

Before introducing the recursive elimination method, let us briefly explain the concept of elimination ideals. Let us consider an ideal \( I \subset R[X,Y] \) in indeterminates \( X = [X_1 \cdots X_m]^T \) and \( Y = [Y_1 \cdots Y_m]^T \). If this ideal contains polynomials depending only on \( Y \), the set of all such polynomials \( I \cap R[Y] \) is not empty. This subset \( I \cap R[Y] \) is also an ideal [30] and is called the elimination ideal of \( I \) with respect to the variable \( X \). The algebraic sets \( V(I) \) and \( V(I \cap R[Y]) \) are related by the following lemma [30].

**Lemma 1** For an ideal \( I \subset R[X,Y], \)

\[
\pi_Y (V(I)) \subset V(I \cap R[Y])
\]

holds, where \( \pi_Y : R^n \times R^m \to R^m \) defined by \((X, Y) \mapsto Y \) is the projection map onto the \( Y \)-space, i.e., \( \pi_Y (V(I)) = \{ Y \in R^m \mid \exists X \in R^n \ s.t. \ (X, Y) \in V(I) \} \).

This lemma explains how the computation of elimination ideals corresponds to the variable elimination from multivariate algebraic equations; indeed, the algebraic sets \( V(I) \) and \( V(I \cap R[Y]) \) can also be viewed as the root sets of generators of \( I \) and \( I \cap R[Y] \), respectively. In other words, Lemma 1 guarantees that, by computing generators of \( I \cap R[Y] \), we can obtain a set of polynomials involving only the variable \( Y \), whose roots contain the \( Y \)-coordinates of all the roots of generators of \( I \). Note
Algorithm 1 Recursive Elimination Method for Joint MAP Filtering Problem

Input: Algebraic equations (20)–(22) for joint MAP estimation
Output: Algebraic equations $G_0(x_N,J_0;\mu,u_{0:N-1},y_{0:N}) = 0$

1: Let $f_k$ be ideal generated by set of polynomials $F_0 \subseteq \mathbb{R}[x_0,x_1,\mu,u_{0:N-1},y_0]$.
2: while $k < N$ do
3: \quad $I_k := f_k + (F_k) \subseteq \mathbb{R}[x_{k-1},x_k,\mu,u_{0:(k-1)},y_{0:k}];\mu_{k-1},y_{k}]
4: \quad f_k := I_k \cap \mathbb{R}[x_0,x_k,\mu,u_{0:k},y_{0:k},d_k]
5: \quad k \leftarrow k + 1
6: end while

7: $N := \bigcap_{j=1}^{N} f_j + (F_j) \subseteq \mathbb{R}[x_{0:N-1},\mu,u_{0:N-1},\mu_{0:N-1},y_{0:N},d_{N-1}]
8: f_N := I_N \cap \mathbb{R}[x_N,\mu,u_{0:N-1},y_{0:N},d_N]
9: Let $G_N$ be a set of generators of $f_N$ that is, $f_N = \langle G_N \rangle$

that generators of $f \cap \mathbb{R}[Y]$ can be computed from generators of $f$ by using Gröbner bases, which are sets of generators satisfying certain conditions and having good properties for symbolic computation [30].

By using elimination ideals, we can eliminate the variables $x_{0:(N-1)}$ from (20)–(22). However, it is known that the computational complexity of a Gröbner basis is doubly exponential in the number of variables in the worst case [34],[35]. Therefore, the number of variables that are handled at the same time during the elimination process needs to be reduced. Now, recall that each set of equations at time step $k$ in (21) includes only the state at the time step $x_k$ and its neighbors $x_{k-1}$ and $x_{k+1}$. This adjacently coupled structure of (21) allows us to perform the elimination process recursively as Algorithm 1.

Remark 3 Due to the expensive computational complexity of a Gröbner basis, the complexity of Algorithm 1 mainly results from the computations of elimination ideals and can be quite high. However, computation of each elimination ideal $f_j$ from $f_k$ in Algorithm 1 involves much fewer variables than in all equations (20)–(22), which means that using Algorithm 1 is more efficient than directly eliminating the state history $x_{0:(N-1)}$ from the whole set of equations (20)–(22) all at once.

From Lemma 1, we can derive a relationship between the algebraic set defined by $f_k$ in Algorithm 1 and the solution set of (20)–(22); by letting $G_N(x_N,\mu,u_{0:N-1},y_{0:N},d_N) \subseteq \mathbb{R}[x_0,x_N,\mu,u_{0:N-1},y_{0:N},d_N]$ be a set of generators of $f_N$, this relationship is stated as follows.

Lemma 2 Let sequence $x_{N}^{*}$ be the joint MAP trajectory with respect to a certain parameter $x_{N}^{*}$ and histories of inputs $u_{0:N-1}$ and measurements $y_{0:N}$. Moreover, let $\bar{u}_{0:N}$ be a sequence that satisfies (20)–(22) with $x_{N}^{*}$, $\mu$, $\bar{u}_{0:N-1}$, and $y_{0:N}$. Then, $x_{N}^{*}$, $\bar{u}_{0:N-1}$, $y_{0:N}$, and $d_N$ satisfy

$$G_N(x_N;\mu,\bar{u}_{0:N-1},y_0,d_N) = 0. \quad (23)$$

This lemma states that, for any given $\mu$, $\bar{u}_{0:N-1}$, and $y_{0:N}$, (23) can be considered as a necessary condition for optimality of $x_{N}^{*}$. This means that, roughly speaking, (23) can be viewed as an implicit function representation of the optimal estimate $x_{N}^{*}$ as a function of the parameter $\mu$ and the histories of inputs $u_{0:N-1}$ and measurements $y_{0:N}$.

4. Moving Horizon Estimation Algorithm

In the following discussion, symbol $j$ denotes a real-time step, and $k$ denotes a time step over the horizon of the joint MAP estimation problem (1)–(6). Furthermore, the estimate at time step $k$ over the horizon at real-time step $j$ is denoted by $x_{k,j}^{*}$ if necessary.

If we have $\mu = \mu(j)$ and the histories of inputs $u_{0:N-1,j}$ and outputs $y_{0:N,j}$ at real-time step $j$, we can obtain a candidate of the optimal estimate $x_{k,j}^{*}$ by substituting those data into (23) and then solving it. In general, (23) is nonlinear in $x_k$ and $d_N$ and has more than one solution, which means the obtained candidate is not necessarily optimal. However, finding the global estimate is extremely difficult and, even when possible, computationally demanding. Therefore, we need to find a better solution in a certain sense by considering the trade-off between optimality and computational cost.

One approach is solving (23) with several initial guesses that are different from each other. These initial guesses would yield several candidates for the optimal estimate, and then we can choose the best one. Note that, in general, the indeterminates of (23) are $x_k$ and $d_N$, which means we have to give initial guesses not only for $x_k$ but also for $d_N$. If we have an initial guess $x_{init}$, then the corresponding initial guess $d_{init}$ can be given from (19) at $k = N$ as

$$d_{init} = 1/ \det D_N(x_{N,j-1}^{*},x_{init},u_{N-1},y_N). \quad (24)$$

where $x_{N,j-1}^{*}$ is the optimal estimate at the previous real-time step. To determine which candidate is better, we compare the value of the PDF of $x_k$ conditioned on $x_{N-1}$ and $y_N$, which is obtained as follows:

$$p(x_k | x_{N-1}, y_N) = \frac{p(x_k,y_N | x_{N-1})}{p(y_N)} = \frac{p(y_N | x_k)p(x_k | x_{N-1})}{p(y_N)} \approx p(y_N | x_k)p(x_k | x_{N-1}). \quad (25)$$

When $x_{N-1} = x_{N-1}^{*}$, $x_k$ maximizing the PDF (25) among the solutions of $G_N = 0$ is the optimal estimate $x_{k}^{*}$. In MHE, the previous optimal estimate $x_{N-1}^{*}$ can be regarded as a good estimate of $x_{N-1}^{*}$. Therefore, we can approximate $p(x_{N,j} | x_{N-1,j})$ in (25) by $p(x_{N,j} | x_{N-1,j})$. In the end, by taking the logarithm, we can determine the optimal solution by computing and comparing the values of the following function:

$$\log p(x_{N,j} | x_{N-1,j}, y_{N,j}) \propto \log p_x(y_{N,j} - h_N(x_{N,j})) + \log p_u(x_{N,j} - f_{N-1}(x_{N,j-1}, u_{N-1,j})) + c, \quad (26)$$

where $c$ is a constant derived from $p(y_N)$ and then neglected in the comparison. Finally, the algorithm to perform MHE using the derived equation $G_N = 0$ is summarized as Algorithm 2.

Remark 4 The stability of Algorithm 2, in the sense that the error between the estimate and the true state at each time step is finite, is quite difficult to investigate because we consider non-linear systems and non-Gaussian noise and do not assume the boundedness of noise. In particular, the boundedness of noise is usually assumed to guarantee the stability and convergence of algorithms for MHE [20]–[23]. This boundedness can be taken into account in Algorithm 2 by a slight modification of the candidate evaluation (lines 6–9). The estimates of noise can be computed from a candidate $\hat{x}$, the previous estimate, and the observed output via the state equation (2) and observation.
Algorithm 2 Moving Horizon Estimation Using Implicit Function Representation

Input: Equation $G(z(x_j; \mu, u_{0:N-1}, y_{0:N-1}), d_z) = 0$, parameter $\mu(j)$ as function of real-time step $j$, histories of inputs $u_{0:N-1,j}$ and $y_{0:N-1,j}$, and set of initial guesses $X_{\text{init}}$

Output: Optimal estimate $x_{j,j}$ at real-time step $j$

1: $l_{\text{max}} \leftarrow -\infty$
2: for each $x_{\text{init}} \in X_{\text{init}}$ do
3: Compute initial guess $d_{\text{init}}$ from equation (24)
4: Compute solution $\hat{x}$ and $\hat{d}$ of $G_x = 0$ with initial guess $x_{\text{init}}$ and $d_{\text{init}}$
5: Compute value of PDF $l(\hat{x})$ by substituting (26)
6: if $l(\hat{x}) > l_{\text{max}}$ then
7: $l_{\text{max}} \leftarrow l(\hat{x})$
8: $x_{0:j,j} \leftarrow \hat{x}$
9: end if
10: end for
11: return $x_{0:j,j}$

The candidate yielding invalid noise estimates can then be discarded. Although this modification would be useful to guarantee the stability or convergence of Algorithm 2, the proof is still part of our future work.

5. Numerical Example

A numerical example is provided to show the efficiency of the proposed method. Algorithm 1 is implemented using Maple on a PC (Intel Core i7-8550U 1.80 GHz, RAM: 8 GB), and Algorithm 2 and other algorithms such as the UKF, and PF are implemented using Python on the same PC.

Consider the following one-dimensional nonlinear system:

$$x_{j+1} = \frac{1}{100} x_{j}^2 + u_j + w_j,$$

$$y_j = x_j + v_j,$$  \hspace{1cm} (27)

The system disturbance $w_j$ is sampled from a Gaussian distribution $N(0, 3)$, while the measurement noise $v_j$ is sampled from the standard Cauchy distribution, whose PDF is defined as

$$p_{v}(v_j) = \frac{1}{\pi (1 + v_j^2)}.$$ \hspace{1cm} (29)

The Cauchy distribution is a heavy-tailed distribution and does not have a mean and variance. It is known that such heavy-tailed distributions can model impulsive noise such as atmospheric and underwater acoustic noise, which are the dominant sources of the noise observed in radar and sonar applications [12]. Hence, the state estimator for dynamic systems with Cauchy noise has been intensively studied these days [11,15].

Remark 5 Due to the high computational cost of symbolic computation in Algorithm 1, the length of the horizon for this example is limited to $N = 1$. The computational cost of a Gröbner basis highly and complicatedly depends on the input polynomials and the order of variables used in the computation [30]. Hence, an order of variables that can reduce the computational cost of Algorithm 1 or a class of problems to which the algorithm can be efficiently applied could be investigated in future work.

On the other hand, for comparison, we also performed the UKF and PF to solve the same state estimation problem. The UKF requires the variance of observation noise, which cannot be defined for the Cauchy distribution. We used several different variances of 1, 100, and 10,000 for the UKF to show how the variance affects its performance. The parameter $\kappa$ of the UKF is set to 0, which, after several trials with different $\kappa$, turned out not to affect the performance in this example. For the PF, the number of particles is set to 80, which shows the same performance as the proposed method in the root mean squared error (RMSE). Although the EKF is also one of the most popular methods, to which the proposed method is compared, it is omitted in this example because the system (27) has no linear term, which yields an inaccurate linear-approximation model in the EKF and results in large estimation errors. Moreover, to show how the off-line variable elimination by Algorithm 1 contributes to the performance of the proposed method, we also perform MHE by solving the stationary conditions (13)–(15).
Fig. 2 Comparison of RMSEs for proposed method, UKF, PF, and naive MHE method.

Since the stationary conditions contain indeterminates of $x_0$ and $x_1$, a pair of initial guesses are needed to solve (13)–(15). We use the Cartesian product $X_{\text{init}} \times X_{\text{init}}$ for the set of initial guesses, obtain several optimal estimate candidates from those initial guesses, and select the best one with the same criterion as that of the proposed method. Hereafter, this method of solving the stationary conditions directly is referred to as the naive MHE method.

For 300 realizations of the system (27)–(28), state estimation is carried out by using the four methods mentioned before: the proposed method, naive MHE, UKF, and PF. Figure 2 shows the RMSEs between the true state and the estimates averaged for all realizations at each time step. It can be seen that the proposed method, PF, and the naive MHE method have almost the same performance and outperform the UKFs. Although, as can be seen in Fig. 2(a), the RMSEs of UKFs become smaller as the measurement variance increases, they cannot reach the RMSEs of the other three methods, which is indicated in Fig. 3. This figure shows the estimated trajectories that were derived from the proposed method, PF, naive MHE method, and UKF with a measurement variance of 10,000 for the same realization as in Fig. 1. All methods except for the UKF yield almost the same trajectories that are close to the realized one, while the trajectory derived from the UKF depicts a sinusoidal curve. This is because the UKF ignores all measurements due to the quite large measurement variance, which indicates that the performance of UKF cannot be improved by increasing the measurement variance further.

To show the efficiency of the proposed method, we next compare the computational times for each method. Figure 4 shows the boxplots of computational times for all time steps and realizations. Although the UKF is fastest, its performance in RMSE is much worse than those of other methods as mentioned above. On the other hand, the slowest is the naive MHE method, whose median of computational times is about six times larger than that of the PF. The proposed method is almost 30% faster than the PF when comparing the median of computational times in spite of their comparable RMSEs, which shows the efficiency of the proposed method.

Remark 6 Note that the PF is performed on a single thread in this example and thus can be accelerated by exploiting the parallel computation. Note also that lines 3–5 of Algorithm 2 can also be parallelized over all initial guesses in $X_{\text{init}}$, which will result in further acceleration of the proposed method.

Remark 7 The maximum computational time for the proposed method is about 0.03 s, which must be shorter than the period of measurements to execute Algorithm 2 in real time. Although this maximum may be undesirable for some applications, it can be shortened by limiting the number of iterations in the numerical algorithm to solve the equation $G_N = 0$ (line 4 of Algorithm 2).

6. Conclusion

In this paper, we proposed a recursive elimination method for the joint MAP estimation of a class of nonlinear systems, where the stationary conditions are written as equations consisting of rational functions. First, the estimation problem is reformulated as an unconstrained nonlinear optimization problem, and then the stationary conditions are derived. Next, the
stationary conditions are converted into algebraic equations by introducing additional variables to deal with rational functions in those conditions. By applying the concept of elimination ideals from commutative algebra, the proposed method eliminates the past state variables from the algebraic equations and obtains an implicit function representation of the optimal estimate as a function of several parameters including the measurement sequence. An MHE algorithm using this implicit representation is proposed, whose efficiency compared to other estimation methods including the particle filter is shown in a numerical example.

For future work, the proof of stability or convergence to the true solution of the proposed method will be considered. For another direction of further studies, new numerical algorithms oriented to solve the implicit function representation derived from the proposed method can be investigated.

Acknowledgments

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

References

[1] R.E. Kalman: A new approach to linear filtering and prediction problems, Journal of Basic Engineering, Vol. 82, No. 1, pp. 35–45, 1960.
[2] R.E. Kalman and R.S. Bucy: New results in linear filtering and prediction theory, Journal of Basic Engineering, Vol. 83, No. 1, pp. 95–108, 1961.
[3] A. Monin: Modal trajectory estimation using maximum Gaussian mixture, IEEE Transactions on Automatic Control, Vol. 58, No. 3, pp. 763–768, 2013.
[4] I. Rusnak: Maximum likelihood optimal estimator of non-autonomous nonlinear dynamic systems, European Control Conference, pp. 909–914, 2015.
[5] X. Luo, Y. Jiao, W.L. Chiou, and S.S. Yau: A novel suboptimal method for solving polynomial filtering problems, Automatica, Vol. 62, pp. 26–31, 2015.
[6] A.H. Jazwinski: Stochastic Processes and Filtering Theory, Elsevier Science, 1970.
[7] S. Thrun, W. Burgard, D. Fox, and R.C. Arkin: Probabilistic Robotics, MIT Press, 2005.
[8] S.J. Julier and J.K. Uhlmann: New extension of the Kalman filter to nonlinear systems, SPIE, Vol. 3068, 1997.
[9] J.L. Crassidis and F.L. Markley: Unscented filtering for spacecraft attitude estimation, Journal of Guidance, Control, and Dynamics, Vol. 26, No. 4, pp. 536–542, 2003.
[10] B. Chen and G. Hu: Nonlinear state estimation under bounded noises, Automatica, Vol. 98, pp. 159–168, 2018.
[11] J.H. Fernández, J.L. Speyer, and M. Idan: Stochastic estimation for two-state linear dynamic systems with additive Cauchy noises, IEEE Transactions on Automatic Control, Vol. 60, No. 12, pp. 3367–3372, 2015.
[12] E.E. Kurutolu, W.J. Fitzgerald, and P.J. Rayner: Near optimal detection of signals in impulsive noise modeled with a symmetric α-stable distribution, IEEE Communications Letters, Vol. 2, No. 10, pp. 282–284, 1998.
[13] P. Reeves: A non-Gaussian turbulence simulation, Technical Report AFFDL-TR-69-67, Air Force Flight Dynamics Laboratory, 1969.
[14] M. Idan and J.L. Speyer: Cauchy estimation for linear scalar systems, IEEE Transactions on Automatic Control, Vol. 55, No. 6, pp. 1329–1342, 2010.
[15] M. Idan and J.L. Speyer: State estimation for linear scalar dynamic systems with additive Cauchy noises: Characteristic function approach, SIAM Journal on Control and Optimization, Vol. 50, No. 4, pp. 1971–1994, 2012.
[16] G. Kitagawa: Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space model, Institute of Statistical Mathematics Research Memorandum, Vol. 462, 1993.
[17] G. Kitagawa: Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, Journal of Computational and Graphical Statistics, Vol. 5, No. 1, pp. 1–25, 1996.
[18] N.J. Gordon, D.J. Salmond, and A.F.M. Smith: Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEEE Proceedings F: Radar and Signal Processing, Vol. 140, No. 2, pp. 107–113, 1993.
[19] V.A. Bavdekar, R.B. Gopaluni, and S.L. Shah: A comparison of moving horizon and Bayesian state estimators with an application to a pH process, IFAC Proceedings Volumes (IFAC-PapersOnline), Vol. 46, pp. 160–165, 2013.
[20] C.V. Rao, J.B. Rawlings, and D.Q. Mayne: Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations, IEEE Transactions on Automatic Control, Vol. 48, No. 2, pp. 246–258, 2003.
[21] A. Alessandri, M. Baglietto, and G. Battistelli: Moving horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes, Automatica, Vol. 44, No. 7, pp. 1753–1765, 2008.
[22] A. Alessandri, M. Baglietto, G. Battistelli, and V. Zavala: Advances in moving horizon estimation for nonlinear systems, Proceedings of the IEEE Conference on Decision and Control, pp. 5681–5688, 2010.
[23] A. Wynn, M. Vukov, and M. Diehl: Convergence guarantees for moving horizon estimation based on the real-time iteration scheme, IEEE Transactions on Automatic Control, Vol. 59, No. 8, pp. 2215–2221, 2014.
[24] G. Goodwin, M.M. Seron, and J.A. de Doná: Constrained Control and Estimation: An Optimisation Approach, Springer London, 2006.
[25] R.E. Mortensen: Maximum-likelihood recursive nonlinear filtering, Journal of Optimization Theory and Applications, Vol. 2, No. 6, pp. 386–394, 1968.
[26] J.A.E. Bryson and Y.-C. Ho: Applied Optimal Control, 1st edition, John Wiley & Sons, 1975.
[27] T. Ohtsuka: A recursive elimination method for finite-horizon optimal control problems of discrete-time rational systems, IEEE Transactions on Automatic Control, Vol. 59, No. 11, pp. 3081–3086, 2014.
[28] T. Iori, Y. Kawano, and T. Ohtsuka: Algebraic approach to nonlinear optimal control problems with terminal constraints: Sufficient conditions for existence of algebraic solutions, SICE Journal of Control, Measurement, and System Integration, Vol. 11, No. 3, pp. 198–206, 2018.
[29] T. Iori and T. Ohtsuka: Recursive elimination method for moving horizon estimation of discrete-time polynomial systems, Proceedings of the IEEE Conference on Decision and Control, pp. 1076–1082, 2019.
[30] D.A. Cox, J. Little, and D. O‘Shea: Ideals, Varieties, and Algorithms, 4th edition, Springer International Publishing, 2015.
[31] C.V. Rao: Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems, Ph.D. dissertation, University of Wisconsin-Madison, 2000.
[32] T. Iori and Y. Kawano: A dynamic programming approach to trajectory estimation, IEEE Transactions on Automatic Control, Vol. 11, No. 3, pp. 537–540, 1966.
[33] T. Iori, Y. Kawano, and T. Ohtsuka: Algebraic approach to nonlinear finite-horizon optimal control problems with terminal constraints, 2017 Asian Control Conference, ASCC 2017, 2018.
[34] K. Kühnle and E. Mayr: Exponential space computation of Gröbner bases, International Symposium on Symbolic and Algebraic Computation, pp. 63–71, 1996.
[35] S. Gao: Counting zeros over finite fields using Gröbner bases, Ph.D. dissertation, Carnegie Mellon University, 2009.

Tomoyuki Iori (Student Member)
He received his Bachelor of Engineering and Master of Informatics degrees from Kyoto University, Japan, in 2016 and 2018. Since 2018, he has been pursuing a Doctor of Informatics degree at the Graduate School of Informatics, Kyoto University, Japan. He is a JSPS research fellow (DC1). His research interests include nonlinear control theory. He is a student member of ISCIE.

Toshiyuki Ohtsuka (Member)
He received his B. Eng., M. Eng., and D. Eng. from Tokyo Metropolitan Institute of Technology, Japan, in 1990, 1992, and 1995. 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 ISCIE, JSME, and JSASS and a senior member of IEEE and AIAA.