Signal Processing 161 (2019) 268–288

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Optimal joint estimation and identification theorem to linear Gaussian system with unknown inputs

Shixiong Wang*, Chongshou Li, Andrew Lim

Department of Industrial Systems Engineering and Management, National University of Singapore, 117576 Singapore

ARTICLE INFO

Article history:
Received 4 November 2018
Revised 8 February 2019
Accepted 4 March 2019
Available online 5 March 2019

Keywords:
Linear gaussian system
Unknown inputs
State estimation
System identification
EM Algorithm

ABSTRACT

Joint estimation and identification to the linear system with unknown input(s) (UI, UIs) is critical in the control community as well as signal processing. In this paper we present the solution to the problem based on the expectation-maximization (EM) method to alternately estimate system states and identify the UIs. The dominant advantage of the proposed method is that we could handle the UI(s) in not only the system dynamics model but also the measurement model. Specifically we make the following contributions: (1) providing the rigorous mathematical definitions of the problem, (2) theoretically proving the existence and uniqueness of the solution to the joint estimation and identification problem, (3) presenting the theoretical proof of convergence and effectiveness of the EM-based algorithm, and (4) supplying with sufficiently insightful explanations for the mathematical derivation.

1. Introduction

Although theories of linear system analysis seems at times less effective compared to some non-linear analysis techniques for certain groups of problems with non-linear properties [1–3], it is undeniable that presently for relatively large numbers of existing real industrial ones, e.g. those being with natural linear-pattern dynamics, or being without a prior knowledge of dynamics so as to unavoidably use a linear model to refactor, linear system theory is still irreplaceable, and needs to be further developed.

One of the open and hot problems in the optimal estimation and system identification is system analysis with unknown inputs in not only (either) dynamics model but also (or) measurement model. This phenomenon was partially noticed and reported many times since 1975 [4]. Mathematically, according to Lan et al. [5], the discrete-time linear stochastic system with unknown inputs is given in (1)

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{F}_k\mathbf{x}_k + \mathbf{\Gamma}_k\mathbf{q}_k + \mathbf{M}_k^0\mathbf{u}_k + \mathbf{M}_k\mathbf{a}_k \\ \mathbf{y}_{k+1} = \mathbf{H}_{k+1}\mathbf{x}_{k+1} + \mathbf{v}_{k+1} + \mathbf{N}_{k+1}^0\mathbf{u}_{k+1} + \mathbf{N}_{k+1}\mathbf{b}_{k+1}, \end{cases} \tag{1}$$

where $\mathbf{x}_k, \mathbf{x}_{k+1} \in \mathbf{R}^n$ and $\mathbf{y}_{k+1} \in \mathbf{R}^m$ are n-dimensional state vector and m-dimensional measure vector, respectively. The $n \times n$ system dynamics matrix $\mathbf{F}_k$, $n \times s$ noise-driving matrix $\mathbf{\Gamma}_k$, input-driving matrix $\mathbf{M}_k$ of dimension $n \times d$, matrix $\mathbf{N}_{k+1}$ of dimension $m \times r$ and $m \times n$ measure matrix $\mathbf{H}_{k+1}$ are all known. Besides, $e \times 1$ vector $\mathbf{u}_k$ is known input signal, and $\mathbf{M}_k^0$, $\mathbf{N}_{k+1}^0$ are input driven matrix with proper dimensions for $\mathbf{u}_k$. All matrices of $\mathbf{F}_k$, $\mathbf{\Gamma}_k$, $\mathbf{H}_{k+1}$, $\mathbf{M}_k^0$, $\mathbf{N}_{k+1}^0$, $\mathbf{M}_k$ and $\mathbf{N}_{k+1}$ are of full column rank. The s-dimensional process noise $\mathbf{q}_k$ and m-dimensional measure noise $\mathbf{v}_{k+1}$ are all zero-mean white Gaussian noises with known $s \times s$ covariance matrix $\mathbf{Q}_k > 0$ and $m \times m$ covariance matrix $\mathbf{R}_{k+1} > 0$, respectively. The initial state $\mathbf{x}_k|_{k=0}$ is Gaussian distributed with pre-given mean $\mathbf{x}_0$ and co-variance $\mathbf{\Sigma}_0$. Moreover, $\mathbf{q}$, $\mathbf{v}$ and $\mathbf{x}_0$ are assumed to be mutually independent since many industrial processes show this property. Finally, we note that d-dimensional vector $\mathbf{a}_k \in \mathbf{R}^d$ and r-dimensional vector $\mathbf{b}_{k+1} \in \mathbf{R}^r$ are unknown inputs. For simplicity and without loss of generality, we in this paper treat the known deterministic input signal $\mathbf{u}_k$ as zero through the time.

The unknown inputs, according to [5–11], could typically be process noises, modelling errors, sensors' fault, actuator faults, sharp manoeuvres in target tracking problems, and man-made jams as in electronic countermeasure. When these uncertainties appear in practical problems, the unawareness may cause disasters. Because the traditional Kalman filer is no longer powerful for those problems so that the estimate error would be out of control or even diverge [12–14]. The problem of jointly estimating the system states and identifying the system particulars is termed here as linear joint estimation and identification problem (LJEIP) and formatted by the Eq. (1). Obviously, the objective of solving the LJEIP is to estimate system states $\mathbf{x}_k$ and identify system particulars, $\mathbf{a}_k$ and $\mathbf{b}_k$, simultaneously.

* Corresponding author.
E-mail addresses: s.wang@u.nus.edu (S. Wang), iselc@nus.edu.sg (C. Li), isealim@nus.edu.sg (A. Lim).

https://doi.org/10.1016/j.sigpro.2019.03.004
0165-1684/

In engineering, literatures pay multiple attentions on various specific problems derived from the system (1). Because the unknown inputs $\mathbf{a}_k$ and $\mathbf{b}_{k+1}$ could be deterministic [15], stochastic [12] or mixed [10,16,17] signals. Different from this perspective, another novel angle is that the unknown inputs could exist merely in system model [4,12], measurement model [8,18] or both [5,10,17]. Obviously, different settings mean different oriented solutions because none of the existing methods is optimal for all problems, in consideration of solution accuracy, convergence speed, robustness and so on. Among them, worthy of mentioning is the following five classes of methodologies: (a) unknown input observer [19–22]; (b) adaptive and robust filter [12,13,23,24]; (c) multiple model [22,25–29]; (d) minimum upper bound filter [12,30]; and (e) linear minimum variance estimator [18,31–33]. They aim to settle different categories of problems with different formulations (or assumptions) or different optimality criterions. For example, the robust filter concerns more on the infinity norm, while linear minimum variance estimator cares the two norm (Euclid norm). For specific application scenarios, oriented solutions stand out.

That the uncertainties exist both in system model and measurement model is more general and getting increasing attention from scholars over the time [5,10,34]. In 2013, Lan et al. [5] firstly proposed a method based on expectation-maximization (EM) idea to tackle this problem and the proposed solution worked very well in practice. Unfortunately, the authors did not provide comprehensive theoretical investigation for this important problem. As far as we know, there neither exist rigorous theoretical studies for it in the existing literatures [35]. Due to the significance of this problem, we aim to make it up and provide the extension with theoretical results of optimal joint estimation and identification of linear stochastic system with unknown inputs. Our contributions are summarized as follows.

(1) We provide rigorous mathematical definitions and derivations in using an EM based solution to calculate $\mathbf{x}_k$, $\mathbf{a}_k$ and $\mathbf{b}_k$ in the LJEIP. We also identify and fix mathematical errors in [5];

(2) We consider a more general case that $\mathbf{\Gamma}_{i-1}\mathbf{Q}_{i-1}\mathbf{\Gamma}_{i-1}^T$ could be non-invertible, which was not discussed by Lan et al. in [5]. For detail, see Remark 5;

(3) We theoretically prove the existence and uniqueness of the solution (the estimates to $\mathbf{x}_k$, $\mathbf{a}_k$ and $\mathbf{b}_k$, that is, $\hat{\mathbf{x}}_k$, $\hat{\mathbf{a}}_k$ and $\hat{\mathbf{b}}_k$) to LJEIP (1);

(4) The effectiveness and convergence of the EM-based solution to the LJEIP is theoretically proved;

(5) We replace the optimal smoother from forward-backward algorithm with the RTS algorithm to improve the computational efficiency;

(6) Insightful explanations are given for deriving theories to the LJEIP.

For simplicity, we in this paper still only concern the case that $\mathbf{a}_k$ and $\mathbf{b}_{k+1}$ are unknown deterministic signals, not random variables. Thus the system (1) could be a Gaussian one. For more on this point, we give Assumption 1 as a premise of our mathematical derivation.

Assumption 1. It should be clearly noted that in this paper, specifically in (1), $\mathbf{a}_k$ and $\mathbf{b}_{k+1}$ are not random variables. They instead are deterministic signals although we do not know their real values. In this sense, the system (1) is still a Gaussian system, since the existing random variables $\mathbf{q}_k$ and $\mathbf{v}_{k+1}$ are normally distributed. Alternatively, we have the equivalent analysis on this point. We could treat the term $\mathbf{\Gamma}_k\mathbf{q}_k + \mathbf{M}_k\mathbf{a}_k$ of (1) in state model as a whole part, notated as a new Gaussian variable $\mathbf{Z}_k$. Then $\mathbf{Z}_k$ is a normal random variable with mean of $\mathbf{M}_k\mathbf{a}_k$, and the variance of $\mathbf{\Gamma}_k\mathbf{Q}_k\mathbf{\Gamma}_k^T$. The story keeps similar to the term $\mathbf{v}_{k+1} + \mathbf{N}_{k+1}\mathbf{b}_{k+1}$ of (1) in measurement model. It is with the mean of $\mathbf{N}_{k+1}\mathbf{b}_{k+1}$, and the variance of $\mathbf{R}_{k+1}$.

As a snapshot, we emphasize here that the main contribution of this paper, the solution to the LJEIP, is termed as the optimal linear joint estimation and identification theorem (LJEIE) to the linear Gaussian system with unknown inputs, which is detailed in Theorem 4.

The remaining paper is structured as follows. We illustrate the relationship between current study and Lan et al. [5] in Section 2. Following it, Section 3 introduces the Rauch-Tung-Striebel (TRS) [36] fixed-interval smoother and analyzes its properties. We conduct theoretical study to the LJEIP in Section 4 which includes proof of existence, uniqueness of the solution to the LJEIP, and the proof of effectiveness, convergence of the EM-based frame to find the solution. In order to demonstrate the effectiveness and efficiency of the proposed LJEIT, we design a simulation experiment of target tracking in Section 5 and analyze the corresponding simulation results. In the end, we conclude the whole work and discuss the possible future work in Section 6.

Before proceeding to the next section, we present several remarks.

Remark 1. Joint estimation and identification problem to linear Gaussian system is actually the inverse problem of a linear system. Because, obviously, a regular system maps the inputs into system states and outputs. Contrarily, the joint estimation and identification problem maps the outputs into system states and inputs.

Remark 2. A typical scenario that could be modelled as problem (1) would be given as: suppose the real system model is $\mathbf{F}$, and the modelling result is $\hat{\mathbf{F}}$. Thus there is a difference $\Delta\mathbf{F}$ between $\mathbf{F}$ and $\hat{\mathbf{F}}$, that is, $\Delta\mathbf{F} = \mathbf{F} - \hat{\mathbf{F}}$. In this case, the estimate to states would have a bias $\Delta\mathbf{F}_k\mathbf{x}_k$. However, we could use the model $\mathbf{M}_k\mathbf{a}_k$ to eliminate or at least weaken this bias, since $\mathbf{M}_k$ could be determined with the experience knowledge in engineering. At worst, $\mathbf{M}_k$ could always be treated as $n \times n$ identical matrix $\mathbf{I}$.

Remark 3. In engineering, requiring the matrices $\mathbf{F}$, $\mathbf{\Gamma}$, $\mathbf{H}$, $\mathbf{M}$ and $\mathbf{N}$ are of full column rank is reasonable and even necessary. Because it means all the information contained in states variables or inputs will be used.

Remark 4. In engineering, requiring the matrices $\mathbf{Q}$ and $\mathbf{R}$ are of full-rank (namely invertible) are reasonable and practical. For $\mathbf{R}$, it is without any doubt [36]. For $\mathbf{Q}$, if it is rank deficiency, then we could always adjust the form of $\mathbf{\Gamma}$ to let $\hat{\mathbf{Q}}$ be of full rank. For example, if $\mathbf{\Gamma} = [\mathbf{\Gamma_1}, \mathbf{\Gamma_2}, \mathbf{\Gamma_3}]$, and $\mathbf{Q} = \mathbf{diag}\{\mathbf{0}, \sigma_1^2, \sigma_2^2\}$, then we could let $\hat{\mathbf{\Gamma}} = [\mathbf{\Gamma_2}, \mathbf{\Gamma_3}]$, and $\hat{\mathbf{Q}} = \mathbf{diag}\{\sigma_1^2, \sigma_2^2\}$. After this transformation, the $\mathbf{Q}$ could always be of full rank and invertible.

2. Review of lan et al. (2013) [5]

Current work is based on Lan et al. [5] and a re-investigation of the studied problem. In this section, we comprehensively review the work of Lan et al. (2013). In [5], the authors first defined the complete-data log-likelihood function L_{k-l}^k based on observations $\{\mathbf{y}_{k-l}, \mathbf{y}_{k-l+1}, \ldots, \mathbf{y}_k\}$ and assumed unknown variables $\{\mathbf{a}_{k-l}, \mathbf{a}_{k-l+1}, \ldots, \mathbf{a}_k\} \cup \{\mathbf{b}_{k-l}, \mathbf{b}_{k-l+1}, \ldots, \mathbf{b}_k\}$, in which k is the current sampling time and $k-l$ means l steps before from current time k. Subsequently, employing EM-based algorithms, Lan et al. (2013) proposed the solution to the LJEIP.

Obviously, Lan et al. (2013) have studied important problems and made significant contributions to the community. The authors, however, only gave a practically workable solution and performed convicing simulation study by setting some variables to be special forms. They did not present strong theoretical investigation. For instance, they did not prove the existence and uniqueness of $\hat{\mathbf{x}}$, $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$. They neither present the proof of the effectiveness and convergence of the EM-based algorithm. Moreover, because of an error made in Eq. (25) in [5], the authors simply viewed $\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^T$ as invertible, which is not sufficient and this matrix could actually be non-invertible. We also point out that the optimal smoother used to solve a fixed interval smoothing problem could be actually replaced by the RTS filter and the computational efficiency can be improved. In the end, we note that there are errors in mathematical derivation in [5]. As a note, we have fixed them in Appendix A.1.

3. RTS Smoother

In this section, we introduce the RTS smoother which will be used to support our linear joint estimation and identification theories. The RTS smoother is an optimal smoother applied in a fixed interval. In current study, we only present the RTS algorithm and do not introduce it in detail. We invite interested readers to refer to Section 9.4 in [36] for more information on this algorithm.

Theorem 1. *[RTS smoother] For a linear Gaussian system defined as (1) in time interval* $[1, N]$, *if* $\{\mathbf{y}_k | k \in \{1, 2, \cdots, N\}\}$, $\{\mathbf{a}_k, \mathbf{b}_k | k \in \{1, 2, \cdots, N\}\}$, $\mathbf{x}_0$ *and* $\boldsymbol{\Sigma}_0$ *are uniquely pre-given, then, for any* $k \in \{1, 2, \cdots, N\}$, *the estimate* $\hat{\mathbf{x}}_k$ *of the system state* $\mathbf{x}_k$ *uniquely exist, in the sense of linear unbiased minimum variance. Besides, the value of* $\hat{\mathbf{x}}_k$ *is given by Algorithm 1.*

Algorithm 1 RTS Smoother. See Ref. [36] is in Algorithm.

Input: $\{\mathbf{y}_k | k \in \{1, 2, \ldots, N\}\}$, $\{\mathbf{a}_k, \mathbf{b}_k | k \in \{1, 2, \ldots, N\}\}$, $\mathbf{x}_0$ and $\boldsymbol{\Sigma}_0$

1: **Initialize the forward filter:** compute $\hat{\mathbf{x}}_{f0}$ and $\mathbf{P}_{f0}^{+}$ given by Eqs. (9.135) in [36]

2: **Execute the forward filter for** $k = 1, 2, 3, \ldots, N$**:** compute $K_{fk}$and $\mathbf{P}_{\mathbf{fk}}^{+}$ given in Eqs. (9.136) in [36] and

$$\mathbf{P}_{fk}^{-} = \mathbf{F}_{k-1}\mathbf{P}_{f,k-1}^{+}\mathbf{F}_{k-1}^T + \boldsymbol{\Gamma}_{k-1}\mathbf{Q}_{k-1}\boldsymbol{\Gamma}_{k-1}^T \tag{2}$$

$$\hat{\mathbf{x}}_{fk}^{-} = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{f,k-1}^{+} + \mathbf{M}_{k-1}\mathbf{a}_{k-1} \tag{3}$$

$$\hat{\mathbf{x}}_{fk}^{+} = \hat{\mathbf{x}}_{fk}^{-} + \mathbf{K}_{fk}\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{fk}^{-} - \mathbf{N}_k\mathbf{b}_k\right) \tag{4}$$

3: **Initialize the backward filter:** compute $\hat{\mathbf{x}}_N$ and $\mathbf{P}_N$ given by Eqs. (9.137) in [36]

4: **Execute the backward filter for** $k = N-1, \ldots, 0$**:** compute all quantities given by Eqs. (9.138) in [36]

Output: $\{\hat{\mathbf{x}}_k | k \in \{1, 2, \cdots, N\}\}$

Proof. According to the principle of the RTS smoother, it is obvious. For detail on the RTS smoother, see [36]. □

For the sense of linear unbiased minimum valiance, we detail in Remark 7.

Theorem 2. *If* $\{\mathbf{y}_k | k \in \{1, 2, \ldots, N\}\}$ *are uniquely pre-given, letting* $\boldsymbol{\rho}_k = [\mathbf{a}_k, \mathbf{b}_k]^T$, *then* $\hat{\mathbf{x}}_k$ *is continuous in the definition domain of* $\boldsymbol{\rho}_k$.

Proof. Theorem 1 and [36] uphold that the mapping from $\mathbf{y}_k$, $\boldsymbol{\rho}_k$ to $\hat{\mathbf{x}}_k$ is no wonder a function (one-to-one mapping). Thus discussing the continuity of $\hat{\mathbf{x}}_k$ over $\boldsymbol{\rho}_k$ is meaningful. Let $D_d[\boldsymbol{\rho}_k]$ be the definition domain of $\boldsymbol{\rho}_k$. (1) If D_d is a continuous set, according to Theorem 1 and Algorithm 1, this theorem holds; (2) If D_d is a discrete set, we just need to prove that $\hat{\mathbf{x}}_k$ is continuous at every separate points in the definition domain of $\boldsymbol{\rho}_k$, because it is a discrete set [37]. By the fact that for every $\boldsymbol{\rho}_k \in D_d$ and $\forall \varepsilon > 0$, it is true that $\exists \delta > 0$ and $\exists \boldsymbol{\rho} \in D_d$, if $\|\rho_k - \rho\| < \delta$, then $\left\|\hat{\mathbf{x}}_k(\rho_k) - \hat{\mathbf{x}}_k(\rho)\right\| < \varepsilon$. The fact stands because $\boldsymbol{\rho}$ could always be $\boldsymbol{\rho}_k$. Thus according to the definition of continuity in functional analysis (or in real analysis) [37], the theorem stands. □

4. Solution to linear joint estimation and identification problem

For a physical linear system, its states $\mathbf{x}_k$ and unknown inputs $\mathbf{a}_k$, $\mathbf{b}_k$ definitely exist (because $\mathbf{a}_k$ and $\mathbf{b}_k$ could be always zero if need). And our aim is to estimate those unknown variables. From the viewpoint of statistics, to construct a statistics as an estimate of a variable, we must depend on enough measures with respect to it. Therefore, we consider the measure set from time step $k-l$ to k, where k indicates the current sampling time and $k-l$ means l steps before from current time k. The LJEIP (1) seems plausible to be solved based on EM-frame, in consideration of that EM-algorithm could alternately and simultaneously optimize the estimates of $\mathbf{x}_k$, $\mathbf{a}_k$, and $\mathbf{b}_k$, and give the definite numerical values of them. The mechanism of EM algorithm actually is to get the maximum likelihood estimate of random variables by alternating optimization in different variables. Thus we should first construct the likelihood function for the LJEIP defined in (1). As a summary, we give the motivation of using the EM-based frame as our technique in Motivation 1.

Motivation 1. In LJEIP (1), we actually have two different kinds of signals (the system states and the unknown inputs) to be estimated. According to the RTS fixed interval smoother, however, if we have the estimates of unknown inputs, we then could have an unique estimate to the system state. Thus only the unknown inputs are the independent underlying variables to the stochastic system (1). Thus if we treat the unknown inputs as our underlying parameters, the system outputs $\mathbf{y}_i$ as our observations and the system dynamics (1) as our probability distribution, we then could use the EM algorithm to maximize the likelihood function (from the unknown inputs to the outputs) while optimizing (estimating) the underlying parameters (unknown inputs).

Definition 1. Let $\mathbf{X}_{k-l}^k = \{\mathbf{x}_{k-l}, \ldots, \mathbf{x}_k\}$, $\mathbf{Y}_{k-l}^k = \{\mathbf{y}_{k-l}, \ldots, \mathbf{y}_k\}$, $\mathbf{A}_{k-l}^k = \{\mathbf{a}_{k-l}, \ldots, \mathbf{a}_k\}$, $\mathbf{B}_{k-l}^k = \{\mathbf{b}_{k-l}, \ldots, \mathbf{b}_k\}$, and $\rho_{k-l}^k = [\mathbf{A}_{k-l}^k, \mathbf{B}_{k-l}^k]^T$, respectively. Plus, using $\hat{\rho}_{k-l}^k$ as an estimate to ρ_{k-l}^k and $\hat{\mathbf{x}}_{i|k-l:k} = E[\mathbf{x}_i | \mathbf{Y}_{k-l}^k, \rho_{k-l}^k, \mathbf{Y}_1^{k-l-1}, \rho_1^{k-l-1}, \mathbf{x}_0, \boldsymbol{\Sigma}_0]$ as the conditional expectation of $\mathbf{x}_i$ (Obviously, $\hat{\mathbf{x}}_{i|k-l:k}$ is smoothed result of $\mathbf{x}_i$ in the interval $[k-l, k]$). Together with $\hat{\mathbf{x}}_{i|k-l:k}$, let $\mathbf{P}_{i,j|k-l:k} = \mathrm{cov}(\hat{\mathbf{x}}_{i|k-l:k}, \hat{\mathbf{x}}_{j|k-l:k})$ define its co-variance matrix with $\hat{\mathbf{x}}_{j|k-l:k}$. Also, let $\tilde{\mathbf{x}}_k$ be the estimate error of $\mathbf{x}_k$, that is $\mathbf{x}_k = \hat{\mathbf{x}}_k + \tilde{\mathbf{x}}_k$.

To apply the EM-frame, we should firstly define the likelihood function and its conditional expectation [38–40]. Obviously, in the LJEIP, observable data set is $\mathbf{Y}_{k-l}^k$, which is also referred to as "incomplete data"; likewise, unobservable data set is $\mathbf{X}_{k-l}^k$, also known as "complete data" [5,39,40].

Definition 2. Let J_{k-l}^k be the incomplete data log-likelihood function in discrete time interval $[k-l, k]$, that is $J_{k-l}^k = \log p[\mathbf{Y}_{k-l}^k | (\rho_{k-l}^k, \mathbf{Y}_1^{k-l-1}, \hat{\rho}_1^{k-l-1}, \mathbf{x}_0, \boldsymbol{\Sigma}_0)]$, and T_{k-l}^k be conditional expectation of J_{k-l}^k, that is $T_{k-l}^k = E_{\hat{\mathbf{X}}_{k-l}^k}[J_{k-l}^k(\rho_{k-l}^k) | \hat{\rho}_{k-l}^k, \mathbf{Y}_{k-l}^k]$; Let $\hat{\mathbf{x}}_{k-l-1} = E[\mathbf{x}_{k-l-1} | (\mathbf{Y}_1^{k-l-1}, \hat{\rho}_1^{k-l-1}, \mathbf{x}_0, \boldsymbol{\Sigma}_0)]$ be the conditional estimate of $\mathbf{x}_{k-l-1}$ and $\mathbf{P}_{k-l-1} = \mathrm{cov}(\mathbf{x}_{k-l-1}, \mathbf{x}_{k-l-1})$ is its co-variance matrix. The operator $\mathbf{D}$ is defined as $\mathbf{D}(\mathbf{x}, \mathbf{P}) = \mathbf{x}^T\mathbf{P}^{-1}\mathbf{x}$, and $\mathbf{C}$ is as $\mathbf{C}(\mathbf{x}) = \mathbf{x}\mathbf{x}^T$. $E_{\mathbf{x}}(\mathbf{y})$ means conditional expectation of $\mathbf{y}$ in presence of $\mathbf{x}$; $Tr(\mathbf{A})$ means calculating the trace of matrix $\mathbf{A}$; and $\|\mathbf{x}\|$ means any practically proper types of norm of $\mathbf{x}$, typically the 2-norm $\|\mathbf{x}\|_2$.

With Definition 1 and Definition 2, J_{k-l}^k could be further displayed in Eq. (5).

$$J_{k-l}^{k} = \log\left[\frac{p\left[\left(\mathbf{X}_{k-l}^{k}, \mathbf{Y}_{k-l}^{k}\right)\big|\left(\rho_{k-l}^{k}, \mathbf{Y}_{1}^{k-l-1}, \hat{\rho}_{1}^{k-l-1}, \mathbf{x}_{0}, \boldsymbol{\Sigma}_{0}\right)\right]}{p\left[\mathbf{X}_{k-l}^{k}\big|\mathbf{Y}_{k-l}^{k}, \left(\rho_{k-l}^{k}, \mathbf{Y}_{1}^{k-l-1}, \hat{\rho}_{1}^{k-l-1}, \mathbf{x}_{0}, \boldsymbol{\Sigma}_{0}\right)\right]}\right]$$
$$= \log\left\{p\left[\left(\mathbf{X}_{k-l}^{k}, \mathbf{Y}_{k-l}^{k}\right)\big|\left(\rho_{k-l}^{k}, \mathbf{Y}_{1}^{k-l-1}, \hat{\rho}_{1}^{k-l-1}, \mathbf{x}_{0}, \boldsymbol{\Sigma}_{0}\right)\right]\right\}$$
$$- \log\left\{p\left[\mathbf{X}_{k-l}^{k}\big|\mathbf{Y}_{k-l}^{k}, \left(\rho_{k-l}^{k}, \mathbf{Y}_{1}^{k-l-1}, \hat{\rho}_{1}^{k-l-1}, \mathbf{x}_{0}, \boldsymbol{\Sigma}_{0}\right)\right]\right\}$$
$$=: L_{k-l}^{k} - O_{k-l}^{k} \tag{5}$$

where

$$L_{k-l}^{k} = \log p\left[\left(\mathbf{X}_{k-l}^{k}, \mathbf{Y}_{k-l}^{k}\right)\big|\left(\rho_{k-l}^{k}, \mathbf{Y}_{1}^{k-l-1}, \hat{\rho}_{1}^{k-l-1}, \mathbf{x}_{0}, \boldsymbol{\Sigma}_{0}\right)\right]$$
$$= \log p\left[\mathbf{x}_{k-l-1}\big|\left(\mathbf{Y}_{1}^{k-l-1}, \hat{\rho}_{1}^{k-l-1}, \mathbf{x}_{0}, \boldsymbol{\Sigma}_{0}\right)\right]$$
$$+ \sum_{i=k-l}^{k} \log p\left[\mathbf{x}_{i}|\mathbf{x}_{i-1}, \left(\rho_{k-l}^{k}, \hat{\mathbf{x}}_{k-l-1}\right)\right]$$
$$+ \sum_{i=k-l}^{k} \log p\left[\mathbf{y}_{i}|\mathbf{x}_{i}, \left(\rho_{k-l}^{k}\right)\right]. \tag{6}$$

As for O_{k-l}^{k}, its specific pattern should be determined by the smoother we used, like forward-backward smoother or RTS [36]. However, it has a common type showed in (7)

$$O_{k-l}^{k} = \log p\left[\mathbf{X}_{k-l}^{k}\big|\mathbf{Y}_{k-l}^{k}, \left(\rho_{k-l}^{k}, \mathbf{Y}_{1}^{k-l-1}, \hat{\rho}_{1}^{k-l-1}, \mathbf{x}_{0}, \boldsymbol{\Sigma}_{0}\right)\right]$$
$$= \log p\left[\mathbf{x}_{k-l-1}\big|\left(\mathbf{Y}_{1}^{k-l-1}, \hat{\rho}_{1}^{k-l-1}, \mathbf{x}_{0}, \boldsymbol{\Sigma}_{0}\right)\right]$$
$$+ \sum_{i=k-l}^{k} \log p\left[\mathbf{x}_{i}|\mathbf{x}_{i-1}, \left(\mathbf{Y}_{k-l}^{k}, \rho_{k-l}^{k}, \hat{\mathbf{x}}_{k-l-1}\right)\right]. \tag{7}$$

Besides, the conditional expectation T_{k-l}^{k} of J_{k-l}^{k} is

$$T_{k-l}^{k}\left(\rho_{k-l}^{k}\big|\hat{\rho}_{k-l}^{k}, \mathbf{Y}_{k-l}^{k}\right) = E_{\hat{\mathbf{X}}_{k-l}^{k}}\left[J_{k-l}^{k}\left(\rho_{k-l}^{k}\right)\big|\left(\hat{\rho}_{k-l}^{k}, \mathbf{Y}_{k-l}^{k}\right)\right]$$
$$= E_{\hat{\mathbf{X}}_{k-l}^{k}}\left[L_{k-l}^{k}\left(\rho_{k-l}^{k}\right)\big|\left(\hat{\rho}_{k-l}^{k}, \mathbf{Y}_{k-l}^{k}\right)\right]$$
$$- E_{\hat{\mathbf{X}}_{k-l}^{k}}\left[O_{k-l}^{k}\left(\rho_{k-l}^{k}\right)\big|\left(\hat{\rho}_{k-l}^{k}, \mathbf{Y}_{k-l}^{k}\right)\right]$$
$$=: G_{k-l}^{k}\left(\rho_{k-l}^{k}\big|\hat{\rho}_{k-l}^{k}, \mathbf{Y}_{k-l}^{k}\right)$$
$$- H_{k-l}^{k}\left(\rho_{k-l}^{k}\big|\hat{\rho}_{k-l}^{k}, \mathbf{Y}_{k-l}^{k}\right). \tag{8}$$

Remark 5. For some Linear Gaussian system in engineering, $\boldsymbol{\Gamma}$ should be of full rank and thus $\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^{\mathbf{T}}$ should be invertible. However, for types of special ones, $\boldsymbol{\Gamma}$ may be rank deficiency, which leads to $\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^{\mathbf{T}}$ is non-invertible. For example, in target tracking community [8,41], the noise driven matrix $\boldsymbol{\Gamma}$ is like

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} T_s^2/2 \\ T_s \end{bmatrix},$$

where $\otimes$ refers to the Kronecher product. The $\boldsymbol{\Gamma}$ now is with the dimension of 4 by 2, and the covariance matrix $\mathbf{Q}$ of $\mathbf{q}$ is 2 by 2. Obviously, in this case the corresponding matrix $\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^{\mathbf{T}}$ is rank deficiency (the rank is 2 rather than 4). For completeness, we must discuss the case that $\boldsymbol{\Gamma}$ is rank deficiency, which is placed in Appendix A.3. In following contexts, for briefness, we mainly display the case of that $\boldsymbol{\Gamma}$ is of full rank.

Remark 6. Due to the definition of iteration process of EM algorithm is only based on L_{k-l}^{k} [38–40], having nothing to do with O_{k-l}^{k}, Lan et al. [5] therefore exclusively paid their attention on L_{k-l}^{k} rather than J_{k-l}^{k}, O_{k-l}^{k} and so on. This should be an ambiguity and easy to misunderstand the readers. Besides, in order to prove the convergence and effectiveness, it is necessary to take all of those items into consideration.

After clarifying the necessary definitions and notations, we then need to discuss the generation of the EM solution to our LJEIP. We in Motivation 2 give the general ideas of how to derive the specific solution, and how to prove its convergence and effectiveness.

Motivation 2. In order to obtain the corresponding EM solution to the LJEIP, we then need to specify the likelihood function (LLF), derive its conditional expectation and construct the EM sequence (including Expectation step and Maximum Step) [5,39]. As for the convergence and effectiveness proof, the Theorem 2 and Corollary 1 in [39] works as long as we can show the continuity of the conditional expectation G_{k-l}^{k} of the LLF of the complete-data set, and the unimodality of the conditional expectation T_{k-l}^{k} of the LLF of the incomplete-data set. Thus in Lemma 1 we figure out the conditional expectations of the both complete-data LLF and incomplete-data LLF, and in Lemma 2, 3, and 4 we show the continuity and unimodality aforementioned. Next, in Lemma 5 and 6, we display why the EM sequence defined in Definition 3 converges to a global optimal solution for the interested problem. In the end, in Theorem 3 and 4 we make clear why the proposed EM solution works for our LJEIP.

In view of the fact given by Eqs. (9)∼(12),

$$p\left[\mathbf{x}_{k-l-1}\big|\left(\mathbf{Y}_{1}^{k-l-1}, \hat{\rho}_{1}^{k-l-1}, \mathbf{x}_{0}, \boldsymbol{\Sigma}_{0}\right)\right] = \mathbf{N}\left(\hat{\mathbf{x}}_{k-l-1}, \mathbf{P}_{k-l-1}\right) \tag{9}$$

$$p\left[\mathbf{x}_{i}|\mathbf{x}_{i-1}, \left(\rho_{k-l}^{k}, \hat{\mathbf{x}}_{k-l-1}\right)\right]$$
$$= \mathbf{N}\left(\mathbf{F}_{i-1}\mathbf{x}_{i-1} + \mathbf{M}_{i-1}\mathbf{a}_{i-1}, \boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right) \tag{10}$$

$$p\left[\mathbf{y}_{i}|\mathbf{x}_{i}, \left(\rho_{k-l}^{k}\right)\right] = \mathbf{N}(\mathbf{H}_{i}\mathbf{x}_{i} + \mathbf{N}_{i}\mathbf{b}_{i}, \mathbf{R}_{i}) \tag{11}$$

$$p\left[\mathbf{x}_{i}\big|\mathbf{x}_{i-1}, \left(\mathbf{Y}_{k-l}^{k}, \rho_{k-l}^{k}, \hat{\mathbf{x}}_{k-l-1}\right)\right] = \mathbf{N}\left(\hat{\mathbf{x}}_{i|k-l:k}, \mathbf{P}_{i,i|k-l:k}\right), \tag{12}$$

where $\mathbf{N}(\mu, \mathcal{D})$ means Gaussian distribution with mean μ and variance $\mathcal{D}$. For the case that $\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}$ is non-invertible, see Appendix A.3.

Eq. (6) could be further given as

$$L_{k-l}^{k} = L_{0,k-l}^{k} + L_{1,k-l}^{k} + L_{2,k-l}^{k} + L_{3,k-l}^{k}, \tag{13}$$

where

$$L_{0,k-l}^{k} = -\frac{2n + m + l(m+n)}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{P}_{k-l-1}|$$
$$- \frac{1}{2}\sum_{i=k-l}^{k}\left(\log\left|\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right| + \log|\mathbf{R}_{i}|\right) \tag{14}$$

$$L_{1,k-l}^{k} = -\frac{1}{2}\mathbf{D}\left(\mathbf{x}_{k-l-1} - \hat{\mathbf{x}}_{k-l-1}, \mathbf{P}_{k-l-1}\right) \tag{15}$$

$$L_{2,k-l}^{k} = -\frac{1}{2}\sum_{i=k-l}^{k}\mathbf{D}\left(\mathbf{x}_{i} - \mathbf{F}_{i-1}\mathbf{x}_{i-1} - \mathbf{M}_{i-1}\mathbf{a}_{i-1}, \boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right) \tag{16}$$

$$L_{3,k-l}^{k} = -\frac{1}{2}\sum_{i=k-l}^{k}\mathbf{D}(\mathbf{y}_{i} - \mathbf{H}_{i}\mathbf{x}_{i} - \mathbf{N}_{i}\mathbf{b}_{i}, \mathbf{R}_{i}), \tag{17}$$

and Eq. (7) could be rewritten as

$$O_{k-l}^{k} = O_{0,k-l}^{k} + O_{1,k-l}^{k} + O_{2,k-l}^{k}, \tag{18}$$

where

$$O_{0,k-l}^{k} = -\frac{n + n(l+1)}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{P}_{k-l-1}|$$
$$- \frac{1}{2}\sum_{i=k-l}^{k}\log\left|\mathbf{P}_{i,i|k-l:k}\right| \tag{19}$$

$$O_{1,k-l}^{k} = -\frac{1}{2}\mathbf{D}\left(\mathbf{x}_{k-l-1} - \hat{\mathbf{x}}_{k-l-1}, \mathbf{P}_{k-l-1}\right) \tag{20}$$

$$O^k_{2,k-l} = -\frac{1}{2}\sum_{i=k-l}^{k} \mathbf{D}\left(\mathbf{x}_i - \hat{\mathbf{x}}_{i,i|k-l:k}, \mathbf{P}_{i,i|k-l:k}\right), \tag{21}$$

in which the operator $\mathbf{D}$ is defined as $\mathbf{D}(\mathbf{x},\mathbf{P}) = \mathbf{x}^T\mathbf{P}^{-1}\mathbf{x}$ and $|\mathbf{A}|$ means determinant of matrix $\mathbf{A}$. For the case that $\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}$ is non-invertible, see Appendix A.3.

Lemma 1. *The conditional expectation G^k_{k-l} of L^k_{k-l} regarding LJEIP (1) is given as*

$$\begin{aligned} G^k_{k-l} &= E_{\hat{\mathbf{X}}^k_{k-l}}\left[L^k_{k-l}\left(\rho^k_{k-l}\right)\middle|\left(\mathbf{Y}^k_{k-l}, \hat{\rho}^k_{k-l}, \hat{\mathbf{x}}_{k-l-1}\right)\right] \\ &= G^k_{0,k-l} + G^k_{1,k-l} + G^k_{2,k-l} + G^k_{3,k-l}, \end{aligned} \tag{22}$$

where

$$G^k_{0,k-l} = L^k_{0,k-l} \tag{23}$$

$$G^k_{1,k-l} = -\frac{n}{2}, \tag{24}$$

$$\begin{aligned} G^k_{2,k-l} = -\frac{1}{2}Tr\Bigg\{ &\sum_{i=k-l}^{k}\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}\right]^{-1} \\ &\cdot\Big[\mathbf{C}(\hat{\mathbf{x}}_{i|k-l:k} - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1|k-l:k} - \mathbf{M}_{i-1}\mathbf{a}_{i-1}) + \mathbf{P}_{i,i|k-l:k} \\ &-\mathbf{P}_{i,i-1|k-l:k}\mathbf{F}^T_{i-1} - \mathbf{F}_{i-1}\mathbf{P}_{i-1,i|k-l:k} + \mathbf{F}_{i-1}\mathbf{P}_{i-1,i-1|k-l:k}\mathbf{F}^T_{i-1}\Big]\Bigg\} \end{aligned} \tag{25}$$

$$G^k_{3,k-l} = -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\mathbf{R}_i^{-1}\cdot\left[\mathbf{C}\left(\mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i|k-l:k} - \mathbf{N}_i\mathbf{b}_i\right) - \mathbf{H}_i\mathbf{P}_{i,i|k-l:k}\mathbf{H}_i^T\right]\right\} \tag{26}$$

and the conditional expectation H^k_{k-l} of O^k_{k-l} is as

$$\begin{aligned} H^k_{k-l} &= E_{\hat{\mathbf{X}}^k_{k-l}}\left[O^k_{k-l}\left(\rho^k_{k-l}\right)\middle|\left(\mathbf{Y}^k_{k-l}, \hat{\rho}^k_{k-l}, \hat{\mathbf{x}}_{k-l-1}\right)\right] \\ &= H^k_{0,k-l} + H^k_{1,k-l} + H^k_{2,k-l}, \end{aligned} \tag{27}$$

where

$$H^k_{0,k-l} = O^k_{0,k-l} \tag{28}$$

$$H^k_{1,k-l} = -\frac{n}{2} \tag{29}$$

$$H^k_{2,k-l} = -\frac{n(l+1)}{2}, \tag{30}$$

and the operator $\mathbf{C}$ is defined as $\mathbf{C}(\mathbf{x}) = \mathbf{x}\mathbf{x}^T$. For the case that $\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}$ is non-invertible, see Appendix A.3.

Proof. See Appendix A.2. □

To apply EM frame, we now discuss the properties of G^k_{k-l}. Below are conclusions.

Lemma 2. *G^k_{k-l} is continuous over the definition domains of both $\hat{\rho}^k_{k-l}$ and ρ^k_{k-l}.*

Proof. No wonder, G^k_{k-l} is continuous to $\hat{\mathbf{x}}_{i|k-l:k}$ and ρ^k_{k-l}, because the definition of G^k_{k-l} is basic. Besides, according to Theorem 2, $\hat{\mathbf{x}}_{i|k-l:k}$ is continuous to $\hat{\rho}^k_{k-l}$, thus in consideration of theory of functional analysis (or real analysis) [37], G^k_{k-l} is continuous to $\hat{\rho}^k_{k-l}$. Therefore, the lemma stands. □

Lemma 3. *G^k_{k-l} is a concave function on the definition domain of $\rho^k_{k-l} = [\mathbf{a}^k_{k-l}, \mathbf{b}^k_{k-l}]^T$, and the peak is reached at its unique stationary point (meaning local maximum here), if $\mathbf{M}_i$, $\mathbf{N}_i$ are all of full column rank.*

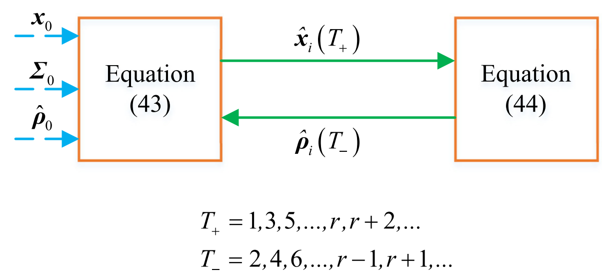


Fig. 1. Iteration process defined in Theorem 4.

Proof. The Hessian matrix of G^k_{k-l} on ρ^k_{k-l} is

$$\frac{\partial^2 G^k_{k-l}}{\partial\left(\rho^k_{k-l}\right)\partial\left(\rho^k_{k-l}\right)^T} = \mathbf{H}_{hes} = \begin{bmatrix}\mathbf{H}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{H}_2\end{bmatrix}, \tag{31}$$

where

$$\mathbf{H}_1 = diag\left\{-\mathbf{M}^T_{i-1}\cdot\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}\right]^{-1}\cdot\mathbf{M}_{i-1}\right\}_{i=k-l,\ldots,k} \tag{32}$$

$$\mathbf{H}_2 = diag\left\{-\mathbf{N}^T_i\cdot\mathbf{R}^{-1}_i\cdot\mathbf{N}_i\right\}_{i=k-l,\ldots,k}. \tag{33}$$

It is obvious that both $\mathbf{H}_1$ and $\mathbf{H}_2$ are negative definite. Thus $\mathbf{H}_{hes}$ is negative definite. Besides

$$\frac{\partial G^k_{k-l}}{\partial\rho^k_{k-l}} = \begin{bmatrix}\mathbf{D}_1\\ \mathbf{D}_2\end{bmatrix} \tag{34}$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are given by Eqs. (35) and (36).

$$\begin{aligned}\mathbf{D}_1 = col\Big\{&\mathbf{M}^T_{i-1}\cdot\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}\right]^{-1} \\ &\cdot\left(\hat{\mathbf{x}}_{i|k-l:k} - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1|k-l:k} - \mathbf{M}_{i-1}\mathbf{a}_{i-1}\right)\Big\}_{i=k-l,\ldots,k}\end{aligned} \tag{35}$$

$$\mathbf{D}_2 = col\left\{\mathbf{N}^T_i\cdot\mathbf{R}^{-1}_i\cdot\left(\mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i|k-l:k} - \mathbf{N}_i\mathbf{b}_i\right)\right\}_{i=k-l,\ldots,k} \tag{36}$$

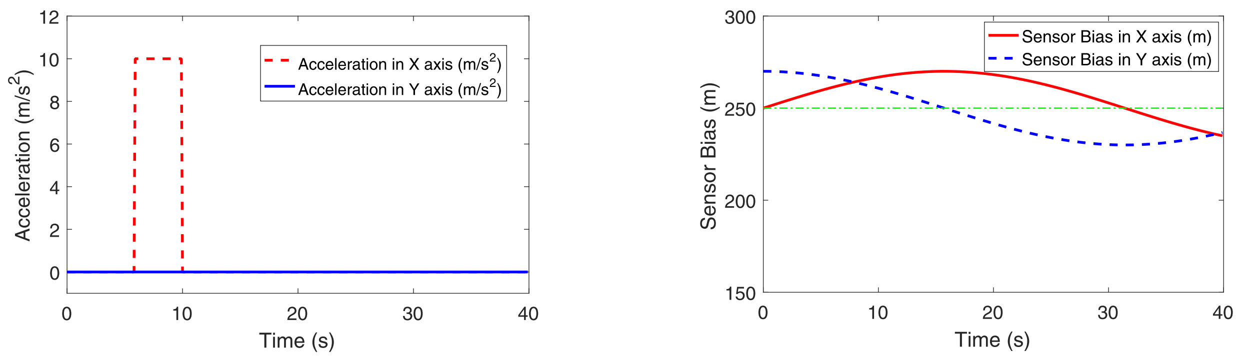
In consideration of that $\mathbf{M}_i$, $\mathbf{N}_i$ are all of full column rank, that is, the matrices defined by $\left\{\mathbf{N}^T_i\cdot\mathbf{R}^{-1}_i\cdot\mathbf{N}_i\right\}$ and $\left\{\mathbf{M}^T_{i-1}\cdot\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}\right]^{-1}\cdot\mathbf{M}_{i-1}\right\}$ are all positive definite and thus invertible, the lemma stands. For the case that $\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}$ is non-invertible, see Appendix A.3. □

Lemma 4. *T^k_{k-l} is a concave function and the unique point reaching its peak T^* is same with the point reaching the peak of G^k_{k-l}, if $\mathbf{M}_i$, $\mathbf{N}_i$ are all of full column rank. As for the point $\boldsymbol{\rho}^*$ reaching the peak mentioned above, it is given by (37). For the case that $\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}$ is non-invertible, see Appendix A.3.*

$$\begin{cases}\mathbf{a}^*_{i-1} = \mathcal{A}^{-1}\cdot\mathbf{M}^T_{i-1}\cdot\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}\right]^{-1}\cdot\left(\hat{\mathbf{x}}_{i|k-l:k} - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1|k-l:k}\right)\\ \mathbf{b}^*_i = \mathcal{B}^{-1}\cdot\mathbf{N}^T_i\cdot\mathbf{R}^{-1}_i\cdot\left(\mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i|k-l:k}\right),\end{cases} \tag{37}$$

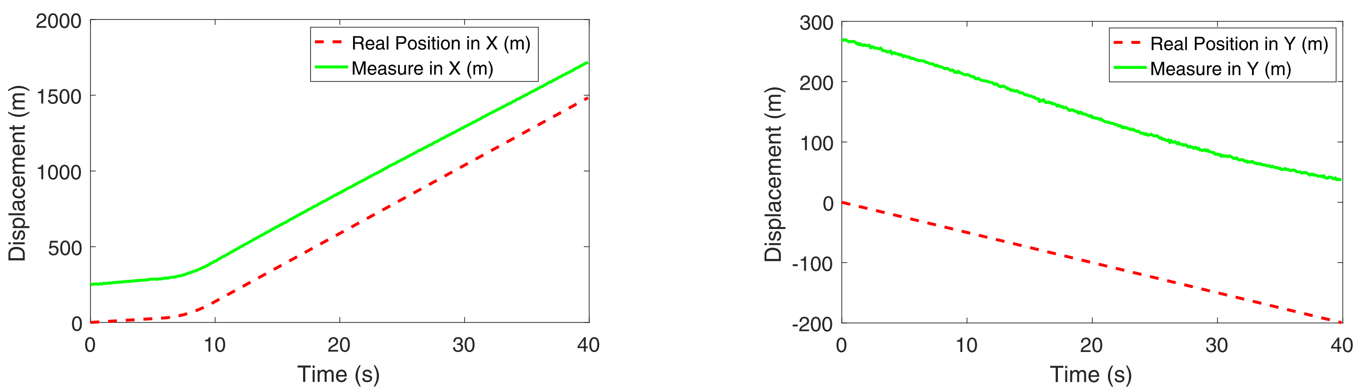
where $\mathcal{A} = \left\{\mathbf{M}^T_{i-1}\cdot\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}^T_{i-1}\right]^{-1}\cdot\mathbf{M}_{i-1}\right\}$, $\mathcal{B} = \left\{\mathbf{N}^T_i\cdot\mathbf{R}^{-1}_i\cdot\mathbf{N}_i\right\}$, $\rho_i{}^* = \left[\mathbf{a}^*_{i-1}, \mathbf{b}^*_i\right]^T$ and $k-l \le i \le k$.

Proof. In consideration of that if $\hat{\rho}^k_{k-l}$ and $\mathbf{Y}^k_{k-l}$ are pre-given, then $\mathbf{X}^k_{k-l}$ would be determined by RTS Smoother. Thus H^k_{k-l} has nothing to do with ρ^k_{k-l}. Specifically, see Eqs. (28)~(30). That means



(a) Unknown input in system model (target manoeuvre) (b) Unknown input in measurement model (radar bias)

Fig. 2. The unknown inputs in system model and measurement model.



(a) Real and measured position in x axis (b) Real and measured position in y axis

Fig. 3. Real and measured position in x axis and y axis.

$$\frac{\partial^2 G_{k-l}^k}{\partial\left(\rho_{k-l}^k\right)\partial\left(\rho_{k-l}^k\right)^T} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix} = \frac{\partial^2 T_{k-l}^k}{\partial\left(\rho_{k-l}^k\right)\partial\left(\rho_{k-l}^k\right)^T} \tag{38}$$

$$\frac{\partial G_{k-l}^k}{\partial \rho_{k-l}^k} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix} = \frac{\partial T_{k-l}^k}{\partial \rho_{k-l}^k}, \tag{39}$$

and $\left\{\mathbf{M}_{i-1}^T \cdot \left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^T\right]^{-1} \cdot \mathbf{M}_{i-1}\right\}$, $\left\{\mathbf{N}_i^T \cdot \mathbf{R}_i^{-1} \cdot \mathbf{N}_i\right\}$ are all invertible, if further we let $\mathbf{D}_1 = \mathbf{0}$ and $\mathbf{D}_2 = \mathbf{0}$, the lemma stands. □

Definition 3. Let $r = 0, 1, 2, 3, \ldots$ be the index of the array $\{\hat{\rho}_r\} := \{\hat{\rho}_{k-l}^k(r)\}$. The element $\hat{\rho}_{k-l}^k(r)|_{r>0}$ of the array is recursively defined as

$$\begin{aligned} \hat{\rho}_{k-l}^k(r+1) &= \arg\max_{\rho_{k-l}^k} G_{k-l}^k(\rho_{k-l}^k | \mathbf{Y}_{k-l}^k, \hat{\rho}_{k-l}^k(r)) \\ &= \arg\max_{\rho_{k-l}^k} E_{\hat{\mathbf{X}}_{k-l}^k}[L_{k-l}^k(\rho_{k-l}^k) | (\mathbf{Y}_{k-l}^k, \hat{\rho}_{k-l}^k(r))], \end{aligned} \tag{40}$$

with the initial value $\hat{\rho}_{k-l}^k(0) = \hat{\rho}_0$ $(\left\|\hat{\rho}_0\right\| < \infty)$.

The Eq. (40) in Definition 3 actually gives Maximum step in EM frame, meaning obtaining the maxima of the conditional expectation of the complete-data set likelihood function G_{k-l}^k. Correspondingly, Lemma 1, which calculates G_{k-l}^k, gives the Expectation step in EM frame. For the detailed concepts of Expectation step and Maximum step of EM frame, see [5,38–40].

Lemma 5. *For any instance of the array $\{\hat{\rho}_r\} := \{\hat{\rho}_{k-l}^k(r)\}$ defined in Definition 3, the inequalities*

$$T_{k-l}^k\left(\hat{\rho}_{k-l}^k(r+1)\right) \geq T_{k-l}^k\left(\hat{\rho}_{k-l}^k(r)\right) \tag{41}$$

and

$$H_{k-l}^k\left(\hat{\rho}_{k-l}^k(r+1)\right) \leq H_{k-l}^k\left(\hat{\rho}_{k-l}^k(r)\right) \tag{42}$$

always hold.

Proof. See [39,40]. □

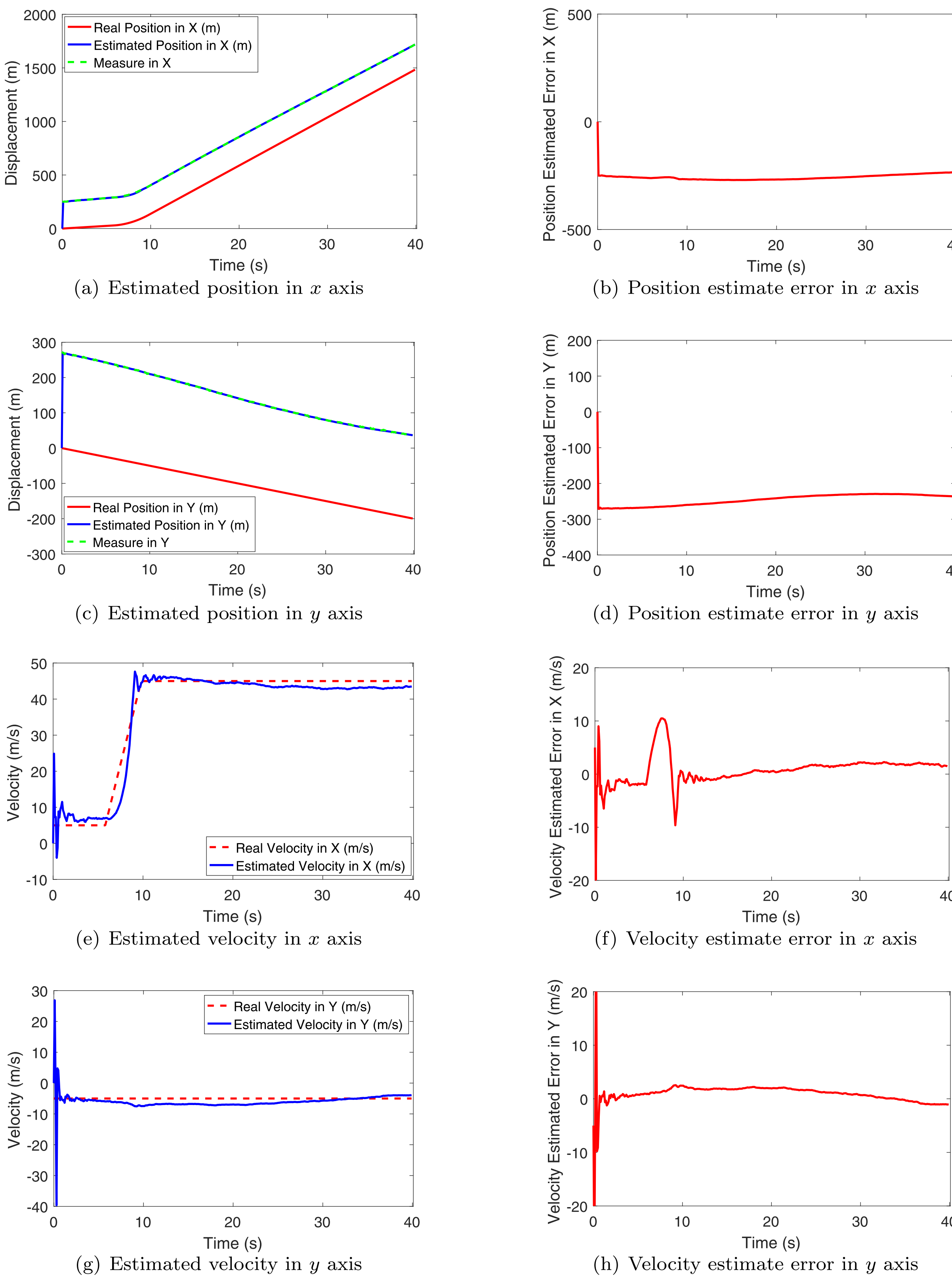
Lemma 6. *If G_{k-l}^k is continuous over both $\hat{\rho}_{k-l}^k$ and ρ_{k-l}^k, then all the limit points of any instance of the array $\{\hat{\rho}_r\} = \{\hat{\rho}_{k-l}^k(r)\}$ defined in Definition 3 are stationary points (local maxima) of T_{k-l}^k, and $T_{k-l}^k\left(\hat{\rho}_{k-l}^k(r)\right)$ converges monotonically to its peaks $T^* = T(\rho^*)$ for some stationary points $\boldsymbol{\rho}^*$. Further, if $T_{k-l}^k\left(\rho_{k-l}^k\right)$ is unimodal, then $\hat{\rho}_{k-l}^k(r) \to \rho^*$.*

Proof. According to Lemma 2, Lemma 4 (Concave is sufficient to unimodal), Lemma 5, *Theorem 2* and *Corollary 1* in [39], this lemma holds. □

4.1. Existence and uniqueness of solution to LJEIP

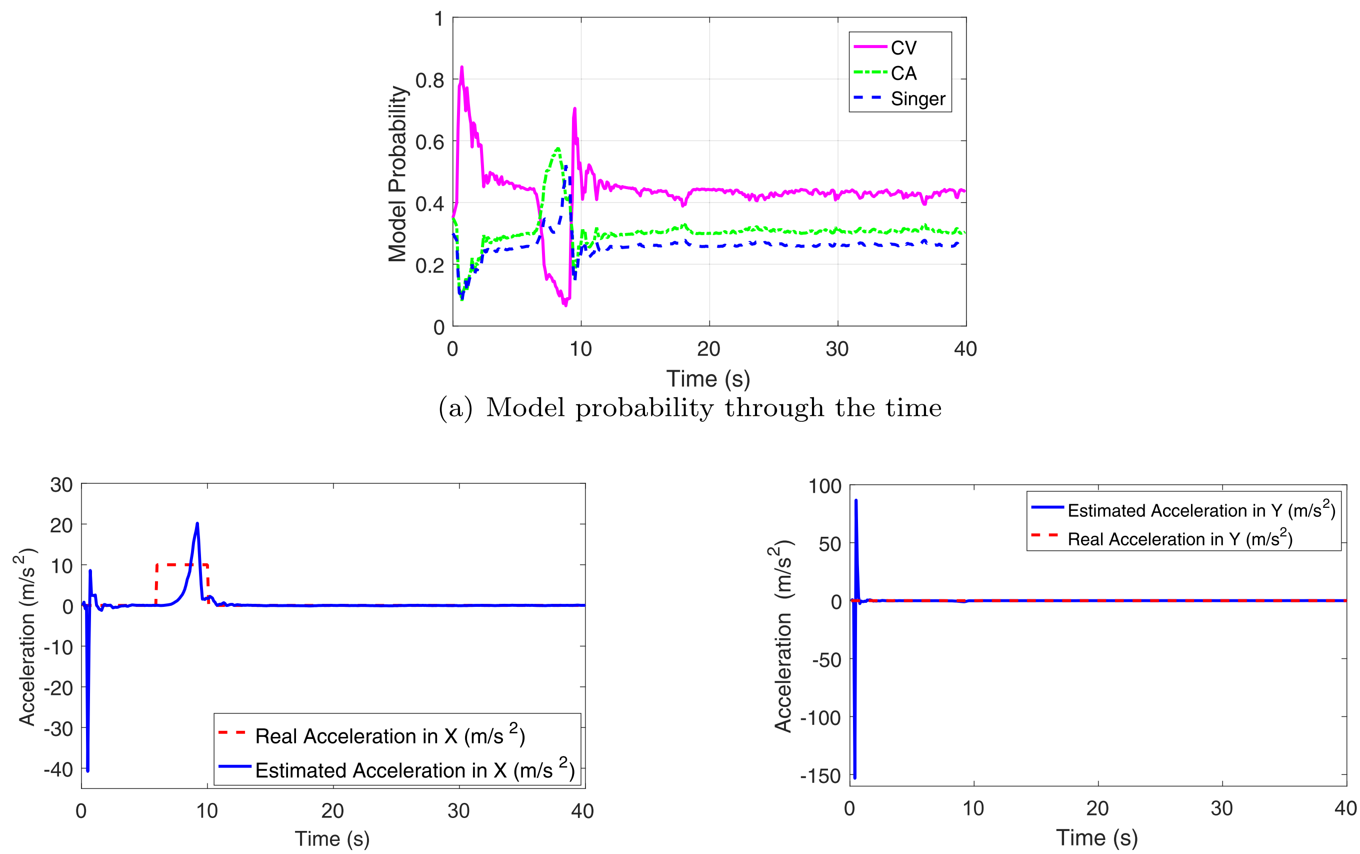
In this section, we prove the existence and uniqueness of the solution to the LJEIP.

Theorem 3. *For the LJEIP defined by (1), if the system measures $\mathbf{y}_i, i = 1, 2, 3, \ldots N$ are uniquely pre-given, and the input-driving matrices $\mathbf{M}_i$, $\mathbf{N}_i$ are all of full column rank, then the solutions to LJEIP, that is, $\hat{\mathbf{x}}_i$, $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{b}}_i$, uniquely exist, with the initial conditions $\mathbf{x}_0$, $\boldsymbol{\Sigma}_0$ and arbitrarily given norm-finite $\hat{\rho}_0$ $(\left\|\hat{\rho}_0\right\| < \infty)$, in the sense of linear unbiased minimum variance, no matter whether $\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^T$ is*



(a) Estimated position in x axis
(b) Position estimate error in x axis
(c) Estimated position in y axis
(d) Position estimate error in y axis
(e) Estimated velocity in x axis
(f) Velocity estimate error in x axis
(g) Estimated velocity in y axis
(h) Velocity estimate error in y axis

Fig. 4. The tracking results given by the IMM method.



(a) Model probability through the time

(b) Real and identified acceleration in x axis

(c) Real and identified acceleration in y axis

Fig. 5. The model probability and target manoeuvre identification results of the IMM method.

invertible (that is, whether Γ_{i-1} is of full rank) or not. Here $\hat{\mathbf{x}}_i$, $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{b}}_i$ is the estimates of $\mathbf{x}_i$, $\mathbf{a}_i$ and $\mathbf{b}_i$, respectively.

Proof. To be identical with the definitions given in previous sections, let $k-l=1$ and $k=N$, if we using Definition 3 to generate a sequence $\hat{\rho}^k_{k-l}(r)$, then according to Lemma 6, $\hat{\mathbf{x}}_i$, $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{b}}_i$ uniquely exists, and $\hat{\mathbf{a}}_i$, $\hat{\mathbf{b}}_i$ are given by $\rho^* = [\mathbf{a}^*, \mathbf{b}^*]^T$. Also, according to Theorem 1, $\hat{\mathbf{x}}_i$ is uniquely determined by RTS smoother within $[k-l, k]$ (Algorithm 1). In summary, the theorem stands. □

4.2. Linear joint estimation and identification theorem (LJEIT)

Theorem 4. *For Linear Joint Estimation and Identification Problem (LJEIP) defined by (1), if the system measures $\mathbf{y}_i, i = 1, 2, 3, \ldots N$ are uniquely pre-given, and the input-driving matrices $\mathbf{M}_i$, $\mathbf{N}_i$ are all of full column rank, then the solutions to LJEIP, that is, $\hat{\mathbf{x}}_i$, $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{b}}_i$, uniquely exist, in the sense of linear unbiased minimum variance, no matter whether $\mathbf{\Gamma}_{i-1}\mathbf{Q}_{i-1}\mathbf{\Gamma}^T_{i-1}$ is invertible (namely, whether Γ_{i-1} is of full rank) or not. Plus, the numerical values of $\hat{\mathbf{x}}_i$, $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{b}}_i$ could be given by a iteratively convergent process (43) and (44) (see also Fig. 1):*

$$\hat{\mathbf{x}}_i(r) = \mathrm{RTS}\big(\{\mathbf{y}_i\}|_{i=1,2,..,N}, \mathbf{x}_0, \mathbf{\Sigma}_0, \hat{\rho}(r)\big) \tag{43}$$

$$\begin{cases} \hat{\mathbf{a}}_{i-1}(r+1) = \mathcal{A}^{-1} \cdot \mathbf{M}^T_{i-1} \cdot \left[\mathbf{\Gamma}_{i-1}\mathbf{Q}_{i-1}\mathbf{\Gamma}^T_{i-1}\right]^{-1} \\ \quad \cdot \left[\hat{\mathbf{x}}_i(r) - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1}(r)\right] \\ \hat{\mathbf{b}}_i(r+1) = \mathcal{B}^{-1} \cdot \mathbf{N}^T_i \cdot \mathbf{R}^{-1}_i \cdot \left[\mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_i(r)\right], \end{cases} \tag{44}$$

where $\mathcal{A} = \left\{\mathbf{M}^T_{i-1} \cdot \left[\mathbf{\Gamma}_{i-1}\mathbf{Q}_{i-1}\mathbf{\Gamma}^T_{i-1}\right]^{-1} \cdot \mathbf{M}_{i-1}\right\}$, $\mathcal{B} = \left\{\mathbf{N}^T_i \cdot \mathbf{R}^{-1}_i \cdot \mathbf{N}_i\right\}$, $i = 1, 2, \ldots, N$, with the initial conditions $\mathbf{x}_0$, $\mathbf{\Sigma}_0$ and arbitrarily given norm-finite $\hat{\rho}_0$ ($\|\hat{\rho}_0\| < \infty$).

That is, if $r \to \infty$, then

$$\begin{cases} \hat{\mathbf{x}}_i(r) \to \hat{\mathbf{x}}_i \\ \hat{\mathbf{a}}_i(r) \to \hat{\mathbf{a}}_i \\ \hat{\mathbf{b}}_i(r) \to \hat{\mathbf{b}}_i. \end{cases} \tag{45}$$

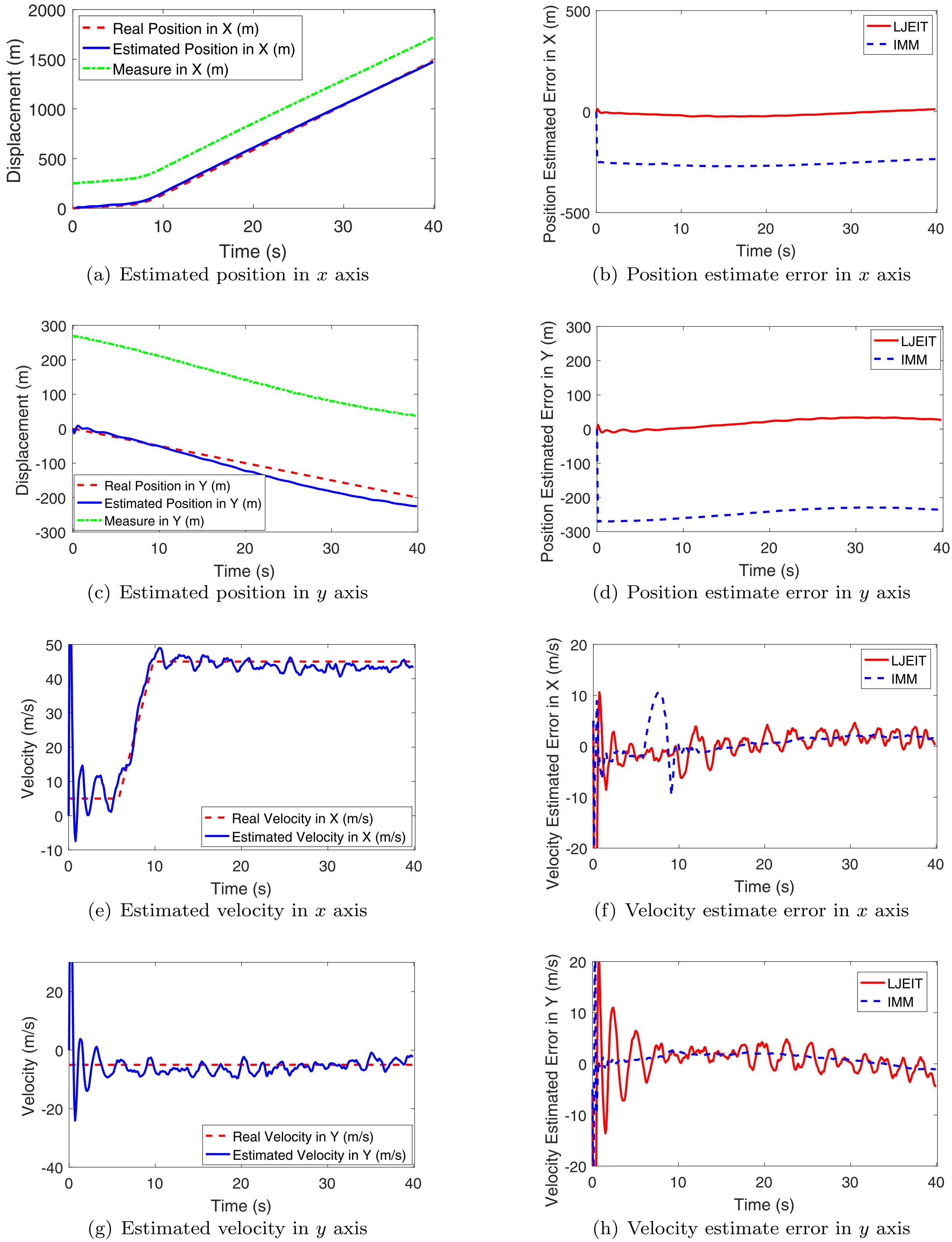
In Eqs. (43)∼(45), $r = 1, 2, 3, \ldots$ indicates the r^{th} iteration, and the function RTS means executing the RTS smoother defined by Algorithm 1. For the case that $\mathbf{\Gamma}_{i-1}\mathbf{Q}_{i-1}\mathbf{\Gamma}^T_{i-1}$ is non-invertible, see Appendix A.3.

Proof. According to Lemma 4 and Theorem 3, the theorem stands. □

Remark 7. To stress that Theorem 3 and Theorem 4 stand in the sense of linear unbiased minimum variance is because the RTS smoother works in the sense of linear unbiased minimum variance. Generally, an estimate of x_i, denoted by $\hat{x}_i$, is said to be optimal in the minimum mean square error (MMSE) sense, if $\hat{x}_i$ minimizes $E[\|x_i - \hat{x}_i\|^2]$. Since the unknown input $\mathbf{a}_{i-1}$ and $\mathbf{b}_i$ are deterministic, not random variables, the estimates to them, $\hat{\mathbf{a}}_{i-1}$ and $\hat{\mathbf{b}}_i$, should be dependent on the properties of $\hat{\mathbf{x}}_i$. Specifically, according to Eq. (44), the estimates to $\mathbf{a}_{i-1}$ and $\mathbf{b}_i$ are also given in the sense of linear unbiased minimum variance, with mean of $\hat{\mathbf{a}}_{i-1}$ and $\hat{\mathbf{b}}_i$, respectively.

The diagram of iteration process defined in Theorem 4 is illustrated in Fig. 1.

For a real system in engineering, the measures $\{\mathbf{y}_i | i = 1, 2, \ldots\}$ are always got in sequence in accordance with time step, that is,



(a) Estimated position in x axis
(b) Position estimate error in x axis
(c) Estimated position in y axis
(d) Position estimate error in y axis
(e) Estimated velocity in x axis
(f) Velocity estimate error in x axis
(g) Estimated velocity in y axis
(h) Velocity estimate error in y axis

Fig. 6. The tracking results given by the LJEIT proposed in this paper.

Algorithm 2 The algorithm of solving the LJEIP in sequence. See Remark 9 and Eqs. (43) and (44) is in Algorithm.

Definition: $r_{\max}$ is maximum iteration steps; δ_L is desired precision to, if reached, end the iteration process; δ_ρ is threshold to decide whether $\hat{\rho}$ should be zero.
Initialize: $r \leftarrow 0$, $k \leftarrow 0$, $\hat{\rho}_0 \leftarrow \mathbf{0}$
Input: $\mathbf{x}_0$, $\mathbf{\Sigma}_0$, $r_{\max}$, δ_L, δ_ρ, l and y_k, $k = 1, 2, 3, \ldots$

while true **do**
  $k \leftarrow k+1$
  if $k \le l$ **then**
    Continue While
  end if
  $\hat{\rho}(0) \leftarrow \hat{\rho}_0$
  $r \leftarrow 0$
  repeat
    The iteration process (43) and (44)
    $r \leftarrow r+1$
  until $r > r_{\max}$ or $\left\| \hat{\rho}_i(r+1) - \hat{\rho}_i(r) \right\| < \delta_L$
  // To check if the unknown input is zero or not
  for each ρ in $\hat{\rho}(r+1)\big|_{k-l \le i \le k}$ **do**
    if $\rho < \delta_\rho$ **then**
      $\rho \leftarrow 0$
    end if
  end for
  // To accelerate the iteration process, use parts of the $\hat{\rho}_{k-l-1}^{k-1}$ as initial conditions. See Remark 9.

$$\hat{\rho}_0 \leftarrow [\,\hat{\rho}(r+1)|_{k-l-1 \le i \le k-1}\{2:l+1\}, \hat{\rho}(r+1)|_{k-l-1 \le i \le k-1}\{l+1\}\,] \tag{46}$$

  // Note that k here actually means the $(k+1)$th iteration. Because in the beginning of "While" statement, we have $k \leftarrow k+1$.
  Record: $\hat{\mathbf{x}}_k$, $\hat{\mathbf{a}}_k$, $\hat{\mathbf{b}}_k$
  if end of getting $\mathbf{y_k}$ **then**
    Break While
  end if
end while
Output: $\{(\hat{\mathbf{x}}_k, \hat{\mathbf{a}}_k, \hat{\mathbf{b}}_k)|k = 1, 2, 3, \ldots,\}$

$\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots\}$, instead of in block. Thus we provide Algorithm 2 to solve the LJEIP.

Remark 8. In Algorithm 2 two practical skills are used: (a) Use a threshold to set the estimates of some unknown inputs to zero. Because due to the existence of noise, even though there is no unknown input (that is $\boldsymbol{\rho} \equiv \mathbf{0}$), $\hat{\rho}$ would not always be zero [5]; (b) Use the estimated $\hat{\rho}$ in time interval $[k-l-1, k-1]$ as parts of the initial conditions to execute iteration process in time interval $[k-l, k]$. In detail, $\hat{\rho}_0$ in time interval $[k-l, k]$ is constructed as $\hat{\rho}_0\big|_{[k-l,k]} = [\ \hat{\rho}_{pre}\{2:l+1\},\ \ \hat{\rho}_{pre}\{l+1\}\ ]$, where $\hat{\rho}_{pre}$ means the estimate to $\boldsymbol{\rho}$ in time interval $[k-l-1, k-1]$, $\{2:l+1\}$ means extract the columns from column 2 to column $l+1$, and $\{l+1\}$ means the last column. We can also construct the initial value $\hat{\rho}_0$ by $\hat{\rho}_0\big|_{[k-l,k]} = [\ \hat{\rho}_{pre}\{l+1\}\ \ \hat{\rho}_{pre}\{l+1\}, \ldots,\ \ \hat{\rho}_{pre}\{l+1\}\ \ \hat{\rho}_{pre}\{l+1\}\ ]$, just as Lan et al. does in [5].

Remark 9. For Algorithm 2, another one important trick in practice is that we can let the estimates to the unknown inputs keeps constant within $[k-l, k]$, meaning taking the average of all the estimates at different steps within this window so that we can weaken the negative influences introduced by noises, just as Lan et al. does in [5,8]. We in the simulation of this paper also use this trick.

5. Simulation experiments and results analysis

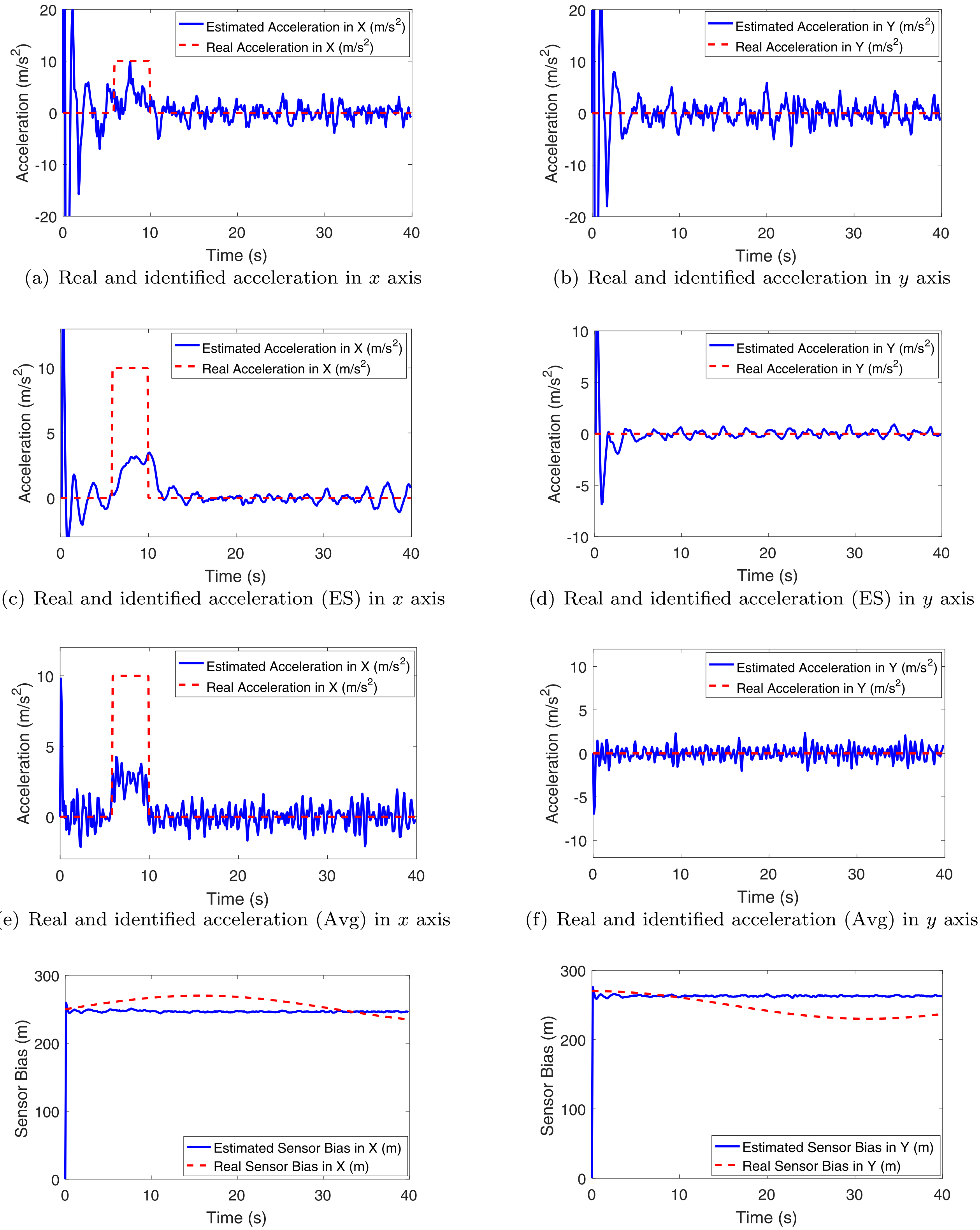
In Lan et al. [5], the authors demonstrated the simulation of tracking a manoeuvring target in the presence of the RGPO (range gate pull-off). In the experiment, the target manoeuvre exists as unknown input in system dynamics model and the RGPO as unknown input in measurement model. The experiment results illustrated the proposed method could simultaneously estimate the system states (position and velocity of the target) and identify the unknown inputs both in the system model and the measurement model. Notably, it is the identifications of unknown inputs that improve the estimate accuracy of the system states. Overall, the proposed method outperformed the canonical Interactive Multi-model (IMM) method, because IMM is powerless for the unknown inputs appearing in measurement model. As a corroboration to [5], we in this paper consider another scenario of tracking a manoeuvring target. The simulation is based on a desktop having the following configurations:

- **Operation System:** Windows 10 Education;
- **CPU:** Intel Core i7 3.2GHz x64;
- **RAM:** 8G.

5.1. Simulation scenario and problem formulation

The scenario in this paper is with following conditions:

- The simulation evolves in total $t_{\max} = 40s$, with the sampling time $T_s = 0.1s$. Thus the discrete time span is $0 \le k \le 400$;
- The target moves in the ground (a 2D-plane with x axis and y axis) with a constant velocity. Its initial position is $[0, 0]^T m$ and initial velocity is $[5,\ -5]^T m/s$. However, during the time slot $60 \le k \le 100$ $(6 \le t \le 10)$, it manoeuvres in the x axis with the acceleration of $10 m/s^2$;
- The ranging radar system (the sensor) could directly obtain the position of a mobile target. However, there exist radar biases in



(a) Real and identified acceleration in x axis

(b) Real and identified acceleration in y axis

(c) Real and identified acceleration (ES) in x axis

(d) Real and identified acceleration (ES) in y axis

(e) Real and identified acceleration (Avg) in x axis

(f) Real and identified acceleration (Avg) in y axis

(g) Real and identified radar biases in x axis

(h) Real and identified radar biases in y axis

Fig. 7. The unknown inputs identification results of the LJEIT.

measures because of the sensor faults, electronic countermeasures, initial position configuration errors, and/or radar self-localization errors. Mathematically, in our simulation the radar biases are given as

$$B_x = A_0 + A\sin(0.01t)$$
$$B_y = A_0 + A\cos(0.01t)$$

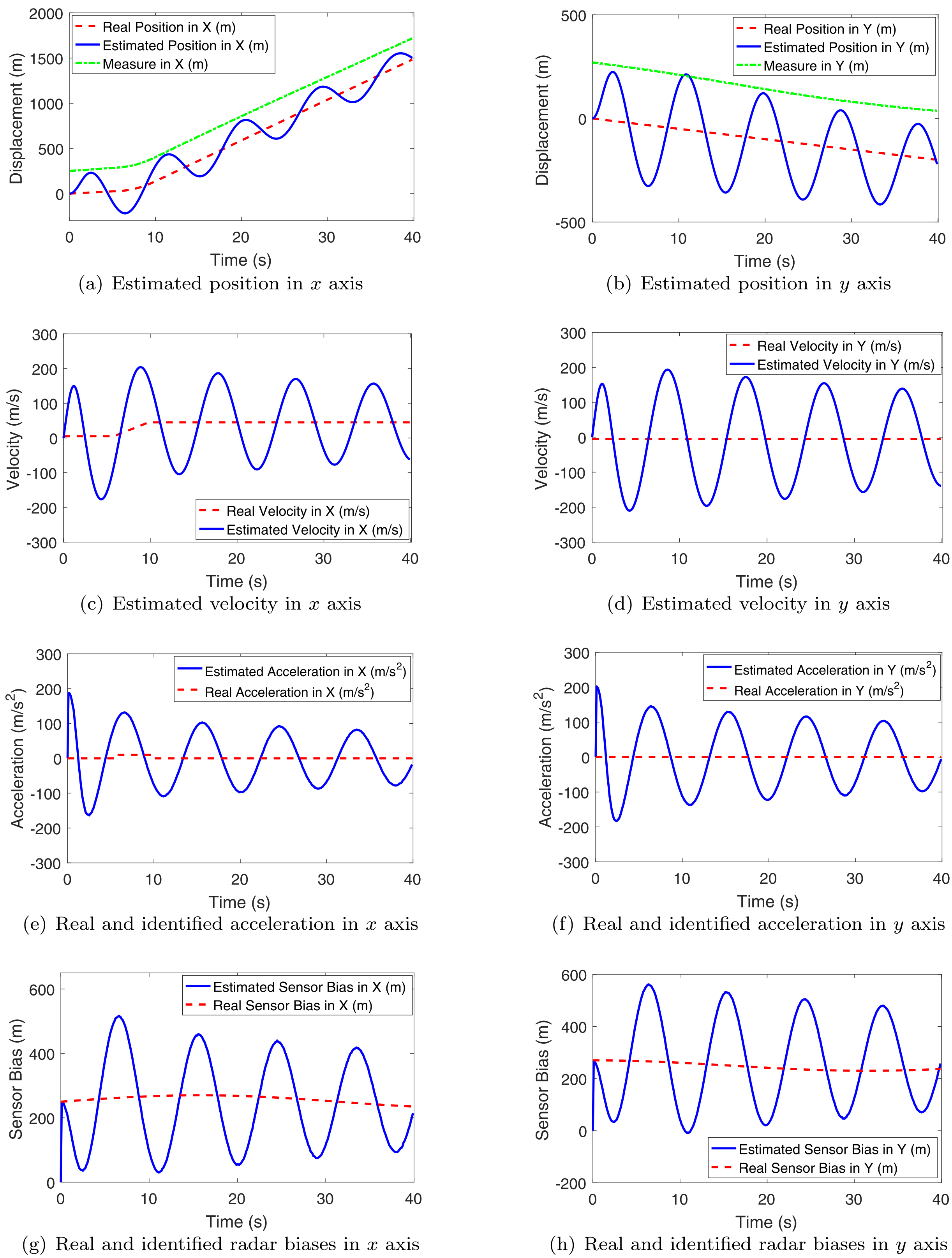
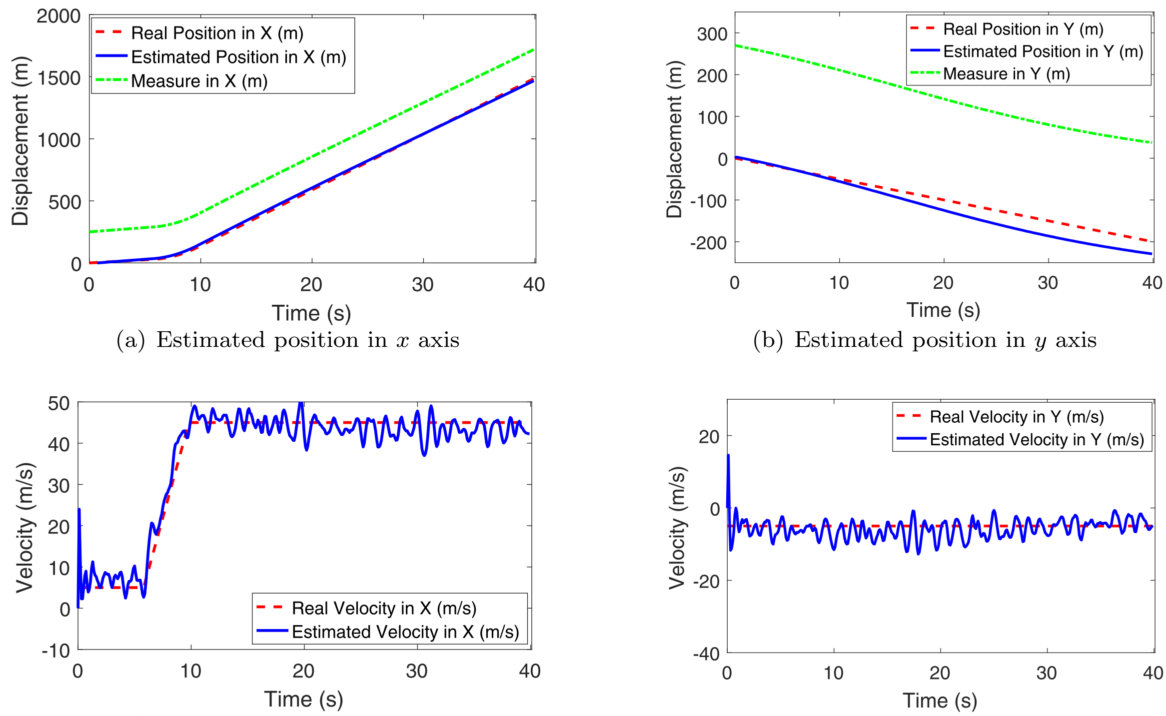


Fig. 8. The tracking results given by the LJEIT when $l = 1$.

where t denotes the time in seconds; B_x is radar bias in x axis and B_y is radar bias in y axis. The constants $A_0 = 250$m and $A = 20$m. The measurement noise is $1m$ in both x and y axis, that is, the covariance matrix in (1) is $\mathbf{R}_{k+1} = diag\{1, \ 1\}$.

Obviously in our simulation, the system states $\mathbf{x}_k$ are real-time position and velocity of the target. The manoeuvre (acceleration within $60 \le k \le 100$) is the unknown input $\mathbf{a}_k$ in system model and the radar bias is unknown input $\mathbf{b}_k$ in measurement model. Henceforth, in the following we should estimate the position and velocity



(a) Estimated position in x axis

(b) Estimated position in y axis

(c) Estimated velocity in x axis

(d) Estimated velocity in y axis

Fig. 9. The tracking results given by the LJEIT when $l = 5$ (Part I).

of the target and identify the unknown inputs both in the system model and measurement model. The unknown inputs in the system model and measurement model is graphically given in Fig. 2.

As a result, we have the real position and measured position in x axis and y axis in Fig. 3, respectively.

From Fig. 3, we can find that the sensor biases and target manoeuvre introduce the significant influences to target tracking and target trajectory. Although canonical filters of Kalman family, including the IMM method, could to some extent handle the problem caused by target manoeuvre, they could do nothing to sensor bias so that the estimate error could be intolerable. Thus in order to further improve the tracking accuracy, we must identify the unknown inputs. This is why our story comes. In Subsection 5.2, we firstly display the tracking results given by the IMM method, just as an intuitive understanding to the insufficiency of canonical filters, and in Subsection 5.3 we show the tracking results given by our Algorithm 2, namely the LJEIT (Theorem 4) in this paper.

5.2. Tracking results by the IMM method

We consider three different tacking models in this part, constant velocity model (CV), constant acceleration model (CA) and Singer model (Singer). The mathematical expressions of the three models could be found in [41]. For briefness, we omit them here. In the Singer model, we set the key parameter (the reciprocal of the maneuver time constant τ) as $\alpha = 1/20$, as suggested in [42]. Additionally, the initial model probability vector is $[CV,\ CA,\ Singer]^T = [0.35,\ 0.35,\ 0.3]^T$, and the model probability transition matrix is

$$P_{ij} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}.$$

The initial system state $\mathbf{x}_0$ is set as $\mathbf{x}_0 = \mathbf{0}_{6\times1}$, meaning a 6-dimension vector having all elements of zero; the initial covariance $\Sigma_0 = \mathbf{I}_{6\times6} \times 10^5$, meaning a 6-dimension identical matrix multiplied by a large number.

With all simulation conditions aforementioned, the tracking results by the IMM method are illustrated in Fig. 4.

Besides, we have the model probability through the time, and manoeuvre (unknown input in system model) identification results in Fig. 5. From Fig. 5, we can see that the IMM method could identify the target manoeuvre to some degree, that is, the unknown input in system model. Because the CA model stands out during $6s \le t \le 10s$ $(60 \le k \le 100)$. However, it can do nothing to sensor errors. Thus, Fig. 4 presents the significant position estimate error.

5.3. Tracking results by the LJEIT in this paper

In this part, we use Algorithm 2 presented in this paper to estimate the system state and identify the unknown inputs. The system model $\mathbf{F}_k$ we use in (1) is the CV model [41], that is

$$\mathbf{F}_k = \begin{bmatrix} 1 & Ts & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & Ts \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

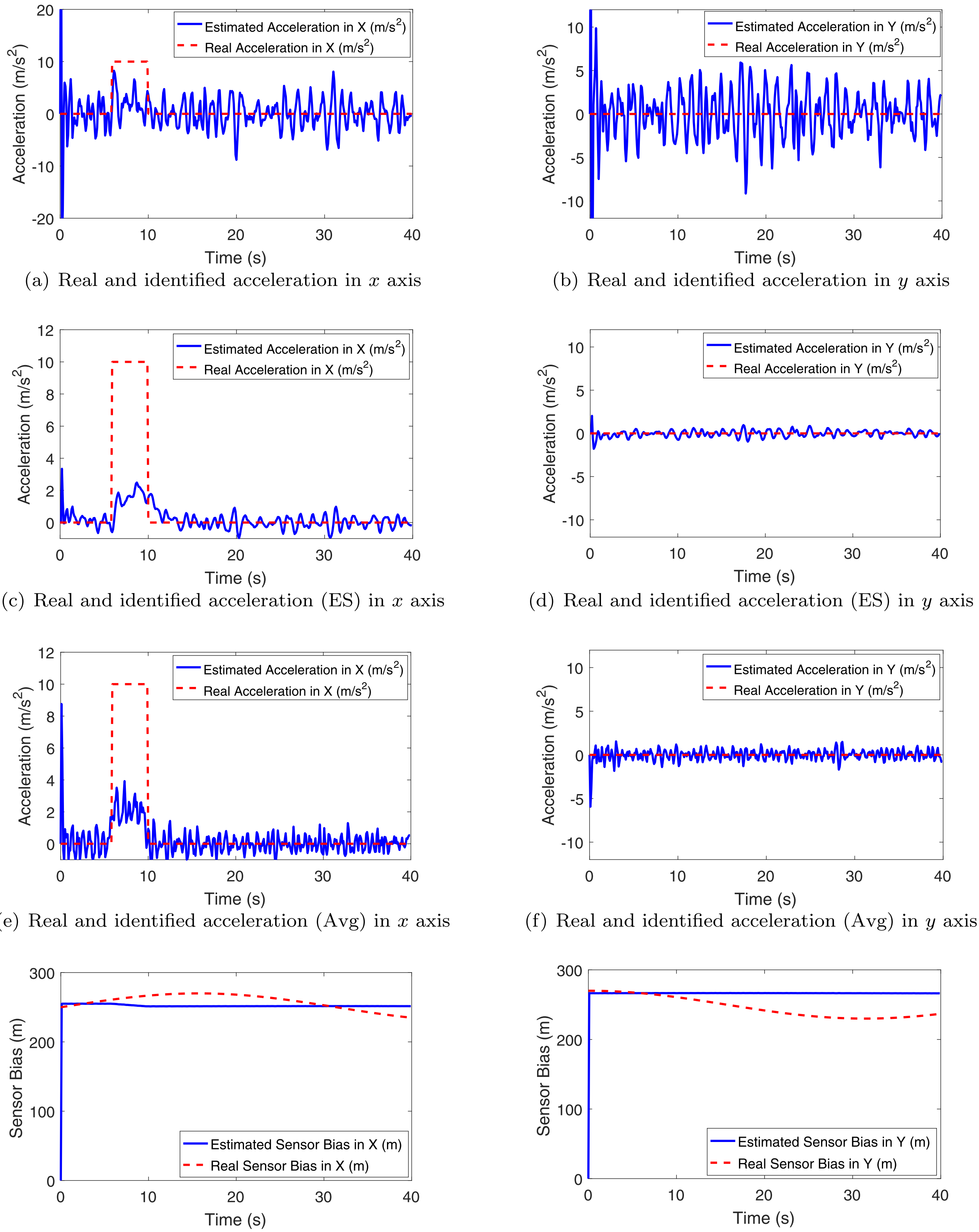
Thus we have the modelling error because we do not take the unknown input, existed as acceleration, into our consideration of modelling. However, we could instead model the unknown acceleration as unknown input $\mathbf{a}_k$ which is driven by input-driven matrix $\mathbf{M}_k$, where

$$\mathbf{M}_k = \begin{bmatrix} Ts & 0 \\ 1 & 0 \\ 0 & Ts \\ 0 & 1 \end{bmatrix}.$$

Besides, we should have the measurement model

$$\mathbf{H}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Similarly, the measurement model is also insufficient because it failed to explain the radar biases. Therefore we model the radar



(a) Real and identified acceleration in x axis
(b) Real and identified acceleration in y axis
(c) Real and identified acceleration (ES) in x axis
(d) Real and identified acceleration (ES) in y axis
(e) Real and identified acceleration (Avg) in x axis
(f) Real and identified acceleration (Avg) in y axis
(g) Real and identified radar biases in x axis
(h) Real and identified radar biases in y axis

Fig. 10. The tracking results given by the LJEIT when $l = 5$ (Part II).

biases as unknown input $\mathbf{b}_k$ in measurement model driven by the matrix $\mathbf{N}_k = \mathbf{I}_{2\times 2}$.

Plus, we give the initial settings: the initial system state $\mathbf{x}_0 = \mathbf{0}_{4\times 1}$, meaning a 4-dimension vector having all elements of zero; the initial co-variance as $\Sigma_0 = \mathbf{I}_{4\times 4} \times 10^5$, meaning a 4-dimension identical matrix multiplied by a large number. The process noise covariance is $\mathbf{Q}_k = diag\{1,\ 0.2,\ 1,\ 0.2\}$ (cm), and the noise driven matrix is $\mathbf{\Gamma}_k = \mathbf{I}_{4\times 4}$; the $\mathbf{R}_k$ is given in Subsection 5.1.

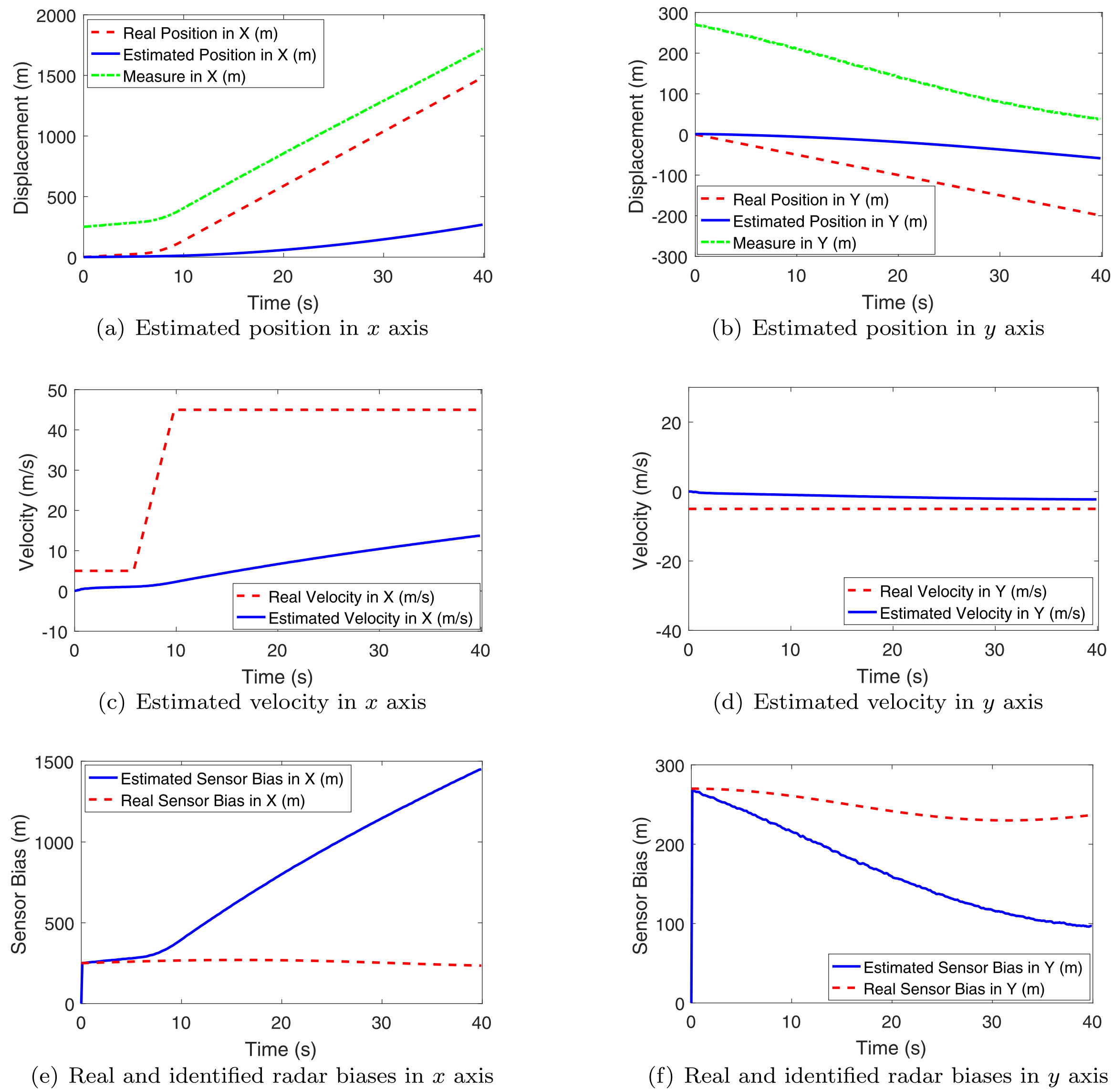


Fig. 11. Estimate radar biases regardless of manoeuvring acceleration.

As a demonstration, we set our parameters in Algorithm 2 as $r_{\max} = 45$, $\delta_L = 10^{-3}$, $\delta_\rho = 10^{-4}$, $l = 3$. Then the tracking results are illustrated in Fig. 6.

In addition, we have the manoeuvre (unknown input in system model) and radar biases (unknown input in measurement model) identification results in Fig. 7. In Fig. 7, (a), (b), (g) and (h) are generated by the LJEIT. And (c) and (d) are produced by the first-order exponential smoothing (ES) from (a) and (b), respectively, with the filter coefficient $\alpha = 0.07$. We use the ES to weaken the influence of noises, in order to make interested signals outstanding so that it is easy for us to identify signals by eyes. Alternatively, we do other 100 times of Monte Carlo simulations and average the estimates to the manoeuvre (Avg), which gives (e) and (f).

As a comparison, we set the window length $l = 1$ and $l = 5$. We in turn have the tracking results showed in Figs. 8–10. For better typesetting layout, we split the results of $l = 5$ into two parts, placed in Fig. 9 (Part I) and Fig. 10 (Part II), respectively.

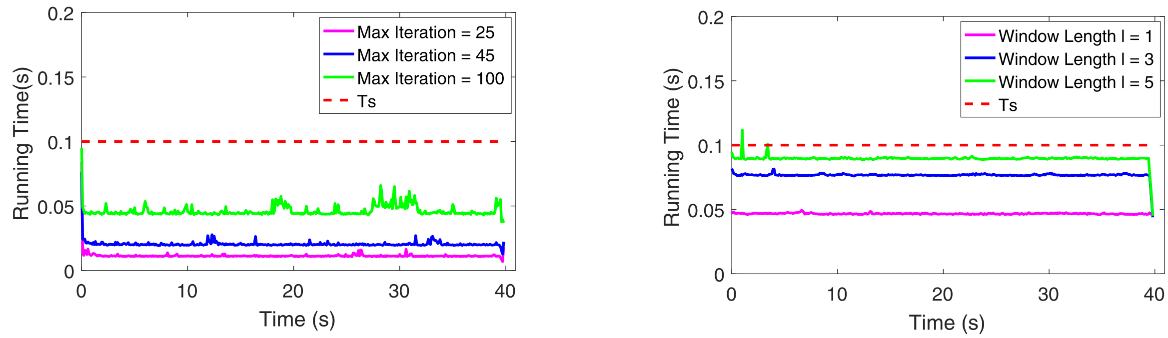
In order to testify the mutual dependence between estimating manoeuvring acceleration (unknown input in the system model) and estimating radar biases (unknown input in the measurement model), we do not consider the estimate to the manoeuvre in the scenario of $l = 3$ and $r_{\max} = 45$, meaning the estimation to the manoeuvring acceleration is always set as zero. Then we have Fig. 11.

5.4. Results analysis

From Figs. 6 and 7. We can see the position estimate of the LJEIT significantly outperforms that of the IMM. Besides, in the x axis and during the target manoeuvring period, the velocity estimate error of the LJEIT is also notably small than that of the IMM.

Additionally, Figs. 8–10 indicate that different window length means different tracking performance. It is obvious that $l = 1$ is not sufficient to our problem and it can only tell apart the main trend of the changing laws of position, velocity, and unknown inputs. $l = 3$ and $l = 5$ are comparatively more suitable.

Besides, we can find in Fig. 11 that the algorithm no longer converges (or maybe it converges within an extremely long time span). This indicates that actually the estimation to the manoeuvring acceleration is not independent of the estimation to the radar



(a) Window Length $l = 5$ (b) Max Iteration $r_{max} = 200$

Fig. 12. The computational burden of the algorithm with different parameters.

biases, meaning the identification of unknown inputs are highly coupled, no matter the inputs are in the system model, in the measurement model or in both. This phenomena is not confusing since the LJEIT requires the conjunct participation of the system states and all of the unknown inputs.

Unsurprisingly, the identification error still also exits in our method. This is dominantly because the algorithm has limited iteration steps ($r_{\max} = 45$), which leads to the truncation error. However this is unavoidable in practice. As long as the accuracy of the results is feasible, the less of the maximum iteration steps, the better the result is.

5.5. Computational burden of LJEIT

Although the proposed LJEIT performs well, as a numerical algorithm, we must concern its computational efficiency, since we only accept the real-time methods in engineering at least in the signal processing community. The real-time algorithm means an algorithm that could return a feasible solution to the problem within a limited time slot like 0.02s, 0.1s or 1s. Now that the sampling time in this paper is $T_s = 0.1s$. The real-time accordingly means the LJEIT could estimate the system states and identify the unknown inputs at each step k within 0.1s. Obviously in our algorithm, the window length l and the maximum iteration steps $r_{\max}$ dominate the main part of the running time. We give the typical testing results of our algorithm in Fig. 12. In Fig. 12, all the used parameters combination are verified to be efficient to generate a feasible solution. For briefness, we neglect the solutions they gave, only providing the computation performances.

From Fig. 12, we can see that even if the maximum iteration step is set as $r_{\max} = 200$, or window length as $l = 5$, all the computations are meaningful since they all terminated within $T_s = 0.1s$. In practice, at some steps the algorithm may not terminate within the limited time slot T_s, for example, the case of $l = 5$ in Fig. 12 (b). What we should do is just to mandatorily stop its execution and return the current truncated solution. In a long run, as long as the number of these kinds of mandatory stop is small, it would not create disaster.

6. Conclusion and future work

In this paper, we study the important LJEIP and extend Lan et al. [5] by providing strong theoretical results. They mainly include the existence and uniqueness proof for solutions of the LJEIP, and effectiveness and convergence proof for the EM-based algorithm generated by the LJEIT. Simulation results validates the effectiveness and efficiency of our method detailed in Theorem 4 and Algorithm 2. Although powerful in many senses, our algorithm still faces several challenges which deserve further study as follows.

- During designing our simulation experiments, we found that, different input-driven matrix ($\mathbf{M}_k$ and/or $\mathbf{N}_k$) means different computational performance in generating the solution, for example the average running time at each step, and truncation error (algorithm accuracy) when given the fixed maximum iteration steps, and so on. Thus the first open problem is to discover the pattern of how the selected input-driven matrix influence the algorithm performances and which family of input-driven matrices could optimize the algorithm efficiency under the fixed parameters (l, $r_{\max}$ etc.);
- Figs. 6–10 indicate that in order to improve the effectiveness of the solution, the window length should be sufficient. However, it does not mean that, the longer the window length is, the better the result is. Instead, it possibly introduces extra computational burden. Thus, the second open problem is to find out, ***under the given maximum iteration steps***, what the best (or feasible) window length should be. Because, in Theorem 4, we do not take into account the specific window length, just asserting the convergence when the iteration number at each step tends to infinity;
- The third open problem is, besides the deterministic signals as the unknown inputs assumed in Assumption 1, which classes of unknown signals the proposed method can handle, and what are the corresponding solutions to them. Because the underlying philosophy of the EM method is essentially the maximum likelihood estimation. In theory, as long as we can find the corresponding exact/approximated probability distribution of interested joint and marginal distributions [38,40], we can carry out the EM method to construct the numerical solution.

Hence here, we invite interested readers in the community to study the above open problems.

Acknowledgment

The corresponding author would like to thank Assoc. Prof. Caihua Chen (chchen@nju.edu.cn, NJU, China) and Dr. Yue Zhao (yuezhao@u.nus.edu, NUS, Singapore) for their helpful discussions with him. Plus, we appreciate Assoc. Prof. Hua Lan (lanhua@nwpu.edu.cn, NWPU, China) for sharing his source code of [8] with us for references so that we can finish our simulation experiment in this paper in a relatively short time.

Declarations of interest

The authors declare that there is no any potential competing interests.

Appendix A

A1. Errors in Lan et al. [5]

In this section, we list and fix some minor errors in Lan et al. [5] in the following Table A1. In this Table, the first column is the error sequence number; the second column list the error equation number in Lan et al. [5]; and the last column is the correct equation no in current paper.

Table A1
Minor Errors in [5] and Corrections in this paper.

No.	Error equations in [5]	Corrections in this paper
1	Eq. (2)	Eq. (47)
2	Eq. (4)	Eq. (48)
3	Eq. (6)	Eq. (49)
4	Eq. (10)	Eq. (50)
5	Eq. (11)	Eq. (51)
6	Eq. (24)	Eq. (52)
7	Eq. (25)	Eq. (53)

$$L_{k-l}^{k} = \log p(\mathbf{x}_{k-l},\ldots,\mathbf{x}_k,\mathbf{y}_{k-l},\ldots,\mathbf{y}_k \big| \rho_{k-l}^{k},\hat{\rho}_1^{k-l-1},\mathbf{Y}_1^{k-l-1},\mathbf{x}_0,\boldsymbol{\Sigma}_0) = \log p(\mathbf{x}_{k-l-1}\big|\hat{\rho}_1^{k-l-1},\mathbf{Y}_1^{k-l-1},\mathbf{x}_0,\boldsymbol{\Sigma}_0) + \sum_{i=k-l}^{k}\log p(\mathbf{x}_i|\mathbf{x}_{i-1},\rho_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}) + \sum_{i=k-l}^{k}\log p(\mathbf{y}_i|\mathbf{x}_i,\rho_{k-l}^{k}), \tag{47}$$

where log is natural logarithmic function.

$$L_{0,k-l}^{k} = -\frac{n+(l+1)(m+n)}{2}\log(2\pi) - \frac{1}{2}\sum_{i=k-l}^{k}\left(\log\left|\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right| + \log|\mathbf{R}_i|\right), \tag{48}$$

where $|\mathbf{A}|$ means determinant of matrix $\mathbf{A}$, which is left unspecified in [5].

$$L_{2,k-l}^{k} = -\frac{1}{2}\sum_{i=k-l}^{k} D\left(\mathbf{x}_i - \mathbf{F}_{i-1}\mathbf{x}_{i-1} - \mathbf{M}_{i-1}\mathbf{a}_{i-1},\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right) \tag{49}$$

$$G_{2,k-l}^{k} = -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right]^{-1}\cdot\left[C\left(\hat{\mathbf{x}}_{i|k-l:k} - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1|k-l:k} - \mathbf{M}_{i-1}\mathbf{a}_{i-1}\right) + \mathbf{P}_{i,i|k-l:k} - \mathbf{P}_{i,i-1|k-l:k}\mathbf{F}_{i-1}^{T} - \mathbf{F}_{i-1}\mathbf{P}_{i-1,i|k-l:k} + \mathbf{F}_{i-1}\mathbf{P}_{i-1,i-1|k-l:k}\mathbf{F}_{i-1}^{T}\right]\right\} \tag{50}$$

Eq. (29) in Lan et al. [5] should be fixed similarly.

$$G_{3,k-l}^{k} = -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\mathbf{R}_i^{-1}\cdot\left[C\left(\mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i|k-l:k} - \mathbf{N}_i\mathbf{b}_i\right) - \mathbf{H}_i\mathbf{P}_{i,i|k-l:k}\mathbf{H}_i^{T}\right]\right\} \tag{51}$$

Eq. (31) in Lan et al. [5] should be fixed similarly.

$$p(\mathbf{x}_{k-l-1}\big|\mathbf{Y}_1^{k-l-1},\hat{\rho}_1^{k-l-1},\mathbf{x}_0,\boldsymbol{\Sigma}_0) = \mathbf{N}(\hat{\mathbf{x}}_{k-l-1},\mathbf{P}_{k-l-1}) \tag{52}$$

$$p(\mathbf{x}_i\big|\mathbf{x}_{i-1},\rho_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}) = \mathbf{N}(\mathbf{F}_{i-1}\mathbf{x}_{i-1} + \mathbf{M}_{i-1}\mathbf{a}_{i-1},\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}) \tag{53}$$

Besides those minor errors listed in the above Table A.1, we note several other important errors in Lan et al. [5].

Eqs. (32) and (33) in [5] are wrong by the definition of derivative operator of matrices. As a result, Eqs. (14) to (17) are neither right in [5] which were derived from Eqs. (32) and (33). We also note that Eq. (32) of Lan et al. [5] holds if and only if there is a premise that $\mathbf{a}_{i-1}$ remains invariant in time interval $[k-l,k]$. We fix Eqs. (32) and (33) in [5]. And corrected equations are (54) and (55) in current paper, respectively.

$$\frac{\partial G_{k-l}^{k}}{\partial \mathbf{a}_{i-1}} = \mathbf{M}_{i-1}^{T}\cdot\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right]^{-1}\cdot\left(\hat{\mathbf{x}}_{i|k-l:k} - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1|k-l:k} - \mathbf{M}_{i-1}\mathbf{a}_{i-1}\right) \tag{54}$$

$$\frac{\partial G_{k-l}^{k}}{\partial \mathbf{b}_i} = \mathbf{N}_i^{T}\cdot\mathbf{R}_i^{-1}\cdot\left(\mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i|k-l:k} - \mathbf{N}_i\mathbf{b}_i\right) \tag{55}$$

A2. Proof of lemma 1

Due to

$$L_{0,k-l}^{k} \equiv const, \tag{56}$$

$G_{0,k-l}^{k}$ could be uniquely determined; As for $G_{1,k-l}^{k}$, $G_{2,k-l}^{k}$ and $G_{3,k-l}^{k}$, given in (57), (58), and (59), respectively.

$$\begin{aligned} G_{1,k-l}^{k} &= E_{\hat{\mathbf{x}}_{k-l-1}}\left[L_{1,k-l}^{k}\left(\rho_{k-l}^{k}\right)\middle|\left(\hat{\rho}_{1}^{k-l-1},\mathbf{Y}_{1}^{k-l-1},\mathbf{x}_0,\boldsymbol{\Sigma}_0\right)\right] \\ &= -\frac{1}{2}Tr\left\{\mathbf{P}_{k-l-1}^{-1}E\left[\mathbf{C}\left(\mathbf{x}_{k-l-1}-\hat{\mathbf{x}}_{k-l-1}\right)\middle|\left(\mathbf{Y}_{1}^{k-l-1},\hat{\rho}_{1}^{k-l-1},\mathbf{x}_0,\boldsymbol{\Sigma}_0\right)\right]\right\} = -\frac{1}{2}Tr\{\mathbf{I}_{n\times n}\} = -\frac{n}{2} \end{aligned} \tag{57}$$

Eq. (57) holds due to $\mathbf{D}(\tilde{\mathbf{x}},\mathbf{P}) = \tilde{\mathbf{x}}^T\mathbf{P}^{-1}\tilde{\mathbf{x}} = Tr(\tilde{\mathbf{x}}^T\mathbf{P}^{-1}\tilde{\mathbf{x}}) = Tr(\mathbf{P}^{-1}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T) = Tr(\mathbf{P}^{-1}\mathbf{C}(\tilde{\mathbf{x}}))$, and $E\mathbf{D}(\tilde{\mathbf{x}},\mathbf{P}) = Tr(\mathbf{P}^{-1}E\mathbf{C}(\tilde{\mathbf{x}})) = Tr(\mathbf{P}^{-1}\mathbf{P}) = Tr(\mathbf{I}_{n\times n})$, where $E(\cdot)$, briefed as E here, means computing the mathematical expectation of the related random variable ($\tilde{\mathbf{x}}$ in this case).

$$\begin{aligned} G_{2,k-l}^{k} &= E_{\hat{\mathbf{X}}_{k-l}^{k}}\left[L_{2,k-l}^{k}\left(\rho_{k-l}^{k}\right)\middle|\left(\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right)\right] \\ &= -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right]^{-1}E\left[\mathbf{C}(\mathbf{x}_i-\mathbf{F}_{i-1}\mathbf{x}_{i-1}-\mathbf{M}_{i-1}\mathbf{a}_{i-1})\middle|\left(\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right)\right]\right\} \\ &= -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right]^{-1}E\left\{\mathbf{C}\left[\left(\hat{\mathbf{x}}_{i|k-l:k}+\tilde{\mathbf{x}}_{i|k-l:k}\right)-\mathbf{F}_{i-1}\left(\hat{\mathbf{x}}_{i-1|k-l:k}+\tilde{\mathbf{x}}_{i-1|k-l:k}\right)\ -\mathbf{M}_{i-1}\mathbf{a}_{i-1}\middle|\left[\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right]\right\}\right\} \\ &= -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right]^{-1}\left\{E\left[\mathbf{C}\left(\hat{\mathbf{x}}_{i|k-l:k}-\mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1|k-l:k}-\mathbf{M}_{i-1}\mathbf{a}_{i-1}\right)\right|\right.\right. \\ &\qquad \left.\left.\left(\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right)\right]+E\left[\mathbf{C}\left(\tilde{\mathbf{x}}_{i|k-l:k}-\mathbf{F}_{i-1}\tilde{\mathbf{x}}_{i-1|k-l:k}\right)\middle|\left(\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right)\right]\right\}\right\} \\ &= -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\left[\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}\right]^{-1}\cdot\left[\mathbf{C}\left(\hat{\mathbf{x}}_{i|k-l:k}-\mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1|k-l:k}-\mathbf{M}_{i-1}\mathbf{a}_{i-1}\right)\right.\right. \\ &\qquad \left.\left.+\mathbf{P}_{i,i|k-l:k}-\mathbf{P}_{i,i-1|k-l:k}\mathbf{F}_{i-1}^{T}-\mathbf{F}_{i-1}\mathbf{P}_{i-1,i|k-l:k}+\mathbf{F}_{i-1}\mathbf{P}_{i-1,i-1|k-l:k}\mathbf{F}_{i-1}^{T}\right]\right\} \end{aligned} \tag{58}$$

$$\begin{aligned} G_{3,k-l}^{k} &= E_{\hat{\mathbf{X}}_{k-l}^{k}}\left[L_{3,k-l}^{k}\left(\rho_{k-l}^{k}\right)\middle|\left(\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right)\right] \\ &= -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\mathbf{R}_i^{-1}E\left[\mathbf{C}(\mathbf{y}_i-\mathbf{H}_i\mathbf{x}_i-\mathbf{N}_i\mathbf{b}_i)\middle|\left(\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right)\right]\right\} \\ &= -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\mathbf{R}_i^{-1}E\left[\mathbf{C}\left(\mathbf{y}_i-\mathbf{H}_i\left(\hat{\mathbf{x}}_{i|k-l:k}+\tilde{\mathbf{x}}_{i|k-l:k}\right)-\mathbf{N}_i\mathbf{b}_i\right)\middle|\left(\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right)\right]\right\} \\ &= -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\mathbf{R}_i^{-1}\left\{E\left[\mathbf{C}\left(\mathbf{y}_i-\mathbf{H}_i\hat{\mathbf{x}}_{i|k-l:k}-\mathbf{N}_i\mathbf{b}_i\right)\middle|\left(\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right)\right]-E\left[\mathbf{C}\left(\mathbf{H}_i\tilde{\mathbf{x}}_{i|k-l:k}\right)\middle|\left(\hat{\rho}_{k-l}^{k},\mathbf{Y}_{k-l}^{k},\hat{\mathbf{x}}_{k-l-1}\right)\right]\right\}\right\} \\ &= -\frac{1}{2}Tr\left\{\sum_{i=k-l}^{k}\mathbf{R}_i^{-1}\cdot\left[\mathbf{C}\left(\mathbf{y}_i-\mathbf{H}_i\hat{\mathbf{x}}_{i|k-l:k}-\mathbf{N}_i\mathbf{b}_i\right)-\mathbf{H}_i\mathbf{P}_{i,i|k-l:k}\mathbf{H}_i^{T}\right]\right\} \end{aligned} \tag{59}$$

With similar principle, $H_{0,k-l}^{k}$, $H_{1,k-l}^{k}$, and $H_{2,k-l}^{k}$ could be obtained. For the case that $\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}$ is non-invertible, see Appendix A.3.

A3. discussion on that $\boldsymbol{\Gamma}_{i-1}|_{k-l\le i\le k}^{(n\times s)}$ is rank deficiency

If $\boldsymbol{\Gamma}_{i-1}|_{k-l\le i\le k}^{(n\times s)}$ is not of full rank (namely $\boldsymbol{\Gamma}_{i-1}\mathbf{Q}_{i-1}\boldsymbol{\Gamma}_{i-1}^{T}$ is non-invertible, since $\mathbf{Q}_{i-1}$ is positive definite), then $\boldsymbol{\Gamma}_{i-1}$ must have a full rank sub-matrix which is with dimensions $s\times s$. Because $\boldsymbol{\Gamma}_{i-1}$ is of full column rank. Assume the full-rank sub-matrix is $\boldsymbol{\Gamma}_{i-1}^{0}$, then $\boldsymbol{\Gamma}_{i-1}^{0}$ should be composed of s row vectors from $\boldsymbol{\Gamma}_{i-1}$. Re-sort all n row vectors of $\boldsymbol{\Gamma}_{i-1}$ to let the s row vectors, which compose the full-rank sub-matrix, gather together and to be at head of $\boldsymbol{\Gamma}_{i-1}$. That is, reassign $\boldsymbol{\Gamma}_{i-1}$ to be with the form showed in (60).

$$\begin{aligned} \bar{\boldsymbol{\Gamma}}_{i-1} &= \begin{bmatrix}\boldsymbol{\Gamma}_{i-1}^{0} & \cdots & \boldsymbol{\Gamma}_{i-1}^{j} & \cdots\end{bmatrix}^T \\ &= \begin{bmatrix}\bar{\boldsymbol{\Gamma}}_{i-1}^{1} & \bar{\boldsymbol{\Gamma}}_{i-1}^{2} & \cdots & \bar{\boldsymbol{\Gamma}}_{i-1}^{s} & \bar{\boldsymbol{\Gamma}}_{i-1}^{s+1} & \cdots & \bar{\boldsymbol{\Gamma}}_{i-1}^{k} & \cdots & \bar{\boldsymbol{\Gamma}}_{i-1}^{n}\end{bmatrix}^T \\ &= \begin{bmatrix}\bar{\boldsymbol{\Gamma}}_{i-1}^{0} & \bar{\boldsymbol{\Gamma}}_{i-1}^{s+1} & \cdots & \bar{\boldsymbol{\Gamma}}_{i-1}^{k} & \cdots & \bar{\boldsymbol{\Gamma}}_{i-1}^{n},\end{bmatrix}^T, \end{aligned} \tag{60}$$

where $\boldsymbol{\Gamma}_{i-1}^{j}$ is a row vector from $\boldsymbol{\Gamma}_{i-1}$ but is not a row of $\boldsymbol{\Gamma}_{i-1}^{0}$.

Let κ_k be an indicator mapping the row index of $\bar{\boldsymbol{\Gamma}}_{i-1}$ to the row index of $\boldsymbol{\Gamma}_{i-1}$. For example, $\kappa_k\big|_{1\le k\le n} = \tau|_{1\le\tau\le n}$ means the kth row vector in $\bar{\boldsymbol{\Gamma}}_{i-1}$ is actually the τth row vector in $\boldsymbol{\Gamma}_{i-1}$; $\kappa_{1:s}$ means the $(\kappa_1,\kappa_2,\ldots\kappa_{s-1},\kappa_s)^{th}$ row vectors in $\boldsymbol{\Gamma}_{i-1}$ which are accordingly the $(1,2,\ldots,s-1,s)th$ row vectors in $\bar{\boldsymbol{\Gamma}}_{i-1}$; $\bar{\boldsymbol{\Gamma}}_{i-1}^{k}$ means the kth row vector in $\bar{\boldsymbol{\Gamma}}_{i-1}$.

Let $\mathbf{w}_{i-1} = \mathbf{\Gamma}_{i-1}\mathbf{q}_{i-1}$, then the covariance matrix of $\mathbf{w}_{i-1}$ should be

$$\mathbf{W}_{i-1} = \mathbf{\Gamma}_{i-1}\mathbf{Q}_{i-1}\mathbf{\Gamma}_{i-1}^{T}. \tag{61}$$

Obviously, $\mathbf{W}_{i-1}$ is non-invertible. It means the row elements of $\mathbf{w}_{i-1}$ are not mutually independent. That is, some row elements of $\mathbf{w}_{i-1}$ could be linearly expressed by others.

Rewritten $\mathbf{w}_{i-1}$ as

$$\bar{\mathbf{w}}_{i-1} = \bar{\mathbf{\Gamma}}_{i-1}\bar{\mathbf{q}}_{i-1}, \tag{62}$$

then we have

$$\bar{\mathbf{w}}_{i-1}^{k}|_{n\geq k>s} = \bar{\mathbf{\Omega}}_{i-1}^{k}\begin{bmatrix} \bar{\mathbf{w}}_{i-1}^{1} & \bar{\mathbf{w}}_{i-1}^{2} & \cdots & \bar{\mathbf{w}}_{i-1}^{s} \end{bmatrix}^{T}, \tag{63}$$

where $\bar{\mathbf{\Omega}}_{i-1}^{k}\big|_{(1\times s)}$ is a constant matrix (linearly expressed). Also, it is easy to know that

$$\bar{\mathbf{\Omega}}_{i-1}^{k} = \bar{\mathbf{\Gamma}}_{i-1}^{k}\cdot\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{-1}. \tag{64}$$

Since

$$p\left(\bar{\mathbf{w}}_{i-1}^{1}, \bar{\mathbf{w}}_{i-1}^{2}, \ldots, \bar{\mathbf{w}}_{i-1}^{s}\right) = \mathbf{N}\left[\mathbf{0}, \bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\right], \tag{65}$$

and (66),

$$p\left(\bar{\mathbf{w}}_{i-1}^{k}\middle|\bar{\mathbf{w}}_{i-1}^{1}, \bar{\mathbf{w}}_{i-1}^{2}, \ldots, \bar{\mathbf{w}}_{i-1}^{s}\right) = \mathbf{N}\left[\mathbf{0}, \bar{\mathbf{\Omega}}_{i-1}^{k}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\left(\bar{\mathbf{\Omega}}_{i-1}^{k}\right)^{T}\right], \tag{66}$$

we have (67).

$$\begin{aligned}
&p\left(\mathbf{w}_{i-1}^{1}, \mathbf{w}_{i-1}^{2}, \ldots, \mathbf{w}_{i-1}^{n}\right) = p\left(\bar{\mathbf{w}}_{i-1}^{1}, \bar{\mathbf{w}}_{i-1}^{2}, \ldots, \bar{\mathbf{w}}_{i-1}^{s}, \bar{\mathbf{w}}_{i-1}^{s+1}, \ldots, \bar{\mathbf{w}}_{i-1}^{n}\right)\\
&\quad = p\left(\bar{\mathbf{w}}_{i-1}^{1}, \bar{\mathbf{w}}_{i-1}^{2}, \ldots, \bar{\mathbf{w}}_{i-1}^{s}\right)\times p\left(\bar{\mathbf{w}}_{i-1}^{s+1}\middle|\bar{\mathbf{w}}_{i-1}^{1}, \bar{\mathbf{w}}_{i-1}^{2}, \ldots, \bar{\mathbf{w}}_{i-1}^{s}\right)\times\cdots\times p\left(\bar{\mathbf{w}}_{i-1}^{n}\middle|\bar{\mathbf{w}}_{i-1}^{1}, \bar{\mathbf{w}}_{i-1}^{2}, \ldots, \bar{\mathbf{w}}_{i-1}^{s}\right)\\
&\quad = \mathbf{N}\left[\mathbf{0}, \bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\right]\times\mathbf{N}\left[\mathbf{0}, \bar{\mathbf{\Omega}}_{i-1}^{s+1}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\left(\bar{\mathbf{\Omega}}_{i-1}^{s+1}\right)^{T}\right]\times\cdots\times\mathbf{N}\left[\mathbf{0}, \bar{\mathbf{\Omega}}_{i-1}^{n}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\left(\bar{\mathbf{\Omega}}_{i-1}^{n}\right)^{T}\right]
\end{aligned} \tag{67}$$

Thus, Eq. (10) in this case should be (68)

$$\begin{aligned}
&p\left[\mathbf{x}_{i}|\mathbf{x}_{i-1}, \left(\mathbf{Y}_{k-l}^{k}, \rho_{k-l}^{k}, \hat{\mathbf{x}}_{k-l-1}\right)\right] = \mathbf{N}\left[(\mathbf{F}_{i-1}\mathbf{x}_{i-1} + \mathbf{M}_{i-1}\mathbf{a}_{i-1})|_{\kappa_{1:s}}, \bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\right]\\
&\quad\times\mathbf{N}\left[\bar{\mathbf{\Omega}}_{i-1}^{s+1}\cdot(\mathbf{F}_{i-1}\mathbf{x}_{i-1} + \mathbf{M}_{i-1}\mathbf{a}_{i-1})|_{\kappa_{1:s}}, \bar{\mathbf{\Omega}}_{i-1}^{s+1}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\left(\bar{\mathbf{\Omega}}_{i-1}^{s+1}\right)^{T}\right]\times\cdots\\
&\quad\times\mathbf{N}\left[\bar{\mathbf{\Omega}}_{i-1}^{n}\cdot(\mathbf{F}_{i-1}\mathbf{x}_{i-1} + \mathbf{M}_{i-1}\mathbf{a}_{i-1})|_{\kappa_{1:s}}, \bar{\mathbf{\Omega}}_{i-1}^{n}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\left(\bar{\mathbf{\Omega}}_{i-1}^{n}\right)^{T}\right],
\end{aligned} \tag{68}$$

in which, the operator $[\cdot]\big|_{\kappa_{1:s}}$ means taking the corresponding rows, that is from κ_1 to κ_s, of $[\cdot]$.

As a result, Eqs. (14) and (16) should be re-given as (69) and (70).

$$\begin{aligned}
L_{0,k-l}^{k} = &-\frac{n+(l+1)(m+n)}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{P}_{k-l-1}|\\
&-\frac{1}{2}\sum_{i=k-l}^{k}\left\{\log\left|\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\right| + \sum_{t=s+1}^{n}\log\left[\bar{\mathbf{\Omega}}_{i-1}^{t}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\left(\bar{\mathbf{\Gamma}}_{i-1}^{0}\right)^{T}\left(\bar{\mathbf{\Omega}}_{i-1}^{t}\right)^{T}\right] + \log|\mathbf{R}_{i}|\right\}
\end{aligned} \tag{69}$$

$$L_{2,k-l}^{k} = -\frac{1}{2}\sum_{i=k-l}^{k}\Bigg\{\mathbf{D}\Big((\mathbf{x}_i - \mathbf{F}_{i-1}\mathbf{x}_{i-1} - \mathbf{M}_{i-1}\mathbf{a}_{i-1})|_{\kappa_{1:s}}, \bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\Big) + \sum_{t=s+1}^{n}\frac{1}{\bar{\mathbf{\Omega}}_{i-1}^{t}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\big(\bar{\mathbf{\Omega}}_{i-1}^{t}\big)^T}\cdot\Big[\mathbf{x}_i^{\kappa_t} - \bar{\mathbf{\Omega}}_{i-1}^{t}\cdot(\mathbf{F}_{i-1}\mathbf{x}_{i-1} + \mathbf{M}_{i-1}\mathbf{a}_{i-1})|_{\kappa_{1:s}}\Big]^2\Bigg\} \tag{70}$$

where $\mathbf{x}_i^{\kappa_t}$ means the $\kappa_t{}^{th}$ row of $\mathbf{x}_i$; $\bar{\mathbf{\Omega}}_{i-1}^{t}$ means the t^{th} row of $\bar{\mathbf{\Omega}}_{i-1}$.

Accordingly, Eq. (25) should be re-given as (71).

$$\begin{aligned} G_{2,k-l}^{k} = & -\frac{1}{2}Tr\Bigg\{\sum_{i=k-l}^{k}\Big[\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\Big]^{-1}\cdot\Big\{\mathbf{C}\Big[\big(\hat{\mathbf{x}}_i - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1} - \mathbf{M}_{i-1}\mathbf{a}_{i-1}\big)|_{\kappa_{1:s}}\Big] \\ & + \mathbf{P}_{i,i|k-l:k}|_{\kappa_{1:s}} - \big(\mathbf{P}_{i,i-1|k-l:k}\mathbf{F}_{i-1}^T\big)|_{\kappa_{1:s}} - \big(\mathbf{F}_{i-1}\mathbf{P}_{i-1,i|k-l:k}\big)\big|_{\kappa_{1:s}} + \big(\mathbf{F}_{i-1}\mathbf{P}_{i-1,i-1|k-l:k}\mathbf{F}_{i-1}^T\big)|_{\kappa_{1:s}}\Big\}\Bigg\} \\ & -\frac{1}{2}\sum_{i=k-l}^{k}\Bigg\{\sum_{t=s+1}^{n}\frac{\Big[\hat{\mathbf{x}}_i^{\kappa_t} - \bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\big(\mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1} + \mathbf{M}_{i-1}\mathbf{a}_{i-1}\big)|_{\kappa_{1:s}}\Big]^2 + \mathbf{P}_{i-1,\ t}^{\Theta}}{\bar{\mathbf{\Omega}}_{i-1}^{t}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\big(\bar{\mathbf{\Omega}}_{i-1}^{t}\big)^T}\Bigg\}, \end{aligned} \tag{71}$$

where

$$\mathbf{P}_{i-1,\ t}^{\Theta} = E_{\hat{\mathbf{X}}_{k-l}^{k}}\Big[\tilde{\mathbf{x}}_i^{\kappa_t} - \bar{\mathbf{\Omega}}_{i-1}^{t}\cdot(\mathbf{F}_{i-1}\tilde{\mathbf{x}}_{i-1})|_{\kappa_{1:s}}\Big]^2 = E_{\hat{\mathbf{X}}_{k-l}^{k}}\Big[\tilde{\mathbf{x}}_i^{\kappa_t} - \bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\mathbf{F}_{i-1}|_{\kappa_{1:s}}\cdot\tilde{\mathbf{x}}_{i-1}|_{\kappa_{1:s}}\Big]^2 \tag{72}$$

Eq. (71) holds because $E_{\hat{\mathbf{X}}_{k-l}^{k}}[\tilde{\mathbf{x}}_i^{\kappa_t}] = 0$.

Besides, Eq. (35) should be rewritten as (73). Let $\mathbf{D}_1 = \mathbf{0}$, then $\mathbf{a}_{i-1}^{*}$ in Eq. (37) should be (74). It should be noted that, due to $\mathbf{M}_{i-1}$ is of full column rank, thus $\bar{\mathcal{B}}$ is no wonder invertible. That is to say $\mathbf{a}_{i-1}$ can be uniquely determined.

$$\begin{aligned} \mathbf{D}_1 = & \mathop{col}_{i=k-l,\dots,k}\Bigg\{(\mathbf{M}_{i-1}|_{\kappa_{1:s}})^T\cdot\Big[\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\Big]^{-1}\cdot\Big[\big(\hat{\mathbf{x}}_i - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1} - \mathbf{M}_{i-1}\mathbf{a}_{i-1}\big)|_{\kappa_{1:s}}\Big] \\ & + \sum_{t=s+1}^{n}\frac{\Big[\bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\mathbf{M}_{i-1}|_{\kappa_{1:s}}\Big]^T}{\bar{\mathbf{\Omega}}_{i-1}^{t}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\big(\bar{\mathbf{\Omega}}_{i-1}^{t}\big)^T}\cdot\Big[\hat{\mathbf{x}}_i^{\kappa_t} - \bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\big(\mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1} + \mathbf{M}_{i-1}\mathbf{a}_{i-1}\big)|_{\kappa_{1:s}}\Big]\Bigg\} \\ = & \mathop{col}_{i=k-l,\dots,k}\Bigg\{(\mathbf{M}_{i-1}|_{\kappa_{1:s}})^T\cdot\Big[\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\Big]^{-1}\cdot\Big[\big(\hat{\mathbf{x}}_i - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1}\big)|_{\kappa_{1:s}} - \mathbf{M}_{i-1}|_{\kappa_{1:s}}\cdot\mathbf{a}_{i-1}\Big] \\ & + \sum_{t=s+1}^{n}\frac{\Big[\bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\mathbf{M}_{i-1}|_{\kappa_{1:s}}\Big]^T}{\bar{\mathbf{\Omega}}_{i-1}^{t}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\big(\bar{\mathbf{\Omega}}_{i-1}^{t}\big)^T}\cdot\Big[\hat{\mathbf{x}}_i^{\kappa_t} - \bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\big(\mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1}\big)|_{\kappa_{1:s}} - \bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\mathbf{M}_{i-1}|_{\kappa_{1:s}}\cdot\mathbf{a}_{i-1}\Big]\Bigg\} \end{aligned} \tag{73}$$

$$\mathbf{a}_{i-1}^{*} = \bar{\mathcal{B}}^{-1}\cdot\bar{\mathcal{A}}, \tag{74}$$

where

$$\bar{\mathcal{A}} = (\mathbf{M}_{i-1}|_{\kappa_{1:s}})^T\cdot\Big[\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\Big]^{-1}\cdot\Big[\big(\hat{\mathbf{x}}_i - \mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1}\big)|_{\kappa_{1:s}}\Big] + \sum_{t=s+1}^{n}\frac{\Big[\bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\mathbf{M}_{i-1}|_{\kappa_{1:s}}\Big]^T}{\bar{\mathbf{\Omega}}_{i-1}^{t}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\big(\bar{\mathbf{\Omega}}_{i-1}^{t}\big)^T}\cdot\Big[\hat{\mathbf{x}}_i^{\kappa_t} - \bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\big(\mathbf{F}_{i-1}\hat{\mathbf{x}}_{i-1}\big)|_{\kappa_{1:s}}\Big], \tag{75}$$

and

$$\bar{\mathcal{B}} = (\mathbf{M}_{i-1}|_{\kappa_{1:s}})^T\cdot\Big[\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\Big]^{-1}\cdot\mathbf{M}_{i-1}|_{\kappa_{1:s}} + \sum_{t=s+1}^{n}\frac{(\mathbf{M}_{i-1}|_{\kappa_{1:s}})^T\cdot\big(\bar{\mathbf{\Omega}}_{i-1}^{t}\big)^T\cdot\bar{\mathbf{\Omega}}_{i-1}^{t}\cdot\mathbf{M}_{i-1}|_{\kappa_{1:s}}}{\bar{\mathbf{\Omega}}_{i-1}^{t}\bar{\mathbf{\Gamma}}_{i-1}^{0}\bar{\mathbf{Q}}_{i-1}^{0}\big(\bar{\mathbf{\Gamma}}_{i-1}^{0}\big)^T\big(\bar{\mathbf{\Omega}}_{i-1}^{t}\big)^T}. \tag{76}$$

References

[1] S. Sarkka, J. Hartikainen, On gaussian optimal smoothing of non-linear state space models, IEEE Trans. Automat. Contr. 55 (2010) 1938–1941.
[2] U. Gntrkn, J.P. Reilly, T. Kirubarajan, H. deBruin, Recursive hidden input estimation in nonlinear dynamic systems with varying amounts of a priori knowledge, Signal Process. 99 (2014) 171–184.
[3] H. Geng, Z. Wang, Y. Cheng, F.E. Alsaadi, A.M. Dobaie, State estimation under non-gaussian lévy and time-correlated additive sensor noises: a modified tobit kalman filtering approach, Signal Process. 154 (2019) 120–128.
[4] S. Wang, E. Wang, P. Dorato, Observing the states of systems with unmeasurable disturbances, IEEE Trans. Automat. Contr. 20 (1975) 716–717.
[5] H. Lan, Y. Liang, F. Yang, Z. Wang, Q. Pan, Joint estimation and identification for stochastic systems with unknown inputs, IET Control Theory Appl. 7 (2013) 1377–1386.
[6] M. Zhong, J. Guo, Q. Cao, On designing pmi kalman filter for ins/gps integrated systems with unknown sensor errors, IEEE Sens. J. 15 (1) (2015) 535–544.
[7] J. Zhou, Y. Liang, F. Yang, L. Xu, Q. Pan, Multi-sensor consensus estimation of state, sensor biases and unknown input, Sensors 16 (9) (2016) 1407.
[8] H. Lan, A.N. Bishop, Q. Pan, Distributed joint estimation and identification for sensor networks with unknown inputs, in: Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2014 IEEE Ninth International Conference on, IEEE, 2014, pp. 1–6.
[9] H. Geng, Y. Liang, Y. Liu, F.E. Alsaadi, Bias estimation for asynchronous multi--rate multi-sensor fusion with unknown inputs, Inf. Fus. 39 (2018) 139–153.
[10] H. Geng, Y. Liang, L. Xu, Fault detection for multi-rate sensor fusion under multiple uncertainties, IET Control Theory Appl. 9 (11) (2015) 1709–1716.
[11] Y. Liang, D.X. An, D.H. Zhou, Q. Pan, A finite-horizon adaptive kalman filter for linear systems with unknown disturbances, Signal Process. 84 (11) (2004) 2175–2194.
[12] Y. Liang, D. Zhou, L. Zhang, Q. Pan, Adaptive filtering for stochastic systems with generalized disturbance inputs, IEEE Signal Process. Lett. 15 (2008) 645–648.
[13] Y. Liang, D.H. Zhou, Q. Pan, et al., A finite-horizon adaptive kalman filter for linear systems with unknown disturbances, Signal Process. 84 (11) (2004) 2175–2194.
[14] Y. Liang, L. Zhang, D. Zhou, Q. Pan, Estimation of systems with statistically–constrained inputs, Appl. Math. Comput. 217 (6) (2010) 2644–2656.
[15] F.J. Bejarano, Functional unknown input reconstruction of descriptor systems: application to fault detection, Automatica 57 (2015) 145–151.
[16] H. Geng, Y. Liang, F. Yang, L. Xu, Q. Pan, Model-reduced fault detection for multi-rate sensor fusion with unknown inputs, Inf. Fus. 33 (2017) 1–14.
[17] H. Geng, Y. Liang, F. Yang, L. Xu, Q. Pan, The joint optimal filtering and fault detection for multi-rate sensor fusion under unknown inputs, Information Fusion 29 (2016) 57–67.
[18] L. Zhou, Y. Liang, J. Zhou, F. Yang, Q. Pan, Linear minimum mean squared estimation of measurement bias driven by structured unknown inputs, IET Radar, Sonar Navigat. 8 (8) (2014) 977–986.
[19] M. Lungu, R. Lungu, Full-order observer design for linear systems with unknown inputs, Int. J. Control. 85 (10) (2012) 1602–1615.
[20] D. Rotondo, A. Cristofaro, T.A. Johansen, F. Nejjari, V. Puig, State estimation and decoupling of unknown inputs in uncertain lpv systems using interval observers, Int. J. Control. 91 (8) (2018) 1944–1961.
[21] A. Chakrabarty, M.J. Corless, G.T. Buzzard, S.H. Zak, A.E. Rundell, State and unknown input observers for nonlinear systems with bounded exogenous inputs, IEEE Trans. Automat. Contr. 62 (11) (2017) 5497–5510.
[22] A. Chakrabarty, A.E. Rundell, S.H. Zak, F. Zhu, G.T. Buzzard, Unknown input estimation for nonlinear systems using sliding mode observers and smooth window functions, SIAM J. Control Optim. 56 (5) (2018) 3619–3641.
[23] S.S. Delshad, A. Johansson, M. Darouach, T. Gustafsson, Robust state estimation and unknown inputs reconstruction for a class of nonlinear systems: multiobjective approach, Automatica 64 (2016) 1–7.
[24] K.M.D. Motchon, K.M. Pekpe, J.-P. Cassar, Robust discerning controls for the operating modes of linear switched systems subject to bounded unknown inputs, Automatica 96 (2018) 159–165.
[25] X.R. Li, V.P. Jilkov, Survey of maneuvering target tracking. part v. multiple–model methods, IEEE Trans. Aerosp. Electron. Syst. 41 (4) (2005) 1255–1321.
[26] S. Bezzaoucha, H. Voos, M. Darouach, A new polytopic approach for the unknown input functional observer design, Int. J. Control. 91 (3) (2018) 658–677.
[27] J. Yang, Y. Chen, F. Zhu, F. Wang, Simultaneous state and output disturbance estimations for a class of switched linear systems with unknown inputs, Int. J. Syst. Sci. 48 (1) (2017) 22–33.
[28] V. Vu, W. Wang, H. Chen, J.M. Zurada, Unknown input-based observer synthesis for a polynomial t-s fuzzy model system with uncertainties, IEEE Trans. Fuzzy Syst. 26 (3) (2018) 1447–1458.
[29] X. Xie, D. Yue, C. Peng, Multi-instant observer design of discrete-time fuzzy systems: a ranking-based switching approach, IEEE Trans. Fuzzy Syst. 25 (5) (2017) 1281–1292.
[30] Y. Qin, Y. Liang, Y. Yang, Q. Pan, F. Yang, Minimum upper-bound filter of markovian jump linear systems with generalized unknown disturbances, Automatica 73 (2016) 56–63.
[31] C.-S. Hsieh, On the global optimality of unbiased minimum-variance state estimation for systems with unknown inputs, Automatica 46 (4) (2010) 708–715.
[32] M. Darouach, M. Zasadzinski, M. Boutayeb, Extension of minimum variance estimation for systems with unknown inputs, Automatica 39 (5) (2003) 867–876.
[33] Y. Cheng, H. Ye, Y. Wang, D. Zhou, Unbiased minimum-variance state estimation for linear systems with unknown input, Automatica 45 (2) (2009) 485–491.
[34] N. Gao, M. Darouach, H. Voos, M. Alma, New unified h-infinity dynamic observer design for linear systems with unknown inputs, Automatica 65 (2016) 43–52.
[35] H. Lan, X. Wang, Q. Pan, F. Yang, Z. Wang, Y. Liang, A survey on joint tracking using expectation–maximization based techniques, Inf. Fus. 30 (2016) 52–68.
[36] D. Simon, Optimal State Estimation: Kalman, H-Infinity and Nonlinear Approaches, Wiley InterScience, Hoboken, 2006.
[37] P.K. Jain, O.P. Ahuja, K. Ahmad, Functional Analysis, New Age, New Delhi, 2007.
[38] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, Wiley Interscience, Hoboken, 2008.
[39] C.F.J. Wu, On the convergence properties of the em algorithm, Annal. Stat. 11 (1983) 95–108.
[40] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the em algorithm, J. R. Stat. Soc. Ser B (Methodol.) 39 (1977) 1–38.
[41] X.R. Li, V.P. Jilkov, Survey of maneuvering target tracking. part i. dynamic models, IEEE Trans. Aerosp. Electron. Syst. 39 (4) (2003) 1333–1364.
[42] R.A. Singer, Estimating optimal tracking filter performance for manned maneuvering targets, IEEE Trans. Aerosp. Electron. Syst. (4) (1970) 473–483.