

Relations Between Full Information and Kalman-based Estimation

Ming Ge and Eric C. Kerrigan

Abstract—For nonlinear state space systems with additive noises, sometimes the number of process noise signals could be less than the dimension of the state space. In order to improve the accuracy and stability of nonlinear state estimation, this paper provides for the first time the derivation of the full information estimator (FIE) for such nonlinear systems. We verify our derivation of the FIE by firstly proving the unbiasedness and minimum-variance of the FIE for linear time varying (LTV) systems, then showing the equivalence between the Kalman filter/smoothen and the FIE for LTV systems. Finally, we prove that the FIE will provide more accurate state estimates than the extended Kalman filter (EKF) and smoothen (EKS) for nonlinear systems.

I. INTRODUCTION

Given the system dynamics, people often want to forecast future system outputs based on current information. For example, in defense, the arrival time and exact target of an incoming cruise missile have to be predicted, in order to eliminate the threat as early as possible. Before predicting a system's outputs, we have to know the system's internal behavior, the so-called "states", which are usually unknown. If systems are disturbed by noise, which they usually are, then the situation will become more difficult.

The Kalman filter is the most notable and widely used linear state estimation algorithm of the past few decades. Given previous system outputs, the algorithm recursively estimates states in a noise-driven linear dynamical system by minimizing the mean-square error between the true and estimate state. If the complete outputs are available, the state estimation error can be further reduced by the Kalman smoother [12]. An "forward-backward" smoothing algorithm was introduced in [12], where the forward step is a Kalman filter and the backward step determines smoothed state estimates by maximizing the likelihood of outputs given states.

For unconstrained nonlinear systems, the extended Kalman filter (EKF) and smoothen (EKS) estimate state error covariance by linearizing nonlinear dynamics with a first-order Taylor expansion around the current estimate. However, when the error in the higher order terms neglected by the linear (first-order) model are significant, the EKF could exhibit poor convergence characteristics and biased estimates [14]. Moreover, general recursive solutions, such as Kalman filtering, are unavailable if there exists physical limits or algebraic constraints on system states. In order to solve these two challenges, for additive system noises, the

full information estimator (FIE) avoids model linearization by reformulating the original estimation as an optimization problem, which allows for the natural addition of inequality constraints [11].

In [4], the Kalman filter as well as the EKF are considered from a different point of view, where the algorithms can be derived from a single iteration of Newton's method on a certain quadratic form with a judiciously chosen initial guess [4]. This idea is very important and allows us to establish a relationship between Kalman-based algorithms and the FIE, hence we could compare their performance for both linear and nonlinear systems.

For a noise-driven nonlinear system, the number of additive process noises signals can be less than the number of state variables, hence the process noise has to multiply a tall matrix, before being added to the nonlinear dynamics¹. The main problem of applying the FIE to such a system is that the probability density function of the successor state given current state does not exist due to the covariance matrix not being positive definite. The main contribution of this paper is therefore to provide the derivation of the FIE for nonlinear systems with additive noises, where the sizes of both process and output noises are smaller or equal to the size of state and output, respectively.

In order to verify our derivation of the FIE for nonlinear systems with additive noises, we replace the nonlinear dynamics with linear time-varying (LTV) terms and prove the unbiasedness and minimum-variance of the FIE for such LTV systems, followed by several results that show the equivalence between the Kalman filter/smoothen and the FIE for linear unconstrained systems. Finally, we prove that the FIE will provide more accurate state estimates than the EKF and EKS for nonlinear unconstrained systems.

This paper is organized as follows: Section II provides the derivation of the FIE for nonlinear systems with additive noises. Section III shows the relationship between the Kalman based state estimators and the FIE. Finally, we draw conclusions in Section IV.

II. FULL INFORMATION ESTIMATION FOR NONLINEAR SYSTEMS WITH ADDITIVE NOISES

Consider a discrete-time nonlinear state space model:

$$\begin{aligned}x_{k+1} &:= f(x_k) + G_k w_k \\ y_k &:= h(x_k) + H_k v_k\end{aligned}\tag{1}$$

where $x_k \in \mathbb{X}_k$ is the unknown state, $y_k \in \mathbb{R}^p$ is the output measurement. $G_k \in \mathbb{R}^{n \times r}$ and $H_k \in \mathbb{R}^{p \times q}$ are two full

¹An example can be found in [9].

Department of Electrical & Electronic Engineering, Imperial College London, London SW7 2AZ, UK ming.ge08@imperial.ac.uk

Eric C. Kerrigan is with Department of Electrical & Electronic Engineering and Department of Aeronautics, Imperial College London, London SW7 2AZ, UK e.kerrigan@imperial.ac.uk

column rank time-varying matrices². w_k and v_k are two unknown noise terms, which affect the state and output, respectively.

Assumption 1: The noise sequences $(w_k)_{k=1}^M$ and $(v_k)_{k=1}^M$ are two random variables having Gaussian (or normal) distributions $\mathcal{N}(0, Q)$ and $\mathcal{N}(0, R)$, respectively, with zero mean and positive-definite covariance matrices Q and R .

Assumption 2: The probability distribution $p(x_1)$ of the initial state has Gaussian distribution $p(x_1) \sim \mathcal{N}(\tilde{x}_1, P_{1|0})$, where \tilde{x}_1 is the *a priori* most likely value of x_1 and $P_{1|0}$ is the corresponding error covariance.

Assumption 3: Functions $f(\cdot)$ and $h(\cdot)$ are twice differentiable. The discrete-time nonlinear model (1) is uniformly observable [7] and there exists a stable state observer for (1) with nonempty feasible region.

Given measurements $(y_k)_{k=1}^M$ and initial conditions, we would like to find the state x_k that maximizes the conditional probability density $p(x_k|y_{1:M})$. Unfortunately, $p(x_k|y_{1:M})$ is difficult to obtain exactly for a nonlinear model, thus we focus our attention on the entire trajectory of states $(x_k)_{k=1}^M$, rather than just one single state x_k by maximizing the probability density $p(x_{1:M}|y_{1:M})$ of the whole state sequence [13].

Lemma 1: Consider two random vectors $w \in \mathbb{R}^{r \times 1}$ and $x \in \mathbb{R}^{n \times 1}$, $n > r$, where $x = m + Gw$, m is a constant vector. If G is a full column rank matrix, $w \sim \mathcal{N}(0, Q)$ and $Q \succ 0$, then the probability distribution of x is given by

$$p(x) \propto \frac{1}{\sqrt{(2\pi)^r |W|}} e^{-\frac{1}{2} \|S(x-m)\|_{W^{-1}}^2},$$

where W is positive definite matrix and S is a constant matrix.

Proof: Because G is a full column rank matrix, the random vector x has a singular joint normal distribution, such that $p(x) \sim \mathcal{N}(0, GQG^\top)$, where $GQG^\top \succeq 0$. By using the singular value or eigenvalue decomposition, we have

$$GQG^\top = UPU^\top = U \begin{bmatrix} \tilde{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} U^\top,$$

where $\tilde{Q} \in \mathbb{R}^r$ is a nonsingular matrix, U is a unitary matrix, such that $U^{-1} = U^\top$. Define a random vector z , such that

$$z := [z_1^\top \quad z_2^\top]^\top = U^{-1}x = U^{-1}Gw,$$

where $z_1 \in \mathbb{R}^{r \times 1}$. Since $z \sim \mathcal{N}(U^{-1}m, P)$, by definition of the singular joint normal distribution [3, pp. 376-377], we have

$$p(z) = \frac{\delta(z_2 - \mathcal{M}_{r+1}^{n-r,n} U^{-1}m)}{\sqrt{(2\pi)^r |\tilde{Q}|}} e^{-\frac{1}{2} \|z_1 - \mathcal{M}_1^{r,n} U^{-1}m\|_{\tilde{Q}^{-1}}^2},$$

because the Dirac delta function $\delta(z_2 - \mathcal{M}_{r+1}^{n-r,n} U^{-1}m)$ is the probability mass function of the degenerate (deterministic) variable z_2 , we have

$$p(z) \propto \frac{1}{\sqrt{(2\pi)^r |\tilde{Q}|}} e^{-\frac{1}{2} \|z_1 - \mathcal{M}_1^{r,n} U^{-1}m\|_{\tilde{Q}^{-1}}^2}.$$

²This condition is required for uniqueness of the conditional densities $p(x_{k+1}|x_k)$ and $p(y_k|x_k)$ to be defined later.

Using transformation of variables [15, p. 22] gives

$$p(x) = p(z) |\mathcal{J}(U^{-1}x)| \propto \frac{1}{\sqrt{(2\pi)^r |W|}} e^{-\frac{1}{2} \|S(x-m)\|_{W^{-1}}^2},$$

where $\mathcal{J}(\cdot)$ denotes the Jacobian matrix, $W = \tilde{Q}$ and $S = \mathcal{M}_1^{r,n} U^{-1}$. ■

Proposition 1: For nonlinear system (1), given output measurements $(y_k)_{k=1}^M$; if Assumption 1 and 2 hold, G_k and H_k are full column rank matrices, then $p(x_{1:M}|y_{1:M})$ can be written as a function of x_1 , $(w_k)_{k=1}^M$ and $(v_k)_{k=1}^M$.

Proof: From Bayes' theorem we have

$$p(x_{1:M}|y_{1:M}) \propto p(y_{1:M}|x_{1:M})p(x_{1:M}),$$

where

$$p(y_{1:M}|x_{1:M}) = \prod_{k=1}^M p(y_k|x_k),$$

$$p(x_{1:M}) = p(x_1) \prod_{k=1}^{M-1} p(x_{k+1}|x_k).$$

By Assumption 1 and Lemma 1, we have

$$p(y_k|x_k) \propto \frac{1}{\sqrt{(2\pi)^q |\tilde{R}_k|}} e^{-\frac{1}{2} \|\mathcal{M}_1^{q,p} \mathcal{Y}_k^{-1}(y_k - h(x_k))\|_{\tilde{R}_k^{-1}}^2}, \quad (2a)$$

$$p(x_{k+1}|x_k) \propto \frac{1}{\sqrt{(2\pi)^r |\tilde{Q}_k|}} e^{-\frac{1}{2} \|\mathcal{M}_1^{r,n} \mathcal{X}_k^{-1}(x_{k+1} - f(x_k))\|_{\tilde{Q}_k^{-1}}^2}, \quad (2b)$$

where \mathcal{X}_k and \mathcal{Y}_k are unitary matrices, $\tilde{Q}_k \succ 0$ and $\tilde{R}_k \succ 0$,

$$G_k Q G_k^\top = \mathcal{X}_k P_k^x \mathcal{X}_k^\top = \mathcal{X}_k \begin{bmatrix} \tilde{Q}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathcal{X}_k^\top,$$

$$H_k R H_k^\top = \mathcal{Y}_k P_k^y \mathcal{Y}_k^\top = \mathcal{Y}_k \begin{bmatrix} \tilde{R}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathcal{Y}_k^\top.$$

Rearranging $\|\mathcal{M}_1^{r,n} \mathcal{X}_k^{-1}(x_{k+1} - f(x_k))\|_{\tilde{Q}_k^{-1}}^2$ in (2a) gives

$$\begin{aligned} \|\mathcal{M}_1^{r,n} \mathcal{X}_k^{-1}(x_{k+1} - f(x_k))\|_{\tilde{Q}_k^{-1}}^2 &= \|\mathcal{M}_1^{r,n} \mathcal{X}_k^{-1} G_k w_k\|_{\tilde{Q}_k^{-1}}^2 \\ &= w_k^\top G_k^\top \mathcal{X}_k^{-\top} (P_k^x)^\dagger \mathcal{X}_k^{-1} G_k w_k \\ &= w_k^\top G_k^\top (\mathcal{X}_k P_k^x \mathcal{X}_k^\top)^\dagger G_k w_k = w_k^\top G_k^\top (G_k Q G_k^\top)^\dagger G_k w_k \\ &= w_k^\top Q^{-1} w_k, \end{aligned} \quad (3a)$$

where $(\cdot)^\dagger$ denotes the MoorePenrose pseudo-inverse. Similarly for the term $\|\mathcal{M}_1^{q,p} \mathcal{Y}_k^{-1}(y_k - h(x_k))\|_{\tilde{R}_k^{-1}}^2$ in (2b),

$$\|\mathcal{M}_1^{q,p} \mathcal{Y}_k^{-1}(y_k - h(x_k))\|_{\tilde{R}_k^{-1}}^2 = v_k^\top R^{-1} v_k. \quad (3b)$$

Because $|\tilde{Q}_k|$ and $|\tilde{R}_k|$ can be written as a linear function of $|Q|$ and $|R|$, respectively. Substituting (3) into (2) gives

$$p(x_{k+1}|x_k) \propto \frac{1}{\sqrt{(2\pi)^r |Q|}} e^{-\frac{1}{2} \|w_k\|_{Q^{-1}}^2},$$

$$p(y_k|x_k) \propto \frac{1}{\sqrt{(2\pi)^q |R|}} e^{-\frac{1}{2} \|v_k\|_{R^{-1}}^2}.$$

Hence, we have

$$p(y_{1:M}|x_{1:M}) \propto \prod_{k=1}^M \frac{1}{\sqrt{(2\pi)^q |R|}} e^{-\frac{1}{2} v_k^\top R^{-1} v_k}$$

$$p(x_{1:M}) \propto p(x_1) \prod_{k=1}^{M-1} \frac{1}{\sqrt{(2\pi)^r |Q|}} e^{-\frac{1}{2} w_k^\top Q^{-1} w_k}.$$

Therefore, $p(x_{1:M}|y_{1:M})$ can be written as a function of x_1 , $(w_k)_{k=1}^M$ and $(v_k)_{k=1}^M$. ■

By taking the logarithm on both sides of $p(x_{1:M}|y_{1:M})$ and removing the constant terms, yields the FIE for (1) [13]

$$\min_{\mathcal{X}_{1,M}} \frac{1}{2} \|w_0\|_{P_{1|0}}^2 + \frac{1}{2} \sum_{k=1}^{M-1} \|w_k\|_{Q^{-1}}^2 + \frac{1}{2} \sum_{k=1}^M \|v_k\|_{R^{-1}}^2$$

$$\text{s.t. } x_{k+1} = f(x_k) + G_k w_k, \quad k = 1, \dots, M-1$$

$$y_k = h(x_k) + H_k v_k, \quad k = 1, \dots, M \quad (4)$$

where the decision variables and their optimal solutions are

$$\mathcal{X}_{1,M} := [x_1^\top \ w_1^\top \ v_1^\top \ x_2^\top \ \dots \ x_{M-1}^\top \ w_{M-1}^\top \ v_{M-1}^\top \ x_M^\top]^\top,$$

$$\mathcal{X}_{1,M}^* := [\hat{x}_{1|M}^\top \ \hat{w}_{1|M}^\top \ \hat{v}_{1|M}^\top \ \hat{x}_{2|M}^\top \ \dots \ \hat{w}_{M-1}^\top \ \hat{v}_{M-1}^\top \ \hat{x}_M^\top]^\top,$$

$w_0 := x_1 - \tilde{x}_1$, $(\hat{x}_{k|M})_{k=1}^M$, $(\hat{w}_{k|M})_{k=1}^M$ and $(\hat{v}_{k|M})_{k=1}^M$ are the estimates of the state and system noise sequence $(x_k)_{k=1}^M$, $(w_k)_{k=0}^M$ and $(v_k)_{k=1}^M$, respectively, with $(\hat{\cdot})_M := (\hat{\cdot})_{M|M}$, given output sequence $y_{1:M}$.

III. RELATIONS BETWEEN KALMAN-BASED STATE ESTIMATION AND FULL INFORMATION ESTIMATION

Theorem 1: Consider the LTV system

$$x_{k+1} = A_k x_k + G_k w_k$$

$$y_k = C_k x_k + H_k v_k \quad (5)$$

where A_k and C_k are full rank dynamics and sensor matrices, G_k and H_k are full column rank matrices, given output measurements $(y_k)_{k=1}^M$, $1 < k \leq M$. If Assumptions 1 and 2 hold, then the FIE (4) is a linear unbiased estimator [6, p. 11] of the state sequence $(x_k)_{k=1}^M$ regardless of the choice of covariances P_1 , Q and R ; if covariances are accurate, then the FIE is a best linear unbiased estimator (BLUE) [6, p. 555] of the state sequence $(x_k)_{k=1}^M$.

Proof: If both G_k and H_k are invertible matrices, then (4) becomes a weighted linear least squares problem and a proof of statement that the FIE is a unbiased estimator and BLUE can found in [1]. However, if both G_k and H_k are not invertible, a different approach is required, as presented here.

By involving Lagrange multipliers $(\alpha_k)_{k=0}^{\bar{k}-1}$ and $(\beta_k)_{k=1}^{\bar{k}}$ for equality constraints and the initial condition, respectively, the FIE problem (4) can be written as the following unconstrained optimization problem:

$$Z_k^* := \arg \min_{Z_k} S_k, \quad (6)$$

where

$$S_1 := \frac{1}{2} \|w_0\|_{P_{1|0}}^2 + \gamma_1^\top (\tilde{x}_1 - x_1 - w_0) + \frac{1}{2} \|v_1\|_{R^{-1}}^2$$

$$+ \beta_1^\top (y_1 - C_1 x_1 - H_1 v_1),$$

$$S_{\bar{k}} := S_{\bar{k}-1} + \frac{1}{2} \|v_{\bar{k}}\|_{R^{-1}}^2 + \beta_{\bar{k}}^\top (y_{\bar{k}} - C_{\bar{k}} x_{\bar{k}} - H_{\bar{k}} v_{\bar{k}}) +$$

$$\frac{1}{2} \|w_{\bar{k}-1}\|_{Q^{-1}}^2 + \alpha_{\bar{k}-1}^\top (x_{\bar{k}} - A_{\bar{k}-1} x_{\bar{k}-1} - G_{\bar{k}-1} w_{\bar{k}-1}),$$

$$Z_{\bar{k}} := [w_0^\top \ v_1^\top \ \alpha_0^\top \ \beta_1^\top \ x_1^\top \ w_1^\top \ v_2^\top \ \alpha_1^\top \ \dots \ \beta_{\bar{k}}^\top \ x_{\bar{k}}^\top]^\top,$$

$$Z_{\bar{k}}^* := [\hat{w}_{0|\bar{k}}^\top \ \hat{v}_{1|\bar{k}}^\top \ \hat{\alpha}_{0|\bar{k}}^\top \ \hat{\beta}_{1|\bar{k}}^\top \ \hat{x}_{1|\bar{k}}^\top \ \hat{w}_{1|\bar{k}}^\top \ \hat{v}_{2|\bar{k}}^\top \ \dots \ \hat{\beta}_{\bar{k}}^\top \ \hat{x}_{\bar{k}}^\top]^\top.$$

The necessary condition of $Z_{\bar{k}}^*$ being an optimum of (6) is

$$\frac{\partial S_{\bar{k}}(Z_{\bar{k}})}{\partial Z_{\bar{k}}} = \mathcal{J}_{S_{\bar{k}}(Z_{\bar{k}})} = \mathcal{H}_{S_{\bar{k}}} Z_{\bar{k}} - b_{S_{\bar{k}}} = 0, \quad (7)$$

where the Hessian matrix $\mathcal{H}_{S_{\bar{k}}}$ is defined by

$$\mathcal{H}_{S_{\bar{k}}} := \begin{bmatrix} \mathcal{H}_{S_{\bar{k}-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{N_0} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{N_{\bar{k}-1}-n} & \mathbf{0} \\ \mathbf{0} & \mathcal{H}_{S_{\bar{k}-1}}^+ \end{bmatrix}.$$

$N_0 := r + q + 2n + p$, $N_{\bar{k}-1}$ is the number of rows/columns of the previous Hessian matrix $\mathcal{H}_{S_{\bar{k}-1}}$; matrices \mathcal{H}_{S_1} and $\mathcal{H}_{S_{\bar{k}-1}}^+$ are given by

$$\mathcal{H}_{S_1} := \begin{bmatrix} P_{1|0}^{-1} & \mathbf{0} & -I_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & R^{-1} & \mathbf{0} & -H_1^\top & \mathbf{0} \\ -I_n & \mathbf{0} & \mathbf{0} & \mathbf{0} & -I_n \\ \mathbf{0} & -H_1 & \mathbf{0} & \mathbf{0} & -C_1 \\ \mathbf{0} & \mathbf{0} & -I_n & -C_1^\top & \mathbf{0} \end{bmatrix}$$

and $\mathcal{H}_{S_{\bar{k}-1}}^+ :=$

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -A_{\bar{k}-1}^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Q^{-1} & \mathbf{0} & -G_{\bar{k}-1}^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & R^{-1} & \mathbf{0} & -H_{\bar{k}}^\top & \mathbf{0} \\ -A_{\bar{k}-1} & -G_{\bar{k}-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & I_n \\ \mathbf{0} & \mathbf{0} & -H_{\bar{k}} & \mathbf{0} & \mathbf{0} & -C_{\bar{k}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I_n & -C_{\bar{k}}^\top & \mathbf{0} \end{bmatrix},$$

respectively; the constant vector $b_{S_{\bar{k}}}$ is defined by

$$b_{S_{\bar{k}}} := [\mathbf{0}_{1,n+q} - \tilde{x}_1^\top - y_1^\top \ \mathbf{0}_{1,n} \ \dots \ -y_{\bar{k}}^\top \ \mathbf{0}_{1,n}]^\top$$

$$= \mathcal{H}_{S_{\bar{k}}} Z_{\bar{k}} = \mathcal{H}_{S_{\bar{k}}} Z_{\bar{k}}^x + \mathcal{B}_{S_{\bar{k}}} Z_{\bar{k}}^d,$$

where $\mathcal{B}_{S_{\bar{k}}} :=$

$$\mathcal{H}_{S_{\bar{k}}} \times \text{diag} \left(\begin{bmatrix} I_{n+q} \\ \mathbf{0}_{2n+p,n+q} \end{bmatrix}, I_{\bar{k}-1} \otimes \begin{bmatrix} I_{r+q} \\ \mathbf{0}_{2n+p,q+r} \end{bmatrix} \right)$$

and

$$Z_{\bar{k}}^d := [w_0^\top \ v_1^\top \ w_1^\top \ v_2^\top \ w_2^\top \ \dots \ w_{\bar{k}-1}^\top \ v_{\bar{k}}^\top]^\top,$$

$$Z_{\bar{k}}^x := [\mathbf{0} \ \alpha_0^\top \ \beta_1^\top \ x_1^\top \ \dots \ \mathbf{0} \ \alpha_{\bar{k}-1}^\top \ \beta_{\bar{k}}^\top \ x_{\bar{k}}^\top]^\top.$$

Because A_k and C_k are full rank matrices, $\mathcal{H}_{S_{\bar{k}}}$ is non-singular³, $Z_{\bar{k}}^*$ can be uniquely determined using (7), such that

$$Z_{\bar{k}}^* = \mathcal{H}_{S_{\bar{k}}}^{-1} b_{S_{\bar{k}}}. \quad (8)$$

³The non-singularity can be proved by recursively showing that the determinant of $\{\mathcal{H}_{S_k}\}_{k=1}^{\bar{k}}$ is non-zero using [2, Fact 2.14.9].

Taking the expectation on both sides yields

$$\mathcal{H}_{S_k}^{-1} \mathbb{E} \{ \mathcal{H}_{S_k} Z_k^x + \mathcal{B}_{S_k} Z_k^d \} = Z_k^x.$$

Now, if we build the matrix \mathcal{H}_{S_k} based on inaccurate $P_{1|0}$, Q and R , and called it $\tilde{\mathcal{H}}_{S_k}$, $\tilde{P}_{1|0}$, \tilde{Q} and \tilde{R} , respectively, then there exists a matrix D_H such that $D_H := \mathcal{H}_{S_k} - \tilde{\mathcal{H}}_{S_k}$, hence we have

$$\tilde{Z}_k^* = \tilde{\mathcal{H}}_{S_k}^{-1} b_{S_k}.$$

Taking the expectation on both sides yields

$$\tilde{\mathcal{H}}_{S_k}^{-1} \left(\tilde{\mathcal{H}}_{S_k} + D_H \right) Z_k^x = Z_k^x + \tilde{\mathcal{H}}_{S_k}^{-1} D_H Z_k^x = Z_k^x.$$

Since $(x_k)_{k=1}^{\bar{k}}$ are parts of vector Z_k^x , the FIE is an unbiased linear estimator of the state sequence $(x_k)_{k=1}^{\bar{k}}$, regardless the choice of covariances $P_{1|0}$, Q and R . The covariance of Z_k^* is given by

$$\mathbb{C}\{Z_k^*\} = \mathcal{H}_{S_k}^{-1} \mathcal{B}_{S_k} \Sigma \mathcal{B}_{S_k}^\top \mathcal{H}_{S_k}^{-\top},$$

where $\Sigma = \text{diag} (P_{1|0}, R, I_{k-1} \otimes \text{diag} (Q, R))$. There exists a matrix $D_{H^{-1}} := \mathcal{H}_{S_k}^{-1} - \tilde{\mathcal{H}}_{S_k}^{-1}$, hence we have

$$\begin{aligned} \mathbb{C}\{\tilde{Z}_k^*\} &= \\ & \left(\mathcal{H}_{S_k}^{-1} + D_{H^{-1}} \right) \mathcal{B}_{S_k} \Sigma \mathcal{B}_{S_k}^\top \left(\mathcal{H}_{S_k}^{-1} + D_{H^{-1}} \right)^\top = \mathbb{C}\{Z_k^*\} \\ & + 2\mathcal{H}_{S_k}^{-1} \mathcal{B}_{S_k} \Sigma \mathcal{B}_{S_k}^\top D_{H^{-1}}^\top + D_{H^{-1}} \mathcal{B}_{S_k} \Sigma \mathcal{B}_{S_k}^\top D_{H^{-1}}. \end{aligned}$$

Since $D_{H^{-1}} \mathcal{B}_{S_k} \Sigma \mathcal{B}_{S_k}^\top D_{H^{-1}}^\top \succeq 0$ and

$$\begin{aligned} & \mathcal{H}_{S_k}^{-1} \mathcal{B}_{S_k} \Sigma \mathcal{B}_{S_k}^\top \\ & = \text{diag} \left(\begin{bmatrix} \mathcal{E} \\ \mathbf{0}_{2n+p, 3n+p+q} \end{bmatrix}, \bigoplus_{k=2}^{\bar{k}} \begin{bmatrix} \mathcal{E}_k \\ \mathbf{0}_{2n+p, N_0} \end{bmatrix} \right), \end{aligned}$$

where $\bigoplus_{k=1}^N G_k := \text{diag} (G_1, \dots, G_N)$ and

$$\begin{aligned} \mathcal{E} &:= \begin{bmatrix} I_n & \mathbf{0} & -P_{1|0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_p & \mathbf{0} & -H_1 R & \mathbf{0} \end{bmatrix}, \\ \mathcal{E}_k &:= \begin{bmatrix} I_n & \mathbf{0} & -G_{k-1} Q & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_p & \mathbf{0} & -H_k R & \mathbf{0} \end{bmatrix}, \end{aligned}$$

covariance of $(x_k)_{k=1}^{\bar{k}}$ is minimized if $D_{H^{-1}} = 0$. Thus, the FIE is a best linear unbiased estimator (BLUE) of the state sequence $(x_k)_{k=1}^{\bar{k}}$, if $P_{1|0}$, Q and R are all accurate. ■

Proposition 2: For the LTV system (5), given output measurements $(y_k)_{k=1}^{\bar{k}}$, $1 < \bar{k} \leq M$, if Assumptions 1 and 2 hold, then the FIE (4) and Kalman filter [5] are equivalent methods for estimating the current state $\hat{x}_{\bar{k}}$.

Proof: The proof should be straightforward by slightly modify the results given in [4]. ■

Proposition 3: For the LTV system (5), given output measurements $(y_k)_{k=1}^M$ and all estimated states $(\hat{x}_k)_{k=1}^M$, if Assumptions 1 and 2 hold, then the FIE and Kalman smoother [12] are equivalent methods for estimating the smoothed state $\hat{x}_{k|M}$, $1 < \bar{k} < M$.

Proof: Because $y_{1:M}$ and $\hat{x}_{1:M}$ are all given, the smoothed state $\hat{x}_{\bar{k}|M}$ can be estimated by maximizing $p(x_{\bar{k}:M} | y_{1:M})$, such that [12].

$$\begin{aligned} p(x_{\bar{k}:M} | y_{1:M}) &\propto p(x_{\bar{k}+1:M}, y_{\bar{k}+1:M} | x_{\bar{k}}, y_{1:\bar{k}}) p(x_{\bar{k}} | y_{1:\bar{k}}), \\ &\propto p(x_{\bar{k}+1:M}, y_{\bar{k}+1:M} | x_{\bar{k}}) p(x_{\bar{k}} | y_{1:\bar{k}}), \\ &\propto p(y_{\bar{k}+1:M} | x_{\bar{k}+1:M}) p(x_{\bar{k}+1:M} | x_{\bar{k}}) p(x_{\bar{k}} | y_{1:\bar{k}}), \end{aligned}$$

hence, $\hat{x}_{M-1|M}$ can be estimated by

$$\begin{aligned} & \min_{x_{M-1}, v_M} \|v_M\|_{R^{-1}}^2 + \|w_{M-1}\|_{Q^{-1}}^2 + \|x_{M-1} - \hat{x}_{M-1}\|_{P_{M-1}^{-1}}^2 \\ & \text{s.t. } x_M = A_{M-1} x_{M-1} + G_{M-1} w_{M-1}, \\ & y_M = C_M x_M + H_k v_M, \end{aligned} \quad (9)$$

By involving Lagrange multipliers α_{M-1} and β_M for equality constraints, (9) can be written as the following unconstrained optimization problem:

$$\begin{aligned} Z_{M-1, M}^* &:= \arg \min_{Z_{M-1, M}} \|x_{M-1} - \hat{x}_{M-1}\|_{P_{M-1}^{-1}}^2 + \|v_M\|_{R^{-1}}^2 \\ & + \beta_M^\top [y_M - C_M x_M - H_M v_M] + \|w_{M-1}\|_{Q^{-1}}^2 \\ & + \alpha_{M-1}^\top [x_M - A_{M-1} x_{M-1} - G_{M-1} w_{M-1}]. \end{aligned}$$

$$\begin{aligned} \text{where } Z_{M-1, M} &:= [x_{M-1}^\top \ w_{M-1}^\top \ v_M^\top \ \alpha_{M-1}^\top \ \beta_M^\top \ x_M^\top]^\top, \\ Z_{M-1, M}^* &:= [\hat{x}_{M-1|M}^\top \ \hat{w}_{M-1|M}^\top \ \hat{v}_M^\top \ \hat{\alpha}_{M-1|M}^\top \ \hat{\beta}_M^\top \ \hat{x}_M^\top]^\top. \end{aligned}$$

Since all estimated states $(\hat{x}_k)_{k=1}^M$ are given, v_M is deterministic and β_M becomes a free variable, hence we have

$$\begin{aligned} & \min_{Z_{M-1|M}} \tilde{S}_{M-1|M} = \\ & \min_{Z_{M-1|M}} \|x_{M-1} - \hat{x}_{M-1}\|_{P_{M-1}^{-1}}^2 + \left\{ \|w_{M-1}\|_{Q^{-1}}^2 \right. \\ & \left. + \alpha_{M-1}^\top [\hat{x}_M - A_{M-1} x_{M-1} - G_{M-1} w_{M-1}] \right\}, \quad (10) \end{aligned}$$

where $Z_{M-1|M} = [x_{M-1}^\top \ w_{M-1}^\top \ \alpha_{M-1}^\top]^\top$, $Z_{M-1|M}^* =: [\hat{x}_{M-1|M}^\top \ \hat{w}_{M-1|M}^\top \ \hat{\alpha}_{M-1|M}^\top]^\top$. The necessary condition of $Z_{M-1|M}^*$ being an optimum of (9) is that

$$\frac{\partial S_{M-1|M}(Z_{M-1|M})}{\partial Z_{M-1|M}} = \mathcal{H}_{S_{M-1|M}} Z_{M-1|M} - b_{S_{M-1|M}} = 0,$$

where the Hessian matrix $\mathcal{H}_{S_{M-1|M}}$ and constant vector $b_{S_{M-1|M}}$ are defined by

$$\mathcal{H}_{S_{M-1|M}} := \begin{bmatrix} P_{M-1}^{-1} & \mathbf{0} & -A_{M-1}^\top \\ \mathbf{0} & Q^{-1} & -G_{M-1}^\top \\ -A_{M-1} & -G_{M-1} & \mathbf{0} \end{bmatrix}$$

and

$$b_{S_{M-1|M}} := [\hat{x}_{M-1}^\top P_{M-1}^{-1} \ \mathbf{0} \ -\hat{x}_M^\top]^\top,$$

respectively. Since P_{M-1}^{-1} is nonsingular, by using [2, Fact 2.17.3], we have

$$\begin{aligned} \hat{x}_{M-1|M} &= \mathcal{M}_1^{n, 2n+r} \mathcal{H}_{S_{M-1|M}}^{-1} b_{S_{M-1|M}} \\ &= \hat{x}_{M-1} + P_{M-1} A_{M-1}^\top P_{M-1}^{-1} (\hat{x}_M - A_{M-1} \hat{x}_{M-1}). \end{aligned}$$

Now, for estimating $\hat{x}_{M-2|M}$, we have

$$\begin{aligned} & \min_{Z_{M-2,M}} \|x_{M-2} - \hat{x}_{M-2}\|_{P_{M-2}^{-1}}^2 \\ & + \sum_{k=M-1}^M [\|v_k\|_{R^{-1}}^2 + \beta_k^\top (y_k - C_k x_k - H_k v_k)] \\ & + \sum_{k=M-2}^{M-1} [\|w_k\|_{Q^{-1}}^2 + \alpha_k^\top (x_{k+1} - A_k x_k - G_k w_k)]. \end{aligned}$$

Since $\hat{x}_{M-1|M}$, $(\hat{x}_k)_{k=1}^M$ and $(P_k)_{k=1}^M$ are given, $(v_k)_{k=M-1}^M$ and w_{M-1} are deterministic, $(\beta_k)_{k=M-1}^M$ and α_{M-1} are free variable, hence we have

$$\begin{aligned} & \min_{Z_{M-2|M}} \|x_{M-2} - \hat{x}_{M-2}\|_{P_{M-2}^{-1}}^2 + [\|w_{M-2}\|_{Q^{-1}}^2 \\ & + \alpha_{M-2}^\top (\hat{x}_{M-1|M} - A_{M-2} x_{M-2} - G_{M-2} w_{M-2})], \end{aligned}$$

and

$$\hat{x}_{M-2|M} = \hat{x}_{M-2} + U_{M-2} (\hat{x}_{M-1|M} - A_{M-2} \hat{x}_{M-2}),$$

where $U_{M-2} := P_{M-2} A_{M-2}^\top P_{M-1|M-2}^{-1}$. Therefore, by induction, we have

$$\hat{x}_{k|M} = \hat{x}_k + P_k A_k^\top P_{k+1|k}^{-1} (\hat{x}_{k+1|M} - A_k \hat{x}_k),$$

which is the same as the Kalman smoother [12]. ■

Proposition 4: For the nonlinear system (1), where G_k and H_k are full column rank matrices, given output measurements $(y_k)_{k=1}^{\bar{k}}$, $1 < \bar{k} \leq M$, if Assumptions 1 to 3 hold, then state estimation using the FIE (4) will have a smaller estimation error $e_{\bar{k}}$ compared to the EKF.

Proof: In [4], it was proven that the EFK is equivalent to solving the FIE problem (4) by only one single Newton step with a carefully chosen initial guess; hence, by optimality, the state estimation using the FIE will have smaller estimation error $e_{\bar{k}}$ than the EKF. ■

Proposition 5: For the nonlinear system (1), given output measurements $(y_k)_{k=1}^M$, all estimated states $(\hat{x}_k)_{k=1}^M$ and noise covariances, if Assumptions 1 to 3 hold, then the state estimate using the FIE (4) will have smaller error $e_{k|M} := \|x_k - \hat{x}_{k|M}\|$ compared to the extended Kalman smoother (EKS), such that

$$\hat{x}_{k|M} = \hat{x}_k + P_k A_k^\top P_{k+1|k}^{-1} (\hat{x}_{k+1|M} - f(\hat{x}_k)),$$

where $A_k := \left. \frac{\partial f(\cdot)}{\partial x_k} \right|_{x_k = \hat{x}_k, w_k = 0}$.

Proof: Similarly to Proposition 3, $\hat{x}_{M-1|M}$ can be estimated by

$$\begin{aligned} & \min_{X_{M-1,M}} \|v_M\|_{R^{-1}}^2 + \|w_{M-1}\|_{Q^{-1}}^2 + \|x_{M-1} - \hat{x}_{M-1}\|_{P_{M-1}^{-1}}^2 \\ & \text{s.t. } x_M = f(x_{M-1}) + G_{M-1} w_{M-1}, \\ & y_M = h(x_M) + H_M v_M. \end{aligned} \quad (11)$$

By involving Lagrange multipliers α_{M-1} and β_M for equality constraints, (11) can be written as the following unconstrained optimization problem:

$$\begin{aligned} Z_{M-1,M}^* & := \arg \min_{Z_{M-1,M}} \|x_{M-1} - \hat{x}_{M-1}\|_{P_{M-1}^{-1}}^2 \\ & + \|v_M\|_{R^{-1}}^2 + \beta_M^\top [y_M - h(x_M) - H_M v_M] \\ & + \|w_{M-1}\|_{Q^{-1}}^2 + \alpha_{M-1}^\top [x_M - f(x_{M-1}) - G_{M-1} w_{M-1}]. \end{aligned}$$

where $Z_{M-1,M} = [x_{M-1}^\top \ w_{M-1}^\top \ v_M^\top \ \alpha_{M-1}^\top \ \beta_M^\top \ x_M^\top]^\top$. Since all estimated states $(\hat{x}_k)_{k=1}^M$ are given, v_M is deterministic and β_M becomes a free variable, hence we have

$$\begin{aligned} & \min_{Z_{M-1|M}} \tilde{S}_{M-1|M} = \\ & \min_{Z_{M-1|M}} \frac{1}{2} \|x_{M-1} - \hat{x}_{M-1}\|_{P_{M-1}^{-1}}^2 + \left\{ \frac{1}{2} \|w_{M-1}\|_{Q^{-1}}^2 \right. \\ & \left. + \alpha_{M-1}^\top [\hat{x}_M - f(x_{M-1}) - G_{M-1} w_{M-1}] \right\}. \end{aligned} \quad (12)$$

Because (12) is an unconstrained nonlinear optimization problem, in order to ensure the convergence of (12), one could use a backtracking line search combined with Newton's method. Moreover, for a non-convex optimization problem like (12), a pure Newton method is not guaranteed to produce a descent direction when the current iterate is not close to the solution [8, p. 31]. One has to keep monitoring the positive definiteness of the Hessian $\mathcal{H}_{\tilde{S}_{M-1|M}}$ and modify the Hessian when necessary [8, pp. 49–56]. Finally, the optimum⁴ of (12) can be found by recursively calculating

$$Z_{M-1|M}^{(i+1)} = Z_{M-1|M}^{(i)} - a_i \mathcal{H}_{\tilde{S}_{M-1|M}}^{-1} \mathcal{J}_{\tilde{S}_{M-1|M}},$$

given initial guess $Z_{M-1|M}^{(1)}$, until $Z_{M-1|M}^{(i+1)}$ has converged to its optimal value $Z_{M-1|M}^*$, where the scalar a_i is called step length, which is determined using a backtracking line search method, with initial value $a_1 = 1$ [8, p. 37]. The expressions for $\mathcal{J}_{\tilde{S}_{M-1|M}}$ and $\mathcal{H}_{\tilde{S}_{M-1|M}}$ are given by

$$\begin{aligned} \mathcal{J}_{\tilde{S}_{M-1|M}} & := \begin{bmatrix} P_{M-1}^{-1} (x_{M-1} - \hat{x}_{M-1}) - \frac{\partial f^\top}{\partial x_{M-1}} \alpha_{M-1} \\ Q^{-1} w_{M-1} - G_{M-1} \alpha_{M-1} \\ \hat{x}_M - f(x_{M-1}) - G_{M-1} w_{M-1} \end{bmatrix}, \\ \mathcal{H}_{\tilde{S}_{M-1|M}} & \approx \begin{bmatrix} P_{M-1}^{-1} & \mathbf{0} & -\frac{\partial f^\top}{\partial x_{M-1}} \\ \mathbf{0} & Q^{-1} & -G_{M-1}^\top \\ -\frac{\partial f}{\partial x_{M-1}} & -G_{M-1} & \mathbf{0} \end{bmatrix}, \end{aligned}$$

where all the second derivatives of $f(\cdot)$ are ignored.

Since the system dynamics are nonlinear, the global optimum Z_k^* is not guaranteed to be found and line search method usually takes several iterations to find a local optimum. Let us pick an initial guess of the optimum, such that $Z_{M-1|M}^{(1)} := [\hat{x}_{M-1|M}^{(1)\top} \ \mathbf{0} \ \mathbf{0}]^\top$, where $\hat{x}_{M-1|M}^{(1)} := \hat{x}_{M-1}$.

⁴In order to prevent converging to a stationary point rather than a minimizer, negative curvature information from the Hessian $\mathcal{H}_{\tilde{S}_{M-1|M}}$ may be required [8, p. 40].

Hence,

$$Z_{M-1|M}^{(2)} = \begin{bmatrix} \hat{x}_{M-1}^\top \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} P_{M-1}^{-1} & \mathbf{0} & -A_{M-1}^\top \\ \mathbf{0} & Q^{-1} & -G_{M-1}^\top \\ -A_{M-1} & -G_{M-1} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \hat{x}_M^- \end{bmatrix},$$

where $\hat{x}_M^- := \hat{x}_M - f(\hat{x}_{M-1})$. Hence we have

$$\hat{x}_{M-1|M}^{(2)} = \hat{x}_{M-1|M}^{(1)} + P_{M-1} A_{M-1}^\top P_{M-1}^{-1} \hat{x}_M^-.$$

It is clear to see that for the smoothed estimate $\hat{x}_{M-1|M}$, the EKS is equivalent to solving (11) by only one line search step with a carefully chosen initial guess $Z_{M-1|M}^{(1)}$; hence, by optimality, the optimum $\hat{x}_{M-1|M}$ obtained by solving (11) will have a smaller $e_{M-1|M}$ than the EKS.

Therefore, by induction, the EKS is solving

$$\begin{aligned} \min_{X_{1,M}} \|\varepsilon_1\|_{P_1^{-1}}^2 + \sum_{k=1}^M \|v_k\|_{R^{-1}}^2 + \sum_{k=1}^{M-1} \|w_k\|_{Q^{-1}}^2 \\ \text{s.t. } x_{k+1} = f(x_k) + G_k w_k, \quad k = 1, \dots, M-1 \\ y_k = h(x_k) + H_k v_k, \quad k = 1, \dots, M \end{aligned}$$

by single backtracking step, given $(\hat{x}_{k|M})_{k=k_s+1}^M$, $(\hat{x}_k)_{k=1}^M$ and a carefully chosen initial guess

$$Z_{1|M}^{(1)} := [\mathbf{0} \quad \hat{x}_1^\top \quad \mathbf{0} \quad \hat{x}_2^\top \quad \dots \quad \mathbf{0} \quad \hat{x}_M^\top]^\top.$$

Hence, by optimality, the state estimate using the FIE will have a smaller error $e_{k|M}$ than the EKS. ■

IV. CONCLUSIONS

In this paper, we used the definition of singular joint probability density to prove that the joint probability density function of a full column rank matrix G multiplying with a normal distributed random vector w is directly proportional to the joint probability density function of that random vector w ; hence provided the derivation of the FIE for nonlinear systems with additive noises, where G_k and H_k are full column rank matrices. We compared the FIE with Kalman based algorithms for both linear and nonlinear systems. We proved that, for a linear system, the FIE is a BLUE of the unknown system states, if initial and noise statistics are accurate. The FIE and Kalman filter/smoothing are completely equivalent for a linear system and FIE will provide better estimates for nonlinear systems compared to both the EKF and EKS. Future work could involve deriving the FIE for the nonlinear system

$$\begin{aligned} x_{k+1} &:= f(x_k, w_k), \\ y_k &:= h(x_k, v_k). \end{aligned}$$

Because noise terms w_k and v_k are not additive to the nonlinear dynamics, evaluating the state and output transition probability density functions $p(x_{1:M})$ and $p(y_{1:M}|x_{1:M})$, respectively, will become difficult, which requires the solution of a functional difference equation, namely the discrete-time analog of the Fokker-Planck equation [10]. Alternatively,

$p(x_{k+1}|x_k)$ and $p(y_k|x_k)$ can be approximated by first order Taylor series:

$$f(x_k, 0) = f(\hat{x}_k, 0) + A_k(x_k - \hat{x}_k) + e_{f_0}, \quad (13a)$$

$$h(x_k, 0) = h(\hat{x}_k, 0) + C_k(x_k - \hat{x}_k) + e_{h_0} \quad (13b)$$

and

$$f(x_k, w_k) = f(\hat{x}_k, 0) + A_k(x_k - \hat{x}_k) + G_k w_k + e_f, \quad (14a)$$

$$h(x_k, v_k) = h(\hat{x}_k, 0) + C_k(x_k - \hat{x}_k) + H_k v_k + e_h, \quad (14b)$$

where $C_k := \left. \frac{\partial h(\cdot)}{\partial x_k} \right|_{x_k=\hat{x}_k, v_k=0}$, $G_k := \left. \frac{\partial f(\cdot)}{\partial w_k} \right|_{x_k=\hat{x}_k, v_k=0}$ and $H_k := \left. \frac{\partial h(\cdot)}{\partial v_k} \right|_{x_k=\hat{x}_k, v_k=0}$. Substituting (13) to (14) gives

$$f(x_k, w_k) = f(x_k, 0) + G_k w_k + e_f - e_{f_0},$$

$$h(x_k, v_k) = h(x_k, 0) + H_k v_k + e_h - e_{h_0}.$$

By ignoring all error terms, if G_k and H_k are both full column rank matrices, then

$$f(x_k, w_k) \approx f(x_k, 0) + G_k w_k,$$

$$h(x_k, v_k) \approx h(x_k, 0) + H_k v_k.$$

So that $p(x_{1:M})$ and $p(y_{1:M}|x_{1:M})$ can be approximated using Proposition 1.

REFERENCES

- [1] A. C. Aitken. On the least squares and linear combinations of observations. *Proceedings of Royal Society Edinburgh*, 55:42–48, 1935.
- [2] Dennis S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory (2nd Edition)*. Princeton University Press, 41 William Street, Princeton, New Jersey, U.S., 2009.
- [3] Michael D. Graham and James B. Rawlings. *Modelling and Analysis Principles for Chemical and Biological Engineers*. Nob Hill Publishing, 2013.
- [4] Jeffrey Humpherys, Preston Redd, and Jeremy West. A fresh look at the Kalman filter. *SIAM Review*, 54(4):801–823, 2012.
- [5] Rudolf Emil Kalman. Journal of basic engineering. *A New Approach to Linear Filtering and Prediction Problems*, 82(1):35–45, 1960.
- [6] Lennart Ljung. *System Identification Theory for the User (2nd Edition)*. Prentice Hall, Upper Saddle River, New Jersey, U.S., 2009.
- [7] P. E. Moraal and J. W. Grizzle. Observer design for nonlinear systems with discrete-time measurement. *IEEE Transactions on Automatic Control*, 40(3):395–404, 1995.
- [8] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, U.S., 2nd edition, 2006.
- [9] M. R. Rajamani and J. B. Rawlings. Application of a new data-based covariance estimation technique to a nonlinear industrial blending drum. Technical Report 2007-03, University of Wisconsin-Madison, September 2007.
- [10] Christopher. V. Rao and James. B. Rawlings. Constrained process monitoring: Moving-horizon approach. *AIChE Journal*, 48(1):97–109, 2002.
- [11] Christopher. V. Rao, James. B. Rawlings, and J. H. Lee. Constrained linear state estimation — a moving horizon approach. *Automatica*, 37(10):1619–1628, 2001.
- [12] H. E. Rauch, F. Tung, and C. T. Striebel. AIAA journal. *Maximum Likelihood Estimates of Linear Dynamic System*, 3(80):1445–1450, 1965.
- [13] James B. Rawlings and Bhavik R. Bakshi. Particle filtering and moving horizon estimation. *Computer and Chemical Engineering*, 30(2006):1529–1541, 2006.
- [14] D. G. Robertson, J. H. Lee, and James. B. Rawlings. A moving horizon-based approach least-squares estimation. *AIChE Journal*, 42(8):2209–2224, 1996.
- [15] Hisashi Tanizaki. *Computational Methods in Statistics and Econometrics*. Marcel Dekker, Inc., New York, U.S., 2004.