6th International Symposium on
Advanced Control of Industrial Processes (AdCONIP)
May 28-31, 2017. Taipei, Taiwan

On Initialization of the Kalman Filter *

Shunyi Zhao and Biao Huang

Abstract— **As a recursive algorithm, the Kalman filter (KF) assumes the initial state distribution is known *a priori*, while the initial distributions used in practice are commonly treated as design parameters. In this paper, the influences of initial states are analyzed under the KF framework. That is, we address the questions about how the initial mean and variance affect the subsequent estimates and how much performance is sacrificed if incorrect values are used. Based upon this, two initialization methods are developed for the cases with large initial uncertainties. A drafting stochastic resonator model is employed to verify the theoretical analysis result as well as the proposed initialization approach.**

I. INTRODUCTION

The Kalman filter (KF) is known as optimal for the linear Gaussian state-space model in the minimum mean square error (MMSE) sense [1]. Due to its global optimality and simple recursive structure, a large number of methods are proposed based on the KF to modify or extend the original algorithm to suit certain "unsatisfied" conditions [2]–[5].

According to the Kalman filtering theory, the distribution of initial state x_0 is assumed to be known, and the filtering procedure generally starts from the mean and variance of x_0. When applying the KF, however, we often face a dilemma that the available information about $p(x_0)$ is insufficient or even not available. Then, one will eventually end up with a "guess" of the mean value and increasing the variance artificially to accommodate for the uncertainties. Obviously, this is not a systematic solution. When the initial guess is poor, the errors will propagate through the recursions as an undiscovered bias [6], resulting in some long lasting transients. In the applications where state estimate is required to be fast, this phenomenon is barely tolerated [7].

In view of this, some improved methods have been proposed. For example, a strategy that using the conditional mean value of state on the first measurement was suggested in [6]. By tracking the inverse of error covariance, the information filter can handel the cases with infinite covariance [1], [8]. A neural network and a fuzzy logic were combined in [9] to accelerate the convergence rate in presence of incorrect initial guessing. In [10], a method based on state partitioning was given, and the approach for a specific problem can be found in [11]. Although these methods alleviate the poor initialization problem to some extend, none of them really eliminates the negative effect of incorrect initial guessing in a rigorous way [12]. The main problem we believe is how the initial states affect the estimate at each time step in the KF and how much performance is sacrificed with imprecise initial guesses have not been well answered.

In this paper, an batch form of the KF is first derived, from which we search for the answers for the above questions. Then, two initialization strategies are given for the cases with insufficient *a priori* knowledge about the initial state distribution.

II. PRELIMINARIES AND PROBLEM FORMULATION

Consider a general linear state-space model

$$x_n = F_n x_{n-1} + G_n w_n \,, \tag{1}$$

$$y_n = H_n x_n + v_n \,, \tag{2}$$

where $x_n \in \mathbb{R}^K$ denotes the state vector at time n, $y_n \in \mathbb{R}^P$ is the measurement vector, $G_n \in \mathbb{R}^{K\times L}$ is the noise matrix, $F_n \in \mathbb{R}^{K\times K}$ and $H_n \in \mathbb{R}^{P\times K}$ denote the state transition matrix and measurement matrix respectively, and $w_n \in \mathbb{R}^L$ and $v_n \in \mathbb{R}^P$ are the white Gaussian noises with zero means and known variances, i.e., $w_n \sim \mathcal{N}(0, Q_n)$ and $v_n \sim \mathcal{N}(0, R_n)$. It is assumed that the initial state x_0, system noise w_n, and measurement noise v_n are mutually uncorrelated at every time step. A control input u_n can be included in this model, but it is considered as a known variable and can be omitted for the sake of clarity. To have a stable KF, $[F_n\, H_n]$ is assumed to be uniformly detectable and $[F_n\, H_n Q_n^{1/2}]$ is uniformly stabilizable with $Q_n^{1/2}(Q_n^{1/2})^T = Q_n$ [8].

With model (1) and (2), the KF gives the optimal estimates in the MMSE sense through the following three steps:

Initialization: With a known initial distribution $p(x_0) \sim \mathcal{N}(\bar{x}_0, P_0)$, start the recursions from $\hat{x}_{0|0} = \bar{x}_0$ and $P_{0|0} = P_0$.

Predication: Use the transition equation and the posterior estimate associated with the previous state to compute the prediction probability density function (pdf):

$$\hat{x}_{n|n-1} = F_n \hat{x}_{n-1|n-1} \,, \tag{3}$$

$$P_{n|n-1} = F_n P_{n-1|n-1} F_n^T + G_n Q_n G_n^T \,, \tag{4}$$

Updating: Update the prediction pdf with the likelihood:

$$\hat{x}_{n|n} = \hat{x}_{n|n-1} + K_n(y_n - H_n \hat{x}_{n|n-1}) \,, \tag{5}$$

$$P_{n|n} = P_{n|n-1} - K_n H_n P_{n|n-1} \,, \tag{6}$$

$$K_n = P_{n|n-1} H_n^T \left(H_n P_{n|n-1} H_n^T + R_n\right)^{-1} . \tag{7}$$

Here, $\hat{x}_{n|n-1}$ and $P_{n|n-1}$ are the predicted (or prior) state estimation and variance given observations up to and including time $n-1$, K_n is the Kalman gain, and $\hat{x}_{n|n}$ and $P_{n|n}$ are the updated (or posterior) state estimation and variance given observations up to and including time n.

*This work was supported by NSERC and AITF.

S. Zhao and B. Huang are with Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta, AB, Canada T6G 2G6, shunyi@ualberta.ca, bhuang@ualberta.ca

As can be seen, in KF, the mean $\bar{x}_0$ and variance P_0 are assumed to be known, which is intuitively demanding. In most cases, the initialization values used are more likely to be $\tilde{x}_0 = \bar{x}_0 + e_0$ and $\tilde{P}_0 = P_0 + P_e$, where e_0 and P_e denote the errors associated with mean $\bar{x}_0$ and variance P_0 respectively. The problem considered in this paper can now be formulated as following. Given the models (1) and (2) with inaccurate initial distribution $p(x_0) \sim \mathcal{N}(\tilde{x}_0\,, \tilde{P}_0)$, we will address the following two questions: (i) how the initial values $\bar{x}_0$ and P_0 affect the subsequent estimates from analytical point of view in the KF? (ii) what is the price (e.g. in terms of accuracy) one has to pay for if we use $\tilde{x}_0$ and $\tilde{P}_e$ instead of $\bar{x}_0$ and P_0? On the basis of these, we will develop an initialization method to reduce the negative influence of e_0 and P_e.

III. Propagation of Initial State

In order to show how $\bar{x}_0$ and P_0 affect the estimate $\hat{x}_{n|n}$ from an analytical point of view, an extended state-space model that represents the original model (1) and (2) on the time interval $[0\,, n]$ is first constructed, as given below.

A. Extended State-Space Model

Using the forward-in-time solution illustrated in [13], the state equations at different time steps can be listed as follows:

$$\begin{aligned} x_n &= F_n x_{n-1} + G_n w_n\,, \\ x_{n-1} &= F_{n-1} x_{n-2} + G_{n-1} w_{n-1}\,, \\ &\vdots \\ x_1 &= F_1 x_0 + G_1 w_1\,. \end{aligned}$$

Substituting x_i with the equation associated with x_{i-1}, where i changes from 2 to $n-1$, and following a similar trick for the measurement equation, it is not difficult to find the extended state-space model on the time interval $[0\,, n]$ given by

$$X_{n,1} = F_{n,1} x_0 + G_{n,1} W_{n,1}\,, \tag{8}$$

$$Y_{n,1} = H_{n,1} x_0 + L_{n,1} W_{n,1} + V_{n,1}\,, \tag{9}$$

where $X_{n,1} \in \mathbb{R}^{nK}$, $W_{n,1} \in \mathbb{R}^{nL}$, $Y_{n,1} \in \mathbb{R}^{nP}$, and $V_{n,1} \in \mathbb{R}^{nP}$ are specified by $X_{n,1} = [x_n^T\,, x_{n-1}^T\,, \cdots, x_1^T]^T$, $W_{n,1} = [w_n^T\,, w_{n-1}^T\,, \cdots, w_1^T]^T$, $Y_{n,1} = [y_n^T\,, y_{n-1}^T\,, \cdots, y_1^T]^T$, and $V_{n,1} = [v_n^T\,, v_{n-1}^T\,, \cdots, v_1^T]^T$. The extended system matrices $F_{n,1} \in \mathbb{R}^{nK\times K}$, $G_{n,1} \in \mathbb{R}^{nK\times nL}$, $H_{n,1} \in \mathbb{R}^{nP\times K}$, and $L_{n,1} \in \mathbb{R}^{nP\times nL}$ are given as $F_{n,1} = [\mathcal{F}_{n,1}^T\,, \mathcal{F}_{n-1,1}^T\,, \cdots, \mathcal{F}_{1,1}^T]^T$, $H_{n,1} = \bar{H}_{n,1} F_{n,1}$, and $L_{n,1} = \bar{H}_{n,1} G_{n,1}$ with

$$G_{n,1} = \begin{bmatrix} G_n & F_n G_{n-1} & \cdots & \mathcal{F}_{n,3} G_2 & \mathcal{F}_{n,2} G_1 \\ 0 & G_{n-1} & \cdots & \mathcal{F}_{n-1,3} G_2 & \mathcal{F}_{n-1,2} G_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & G_2 & \mathcal{F}_{2,2} G_1 \\ 0 & 0 & \cdots & 0 & G_1 \end{bmatrix},$$

with $\bar{H}_{n,1} = \mathrm{diag}\big(H_n\,, H_{n-1}\,, \cdots, H_1\big)$ and $\mathcal{F}_{j,i} = F_j F_{j-1} \cdots F_i$, $i \leqslant j$.

Note that the dimensions of all the vectors and matrices in (8) and (9) grow linearly with time n.

B. Batch Optimal Estimates

With the extended state-space model available, we illustrate the effect of initial states $\bar{x}_0$ and P_0 used in the KF next.

Similar to the least square estimate that handles all the observations at one time, we construct a linear estimator of x_n using the extended measurement vector $Y_{n,1}$ and the initial mean $\bar{x}_0$ by

$$\hat{x}_{n|n} = K_{n,1} Y_{n,1} + S_{n,1} \bar{x}_0\,, \tag{10}$$

where the gains $K_{n,1}$ and $S_{n,1}$ are determined with some predefined cost functions. Note that we use the error-free value $\bar{x}_0$ instead of $\tilde{x}_0$ here in order to be consistent with the KF (3)-(7). The case of $\tilde{x}_0$ will be demonstrated latter.

According to the orthogonality principle [13], the estimate (10) achieves the optimality in the MMSE sense if and only if

$$\mathbb{E}\big\{x_n - \hat{x}_{n|n}\big\} = 0\,, \tag{11}$$

$$\mathbb{E}\big\{\big(x_n - \hat{x}_{n|n}\big) Y_{n,1}^T\big\} = 0\,. \tag{12}$$

From (8), the state x_n can be rewritten as

$$x_n = \mathcal{F}_{n,1} x_0 + \bar{G}_{n,1} W_{n,1}\,, \tag{13}$$

where $\bar{G}_{n,1} \triangleq [G_n \ F_n G_{n-1} \ \cdots \ \mathcal{F}_{n,3} \ \mathcal{F}_{n,2} G_2]$ is the first row vector of $G_{n,1}$. Substituting (13) and (10) into the first condition (11), and taking the expectation, we have

$$\begin{aligned} \mathcal{F}_{n,1} \mathbb{E}\{x_0\} &= K_{n,1} H_{n,1} \mathbb{E}\{x_0\} + S_{n,1} \bar{x}_0 \\ &= (K_{n,1} H_{n,1} + S_{n,1}) \mathbb{E}\{x_0\}\,, \end{aligned}$$

where the facts that $\mathbb{E}\{W_{n,1}\} = 0$, $\mathbb{E}\{V_{n,1}\} = 0$, and $\bar{x}_0 = \mathbb{E}\{x_0\}$ are used. Since $\mathbb{E}\{x_0\}$ cannot always be zero, one consequently gets $S_{n,1} = \mathcal{F}_{n,1} - K_{n,1} H_{n,1}$, which enables us to represent (10) equivalently by

$$\hat{x}_{n|n} = K_{n,1} Y_{n,1} + (\mathcal{F}_{n,1} - K_{n,1} H_{n,1}) \bar{x}_0\,. \tag{14}$$

Now, we consider the second condition (12) that implies the estimation error $x_n - \hat{x}_{n|n}$ should be orthogonal to the extended measurement vector $Y_{n,1}$. Replacing x_n with (13), $\hat{x}_{n|n}$ with (14), and $Y_{n,1}$ with (9), yields

$$\begin{aligned} &\mathbb{E}\big\{\big[\mathcal{F}_{n,1} x_0 + \bar{G}_{n,1} W_{n,1} - K_{n,1} Y_{n,1} - S_{n,1} \bar{x}_0\big] \\ &\quad \times \big[H_{n,1} x_0 + L_{n,1} W_{n,1} + V_{n,1}\big]^T\big\} = 0\,, \end{aligned}$$

which further leads to

$$\begin{aligned} 0 &= \mathbb{E}\big\{\big[(K_{n,1} H_{n,1} - \mathcal{F}_{n,1})(x_0 - \bar{x}_0) \\ &\quad + (K_{n,1} L_{n,1} - \bar{G}_{n,1}) W_{n,1} + K_{n,1} V_{n,1}\big] \\ &\quad \times \big[H_{n,1} x_0 + L_{n,1} W_{n,1} + V_{n,1}\big]^T\big\} \\ &= (K_{n,1} H_{n,1} - \mathcal{F}_{n,1}) P_0 H_{n,1}^T \\ &\quad + (K_{n,1} L_{n,1} - \bar{G}_{n,1}) Q_{n,1} L_{n,1}^T \\ &\quad + K_{n,1} R_{n,1}\,, \end{aligned} \tag{15}$$

where

$$\begin{aligned}
P_0 &\triangleq \mathbb{E}\{(x_0-\bar{x}_0)x_0^T\}\\
&= \mathbb{E}\{x_0x_0^T\}-\bar{x}_0\mathbb{E}\{x_0^T\}-\mathbb{E}\{x_0\}\bar{x}_0^T+\bar{x}_0\bar{x}_0^T\\
&= \mathbb{E}\{(x_0-\bar{x}_0)(x_0-\bar{x}_0)^T\}\,,\\
Q_{n,1} &\triangleq \mathbb{E}\{W_{n,1}W_{n,1}^T\}=\mathrm{diag}(Q_n\,,Q_{n-1}\,,\cdots,Q_n)\,,\\
R_{n,1} &\triangleq \mathbb{E}\{V_{n,1}V_{n,1}^T\}=\mathrm{diag}(R_n\,,R_{n-1}\,,\cdots,R_n)\,.
\end{aligned}$$

After some arrangements, (15) becomes

$$K_{n,1}\Sigma_{n,1}=\mathcal{F}_{n,1}P_0H_{n,1}^T+\bar{G}_{n,1}Q_{n,1}L_{n,1}^T\,,$$

from which we get the required gain $K_{n,1}$ by dividing it with $\Sigma_{n,1}$ from both-hand sides as

$$K_{n,1}=(\mathcal{F}_{n,1}P_0H_{n,1}^T+\bar{G}_{n,1}Q_{n,1}L_{n,1}^T)\Sigma_{n,1}^{-1}\,, \tag{16}$$

where

$$\Sigma_{n,1}=H_{n,1}P_0H_{n,1}^T+L_{n,1}Q_{n,1}L_{n,1}^T+R_{n,1}\,. \tag{17}$$

At this point, we can see that (14) with $K_{n,1}$ specified by (16) provides us the MMSE estimate, where the coefficient $S_{n,1}=\mathcal{F}_{n,1}-K_{n,1}H_{n,1}$ governs the effect of $\bar{x}_0$ on the estimate $\hat{x}_{n|n}$, and the counterpart of P_0 is reflected by $K_{n,1}$. However, we still cannot answer the target equations yet, as the relationship between (14) and the KF (3)-(7) is not clear. Considering the fact that both methods minimize the estimation error variance, one naturally expects the equivalence (or transformation) between them, which will be demonstrated next.

C. Equivalence to the KF

The algorithmic structure of KF is recursive, while (14) is a batch form. To show their equivalence, (14) will also be realized in a recursive form.

By introducing intermediate variables $\hat{x}_{n|n}^{(a)}\triangleq K_{n,1}Y_{n,1}$ and $\hat{x}_{n|n}^{(b)}\triangleq(\mathcal{F}_{n,1}-K_{n,1}H_{n,1})\bar{x}_0$, we decompose $\hat{x}_{n|n}$ by

$$\hat{x}_{n|n}=\hat{x}_{n|n}^{(a)}+\hat{x}_{n|n}^{(b)}\,. \tag{18}$$

1) Recursion of $\hat{x}_{n|n}^{(a)}$: From (16), one can accordingly get

$$K_{n,1}=D_{n,1}\Sigma_{n,1}^{-1}\,,$$

where

$$D_{n,1}=\underbrace{\mathcal{F}_{n,1}P_0H_{n,1}^T}_{D_{n,1}^{(a)}}+\underbrace{\bar{G}_{n,1}Q_{n,1}L_{n,1}^T}_{D_{n,1}^{(b)}}\,.$$

Using the definitions of matrices $H_{n,1}$ and $L_{n,1}$, the first-two components in $\Sigma_{n,1}$ can be decomposed as, respectively,

$$\begin{aligned}
\Sigma_{n,1}(P) &\triangleq H_{n,1}P_0H_{n,1}^T\\
&= \begin{bmatrix} H_n\mathcal{F}_{n,1}P_0\mathcal{F}_{n,1}^TH_n^T & H_nF_nD_{n-1,1}^{(a)}\\ (D_{n-1,1}^{(a)})^TF_n^TH_n^T & \Sigma_{n-1,1}(P)\end{bmatrix}\,, \qquad (19)\\
\Sigma_{n,1}(Q) &\triangleq L_{n,1}Q_{n,1}L_{n,1}^T\\
&= \begin{bmatrix} H_n\bar{G}_{n,1}Q_{n,1}\bar{G}_{n,1}^TH_n^T & H_nF_nD_{n-1,1}^{(b)}\\ (D_{n-1,1}^{(b)})^TF_n^TH_n^T & \Sigma_{n-1,1}(Q)\end{bmatrix}\,, \qquad (20)
\end{aligned}$$

where the equality

$$\bar{G}_{n,1}Q_{n,1}\bar{G}_{n,1}^T=G_nQ_nG_n^T+F_n\bar{G}_{n-1,1}Q_{n-1,1}\bar{G}_{n-1,1}^TF_n^T$$

is used. Substituting (19) and (20) into (17), and considering $R_{n,1}$ is a diagonal matrix, $\Sigma_{n,1}$ can be rewritten by

$$\Sigma_{n,1}=\Delta_{n,1}+\Psi_{n,1}\,,$$

where

$$\begin{aligned}
\Delta_{n,1} &= \begin{bmatrix} R_n & 0\\ 0 & \Sigma_{n-1,1}\end{bmatrix}\,,\\
\Psi_{n,1} &= \begin{bmatrix} H_nU_{n,1}H_n^T & H_nF_nD_{n-1,1}\\ D_{n-1,1}^TF_n^TH_n^T & 0\end{bmatrix}\,,\\
U_{n,1} &= \mathcal{F}_{n,1}P_0\mathcal{F}_{n,1}^T+\bar{G}_{n,1}Q_{n,1}\bar{G}_{n,1}^T\,. \qquad (21)
\end{aligned}$$

Using the inverse matrix lemma [14], we get the inverse of $\Sigma_{n,1}$ required in $K_{n,1}$ by

$$\Sigma_{n,1}^{-1}=\Delta_{n,1}^{-1}-\Delta_{n,1}^{-1}\big(I+\Psi_{n,1}\Delta_{n,1}^{-1}\big)^{-1}\Psi_{n,1}\Delta_{n,1}^{-1}\,, \tag{22}$$

as $\Delta_{n,1}$ is invertible.

Another recursion needed is for the matrix $D_{n,1}$, whose components $D_{n,1}^{(a)}$ and $D_{n,1}^{(b)}$ satisfy the following equalities:

$$\begin{aligned}
D_{n,1}^{(a)} &= \big[\mathcal{F}_{n,1}P_0\mathcal{F}_{n,1}^TH_n^T \quad \mathcal{F}_{n,1}P_0H_{n-1,1}^T\big]\,,\\
D_{n,1}^{(b)} &= \big[\bar{G}_{n,1}Q_{n,1}\bar{G}_{n,1}^TH_n^T \quad F_n\underbrace{\bar{G}_{n-1,1}Q_{n-1,1}L_{n-1,1}^T}_{D_{n-1,1}^{(b)}}\big]\,.
\end{aligned}$$

Combining these two equations gives us the recursion of $D_{n,1}$ as

$$D_{n,1}=\big[U_{n,1}H_n^T \quad F_nD_{n-1,1}\big]\,. \tag{23}$$

At this point, the gain $K_{n,1}$ becomes

$$\begin{aligned}
K_{n,1} &= \big[U_{n,1}H_n^T \quad F_nD_{n-1,1}\big]\\
&\quad\times\big(\Delta_{n,1}^{-1}-\Delta_{n,1}^{-1}\big(I+\Psi_{n,1}\Delta_{n,1}^{-1}\big)^{-1}\Psi_{n,1}\Delta_{n,1}^{-1}\big)\\
&= \big[U_{n,1}H_n^TR_n^{-1} \quad F_nD_{n-1,1}\Sigma_{n-1,1}^{-1}\big]\\
&\quad\times\big(I-\big(I+\Psi_{n,1}\Delta_{n,1}^{-1}\big)^{-1}\Psi_{n,1}\Delta_{n,1}^{-1}\big)\\
&= \big[U_{n,1}H_n^TR_n^{-1} \quad F_nK_{n-1,1}\big]\Lambda_{n,1}^{-1}\,, \qquad (24)
\end{aligned}$$

where

$$\Lambda_{n,1}\triangleq I+\Psi_{n,1}\Delta_{n,1}^{-1}=\begin{bmatrix}\Lambda_{n,1}^{(1,1)} & \Lambda_{n,1}^{(1,2)}\\ \Lambda_{n,1}^{(2,1)} & \Lambda_{n,1}^{(2,2)}\end{bmatrix}\,,$$

with the corresponding elements specified as

$$\begin{aligned}
\Lambda_{n,1}^{(1,1)} &= I+H_nU_{n,1}H_n^TR_n^{-1}\,,\\
\Lambda_{n,1}^{(1,2)} &= H_nF_nK_{n-1}\,,\\
\Lambda_{n,1}^{(2,1)} &= D_{n-1,1}^TF_n^TH_n^TR_n^{-1}\,,\\
\Lambda_{n,1}^{(2,2)} &= I\,.
\end{aligned}$$

Since $\Lambda_{n,1}^{(2,2)}$ is invertible, the Schur complement [15] can be employed, by computing the inverse of $\Lambda_{n,1}$ with

$$\Lambda_{n,1}^{-1}=\begin{bmatrix}\tilde{\Lambda}_{n,1}^{(1,1)} & -\tilde{\Lambda}_{n,1}^{(1,1)}\Lambda_{n,1}^{(1,2)}\\ -\Lambda_{n,1}^{(2,1)}\tilde{\Lambda}_{n,1}^{(1,1)} & I+\Lambda_{n,1}^{(2,1)}\tilde{\Lambda}_{n,1}^{(1,1)}\Lambda_{n,1}^{(1,2)}\end{bmatrix}\,, \tag{25}$$

where

$$\begin{aligned}\tilde{\Lambda}_{n,1}^{(1,1)} &= (\Lambda_{n,1}^{(1,1)} - \Lambda_{n,1}^{(1,2)}\Lambda_{n,1}^{(2,1)})^{-1} \\ &= (I + H_n N_n H_n^T R_n^{-1})^{-1}, \\ N_n &= U_{n,1} - F_n K_{n-1,1} D_{n-1,1}^T F_n^T . \end{aligned} \quad (26)$$

Substituting (25) into (24), performing the multiplication, and making some arrangements, we get

$$K_{n,1} = \begin{bmatrix} J_n & F_n K_{n-1,1} - J_n H_n F_n K_{n-1,1} \end{bmatrix},$$

where

$$J_n = N_n H_n^T \big(R_n + H_n N_n H_n^T\big)^{-1} . \quad (27)$$

Using the definition of $\hat{x}_{n|n}^{(a)}$, the corresponding recursion turns out to be

$$\hat{x}_{n|n}^{(a)} = F_n \hat{x}_{n-1|n-1}^{(a)} + J_n\big(y_n - H_n F_n \hat{x}_{n-1|n-1}^{(a)}\big) . \quad (28)$$

2) Recursion of $\hat{x}_{n|n}^{(b)}$: Following a similar line, the second term $\hat{x}_{n|n}^{(b)}$ can be computed by

$$\begin{aligned}\hat{x}_{n|n}^{(b)} &= F_n(\mathcal{F}_{n-1,1} - K_{n-1,1}H_{n-1,1})\bar{x}_0 \\ &\quad -J_n H_n F_n\big(\mathcal{F}_{n-1,1} - K_{n-1,1}H_{n-1,1}\big)\bar{x}_0 \\ &= (I - J_n H_n) F_n \hat{x}_{n-1|n-1}^{(b)} . \end{aligned} \quad (29)$$

3) Recursion of $\hat{x}_{n|n}$: Combining (28) and (29), the recursive implementation of $\hat{x}_{n|n}$ is

$$\begin{aligned}\hat{x}_{n|n} &= F_n(\hat{x}_{n-1|n-1}^{(a)} + \hat{x}_{n-1|n-1}^{(b)}) \\ &\quad +J_n(y_n - H_n F_n(\hat{x}_{n-1|n-1}^{(a)} + \hat{x}_{n-1|n-1}^{(b)})) \\ &= F_n \hat{x}_{n-1|n-1} + J_n(y_n - H_n F_n \hat{x}_{n-1|n-1}) \end{aligned} \quad (30)$$

which has the same structure as the KF. That is, the term $F_n\hat{x}_{n-1|n-1}$ predicts the estimate from the previous estimation, and the filter gain J_n corrects the predicted values. By setting $n = n+1$ and using (21) and (23) again, the recursive expression of N_n involved in J_n can be found as

$$N_n = F_n N_{n-1} F_n^T + G_n Q_n G_n^T - F_n J_{n-1} H_{n-1} N_{n-1} F_n^T , \quad (31)$$

with initial value $N_1 = F_1 P_0 F_1^T + G_1 Q_1 G_1^T$.

Comparing (27), (30) and (31) with the KF, we complete our proof that the batch estimate (14) is equivalent to the KF with the following notational transformations:

$$J_n \Leftrightarrow K_n \ , N_n \Leftrightarrow P_{n|n-1} .$$

D. Effect of Initial States

Now, we summarize the main results and address the questions by the following theorem.

Thorem 1: Given the linear state-space model (1) and (2) with known initial distribution $p(x_0) \sim \mathcal{N}(\bar{x}_0, P_0)$, and uncorrelated white Gaussian noises $w_n \sim \mathcal{N}(0, Q_n)$ and $v_n \sim \mathcal{N}(0, R_n)$, the traditional KF can be equivalently transformed to $\hat{x}_{n|n} = K_{n,1}Y_{n,1} + \big(\mathcal{F}_{n,1} - K_{n,1}H_{n,1}\big)\bar{x}_0$, where $K_{n,1}$ is specified by (16). The term $\hat{x}_{n|n}^{(b)} = (\mathcal{F}_{n,1} - K_{n,1}H_{n,1})\bar{x}_0$ implies the effect of $\bar{x}_0$ on the estimate $\hat{x}_{n|n}$, whose recursion form is provided by

$$\hat{x}_{n|n}^{(b)} = \big(I - K_n H_n\big) F_n \hat{x}_{n-1|n-1}^{(b)} , \quad (32)$$

where K_n is the Kalman gain calculated at time n.

Proof 1: The proof has been given from section III-A to section III-C.

Considering (14), it shows that the influence of P_0 is reflected through the gain $K_{n,1}$ in an indirect way, while $\bar{x}_0$ affects $\hat{x}_{n|n}$ directly. This is why the KF is more sensitive to the initial mean used, compared with the variance. In the cases with imprecise initial distributions $\mathcal{N}(\tilde{x}_0, \tilde{P}_0)$, accordingly, the corrupted estimate $\hat{x}'_{n|n}$ can be obtained by

$$\hat{x}'_{n|n} = \tilde{K}_{n,1}Y_{n,1} + \big(\mathcal{F}_{n,1} - \tilde{K}_{n,1}H_{n,1}\big)\tilde{x}_0 ,$$

where $\tilde{K}_{n,1}$ is specified as (14) by replacing P_0 with $\tilde{P}_0$. Then, the the price in term of accuracy we pay for the incorrect initialization can be evaluated by

$$\Delta x_{n|n} = \hat{x}'_{n|n} - \hat{x}_{n|n} .$$

Up to now, the required solutions are all found. A step forward is to eliminate the negative effect of $\tilde{x}_0$ and $\tilde{P}_0$ to get satisfied estimates as quick as possible after filtering starts.

IV. Initialization Strategies

From Theorem 1, we see that the sensitiveness of $\hat{x}_{n|n}$ to $\bar{x}_0$ is due to the term $S_{n,1}\bar{x}_0$, which acts as a bias with a direct impact. To overcome this problem, we suggest two different strategies below.

A. Strategy One

The first strategy is to minimize the MMSE with the unbiasedness constraint (11). That is, solve the optimization problem

$$\arg\min_{K_{n,1}} \operatorname{tr} \mathbb{E}\big\{[x_n - \hat{x}_{n|n}][x_n - \hat{x}_{n|n}]^T\big\}$$

subject to $\mathcal{F}_{n,1} = K_{n,1}H_{n,1}$, where tr means the trace operation and $\hat{x}_{n|n}$ is specified by (14) that can be further rewritten as $\hat{x}_{n|n} = K_{n,1}Y_{n,1}$ due to the constraint. This problem has been solved in our recent paper [16], leading to a linear optimal unbiased filter.

B. Strategy Two

The second strategy is that we remove $S_{n,1}\bar{x}_0$ artificially, and get a new linear estimator as

$$\hat{x}_{n|n} = \bar{K}_{n,1}Y_{n,1} . \quad (33)$$

Here, the compensation used for the neglected term is to recompute a new gain $\bar{K}_{n,1}$ by solving

$$\arg\min_{\bar{K}_{n,1}} \operatorname{tr} \mathbb{E}\big\{[x_n - \hat{x}_{n|n}][x_n - \hat{x}_{n|n}]^T\big\} , \quad (34)$$

Substituting x_n with (13) and $\hat{x}_{n|n}$ with (33), we get

$$\begin{aligned}\arg\min_{\bar{K}_{n,1}} \operatorname{tr} \mathbb{E}\big\{&\big[(\bar{K}_{n,1}H_{n,1} - \mathcal{F}_{n,1})x_0 \\ &+(\bar{K}_{n,1}L_{n,1} - \bar{G}_{n,1})W_{n,1} + \bar{K}_{n,1}V_{n,1}\big]\big[\cdots\big]^T\big\} .\end{aligned}$$

Comparing with the strategy one, the first term $\bar{K}_{n,1}H_{n,1} - \mathcal{F}_{n,1}$ cannot be omitted since we do not have the constraint $\bar{K}_{n,1}H_{n,1} = \mathcal{F}_{n,1}$. On the other hand, comparing it with the former minimization problem associated with the estimator (14),

$$\begin{aligned}&\arg\min_{K_{n,1}} \operatorname{tr}\mathbb{E}\big\{\big[(K_{n,1}H_{n,1} - \mathcal{F}_{n,1})(x_0 - \bar{x}_0)\\&+(K_{n,1}L_{n,1} - \bar{G}_{n,1})W_{n,1} + \bar{K}_{n,1}V_{n,1}\big]\big[\cdots\big]^T\big\},\end{aligned}$$

one can see that in the KF, the gain $K_{n,1}$ is designed for $x_0 - \bar{x}_0$ as $\bar{x}_0$ is assumed to be available, while $\bar{K}_{n,1}$ is designed for x_0 itself in the proposed method (33). With this difference, the direct influence of $\bar{x}_0$ will be changed to be indirect (same as P_0), and better robustness is thus achieved.

Using the lemma given in [16], the solution to (34) can be obtained by

$$\bar{K}_{n,1} = (\mathcal{F}_{n,1}\Theta_{x_0}H_{n,1}^T + \bar{G}_{n,1}Q_{n,1}L_{n,1}^T)\bar{\Sigma}_{n,1}^{-1}, \quad (35)$$

where

$$\begin{aligned}\bar{\Sigma}_{n,1} &= H_{n,1}\Theta_{x_0}H_{n,1}^T + L_{n,1}Q_{n,1}L_{n,1}^T + R_{n,1}, &(36)\\ \Theta_{x_0} &= \mathbb{E}\{x_0x_0^T\} = \bar{x}_0\bar{x}_0^T + P_0. &(37)\end{aligned}$$

Since $\bar{K}_{n,1}$ has the same structure as $K_{n,1}$, we can get its recursion form by (28) easily. Here, however, more attentions should be paid on the initial values of $\bar{K}_{n,1}$. Using the definitions of $H_{n,1}$, $L_{n,1}$ and $\bar{G}_{n,1}$ with $n = 1$, we have

$$\begin{aligned}\bar{K}_{1,1} &= (F_1\Theta_{x_0}F_1^T + G_1Q_1G_1^T)H_1^T\big(H_1(F_1\Theta_{x_0}F_1^T\\ &\quad +G_1Q_1G_1^T)H_1^T + R_1\big)^{-1}, &(38)\end{aligned}$$

where $\Theta_{x_0} = \tilde{x}_0\tilde{x}_0^T + \tilde{P}_0$ as the true value (37) is not known. Then, the first estimate $\hat{x}_{1|1}$ can be calculated by (33) using $Y_{n,1} = y_1$. Once $\hat{x}_{1|1}$ is obtained, the subsequent estimates can be recursively computed with (28) starting from $N_1 = F_1\Theta_{x_0}F_1^T + G_1Q_1G_1^T$.

V. SIMULATIONS

In this section, we test the proposed initialization methods (mainly on the strategy two since the performance of strategy one has been demonstrated in [16]) using a randomly drifting stochastic resonator [17] as an example, which is specified by (1) with $G_n = I$ and

$$F_n = \begin{bmatrix}1 & 0 & 0\\ 0 & \cos(\alpha_n\Delta_n) & \sin(\alpha_n\Delta_n)/\alpha_n\\ 0 & -\alpha_n\sin(\alpha_n\Delta_n) & \cos(\alpha_n\Delta_n)\end{bmatrix},$$

where $\Delta_n = 0.5$s is the sampling interval and α_n takes value from $\{0.05, 0.06\}$ according to a Markov chain with a transition probability matrix whose diagonal elements are 0.9. The noise in resonator is zero mean white Gaussian with variance $\delta_{w1}^2 = \delta_{w2}^2 = \delta_{w3}^2 = 1$.

The measurement equation is given by (2) with $H_n = [1\ 1\ 0]$, where the measurement noise v_n is white Gaussian with zero mean and variance $\delta_v^2 = 100$. The process starts from $p(x_0) \sim \mathcal{N}(\bar{x}_0, P_0)$ with $\bar{x}_0 = [8, -2, -2]^T$ and $P_0 = [P_{0,1}, P_{0,2}, P_{0,3}]$ with $P_{0,1} = [5, 0, 0]^T$, $P_{0,2} = [0, 5, 0]^T$, and $P_{0,3} = [0, 0, 5]^T$.

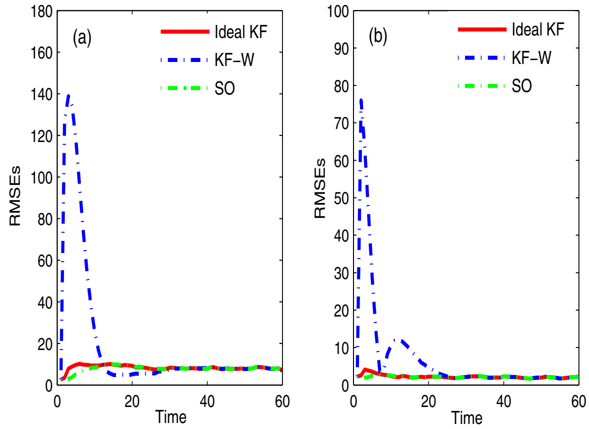

Fig. 1. RMSEs of different algorithms with $\tilde{x}_0$ and $\tilde{P}_0$ based on 100 Monte-Carlo runs: (a) the second state and (b) the third state, where "SO" is the abbreviation of strategy one.

As can be seen, the first state is uncorrelated with the other two states, and thus is generally considered as a drafting source in a resonator model. Therefore, we mainly focus on the estimation of the second and third states. Considering that $\bar{x}_0$ and P_0 are not precisely known and we use a poor initial guess $\tilde{x}_0 = [8, 100, 80]$ and $\tilde{P}_0 = [\tilde{P}_{0,1}, \tilde{P}_{0,2}, \tilde{P}_{0,3}]$ with $\tilde{P}_{0,1} = [5\ 0\ 0]^T$, $\tilde{P}_{0,2} = [0\ 10\ 0]^T$, and $\tilde{P}_{0,3} = [0\ 0\ 10]^T$ to test both the strategy one (SO) and KF (denoted as KF-W) to demonstrate the trade-off.

The average root mean square errors (RMSEs) of different methods based on 100 Monte-Carlo runs are provided in Fig. 1, where the ideal KF with correct initial distribution is also employed as a benchmark. It shows that, considerable improvements on accuracy are achieved by the SO, especially in the initial estimation phase, while the Kalman estimates converge to the true states at a slower rate, due to the long lasting negative effect of poor $\tilde{x}_0$. One also immediately notices that the difference between SO and the ideal KF is minor, which can be neglected completely in practical applications. This provides us another proof from the opposite angle that the SO can be used in place of the KF when initialization is uncertain.

By introducing an coefficient as $q\tilde{P}_0$, we next compare SO with KF using different initial covariances, where q takes value from $\{5, 30, 80\}$. The main reason for this design is that the initial performance of KF with incorrect initial states can be improved by amplifying the corresponding covariance, and we would like to see whether this rule is valid for the proposed algorithm or not. Based on 100 Mento-Carlo runs, the average RMSEs with respect to different q are shown in Fig. 2. As expected, the KF is sensitive to the q used, and the negative effect of $\tilde{x}_0$ in the KF decreases with the increase of q considerably. For the SO, although similar phenomenon can be observed in Fig. 2(b), it is not significant. Comparing it with the KF, we can see

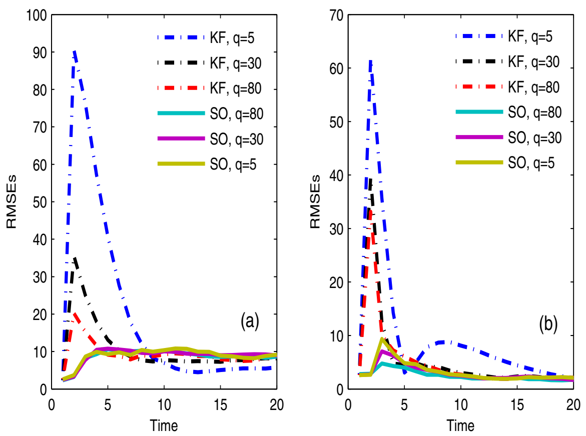

Fig. 2. Average RMSEs of different algorithms with respect to q: (a) the second state and (b) the third state.

that the accuracy of SO with $p = 5$ is even higher than the counterpart of KF with $p = 80$, which can be considered as a distinctive advantage of the proposed method.

VI. CONCLUSIONS

In this paper, the effect of initial states used in the KF is analyzed. It is shown that comparing to the initial variance P_0, the initial mean $\bar{x}_0$ introduces a bias in the state estimation in a more direct way. Two initialization strategies are introduced. One is to minimize the error variance with the unbiasedness constraint, and the other is to remove the term with respect to the initial mean artificially and modify the gain optimally. Using a drafting stochastic resonator as an example, we demonstrate that the proposed approach can be used as an alternative to the KF when there is uncertainty in the initial conditions.

REFERENCES

[1] D. Simon, *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*, Hoboken, NJ: J. Wiley & Sons, 2006.

[2] H. Cox, "On the estimation of state variables and parameters for noisy dynamic systems, *IEEE Trans. Autom. Contr.*, vol. 9, no. 1, pp. 5C12, Jan. 1964.

[3] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems, in *Proc. of SPIE*, vol. 3068, pp. 182-193, 1997.

[4] B. Ait-El-Fquih and F. Desbouvries, "Kalman Filtering in Triplet Markov Chains, *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2957C2963, Aug. 2006.

[5] Y. Becis-Aubry, M. Boutayeb, and M. Darouach, "State estimation in the presence of bounded disturbances, *Automatica*, vol. 44, no. 7, pp. 1867C1873, Jul. 2008.

[6] L. B. Weiner, "Kalman filter initialization with large initial uncertainty and strong measurement nonlinearity, in *Proceedings of the Region 3 Conference and Exhibit*, Huntsville, AL; United States, pp. 150-151, 1981.

[7] U. Frese, B. Baeuml, G. Schreiber, and I. Schaefer, et al, "Off-the-shelf vision for a robotic ball catcher, In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Oktober, Maui, pp. 1623-1629, 2011.

[8] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prenrice-Hall, 1979.

[9] J. M. Roberts, D. J. Mills, D. Charnley, and C. J. Harris, "Improved Kalman filter initialisation using neurifuzzy estimation, *Fourth Int. Conf. on Artificial Neural Networks*, pp. 26C28, 1995.

[10] M. Linderoth, K. Soltesz, A. Robwetsson, and R. Johansson, "Initialization of the Kalman filter without assumputations on the initial state, *IEEE Int. Conf. on Robotics and Automation*, 2011, pp. 4992-4997.

[11] J. Hyland, "An iterated-extended Kalman filter algorithm for tracking surface and sub-surface targets,*OCEANS 02 MTS/IEEE*, vol. 3, pp. 1283C1290, 2002.

[12] D. Liu and X. Liu, "Optimal state estimation without the requirement of a prior statistics information of the initial state, *IEEE Trans. Automat. Contr.*, vol. 39, no. 10, pp. 2087-2091, Oct. 1994.

[13] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice Hall, 1994.

[14] G. H. Golub and G. F. Van Loan, *Matrix Computation*, 3rd Ed. ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[15] F. L. Lewis, *Optimal Control*, John Wiley & Sons, Inc., 1986.

[16] S.Y. Zhao, B. Huang, and F. Liu, "Linear optimal unbiased filter for time-variant systems without apriori information on initial conditions," *IEEE Trans. Autom. Control*, doi: 10.1109/ TAC.2016.2557999, 2016.

[17] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Trans. Auto. Contr.*, vol. 54, no. 3, pp. 596-600, Mar. 2009.