Derivation of the Kalman filter in a Bayesian filtering perspective

Ramakrishna Gurajala
Dept. of E.C.E., Andhra University
Visakhapatnam, India
ramagurajala@gmail.com

Praveen B. Choppala
Dept. of E.C.E., WISTM, Andhra Univ.
Visakhapatnam, India
praveen.b.choppala@gmail.com

James Stephen Meka
Dept. of C.S.E., WISTM, Andhra Univ.
Visakhatpatnam, India
jamesstephenm@gmail.com

Paul D. Teal
School of E.C.S., Victoria Univ. of Wellington
Welington, New Zealand
paul.teal@vuw.ac.nz

***Abstract*—The Kalman filter is popularly known to be an optimal recursive implementation of the Bayesian prediction and correction in the sense that it minimises the estimated error covariance. The filter has been originally derived in this error minimising framework and there is extensive literature on the same. The Kalman filter has also been derived under other frameworks, like the maximum likelihood approach, etc., which all converge to the true posterior. In this paper we present a purely Bayesian filtering approach to the Kalman filter. We first build an analogy to the principles of Bayesian estimation and then present a step-by-step derivation for the Kalman filter following the Bayesian principles. From this derivation, we show that the Kalman filter gives a tractable solution to the Bayesian filtering process by computing the underlying probability densities exactly. This derivation is known to some in the research community but no formal article in the literature presents it in detail. This paper fills this gap and will be a good read for Bayesian enthusiasts. The filter is simulated in the proposed framework on a simple 4-D linear Gaussian model.**

***Index Terms*—Bayesian state estimation, Kalman filter, linear Gaussian models, prediction, update.**

I. Introduction

The Bayesian state estimation is an important solution to estimate hidden dynamic target states and is used widely in radar/sonar tracking applications [1]. The Bayesian state estimation approach sequentially builds a posterior probability density function (pdf) of the target state conditioned on all the sensor observations by predicting a probable target state from the state transition model and then weighing that prediction using the sensor observation model [2]. The Kalman filter, proposed by R.E. Kalman in 1960, is a popular Bayesian filtering approach that provides an analytical solution to estimate the target state in a way that minimises the error associated with the estimation [3]. Therefore, the filter is theoretically optimal for linear Gaussian state space models. Despite the filter being limited to linear Gaussian models, it continues to dominate Bayesian state estimation applications inluding the recent search for the missing aircraft MH370 [4].

There are different and varied mathematical perspectives to the Kalman filter. The filter can be seen as a least square estimator and also a minimum variance estimator under Gaussian assumptions that minimises the mean square error (MSE) in the estimated parameters [5]. The founding principles for Kalman filtering have been adopted from least squares estimation and then applied to the problem of sequential estimation [6], [7]. That said, the Kalman filter has been derived from different perspectives. The filter was first derived using orthogonal projection method by R.E. Kalman in [8] and was later shown to be equivalent to the time variant Weiner filter [9], [10]. The innovations approach to the Kalman filter derivation was first presented in [11] and thereafter several theoretical implementations were developed for minimising the distance between the predicted target state translated to the observation space and the observation itself by scaling with the Kalman gain. The Bayesian approach to the innovations based Kalman filter derivation is recently given in [12]. More recently, the Kalman filter was derived using the maximum likehood estimation and Newton optimisation methods but here it requires a carefully chosen initial guess of the target state [13], [14].

The Kalman filter and its several variants generally used in the literature provide an optimal recursive implementation of the prediction and the correction in the sense that it minimises the estimated error covariance. However a Kalman filter derivation in a purely Bayesian sense is not readily available as a research article. This paper fills that gap. In this paper, we present a purely Bayesian perspective to the Kalman filter and derive it as an analogous to the Bayesian sequential filter. The rest of the paper is organised as follows. Section II sets the notation and gives a detailed derivation to the Bayesian state estimation process. Section III presents the Bayesian Kalman filter in a purely Bayesian perspective. This is followed by a simulation study in section IV and concluding remarks in section V

II. Bayesian state estimation

In this section we set the notation for the discrete time state space model and derive the Bayesian state estimation methodology [15]. The state of the target at time t, denoted as $\mathbf{x}_t \in \mathbb{R}^{d_\mathbf{x}}$, characterises all the dynamics of the target like the position, velocity, etc. This state evolves in time via a hidden Markov process defined as

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{a}_t) \tag{1}$$

where $d_{\mathbf{x}}$ is the dimension of the target state, $f(.)$ is a Markov function the describes how the target heads from $t-1$ to t and $\mathbf{a}_t$ is the noise associated with the target heading. The sensor observation of the hidden target state, denoted as $\mathbf{y}_t \in \mathbb{R}^{d_{\mathbf{y}}}$, is obtained using the observation model as

$$\mathbf{y}_t = h(\mathbf{x}_t, \mathbf{e}_t) \tag{2}$$

where $d_{\mathbf{y}}$ is the dimension of the observation vector, $h(.)$ is observation function and $\mathbf{e}_t$ is the observation noise. The set of target states and measurements are denoted as $\mathbf{x}_{1:T} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_T\}$ and $\mathbf{y}_{1:T} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t, \ldots, \mathbf{y}_T\}$. This state space model is pictorially shown in Fig. 1. The hidden time varying state $\mathbf{x}_t$ of the system at time t can be observed as $\mathbf{y}_t$ by the observation model.

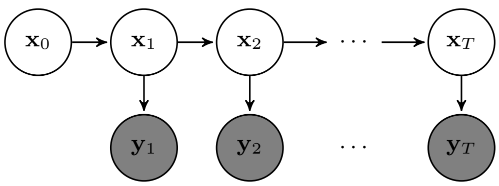

Figure 1. Pictorial representation of the conventional state space model.

The aim of Bayesian state estimation is to estimate the hidden target state over time using all available observations, i.e.,

$$\mathbf{x}_0 \longrightarrow \mathbf{x}_1|\mathbf{y}_1 \longrightarrow \ldots \longrightarrow \mathbf{x}_{T-1}|\mathbf{y}_{1:T-1} \longrightarrow \mathbf{x}_T|\mathbf{y}_{1:T} \tag{3}$$

In a probabilistic sense, we aim to obtain the pdfs

$$p(\mathbf{x}_0) \longrightarrow p(\mathbf{x}_1|\mathbf{y}_1) \longrightarrow \ldots \longrightarrow p(\mathbf{x}_T|\mathbf{y}_{1:T}) \tag{4}$$

That is, at time step t, we aim to obtain the posterior pdf $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ [15], [16].

Defn. II.1: Bayesian filtering

Lemma: If the pdf at time $t-1$, $p(\mathbf{x}_{t-1}|\mathbf{y}_{t:-1})$ is available, then the pdf at time t is given by

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|y_{1:t-1})\, d\mathbf{x}_{t-1}$$

Proof: In a Bayesian sense, the state space estimation (also called filtering) is two-fold, (a) prediction, and (b) update. If $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ is available, then this two-fold estimation is described as

$$\underbrace{p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})}_{\text{posterior at time } t-1} \longrightarrow \underbrace{p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}_{\text{prediction at time } t} \longrightarrow \underbrace{p(\mathbf{x}_t|\mathbf{y}_{1:t})}_{\text{updated posterior at time } t} \tag{5}$$

where the predicted pdf is derived as

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) &= \int p(\mathbf{x}_t, \mathbf{x}_{t-1}|y_{1:t-1})\, d\mathbf{x}_{t-1} \\ &\overset{(a)}{=} \int p(\mathbf{x}_t|\mathbf{x}_{t-1}, y_{1:t-1}) p(\mathbf{x}_{t-1}|y_{1:t-1})\, d\mathbf{x}_{t-1} \\ &\overset{(b)}{=} \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|y_{1:t-1})\, d\mathbf{x}_{t-1} \end{aligned} \tag{6}$$

where $\overset{(a)}{=}$ follows from $P(A,B) = P(A|B)P(B)$, $\overset{(b)}{=}$ follows because from the state transtion density we infer that $\mathbf{x}_t$ is independent of $\mathbf{y}_{1:t-1}$ given $\mathbf{x}_{t-1}$ as $\mathbf{x}_t \perp\!\!\!\perp \mathbf{y}_{1:t-1}|\mathbf{x}_{t-1}$.

The updated pdf is derived as

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{1:t}) &\overset{(a)}{=} \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_t)p(\mathbf{x}_t)}{p(\mathbf{y}_{1:t})} \\ &= \frac{p(\mathbf{y}_t, \mathbf{y}_{1:t-1}|\mathbf{x}_t)p(\mathbf{x}_t)}{p(\mathbf{y}_t, \mathbf{y}_{1:t-1})} \\ &\overset{(b)}{=} \frac{p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \mathbf{x}_t)p(\mathbf{y}_{1:t-1}|\mathbf{x}_t)p(\mathbf{x}_t)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})p(\mathbf{y}_{1:t-1})} \\ &\overset{(c)}{=} \frac{p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})p(\mathbf{y}_{1:t-1})p(\mathbf{x}_t)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})p(\mathbf{y}_{1:t-1})p(\mathbf{x}_t)} \\ &= \frac{p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \\ &\overset{(d)}{=} \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \end{aligned} \tag{7}$$

$$\overset{(e)}{\propto} p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \tag{8}$$

where $\overset{(a)}{=}$ follows from Bayes' rule $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$, $\overset{(b)}{=}$ follows from $P(A,B) = P(A|B)P(B)$, $\overset{(c)}{=}$ follows from applying the Bayes' rule on $p(\mathbf{y}_{1:t-1}|\mathbf{x}_t)$, $\overset{(d)}{=}$ follows because $\mathbf{y}_t \perp\!\!\!\perp \mathbf{y}_{1:t-1}|\mathbf{x}_t$ and $\overset{(e)}{\propto}$ follows by neglecting the normalising constant. The denominator in $\overset{(d)}{=}$ is the normalising constant and is given by

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t \tag{9}$$

In summary, if the posterior pdf at time $t-1$ is known as $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ then the posterior pdf at time t is obtained by substituting (6) and (9) in (7) as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|y_{1:t-1})\, d\mathbf{x}_{t-1}}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t} \tag{10}$$

$$= \frac{p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|y_{1:t-1})\, d\mathbf{x}_{t-1}}{\int_{\mathbf{x}_t, \mathbf{x}_{t-1}} p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|y_{1:t-1})\, d(\mathbf{x}_t, \mathbf{x}_{t-1})} \tag{11}$$

$$\propto p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|y_{1:t-1})\, d\mathbf{x}_{t-1} \tag{12}$$

The equation (12) which summarises the principle of Bayesian state estimation implies that to move from $t-1$ to t, we predict a hypothetical target state $\mathbf{x}_t$ at time t using the state transition pdf $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and then integrate out the previous target state $\mathbf{x}_{t-1}$. Once the prediction is available, the belief that the predicted hypothesis is close to the actual target state is then computed by the measurement pdf $p(\mathbf{y}_t|\mathbf{x}_t)$, and any correction in the hypothesis, if required, is applied therein. This sequential two-fold estimation forms the basis for the Bayesian filtering process.

Once the updated pdf is available, the hidden target state can be estimated using the well known expected *a posteriori* (EAP) [2] as

$$\hat{\mathbf{x}}_t^{\text{EAP}} = \mathbb{E}(p(\mathbf{x}_t|\mathbf{y}_{1:t})) = \int \mathbf{x}_t p(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t \tag{13}$$

or the maximum *a posteriori* (MAP) as

$$\hat{\mathbf{x}}_t^{\text{MAP}} = \arg\ \max_{\mathbf{x}} p(\mathbf{x}_t|\mathbf{y}_{1:t}) \tag{14}$$

In the next section, we present the derivation for the Kalman filter in the Bayesian recursion set up described in this section.

III. The Kalman filter

The Kalman filter, which is by far the most popular Bayesian estimation method, operates by assuming that the pdfs are Gaussian in nature and the underlying state space model is linear (or closely linear) [3], [16]. The filter is originally derived as a minimum mean square error estimator that minimises the distance between the predicted hypothesis and the measurement via a scaling parameter called the Kalman gain. This derivation is the most studied approach in the literature. In this section, we present a Kalman filter derivation that is purely Bayesian is its framework.

Consider a linear Gausian state space model described as

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{a}_t) = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}\mathbf{a}_t \tag{15}$$

$$\mathbf{y}_t = h(\mathbf{x}_t, \mathbf{e}_t) = \mathbf{H}\mathbf{x}_t + \mathbf{e}_t \tag{16}$$

where in (15), $\mathbf{F}$ is the state transition matrix, $\mathbf{G}$ is the control matrix and the state transition noise $\mathbf{a}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ is zero mean white Gaussian with covariance $\mathbf{Q}$. In (16), $\mathbf{H}$ is the matrix that translates the target from the state space to the observation space and the observation noise $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ is zero mean Gaussian with covariance $\mathbf{R}$. The noise vectors are assumed to be independent and identically distributed (i.i.d).

In the Kalman filter, the posterior pdfs in the Bayesian scheme in (5) are Gaussians and are denoted as [16]

$$\begin{aligned} p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) &= \mathcal{N}(\mathbb{E}(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}), \mathbb{V}(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})) \\ &= \mathcal{N}(\hat{\mathbf{x}}(t-1|t-1), \hat{\mathbf{S}}(t-1|t-1)) \end{aligned} \tag{17}$$

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) &= \mathcal{N}(\mathbb{E}(\mathbf{x}_t|\mathbf{y}_{1:t-1}), \mathbb{V}(\mathbf{x}_t|\mathbf{y}_{1:t-1})) \\ &= \mathcal{N}(\hat{\mathbf{x}}(t|t-1), \hat{\mathbf{S}}(t|t-1)) \end{aligned} \tag{18}$$

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{1:t}) &= \mathcal{N}(\mathbb{E}(\mathbf{x}_t|\mathbf{y}_{1:t}), \mathbb{V}(\mathbf{x}_t|\mathbf{y}_{1:t})) \\ &= \mathcal{N}(\hat{\mathbf{x}}(t|t), \hat{\mathbf{S}}(t|t)) \end{aligned} \tag{19}$$

where (17), (18) and (19) are our notations for the means and variances of the Gaussians. The Kalman filter hence can be treated as a sequential estimator that propagates the means and the covariances of the pdfs over time. That is, if the posterior at $t-1$ is known meaning that the expectation $\hat{\mathbf{x}}(t-1|t-1)$ and the covariance $\hat{\mathbf{S}}(t-1|t-1)$ estimates are known, then it is enough to compute the predicted and updated means and covariances.

Defn. III.1: Kalman filter prediction

Lemma: If at time $t-1$ the covariance $\hat{\mathbf{S}}(t-1|t-1)$ and expectation $\hat{\mathbf{x}}(t-1|t-1)$ are known, then the predicted covariance and expectation are given by

$$\hat{\mathbf{S}}(t|t-1) = \mathbf{F}\hat{\mathbf{S}}(t-1|t-1)\mathbf{F}^\top + \mathbf{G}\mathbf{Q}\mathbf{G}^\top$$
$$\hat{\mathbf{x}}(t|t-1) = \mathbf{F}\hat{\mathbf{x}}(t-1|t-1)$$

Proof: The predicted expectation is derived as follows,

$$\begin{aligned} \hat{\mathbf{x}}(t|t-1) &= \mathbb{E}\left(\mathbf{x}_t|\mathbf{y}_{1:t-1}\right) \\ &= \mathbb{E}\left((\mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}\mathbf{a}_t)|\mathbf{y}_{1:t-1}\right) \\ &\overset{(a)}{=} \mathbb{E}\left(\mathbf{F}\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right) + \mathbb{E}\left(\mathbf{G}\mathbf{a}_t|\mathbf{y}_{1:t-1}\right) \\ &\overset{(b)}{=} \mathbf{F}\ \mathbb{E}\left(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right) + \mathbf{G}\ \mathbb{E}\left(\mathbf{a}_t|\mathbf{y}_{1:t-1}\right) \\ &\overset{(c)}{=} \mathbf{F}\hat{\mathbf{x}}(t-1|t-1) + \mathbf{G}\ E\left(\mathbf{a}_t\right) \\ &\overset{(d)}{=} \mathbf{F}\hat{\mathbf{x}}(t-1|t-1) \end{aligned} \tag{20}$$

where $\overset{(a)}{=}$ and $\overset{(b)}{=}$ follows from the linearity property of expectation $\mathbb{E}(nA + mB) = n\mathbb{E}(A) + m\mathbb{E}(B)$, $\overset{(c)}{=}$ follows because $\mathbf{a}_t$ is independent and $\overset{(d)}{=}$ follows because $E(\mathbf{a}_t) = 0$. Then the predicted covariance is derived according to

$$\begin{aligned} \hat{\mathbf{S}}(t|t-1) &= \mathbb{V}\left(\mathbf{x}_t|\mathbf{y}_{1:t-1}\right) \\ &= \mathbb{V}\left((\mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}\mathbf{a}_t)|\mathbf{y}_{1:t-1}\right) \\ &\overset{(a)}{=} \mathbb{V}\left(\mathbf{F}\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right) + \mathbb{V}\left(\mathbf{G}\mathbf{a}_t|\mathbf{y}_{1:t-1}\right) \\ &\overset{(b)}{=} \mathbf{F}\ \mathbb{V}\left(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right)\ \mathbf{F}^\top + \mathbf{G}\ \mathbb{V}\left(\mathbf{a}_t|\mathbf{y}_{1:t-1}\right)\ \mathbf{G}^\top \\ &\overset{(c)}{=} \mathbf{F}\hat{\mathbf{S}}(t-1|t-1)\mathbf{F}^\top + \mathbf{G}\ \mathbb{V}\left(\mathbf{a}_t\right)\ \mathbf{G}^\top \\ &\overset{(d)}{=} \mathbf{F}\hat{\mathbf{S}}(t-1|t-1)\mathbf{F}^\top + \mathbf{G}\mathbf{Q}\mathbf{G}^\top \end{aligned} \tag{21}$$

where $\overset{(a)}{=}$ follows from the linearity property of the variance $\mathbb{V}(A + B) = \mathbb{V}(A) + \mathbb{V}(B)$, $\overset{(b)}{=}$ follows from the scaling property of variance $\mathbb{V}(mA) = m^2\mathbb{V}(A)$, $\overset{(c)}{=}$ follows because $\mathbf{a}_t$ is independent and $\overset{(d)}{=}$ follows because $V(\mathbf{a}_t) = Q$.

Defn. III.2: Kalman filter update

Lemma: If at time t the predicted covariance $\hat{\mathbf{S}}(t|t-1)$ and expectation $\hat{\mathbf{x}}(t|t-1)$ are known, then the updated covariance and expectation are given by

$$\hat{\mathbf{S}}(t|t) = (\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H} + \hat{\mathbf{S}}(t|t-1)^{-1})^{-1}$$
$$\hat{\mathbf{x}}(t|t) = \hat{\mathbf{S}}(t|t)(\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{y}_t + \hat{\mathbf{S}}(t|t-1)^{-1}\hat{\mathbf{x}}(t|t-1))$$

Proof: The update step follows from (8) and can be expressed as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \tag{22}$$
$$= \mathcal{N}(\mathbf{H}\mathbf{x}_t, \mathbf{R})\mathcal{N}(\hat{\mathbf{x}}(t|t-1), \hat{\mathbf{S}}(t|t-1)) \tag{23}$$

The measurement density in (23) can be expanded as

$$\begin{aligned}
&p(\mathbf{y}_t|\mathbf{x}_t)\\
&= \mathcal{N}(\mathbf{H}\mathbf{x}_t, \mathbf{R})\\
&= \exp\left(-\frac{1}{2}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)^\top \mathbf{R}^{-1}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)\right)\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{y}_t^\top \mathbf{R}^{-1}\mathbf{y}_t - \mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{y}_t - \mathbf{y}_t^\top \mathbf{R}^{-1}\\
&\qquad \mathbf{H}\mathbf{x}_t + \mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H}\mathbf{x}_t\bigg)\bigg)\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{y}_t^\top \mathbf{R}^{-1}\mathbf{y}_t - \mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{y}_t - \mathbf{x}_t^\top \mathbf{H}^\top\\
&\qquad \mathbf{R}^{-1}\mathbf{y}_t + \mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H}\mathbf{x}_t\bigg)\bigg)\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H}\mathbf{x}_t - 2\mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{y}_t\\
&\qquad + \mathbf{y}_t^\top \mathbf{R}^{-1}\mathbf{y}_t\bigg)\bigg)\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H}\mathbf{x}_t - 2\mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{y}_t + c\bigg)\bigg)
\end{aligned} \tag{24}$$

Similarly the predicted pdf in (23) can be expanded as

$$\begin{aligned}
&p(\mathbf{x}_t|\mathbf{y}_{1:t-1})\\
&= \mathcal{N}(\hat{\mathbf{x}}(t|t-1), \hat{\mathbf{S}}(t|t-1))\\
&= \exp\bigg(-\frac{1}{2}\bigg((\mathbf{x}_t - \hat{\mathbf{x}}(t|t-1))^\top \hat{\mathbf{S}}(t|t-1)^{-1}\\
&\qquad (\mathbf{x}_t - \hat{\mathbf{x}}(t|t-1)\bigg)\bigg)\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\mathbf{x}_t - \mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\\
&\qquad \hat{\mathbf{x}}(t|t-1) - \hat{\mathbf{x}}(t|t-1)^\top \hat{\mathbf{S}}(t|t-1)^{-1}\mathbf{x}_t\\
&\qquad + \hat{\mathbf{x}}(t|t-1)^\top \hat{\mathbf{S}}(t|t-1)^{-1}\hat{\mathbf{x}}(t|t-1)\bigg)\bigg)\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\mathbf{x}_t - \mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\\
&\qquad \hat{\mathbf{x}}(t|t-1) - \mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\hat{\mathbf{x}}(t|t-1)\\
&\qquad + \hat{\mathbf{x}}(t|t-1)^\top \hat{\mathbf{S}}(t|t-1)^{-1}\hat{\mathbf{x}}(t|t-1)\bigg)\bigg)\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\mathbf{x}_t - 2\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\\
&\qquad \hat{\mathbf{x}}(t|t-1) - \hat{\mathbf{x}}(t|t-1)^\top \hat{\mathbf{S}}(t|t-1)^{-1}\hat{\mathbf{x}}(t|t-1)\bigg)\bigg)\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\mathbf{x}_t - 2\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\\
&\qquad \hat{\mathbf{x}}(t|t-1) + c\bigg)\bigg)
\end{aligned} \tag{25}$$

By substituting (24) and (25) in (23) we obtain

$$\begin{aligned}
&p(\mathbf{x}_t|\mathbf{y}_{1:t})\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H}\mathbf{x}_t - 2\mathbf{x}_t^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{y}_t + c\bigg)\bigg)\times\\
&\quad \exp\bigg(-\frac{1}{2}\bigg(\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\mathbf{x}_t - 2\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t-1)^{-1}\hat{\mathbf{x}}(t|t-1) + c\bigg)\bigg)\\
&= \exp\bigg(-\frac{1}{2}\bigg(\mathbf{x}_t^\top (\mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H} + \hat{\mathbf{S}}(t|t-1)^{-1})\mathbf{x}_t\\
&\quad - 2\mathbf{x}_t^\top (\mathbf{H}^\top \mathbf{R}^{-1}\mathbf{y}_t + \hat{\mathbf{S}}(t|t-1)^{-1}\hat{\mathbf{x}}(t|t-1)) + c\bigg)\bigg)
\end{aligned} \tag{26}$$

Note that the constant terms in (25) and (26) that does not include the $\mathbf{x}_t$ term are grouped as some constant c. The whole idea of expanding the exponentials in (23) and rearranging the terms as in (24) and (25) to arrive at (26) is to express as a Gaussian pdf in terms of the random variable $\mathbf{x}_t$. By this, we can then compare with the expanded version of the updated pdf as

$$
\begin{aligned}
&p(\mathbf{x}_t|\mathbf{y}_{1:t})\\
&= \mathcal{N}(\hat{\mathbf{x}}(t|t), \hat{\mathbf{S}}(t|t))\\
&\propto \exp\Big(-\frac{1}{2}((\mathbf{x}_t - \hat{\mathbf{x}}(t|t))^\top \hat{\mathbf{S}}(t|t)^{-1}(\mathbf{x}_t - \hat{\mathbf{x}}(t|t))\Big)\\
&\propto \exp\Big(-\frac{1}{2}(\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t)^{-1}\mathbf{x}_t - \mathbf{x}_t^\top \hat{\mathbf{S}}(t|t)^{-1}\hat{\mathbf{x}}(t|t)\\
&\qquad - \hat{\mathbf{x}}(t|t)^\top \hat{\mathbf{S}}(t|t)^{-1}\mathbf{x}_t + \hat{\mathbf{x}}(t|t)^\top \hat{\mathbf{S}}(t|t)^{-1}\hat{\mathbf{x}}(t|t))\Big)\\
&\propto \exp\Big(-\frac{1}{2}(\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t)^{-1}\mathbf{x}_t - \mathbf{x}_t^\top \hat{\mathbf{S}}(t|t)^{-1}\hat{\mathbf{x}}(t|t)\\
&\qquad - \mathbf{x}_t^\top \hat{\mathbf{S}}(t|t)^{-1}\hat{\mathbf{x}}(t|t) + \hat{\mathbf{x}}(t|t)^\top \hat{\mathbf{S}}(t|t)^{-1}\hat{\mathbf{x}}(t|t))\Big)\\
&\propto \exp\Big(-\frac{1}{2}(\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t)^{-1}\mathbf{x}_t - 2\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t)^{-1}\hat{\mathbf{x}}(t|t)\\
&\qquad + \hat{\mathbf{x}}(t|t)^\top \hat{\mathbf{S}}(t|t)^{-1}\hat{\mathbf{x}}(t|t))\Big)\\
&\propto \exp\Big(-\frac{1}{2}(\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t)^{-1}\mathbf{x}_t - 2\mathbf{x}_t^\top \hat{\mathbf{S}}(t|t)^{-1}\hat{\mathbf{x}}(t|t) + c)\Big)
\end{aligned} \tag{27}
$$

By comparing (26) and (27) we observe that the coefficient of $\mathbf{x}_t^\top(.)\mathbf{x}_t$ is the inverse of the covariance. Therefore the updated covariance is

$$\hat{\mathbf{S}}(t|t) = (\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H} + \hat{\mathbf{S}}(t|t-1)^{-1})^{-1} \tag{28}$$

We also observe that the coefficient of $-2\mathbf{x}_t^\top(.)$ is the product of the precision matrix (inverse of the covariance) and the mean. Therefore the updated mean can be written as

$$\hat{\mathbf{x}}(t|t) = \hat{\mathbf{S}}(t|t)(\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{y}_t + \hat{\mathbf{S}}(t|t-1)^{-1}\hat{\mathbf{x}}(t|t-1)) \tag{29}$$

The Kalman filter, thus derived is the optimal Bayesian estimator for linear Gaussian hidden state space models because it gives a tractable solution to computing the pdfs.

IV. A SIMPLE SIMULATION EXAMPLE

In this section we present a simple 4-D simulation example of the Kalman filter. The target state is defined as $\mathbf{x}_t = (x_t, y_t, v_{x_t}, v_{y_t})^\top \in \mathbb{R}^4$ where the first two entries are the $x-y$ positions of the target and the next are the corresponding velocities. The target moves in the $x-y$ plane via constant velocity (CV) motion model given by

$$\mathbf{x}_t = \begin{bmatrix} 1 & 0 & \Delta t & 0\\ 0 & 1 & 0 & \Delta t\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\mathbf{x}_{t-1} + \begin{bmatrix} \Delta t^2/2 & 0\\ 0 & \Delta t^2/2\\ \Delta t & 0\\ 0 & \Delta t\end{bmatrix}\mathbf{a}_t$$

Here Δt is the interval between two observations and $\mathbf{a}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$. The observation model is

$$\mathbf{y}_t = \begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix}\mathbf{x}_t + \mathbf{e}_t$$

where the observation noise is $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$.

The Figure 2 shows the Kalman filter estimate for a $T = 100$ time step run and it can be observed that the filter locks the target fairly early and tracks it with good accuracy.

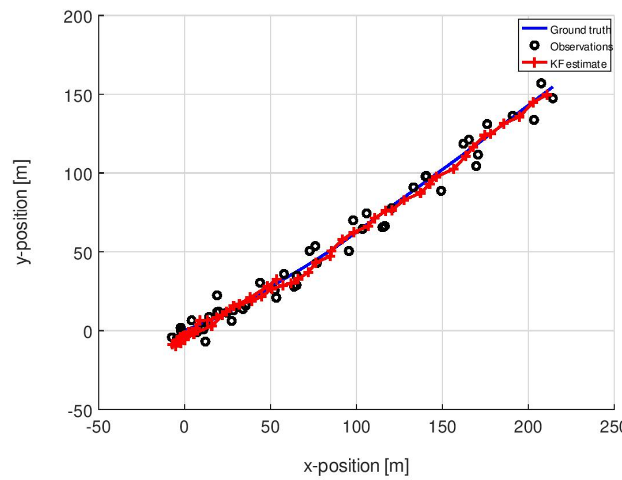

Figure 2. The demonstation of the Kalman filter for the linear Gaussian model.

A measure to test the Kalman filter is to verify if the estimated covariance converges. Figure 3 shows the estimated covariances for the position and velocity components and it can be observed that they indeed converge and thus indicating that the filter has converged to the true posterior.

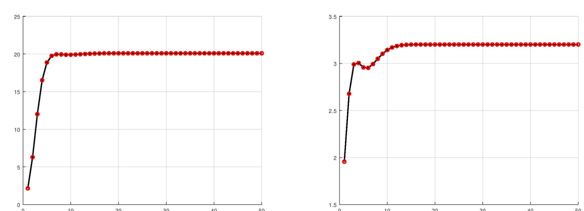

Figure 3. The left panel shows the covariance estimates $\sigma^2_{1,1}, \sigma^2_{2,2}$ corresponding to the position components. The left panel shows the covariance estimates $\sigma^2_{3,3}, \sigma^2_{4,4}$ corresponding to the velocity components.

V. CONCLUSION

In this paper, we derived the popular Kalman filter in a purely Bayesian perspective. The common dervations available in the literature to date are more focussed on presenting the Kalman filter as a MSE minimising filter. In this paper, we show the analogy of the Bayesian filter with the Kalman filter and present a detailed derivation for the latter. This derivation is not formally available as an article. Hence it can be regarded that this paper will be a good reading to obtain a deep Bayesian view of the Kalman filter. We also presented the MATLAB implementation and the simulation results for a simple 4-D linear Gaussian example.

References

[1] Bar-Shalom, Yaakov, X. Rong Li, and Thiagalingam Kirubarajan, "Estimation with applications to tracking and navigation: theory algorithms and software," John Wiley & Sons, 2004.

[2] R.P.S. Mahler, "Advances in statistical multisource-multitarget information fusion," Artech House, 2014.

[3] R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," ASME. J. Basic Eng., vol 82(1) pp. 35–45, 1960.

[4] Davey, Sam, Neil Gordon, Ian Holland, Mark Rutten, and Jason Williams, "Bayesian Methods in the Search for MH370," Springer (Nature) Briefs in Electrical and Computer Engineering, 2016.

[5] B. Anderson, and John B. Moore, "Optimal filtering," Courier Corporation, 2012.

[6] H. Sorenson, "Least-squares estimation: from Gauss to Kalman," IEEE spectrum, vol 7, no. 7 pp. 63–68, 1970.

[7] N. Shimkin, "Derivations of the Discrete-Time Kalman Filter," Lecture Notes, 2009.

[8] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," in Trans. ASME Ser. D, J. Basic Engineering, vol 83, pp. 95–107, 1961.

[9] R. E. Kalman, "New methods in Wiener filtering theory," in Proc. 1st Symp. on Engrg. Appl. of Random Function Theory and Probability, 1963.

[10] Anderson, Brian DO, and John B. Moore, "The Kalman-Bucy filter as a true time-varying Wiener filter," in IEEE Trans. on Systems, Man, and Cybernetics, vol 2, pp. 119–128, 1971.

[11] T. Kailath, "The innovations approach to detection and estimation theory," Proc. of the IEEE, vol 58, no. 5, pp. 680–695, 1970.

[12] Hamed Masnadi-Shirazi, Alireza Masnadi-Shirazi, Mohammad-Amir Dastgheib, "A Step by Step Mathematical Derivation and Tutorial on Kalman Filters ," *arXiv:1910.03558v1*, 2019.

[13] Yan-Xia Lin, "An alternative derivation of the Kalman filter using the quasi-likelihood method," Elsevier J. of Statistical Planning and Inference, vol 137, no. 5, pp. 1627–1633, 2007.

[14] J. Humpherys, and Jeremy West, "Kalman filtering with Newton's method," Proc. IEEE Control Systems Magazine, vol 30, no. 6, pp. 101–106, 2010.

[15] P. Choppala, "Bayesian multiple target tracking," PhD Thesis, Victoria University of Wellington, New Zealand, 2014.

[16] N. Gordon, B. Ristic, and S. Arulampalam, "Beyond the Kalman filter: Particle filters for tracking applications," Artech House, 2004.