Intrinsically Bayesian Robust Kalman Filter: An Innovation Process Approach

Roozbeh Dehghannasiri, Student Member, IEEE, Mohammad Shahrokh Esfahani, Member, IEEE, and Edward R. Dougherty, Fellow, IEEE

Abstract—In many contemporary engineering problems, model uncertainty is inherent because accurate system identification is virtually impossible owing to system complexity or lack of data on account of availability, time, or cost. The situation can be treated by assuming that the true model belongs to an uncertainty class of models. In this context, an intrinsically Bayesian robust (IBR) filter is one that is optimal relative to the cost function (in the classical sense) and the prior distribution over the uncertainty class (in the Bayesian sense). IBR filters have previously been found for both Wiener and granulometric morphological filtering. In this paper, we derive the IBR Kalman filter that performs optimally relative to an uncertainty class of state-space models. Introducing the notion of Bayesian innovation process and the Bayesian orthogonality principle, we show how the problem of designing an IBR Kalman filter can be reduced to a recursive system similar to the classical Kalman recursive equations, except with “effective” counterparts, such as the effective Kalman gain matrix. After deriving the recursive IBR Kalman equations for discrete time, we use the limiting method to obtain the IBR Kalman-Bucy equations for continuous time. Finally, we demonstrate the utility of the proposed framework for two real world problems: sensor networks and gene regulatory network inference.

Index Terms—Intrinsically Bayesian robust, Kalman filter, innovation process, orthogonality principle.

I. INTRODUCTION

Optimal filtering has been at the core of attention in signal processing community since the seminal works of Kolmogorov, Wold, Wiener, and Masani [1]; however, in practice, full knowledge of the necessary model statistics is often lacking. For instance, in the case of the Wiener filter it may be unwarranted to assume precise knowledge of the covariance structure relating to the observation and true signals. When faced with this kind of uncertainty, it is prudent to frame optimality in terms of an uncertainty class of signal models. This issue was first addressed during the late 1970s for Wiener filtering from a minimax perspective [2], [3].

Since its introduction in 1960, Kalman filter has been one of the most popular filtering approaches [4]. Classical Kalman filtering requires exact knowledge of noise statistics. However, in many practical applications, this condition does not always hold. Unknown noise covariance matrices can have a considerable impact on the performance of the Kalman filter. To address this issue, a number of Kalman filter design strategies have been proposed in the literature. One general approach when dealing with unknown noise covariances, called adaptive Kalman filtering, is to simultaneously estimate the noise covariances along with the state estimation [5]–[9]. Also, finite impulse response (FIR) counterparts of the Kalman filter, called “Kalman-like unbiased FIR filters”, have been proposed [10]–[13] and have been shown to be more robust than the ordinary Kalman filter when the noise covariances are not known or the noise is not white [10]. Another approach is to design a Kalman filter under the minimax criterion, the aim being to design a filter with the minimal upper bound error for all possible values of the noise covariances [14]–[16].

In this paper, we consider Kalman filtering when the statistical information regarding process and observation noise is missing by extending, via an innovation-process approach, the notion of Kalman filtering for a known model to the intrinsically Bayesian robust (IBR) Kalman filtering for an uncertainty class of models. The IBR concept was introduced in [17], where a fully optimal solution was found for a class of filtering problems that included linear filtering and nonlinear binary filtering. The setting is referred to as “Bayesian” because it assumes a prior distribution governing the unknown model parameters, which differs from the earlier minimax approaches taken for Wiener filtering. It is called “intrinsic” because optimization is over the entire class of filters rather than over the subclass of filters that are solutions to individual models in the uncertainty class, as were earlier constrained solutions for Wiener [18] and binary [19] filtering (see [17] for a more detailed discussion). The IBR terminology is employed in [20], where its structural advantage over classical Bayesian linear regression when there is additional data is demonstrated in the Gaussian model.

Perhaps the notion of innovation process was implicitly inferred from the Wold’s decomposition [1], where a (finite variance) discrete stochastic process $y_k$ was decomposed into two deterministic and random parts, while the random part was then named a moving average by Kolmogorov, who completed Wold’s work and found the best linear estimation of the process at time $k+1$ based on observation up to time $k$. The name “innovation process” was later coined by Wiener and Masani [1]. The seminal paper by Kailath on the innovation process [21] is the main inspiration of the approach in this paper. In [21], an innovation approach is proposed for the problem of least-squares estimation in additive white-noise. In a follow-up study [22], the problem of smoothing via an innovation approach, in a similar setting, was introduced.
Finally, the method in [23] generalized the case to nonlinear estimation (non-Gaussian processes).

To address the problem of designing an IBR Kalman filter, all definitions and properties of the original Kalman filtering need to be revisited. Utilizing the Bayesian orthogonality principle and the Bayesian innovation process, we design the IBR Kalman filter by extending concepts such as Kalman gain matrix, noise covariance matrices, and the estimation error covariance matrix to their “effective” counterparts. Using these effective matrices, the IBR Kalman filter for an uncertainty class of models can be expressed in a similar recursive form as the classical Kalman filter.

This paper is organized as follows. In Section II, we briefly review the original Kalman filter and the concept of intrinsically Bayesian robust filtering. In Section III, we introduce the Bayesian orthogonality principle, define the Bayesian innovation process, and derive the recursive equations for intrinsically Bayesian robust Kalman filtering. Then we extend the results to the continuous-time domain where the filter is called the intrinsically Bayesian robust Kalman-Bucy filter. Section IV contains two numerical examples to demonstrate the utility and the effectiveness of the proposed Kalman filtering framework. Finally, we conclude the paper in Section V.

II. BACKGROUND

Let us first summarize some notation employed in the paper. Lowercase and uppercase boldface letters denote vectors and matrices, respectively. For a matrix $M$, $m(i,j)$ denotes the element at row $i$ and column $j$. For a vector $v$, $v(i)$ denotes the $i$-th element of the vector. $\text{Tr}(M)$ and $M^T$ represent the trace and transpose operators, respectively. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Throughout the paper, expectation operator without subscript $E[\bullet]$ denotes the expectation of random variable (vector) relative to its corresponding probability measure $P$. Also, $\text{cov}[\bullet]$ represents the covariance matrix of a random vector. Finally, the over-dot operator represents the derivative taken with respect to the time index $t$, $\dot{x}(t) = \frac{dx(t)}{dt}$.

A. Kalman Filtering

A fundamental problem in engineering is the estimation of an unobserved multivariate random process $x_k$, an $n \times 1$ vector indexed by $k = 0, 1, 2, \ldots$ based on observing another random process $y_k$, an $m \times 1$ vector. Let $\psi_k(y)$ denote the estimate of $x_k$ as a function (filter) of $y_k$, $l \in \{1, \ldots, L\}$, $L$ being the window of observations used for the estimation. The performance of $\psi$ is evaluated via a cost function $\xi(\hat{x}_k, \psi_k(y))$, which is usually the mean-square error (MSE),

$$\xi(\hat{x}_k, \psi_k(y)) = E[(\hat{x}_k - \psi_k(y))^T(\hat{x}_k - \psi_k(y))].$$

(1)

We desire a filter $\psi$ that minimizes the cost function.

The Kalman approach [4] formulates the problem in a state-space model and provides the solution in the time domain. The state-space model contains a pair of equations, the state and observation equations. The state equation is defined as

$$x_{k+1} = F_k x_k + \Gamma_k u_k,$$

(2)

where $x_k$ is called the state vector, $F_k$ is an $n \times n$ matrix called the state transition matrix, $u_k$ is a $p \times 1$ vector called the process noise, and $\Gamma_k$ is an $n \times p$ matrix called the process noise transition matrix. The observation equation is defined as

$$y_k = H_k x_k + v_k.$$

(3)

In this model, $y_k$ is called the observation vector, $H_k$ is an $m \times n$ matrix called the observation transition matrix, and $v_k$ is an $m \times 1$ vector called the observation noise vector. We employ the notation $z_k = H_k x_k$. The process and observation noise $u_k$ and $v_k$ are zero-mean discrete white-noise processes with

$$E[u_k u_k^T] = Q_k \delta_{kl}, \quad \forall k, l = 0, 1, 2, \ldots$$

(4a)

$$E[v_k v_k^T] = R_k \delta_{kl}, \quad \forall k, l = 0, 1, 2, \ldots$$

(4b)

$$E[v_k x_k^T] = 0_{m \times n}, \quad \forall k, l = 0, 1, 2, \ldots$$

(4c)

$$E[u_k v_k^T] = 0_{p \times m}, \quad \forall k, l = 0, 1, 2, \ldots$$

(4d)

$$E[u_k y_k^T] = 0_{p \times m}, \quad 0 \leq l \leq k,$$

(4e)

where $0_{m \times n}$ is a zero matrix of size $m \times n$. The aim is to find an estimate $\hat{x}_k$ as a function of the observations $y_l$, $l \leq k - 1$, such that $E[(x_k - \hat{x}_k)^T(x_k - \hat{x}_k)]$ is minimized. This estimate is called the least-squares estimate for $x_k$.

In [21], the notion of an innovation process, $\tilde{z}_k$, has been introduced to derive the Kalman filter recursive equations for the state-space model in (2) and (3) as follows:

$$\tilde{z}_k = y_k - \tilde{z}_k,$$

(5a)

$$K_k = P_k^* H_k^T \left( P_k^* + R_k \right)^{-1},$$

(5b)

$$\tilde{x}_{k+1} = \Phi_k \tilde{x}_k + \Phi_k K_k \tilde{z}_k,$$

(5c)

$$P_{k+1} = (\Phi_k - \Phi_k K_k H_k) P_k^* \Phi_k^T + \Gamma_k Q_k \Gamma_k^T.$$

(5d)

The innovation process $\tilde{z}_k$ obtained in (5a) is a discrete white-noise process and is used in (5c) to find the update for the least-squares estimate $\hat{x}_k$. The matrix $K_k$ in (5b) is called the Kalman gain matrix. The matrix $P_k^*$ in (5d) is the optimal estimation error covariance matrix at time $k$, $P_k^* = E[(x_k - \tilde{x}_k)^T(x_k - \tilde{x}_k)]$. Note that in (5a), $\tilde{z}_k = H_k \tilde{x}_k$ and in (5b), $P_k^*$ is the optimal estimation error covariance matrix relative to $z_k$, $P_k^* = E[(z_k - \tilde{z}_k)^T(z_k - \tilde{z}_k)]$. It can be easily seen that $P_k^* = H_k P_k^* H_k^T$. (5d) is the discrete Riccati recursion [24]. For time-invariant models, $P_k^*$ in (5d) converges to a steady-state value $P_\infty$ if the model is asymptotically stable, meaning that all eigenvalues of $\Phi_k$ have a norm smaller than one [25]. It should be recognized that when noise processes $u_k$ and $v_k$ are white and Gaussian, the Kalman filter is optimal in the MSE sense among all linear and non-linear filters. Otherwise, the Kalman filter is the optimal linear filter.

For the continuous-time domain, where the filter is called the Kalman-Bucy filter [26], the state and observation equations in (2) and (3) become

$$\dot{x}_t = F x_t + \Gamma u_t, \quad t \in [a, b],$$

(6)

$$y_t = H x_t + v_t, \quad t \in [a, b],$$

(7)
where \( E[u_k u_k^T] = Q_\delta(t - s) \) and \( E[v_k v_k^T] = R_\delta(t - s) \) for \( t, s \in [a, b] \). Analogous to the discrete-time equations, the continuous-time recursive equations are [21], [26]

\[
\tilde{z}_t = y_t - \tilde{x}_t, \\
K_t = P_t^x H_t^{-1}, \\
\tilde{x}_t = F_t \tilde{x}_t + K_t \tilde{z}_t, \\
P_t^x = F_t P_t^x + P_t^x F_t^T - K_t R_t K_t^T + \Gamma_t Q_2 \Gamma_t^T.
\]

B. Intrinsically Bayesian Robust Filtering

When it is not realistic to assume that all parameters of the system model are known, it is prudent to design a robust filter by taking into account all possible models consistent with the partial prior knowledge. Qualitatively, a filter is robust if its performance degradation is acceptable for models close to the model for which it has been designed. Assume that the model is parameterized by parameter \( \theta \in \Theta \), \( \Theta \) being the set of all possible parameters and is called an uncertainty class. Robust filtering has been treated from two different perspectives: minimax and Bayesian. In minimax robustness, the aim is to find a filter that has the best worst-case performance. The minimax criterion has been used in [3], [27]–[29] to design robust linear filters, such as robust Kalman, robust Wiener, and robust matched filters. Minimax is highly conservative and tends to give too much weight to models with low likelihood.

A Bayesian robust filter is designed to achieve the best average performance across the uncertainty class. This approach utilizes a prior distribution \( \pi(\theta) \) over \( \Theta \). Bayesian robustness has been used in different engineering applications such as filtering (Wiener [17], binary [19], granulometric [30]), classification [31], signal compression [32], blind image deconvolution [33], and Bayesian hypothesis testing [34].

In particular, the concept of intrinsically Bayesian robust (IBR) filtering has been introduced in [17]. An IBR filter minimizes the average MSE across the uncertainty class \( \Theta \):

\[
\psi(\theta) = \arg \min_{\psi \in \Psi} E_\theta \left[ \xi(\theta, \Psi(x_k, \psi_k(y))) \right],
\]

where the cost function \( \xi(\cdot, \cdot) \) is computed relative to model \( \theta \), \( \Psi \) is the class of all filters to which the optimization is restricted, and the expectation is taken relative to the prior distribution \( \pi(\theta) \) over \( \Theta \). In [17], the notion of effective characteristics is proposed to find an IBR Wiener filter. The basic idea is to formulate the optimization problem in terms of effective characteristics derived from the underlying random process. Characteristics are functions of the processes under study that determine the filter cost function \( \xi \). In Wiener filtering, the characteristics are the auto- and cross-correlation functions between processes. Using effective characteristics, the average cost of the Wiener filter relative to the uncertainty class is equivalent to the error of applying that filter to the effective characteristics. Thus, designing an IBR Wiener filter for an uncertainty class reduces to designing a Wiener filter for the effective characteristics and the IBR Wiener filter can be found in the same way that the Wiener filter is found for a known model [17].

In this paper, our aim is to find an IBR Kalman filter that shares conceptual similarities with the IBR Wiener filter in [17]. To do so, we state and prove the Bayesian orthogonality principle, define a Bayesian innovation process, and define the effective Kalman gain matrix.

III. IBR Kalman Filter

In this section, first we introduce the notion of Bayesian innovation process as a means to derive recursive equations for the IBR Kalman filter and then derive recursive equations for the discrete-time domain. The equations for the IBR Kalman-Bucy filter are then found via limits of the discrete-time equations.

A. Kalman Filter

Throughout this paper, we assume that the covariance matrices of the process and observation noise are unknown. Let \( \theta_1 \) and \( \theta_2 \) be two unknown parameters for the process and observation noise covariance matrices, respectively:

\[
\begin{align*}
E[x_k^T (u_k^T \pi(\theta))] &= Q_{k1} \delta_{kl}, \\
E[v_k^T (v_k^T \pi(\theta))] &= R_{k2} \delta_{kl},
\end{align*}
\]

where equality with sign \( \pi(\theta) \) implies equality in probability. The state-space model is completely determined by \( \theta = [\theta_1, \theta_2] \in \Theta \), the uncertainty class. Let \( \pi(\theta) \) be a prior distribution governing \( \Theta \). Assuming \( \theta_1 \) and \( \theta_2 \) are independent, i.e., \( \pi(\theta) = \pi(\theta_1) \pi(\theta_2) \), the state-space model is parameterized as

\[
\begin{align*}
x_{k+1}^\theta &= \Phi_k x_k + G_k u_k, \\
y_k^\theta &= H_k x_k + v_k.
\end{align*}
\]

While the state \( x_k^\theta \) depends only on the parameter \( \theta_1 \) of the process noise, the observation \( y_k^\theta \) depends on both the process noise parameter \( \theta_1 \) and the observation noise parameter \( \theta_2 \).

Definition 1: Let \( y_l^\theta = H_k x_k + v_k \), where \( \theta = [\theta_1, \theta_2] \in \Theta \) is the unknown parameter. Assume that random processes \( x_k^\theta \) and \( y_k^\theta \) are vectors of size \( n \times 1 \) and \( m \times 1 \), respectively. Let \( \mathcal{G} \) be the vector space of all \( n \times m \) matrix-valued functions \( G_{k,l} \), where \( G_k, l \in \mathcal{G} \) is a mapping \( G_{k,l} : N \times N \rightarrow \mathbb{R}^{n \times m} \) such that \( \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} ||G_{k,l}||_2 < \infty \), where \( ||\cdot||_2 \) is the \( L_2 \) norm. The Bayesian least-squares estimate at time \( k \) given observations \( y_l^\theta, l \leq k - 1 \), is given by

\[
\hat{x}_k^\theta = \sum_{l=k-1}^\infty G_{k,l}^\theta y_l,
\]

where

\[
G_{k,l}^\theta = \arg \min_{G_{k,l} \in \mathcal{G}} E_\theta \left[ \left( x_k^\theta - \sum_{l=k-1}^\infty G_{k,l} y_l^\theta \right)^T \right.
\times \left. \left( x_k^\theta - \sum_{l=k-1}^\infty G_{k,l} y_l^\theta \right) \right].
\]
mind that $x_{k}^{\theta_{i}}$ depends only on $\theta_{i}$, whereas $\tilde{x}_{k}^{\theta}$ depends on $\theta = [\theta_{1}, \theta_{2}]$. We call it the “Bayesian orthogonality principle” owing to the fact that the outer expectation is relative to the distribution over the uncertainty class.

Theorem 1: (Bayesian Orthogonality Principle) A linear filter with weighting function $G_{k,l}^{\Theta}$, as defined in (14) satisfies (15) (having minimum average MSE across the uncertainty class) if and only if

$$E_{\theta}[E[(x_{k}^{\theta} - \tilde{x}_{k}^{\theta}) (y_{l}^{\theta})^{T}]] = 0_{n \times m}, \quad \forall l \leq k - 1 \quad (16)$$

or, with regard to the components of the matrices,

$$E_{\theta}[E[(x_{k}^{\theta}(i) - \tilde{x}_{k}^{\theta}(i)) y_{l}^{\theta}(j)]] = 0, \quad \forall 1 \leq i \leq n, \quad 1 \leq j \leq m. \quad (17)$$

where $E[\cdot]$ denotes the expectation relative to the probability measure $P$.

Filtering problem can be facilitated by whitening observations. In the classical Kalman filter, this intuition is used to define the innovation process [21]. We extend this notion to the Bayesian innovation process for IBK Kalman filtering.

Definition 2: If $\tilde{x}_{k}^{\theta}$ is the Bayesian least-squares estimate at time $k$ for $x_{k}^{\theta_{i}}$, then

$$\tilde{x}_{k}^{\theta} = y_{k}^{\theta} - H_{k} \tilde{x}_{k}^{\theta}, \quad (18)$$

is a zero-mean process called the Bayesian innovation process.

For the classical Kalman filter, the covariance of the innovation process is $E[\tilde{z}_{k} \tilde{z}_{l}^{T}] = (P_{k}^{z} + R_{k}) \delta_{kl} [21]$. The following proposition extends this to the Bayesian innovation process.

Proposition 1: Let $\tilde{z}_{k}^{\theta}$ be the Bayesian innovation process as defined in (18). Then

$$E_{\theta}[E[\tilde{x}_{k}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] = E_{\theta}[P_{l}^{z,\theta} + R_{l}^{\theta}] \delta_{kl}, \quad (19)$$

where $P_{k}^{z,\theta} = E[(x_{k}^{\theta} - \tilde{x}_{k}^{\theta}) (x_{k}^{\theta} - \tilde{x}_{k}^{\theta})^{T}]$. Note that $\tilde{z}_{k}^{\theta} = H_{k} \tilde{x}_{k}^{\theta}$.

**Proof:** Please refer to Appendix A.

Our aim is to find the Bayesian least-squares estimate $\hat{x}_{k}^{\theta}$ for $x_{k}^{\theta}$ based on the observations $y_{l}^{\theta}$ up to time $k - 1$. The following lemma helps us find the Bayesian least-squares estimate using the Bayesian innovation process.

**Lemma 1:** The Bayesian least-squares estimate for $x_{k}^{\theta}$ using the Bayesian innovation process $\tilde{z}_{l}^{\theta}$, $l \leq k - 1$, is the Bayesian least-squares estimate for $x_{k}^{\theta}$ based upon observations $y_{l}^{\theta}$, $l \leq k - 1$.

**Proof:** Please refer to Appendix B.

We next derive recursive equations for the IBK Kalman filter. According to the lemma, $\tilde{x}_{k}^{\theta}$ can be found using $\tilde{z}_{l}^{\theta}$. Since $\tilde{z}_{l}^{\theta}$ possesses property (19), it simplifies the equations. We can find the Bayesian least-squares estimate $\hat{x}_{k}^{\theta}$ via a linear operator of the form

$$\hat{x}_{k}^{\theta} = \sum_{l \leq k - 1} G_{k,l}^{\Theta} \tilde{z}_{l}^{\theta}. \quad (20)$$

As $G_{k,l}^{\Theta}$ should satisfy the Bayesian orthogonality principle, for $l \leq k - 1$,

$$0_{n \times m} = E_{\theta}[E[x_{k}^{\theta} (x_{l}^{\theta} - \tilde{x}_{k}^{\theta}) (\tilde{z}_{l}^{\theta})^{T}]] = E_{\theta}[E[x_{k}^{\theta} - \sum_{i \leq k - 1} G_{k,i}^{\Theta} \tilde{z}_{i}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] = E_{\theta}[\sum_{i \leq k - 1} G_{k,i}^{\Theta} \tilde{z}_{i}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]. \quad (21)$$

Hence, for $l \leq k - 1$,

$$E_{\theta}[E[x_{k}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] = E_{\theta}[E[\sum_{i \leq k - 1} G_{k,i}^{\Theta} \tilde{z}_{i}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] = \sum_{i \leq k - 1} G_{k,i}^{\Theta} E_{\theta}[E[\tilde{z}_{i}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]]. \quad (22)$$

Using (19), (22) reduces to

$$E_{\theta}[E[x_{k}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] = \sum_{i \leq k - 1} G_{k,i}^{\Theta} E_{\theta}[P_{l}^{z,\theta} + R_{l}^{\theta}] \delta_{li} = G_{k,l}^{\Theta} E_{\theta}[P_{l}^{z,\theta} + R_{l}^{\theta}]. \quad (23)$$

Rearranging (23) yields

$$G_{k,l}^{\Theta} = E_{\theta}[E[x_{k}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] E_{\theta}^{-1}[P_{l}^{z,\theta} + R_{l}^{\theta}]. \quad (24)$$

We compute (20) by plugging in (24):

$$\hat{x}_{k}^{\theta} = \sum_{l \leq k - 1} E_{\theta}[E[x_{k}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] E_{\theta}^{-1}[P_{l}^{z,\theta} + R_{l}^{\theta}] \tilde{z}_{l}^{\theta}. \quad (25)$$

To investigate the existence of the inverse of $E_{\theta}[P_{l}^{z,\theta} + R_{l}^{\theta}]$, we know that $E_{\theta}[P_{l}^{z,\theta}]$ is a positive semi-definite matrix. A sufficient condition that $E_{\theta}^{-1}[P_{l}^{z,\theta} + R_{l}^{\theta}]$ always exists is that $E_{\theta}[R_{l}^{\theta}]$ is a positive definite matrix. This means that the observation of all components of the state vector is noisy.

An update equation for the Bayesian least-squares estimate can be found as

$$\hat{x}_{k+1}^{\theta} = \sum_{l \leq k} E_{\theta}[E[x_{k+1}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] E_{\theta}^{-1}[P_{l}^{z,\theta} + R_{l}^{\theta}] \tilde{z}_{l}^{\theta}$$

$$= \sum_{l \leq k} \delta_{k+1,l} E_{\theta}[E[x_{k+1}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] E_{\theta}^{-1}[P_{l}^{z,\theta} + R_{l}^{\theta}] \tilde{z}_{l}^{\theta}$$

$$= \sum_{l \leq k} \Phi_{k,l} E_{\theta}[E[x_{k+1}^{\theta} (\tilde{z}_{l}^{\theta})^{T}]] E_{\theta}^{-1}[P_{l}^{z,\theta} + R_{l}^{\theta}] \tilde{z}_{l}^{\theta} + \Phi_{k} K_{k}^{\Theta} \tilde{z}_{k}^{\theta}$$

$$= \Phi_{k} \hat{x}_{k}^{\theta} + \Phi_{k} K_{k}^{\Theta} \tilde{z}_{k}^{\theta}, \quad (26)$$

where we used (12) and the fact that $\tilde{z}_{k}^{\theta}$ is independent from $u_{l}^{\theta}$ for $l \leq k$ to get the second equality, in the last line $\hat{x}_{k}^{\theta}$ is obtained using (25) and

$$K_{k}^{\Theta} = E_{\theta}[E[x_{k+1}^{\theta} (\tilde{z}_{k}^{\theta})^{T}]] E_{\theta}^{-1}[P_{k}^{z,\theta} + R_{k}^{\theta}], \quad (27)$$

is called the effective Kalman gain matrix. We wish to reformulate the effective Kalman gain matrix. First, $E_{\theta}[E[x_{k+1}^{\theta} (\tilde{z}_{k}^{\theta})^{T}]]$

$$= E_{\theta}[E[x_{k+1}^{\theta} (x_{k}^{\theta})^{T} H_{k}^{T} + (v_{k}^{\theta})^{T} - (\tilde{x}_{k}^{\theta})^{T} H_{k}^{T}]] = E_{\theta}[E[x_{k+1}^{\theta} (x_{k}^{\theta})^{T}]] H_{k}^{T}, \quad (28)$$
where (4c) is used to obtain the second equality and $\mathbf{x}_k^{c,\theta} = \mathbf{x}_k^\theta - \mathbf{x}_k^{\theta}$ is called the Bayesian least-squares estimation error at time $k$. Then
\[
E_\theta \left[ E \left[ \mathbf{x}_k^{\theta} \left( \mathbf{x}_k^{c,\theta} \right)^T \right] \right] = E_\theta \left[ E \left[ \left( \mathbf{x}_k^{\theta} + \mathbf{x}_k^{c,\theta} \right) \left( \mathbf{x}_k^{c,\theta} \right)^T \right] \right] = E_\theta \left[ E \left[ \mathbf{x}_k^{c,\theta} \left( \mathbf{x}_k^{c,\theta} \right)^T \right] \right] = E_\theta \left[ \mathbf{P}_k^{c,\theta} \right],
\]  
(29)

where $\mathbf{P}_k^{c,\theta}$ denotes the Bayesian estimation error covariance matrix relative to $\theta$ at time $k$. Using (29) in (28) gives the effective Kalman gain matrix as
\[
\mathbf{K}_k^\theta = E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] \mathbf{H}_k^T \mathbf{E}_\theta^{-1} \left[ \mathbf{P}_k^{x,\theta} + \mathbf{R}_k^{\theta_2} \right].
\]  
(30)

Next we find the update equation for $\mathbf{x}_k^{c,\theta}$:
\[
\mathbf{x}_k^{c,\theta} = \mathbf{x}_k^{\theta} - \mathbf{K}_k^\theta \mathbf{H}_k \mathbf{P}_k^{x,\theta} \left( \mathbf{I} - \mathbf{K}_k^\theta \mathbf{H}_k \right)^T \Phi_k^T + \Gamma_k \mathbf{Q}_k^\theta \Gamma_k^T + \Phi_k \mathbf{K}_k^\theta \mathbf{R}_k^{\theta_2} \left( \mathbf{K}_k^\theta \right)^T \Phi_k^T
\]  
(31)

where we used (12) and (26) to obtain the second equality, and used (13) and (18) to get the third equality. Letting $\mathbf{P}_k^{x,\theta}$ denote the set of unknown parameters of $\mathbf{x}_k^{c,\theta}$, we can obtain
\[
\mathbf{P}_k^{x,\theta} = \Phi_k \left( \mathbf{I} - \mathbf{K}_k^\theta \mathbf{H}_k \right) \mathbf{P}_k^{x,\theta} \left( \mathbf{I} - \mathbf{K}_k^\theta \mathbf{H}_k \right)^T \Phi_k^T + \Gamma_k \mathbf{Q}_k^\theta \Gamma_k^T + \Phi_k \mathbf{K}_k^\theta \mathbf{R}_k^{\theta_2} \left( \mathbf{K}_k^\theta \right)^T \Phi_k^T
\]  
(32)

Taking the expectations of both sides in (32) yields
\[
E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] = \Phi_k \left( \mathbf{I} - \mathbf{K}_k^\theta \mathbf{H}_k \right) E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] \left( \mathbf{I} - \mathbf{K}_k^\theta \mathbf{H}_k \right)^T \Phi_k^T + \Gamma_k E_\theta \left[ \mathbf{Q}_k^\theta \right] \Gamma_k^T + \Phi_k \mathbf{K}_k^\theta \mathbf{R}_k^{\theta_2} \left( \mathbf{K}_k^\theta \right)^T \Phi_k^T = \Phi_k \left( \mathbf{I} - \mathbf{K}_k^\theta \mathbf{H}_k \right) E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] \Phi_k^T + \Gamma_k E_\theta \left[ \mathbf{Q}_k^\theta \right] \Gamma_k^T
\]  
(33)

where the second equality results from
\[
\mathbf{K}_k^\theta \mathbf{H}_k E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] \Phi_k^T = E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] \mathbf{H}_k^T = \Phi_k \left( \mathbf{I} - \mathbf{K}_k^\theta \mathbf{H}_k \right) E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] \mathbf{H}_k^T - \mathbf{K}_k^\theta E_\theta \left[ \mathbf{R}_k^{\theta_2} \right].
\]  
(34)

which can be found by rearranging the equation for the effective Kalman gain matrix in (30).

The following equations summarize the IBR Kalman filtering:
\[
\dot{\mathbf{z}}_k^\theta = \mathbf{y}_k^\theta - \mathbf{z}_k^\theta,
\]  
(35a)
\[
\mathbf{K}_k^\theta = E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] \mathbf{H}_k^T E_\theta^{-1} \left[ \mathbf{P}_k^{x,\theta} + \mathbf{R}_k^{\theta_2} \right],
\]  
(35b)
\[
\dot{\mathbf{x}}_{k+1} = \Phi_k \mathbf{x}_k + \mathbf{z}_k, 
\]  
(35c)
\[
E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] = \Phi_k \left( \mathbf{I} - \mathbf{K}_k^\theta \mathbf{H}_k \right) E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] \Phi_k^T + \Gamma_k E_\theta \left[ \mathbf{Q}_k^\theta \right] \Gamma_k^T
\]  
(35d)

The structure of these equations is similar to (5a), (5b), and (5d) for the classical Kalman filter, where $\mathbf{K}_k^\theta$, $\mathbf{P}_k^{x,\theta}$, $\mathbf{Q}_k$, and $\mathbf{R}_k$ are replaced by $\mathbf{K}_k^\theta$, $E_\theta \left[ \mathbf{P}_k^{x,\theta} \right]$, $E_\theta \left[ \mathbf{Q}_k^\theta \right]$, and $E_\theta \left[ \mathbf{R}_k^{\theta_2} \right]$, respectively. This correspondence, in which the gain matrix is replaced by the effective Kalman gain matrix is analogous to the IBR Wiener filtering, where the power spectrum is replaced by the effective power spectrum and for

IBR granulometric (morphological) filtering the granulometric spectrum is replaced by the effective granulometric spectrum [17]. We should point out that since the structure of the proposed IBR Kalman filter is completely similar to that of the ordinary Kalman filter, except the use of effective quantities, its complexity is comparable (almost the same) as that of the ordinary Kalman filter. A block diagram of the proposed IBR Kalman filtering method is given in Figure 1. Initial conditions for the IBR Kalman filter can be set to $E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] = \text{cov}[\mathbf{x}_0]$ and $\mathbf{x}_0 = E \left[ \mathbf{x}_0 \right]$. Note that $\mathbf{x}_0$ does not depend on $\theta$ because it is independent from future process and observation noise.

It should be noted that one can assume that the process noise transition matrix $\Gamma_k$ is also unknown, being parameterized by $\theta_3$. It is easily seen that (35a)-(35c) can be used as the recursive equations of the IBR Kalman filter and that $E_\theta \left[ \mathbf{P}_k^{x,\theta} \right]$ will be updated as
\[
E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] = \Phi_k \left( \mathbf{I} - \mathbf{K}_k^\theta \mathbf{H}_k \right) E_\theta \left[ \mathbf{P}_k^{x,\theta} \right] \Phi_k^T + \mathbf{E}_\theta \left[ \Gamma_k^\theta \mathbf{Q}_k^\theta \left( \Gamma_k^\theta \right)^T \right] + \mathbf{E}_\theta \left[ \Gamma_k^\theta \mathbf{E}_\theta \left[ \mathbf{Q}_k^\theta \right] \Gamma_k^T \right].
\]  
(36)

where we have used the independence assumption on $\theta_1$ and $\theta_3$ in the second equality. Because the effect of $\Gamma_k$ is only in the covariance matrix of the process noise, it can be viewed as part of the noise covariance, in the sense that $\Gamma_k^\theta \mathbf{Q}_k^\theta \left( \Gamma_k^\theta \right)^T$ can be viewed collectively as the covariance matrix of the process noise. Hence, one could employ a parameter $\theta'$ to denote the set of unknown parameters of $\mathbf{Q}_k$ and $\Gamma_k$ together.

B. Using Current Observation for IBR Kalman Filtering

In this subsection, we consider the case that the observation $\mathbf{y}_k^\theta$ is also utilized for estimating $\mathbf{x}_k^\theta$. To avoid confusion, we
use notations $\hat{x}_k^θ$ and $\tilde{x}_k^θ$ to distinguish between estimates with and without using the current observation. We also use $P_{k|k}^x = \text{cov}[\hat{x}_{k|k}^θ - \tilde{x}_{k|k}^θ]$ and $P_{k}^x = \text{cov}[\hat{x}_{k}^θ - \tilde{x}_{k}^θ]$.

With complete knowledge of the noise covariance matrices, the recursive equations for the ordinary Kalman filter are

\begin{align*}
    \hat{x}_{k+1|k+1} &= \hat{x}_{k+1} + K_{k+1} \tilde{z}_{k+1}, \\
    P_{k+1|k+1} &= (I - K_{k+1}H_{k+1})P_{k+1|k}^x.
\end{align*}

(37) (38)

If the state-space model is parameterized by the unknown parameter $\theta = [\theta_1, \theta_2]$ according to (12) and (13), then we desire the recursive equations for IBR Kalman filtering when the current observation $y_k^θ$ is also used for estimation at time $k$. In this regard, (20) becomes

\begin{equation}
    \hat{x}_{k|k}^θ = \sum_{l \leq k} G_{k|l}^θ \tilde{z}_l^θ.
\end{equation}

(39)

Following similar steps as in (21)-(24) but for $l \leq k$ yields

\begin{equation}
    \hat{x}_{k|k}^θ = \sum_{l \leq k} E_{\theta} \left[ E \left[ x_{k|l}^θ (z_l^θ)^T \right] \right] E_{\theta}^{-1} \left[ P_{l|l}^x + R_l^2 \right] \tilde{z}_l^θ.
\end{equation}

(40)

Comparing (40) with the second equality in (26), it is seen that $\hat{x}_{k|k}^θ = \Phi_k \hat{x}_{k|k}^θ$. Now, similar to (26), we find the update equation for $\hat{x}_{k|k}^θ$:

\begin{align*}
    \hat{x}_{k+1|k+1}^θ =& \sum_{l \leq k} E_{\theta} \left[ E \left[ x_{k+1|l}^θ (z_l^θ)^T \right] \right] E_{\theta}^{-1} \left[ P_{l+1|l}^x + R_l^2 \right] \tilde{z}_l^θ \\
    =& \sum_{l \leq k} E_{\theta} \left[ E \left[ x_{k+1|l}^θ (z_l^θ)^T \right] \right] E_{\theta}^{-1} \left[ P_{l+1|l}^x + R_l^2 \right] \tilde{z}_l^θ \\
    &+ E_{\theta} \left[ E \left[ x_{k+1|l}^θ (z_l^θ)^T \right] \right] E_{\theta}^{-1} \left[ P_{l+1|l}^x + R_l^2 \right] \tilde{z}_l^θ \\
    =& \Phi_k \hat{x}_{k|k}^θ + K_{k+1} \tilde{z}_{k+1} \\
    =& \hat{x}_{k|k}^θ + K_{k+1} \tilde{z}_{k+1}.
\end{align*}

(41)

We also wish to find the update equation for $E_{\theta} [P_{k|k}^x]$. To do so, we can see that

\begin{align*}
    x_{k+1|k+1}^θ = x_{k+1}^θ - \hat{x}_{k|k+1}^θ \\
    x_{k+1}^θ = x_{k+1}^θ - \hat{x}_{k|k}^θ \\
    x_{k+1}^θ = x_{k+1}^θ - K_{k+1} (y_{k+1}^θ - H_{k+1} \hat{x}_{k|k}^θ) \\
    x_{k+1}^θ = (I - K_{k+1}H_{k+1}) x_{k+1}^θ + K_{k+1} \tilde{z}_{k+1} \\
    x_{k+1}^θ = (I - K_{k+1}H_{k+1}) x_{k+1}^θ + K_{k+1} \tilde{z}_{k+1}.
\end{align*}

(42)

Therefore, using (42) we have

\begin{align*}
    E_{\theta} [P_{k|k+1}^x] &= E_{\theta} \left[ E \left[ x_{k+1|k+1}^θ (x_{k+1|k+1}^θ)^T \right] \right] \\
    &= (I - K_{k+1}H_{k+1}) E_{\theta} \left[ P_{k+1|k}^x \right] (I - K_{k+1}H_{k+1})^T \\
    &+ K_{k+1} \tilde{z}_{k+1} (K_{k+1} \tilde{z}_{k+1})^T.
\end{align*}

(43)

where we use (34) to obtain the last equality. We see that (41) and (43) are similar to those for the classical Kalman filter (when the current observation is utilized) in (37) and (38) except that we use $E_{\theta} [P_{k|k+1}^x]$ and the effective Kalman gain matrix $K_{k}^θ$.

C. IBR Kalman-Bucy Filter

For the continuous-time domain, the state-space model with unknown process and noise covariance matrices is given by

\begin{align*}
    x_{t|a}^θ = F_t x_{t}^θ + \Gamma_t u_{t}, \\
    y_{t|a}^θ = H_t x_{t|a}^θ + \nu_{t},
\end{align*}

(44) (45)

where for $t, s \in [a, b)$,

\begin{align*}
    E[u_t^θ (u_s^θ)^T] &= \pi(θ_1) Q_{t}^θ δ(t - s), \quad (46) \\
    E[\nu_t^θ (ν_s^θ)^T] &= \pi(θ_2) R_{t}^θ δ(t - s). \quad (47)
\end{align*}

Let $z_{k|a}^θ = H_t x_{t|a}^θ$ and $\tilde{x}_k^θ$ be the Bayesian least-squares estimate for $x_{t|a}^θ$ based on observations $y_{t|a}^θ$, $s < t$,

\begin{equation}
    \tilde{x}_t^θ = \int_a^t G_{t,a}^θ y_s^θ ds,
\end{equation}

(48)

where

\begin{align*}
    G_{t,a}^θ = \arg \min_{G_{t,a} \in G} E_{\theta} \left[ \left( x_{t|a} - \int_a^t G_{t,s} y_s^θ ds \right)^T \times \left( x_{t|a} - \int_a^t G_{t,s} y_s^θ ds \right) \right].
\end{align*}

(49)

In (49), $G$ is the vector space of all $m \times m$ matrix-valued functions $G_{t,s}$, where $G_{t,s} \in G$ is a mapping $G_{t,s} : (a, b) \times [a, b) \to \mathbb{R}^{m \times m}$ such that $\int_a^b \int_a^t ||G_{t,s}||^2 dt ds < \infty$.

The continuous-time Bayesian orthogonality principle is expressed similarly to the discrete-time domain. Let $\tilde{x}_t^θ$ be the estimate obtained using (48). Condition (49) holds if and only if

\begin{equation}
    E_{\theta} \left[ E \left[ (x_{t|a} - \tilde{x}_t^θ) (y_s^θ)^T \right] \right] = 0_{n \times m}, \quad \forall s \in [a, t).
\end{equation}

(50)

The Bayesian innovation process is defined as

\begin{equation}
    \tilde{z}_t^θ = y_t^θ - H_t \tilde{x}_t^θ.
\end{equation}

(51)

The equations for IBR Kalman-Bucy filtering in the continuous domain can be found based on the equations for the discrete case. This is a classic approach to derive the Kalman-Bucy filter in the literature [35]-[38]. If the time corresponding to discrete index $k$ is $t_k$, then the equations for the IBR Kalman-Bucy filter can be derived as an extension of those for the IBR Kalman filter by letting $\Delta t = t_{k+1} - t_k \to 0$.

First we need to find a correspondence between the state equations for the continuous and discrete time domains in (44) and (12). The general solution for the dynamical model described by (44) can be found as [39]

\begin{equation}
    x_t^θ = \Phi(t, a)x_a^θ + \int_a^t \Phi(t, τ) \Gamma_t u_t^θ dτ,
\end{equation}

(52)

where $\Phi(t, a)$, called the state-transition matrix, satisfies $\Phi(t, a) = I$, the identity matrix of appropriate size, and

\begin{equation}
    \frac{∂ \Phi(t, a)}{∂t} = F_t \Phi(t, a).
\end{equation}

(53)

Let $t_k$ be the discrete-time index, where $t_{k+1} - t_k = \Delta t$, $k = 0, 1, 2, \ldots$. Then (52) can be discretized as

\begin{equation}
    x_{t_{k+1}}^θ = \Phi(t_{k+1}, t_k)x_{t_k}^θ + \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, τ) \Gamma_t u_t^θ dτ.
\end{equation}

(54)
States \( x_{k+1}^\theta \), \( k = 0, 1, 2, \ldots \), can be viewed as the output of a state-space model with discrete-time domain. In fact, the continuous-time model can be treated as a discrete-time model when \( \Delta t \to 0 \). To avoid confusion, we use an overline for variables in the discrete-time domain to distinguish them from those for continuous time. If we correspond (54) with (12), the matrices for the equivalent discrete-time state-space model are \( \Phi_k \), \( \upsilon_k \), \( \Gamma_k \), \( H_k \), and \( v_k \). Note that \( \Phi_k = \Phi(t_{k+1}, t_k) \) and \( H_k = H_{\upsilon_k} \), then
\[
\Gamma_k \upsilon_k^\theta = \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, \tau) \Gamma_\tau \upsilon_\tau^\theta \, d\tau,
\]
and
\[
\Gamma_k Q_k^\theta \Gamma_k^T = \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} \begin{bmatrix} \Phi(t_{k+1}, \tau') \Gamma_\tau Q_\tau^\theta \delta(\tau - \tau') \\ \times \Gamma^T_\tau \Phi^T(t_{k+1}, \tau) \end{bmatrix} d\tau d\tau'.
\]
(55)

We desire a differential equation for \( E_\theta[\mathbf{P}_k^x^\theta] \). Based on (35d),
\[
\lim_{\Delta t \to 0} \frac{E_\theta[\mathbf{P}_{k+1}^x^\theta] - E_\theta[\mathbf{P}_k^x^\theta]}{\Delta t} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left\{ (\Phi_k - \Phi_k K_k^\theta H_k) E_\theta[\mathbf{P}_k^x^\theta] \Phi_k^T + \Gamma_k E_{\upsilon_1}(Q_k^\theta) \Gamma_k^T \right\}.
\]
(56)

To compute (56), we need to compute several limits. First, using (55),
\[
\lim_{\Delta t \to 0} \frac{\Gamma_k E_\theta[Q_k^\theta] \Gamma_k^T}{\Delta t} = \Gamma_{t_k} E_{\upsilon_1}(Q_{t_k}^\theta) \Gamma_{t_k},
\]
(57)

In (57), it is assumed that the expectation and integral operator can be interchanged. Regarding \( \Phi(t, t_k) \), we can see that
\[
\left. \frac{\partial \Phi(t, t_k)}{\partial t} \right|_{t=t_k} = \lim_{\Delta t \to 0} \frac{\Phi(t_{k+1}, t_k) - \Phi(t_k, t_k)}{\Delta t} = F_{t_k} \Phi(t_k, t_k) = F_{t_k},
\]
(58)

where we used (53) to derive the second line. This implies
\[
\lim_{\Delta t \to 0} \frac{1}{\Delta t} \Phi(t_{k+1}, t_k) = F_{t_k} + \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbf{I}.
\]
(59)

Substituting (57) and (59) in (56) yields
\[
\lim_{\Delta t \to 0} \frac{E_\theta[\mathbf{P}_{k+1}^x^\theta] - E_\theta[\mathbf{P}_k^x^\theta]}{\Delta t} = \lim_{\Delta t \to 0} \left\{ (F_{t_k} + \frac{1}{\Delta t} \mathbf{I} - \frac{1}{\Delta t} \Phi_k K_k^\theta H_k) E_\theta[\mathbf{P}_k^x^\theta] \times (F_{t_k} \Delta t + \mathbf{I})^T \Gamma_k E_{\upsilon_1}(Q_{t_k}^\theta) \Gamma_k^T - \frac{1}{\Delta t} E_\theta[\mathbf{P}_k^x^\theta] \right\}.
\]
(60)

To proceed, we must recognize that, since continuous white-noise cannot be represented by ordinary functions, a complete mathematical analysis requires generalized functions. As is commonplace with engineering analyses of white noise processes [35], [36], [38], we are approximating continuous white-noise by discrete white-noise and letting \( \Delta t \to 0 \), with the understanding that these limits do not exist in the ordinary sense and must be placed into the framework of generalized functions for a full mathematical treatment. Following [38], this means that the discrete white-noise sequence approximates the continuous white noise process by shrinking the pulse lengths \( \Delta t \) and increasing their amplitude so that the discrete white-noise sequence tends to infinite-valued pulses of zero duration with the area under the impulse autocorrelation function equaling the area under the continuous white-noise impulse autocorrelation function.

In the present situation, since \( E(\mathbf{v}_s^\theta (\mathbf{v}_s^\theta)^T) = R^\theta \delta(t-s) = \infty \) for \( t = s \), when approximating continuous white noise by sampling it at fixed intervals \( \Delta t \), as \( \Delta t \to 0 \), the covariance of the resulting discrete noise process \( \mathbf{R}_k^\theta \) tends to \( \infty \) and we can approximate \( \mathbf{R}_k^\theta \) by \( R^\theta / \Delta t \) as \( \Delta t \to 0 \). Hence, we compute \( \lim_{\Delta t \to 0} \Phi_k^\theta \) in (60) as
\[
\lim_{\Delta t \to 0} \mathbf{K}_k^\theta = \lim_{\Delta t \to 0} E_\theta[\mathbf{P}_k^x^\theta] H_k^T E_{\upsilon_1}^T H_k \mathbf{P}_k^x^\theta H_k^T + \frac{R^\theta}{\Delta t},
\]
(61)

resulting because \( 1/\Delta t \to \infty \) and \( 1/\Delta t \) is in the denominator. For (60), we also need
\[
\lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbf{K}_k^\theta = \lim_{\Delta t \to 0} E_\theta[\mathbf{P}_k^x^\theta] H_k^T E_{\upsilon_1}^T [\mathbf{P}_k^x^\theta + \frac{R^\theta}{\Delta t}],
\]
(62)

Note that \( \lim_{\Delta t \to 0} \Phi_k = \mathbf{I} \). Plugging (61) and (62) into (60) gives
\[
\lim_{\Delta t \to 0} \frac{E_\theta[\mathbf{P}_{k+1}^x^\theta] - E_\theta[\mathbf{P}_k^x^\theta]}{\Delta t} = F_{t_k} E_\theta[\mathbf{P}_k^x^\theta] + E_\theta[\mathbf{P}_k^x^\theta] F_{t_k}^T - E_\theta[\mathbf{P}_k^x^\theta] H_k E_{\upsilon_1}^T R^\theta_{t_k} + \Gamma_{t_k} E_{\upsilon_1}(Q_{t_k}^\theta) \Gamma_{t_k},
\]
(63)
or equivalently,

\[ E[\hat{P}\_t^{\_\theta}] = \Phi_t E[\Phi_t^\top \hat{P}\_t^{\_\theta}] + E[\Phi_t^\top \hat{P}\_t^{\_\theta}] \Phi_t^T - E[\hat{P}\_t^{\_\theta}] \Gamma_{\theta_2} E[\hat{P}\_t^{\_\theta}] + \Gamma_1 E[Q_{\theta_1}^\top \Gamma_1]. \]  

(64)

We also need to find the differential equation for the Bayesian least-squares estimation. Using (35c),

\[ \hat{x}_\theta^{t+1} = \hat{F}_k \hat{x}_\theta^t + \hat{K}_k \hat{z}_\theta^t. \]

Incorporating (59) and (62) yields

\[
\lim_{\Delta t \to 0} \frac{\hat{x}_\theta^{t+1} - \hat{x}_\theta^t}{\Delta t} = \lim_{\Delta t \to 0} \frac{\hat{F}_k \hat{x}_\theta^t + \hat{K}_k \hat{z}_\theta^t - \hat{x}_\theta^t}{\Delta t} = \hat{F}_k \hat{x}_\theta^t + \hat{K}_k \hat{z}_\theta^t,
\]

(65)

which can be rewritten as

\[
\hat{\theta}_t = \hat{F}_k \hat{\theta}_t + E[\hat{P}_t^{\_\theta}] \hat{H}_t E_{\theta_2}^{-1} \hat{R}_t^{\_\theta} \hat{z}_t = \hat{F}_k \hat{\theta}_t + \hat{K}_t \hat{z}_t,
\]

(66)

where

\[ \hat{K}_t^{\_\theta} = E[\hat{P}_t^{\_\theta}] \hat{H}_t E_{\theta_2}^{-1} \hat{R}_t^{\_\theta}, \]

(67)

is the effective Kalman gain matrix.

The equations required for IBR Kalman-Bucy filtering are (51), (67), (66), and (64). If \( x_a \) denotes the random process at time \( a \), then initial conditions are set to \( \hat{x}_\theta^0 = E[\hat{x}_a] \) and \( E[\hat{P}_0^{\_\theta}] = \text{cov}[x_a] \).

For the Bayesian innovation process \( \hat{z}_\theta^t \), we can show that

\[ E[\hat{z}_\theta^t | \hat{z}_\theta^s] = E[\hat{z}_\theta^t] \delta(t-s). \]

(68)

Similar to the discrete case in Proposition 1, it can be shown that \( E[\hat{z}_\theta^t | \hat{z}_\theta^s] = 0 \) except for \( t = s \). For \( t \neq s \), we treat it as a limit of the discrete-time domain when \( \Delta t \to 0 \):

\[
\lim_{\Delta t \to 0} E[\hat{z}_\theta^t | \hat{z}_\theta^s] = \lim_{\Delta t \to 0} E[\hat{z}_\theta^t \hat{H}_k \hat{P}_k^{\_\theta} \hat{H}_k^\top + \hat{R}_k^{\_\theta}] = \lim_{\Delta t \to 0} E[\hat{z}_\theta^t \hat{H}_k \hat{P}_k^{\_\theta} \hat{H}_k^\top + \frac{1}{\Delta t} \hat{R}_k^{\_\theta}] = \lim_{\Delta t \to 0} E\left[ \frac{1}{\Delta t} \hat{R}_k^{\_\theta} \right] = \lim_{\Delta t \to 0} E[\hat{z}_\theta^t | \hat{z}_\theta^s],
\]

(69)

where \( \hat{H}_k \hat{P}_k^{\_\theta} \hat{H}_k^\top \), which is finite, vanishes as \( \Delta t \to 0 \) because \( \frac{1}{\Delta t} \hat{R}_k^{\_\theta} \to \infty \). Based on (69) and the fact that \( E[\hat{z}_\theta^t | \hat{z}_\theta^s] = 0 \) for \( t \neq s \), we can use Dirac delta function notation as used in (68).

**D. Performance of the IBR Kalman Filter**

To illustrate the effectiveness of the IBR Kalman filter, we consider how model-specific Kalman filters perform on average relative to the prior. To find the average MSE of the model-specific Kalman filters, we must consider the performance of an arbitrary model-specific Kalman filter when the parameters used for filter design are different from the parameters of the true model to which the filter will be applied. In the sequel, we use \( \theta' = [\theta'_1, \theta'_2] \) to denote the parameters used to design the Kalman filter and \( \theta \) to denote the parameters of the true model. The weighting function for the model-specific Kalman filter relative to \( \theta' \) is given by

\[ G_{k,l}^{\theta'} = E\left[ \hat{z}_l (\theta') \right] \left( \hat{P}_l^{\theta'} + \hat{R}_l^{\_\theta} \right)^{-1}, \]

(70)

where \( \hat{z}_l (\theta') \) is the innovation process relative to model \( \theta' \) obtained using (5a) and \( \hat{P}_l^{\theta'} = H_l \hat{P}_l^{\theta}(\theta') H_l^\top \), in which \( \hat{P}_l^{\theta'}(\theta') \) is the optimal estimation error covariance matrix for \( \theta' \), similar to the covariance matrix in (5d). Keeping in mind that there is a mismatch between the parameter \( \theta' \) of the filter and the parameter \( \theta \) of the model, the innovation process used for the estimation is

\[ \hat{z}_k (\theta, \theta') = y_k - \hat{z}_k (\theta, \theta'), \]

(71)

where \( \hat{z}_k (\theta, \theta') \) is the estimate when the \( \theta' \)-specific Kalman filter is applied to the \( \theta \) model and is obtained as

\[ \hat{z}_k (\theta, \theta') = \sum_{l \leq k-1} G_{k,l}^{\theta'} \hat{z}_l (\theta, \theta'). \]

(72)

Therefore,

\[ \hat{\theta}_k (\theta, \theta') = \sum_{l \leq k-1} E\left[ \hat{z}_k (\theta, \theta') \right] \left( \hat{P}_l^{\theta'} + \hat{R}_l^{\_\theta} \right)^{-1} \hat{z}_k (\theta, \theta'). \]

(73)

The recursive equation for \( \hat{\theta}_k (\theta, \theta') \) is obtained as

\[ \hat{\theta}_{k+1} (\theta, \theta') = \sum_{l \leq k} E\left[ \hat{z}_k (\theta, \theta') \right] \left( \hat{P}_l^{\theta'} + \hat{R}_l^{\_\theta} \right)^{-1} \hat{z}_k (\theta, \theta'). \]

(74)

Therefore, if \( \hat{P}_k^{\theta'} (\theta, \theta') = E\left[ \hat{z}_k (\theta, \theta') \hat{z}_k (\theta, \theta')^\top \right] \), then using (74)

\[ \hat{P}_k^{\theta'} (\theta, \theta') = \Phi_k (I - K_k^{\theta'} H_k) \hat{P}_k^{\theta'} (\theta, \theta') (I - K_k^{\theta'} H_k)^\top \Phi_k^T + \Gamma_k Q_k^{\theta'} \Gamma_k^T + \Phi_k K_k^{\theta'} R_k^{\_\theta} \Phi_k^{\_\theta}, \]

(75)
Taking expectations in (75) relative to $\theta$ gives
\[
E_{\theta}\left[P_{k+1}^{\infty}(\theta, \theta')\right] = \Phi_{k} (I - K_{k}^{\theta} \mathcal{H}_{k}) E_{\theta}\left[P_{k}^{\infty}(\theta, \theta')\right] (I - K_{k}^{\theta} \mathcal{H}_{k})^{T} \Phi_{k}^{T} + \Gamma_{k} E_{\theta_{1}}[Q_{k}^{\theta}] \Gamma_{k}^{T} + \Phi_{k} K_{k}^{\theta} E_{\theta_{2}}[R_{k}^{\theta}] (K_{k}^{\theta})^{T} \Phi_{k}^{T},
\] (76)

To find $K_{k}^{\theta}$ and use it in (76), we also need to find $P_{k}^{\infty}(\theta')$. The update equation for $P_{k}^{\infty}(\theta')$ can be computed similarly to (5d) and by using matrices relative to parameter $\theta'$, i.e., $K_{k}^{\theta'}$ and $Q_{k}^{\theta'}$. Therefore, to find the average estimation error covariance matrix $E_{\theta}\left[P_{\theta}^{\infty}(\theta, \theta')\right]$ of an arbitrary $\theta'$-specific Kalman filter, one needs to keep track of two different covariance matrices simultaneously, $E_{\theta}\left[P_{k}^{\infty}(\theta, \theta')\right]$ and $P_{k}^{\infty}(\theta')$. The average MSE of the $\theta'$-specific Kalman filter is the trace of $E_{\theta}\left[P_{\theta}^{\infty}(\theta, \theta')\right]$.

We will also compare the IBR Kalman filter to the steady-state minimax Kalman filter defined by
\[
\theta_{\text{minimax}} = \arg \min_{\theta' \in \Theta, \theta \in \Theta} \lim_{k \to \infty} \text{Tr}(P_{k}^{\infty}(\theta, \theta')).
\] (77)
The average MSE of the steady-state minimax Kalman filter is simply the average MSE of the $\theta_{\text{minimax}}$-specific Kalman filter.

E. Uncertainty quantification for Kalman filtering

As stated in Section II-B, the optimal choice given model uncertainty is to use the IBR filter $\psi^{\theta}$. Relative to the true parameter $\theta$, using $\psi^{\theta}$ instead of the model-specific optimal filter $\psi(\theta)$ incurs a cost (MSE in the case of filtering),
\[
U_{\xi, \psi}(k; \theta) = \xi^{\theta}(x_{k}, \psi_{k}^{\theta}(y)) - \xi^{\theta}(x_{k}, \psi_{k}^{\theta}(y)),
\] (78)
called the objective cost of uncertainty (OCU) relative to $\theta$ [40]. The cost is “objective” in the sense that it relates to an operator’s purpose, not the general uncertainty (entropy) in the prior distribution. The expectation of the OCU relative to the prior defines the mean objective cost of uncertainty (MOCU):
\[
\mathcal{M}_{\xi, \psi_{k}}(k) = E_{\theta}[U_{\xi, \psi_{k}}(k; \theta)].
\] (79)
MOCU measures the expected difference between the performance of the IBR filter (optimal option in the presence of uncertainty) and the model-specific optimal filter (optimal option in the absence of uncertainty), and can be used for objective-based uncertainty quantification [40]. MOCU has been used to develop objective-based experimental design methods in which the goal is to find which unknown parameter in the model should be estimated first to have the most improvement in the performance of the IBR operator [41–43].

In the context of the present paper, MOCU can be used to measure uncertainty when the object is Kalman filtering. The MOCU for Kalman filtering can be found as
\[
\mathcal{M}_{\xi, \psi_{k}}(k) = \text{Tr}(E_{\theta}[P_{k}^{\infty}(\theta, \theta')]) - \text{Tr}(E_{\theta}[P_{k}^{\infty}(\theta)]),
\] (80)
where the expectations are obtained using (35d) and (5d), using $\theta$ instead of $\theta'$. We shall not pursue this matter further here on account of space.

IV. NUMERICAL EXPERIMENTS

A. Example: State Estimation in Sensor Networks

In sensor network applications, the exact knowledge regarding noise covariance matrices is often missing due to sensor drift, hardware variability, etc. To demonstrate the utility of IBR Kalman filtering we consider state estimation in sensor networks, specifically, the model used for sensor scheduling in traffic monitoring and management [44]. Consider a vehicle in 2-D space that can move in both dimensions. Acceleration is assumed to be zero except for a small perturbation regarded as the process noise. The state vector contains four components: $p_{x}$ (vehicle position in $x$), $p_{y}$ (vehicle position in $y$), $v_{x}$ (velocity in the $x$ direction), and $v_{y}$ (velocity in the $y$ direction). The discretization step size is denoted by $h$. Assuming that two sensors are utilized to measure vehicle dynamics, the state-space model describing the dynamics is of the following form:
\[
\begin{align*}
x & = \begin{bmatrix} p_{x} \\ p_{y} \\ v_{x} \\ v_{y} \end{bmatrix}, & \Phi_{k} &= \begin{bmatrix} 1 & 0 & h & 0 \\ 0 & 1 & 0 & h \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \\
\Gamma_{k} &= \begin{bmatrix} h^{2}/2 & 0 & 0 & h \\ 0 & h^{2}/2 & 0 & 0 \\ 0 & 0 & h & 0 \\ 0 & 0 & 0 & h \end{bmatrix}, & H_{k} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.
\end{align*}
\]

We set $h = 0.5$. The process noise and observation noise covariance matrices are
\[
Q_{k}^{\theta} = \begin{bmatrix} \theta & 0 \\ 0 & \theta \end{bmatrix}, & R_{k} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix},
\]
where $\theta$ is assumed to be unknown and belongs to $[0.25, 3]$. The initial conditions are set to $\text{cov}[x_{0}] = 12 \times I$ and $E[x_{0}] = 0 \times 1$.

We consider two priors for $\theta$: uniform and beta $B(0.1, 0.9)$ (scaled over the interval $[0.25, 3]$). The average MSEs for the model-specific, minimax, and IBR Kalman filters are given by the traces of $E_{\theta}[P_{k}^{\infty}(\theta, \theta')], E_{\theta}[P_{k}^{\infty}(\theta, \theta_{\text{minimax}})],$ and $E_{\theta}[P_{k}^{\infty}(\theta)]$ (obtained from (35d)), respectively. Figure 2 shows the average MSE obtained for the minimax and IBR Kalman filters over time. The minimax Kalman filter corresponds to $\theta_{\text{minimax}} = 3$, independent of the prior. For both priors,
average MSE for the IBR Kalman filter is always lower than average MSE for the minimax Kalman filter. The difference between the performances of the minimax and IBR Kalman filters is larger for the Beta prior because minimax performs poorly when there are extreme states with low probability.

Figure 3 compares average steady-state MSES for model-specific, minimax, and IBR Kalman filters when \( k \) is large enough that the steady-state is reached. We can see that the minimum average MSE is achieved by applying the IBR Kalman filter.

Figure 4 shows the performance of various Kalman filters over the uncertainty class of the process noise parameter, where MSE values are in the steady-state (approximately \( k \geq 10 \)). Figure 4 (b) shows the MSE surface when the \( \theta \)-specific Kalman filter is applied to the model with parameter \( \theta \). Figures 4 (c) and (d) illustrate the steady-state MSE of the IBR and minimax Kalman filters over the entire uncertainty class when priors are either uniform or beta. The lower band for MSE for each \( \theta \) is determined by the \( \theta \)-specific Kalman filter. The IBR Kalman filter generally does not perform as well as the \( \theta \)-specific Kalman filter at \( \theta \) and sometimes even has worse performance than the minimax filter. This behavior is normal. The IBR filter is optimal relative to the prior and performs well in the regions of the uncertainty class where there is considerable prior probability mass. For example, in Figure 4 (d), the IBR Kalman filter performs very close to the performance of the \( \theta \)-specific Kalman filter and much better than the minimax filter for \( \theta < 2 \), most of the prior probability mass being concentrated in this region.

For this example, we also assume that \( Q^\theta_k \) can be a general covariance matrix of size \( 2 \times 2 \) without any constraints. A suitable prior distribution for this covariance matrix is a Wishart distribution \( W(\Sigma, n) \), \( \Sigma \) and \( n \) being the scale matrix and the degree of freedom, respectively. In our simulations, we consider a Wishart prior distribution with \( \Sigma = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 2 \end{bmatrix} \) and \( n = 2 \). Keeping in mind that \( E_\theta [Q^\theta_k] = n \Sigma \), an IBR Kalman filter can be designed for this settings by computing the effective Kalman gain matrix. Regarding the model-specific Kalman filters, we generate a sample pool of 10,000 covariance matrices from the assumed Wishart distribution and then compute the average MSE of the model-specific Kalman filter designed relative to each generated noise covariance matrix. The minimax Kalman filter is found according to (77) where the search space is limited to the set of generated covariance matrices. In Figure 5, we show the average steady-state MSES for the model-specific Kalman, minimax Kalman, and the IBR Kalman filters. We use box plots to show the average MSE of the model-specific filters, which have been grouped according to the elements \( q(1,1) \) and \( q(1,2) \) of the corresponding covariance matrix in Figures 5 (a) and (b), respectively. The minimum average MSE is achieved when the proposed IBR Kalman filter is utilized.

B. Example: Gene Regulatory Network Construction

In this section, we consider IBR Kalman filtering for inferring a gene regulatory network (GRN) from time series data. GRNs are network models used to discover therapeutic interventions, e.g., drug design and scheduling [45]. Owing to noise in data acquisition and inherent noise in the underlying biological system, robust inference is a critical need. Specifically, we focus on GRN inference when modeling via continuous nonlinear ordinary differential equations [46]. Let \( g_i \) denote the expression value of gene \( i \). Assuming there are \( m \) genes interacting with each other, the dependency between genes can be modeled as a differential equation:

\[
\frac{dg_i}{dt} = f_i(g_1, \ldots, g_m) + v_i, \quad (81)
\]
where $f_i$ is a regulatory function that can take various forms and $v_i$ is the external noise. Following [46], we choose the following form for $f_i$, $1 \leq i \leq m$:

$$f_i(g_1, \ldots, g_m) = \sum_{j=1}^{L_i} [(x_{ij} + u_{ij})\Omega_{ij}(g_1, \ldots, g_m)],$$  \hspace{1cm} (82)

where $\Omega_{ij}(g_1, \ldots, g_m)$ is the $j$-th nonlinear term of the regulatory function $f_i$, $L_i$ is the number of nonlinear terms in $f_i$, and $u_{ij}$ is the parameter noise. We assume that the parameter and external noise are independent and are white-noise processes. Moreover, we assume that $f_i$ is a quadratic function. We extract 12 genes, HAP1, CYB2, CYC7, ROX1, CYT1, HAP2, HAP3, HAP4, CYC1, COX5A, COX5B, and GPD2, from the yeast cell cycle dataset. Genes are named $g_1$ to $g_{12}$ in the order they have been introduced. To avoid unlimited fluctuations for gene expressions, lower and upper bounds for gene expression are set to 0 and 2000, respectively.

We assume that the coefficients $x_{ij}$ are unknown and aim to find them via data generated from the system. In our simulations, we generated synthetic time series data using regulatory equations inferred in [46] according to Table I. There are 54 coefficients in Table I. Based on the methodology proposed in [46], the coefficients can be estimated using Kalman filtering where states are coefficients to be estimated, observations are changes in the values of genes during the time step, and process and observation noise are parameter and external noise, respectively. Therefore, the following state-space model is used for the aforementioned inference problem:

$$
\begin{align*}
\mathbf{x}_{k+1} &= \mathbf{x}_k + \mathbf{u}_k, \\
\mathbf{y}_k &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k.
\end{align*}
$$  \hspace{1cm} (83)

The state vector $\mathbf{x}_k$ is a vector of size $54 \times 1$ containing coefficients to be estimated; the first four elements of the state vector are unknown coefficients for gene $g_1$; the next four elements of the state vector $\mathbf{x}_k$ are related to gene $g_2$, and so on. Similarly, 54 elements of the process noise vector $\mathbf{u}_k$ are assigned to the parameter noises. For example, the first 4 elements are assigned to $u_{11}$, $u_{12}$, $u_{13}$ and $u_{14}$ as shown in Table I. The observation noise vector $\mathbf{v}_k$ contains the external noises $v_i$ as defined in (81). The observation transition matrix $\mathbf{H}_k$ is of size $12 \times 54$ and obtained by arranging the nonlinear terms in the right hand side of Table I in a matrix format. For example, $g_1(k)g_4(k)$, $g_1(k)g_{11}(k)$, $g_4(k)g_8(k)$, and $g_6(k)g_6(k)$ are placed in the first row and columns 1 to 4; $g_1(k)$, $g_3^2(k)$, $g_4(k)g_5(k)$, and $g_3^2(k)$ are put in the second row and columns 5 to 8; and so on. The observation vector $\mathbf{y}_k$ is a $12 \times 1$ vector with $y_k(i) = (g_i(k\Delta t + \Delta t) - g_i(k\Delta t))/\Delta t$, where $\Delta t$ is the step size used to sample the differential equations. We set $\Delta t = 0.5$. Initial conditions are $E[x_0] = 0_{54 \times 1}$ and $\text{cov}[x_0] = 0.001 \times I$. The covariance matrix $\mathbf{Q}$ of the process noise is set to $10^{-10} \times I$. The observation noise covariance matrix $\mathbf{R}$ is $\theta \times I$, where $\theta$ is unknown and belongs to interval $[1, 7]$. We employ two different prior distributions: beta $B(\alpha, \beta)$ with $\alpha = 0.1$ scaled over the interval and uniform.

Figure 6 presents the average MSE for the IBR Kalman and minimax filter when different numbers of observations are used for estimation. For a better visualization, we also plot the average MSE values for more than 20 observations separately inside the graphs. For both priors, the average MSE achieved by IBR Kalman filtering is always lower than that achieved by minimax filtering.

Figure 7 shows the average MSE achieved using the IBR Kalman filter, the minimax Kalman filter, and the model-specific Kalman filters for two different prior distributions when $k = 20$ observations are observed. The difference between the performance of the IBR and minimax filters is larger for the beta distribution because for this distribution.
there is not much prior probability mass around \( \theta_{\text{minimax}} \).

Figure 8 analyzes the performance of the different robust Kalman filter designs for two different prior distributions shown in Figure 8 (a). Figure 8 (b) shows the MSE surface with respect to the Kalman filter designed for \( \theta^* \) but applied to the model with parameter \( \theta \). Figures 8 (c) and (d) show the performances of the IBR and minimax designs over the uncertainty class. The IBR Kalman filter outperforms the minimax Kalman filter for a majority of parameter values, especially in the regions with high prior probability mass.

C. Comparison with adaptive Kalman filtering

When dealing with unknown noise statistics, the aim of adaptive Kalman filtering is to adjust knowledge about noise statistics along with state estimation. Adaptive Kalman filtering approaches can be divided into four different groups [47]: Bayesian, maximum likelihood, correlation, and covariance matching methods. We compare IBR Kalman filtering with two benchmark adaptive methods proposed by Myers [7] and Mehra [9] for the example applications we consider in this paper. First, we briefly explain these two methods. Due to space limitation, we only highlight the main points of these two methods and refer interested readers to [7] and [9] for equations and more explanation.

In the Myers method, the concept of empirical estimators is utilized to estimate noise covariance matrices from observations. Assume that noise covariance matrices \( R \) and \( Q \) are unknown. Some initial values \( R^{(0)} \) and \( Q^{(0)} \) are used to design the classical Kalman filter according to (5a)-(5d). Let the estimated state vectors using this Kalman filter be denoted by \( \hat{x}_{k|k} \) and \( \hat{x}_{k|k} \). These can be used to approximate observation and process noise vectors denoted by \( \tilde{u}_k \) and \( \tilde{v}_k \), respectively. Using \( \hat{v}_k \), an unbiased estimate \( R^{(1)} \) after \( N_s \) observations can be found according to [7, eq. (9)]. Also, regarding the process noise covariance matrix, an unbiased estimate of \( Q^{(1)} \) using \( \hat{u}_k \) is given by [7, eq. (13)]. This procedure of adjusting noise covariance matrices can be repeated after every \( N_s \) observations. Note that in this adaptive method, it is assumed that unknown noise covariance matrices \( R \) and \( Q \) are constant.

The learning process in the Mehra method is based on the autocorrelation function of the innovation process [9]. In this method, when updating noise covariance matrices is not feasible, the Kalman gain matrix is adapted directly. If the Kalman filter is designed relative to the true noise statistics, then the resulting innovation process is a discrete white-noise
process; however, when the Kalman filter is not designed relative to the true noise statistics, the Mehra method uses the autocorrelation of the resulting innovation process to adjust the Kalman filter. As stated in [9], it is assumed that the system under consideration is time invariant, completely controllable, and observable. Therefore, we drop the time index $k$ from the model matrices such as $H$ and $\Phi$ for this method. If $z_k$ is the innovation process resulting from applying the Kalman filter whose corresponding steady-state Kalman gain matrix is denoted by $K^{(0)}$, after $N_s$ samples of the innovation process, the sample autocorrelation function for this process, denoted by $\Sigma_z$ (for the lag $\tau$), can be computed. Keeping in mind that the state vector is of size $n \times 1$, define the matrix

$$A = \begin{bmatrix} H \Phi & H \Phi (I - K^{(0)} H) \Phi \\ & \vdots \\ & H (\Phi (I - K^{(0)} H))^{n-1} \Phi \end{bmatrix}.$$ 

Now the updated Kalman gain matrix $K^{(1)}$ is found according to [9, eq. (33)] Provided that all eigenvalues of $\Phi (I - K^{(1)} H)$ have a norm smaller than 1, the Kalman gain matrix can further be updated recursively according to [9, eqs. (34)-(36)]; otherwise, the method suggests that $K^{(1)}$ is used to filter the observations and update the Kalman gain matrix according to [9, eq. (33)].

Figure 9 compares our proposed method with the Myers and Mehra adaptive methods for the sensor network example considered in Section IV-A using exactly the same settings as in Section IV-A, e.g., unknown process noise covariance matrix with the diagonal element being uniformly distributed over $[0, 25, 3]$. Figures 9 (a) and (b) study the average MSE for different number of observations. Note that these figures give the average MSE for after $k = 10$ observations so that the difference between the performances of different methods becomes visible. The average MSE of the IBR filter seems to be constant because, as Figure 2 suggests, after about $k = 10$ observations, the IBR Kalman filter reaches its steady-state MSE. To find the average MSE for the IBR method we use (35d). To find the average MSE for the adaptive methods, first we generate a pool of assumed true values for the unknown diagonal element of the process noise covariance based on the assigned prior distributions. Then for each assumed true value, we run the adaptive Kalman filtering methods and compute the resulting MSE of state estimation using (75). Then the average MSE is obtained by taking the average of all MSEs found for the generated assumed true values. Because the performance of the adaptive methods depends on the number of observations $N_s$ used for the process of adjusting filter parameters, we show the performance of these two methods for different $N_s$. Should a small $N_s$ be chosen, then the estimates of the noise covariance matrices for the Myers method and the estimate of the innovation autocorrelation for the Mehra method would be highly inaccurate and that would affect the accuracy of the Kalman filter, perhaps even making the output of the filter unstable. On the other hand, if a large $N_s$ is chosen then it would delay the adjusting operation and may cause the filter to run for a large number of observations without being adjusted.

For example, in Figures 9 (a) and (b) we can see that when $N_s$ is set to 150 and 700 for the Myers and Mehra methods, respectively, the performance of the respective adaptive Kalman filter deteriorates as more observations are utilized. These figures suggest that when a large number of observations are available, the adaptive methods are able to estimate the unknown noise covariance matrix (or the Kalman gain matrix for the Mehra method) to some extent and consequently have comparable performance or even outperform the IBR Kalman filter. However, when there is not enough observations the IBR Kalman filter performs much better than the adaptive methods. That is why, compared to Section IV-A, for the performance comparison with the adaptive methods we did simulations for much more observations (up to $k = 2500$), while the IBR Kalman filter has converged to its steady-state performance after $k = 10$ observations.

Figure 10 shows the performance of the proposed IBR and adaptive Kalman filtering methods for the gene regulatory network construction example studied in Section IV-B. Note that for this example the Mehra method is not applicable because it assumes that the state-space model is time invariant, whereas in the state-space model used for GRN construction the observation transition matrix $H$ is time varying, its value
at each time point being determined by the gene expression values. Therefore, we only compare IBR Kalman filtering with the Myers method. All simulation settings are the same as those in Section IV-B. For this application, where the matrix $H_k$ changes with time and the dimensions of the state and observation vector are large, we can see that even for a large number of observations the Myers method fails to estimate the unknown observation noise covariance matrix with high accuracy and the IBR Kalman filter has better performance. We note that for the sake of comparison with the adaptive method we have run simulations for up to $k = 2500$ observations; however, in practice it is not possible to obtain that generous amount of data for GRN identification. As genomic data are expensive, small samples are commonplace in systems biology and bioinformatics, and often we are not able to obtain more than 50 samples.

It is worth mentioning that while the Myers adaptive method assumes stationary noise statistics and the Mehra method is only applicable to time invariant models that are observable and controllable, there is no limitation in the application of the proposed IBR Kalman filtering framework for different state-space models, different noise statistics, and different prior distributions for unknown parameters. Moreover, adaptive methods are suitable when data are plentiful, but not when observations are expensive or limited by subject availability, as in oncology.

### D. Comparison with the sensitivity penalization based estimation

In [48], a regularized least-squares (RLS) framework that takes into account unknown modeling-error parameters has been proposed. In fact, unknown parameters, in the context of this framework, embody the deviation of the model parameters from their nominal values. Another approach proposed in [49] is based on penalizing sensitivity of estimation relative to modeling errors and assumes that the model matrices $H_k$, $\Phi_k$, and $\Gamma_k$ are all differentiable functions of some unknown modeling-error parameter $\theta$. The method has been extended to the situation where the observation can be randomly lost [50] and to GRN structure identification [51] when the GRN model is a nonlinear state-space based on sigmoid functions. For this method, a design parameter $\gamma$ controls the importance of the estimation relative to the nominal model and the sensitivity of the estimation with respect to the modeling error. Considering the fact that matrices $H_k$, $\Phi_k$, and $\Gamma_k$ in [49] and matrix $\Gamma_k$ and noise covariance matrices $Q_k$ and $R_k$ in our proposed method can possess uncertainty, in order to compare these two methods we can only focus on the uncertainty in the process noise transition matrix $\Gamma_k$. We consider the following numerical example:

\[
\begin{align*}
x^\theta_{k+1} &= \begin{bmatrix} 0.8 & 0 \\ 0.9 & 0.2 \end{bmatrix} x^\theta_k + \begin{bmatrix} 0.6 + \theta \\ 0.5 \end{bmatrix} u_k, \\
y^\theta_k &= \begin{bmatrix} 1 & 1 \end{bmatrix} x^\theta_k + v_k.
\end{align*}
\]

This model has been adopted from [52]–[54] where the model was used for sensor networks with packet drops and random measurement delays. We assume that $\theta$ is unknown and uniformly distributed over $[-0.3, 1]$.

Figure 11 presents the results of our method and [49] with $\gamma = 0.7$ and 0.8. In Figure 11 (a), as must be so, the minimum average MSE relative to the prior distribution is achieved by the IBR Kalman filter. Figure 11 (b) shows the performance of both methods over the uncertainty class. The lower bound is determined by the performance of the model-specific Kalman filters designed relative to the true parameter. We can see that except for the small values of $\theta$, for most of the uncertainty class the IBR Kalman filter outperforms [49]. Intuitively, the IBR filter uses knowledge in the prior distribution, whereas the method in [49] considers no prior distribution for the error parameters (or unknown parameters in the context of our paper). In general, the IBR advantage depends on finding suitable prior distributions, which can be a difficult task. In the case of GRNs, methods for constructing prior distributions for Bayesian robust gene-based phenotypic classification [31] have been developed for Gaussian [55] and discrete [56] models, and work continues on this basic engineering problem.

### V. Conclusion

We have introduced the notion of an intrinsically Bayesian robust Kalman filter, which is expressed in a set of recursive equations analogous to those for classical Kalman filtering, with the Kalman gain matrix replaced by the effective Kalman

---

**Figure 10:** Performance comparison of the Myers method and the proposed IBR Kalman filter for the gene regulatory network construction example. The adaptive method is applied after every $N_s = 400$ observations. (a) Uniform prior. (b) Beta prior.

**Figure 11:** Performance comparison of the sensitivity penalization based approach [49] and the proposed IBR Kalman filter. (a) The average dynamic MSE of two methods relative to the prior probability distribution. (b) Performance analysis of two methods over the uncertainty class.
gain matrix. The IBR Kalman filter provides an optimal operator yielding the best performing filter given the available prior knowledge. Having developed the recursive structure of IBR Kalman filtering in the present paper, we are turning our attention to a follow-up paper in which we envision that, via the Bayesian innovation process approach, we will be able to update the priors after observations – a process that is problematic owing to the dependency between observed samples. We believe that the IBR Kalman filtering framework can be extended for a wide variety of problems for which ordinary Kalman filtering has been extended. Future works include studying state-space models in which process and observation noise are not white, developing the Bayesian robust version of the extended Kalman filter (EKF), and applying our proposed IBR Kalman filtering framework to the problem of simultaneous estimation of gene regulatory network structure and gene expression values along with the evaluation of its performance compared to using the sensitivity penalization based approach in [51].

**APPENDIX A**

**Proof of Proposition 1**

In order to prove Proposition 1, first we consider the case that $k \neq l$. Without loss of generality we assume that $k > l$:

$$E_{\theta}[E \left[ \theta \right] E \left[ \theta \right] T] = E_{\theta}[E \left[ \theta \right] E \left[ \theta \right] T] = \sum_{l \leq k} \left[ G_{l,k} \theta \theta \right] \left( \theta \theta - \theta \theta \right) T = 0_{m \times m}, \quad (84)$$

where the first term is zero because of the Bayesian orthogonality principle and the third term is zero because future observation noise is independent from the past observations. Now we turn to the case that $k = l$:

$$E_{\theta}[E \left[ \theta \right] E \left[ \theta \right] T] = \sum_{l \leq k} \left[ G_{l,k} \theta \theta \right] \left( \theta \theta - \theta \theta \right) T + 2E_{\theta}[E \left[ \theta \right] E \left[ \theta \right] T] = E_{\theta}[P_{l,k} \theta \theta + R_{l,k} \theta \theta], \quad (85)$$

where the third term in the second equality is zero because the Bayesian least-squares estimation at time $k$ based upon observations up to time $k - 1$ and $\theta \theta$ are both independent from the observation noise at time $k$.

**APPENDIX B**

**Proof of Lemma 1**

We prove this lemma with the help of the Bayesian orthogonality principle. For each $\theta$ we show that, if $x_{k}^0 = \sum_{l \leq k-1} G_{l,k} \theta \theta$ such that $E_{\theta}[E \left[ \theta \theta - \theta \theta \right] T] = 0_{m \times m}$ for $l \leq k - 1$ and $z_{l}^0 = y_{l}^0 - H \theta x_{l}^0$, then $E_{\theta}[E \left[ \theta \theta - \theta \theta \right] T] = 0_{n \times m}$ also holds. This implies that $x_{k}^0$ is the Bayesian least-squares estimate given the observation process $y_{k}^0$. Now, for any $l \leq k - 1$,

$$0_{n \times m} = E_{\theta}[E \left[ \theta \theta - \theta \theta \right] T] = E_{\theta}[E \left[ \theta \theta - \theta \theta \right] T]$$

where the first term is zero because of the Bayesian orthogonality principle. For each $\theta$ we prove this lemma with the help of the Bayesian orthogonality principle and the third term is zero because future observation noise are not white, developing the Bayesian robust version of the extended Kalman filter (EKF), and applying our proposed IBR Kalman filtering framework to the problem of simultaneous estimation of gene regulatory network structure and gene expression values along with the evaluation of its performance compared to using the sensitivity penalization based approach in [51].

**REFERENCES**


Rouzbeh Dehghannasiri received the B.S. degree from the University of Tehran, Iran, in 2010, the M.A.Sc. degree from the McMaster University, Hamilton, Canada in 2012, and the Ph.D. degree from Texas A&M University, College Station, TX, in 2016, all in electrical engineering. Since his PhD graduation, he has been a post-doctoral research associate in the Department of Electrical and Computer Engineering of Texas A&M University. His current research interests include statistical signal processing, uncertainty quantification in dynamical and computational biology. He is the recipient of the Best Paper Award at the 12th Annual MICBIOS Conference in 2015 and the McMaster Outstanding Thesis Research Award in 2012.

Mohammad Shahrokh Esfahani received the B.S. degree from the University of Tehran, Iran, in 2007, the M.Sc. degree from Sharif University of Technology, in 2009, and the PhD degree from Texas A&M University in 2014, all in electrical engineering. He is currently a post-doctoral research fellow in the Division of Oncology and Center for Cancer Systems Biology in Stanford School of Medicine, Stanford University. Prior to that, he was a post-doctoral research associate in the Center for Bioinformatics and Genomic Systems Engineering, Texas A&M University. His current research interests include small-sample classification, Bayesian statistics, and statistical signal processing, with the applications in non-invasive cancer diagnosis and prognosis and precision medicine.

Edward R. Dougherty is a Distinguished Professor in the Department of Electrical and Computer Engineering at Texas A&M University in College Station, TX, where he holds the Robert M. Kennedy ’26 Chair in Electrical Engineering and is Scientific Director of the Center for Bioinformatics and Genomic Systems Engineering. He holds a Ph.D. in mathematics from Rutgers University and an M.S. in Computer Science from Stevens Institute of Technology, and has been awarded the Doctor Honoris Causa by the Tampere University of Technology. He is a fellow of both IEEE and SPIE, has received the SPIE President's Award, and serves as the editor of the SPIE/IS&T Journal of Electronic Imaging. At Texas A&M University he has received the Association of Former Students Distinguished Achievement Award in Research, been named Fellow of the Texas Engineering Experiment Station, and received the Dwight Look College of Engineering. Prof. Dougherty is author of 16 books and author of more than 300 journal papers.