2015 American Control Conference
Palmer House Hilton
July 1-3, 2015. Chicago, IL, USA

Simultaneous Input and State Estimation for Linear Time-Invariant Continuous-Time Stochastic Systems with Unknown Inputs

Sze Zheng Yong [a] Minghui Zhu [b] Emilio Frazzoli [a]

Abstract— **In this paper, we present an optimal filter for linear time-invariant continuous-time stochastic systems that simultaneously estimates the states and unknown inputs in an unbiased minimum-variance sense. The optimality of the proposed filter is proven by reduction to an equivalent system without unknown inputs. Then, a second proof is given for a special case by limiting case approximations of the optimal discrete-time filter [1], thus establishing the connection between the continuous- and discrete-time filters. Conditions for the existence of a steady-state solution for the proposed filter are also given. Moreover, we show that a principle of separation of estimation and control holds for linear systems with unknown inputs. An example is given to demonstrate these claims.**

I. Introduction

For linear continuous-time stochastic systems with known inputs, the Kalman-Bucy filter [2] is regarded as the optimal solution to extract information about a variable of interest from noisy measurements. However, the accessibility of the unknown/disturbance input is often not possible. This problem of simultaneous state and input estimation is found across many disciplines and applications, from the real-time estimation of mean areal precipitation during a storm [3] to input estimation in physiological and transportation systems [1], [4] to fault detection and diagnosis [5].

Literature review. Initial research with the objective of *concurrently* obtaining minimum-variance unbiased estimates for both the states and the unknown disturbance inputs to the system has been focused on particular classes of linear *discrete-time* systems with unknown inputs [6]–[10]. Recently, a less restrictive framework for optimally estimating both state and unknown input has been proposed in [1]. However, most state and input/fault estimation solutions for *continuous-time* systems are limited to deterministic systems (see, e.g., [11]–[14] and references therein), and they typically assume bounds on the allowed disturbance input or rely on the differentiation of measurements, which may lead to instability when noise is present. Hence, the problem of simultaneous state and input estimation for linear continuous-time stochastic systems remains open.

Contributions. To bridge the gap in knowledge about optimal filters for simultaneous input and state estimation of linear time-invariant continuous-time stochastic systems, we propose an optimal filter, that estimates the unknown input and the system states in the minimum-variance unbiased (MVU) sense, i.e., the estimated signals have zero bias and have variances that are not higher than any other unbiased estimates for all possible values of the signals.

To prove the optimality of the proposed filter, we reduce the linear system with unknown input to an equivalent system without unknown inputs. Thus, we can use existing results of the Kalman-Bucy filter [2] to obtain the proposed optimal filter and also the conditions for the existence of a steady-state solution of the resulting Riccati differential equation. Next, we derive the optimal filter for a special case by using limiting case approximations of the discrete-time optimal filter in [1]. Through this derivation, we gain a better understanding of the similarities and differences of the discrete-time and continuous-time filters. Moreover, we present a second complementary algorithm under the condition that an additional measurement with information about the state derivative is not available and that only the estimation of state is desired. Finally, we show that a principle of separation of estimation and control also exists for linear systems with unknown inputs.

Notation. $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, and $\mathbb{N}$ positive integers. For a vector of random signals, $v \in \mathbb{R}^n$, its expectation is denoted by $\mathbb{E}[v]$. Given a matrix $M \in \mathbb{R}^{p\times q}$, its transpose, inverse and rank are given by $M^\top$, M^{-1} and $\mathrm{rk}(M)$. For a symmetric matrix S, $S \succ 0$ and $S \succeq 0$ denote S is positive definite and positive semidefinite.

II. Problem Statement

For ease of exposition, we consider the following model representation for linear time-invariant continuous-time stochastic system[1]

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gd(t) + w(t), \\ y(t) &= Cx(t) + Du(t) + Hd(t) + v(t), \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ and $y(t) \in \mathbb{R}^l$ are the state and measurement vectors at time t. The known and unknown input vectors, $u(t) \in \mathbb{R}^m$ and $d(t) \in \mathbb{R}^p$, respectively, are assumed to be deterministic and once differentiable (i.e., $\dot{u}(t)$ and $\dot{d}(t)$ exist). The process noise $w(t) \in \mathbb{R}^n$ and the mea-

[a] S.Z Yong and E. Frazzoli are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA (e-mail: szyong@mit.edu, frazzoli@mit.edu).

[b] M. Zhu is with the Department of Electrical Engineering, Pennsylvania State University, 201 Old Main, University Park, PA 16802, USA (e-mail: muz16@psu.edu).

[1]This model representation is sometimes referred to as a nonsuccessful construction [15, Section 3.4]. However, we intentionally choose this more common representation found in the original Kalman-Bucy filter [2] and textbooks (e.g., [16]–[18]) over the more accurate stochastic differential equations [15, Section 7.6]. Since the proofs in this paper, among others, use existing results from the Kalman-Bucy filter which are derived from the duality of the optimal control and estimation problems [2], [15] that is directly applicable to the stochastic differential equation representation, this choice of representation does not affect the results of this paper beyond the slight abuse in model representation.

978-1-4799-8686-6/$31.00 ©2015 AACC

surement noise[2] $v(t) \in \mathbb{R}^l$ are assumed to be mutually uncorrelated, zero-mean, white random processes with known covariance matrices, with noise statistics $\mathbb{E}[w(t_1)w(t_2)^\top] = Q(t_1)\delta(t_1 - t_2)$, $\mathbb{E}[v(t_1)v(t_2)^\top] = R(t_1)\delta(t_1 - t_2)$ and $\mathbb{E}[w(t_1)v(t_2)^\top] = 0$, where $\delta(\cdot)$ is the Dirac delta function, $Q(t) = Q \succeq 0$ and $R(t) = R \succ 0$ for all t (see, e.g., [2], [18] for details about the choice of noise model). The matrices A, B, G, C, D, H, Q and R are known and constant. $x(t_0) = x_0$ is also assumed to be independent of $v(t)$ and $w(t)$ for all t and an estimate $\hat{x}(t_0) := \hat{x}_0$ of the initial state x_0 is available with covariance matrix $\mathcal{P}_0^x$. Without loss of generality, we assume throughout the paper that $n \geq l \geq 1$, $l \geq p \geq 0$ and $m \geq 0$ and the current time $t \geq 0$.

As observed in [19] for deterministic systems with unknown inputs, except for some trivial cases (e.g., H has full rank), derivatives of the measurements are unavoidable when the reconstruction of the unknown input is desired. Thus, we expect stochastic systems to similarly require some form of additional signal information/measurement that is a counterpart of the derivative of $y(t)$ in the deterministic case. The necessity for such a measurement related to the derivative will become obvious after a similarity transformation shown in Section III-A, in which we note that the measurement $y(t)$ does not contain any information about a particular component of the unknown input $d(t)$, if H is rank deficient. However, since numerical differentiation of noisy measurements is in general to be avoided and a higher-order system representation via state augmentation with inputs would not result in a corresponding full rank H' if the original H does not have full rank, we instead assume that we have additional measurements that contain information about $\dot{x}(t)$ given by

$$\overline{y}(t) = \overline{C}\dot{x}(t) + \overline{D}\dot{u}(t) + \overline{H}\dot{d}(t) + \overline{v}(t), \tag{2}$$

with the following noise statistics: $\mathbb{E}[\overline{v}(t)] = 0$, $\mathbb{E}[\overline{v}(t_1)\overline{v}(t_2)^\top] = \overline{R}(t_1)\delta(t_1 - t_2)$, $\mathbb{E}[v(t_1)\overline{v}(t_2)^\top] = 0$ and $\mathbb{E}[w(t_1)\overline{v}(t_2)^\top] = 0$, and $\overline{R} \succ 0$ is constant and known. This assumption of an additional measurement is reasonable in practice, for e.g., accelerations of mechanical systems are typically measured in addition to state (position and velocity) measurements. Note that the measurement $\overline{y}$ is not needed if H is full rank. Finally, to simplify notations, we omit the explicit time-dependence of the signals throughout the paper.

The objective of this paper is to design an optimal recursive filter algorithm (Section III) which simultaneously estimates the system state $x(t)$ and the unknown input $d(t)$ based on an initial estimate $\hat{x}_0$ and the measurements up to time t, $y(\tau)$ and, in the case that H is not full rank, with additional measurements $\overline{y}(\tau)$ for all $0 \leq \tau \leq t$. No prior knowledge of the dynamics of $d(t)$ is assumed. In addition, the special case[3] where $\overline{C} = C$, $\overline{D} = D$ and $\overline{H} = H$ will be studied in relation to the discrete-time filter for developing an intuition about the similarities and subtle differences between continuous- and discrete-time filters.

A secondary goal of this paper is to present a complementary algorithm for the same special case when H is rank deficient and with the additional restriction that the measurement $\overline{y}$ and the signal $\dot{u}(t)$ are not available. In this case, only a particular component of the unknown input can be estimated. However, this case is of interest for applications that only require the estimation of state for systems with unknown inputs, as is done in earlier literature on linear discrete-time systems [3], [20], [21]. Moreover, for reasons that will be expounded in Section IV, this algorithm is only presented for the case that $v(t) = 0$ for all $t \geq 0$.

[2]Note that in Section IV-A, we will consider a heuristic approach for a special case in which we assume that $v(t) = 0$ for all $t \geq 0$.

[3]The additional measurement $\overline{y}(t)$ for this special case can be seen as a pseudo-derivative of the output measurement $y(t)$, which should be distinguished from the output derivative signal $\dot{y}(t)$ that is not well defined due to the differentiation of the discontinuous noise term.

III. Minimum-variance Unbiased Filter for Input and State Estimation

A. Similarity Transformation

Similar to its discrete-time counterpart [1], we first carry out a similarity transformation of the system. Let $\text{rk}(H) = p_H$. Then, we rewrite the direct feedthrough matrix H using singular value decomposition as

$$H = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix}, \tag{3}$$

where $\Sigma \in \mathbb{R}^{p_H \times p_H}$ is a diagonal matrix of full rank, $U_1 \in \mathbb{R}^{l \times p_H}$, $U_2 \in \mathbb{R}^{l \times (l - p_H)}$, $V_1 \in \mathbb{R}^{p \times p_H}$, $V_2 \in \mathbb{R}^{p \times (p - p_H)}$, and $U := \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ and $V := \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ are unitary matrices. Note that in the case with no direct feedthrough, Σ is the zero matrix, U_1 and V_1 are empty matrices, and U_2 and V_2 are arbitrary unitary matrices.

Then, we transform the (completely) unknown input into two orthogonal components:

$$d_1 = V_1^\top d, \quad d_2 = V_2^\top d. \tag{4}$$

Since V is unitary, $d = V_1 d_1 + V_2 d_2$. Next, we decouple the output y using a nonsingular transformation

$$T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} I_{p_H} & -U_1^\top R U_2 (U_2^\top R U_2)^{-1} \\ 0 & I_{(l - p_H)} \end{bmatrix} \begin{bmatrix} U_1^\top \\ U_2^\top \end{bmatrix}, \tag{5}$$

and extract a projection of $\overline{y}$ using $\overline{T}_2 = \overline{U}_2^\top$ to obtain

$$\begin{aligned} \dot{x} &= Ax + Bu + G_1 d_1 + G_2 d_2 + w, \\ y &= Cx + Du + H_1 d_1 + v, \\ z_1 &= T_1 y = C_1 x + D_1 u + \Sigma d_1 + v_1, \\ z_2 &= T_2 y = C_2 x + D_2 u + v_2, \\ \overline{z}_2 &= \overline{T}_2 \overline{y} = \overline{C}_2 \dot{x} + \overline{D}_2 \dot{u} + \overline{v}_2, \end{aligned} \tag{6}$$

where $\overline{U}_2$ is obtained from the singular value decomposition of $\overline{H} = \begin{bmatrix} \overline{U}_1 & \overline{U}_2 \end{bmatrix} \begin{bmatrix} \overline{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \overline{V}_1^\top \\ \overline{V}_2^\top \end{bmatrix}$, while $C_1 := T_1 C$, $C_2 := T_2 C = U_2^\top C$, $\overline{C}_2 := \overline{T}_2 \overline{C} = \overline{U}_2^\top \overline{C}$, $D_1 := T_1 D$, $D_2 := T_2 D = U_2^\top D$, $\overline{D}_2 := \overline{T}_2 \overline{D} = \overline{U}_2^\top \overline{D}$, $G_1 := G V_1$, $G_2 := G V_2$, $H_1 := H V_1 = U_1 \Sigma$, $v_1 := T_1 v$, $v_2 := T_2 v = U_2^\top v$ and $\overline{v}_2 := \overline{T}_2 \overline{v} = \overline{U}_2^\top \overline{v}$. Note that with the above transformation, the signal d_2 does not show up in the measurement y, which means that there is insufficient information in the measurement y to estimate the signal d, specifically the component given by d_2, which is nonempty when H is not full rank. Thus, in order to estimate the unknown input signal d in the case with rank deficient H, the additional measurement given by $\overline{y}$ is necessary, as previously stated in Section II.

The similarity transform was also chosen such that the measurement noise terms for the decoupled outputs are uncorrelated. The autocorrelations and correlations of v_1, v_2 and $\overline{v}_2$, with the initial state and process noise are:

$$\begin{aligned}
\mathbb{E}[v_1(t_1)v_1(t_2)^\top] &= T_1R(t_1)T_1^\top\delta(t_1-t_2) := R_1(t_1)\delta(t_1-t_2),\\
\mathbb{E}[v_2(t_1)v_2(t_2)^\top] &= T_2R(t_1)T_2^\top\delta(t_1-t_2)\\
&= U_2^\top R(t_1)U_2\delta(t_1-t_2) := R_2(t_1)\delta(t_1-t_2),\\
\mathbb{E}[v_1(t_1)v_2(t_2)^\top] &= \delta(t_1-t_2)(U_1^\top R(t_1)U_2^\top\\
&\quad - U_1^\top R(t_1)U_2(U_{2,k}^\top R(t_1)U_2)^{-1}U_2^\top R(t_1)U_2) = 0,\\
\mathbb{E}[\overline{v}_2(t_1)\overline{v}_2(t_2)^\top] &= \overline{T}_2\overline{R}(t_1)\overline{T}_2^\top\delta(t_1-t_2) := \overline{R}_2(t_1)\delta(t_1-t_2),\\
\mathbb{E}[v_1(t_1)\overline{v}_2(t_2)^\top] &= T_1\mathbb{E}[v(t_1)\overline{v}(t_2)^\top]\overline{T}_2^\top = 0,\\
\mathbb{E}[v_2(t_1)\overline{v}_2(t_2)^\top] &= T_2\mathbb{E}[v(t_1)\overline{v}(t_2)^\top]\overline{T}_2^\top = 0,\\
\mathbb{E}[v_1(t_1)w(t_2)^\top] &= T_1\mathbb{E}[v(t_1)w(t_2)^\top] = 0,\\
\mathbb{E}[v_2(t_1)w(t_2)^\top] &= T_2\mathbb{E}[v(t_1)w(t_2)^\top] = 0,\\
\mathbb{E}[v_1(t)x_0^\top] &= T_1\mathbb{E}[v(t)x_0^\top] = 0,\\
\mathbb{E}[v_2(t)x_0^\top] &= T_2\mathbb{E}[v(t)x_0^\top] = 0,\\
\mathbb{E}[\overline{v}_2(t_1)w(t_2)^\top] &= \overline{T}_2\mathbb{E}[\overline{v}(t_1)w(t_2)^\top] = 0,\\
\mathbb{E}[\overline{v}_2(t)x_0^\top] &= \overline{T}_2\mathbb{E}[\overline{v}(t)x_0^\top] = 0,
\end{aligned}$$

where R_1, R_2 and $\overline{R}_2$ are positive definite.

B. Filter Description

We know from the discrete-time version of the MVU filter [1], [21] that only the projection of the output y, corresponding to z_2 in this paper, can be considered in the innovation/residual computation that is then used to update the state estimate. This has been shown to be necessary in order that the state estimate is unbiased. On the other hand, the component of the output corresponding to z_1 is used to estimate d_1. In contrast with the discrete-time filter in which the d_2 component can only be estimated with a unit time delay, in the continuous-time version, we use $\overline{z}_2$ for the estimation of d_2. Since the distinction between the propagation and update steps of discrete-time filtering does not exist in continuous time, unlike [1], only one filter structure is considered, which is as follows:

$$\begin{aligned}
\hat{d}_1 &= M_1(z_1 - C_1\hat{x} - D_1u),\\
\hat{d}_2 &= M_2(\overline{z}_2 - \overline{C}_2A\hat{x} - \overline{C}_2Bu - \overline{C}_2G_1\hat{d}_1 - \overline{D}_2\dot{u}), \quad (7)\\
\hat{d} &= V_1\hat{d}_1 + V_2\hat{d}_2,
\end{aligned}$$

$$\dot{\hat{x}} = A\hat{x} + Bu + G_1\hat{d}_1 + G_2\hat{d}_2 + L(z_2 - C_2\hat{x} - D_2u), \quad (8)$$

where the matrices $L \in \mathbb{R}^{n\times(l-p_H)}$, $M_1 \in \mathbb{R}^{p_H\times p_H}$ and $M_2 \in \mathbb{R}^{(p-p_H)\times(l-p_H)}$ are filter gain matrices that are chosen to minimize the state and input error covariances.

A summary of the optimal continuous-time filter is given in Algorithm 1. This algorithm has the following properties, which we will further describe and prove in Section III-C:

Algorithm 1 Input and State Estimation Algorithm

Initialize: $\hat{x}(t_0) = \hat{x}_0$; $P^x(t_0) = \mathcal{P}_0^x$; $M_1 = \Sigma^{-1}$; $\hat{Q} = Q + G_1M_1R_1M_1^\top G_1^\top$; $\hat{A} = A - G_1M_1C_1$;
while $t < t_f$ **do**
 ▷ Unknown input estimation
 $\tilde{R}_2 = \overline{C}_2(\hat{A}P^x\hat{A}^\top + \hat{Q})\overline{C}_2^\top + \overline{R}_2$;
 $M_2 = (G_2^\top\overline{C}_2^\top\tilde{R}_2^{-1}\overline{C}_2G_2)^{-1}G_2^\top\overline{C}_2^\top\tilde{R}_2^{-1}$;
 $\hat{d}_1 = M_1(z_1 - C_1\hat{x} - D_1u)$;
 $\hat{d}_2 = M_2(\overline{z}_2 - \overline{C}_2A\hat{x} - \overline{C}_2Bu - \overline{C}_2G_1\hat{d}_1 - \overline{D}_2\dot{u})$;
 $\hat{d} = V_1\hat{d}_1 + V_2\hat{d}_2$;
 $P^d = V_1M_1(C_1P^xC_1^\top + R_1)M_1^\top V_1^\top + V_2(G_2^\top\overline{C}_2^\top\tilde{R}_2^{-1}\overline{C}_2G_2)^{-1}V_2^\top + V_1M_1C_1P^x\hat{A}^\top\overline{C}_2^\top M_2^\top V_2^\top + V_2M_2\overline{C}_2\hat{A}P^xC_1^\top M_1^\top V_1^\top - V_1M_1R_1M_1^\top G_1^\top\overline{C}_2^\top M_2^\top V_2^\top - V_2M_2\overline{C}_2G_1M_1R_1M_1^\top V_1^\top$;
 ▷ State estimation
 $\overline{A} = (I - G_2M_2\overline{C}_2)\hat{A}$;
 $\overline{Q} = (I - G_2M_2\overline{C}_2)\hat{Q}(I - G_2M_2\overline{C}_2)^\top + G_2M_2\overline{R}_2M_2^\top G_2^\top$;
 $L = P^xC_2^\top R_2^{-1}$;
 $\dot{\hat{x}} = A\hat{x} + Bu + G_1\hat{d}_1 + G_2\hat{d}_2 + L(z_2 - C_2\hat{x} - D_2u)$;
 $\dot{P}^x = \overline{A}P^x + P^x\overline{A}^\top + \overline{Q} - P^xC_2^\top R_2^{-1}C_2P^x$;
end while

Theorem 1 (Minimum-variance unbiased estimation)**.** *If* $\mathrm{rk}(\overline{C}_2G_2) = p - p_H$ *and* $(\overline{A}, C_2)$ *is detectable, where the matrix* $\overline{A}$ *is as defined in Algorithm 1, then the filter gains,* L*,* M_1 *and* M_2*, given in Algorithm 1 provide the unbiased, best linear estimate (BLUE) of the unknown input and the minimum-variance unbiased estimate of system states.*

Theorem 2 (Convergence to steady-state)**.** *Let* $\mathrm{rk}(\overline{C}_2G_2) = p - p_H$*. Then, with* $P^x(t_0) \succeq 0$*, the filter's Riccati differential equation given by*

$$\dot{P}^x = \overline{A}P^x + P^x\overline{A}^\top + \overline{Q} - P^xC_2^\top R_2^{-1}C_2P^x, \quad (9)$$

with $\overline{A}$ *and* $\overline{Q}$ *given in Algorithm 1, (exponentially) converges to a unique stationary solution if and only if*
(i) The pair $(\overline{A}, C_2)$ *is detectable, and*
(ii) The pair $(\overline{A}, \overline{Q}^{\frac{1}{2}})$ *is stabilizable,*
where the matrices $\overline{A}$ *and* $\overline{Q}$ *are as defined in Algorithm 1.*

Remark 1. *In the special case (*$\overline{C} = C$*,* $\overline{D} = D$*,* $\overline{H} = H$*), from which follows* $\overline{C}_2 = C_2$*, a system property known as strong observability*[4] *implies that the pair* $(\overline{A}, C_2)$ *is observable; and that* C_2 *and* G_2 *have full rank. A full-rank* G_2 *is a necessary condition for* $rank(C_2G_2) = p - p_H$*, while* C_2 *with full rank is also necessary if* $l = p$*. Hence, strong observability is closely related to the fact that a minimum-variance unbiased estimator exists and admits a steady-state solution. A similar condition also holds for the optimal discrete-time filter in [1].*

C. Filter Analysis

We first provide a proof of Theorem 1 by showing an equivalence of the problem to a continuous-time system without unknown inputs and as such, we can apply the results of the Kalman-Bucy filter. Then, we provide an alternative

[4]Strong observability is the condition under which the initial condition x_0 and the unknown input signal history, $d(\tau)$ for all $0 \leq \tau \leq t$ can be uniquely determined from the measured output history $y(\tau)$ for all $0 \leq \tau \leq t$ (see, e.g., [22])

derivation for the special case by means of limiting case approximations of the optimal discrete-time filter [1]. In the process, we gain insight into the subtle difference between the special case continuous-time filter and the discrete-time filter in [1]. Finally, we provide a proof of Theorem 2 based on the steady-state Kalman-Bucy filter properties of the equivalent problem.

1) Proof 1: By Equivalent System without Unknown Inputs: As was observed in [1], the unknown input can be viewed as consisting of a known component given by the input estimate, and a zero-mean random variable with known variance which can be dealt in the same manner as with process and measurement noises:

$$\begin{aligned} d_1 &= \hat{d}_1 + (d_1 - \hat{d}_1) := \hat{d}_1 + \tilde{d}_1, \\ d_2 &= \hat{d}_2 + (d_2 - \hat{d}_2) := \hat{d}_2 + \tilde{d}_2. \end{aligned} \tag{10}$$

From (7) and choosing the matrices M_1 and M_2 such that $M_1 \Sigma = I$ and $M_2 \overline{C}_2 G_2 = I$, which is possible because Σ and $\overline{C}_2 G_2$ have full rank by assumption, we obtain

$$\begin{aligned} \tilde{d}_1 &= -M_1(C_1 \tilde{x} + v_1), \\ \tilde{d}_2 &= -M_2 \overline{C}_2 \hat{A} \tilde{x} + M_2 \overline{C}_2 G_1 M_1 v_1 - M_2 \overline{v}_2 - M_2 \overline{C}_2 w, \end{aligned} \tag{11}$$

where $\hat{A} := A - G_1 M_1 C_1$. Since we can design L such that $\tilde{x} := x - \hat{x}$ tends exponentially towards zero (shown below), and the process and measurement noises have zero mean, both $\tilde{d}_1$ and $\tilde{d}_2$ exponentially tend towards zero-mean random variables with the following (auto-)correlations:

$$\begin{aligned} &\mathbb{E}[\tilde{d}_1(t_1)\tilde{d}_1(t_2)^\top] := P_1^d(t_1)\delta(t_1 - t_2) \\ &= (M_1 R_1 M_1^\top + M_1 C_1 P^x C_1^\top M_1^\top)(t_1)\delta(t_1 - t_2), \end{aligned} \tag{12}$$

$$\begin{aligned} &\mathbb{E}[\tilde{d}_1(t_1)\tilde{d}_2(t_2)^\top] := P_{12}^d(t_1)\delta(t_1 - t_2) \\ &= (M_1 C_1 P^x \hat{A}^\top \overline{C}_2^\top M_2^\top - M_1 R_1 M_1^\top G_1^\top \overline{C}_2^\top M_2^\top)(t_1) \\ &\quad \delta(t_1 - t_2), \end{aligned} \tag{13}$$

$$\begin{aligned} &\mathbb{E}[\tilde{d}_2(t_1)\tilde{d}_2(t_2)^\top] := P_2^d(t_1)\delta(t_1 - t_2) \\ &= (M_2 \tilde{R}_2 M_2^\top)(t_1)\delta(t_1 - t_2), \end{aligned} \tag{14}$$

$$\begin{aligned} &\mathbb{E}[\tilde{d}(t_1)\tilde{d}(t_2)^\top] := P^d(t_1)\delta(t_1 - t_2) \\ &= (V_1 P_1^d V_1^\top + V_1 P_{12}^d V_2^\top + V_2 P_{12}^{d\top} V_1^\top + V_2 P_2^d V_2^\top)(t_1) \\ &\quad \delta(t_1 - t_2), \end{aligned} \tag{15}$$

where we defined $\mathbb{E}[\tilde{x}(t_1)\tilde{x}^\top(t_2)] := P^x(t_1)\delta(t_1 - t_2)$, $\tilde{R}_2 := \overline{C}_2(\hat{A} P^x \hat{A}^\top + \hat{Q})\overline{C}_2^\top + \overline{R}_2$ and $\hat{Q} := Q + G_1 M_1 R_1 M_1^\top G_1^\top$, as well as omitted $\mathbb{E}[\tilde{x}(t_1)v_1^\top(t_2)]$, $\mathbb{E}[\tilde{x}(t_1)\overline{v}_2^\top(t_2)]$ and $\mathbb{E}[\tilde{x}(t_1)w^\top(t_2)]$ due to their negligible contributions to the above correlations.

To obtain the best linear unbiased estimate of both projections of the unknown inputs, $\hat{d}_1$ and $\hat{d}_2$, as in its discrete-time counterpart [1], we choose M_1 and M_2 such that the assumption in the Gauss-Markov Theorem is satisfied, as outlined in [18, pp. 96-98]:

$$M_1 = \Sigma^{-1}, M_2 = (G_2^\top \overline{C}_2^\top \tilde{R}_2^{-1} \overline{C}_2 G_2)^{-1} G_2^\top \overline{C}_2^\top \tilde{R}_2^{-1}. \tag{16}$$

Next, substituting (10) into the system dynamics (6), and using (8) and (11), we obtain the state estimate error system

$$\dot{\tilde{x}} = \dot{x} - \dot{\hat{x}} = \overline{A}\tilde{x} + \overline{w} - L(C_2 \tilde{x} + v_2), \tag{17}$$

where $\overline{A} := (I - G_2 M_2 \overline{C}_2)\hat{A}$ and $\overline{w} := (I - G_2 M_2 \overline{C}_2)w - (I - G_2 M_2 \overline{C}_2)G_1 M_1 v_1 - G_2 M_2 \overline{v}_2$. Note that even during transients, where the $\tilde{d}_1$ and $\tilde{d}_2$ have non-zero means, the terms contributing to these biases are functions of $\tilde{x}$ and are thus absorbed into the $\overline{A}$. More importantly, the state estimate error dynamics in (17) is the same state estimate error dynamics of a Kalman-Bucy filter [2] for a linear system without unknown inputs:

$$\dot{x}_e = \overline{A} x_e + \overline{w}, \quad y_e = C_2 x_e + v_2.$$

Since the objective of both systems is the same, i.e. to obtain an unbiased minimum-variance filter, they are equivalent systems from the perspective of optimal filtering. Hence, the optimal filter is obtained, as in [2], when L is chosen as

$$L = P^x C_2^\top R_2^{-1}, \tag{18}$$

where the state estimate error covariance, P^x, is obtained from the Riccati differential equation given in (9) where the noise intensity, $\overline{Q}$, is obtained from

$$\begin{aligned} &\mathbb{E}[\overline{w}(t_1)\overline{w}^\top(t_2)] := \overline{Q}(t_1)\delta(t_1 - t_2) \\ &= ((I - G_2 M_2 \overline{C}_2)\hat{Q}(I - G_2 M_2 \overline{C}_2)^\top \\ &+ G_2 M_2 \overline{R}_2 M_2^\top G_2^\top)(t_1)\delta(t_1 - t_2). \end{aligned} \tag{19}$$

Moreover, since we assume that $(\overline{A}, C_2)$ is detectable, the eigenvalues of $\overline{A} - LC_2$ can be designed to have negative real parts. Since the linear error system (17) is stable, $\mathbb{E}[\tilde{x}]$ tends to zero, and the effect of the initial condition $\hat{x}_0$ decays exponentially. In summary, the proposed filter provides the best linear unbiased estimate of the unknown input and the minimum-variance unbiased estimate of the state; thus, we have proven the claims of Theorem 1.

2) Proof 2: By Limiting Case Approximations: An alternate derivation of the optimal filter for the special case ($\overline{C} = C$, $\overline{D} = D$, $\overline{H} = H \Rightarrow \overline{C}_2 = C_2$) can be obtained from the optimal discrete-time filter [1] using limiting case approximations. Although this derivation lacks rigor due to various approximations, this is interesting for understanding the connection between the continuous- and discrete-time filters and also from a pedagogical point of view, since this is often used to derive the continuous-time Kalman-Bucy filters in textbooks (e.g., [17], [23]).

If the sampling period $\triangle t$ is small, we can use Euler's approximation to write the discretized version of (6) as

$$\begin{aligned} x_{k+1} &\approx (I + A\triangle t)x_k + B\triangle t u_k + G_1 \triangle t d_{1,k} \\ &\qquad + G_2 \triangle t d_{2,k} + w_k \\ &:= A_d x_k + B_d u_k + G_{1,d} d_{1,k} + G_{2,d} d_{2,k} + w_k, \\ y_k &= C x_k + D u_k + H_1 d_k + v_k \\ &:= C_d x_k + D_d u_k + H_{1,d} d_k + v_k, \\ z_{1,k} &= C_1 x_k + D_1 u_k + \Sigma d_{1,k} + v_{1,k} \\ &:= C_{1,d} x_k + D_{1,d} u_k + \Sigma_d d_{1,k} + v_{1,k}, \\ z_{2,k} &= C_2 x_k + D_2 u_k + v_{2,k} := C_{2,d} x_k + D_{2,d} u_k + v_{2,k}, \end{aligned} \tag{20}$$

where the process and measurement noises are $w_k \sim (0, Q\triangle t)$ and $v_k \sim (0, R/\triangle t)$, in which $Q_d \approx Q\triangle t$ as $\triangle t \to 0$, and the discrete measurement noise is approximated as the average value of the continuous noise [16].

Since the first component of the unknown input can computed pointwise without delay, we expect $M_{1,k} \to M_1$. Thus, we have the estimate $\hat{d}_1$ as in (7) directly from the discrete-time version given by $\hat{d}_{1,k}^D = M_{1,k}(z_{1,k} - C_{1,d}\hat{x}_{k|k} - D_{1,d} u_k)$

[1]. On the other hand, the limiting case approximation of the second component of the unknown input is given by:

$$\begin{aligned}\hat{d}^D_{2,k-1} &= M_{2,k}(z_{2,k} - C_{2,d}(A_d\hat{x}_{k-1|k-1} + B_d u_{k-1} \\ &\quad + G_{1,d}\hat{d}_{1,k-1}) - D_{2,d}u_k) \\ &\approx M_{2,k}\triangle t(\tfrac{z_{2,k} - C_{2,d}\hat{x}_{k-1|k-1} - D_{2,d}u_{k-1}}{\triangle t} - D_2\tfrac{u_k - u_{k-1}}{\triangle t} \\ &\quad - C_2 B u_{k-1} - C_2 A\hat{x}_{k-1|k-1} - C_2 G_1 \hat{d}_{1,k-1}) \\ &= M_{2,k}\triangle t(\tfrac{z_{2,k} - \hat{z}_{2,k-1}}{\triangle t} - C_2 A\hat{x}_{k-1|k-1} - D_2\tfrac{u_k - u_{k-1}}{\triangle t} \\ &\quad - C_2 B u_{k-1} - C_2 G_1 \hat{d}_{1,k-1}),\end{aligned}$$

where the first equation is the discrete-time version from [1], and we defined $M_2 := \lim_{\triangle t\to 0} M_{2,k}\triangle t$ and $\hat{z}_{k-1} := T_{2,d}\hat{y}_{k-1} := T_{2,d}(C_d\hat{x}_{k-1|k-1} + D_d u_{k-1})$, as well as substituted the approximate matrices $A_d \approx I + A\triangle t$, $B_d \approx B\triangle t$, $C_{2,d} \approx C_2$, $D_{2,d} \approx D_2\triangle t$ and $G_{1,d} \approx G_1\triangle t$ from (20). Taking the limit of $\triangle t \to 0$, we obtain

$$\hat{d}_2 = M_2(\overline{z}_2 - C_2 A\hat{x} - C_2 B u - C_2 G_1 \hat{d}_1 - D_2\dot{u}), \quad (21)$$

where we have replaced the term $\lim_{\triangle t\to 0}\frac{z_{2,k} - \hat{z}_{2,k-1}}{\triangle t} = T_2(\lim_{\triangle t\to 0}\frac{y_k - y_{k-1}}{\triangle t} + \frac{C_d\tilde{x}_{k-1|k-1} + v_{k-1}}{\triangle t})$ with $\overline{z}_2 = T_2\overline{y} = T_2(\lim_{\triangle t\to 0}\frac{y_k - y_{k-1}}{\triangle t} + \frac{\overline{v}_d\triangle t - (v_k - v_{k-1})}{\triangle t})$ which we assume is obtained from the noisy measurement of $\overline{y}$ according to (6) (with $\overline{C} = C$, $\overline{D} = D$, $\overline{H} = H$) and where we have defined $\tilde{x}_{k-1|k-1} := x_{k-1} - \hat{x}_{k-1|k-1}$. This indirectly implies that the optimal discrete-time filter "differentiates" the second projection of the output, z_2, using finite difference. Moreover, we can infer that the equivalent discrete-time estimation of $d_{2,k-1}$ corresponding to (21) is given by

$$\begin{aligned}\hat{d}_{2,k-1} &= M_{2,k}(z_{2,k} - C_{2,d}(A_d\hat{x}_{k-1|k-1} + B_d u_{k-1} \quad (22) \\ &+ G_{1,d}\hat{d}_{1,k-1}) - D_{2,d}u_k - C_{2,d}\tilde{x}_{k-1|k-1} + \overline{v}_{2,d}\triangle t - v_{2,k}),\end{aligned}$$

which would be unimplementable in discrete time because the noise terms and the true state are not available. Thus, to obtain the best linear unbiased estimate of both projections of the unknown inputs, $\hat{d}_{1,k} = \hat{d}^D_{1,k}$ and $\hat{d}_{2,k-1}$, we choose $M_{1,k}$ and $M_{2,k}$ as in [1], such that $M_{1,k}\Sigma_d = I$, $M_{2,k}C_{2,d}G_{2,d} = I$ and the assumption in the Gauss-Markov Theorem is satisfied [18, pp. 96-98]:

$$\begin{aligned}M_{1,k} &= \Sigma_d^{-1}, \\ M_{2,k} &= (G_{2,d}^\top C_{2,d}^\top \tilde{R}_{2,k}^{-1} C_{2,d} G_{2,d})^{-1} G_{2,d}^\top C_{2,d}^\top \tilde{R}_{2,k}^{-1},\end{aligned} \quad (23)$$

where $\tilde{R}_{2,k} := C_{2,d}(\hat{A}_d - I)P^x_{k-1|k-1}(\hat{A}_d - I)^\top C_{2,d}^\top + C_{2,d}\hat{Q}_{k-1}C_{2,d}^\top + \overline{R}_2\triangle t$, $\hat{A}_d := A_d - G_{1,d}M_{1,k}C_{1,d}$ and $\hat{Q}_{k-1} := Q_d + G_{1,d}M_{1,k-1}R_{1,d}M_{1,k-1}^\top G_{1,d}^\top$, and we have applied $\mathbb{E}[\overline{v}_{2,d}\overline{v}_{2,d}^\top] := \overline{R}_{2,d} \approx \frac{\overline{R}_2}{\triangle t}$. Then, using the approximate matrices (20), substituting$P^x_{k-1|k-1} \approx \frac{P^x}{\triangle t}$ and $\tilde{R}_{2,k} \approx \frac{\tilde{R}_2}{\triangle t}$, as well as taking the limit of $\triangle t \to 0$, we obtain $M_1 = \lim_{\triangle t\to 0} M_{1,k}$ and $M_2 = \lim_{\triangle t\to 0} M_{2,k}$ as are given in (16).

Furthermore, with $\hat{d}_{1,k} = \hat{d}^D_{1,k}$ and $\hat{d}_{2,k}$ given by (22), the unknown input estimate errors are given by:

$$\begin{aligned}\tilde{d}_{1,k} &:= d_{1,k} - \hat{d}_{1,k} = -M_{1,k}C_{1,d}\tilde{x}_{k|k} - M_{1,k}v_{1,k}, \\ \tilde{d}_{2,k-1} &:= d_{2,k-1} - \hat{d}_{2,k-1} \\ &= -M_{2,k}C_{2,d}(\hat{A}_d - I)\tilde{x}_{k-1|k-1} - M_{2,k}C_{2,d}w_{k-1} \\ &\quad + M_{2,k}C_{2,d}G_{1,d}M_{1,k-1}v_{1,k-1} - M_{2,k}\overline{v}_{2,d}\triangle t.\end{aligned}$$

Hence, the error covariance matrices can be computed as

$$\begin{aligned}P^d_{1,k} &:= \mathbb{E}[\tilde{d}_{1,k}\tilde{d}_{1,k}^\top] = M_{1,k}(C_{1,d}P^x_{k|k}C_{1,d}^\top + R_{1,k})M_{1,k}^\top, \\ P^d_{12,k-1} &:= \mathbb{E}[\tilde{d}_{1,k-1}\tilde{d}_{2,k-1}^\top] \\ &\approx M_{1,k-1}C_{1,d}P^x_{k-1|k-1}(\hat{A}_d^\top - I)C_{2,d}^\top M_{2,k} \\ &\quad - M_{1,k-1}R_{1,k-1}M_{1,k-1}^\top G_{1,d}^\top C_{2,d}^\top M_{2,k}^\top, \\ P^d_{2,k-1} &:= \mathbb{E}[\tilde{d}_{2,k-1}\tilde{d}_{2,k-1}^\top] = M_{2,k}\tilde{R}_{2,k}M_{2,k}^\top.\end{aligned}$$

As above, substituting the approximate matrices and taking the limit of $\triangle t \to 0$, we obtain the expressions for $P^d_1 := \lim_{\triangle t\to 0} P^d_{1,k}\triangle t$, $P^d_{12} := \lim_{\triangle t\to 0} P^d_{12,k-1}\triangle t$ and $P^d_2 := \lim_{\triangle t\to 0} P^d_{2,k-1}\triangle t$ given by (12), (13) and (14).

Next, we derive the continuous-time state estimate and error covariance dynamics from the ULISE variant of the discrete-time filter [1] in a similar manner:

$$\begin{aligned}\hat{x}^\star_{k|k} &= A_d\hat{x}_{k-1|k-1} + B_d u_{k-1} + G_{1,d}\hat{d}_{1,k-1} + G_{2,d}\hat{d}_{2,k-1}, \\ \hat{x}_{k|k} &= \hat{x}^\star_{k|k} + \tilde{L}_k(z_{2,k} - C_{2,d}\hat{x}^\star_{k|k} - D_{2,d}u_k) \\ &\approx (I + A\triangle t)\hat{x}_{k-1|k-1} + B\triangle t u + G_1\triangle t\hat{d}_1 \\ &\quad + G_2\triangle t\hat{d}_2 + \tilde{L}_k(z_2 - C_2(I + A\triangle t)\hat{x}_{k-1|k-1} \\ &\quad - D_2 u - C_2 B\triangle t u + G_1\triangle t\hat{d}_1 + G_2\triangle t\hat{d}_2), \\ \Rightarrow \frac{\hat{x}_{k|k} - \hat{x}_{k-1|k-1}}{\triangle t} &\approx \frac{\tilde{L}_k}{\triangle t}(z_2 - C_2\hat{x}_{k-1|k-1} - D_2 u) \\ &\quad + A\hat{x}_{k-1|k-1} + Bu + G_1\hat{d}_1 + G_2\hat{d}_2,\end{aligned} \quad (24)$$

where we have neglected higher order terms in the latter term in the final equation above. Taking the limit of $\triangle t \to 0$, we obtain the continuous state estimate dynamics as above, i.e. (8), if we define $L := \lim_{\triangle t\to 0}\frac{\tilde{L}_k}{\triangle t}$, as is done for the continuous-time Kalman-Bucy filter. From the state estimate error covariance given by

$$\begin{aligned}P^x_{k|k} &:= \mathbb{E}[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^\top] \\ &= P^{\star x}_{k|k} - \tilde{L}_k C_{2,d}P^{\star x}_{k|k} - P^{\star x}_{k|k}C_{2,d}^\top\tilde{L}_k^\top + \tilde{L}_k\check{R}_k\tilde{L}_k^\top,\end{aligned} \quad (25)$$

where $P^{\star x}_{k|k} := \mathbb{E}[(x_k - \hat{x}^\star_{k|k})(x_k - \hat{x}^\star_{k|k})^\top]$ and $\check{R}_k := C_{2,d}P^{\star x}_{k|k}C_{2,d}^\top + R_{2,d}$, we can find the optimal $\tilde{L}_k$ by minimizing the trace of (25) to obtain

$$\begin{aligned}\tilde{L}_k &= P^{\star x}_{k|k}C_{2,d}^\top(C_{2,d}P^{\star x}_{k|k}C_{2,d}^\top + R_{2,d})^{-1}, \\ \Rightarrow \frac{\tilde{L}_k}{\triangle t} &= P^{\star x}_{k|k}C_{2,d}^\top(C_{2,d}P^{\star x}_{k|k}C_{2,d}^\top\triangle t + R_{2,d}\triangle t)^{-1}.\end{aligned} \quad (26)$$

As $\triangle t \to 0$, if $P^{\star x}_{k|k}$ is finite, and we let $P^{\star x}_{k|k} \to P^x$, then we obtain the filter gain L as given in (18).

Substituting the optimal $\tilde{L}_k$ (26) into (25), we obtain $P^x_{k|k}$, and into the discrete-time optimal state estimation dynamics (24) to derive the state error covariance matrix, $P^{\star x}_{k|k}$ (similar derivation in [1]):

$$P^x_{k|k} = P^{\star x}_{k|k} - P^{\star x}_{k|k}C_{2,d}^\top\check{R}_k^{-1}C_{2,d}P^{\star x}_{k|k}, \quad (27)$$

$$\begin{aligned}P^{\star x}_{k|k} &= (\hat{A}_d - G_{2,d}M_{2,k}C_{2,d}(\hat{A}_d - I))P^x_{k-1|k-1} \\ &\quad (\hat{A}_d - G_{2,d}M_{2,k}C_{2,d}(\hat{A}_d - I))^\top + (I - G_{2,d}M_{2,k}C_{2,d}) \\ &\quad \hat{Q}_{k-1}(I - G_{2,d}M_{2,k}C_{2,d})^\top + G_{2,d}M_{2,k}\overline{R}_{2,d}M_{2,k}^\top G_{2,d}^\top(\triangle t)^2.\end{aligned}$$

Applying the approximation matrices defined in (20) and neglecting higher order terms, we have

$$\begin{aligned}\frac{P^{\star x}_{k|k} - P^{\star x}_{k-1|k-1}}{\triangle t} &\approx \overline{A}P^{\star x}_{k-1|k-1} + P^{\star x}_{k-1|k-1}\overline{A}^\top \\ &+ (I - G_2M_2C_2)\hat{Q}(I - G_2M_2C_2)^\top + G_2M_2\overline{R}_2M_2^\top G_2^\top \\ &- P^{\star x}_{k-1|k-1}C_2^\top(C_2P^{\star x}_{k-1|k-1}C_2^\top\triangle t + R_2)^{-1}C_2P^{\star x}_{k-1|k-1}.\end{aligned}$$

Next, if $P^{\star x}_{k|k}$ is finite, as $\triangle t \to 0$, we obtain the Riccati differential equation governing P^x given in (9), where we applied $P^x = P^{\star x}$, which can be deduced from $\hat{x}_{k|k} \approx \hat{x}^{\star}_{k|k} + G_2 \hat{d}_2 \triangle t$ by taking the limit of $\triangle t \to 0$.

3) Filter Convergence and Optimality: For linear time-invariant systems, the conditions for the convergence of the filter gains to steady-state of the proposed filter are closely related to the existence and the uniqueness of stabilizing solutions of its continuous-time algebraic Riccati equation (CARE), i.e. (9) with $\dot{P^x} = 0$. These conditions are given in Theorem 2 and the proof of the results can be found in [18, Sections 16.7-16.8]. The optimality of the filter also follows from the equivalence of the proposed filter to a Kalman-Bucy filter [2] without unknown inputs. From the perspective of limiting case approximations, the discrete-time filter in [1] is globally optimal and converges to a steady-state solution for arbitrary $\triangle t$, and the Euler approximation converges to the continuous system. So, from the optimality of the Kalman-Bucy filter, it can be inferred that limiting case filter is also optimal. Furthermore, for the special case ($\overline{C}_2 = C_2$) the connection between strong observability and the observability of $(\overline{A}, C_2)$, as well as C_2 and G_2 being full rank (Remark 1) follow directly from

$$\begin{aligned}
n+p &= \mathrm{rk}\begin{bmatrix} sI-A & -G \\ C & H \end{bmatrix} = \mathrm{rk}\begin{bmatrix} sI-A & -G \\ C & U\begin{bmatrix}\Sigma & 0 \\ 0 & 0\end{bmatrix}V^\top \end{bmatrix} \\
&= \mathrm{rk}\begin{bmatrix} I & 0 \\ 0 & T \end{bmatrix}\begin{bmatrix} sI-A & -G \\ C & U\begin{bmatrix}\Sigma & 0 \\ 0 & 0\end{bmatrix}V^\top \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & V \end{bmatrix} \\
&= \mathrm{rk}\begin{bmatrix} sI-A & -GV \\ TC & TU\begin{bmatrix}\Sigma & 0 \\ 0 & 0\end{bmatrix} \end{bmatrix} = \mathrm{rk}\begin{bmatrix} sI-A & -G_1 & -G_2 \\ C_1 & \Sigma & 0 \\ C_2 & 0 & 0 \end{bmatrix} \\
&= \mathrm{rk}\begin{bmatrix} I & G_1\Sigma^{-1} & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}\begin{bmatrix} sI-A & -G_1 & -G_2 \\ C_1 & \Sigma & 0 \\ C_2 & 0 & 0 \end{bmatrix} \\
&= \mathrm{rk}\begin{bmatrix} sI-\hat{A} & 0 & -G_2 \\ C_1 & \Sigma & 0 \\ C_2 & 0 & 0 \end{bmatrix} = \mathrm{rk}\begin{bmatrix} sI-\hat{A} & -G_2 \\ C_2 & 0 \end{bmatrix} + p_H \\
&= \mathrm{rk}\begin{bmatrix} sI-\hat{A} & -G_2 \\ C_2 & 0 \end{bmatrix}\begin{bmatrix} I & 0 \\ -M_2C_2\hat{A} & I \end{bmatrix} + p_H \\
&= \mathrm{rk}\begin{bmatrix} sI-\overline{A} & -G_2 \\ C_2 & 0 \end{bmatrix} + p_H,
\end{aligned}$$

where the first equality is the rank condition for strong observability given in [22], the third from last equality holds because Σ is square and has full rank p_H, and we have assumed that $n \geq l \geq 1$ and $l \geq p \geq 0$.

IV. State-Only Estimation

In this section, we once again consider the special case with the restriction that the measurement $\overline{y}$, as well as $\dot{u}$ are not available. As argued in Section III-A, the unknown input cannot be (fully) estimated, thus the objective of this section is to derive a state-only estimation algorithm that does not require the measurement of $\overline{y}$.

The approach taken for this derivation is to modify the input and state estimation algorithm in the previous section (Algorithm 1) to accommodate the absence of the measurement $\overline{y}$ by replacing $\overline{z}_2$ with $\dot{z}_2 := T_2\dot{y}$ in (7), for which we only have the measurement of y. However, this substitution requires that the measurement noise term v is differentiable, even when $\dot{y}$ is never explicitly computed. Therefore, we further restrict the class of problems to the case in which $v = 0$ (an assumption we will relax in a future work). On the other hand, we note that the computation of L in (18), we require the invertibility of R_2. Thus, a fictitious zero-mean noise v_2 with noise intensity $R_2 \succ 0$ is added to the measurement z_2, i.e. $z_2' = z_2 + v_2$. Finally, to circumvent the need to have direct access to $\dot{y}$ and $\dot{u}$, we propose an algorithm in Section IV-A that produces the same state estimate as (8) with only y and u, which are known.

Algorithm 2 State-Only Estimation Algorithm

Initialize: $\hat{x}(t_0) = \hat{x}_0$; $P^x(t_0) = \mathcal{P}^x_0$; $M_1 = \Sigma^{-1}$; $\hat{A} = A - G_1M_1C_1$; $\hat{B} = B - G_1M_1D_1$;
while $t < t_f$ **do**
  $\tilde{R}_2 = C_2(\hat{A}P^x\hat{A}^\top + Q)C_2^\top$;
  $M_2 = (G_2^\top C_2^\top \tilde{R}_2^{-1} C_2 G_2)^{-1} G_2^\top C_2^\top \tilde{R}_2^{-1}$;
  $\overline{A} = (I - G_2M_2C_2)\hat{A}$;
  $\overline{B} = (I - G_2M_2C_2)\hat{B}$;
  $\overline{G} = (I - G_2M_2C_2)G_1$;
  $\overline{Q} = (I - G_2M_2C_2)Q(I - G_2M_2C_2)^\top$;
  $L = P^xC_2^\top R_2^{-1}$;
  $v_2 \sim \mathcal{N}(0, R_2)$;
  $\dot{\theta} = (\overline{A} - LC_2)\hat{x} + (\overline{B} - LD_2)u + \overline{G}M_1z_1 + L(z_2 + v_2)$;
  $\hat{x} = G_2M_2z_2 - G_2M_2D_2u + \theta$;
  $\dot{P^x} = \overline{A}P^x + P^x\overline{A}^\top + \overline{Q} - P^xC_2^\top R_2^{-1}C_2P^x$;
end while

In brief, the modifications result in an algorithm similar to Algorithm 1, but with $R_1 = 0$ and $\overline{R}_2 = 0$, and with the same state estimate in Section IV-A. A summary of the new algorithm is given in Algorithm 2. Given the equivalence of the algorithms in the special case, it follows directly that Theorem 2 and Remark 1 hold, i.e. the condition $\mathrm{rk}(C_2G_2) = p - p_H$ and the *strong observability* of the system are sufficient conditions for the existence of a steady-state filter. Note, however, that the addition of a fictitious noise may degrade the performance of the proposed filter.

A. State-Only Estimation Algorithm

After rearranging and combing terms, the state estimation (8) can be rewritten as follows:

$$\begin{aligned}
\dot{\hat{x}} &= A\hat{x} + Bu + G_1M_1(z_1 - C_1\hat{x} - D_1u) \\
&\quad + G_2M_2(\dot{z}_2 - C_2A\hat{x} - C_2Bu - C_2G_1M_1(z_1 \\
&\quad - C_1\hat{x} - D_1u) - D_2\dot{u}) + L(z_2' - C_2\hat{x} - D_2u) \\
&= (\overline{A} - LC_2)\hat{x} + (\overline{B} - LD_2)u + \overline{G}M_1z_1 \\
&\quad + Lz_2' + G_2M_2\dot{z}_2 - G_2M_2D_2\dot{u} \\
&:= h(\hat{x}, u, z_1, z_2') + \Phi_1\dot{z}_2 + \Phi_2\dot{u},
\end{aligned} \tag{28}$$

where $\overline{B} := (I - G_2M_2C_2)(B - G_1M_1D_1)$, $\overline{G} = (I - G_2M_2C_2)G_1M_1$, and z_2 is replaced with $z_2' = z_2 + v_2$, where v_2 is a fictitious zero-mean noise with a chosen intensity R_2. Then, to derive an equivalent without $\dot{y}$ and $\dot{u}$, we let

$$\dot{\theta} = h(\check{x}, u, z_1, z_2'),\ \ \check{x} = \Phi_1z_2 + \Phi_2u + \theta. \tag{29}$$

Taking the derivative of $\check{x}$, we have

$$\dot{\check{x}} = \Phi_1\dot{z}_2 + \Phi_2\dot{u} + \dot{\theta} = \Phi_1\dot{z}_2 + \Phi_2\dot{u} + h(\check{x}, u, z_1, z_2').$$

So the output $\breve{x}$ of (29) is identical to that of $\hat{x}$ in (28). However, (29) does not include $\dot{z}_2$ and $\dot{u}$, as desired.

V. Separation Principle

We now investigate the stability of the closed-loop system, when the controller is a state feedback controller with disturbance rejection terms using the estimates from (7),(8):

$$u = -K\hat{x} - J_1\hat{d}_1 - J_2\hat{d}_2, \tag{30}$$

where K is the state feedback gain, and J_1 and J_2 are the disturbance rejection gains ($J_1 = J_2 = 0$ in Algorithm 2). The following theorem shows that a separation principle for linear stochastic systems with unknown input holds, i.e., the designs of the controller and the state and input estimator can be carried out independently.

Theorem 3 (Separation Principle)**.** *The state feedback controller gain K in* (30) *can be designed independently of the state estimator gain L in* (18) *(cf. Algorithms 1 and 2).*

Proof. Substituting (30) into (6) and from (17),(10), (11) (for ALISE, with $\overline{C} = C$, $\overline{D} = D$ and $\overline{H} = H$), we have

$$\begin{aligned}\begin{bmatrix}\dot{x}\\ \dot{\tilde{x}}\end{bmatrix} &= \begin{bmatrix}A-BK & B(K - J_1M_1C_1 - J_2M_2\overline{C}_2\hat{A})\\ 0 & \overline{A} - LC_2\end{bmatrix}\begin{bmatrix}x\\ \tilde{x}\end{bmatrix}\\ &\quad + \begin{bmatrix}G_1 - BJ_1 & G_2 - BJ_2\\ 0 & 0\end{bmatrix}\begin{bmatrix}d_1\\ d_2\end{bmatrix}\\ &\quad + \begin{bmatrix}I - BJ_2M_2\overline{C}_2 & 0 & \begin{matrix}BJ_2M_2\overline{C}_2G_1M_1\\ -BJ_1M_1\end{matrix} & 0 & -BJ_2M_2\\ 0 & I & 0 & -L & 0\end{bmatrix}\begin{bmatrix}w\\ \overline{w}\\ v_1\\ v_2\\ \overline{v}_2\end{bmatrix}.\end{aligned} \tag{31}$$

Since the state matrix is triangular, the eigenvalues are of $A - BK$ and those of $\overline{A} - LC_2$. Thus the stability of the state and input feedback and estimator are independent. ∎

Hence, the state feedback gain K can be independently designed (e.g., with Linear Quadratic Regulator (LQR)) with no effect on the stability of the estimator in Algorithms 1 and 2. Moreover, J_1 and J_2 in Algorithm 1 can be chosen such that the effect of disturbance input on the closed loop system is reduced. However, J_1 and J_2 must be chosen such that u, $\hat{d}_1$ and $\hat{d}_2$ can be uniquely determined, since the equations for $\hat{d}_1$ and $\hat{d}_2$ in (7) become implicit equations and the dependence on $\dot{u}$ in $\hat{d}_2$ also results in a differential equation for u.

VI. Illustrative Example

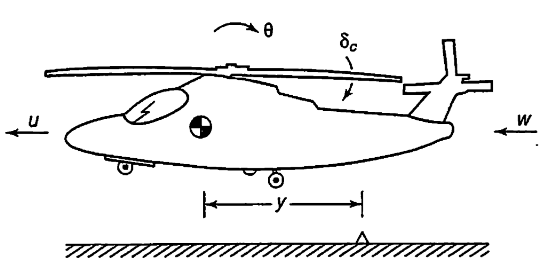

Fig. 1. A helicopter near hover [24].

We consider the linearized longitudinal dynamics of a helicopter [24] depicted in Fig. 1:

$$\begin{aligned}\dot{\theta} &= q,\\ \dot{q} &= -0.415q - 0.011u + 6.27\delta_c - 0.011w_h,\\ \dot{u} &= 9.8\theta - 1.43q - 0.0198u + 9.8\delta_c - 0.0198w_h,\\ \dot{y} &= u,\end{aligned} \tag{32}$$

where the system states are the fuselage pitch angle θ, the pitch rate $\dot{q}$, the horizontal velocity of the center of gravity u and the horizontal distance from the desired hover point y; while the only control input is the tilt angle of the rotor thrust vector δ_c. The variable $w_h = w_d + w$ represents a horizontal wind disturbance with a deterministic time-varying component, w_d, and a stochastic component, w, modeled as one of the following:

(i) A first order Gauss-Markov process, i.e. $\dot{w} = -0.2w + 6\xi$, driven by a zero-mean, continuous time, Gaussian white noise, ξ, with intensity $Q = 5 \times 10^{-4}$.
(ii) A zero-mean, continuous time, Gaussian white noise, w, with intensity $Q = 5 \times 10^{-4}$.

In both cases, we have measurements of y, u and q only, each with a measurement noise intensity of 1×10^{-3}, 1.6×10^{-3} and 0.9×10^{-3}, respectively, while $\overline{y}$ is assumed to be available with $\overline{C} = C$, $\overline{D} = D$ and $\overline{H} = H$ and similarly are noisy with noise intensity of 2×10^{-3}, 1×10^{-3} and 1.9×10^{-3}, respectively. Furthermore, the measurement of u is plagued by the presence of an additive a time-varying bias, e_m, which in this example is a sinusoidal signal.

Since there is a separation principle for the controller and estimator of this system (Section V), we design them independently. The controller we chose is the LQR, whereas the estimator is the filter proposed in this paper. For the LQR gain, K, we have chosen the following cost matrices: $Q_{LQR} = C_{LQR}^\top C_{LQR}$ and $R_{LQR} = 5$, where $C_{LQR} := \begin{bmatrix}0 & 0 & 0 & 1\end{bmatrix}$, while $J_1 = 0$ and $J_2 = -1.943 \times 10^{-3}$ are chosen to minimize $|G_1 - BJ_1|$ and $|G_2 - BJ_2|$.

We implemented the LQR state feedback control law and the filter described above in MATLAB/Simulink on a 2.2 GHz Intel Core i7 CPU for both cases and the results are shown in Fig. 2, when the horizontal wind is modeled as the sum of a deterministic time-varying component and a first order Gauss-Markov process (Case (i)) and in Fig. 3, when the horizontal wind is modeled as the sum of a deterministic time-varying component and a zero-mean Gaussian white noise (Case (ii)). Note that the projections of the unknown input vector, i.e. d_1 and d_2, obtained with the transformation (5), correspond to real unknown signals, in that $d_1 = e_m$ and $d_2 = w_d$. Thus, we observe from the figures that the proposed filter successfully estimates the system states and also the unknown inputs, w_d and e_m, and the traces of the continuous estimate error covariance matrices of both states and unknown inputs converge in less than 0.5 ms. Note that in Fig. 2 and 3, the measured $\dot{y}$ (green) is faulty, hence the estimates (blue) in both cases were shown to be capable of tracking the actual/true value (red) while rejecting the fault.

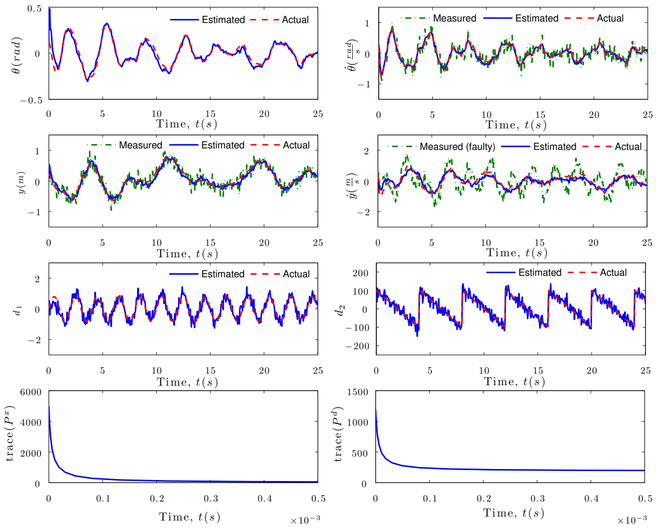

Fig. 2. Actual states θ, q, y, u and their estimates; unknown inputs d_1, d_2, and their estimates; as well as trace of estimate error covariance of states and unknown inputs, when the horizontal wind is modeled as the sum of a deterministic time-varying component and a first order Gauss-Markov process (Case (i)).

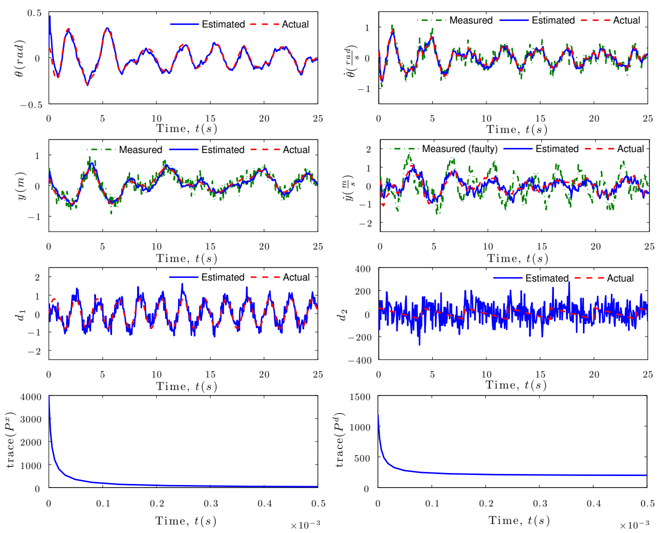

Fig. 3. Actual states θ, q, y, u and their estimates; unknown inputs d_1, d_2, and their estimates; as well as trace of estimate error covariance of states and unknown inputs, when the horizontal wind is modeled as the sum of a deterministic time-varying component and a zero-mean Gaussian white noise (Case (ii)).

VII. Conclusion

This paper presented an optimal filter for linear time-invariant continuous-time stochastic systems that simultaneously estimates the states and unknown inputs in an unbiased minimum-variance sense. The proposed filter was derived by reducing the system to an equivalent system such that the Kalman-Bucy filtering techniques can be directly applied, and by limiting case approximations of the optimal discrete-time filter for a special case. We also provided conditions under which the proposed filter has a steady-state solution, and presented a complementary algorithm for the special case when the 'required' additional measurement is not available. Moreover, a principle of separation of estimation and control was shown to also hold for linear systems with unknown inputs, and the simulation of a helicopter hovering example demonstrates these claims. A possible future direction is the extension of the filtering techniques to linear time-varying continuous-time systems and nonlinear systems.

Acknowledgments

The work presented in this paper was supported by the National Science Foundation, grant #1239182. M. Zhu is partially supported by ARO W911NF-13-1-0421 (MURI).

References

[1] S.Z. Yong, M. Zhu, and E. Frazzoli. A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems. *Automatica*, 2014. Provisionally accepted. Available from: `http://arxiv.org/abs/1309.6627`.

[2] R.E. Kalman and R.S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(3):95–108, 1961.

[3] P.K. Kitanidis. Unbiased minimum-variance linear state estimation. *Automatica*, 23(6):775–778, November 1987.

[4] G. De Nicolao, G. Sparacino, and C. Cobelli. Nonparametric input estimation in physiological systems: Problems, methods, and case studies. *Automatica*, 33(5):851–870, 1997.

[5] R. Patton, R. Clark, and P.M. Frank. *Fault diagnosis in dynamic systems: theory and applications*. Prentice-Hall international series in systems and control engineering. Prentice Hall, 1989.

[6] C.S. Hsieh. Robust two-stage Kalman filters for systems with unknown inputs. *IEEE Transactions on Automatic Control*, 45(12):2374–2378, December 2000.

[7] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*, 43(1):111–116, January 2007.

[8] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, 43(5):934–937, 2007.

[9] H. Fang, Y. Shi, and J. Yi. On stable simultaneous input and state estimation for discrete-time linear systems. *International Journal of Adaptive Control and Signal Processing*, 25(8):671–686, 2011.

[10] S.Z. Yong, M. Zhu, and E. Frazzoli. Simultaneous input and state estimation for linear discrete-time stochastic systems with direct feedthrough. In *Conference on Decision and Control (CDC)*, pages 7034–7039, 2013.

[11] M. Corless and J. Tu. State and input estimation for a class of uncertain systems. *Automatica*, 34(6):757–764, 1998.

[12] Y. Xiong and M. Saif. Unknown disturbance inputs estimation based on a state functional observer design. *Automatica*, 39(8):1389–1398, 2003.

[13] R. Isermann. Model-based fault-detection and diagnosis status and applications. *Annual Reviews in Control*, 29(1):71–85, 2005.

[14] K. Kalsi, S. Hui, and S.H. Zak. Unknown input and sensor fault estimation using sliding-mode observers. In *American Control Conference (ACC), 2011*, pages 1364–1369, June 2011.

[15] K.J. Åström. *Introduction to stochastic control theory*. Mathematics in Science and Engineering. Academic Press, 1970.

[16] J.L. Crassidis and J.L. Junkins. *Optimal Estimation of Dynamic Systems (Chapman & Hall/CRC Applied Mathematics & Nonlinear Science)*. Chapman and Hall/CRC, 1 edition, April 2004.

[17] F.L. Lewis, L. Xie, and D. Popa. *Optimal and robust estimation : with an introduction to stochastic control theory*. Automation and control engineering. CRC Press, Boca Raton, 2008.

[18] T. Kailath, A.H. Sayed, and B. Hassibi. *Linear estimation*. Prentice-Hall information and system sciences series. Prentice Hall, 2000.

[19] M. Hou and R.J. Patton. Input observability and input reconstruction. *Automatica*, 34(6):789–794, 1998.

[20] M. Darouach and M. Zasadzinski. Unbiased minimum variance estimation for systems with unknown exogenous inputs. *Automatica*, 33(4):717–719, 1997.

[21] Y. Cheng, H. Ye, Y. Wang, and D. Zhou. Unbiased minimum-variance state estimation for linear systems with unknown input. *Automatica*, 45(2):485–491, 2009.

[22] M.L.J. Hautus. Strong detectability and observers. *Linear Algebra and its Applications*, 50(0):353–368, 1983.

[23] D. Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 1 edition, August 2006.

[24] A.E. Bryson. *Applied optimal control: examples and algorithms*. Cambridge University Press, 2002.