On the Connection Between Different Noise Structures for LPV-SS Models

Pepijn B. Cox and Roland Tóth

Abstract—Different representations to describe noise processes and finding connections or equivalence between them have been part of active research for decades, in particular for linear time-invariant case. In this paper the linear parameter-varying (LPV) setting is addressed; starting with the connection between an LPV state-space (SS) representation with a general noise structure and the LPV-SS model in an innovation structure, i.e., the Kalman filter. More specifically, the considered LPV-SS representation with general noise structure has static, affine dependence on the scheduling signal; however, we show that its companion innovation structure has a dynamic, rational dependency structure. Following, we would like to highlight the assumption that the system is quadratic stability. Contrary to the LTI case, having quadratic stability for LPV or jump-Markov linear systems is only a sufficient condition, i.e., restrictive condition for stability (e.g., see [3]), and, therefore, the result of [2] does not apply to every stable LPV or jump-Markov linear system. However, in this paper such strong assumption on stability of the system is not required. Hence, we will treat a more general case.

In this paper, we start by providing the connection between the LPV-SS representation with general noise and its LPV-SS innovation structure companion (Sec. II). When moving from the general noise structure with affine, static dependency on the scheduling signal to the innovation form, the resulting error state covariance matrix function, Kalman gain, and the covariance of the innovation noise will have dynamic, rational dependency on the scheduling signal. Then, based upon the stability result of [4], we show that the innovation recursion can recover from an incorrect initialization or disturbance with a guaranteed asymptotic convergence (Sec. III). This guaranteed convergence implies a “fading memory” effect within the innovation recursion and it is used to argue that the Kalman gain can be approximated by only using a partial trajectory of the scheduling signal, in stead of the complete trajectory (Sec. IV).

I. INTRODUCTION

Including general representations of noise processes in system identification is essential for capturing a wide variety of possible noise sources experienced in practice, e.g., unmodelled dynamics, sensor noise, parameter inaccuracies, etc. Hence, active research on different representations, their generality, and connections between them has been going on for decades. Especially, the linear time invariant (LTI) case has a well established connection between an LTI state-space (SS) representation with state and output additive noise and the innovation form, i.e., the Kalman filter (e.g., see [1] and the references therein). In this case, the Kalman filter is asymptotically time invariant, therefore, a suboptimal filter can be found with a constant Kalman matrix.

To the authors knowledge, similar time invariant Kalman filters for linear parameter-varying (LPV), time-varying, or nonlinear systems does not exists. Except [2], for a stochastic jump-Markov linear system a Kalman gain that has affine dependency on the switching signal can be found, under the assumption that the system is quadratic stability. Contrary to the LTI case, having quadratic stability for LPV or jump-Markov linear systems is only a sufficient condition, i.e., restrictive condition for stability (e.g., see [3]), and, therefore, the result of [2] does not apply to every stable LPV or jump-Markov linear system. However, in this paper such strong assumption on stability of the system is not required. Hence, we will treat a more general case.

II. PRELIMINARIES

Notation

We denote a probability space as $(\Xi, \mathcal{F}_\Xi, \mathbf{P})$ where $\mathcal{F}_\Xi$ is the $\sigma$-algebra, defined over the sample space $\Xi$; and $\mathbf{P} : \mathcal{F}_\Xi \to [0, 1]$ is the probability measure defined over the measurable space $(\Xi, \mathcal{F}_\Xi)$. Within this work, we consider random variables that take values on the Euclidean space. More precisely, for the given probability space $(\Xi, \mathcal{F}_\Xi, \mathbf{P})$ we define a random variable $\mathbf{f}$ as a measurable function $\mathbf{f} : \Xi \to \mathbb{R}^n$, which induces a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. As such, a realization $\nu \in \Xi$ of $\mathbf{P}$, denoted $\nu \sim \mathbf{P}$, defines a realization $f(\nu)$ of $\mathbf{f}$, i.e., $f := f(\nu)$. Furthermore, a stochastic process $\mathbf{x}$ is a collection of random variables $\mathbf{x}_t : \Xi \to \mathbb{R}^n$ indexed by the set $t \in \mathbb{Z}$ (discrete time), given as $\mathbf{x} = \{\mathbf{x}_t : t \in \mathbb{Z}\}$. A realization $\nu \in \Xi$ of the stochastic process defines a signal trajectory $x := \{\mathbf{x}_t(\nu) : t \in \mathbb{Z}\}$. We call a stochastic process $\mathbf{x}$ stationary if $\mathbf{x}_t$ has the same probability distribution on each time index as $\mathbf{x}_{t+\tau}$ for all $\tau \in \mathbb{N}$.

In addition, the inequalities $A \succeq B$ and $A \succ B$, for two symmetric matrices $A$ and $B$ of equal dimension, imply that $A - B$ is semi-positive and positive definite, respectively.
\[ x_{t+1} = A(p_t)x_t + B(p_t)u_t + G(p_t)w_t, \]  
\[ y_t = C(p_t)x_t + D(p_t)u_t + H(p_t)v_t, \]  
where \( x : \mathbb{R}^n \rightarrow \mathbb{R}^n \) is the state variable, \( y : \mathbb{R}^n \rightarrow \mathbb{R}^p \) is the measured output signal, \( u : \mathbb{R} \rightarrow \mathbb{R}^m \) is the input signal, \( p : \mathbb{R}^n \rightarrow \mathbb{R} \) is the scheduling variable, \( x_0 \in \mathbb{R}^n \) is the initial state, \( w : \mathbb{R}^n \rightarrow \mathbb{R}^n \) is the discrete time, \( v : \mathbb{R}^n \rightarrow \mathbb{R}^p \) is the random variable of the stochastic process \( v \), respectively, \( Q \in \mathbb{R}^{n \times n} \), \( S \in \mathbb{R}^{p\times p} \), and \( R \in \mathbb{R}^{m \times m} \) are covariance matrices, such that \( \Sigma \) is positive definite. Furthermore, we will assume \( u, p, w, v, y \) to have left compact support to avoid technicalities with initial conditions. The matrix functions \( A(\cdot), \ldots, H(\cdot) \), defining the SS representation \((1)\) are defined as affine combinations:

\[ A(p_t) = A_0 + \sum_{i=1}^{n_0} A_i \psi^{[i]}(p_t), \quad B(p_t) = B_0 + \sum_{i=1}^{n_0} B_i \psi^{[i]}(p_t), \quad C(p_t) = C_0 + \sum_{i=1}^{n_0} C_i \psi^{[i]}(p_t), \]
\[ D(p_t) = D_0 + \sum_{i=1}^{n_0} D_i \psi^{[i]}(p_t), \quad G(p_t) = G_0 + \sum_{i=1}^{n_0} G_i \psi^{[i]}(p_t), \]

where \( \psi^{[i]}(\cdot) : \mathbb{R} \rightarrow \mathbb{R} \) is bounded scalar functions on \( \mathbb{R} \) and \( \{A_i, B_i, C_i, D_i, G_i\}_{i=0}^{n_0} \) are constant matrices with appropriate dimensions. Additionally, for well-posedness, it is assumed that \( \{\psi^{[i]}(\cdot)\}_{i=1}^{n_0} \) are linearly independent over an appropriate function space and are normalized w.r.t. an appropriate norm or inner product \([5]\).

### B. The innovation form

To start, under some mild conditions, the LPV-SS representation \((1)\) has the following equivalent innovation form:

**Lemma 1.** For each given trajectory of the input \( u \) and scheduling \( p \), the LPV data-generating system \((1)\) can be equivalently represented by a \( p \)-dependent innovation form:

\[ \dot{x}_{t+1} = A_t(x_t) + B_t(p_t)u_t + K_t \xi_t, \]  
\[ y_t = C_t(x_t) + D_t(p_t)u_t + H_t(v_t), \]

where \( \xi_t \sim \mathcal{N}(0, \Lambda_t) \) and \( K_t \) can be uniquely determined by:

\[ K_t = [A_t(p_t)P_{t|t-1}C_t(p_t) + G_t(p_t)SRH_t(p_t)]^{-1}, \]
\[ P_{t|t-1} = A_t(p_t)P_{t-1|t-1}A_t^T(p_t) - K_t(p_t)K_t^T + G_t(p_t)QG_t(p_t), \]
\[ \Omega_t = C_t(p_t)P_{t|t-1}C_t(p_t) + H_t(p_t)RH_t(p_t), \]

Under the assumption that \( \exists t_0 \in \mathbb{Z} \) such that \( x_{t_0} = 0 \) and \( \Gamma_t \) is non-singular for all \( t \in [t_0, \infty] \). In \((4a)-(4e)\), the notation of \( \mathcal{K}_t, \mathcal{P}_{t|t-1} \), and \( \Omega_t \) is a shorthand for \( \mathcal{K}_t := (K \circ p_t) \in \mathbb{R}^{n_x \times n_x}, \mathcal{P}_{t|t-1} := (P_{t|t-1} \circ p_t) \in \mathbb{R}^{n_x \times n_x}, \) and \( \Omega_t := (\Omega \circ p_t) \in \mathbb{R}^{n_x \times n_x}. \) The operator \( \circ : (\mathbb{R}, \mathbb{P}) \rightarrow \mathbb{R} \) denotes \( (K_t \circ p_t) = K_t(p_t + \tau_1, \ldots, p_t + \tau_2) \) with \( \tau_1, \tau_2 \in \mathbb{Z}. \) The subscript notation \( t_{\{t\}} \) denotes the matrix function at time \( t + 1 \) depends only on \( p_i \) for \( i = t_0, \ldots, t. \)

**Proof.** See Appendix.

From Lem.\( 1 \) it becomes clear that moving from the LPV-SS system \((1)\) with static, affine dependency to the innovation form comes at the cost of dynamic, rational dependency on the scheduling signal. In Sec.\( III \) guaranteed asymptotic convergence of the covariance matrix \( \mathcal{P}_{t|t-1} \) is proven if the covariance matrix is perturb by an error in the past. That result is used to argue how the Kalman gain \( K_t \) can be approximated by a partial trajectory of the scheduling signal, in Sec.\( IV \).

### III. Guaranteed Asymptotic Convergence of the Covariance Matrix \( \mathcal{P}_{t|t-1} \)

In this section, we will show that an error created on the priori error covariance matrix \( \mathcal{P}_{t|t-1} \) at a certain time will asymptotically decrease to zero when time progresses. To this end, let us introduce some technicalities. Firstly, the stochastic processes \( w \) and \( v \) need to be uncorrelated, hence using the minimum variance estimate of \( w_t \) given by \( \bar{w}_t = w_t - SR^{-1}v_t \), the state equation \((1a)\) is rewritten as (e.g., see \( II \) Section 5.5)

\[ x_{t+1} = (A(p_t) - G(p_t)SR^{-1}H^{-1}(p_t)C(p_t))x_t + (B(p_t) - G(p_t)SR^{-1}H^{-1}(p_t)D(p_t))u_t + (G(p_t)SR^{-1}H^{-1}(p_t)y_t + G(p_t)\tilde{w}_t, \]

where \[ \begin{bmatrix} \bar{w}_t \\ v_t \end{bmatrix} \sim \mathcal{N} \begin{bmatrix} 0 \\ 0 \\ Q - SR^{-1}S \\ 0 \\ R \end{bmatrix}. \]

Define \( \bar{u}_t := [u_t^T \quad y_t^T]^T \), which gives the following scheduling dependent matrices:

\[ \bar{A}_t = A_t - G_tSR^{-1}H^{-1}(p_t)C_t, \]
\[ \bar{B}_t = B_t - G_tSR^{-1}H^{-1}(p_t)D_t, \]
\[ \bar{Q} = Q - SR^{-1}S. \]

Secondly, let \( \bar{B}_t \) and \( \bar{D}_t \) be bounded and assume that:

\[ \alpha_1 I \geq \bar{G}(p_t)\bar{Q}\bar{G}(p_t) \geq \alpha_2 I, \]
\[ \beta_1 I \leq \bar{C}(p_t)^T(\bar{H}(p_t)\bar{R}\bar{H}(p_t)^T)^{-1}\bar{C}(p_t) \leq \beta_2 I, \]
\[ \delta_1 I \geq \bar{A}_t^T\bar{A}_t \geq \delta_2 I, \]

holding for \( i \in \mathbb{T} \), with left compact support \( \mathbb{T} \) of the scheduling signal, \( \alpha_1, \beta_1, \beta_2, \delta_1 > 0, \) and \( \alpha_2, \beta_1, \beta_2, \delta_2 < \infty. \) Conditions \((5)\)-(\( 5h \)) imply that the system \((1)\) is stochastically controllable and observable for all possible variations of \( p \in \mathbb{P} \), i.e., the state
can uniquely be reconstructed, which are not over restrictive assumptions. Define the posterior filter error $e_t = x_t - \hat{x}_{t|t}$ dynamics by
\[ e_{t+1} = (I - K_{t+1}C(p_{t+1}))\hat{A} e_t = W_{t+1}. \tag{5i} \]

If (5f)-(5h) hold then there exist a $\beta_3, \beta_4, \beta_6$ as
\[ \left[ \begin{array}{c} \hat{C}(p_i) \ 0 \\ C(p_{i-1}) \end{array} \right] \left[ \begin{array}{c} \hat{A}_{i-1} \ \hat{A}_{i-2} \\ A_{i-1} \end{array} \right] e_{i-2} \geq \beta_3 \|e_{i-2}\|, \tag{5j} \]
\[ \left[ \begin{array}{c} \hat{R} \\ 0 \end{array} \right] + L_i \left[ \begin{array}{c} P_{t|i-1} \ 0 \\ 0 \ P_{t-1|i-2} \end{array} \right] L_i^\top \geq \beta_5 I, \tag{5k} \]
with
\[ L_i = \left[ \begin{array}{c} C(p_i) \\ 0 \end{array} \right] \left[ \begin{array}{c} \hat{C}(p_i) \\ C_{i-1} \end{array} \right], \]
and
\[ \|W^{-1}_{t-2}W^{-1}_{t-1}e_t\|_2 \geq \beta_6 \|e_t\|_2, \tag{5l} \]
for $i = \mathbb{T}$ where $0 < \beta_3, \beta_5, \beta_6 < \infty$. Using (5j)-(5l), the error created on the priori error covariance matrix $P_{t+1|i}$ at a certain time will asymptotically to zero by:

**Theorem 1.** Let $\hat{B}_i$ and $D(p_i)$ be bounded and (5f)-(5h) hold. Assume that $P_{t|i-1}$ is a positive semi definite matrix function of $t$, which has a left compact support $\mathbb{T}$ and satisfies (4d)-(4e) for a given trajectory of $p$ (with left compact support). Let us consider that there exists a $\tau > 0$ and an associated $\hat{P}_{t|i-1}^{(\tau)}$ constructed with the recursions of (4d) initialized at time $t-\tau$ with
\[ \hat{P}_{t-\tau|i-1} = \hat{A}_i \hat{P}_{0,\tau} \hat{A}_i^\top + \hat{Q}_i \]
where $\hat{P}_{0,\tau}$ is a static matrix defined as
\[ 0 < \hat{P}_{0,\tau} \prec \left( \frac{\alpha_2}{\alpha_2 \beta_2 + 1} \right) I. \tag{6b} \]
Then, the difference between $P_{t|i-1}$ and $\hat{P}_{t|i-1}^{(\tau)}$ has the following bound
\[ \max_{t \in \mathbb{T}} \|P_{t|i-1} - \hat{P}_{t|i-1}^{(\tau)}\|_2 \leq \xi \frac{\delta_1 n_x (\alpha_1 \beta_1 + 2) (\alpha_2 \beta_2 + 1)}{\beta_5 (\alpha_2 \beta_2 + 1)} \]
where $\xi \in (0, 1)$ and given as
\[ \xi = \frac{\beta_5 (\alpha_2 \beta_2 + 1)}{\beta_5 (\alpha_2 \beta_2 + 1) + \alpha_2 \beta_3 \beta_6} \]

The remainder of the section is used to prove Theorem 1. The following proof uses extensively the result of (4), however, Dyest and Price make use of the posterior covariance matrix $P_{t|i}$ in stead of the prior covariance matrix $P_{t|i-1}$. Hence, we will first construct the proof w.r.t. posterior covariance matrix. The covariance matrix is given as
\[ P_{t|i} = (I - K_i C(p_i)) P_{t|i-1}, \tag{7} \]
where it is proven in (4h) to be bounded as
\[ \left( \frac{\alpha_2}{\alpha_2 \beta_2 + 1} \right) I \leq P_{t|i} \leq \left( \frac{1}{\beta_1} + \alpha_1 \right) I, \tag{8} \]
and the prior covariance can be found from the posterior as
\[ P_{t+1|i} = \hat{A}_i P_{t|i} \hat{A}_i^\top + \hat{Q}_i. \tag{9} \]

Remark that $\hat{P}_{0,\tau}$ of (6a) substitutes the posterior covariance matrix on $P_{t-\tau|i-\tau}$ to construct $\hat{P}_{t|i}^{(\tau)}$.

**Lemma 2.** Given (6b) and the error dynamics (5i), it holds that
\[ 0 \leq \left( P_{t|i} - \hat{P}_{t|i}^{(\tau)} \right) \leq \Pi_{t} \hat{W}_{t} \left( P_{t-\tau|i-\tau} - \hat{P}_{0,\tau} \right) \Pi_{t-1} \hat{W}_{t-1}^\top, \tag{10} \]
where $\hat{W}_{t}$ is the filter error dynamics w.r.t. $\hat{K}_{t+1}$ of $\hat{P}_{t|i}^{(\tau)},$.

**Proof.** Let first proof the lemma for $\tau = 1$. As $P_{t|i}$ is the optimal solution for $K_i$, then $P_{t|i}$ is constructed from $K_i$ for $i = t_0, \ldots, t-\tau - 1$ and $K_j^{(\tau)}$ for $j = t-\tau, \ldots, t$ with $K_j \neq K_j^{(\tau)}$. Remark that $P_{t|i}$ is suboptimal; hence, $P_{t|i} \geq P_{t|i}$ [1] Theorem 2.1]. Therefore,
\[ P_{t|i} - \hat{P}_{t|i}^{(\tau)} \geq P_{t|i} - \hat{P}_{t|i}^{(\tau)} \tag{11a} \]
Also see that (7) can be written by using (6d), as
\[ P_{t|i} = (I - K_i C(p_i)) (\hat{A}_i - \hat{A}_t) P_{t-1|i-1} \hat{A}_t^\top + \hat{G}(p_{t-1}) \hat{Q}(p_{t-1}) \]
\[ (I - K_i C(p_i)) \hat{A}_i^\top + K_i \mathcal{K}_{i-1} \mathcal{K}_{i-1}^\top \]
Combining (11a) and (11b) results in
\[ \hat{W}_{t-1} \left( P_{t-1|i-1} - \hat{P}_{0,\tau} \right) \hat{W}_{t-1}^\top \geq \left( P_{t|i} - \hat{P}_{t|i}^{(\tau)} \right) \tag{11c} \]
Then repeating the upper bound (11c) for $\tau$ time steps proves the upper bound, i.e., right-hand side of (10). Similar argument can be made for the lower bound. Now, initialize with $\hat{P}_{0,\tau}$ and use $K_j$ for $j = t-\tau, \ldots, t$ to construct $P_{t|i}$, i.e., find an suboptimal solution with $P_{t|i} \geq \hat{P}_{t|i}^{(\tau)}$. Hence $P_{t|i} - \hat{P}_{t|i}^{(\tau)} \geq P_{t|i} - \hat{P}_{t|i}^{(\tau)}$, which results, for $\tau = 1$, in the following lower bound
\[ \hat{W}_{t-1} \left( P_{t-1|i-1} - \hat{P}_{0,\tau} \right) \hat{W}_{t-1}^\top \geq \left( P_{t|i} - \hat{P}_{t|i}^{(\tau)} \right) \tag{11d} \]
As $P_{t-1|i-1} - \hat{P}_{0,\tau}$ is semi-positive definite (by construction of $\hat{P}_{0,\tau}$), the left-hand side of (11d) is bounded by zero.

Next, let us provide the sufficient conditions for quadratic Lyapunov stability of the filter dynamics (5i) proven in (4):

**Lemma 3.** If the LPV-SS system (1) satisfies conditions (5f)- (5h) then the system (5i) is asymptotically stable. Additionally, there exists a real scalar function $V(e_t, t)$ such that
\[ 0 < \gamma_1 \|e_t\|^2 \leq V(e_t, t) \leq \gamma_2 \|e_t\|^2, e_t \neq 0, \tag{12a} \]
\[ V(e_t, t) - V(e_{t-1}, t-1) \leq \gamma_3 \|e_t\|^2 < 0, e_t \neq 0, \tag{12b} \]
where
\[ \gamma_1 = \frac{\beta_1}{1 + \alpha_1 \beta_1}, \quad \gamma_2 = \frac{1}{\alpha_2} + \beta_2, \quad \gamma_3 = -\beta_3 \beta_5^{-1} \beta_6. \tag{12c} \]

Using Lem. 3 the bound on the error-dynamics is:
Lemma 4. If the LPV-SS system (1) satisfies conditions 53-5b then the τ-step homogeneous error dynamics (5) are bounded as

$$\|e_t\|^2 = \left\| \prod_{t'=t-\tau}^{t-1} W_{t'\tau-t} \right\|^2_2 \leq \xi^\tau \gamma_1^{-1} \gamma_2 \|e_{t-\tau}\|^2_2,$$

where $\xi = \frac{\gamma_2}{\gamma_2 - \gamma_3}$.

**Proof.** To simplify notation, define $V_t := V(e_t, t)$. Substituting the upper bound of (12a) into (12b) gives

$$V_t - V_{t-1} \leq \gamma_3 \|e_t\|^2_2 \leq \gamma_3 \gamma_2^{-1} V_t < 0.$$  (14a)

Therefore the following holds

$$V_t - \xi V_{t-1} \leq 0, \quad \xi = \frac{1}{1 - \gamma_3 \gamma_2^{-1}}.$$  (14b)

where $\xi \in [0, 1]$ as $\gamma_3 \gamma_2^{-1} < 0$. Hence, $V_t - \xi^\tau V_{t-\tau} \leq 0$. Combining this τ-step Lyapunov bound and (12a) gives

$$\gamma_1 \|e_t\|^2_2 \leq V_t \leq \xi \gamma_2 \|e_{t-\tau}\|^2_2.$$  (14c)

The proof is completed by applying the $\ell_2$ norm on (5i) and substituting (14c).

To complete the proof of Thm.1 First, take the eigenvalue decomposition $Z = UDU^T$, where $U$ is a matrix containing the eigenvectors and $D$ the diagonal matrix containing the eigenvalues $\lambda_i \geq 0$ of $Z$. For a positive definite matrix $Z$, it holds that

$$Z = UDU^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T.$$  (15)

Second, take the eigenvalue decomposition of $P_{t|t-\tau} = \sum_{j=1}^{n} \lambda_j \xi_j u_j u_j^T$ and define $y_{j,t} := \lambda_j u_j$. Then, combining the trace of the left-hand side of (10) with (13) gives

$$\sum_{j=1}^{n} \sum_{j'=1}^{n} \xi_j \gamma_1^{-1} \gamma_2 \|y_{j,t}\|^2_2 \leq \xi^\tau \gamma_1^{-1} \gamma_2 \text{Tr}[P_{t|t-\tau} - P_{0|\tau}] T.$$  (16)

Joining (15) and (16) results in

$$P_{t|t} - \hat{P}_{t|t}^{(\tau)} \leq \xi^\tau \gamma_1^{-1} \gamma_2 \text{Tr}[P_{t|t-\tau} - P_{0|\tau}] T.$$  (17)

Left and right multiplying by $\hat{A}_t$ and $\hat{A}_t^T$, respectively, and substituting (9) gives

$$\hat{A}_t \left(P_{t|t} - \hat{P}_{t|t}^{(\tau)}\right) \hat{A}_t^T = \hat{P}_{t+1|t} - \hat{P}_{t+1|t}^{(\tau)} \leq \xi^\tau \gamma_1^{-1} \gamma_2 \text{Tr}[P_{t|t-\tau} - P_{0|\tau}] \hat{A}_t \hat{A}_t^T \leq \xi^\tau \gamma_1^{-1} \gamma_2 \delta_1 \text{Tr}[P_{t|t-\tau} - P_{0|\tau}] \hat{A}_t \hat{A}_t^T.$$  (17)

Taking into account (8) and using (6), the following holds

$$\text{Tr}[P_{t|t-\tau} - P_{0|\tau}] < n_x (\beta_1^{-1} + \alpha_1).$$  (18)

To conclude the proof, the spectral norm of a matrix $A$ is $\|A\|_2 = \sigma_{\text{max}}(A)$. Hence, applying the spectral norm on (17) and substitute (18) results in

$$\|P_{t|t-\tau} - \hat{P}_{t|t-\tau}^{(\tau)}\|_2 \leq \xi^\tau \gamma_1^{-1} \gamma_2 \delta_1 \|A\|_2 (\beta_1^{-1} + \alpha_1),$$  (19)

which is equivalent to (6c) when substituting (12a) and taking into account that the bound is time independent, i.e., it should hold for every $t \in \mathbb{T}$, which concludes the proof of Thm.1.

**IV. APPROXIMATION OF KALMAN GAIN**

The innovation form is a different view on constructing a the Kalman filter for (1). Hence, $\hat{K}_t$ in (4d) can be viewed as the optimal LPV Kalman gain of (1). In the LTI case, the Kalman filter is asymptotically time invariant, therefore, a suboptimal filter can be found with a constant $P$ and $K$ matrix (11). Hence, in the LTI case, the innovation form with constant $K$ and $P$ matrix is viewed as a model description which allows a general noise model. However, for the LPV case, Lem.1 indicates that even if $A(i), \ldots, D(i)$ have, for example, affine dependence on $p_t$ (each $\psi[i](p_t) = p_t^{(i)}$) then $K_t, P_{t|t-\tau}, \Omega_t$ are meromorphic functions, where the nominator and denominator are polynomial functions in the scheduling signal $p$ and its past time-shifts. Hence, the filter, generally speaking, it is not clear that $\hat{K}_t$ will converge to a steady state solution with some constant $K$ matrix, and, therefore, $\hat{K}_t$ is a function of scheduling signal and its past, i.e., $p_t$ with $t \in \mathbb{T}$.

However, a popular model for many subspace identification schemes is the innovation form, e.g., see (6). In the LTI case, the connection between the innovation form and the LTI counterpart of (1), e.g., $A(p) = A$, is well studied. However, it has not been thoroughly investigated in the LPV case. As Lem.1 shows, the LPV-SS representation with general noise (1) is not equivalent to the innovation form with only static, affine matrix functions, commonly used (7), (8). Hence, in this section, we are providing two approximations: i) due to the asymptotic convergence of the innovation filter (Thm.1), the Kalman gain $K_t$ can be approximated by $K^{(\tau)}_t$, which depends only on $p_{t-\tau}, \ldots, p_t$; and ii) in some cases, by sacrificing state minimally, the approximate Kalman gain with dynamic, rational dependence on the scheduling signal can be transformed to an approximate Kalman gain with static, affine dependence (Sec. V.A).

To start with the first approximation, thm.1 highlights that the covariance matrix $P_{t|t-\tau}$ can be arbitrary well approximated by only taking the scheduling signal $p$ from $p_{t-\tau}, \ldots, p_t$ into account, e.g., “fading memory” of the innovation recursions. The approximation error is upper bounded, as given in (6c), and decays to zero if $\tau \to \infty$. Furthermore, the covariance matrix (4d) is not implicitly dependent on the Kalman gain (4c); however, any approximation of $P$ will lead to an approximation of $K$. As $K$ is a rational function, any approximation of $P$ will result in a unique relation in $K$ (up to co-primness of the nominator and denominator).

**Conjecture 1.** Let us consider that there exists a $\tau > 0$, $\hat{P}_{t|t-\tau}^{(\tau)}$ as constructed in Thm.1 and let the associated gain $K^{(\tau)}_t$ be given by (4d) using $\hat{P}_{t|t-\tau}^{(\tau)}$. Then the Kalman gain can be decomposed as

$$K_t = K^{(\tau)}_t + R^{(\tau)}_t,$$  (20a)

where $R^{(\tau)}_t$ is a rational matrix function in $p_t, p_{t-1}, \ldots$. In addition, if $\tau \to \infty$ then $R^{(\tau)}_t \to 0$ and

$$\|R^{(\tau)}_t\|_2 < \|R^{(\tau+1)}_t\|_2.$$  (20b)

where $R^{(\tau+1)}_t$ is the remainder term w.r.t. $K^{(\tau+1)}_t$ and $K^{(\tau+1)}_t$ is constructed by using $\hat{P}_{t|t-\tau}^{(\tau+1)}$.  \hfill \Box
Conj. [1] highlights that $K_t$ can be approximated by $K_t^{(r)}$, which depends only on $p_{t-\tau},\ldots,p_t$. This truncation can be made arbitrarily accurate by choosing an appropriate $\tau$, i.e., $\|R_t^{(\tau+1)}\|_2 \ll \|K_t\|_2$.

A. Static, affine Kalman gain

A popular choice in LPV-SS identification is to identify an LPV-SS innovation form with a static and affine Kalman filter matrix (e.g., see [5], [9]), similarly parametrized as (3). Under the assumption that the Kalman filter function can be arbitrarily well approximated by $K_t^{(r)}$, the dynamic, rational dependence on the scheduling signal may, in some cases, be transformed into a static, affine LPV-SS representation by adding states, i.e., increasing $n_x$.

Example 1. Consider the following LPV representation

\[ y_t = -p_t y_{t-1} + p_t u_{t-1} + e_t + e_{t-1}. \] \hspace{1cm} (21a)

The state minimal LPV-SS realization of (21a) is

\[ x_{t+1} = -p_t x_t + u_t + \frac{1 - p_{t+1}}{p_{t+1}} e_t, \hspace{1cm} y_t = p_t x_t + e_t, \] \hspace{1cm} (21b)

which is rationally and dynamically dependent on the scheduling parameter $p$. Note, it can be shown that there exists no state transformation (not even $p$-dependent) which can turn (21b) into (4a)-(4b) with static, affine dependence $K$ on $p$, i.e., $K_t = K(p_t)$, and keep the minimal state dimension $n_x = 1$ ([10], Def. 3.29). However, the transformation into a static, affine form can be done by introducing an additional state as

\[ \tilde{x}_{t+1} = \begin{bmatrix} -p_t & 1 \\ 0 & 0 \end{bmatrix} \tilde{x}_t + \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} u_t + e_t, \] \hspace{1cm} (21c)

\[ y_t = \begin{bmatrix} -p_t & 1 \end{bmatrix} \tilde{x}_t + e_t. \]

Ex. [1] shows the elimination of dynamic, rational dependence by sacrificing state minimality. Hence, as many LPV-SS identification methods estimate a static, affine functional relation on the scheduling signal [8], [9], the rank relieving property of subspace methods is lost, as additional states are added to preserve the static, affine dependency. As a conclusion, in the LPV case, the Kalman gain $K_t$ should have rational and dynamic dependence on the scheduling signal to enjoy general noise modelling capabilities (Lem. [1]) and minimality of the state dimension. However, in practice, we need to restrict overparameterization to reduce complexity of the estimation method and variance of the model estimates. Hence, the above given analysis is important to understand the trade-off behind these choices.

V. CONCLUSION

We have shown that the innovation form [1] should have a Kalman gain with rational and dynamic dependence on the scheduling signal to represent general noise. However, this function can be approximated by truncating the dynamic dependence. Using this truncation, for some cases, an equivalent LPV-SS representation with affine and static dependency on the scheduling signal can be found by including additional states, resulting in a non-state minimal system.

APPENDIX

INNOVATION REPRESENTATION

The idea of the innovation process $\xi_t$ is such that $\xi_t$ consists of that part of $y_t$ not carried in $y_{t-1}, y_{t-2}, \ldots$, i.e.,

\[ \xi_t = y_t - E^* \{ y_t \mid Y_{t-1} \}, \] \hspace{1cm} (22)

where $E^* \{ \cdot \}$ is the minimum variance estimator and $Y_{t-1}$ indicates the set of observations $\{y_{t-1}, \ldots, y_0\}$. The signal $y_t$ generated by (11) is a sequence of Gaussian random variables as $u_t$ is known exactly. Hence, the output signal $y$ is split into a ‘deterministic’ part of $y_t$ as $\tilde{y}_t = E^* \{ y_t \mid Y_{t-1} \}$ and a white noise $\xi_t$ with Gaussian distribution. The variables $\tilde{y}_t$ and $\xi_t$ are uncorrelated, i.e., $E \{ \tilde{y}_t \xi_t^* \} = 0$ for $i = 0, \ldots, t$ because of the orthogonality property of the minimum variance estimator. Without loss of generality, we assume that $\xi_0 = y_0 - E^* \{ y_0 \}$. Hence, as the initial condition is known, there exists a causal filter of $y_0, \ldots, y_t$ to $\tilde{y}_t$ by writing out (11). The other way around, i.e., that $y_t$ depends on $\xi_0, \ldots, \xi_t$, can be shown in a recursive way (11). Therefore, the dataset $y_0, \ldots, y_t$ and $\xi_0, \ldots, \xi_t$ are uniquely related to each other and the following holds

\[ E \{ y_t \mid y_0, \ldots, y_{t-1} \} = E \{ y_t \mid \xi_0, \ldots, \xi_{t-1} \}. \] \hspace{1cm} (23)

In addition, for any variable $x_t$ which has a joint Gaussian distribution with $y_t$ it holds that

\[ x_t = E \{ x_t \mid y_0, \ldots, y_{t-1} \} = E \{ x_t \mid \xi_0, \ldots, \xi_{t-1} \}. \] \hspace{1cm} (24)

Substituting (23) and (24) into (22) and taking the output equation relation (11) into account, gives

\[ y_t = C(p_t) \tilde{x}_t + D(p_t) u_t + \xi_t. \] \hspace{1cm} (25)

We will assume that the initial state $x_0 = 0$ is known.

As $\xi_0, \ldots, \xi_{t+1}$ are mutually uncorrelated, the conditional expectation (24) can be split up, e.g., see [1] Theorem 2.4, Ch. 5], and combined with (2a), which gives

\[ \tilde{x}_{t+1} = E \{ x_{t+1} \mid \xi_0, \ldots, \xi_{t-1} \} + E \{ x_{t+1} \mid \xi_t \} - E \{ x_{t+1} \}, \]

\[ = A(p_t) \tilde{x}_t + B(p_t) u_t + E \{ x_{t+1} \mid \xi_t \} - E \{ x_{t+1} \}. \] \hspace{1cm} (26)

Note that $x_{t+1}$ is uncorrelated with $\xi_{t-i}$ for $i > 0$. The state $x_{t+1}$ and $\xi_t$ are jointly Gaussian distributed variables, hence

\[ E \{ x_{t+1} \mid \xi_t \} = E \{ x_{t+1} \} + \text{cov} \{ x_{t+1}, \xi_t \} \text{var} \{ \xi_t \}^{-1} \{ \xi_t - E \{ \xi_t \} \}. \] \hspace{1cm} (27)

1Comparable phenomena can be observed in the LTI case. If it is assumed that $S = 0$, however, for the underlying system $S \neq 0$, then an increase of the state dimension is also evident [1].

2This proof can be extended to $x_0 \in N(0, P_0)$. However, it involves additional constraints to ensure that the noise sequences $w$ and $v$ can be causally computed from $y$, see [1] Theorem 3.4, Ch. 9]. For simplicity, this case will not be considered.
by using the minimum variance estimator property, e.g., see [1] Theorem 2.1, Ch. 5]. Define the error of the state estimate by 
\( \hat{x}_t = x_t - \hat{x}_t \). To compute \( \text{cov} [x_{t+1}, \xi_t] \), see:

\[
\text{cov} [x_{t+1}, \xi_t] = \text{cov} [x_{t+1}, C(p_t) \hat{x}_t + H(p_t)v_t] \\
= \mathbb{E} \left[ (A(p_t)(x_t - \mathbb{E} \{x_t\}) + G(p_t)w_t)[C(p_t)\hat{x}_t + H(p_t)v_t]^\top \right] \\
= A(p_t)P_{t|t-1} C(p_t) + G(p_t)SH(p_t),
\]

where \( P_{t|t-1} = \text{var} [\hat{x}_t] \) is the a priori state error covariance. To compute \( \text{var} [\xi_t] \), note that \( (16) \) and \( (25) \) are equal in \( y_t \), hence, by using the transitive property of equality, the variance of \( \xi_t \) is given as

\[
\Omega_t = \text{var} [\xi_t] = \text{var} [C(p_t)\hat{x}_t + H(p_t)v_t] = C(p_t)P_{t|t-1} C(p_t) + H(p_t)\Sigma H(p_t)^\top.
\]

Substituting \( (27) \), \( (28) \), and \( (29) \) in \( (26) \) gives

\[
\hat{x}_{t+1} = A(p_t)\hat{x}_t + B(p_t)u_t + K_t \xi_t, \quad (30a) \\
K_t = [A(p_t)P_{t|t-1} C(p_t) + G(p_t)SH(p_t)]\Omega_t^{-1}. \quad (30b)
\]

Finally, the a priori state error covariance \( P_{t|t-1} \) should be found. Subtracting \( (30a) \) from \( (13) \) gives

\[
\tilde{x}_{t+1} = A(p_t)\tilde{x}_t + G(p_t)w_t - K_t \xi_t \\
= [A(p_t) - K_tC(p_t)] \tilde{x}_t + G(p_t)w_t - K_tH(p_t)v_t. \quad (31)
\]

Then

\[
P_{t+1|t} = \left[ A(p_t) - K_tC(p_t) \right] P_{t|t-1} \left[ A(p_t)^\top - C(p_t)K_t^\top \right] \\
+ G(p_t)QG(p_t)^\top + K_t\Sigma H(p_t)K_t \Omega_t \\
- G(p_t)SH(p_t)K_t^\top - K_t\Sigma H(p_t)S^\top G(p_t) \\
= A(p_t)P_{t|t-1} A(p_t)^\top + G(p_t)QG(p_t)^\top + K_t\Omega_t K_t \\
- K_tC(p_t)P_{t|t-1} A(p_t)^\top + \Sigma H(p_t)S^\top G(p_t) \\
- [A(p_t)P_{t|t-1} C(p_t) + G(p_t)SH(p_t)]K_t^\top, \quad (32)
\]

Combining \( (30b) \) and \( (32) \) gives \( (4d) \).

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, **Optimal Filtering.** Prentice Hall, 1979.

[2] M. Petreczky and R. Vidal, “Realization theory of generalized bilinear systems,” *Submitted to IEEE Trans. on Automatic Control.*

[3] S. Boyd and Q. Yang, “Structured and simultaneous lyapunov functions for system stability problems,” *Int. J. of Control*, vol. 49, no. 6, pp. 2215–2240, 1989.

[4] J. J. Deyst and C. Price, “Conditions for asymptotic stability of the discrete minimum-variance linear estimator,” *IEEE Trans. on Automatic Control*, vol. 13, no. 6, pp. 702–705, Dec 1968.

[5] R. Tóth, H. S. Abbas, and H. Werner, “On the state-space realization of LPV input-output models: practical approaches,” *IEEE Trans. on Control Systems Technology*, vol. 20, no. 1, pp. 139–153, Jan. 2012.

[6] M. Verhaegen and V. Verdult, *Filtering and system identification: a least squares approach.* Cambridge university press, 2007.

[7] V. Verdult and M. Verhaegen, “Subspace identification of multivariable linear parameter-varying systems,” *Automatica*, vol. 38, no. 5, pp. 805–814, 2002.

[8] J. W. van Wingerden and M. Verhaegen, “Subspace identification of bilinear and LPV systems for open- and closed-loop data,” *Automatica*, vol. 45, no. 2, pp. 372–381, 2009.

[9] F. Felici, J. W. van Wingerden, and M. Verhaegen, “Subspace identification of MIMO LPV systems using a periodic scheduling sequence,” *Automatica*, vol. 43, no. 10, pp. 1684–1697, 2006.

[10] R. Tóth, *Modeling and identification of linear parameter-varying systems.* Springer, 2010.