Anomaly Detection Under Multiplicative Noise Model Uncertainty

Venkatraman Renganathan, *Member, IEEE*, Benjamin J. Gravell, *Graduate Student Member, IEEE*, Justin Ruths, and Tyler H. Summers

***Abstract*—State estimators are crucial components of anomaly detectors that are used to monitor cyber-physical systems. Many frequently-used state estimators are susceptible to model risk as they rely critically on the availability of an accurate state-space model. Modeling errors make it more difficult to distinguish whether deviations from expected behavior are due to anomalies or simply a lack of knowledge about the system dynamics. In this letter, we account for model uncertainty through a multiplicative noise framework. Specifically, we propose to use the multiplicative noise LQG based compensator in this setting to hedge against the model uncertainty risk. The size of the residual from the estimator can then be compared against a threshold to detect anomalies. Finally, the proposed detector is validated using numerical simulations. Extension of state-of-the-art anomaly detection in cyber-physical systems to handle model uncertainty represents the main novel contribution of the present work.**

***Index Terms*—Anomaly detection, multiplicative noise, coupled riccatti, LQG, vulnerable CPS.**

I. Introduction

CYBER-PHYSICAL Systems (CPS) are physical processes that are tightly integrated with computation and communication systems for monitoring and control. Though advances in CPS design has equipped them with adaptability, resiliency, safety, and security features that exceed the simple embedded systems of the past, it often leaves open several points for attackers to strike. CPS security problems have attracted the attention of researchers worldwide recently; some state-of-the-art anomaly detection algorithms can be found in [1]–[3].

Manuscript received September 14, 2021; revised November 11, 2021; accepted November 25, 2021. Date of publication December 13, 2021; date of current version December 22, 2021. This work was supported in part by the United States Air Force Office of Scientific Research under Award FA2386-19-1-4073 and in part by the National Science Foundation under Award ECCS-2047040. Recommended by Senior Editor F. Dabbene. *(Venkatraman Renganathan and Benjamin J. Gravell contributed equally to this work.) (Corresponding author: Venkatraman Renganathan.)*

Venkatraman Renganathan is with the Department of Automatic Control, Lund University, 22738 Lund, Sweden (e-mail: venkat@control.lth.se).

Benjamin J. Gravell, Justin Ruths, and Tyler H. Summers are with the Department of Mechanical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: benjamin.gravell@utdallas.edu; jruths@utdallas.edu; tyler.summers@utdallas.edu).

Digital Object Identifier 10.1109/LCSYS.2021.3134944

A common practice is to model a CPS as either a deterministic system or a stochastic system with additive Gaussian uncertainties. Motivated by the recent developments in distributionally robust optimization (DRO) techniques [4]–[6], authors in [7]–[9] have developed DRO anomaly detectors that remove assumptions on specific functional forms of the uncertainties in the stochastic CPS model. On the other hand, it is a common practice to assume that the true CPS dynamics are known exactly. Unfortunately, modeling and sampling errors are inherent and significant in working with real systems due to nonlinearities, learned (system identification, machine learning) models, adaptive models, or simply due to changing environmental conditions or aging. A multiplicative noise framework for capturing model uncertainty offers several compelling advantages over additive noise models. It provides a statistical description of the uncertainty that depends on the control input and state [10]–[12]. Using a multiplicative noise model, however, requires new tools to build and tune anomaly detectors that accommodate the more general functional form of the model.

State estimation is a crucial component in any model-based anomaly detector design, which depends on a state-space model for the system dynamics. This dependency causes limitations on the usage of the classical Kalman filter as it critically relies on the availability of an accurate state-space model, making it susceptible to model risk. Robust Kalman filtering with additive uncertainties was explored in [13], where the uncertain joint distribution of the states and outputs was accounted for. Another robust Kalman filter design was developed using a τ-divergence based family of distributions in [14]. In [15], a Wasserstein distributionally robust Kalman filter (W-DR-KF) was developed to account for distributional uncertainty. However, a procedure for jointly computing a pair of state estimator and feedback gain to guarantee stability in this setting remains unexplored.

Although stochastic modeling of CPS with additive uncertainty is well studied, there are no works to the best of our knowledge which have considered both multiplicative and additive noises together in the CPS security literature. The evolution of non-Gaussian state distributions under the effect of multiplicative noise invalidates use of the standard Kalman filter, as the separation principle available in linear quadratic Gaussian (LQG) setting in [16] no longer holds. Though [10] considered both multiplicative and additive noises in an optimal control setting, a restrictive Gaussian assumption was imposed on the uncertainties. The approach in this

letter builds on the foundation established by [17], where the multiplicative noise-driven LQG (MLQG) problem was solved by posing a set of coupled algebraic Riccati equations, from which the optimal linear output feedback controller and estimator gains were jointly computed.

Contributions: This letter is part of our ongoing work [7], [8] to leverage powerful results in control theory and distributionally robust optimization to design robust anomaly detectors. Specifically, the detector threshold corresponding to a desired false alarm rate in the setting considered in this letter was computed through the moment-based approaches explained [7]. In prior work we addressed detectors robust to non-Gaussian additive noise. In this letter,

1) We design an anomaly detector for stochastic linear cyber-physical systems that is robust to modeling errors. To our knowledge, this is the first paper to consider tuning an anomaly detector for a system model that incorporates model uncertainty. We propose a multiplicative noise framework and integrate the MLQG compensator to compute the residual.
2) We demonstrate our proposed approach using numerical simulations and show that multiplicative noises result in greater anomaly detector thresholds as long as mean square compensatability conditions are satisfied.

The rest of the letter is organized as follows. In Section II, the problem of monitoring an uncertain CPS with model uncertainty is formulated. Then, the multiplicative noise driven LQG compensator is discussed in Section III. Subsequently, the anomaly detector design is presented in Section IV. The proposed idea is then demonstrated using a numerical simulation in Section V. Finally, the letter is closed in Section VI along with directions for future research.

Notations & Preliminaries

The set of real numbers, integers are denoted by $\mathbb{R}, \mathbb{Z}$. The subset of real numbers greater than $a \in \mathbb{R}$ is denoted by $\mathbb{R}_{>a}$. The set of integers between two values $a, b \in \mathbb{Z}$ with $a < b$ is denoted by $[a : b]$. We denote by $\mathbb{S}^n$ the set of symmetric matrices in $\mathbb{R}^{n\times n}$ and the cone of positive definite (semi-definite) matrices on $\mathbb{S}^n$ as $\mathbb{S}^n_{++}(\mathbb{S}^n_{+})$. An identity matrix in dimension n is denoted by I_n. The Kronecker product of two matrices $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times q}$ is denoted by $A \otimes B$ and the vectorization of a matrix $A \in \mathbb{R}^{m\times n}$ is denoted by $\text{vec}(A) \in \mathbb{R}^{mn}$ and the matricization of vector $x \in \mathbb{R}^p$ is denoted by $\text{mat}(x, n, m) \in \mathbb{R}^{n\times m}$ where $n\times m = p$. The trace of a matrix $A \in \mathbb{R}^{n\times n}$ is denoted by $\mathbf{Tr}(A)$. A probability distribution with mean μ and covariance Σ is denoted by $\mathbb{P}(\mu, \Sigma)$, and specifically $\mathcal{N}_d(\mu, \Sigma)$ if the distribution is normal in $\mathbb{R}^d$. Given a matrix $A \in \mathbb{R}^{n\times n}$ and a vector valued random variable $z \in \mathbb{R}^p, p \geq 1$ with $\mathbb{E}[z] = \mu, \mathbb{E}[(z-\mu)(z-\mu)^\top] = \Sigma$, then $\mathbb{E}[z^\top A z] = \mathbf{Tr}(A\Sigma) + \mu^\top A\mu$.

II. Problem Formulation

A. Uncertain CPS Model

We model an uncertain CPS for time $k \in \mathbb{N}$ using a stochastic discrete-time linear time varying (LTV) system:

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \tag{1}$$

$$y_k = C_k x_k + v_k. \tag{2}$$

Here, $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, and $y_k \in \mathbb{R}^p$ are the system state, control input, and output at time k. The next-state $x_{k+1} \in \mathbb{R}^n$ is a random linear combination of the current state and process noise w_k, which is a zero-mean white noise process. Similarly, the output $y_k \in \mathbb{R}^p$ is a random linear combination of the states and the sensor noise v_k, which is a zero-mean white noise process. The initial state is a random variable $x_0 \sim \mathbb{P}_{x_0}(0, \Sigma_{x_0})$. The system matrices are decomposed as

$$A_k = \left(\bar{A} + \hat{A}_k\right),\ B_k = \left(\bar{B} + \hat{B}_k\right),\ C_k = \left(\bar{C} + \hat{C}_k\right),$$
$$\hat{A}_k = \sum_{i=1}^{n_a} \gamma_{ki}\mathcal{A}_i,\ \hat{B}_k = \sum_{j=1}^{n_b} \delta_{kj}\mathcal{B}_j,\ \hat{C}_k = \sum_{l=1}^{n_c} \kappa_{kl}\mathcal{C}_l. \tag{3}$$

where $\bar{A}$, $\bar{B}$, $\bar{C}$ denote the nominal dynamics, control, and output matrices respectively. Given the constants, $n_a, n_b, n_c \in \mathbb{Z}_{>0}$, the multiplicative noise terms are modeled by the i.i.d. across time (white), zero-mean, mutually independent scalar random variables γ_{ki}, δ_{kj}, κ_{kl}, which have variances $\sigma^2_{a,i}$, $\sigma^2_{b,j}$, $\sigma^2_{c,l}$ for $i \in [1 : n_a]$, $j \in [1 : n_b]$, $l \in [1 : n_c]$ respectively. The pattern matrices $\mathcal{A}_i \in \mathbb{R}^{n\times n}$, $\mathcal{B}_j \in \mathbb{R}^{n\times m}$, and $\mathcal{C}_l \in \mathbb{R}^{p\times n}$ specify how each scalar noise term affects the system matrices. It is then evident from (1) and (2) that $\hat{A}_k, \hat{B}_k$, and $\hat{C}_k$ quantify uncertainty about the nominal system matrices $\bar{A}, \bar{B}$, and $\bar{C}$ respectively. The distributions of all the scalar multiplicative noise random variables are assumed to be known. The covariance of the additive noises[1] (Σ_w, Σ_v) are assumed to be known; they may be estimated from collected data via, e.g., bootstrap sample averaging. For simplicity, we assume that x_0 and all the additive, multiplicative noises $w_k, v_k, \{\gamma_{ki}\}_{i=1}^{n_a}, \{\delta_{kj}\}_{j=1}^{n_b}, \{\kappa_{kl}\}_{l=1}^{n_c}$ are mutually independent of each other. We denote the first moment, second moment, and covariance of the state at time k as $\mu_{x_k} = \mathbb{E}[x_k]$, $V_k = \mathbb{E}[x_k x_k^\top]$, and $\Sigma_{x_k} = \mathbb{E}[(x_k - \mu_{x_k})(x_k - \mu_{x_k})^\top]$, respectively. Likewise, we denote the first moment, second moment, and covariance of the output at time k as $\mu_{y_k} = \mathbb{E}[y_k]$, $Y_k = \mathbb{E}[y_k y_k^\top]$, and $\Sigma_{y_k} = \mathbb{E}[(y_k - \mu_{y_k})(y_k - \mu_{y_k})^\top]$, respectively.

B. Review of Concepts

Here, we re-state some definitions from [17] on the mean squared versions of stabilizability, detectability and the resulting compensatability of systems given by (1) and (2).

Definition 1: The system in (1) is *mean-square stable* if $\forall x_0 \in \mathbb{R}^n, \exists V_\infty \in \mathbb{S}^n_+$ such that

$$\lim_{k\to\infty} V_k = \lim_{k\to\infty} \mathbb{E}\left[x_k x_k^\top\right] \to V_\infty.$$

Definition 2: The system in (1) is *mean-square stabilizable* if there exists a control gain matrix $K \in \mathbb{R}^{m\times n}$ such that using controls $u_k = Kx_k$ makes (1) mean-square stable.

Definition 3: The system in (1) and (2) is *mean-square compensatable* if there exist control and filter gain matrices $K \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{n\times p}$ such that the system

$$\begin{bmatrix} x_{k+1} \\ \hat{x}_{k+1} \end{bmatrix} = \begin{bmatrix} A_k & B_k K \\ LC_k & \bar{A} + \bar{B}K - L\bar{C} \end{bmatrix} \begin{bmatrix} x_k \\ \hat{x}_k \end{bmatrix}$$

[1]Even when the primitive random variables w_k, v_k, x_0 are assumed to be Gaussian, the resulting $\mathbb{P}_{x_k}$ at any time step $k > 0$ will be *non-Gaussian* due to the multiplicative noise.

is mean-square stable.

Assumptions:

1) The system given by (1) and (2) is *mean-square compensatable*.
2) The optimal state estimator at any time k given (1) and (2) is an affine[2] function of the output y_k.

Problem 1: Under the above assumptions for a given stochastic CPS model specified by (1), (2), obtain residual data from an appropriate state estimator module that accounts for both multiplicative and additive noises, and subsequently design an anomaly detector threshold such that the worst case false alarm rate does not exceed a desired value.

III. Residuals via Multiplicative Noise LQG

Due to the multiplicative noises in (1) and (2), the state distribution will be non-Gaussian even when all primitive noise distributions are Gaussian. Further, the classical separation principle from the additive noise setting does not hold in presence of multiplicative noises [17]. This necessitates a framework where the optimal controller and the estimator gains are computed *jointly*. Here, we elaborate on obtaining the residual from CPS using the multiplicative noise-driven LQG and show that the residual covariance is a function of both additive and multiplicative noise covariance matrices.

A. Designing Multiplicative Noise-Driven LQG

Under both multiplicative and additive noises in the system, the optimal linear output feedback controller can be exactly computed through the combination of a multiplicative noise KF with a multiplicative noise LQR as described in [10], [17], [18]. We consider the multiplicative noise-driven linear-quadratic Gaussian (MLQG) optimal control problem, which requires finding an output feedback controller $u_k = \pi_k(y_{0:k})$ for a system given by (1) and (2):

$$\underset{\pi_k \in \Pi_k}{\text{minimize}} \ \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_{\mathcal{E}_k}\left[\sum_{k=0}^{T-1} x_k^\top Q x_k + u_k^\top R u_k\right], \quad \text{subject to (1), (2),} \tag{4}$$

where $\mathcal{E}_k = \{x_0, \{\hat{A}_k\}, \{\hat{B}_k\}, \{\hat{C}_k\}, \{w_k\}, \{v_k\}\}$, $Q \succeq 0, R \succ 0$. Then, the optimal linear compensator gain matrices can be computed by solving the following coupled nonlinear matrix Riccati equations in symmetric matrix variables $P_1, P_2, P_3, P_4 \in \mathbb{S}_+^n$:

$$P_1 = Q + \bar{A}^\top P_1 \bar{A} + \sum_{i=1}^{n_a} \sigma_{a,i}^2 \mathcal{A}_i^\top P_1 \mathcal{A}_i - K^\top K_\alpha K + \sum_{i=1}^{n_a} \sigma_{a,i}^2 \mathcal{A}_i^\top P_2 \mathcal{A}_i + \sum_{i=1}^{n_c} \sigma_{c,i}^2 \mathcal{C}_i^\top L^\top P_2 L \mathcal{C}_i, \tag{5}$$

$$P_2 = (\bar{A} - L\bar{C})^\top P_2 (\bar{A} - L\bar{C}) + K^\top K_\alpha K, \tag{6}$$

$$P_3 = \Sigma_w + \bar{A} P_3 \bar{A}^\top - L L_\alpha L^\top + \sum_{i=1}^{n_a} \sigma_{a,i}^2 \mathcal{A}_i P_3 \mathcal{A}_i^\top + \sum_{i=1}^{n_a} \sigma_{a,i}^2 \mathcal{A}_i P_4 \mathcal{A}_i^\top + \sum_{i=1}^{n_b} \sigma_{b,i}^2 \mathcal{B}_i K P_4 K^\top \mathcal{B}_i^\top, \tag{7}$$

[2]It is possible to design a nonlinear state estimator to outperform a given affine estimator in this setting. However, it is out of the scope of this letter.

$$P_4 = (\bar{A} + \bar{B}K) P_4 (\bar{A} + \bar{B}K)^\top + L L_\alpha L^\top, \tag{8}$$

where for notation simplicity, we denote

$$K_\alpha = R + \bar{B}^\top P_1 \bar{B} + \sum_{j=1}^{n_b} \sigma_{b,j}^2 \mathcal{B}_j^\top P_1 \mathcal{B}_j + \sum_{j=1}^{n_b} \sigma_{b,j}^2 \mathcal{B}_j^\top P_2 \mathcal{B}_j \tag{9}$$

$$L_\alpha = \Sigma_v + \bar{C} P_3 \bar{C}^\top + \sum_{j=1}^{n_c} \sigma_{c,j}^2 \mathcal{C}_j P_3 \mathcal{C}_j^\top + \sum_{j=1}^{n_c} \sigma_{c,j}^2 \mathcal{C}_j P_4 \mathcal{C}_j^\top. \tag{10}$$

Then, the associated optimal controller and estimator gains (K, L) are given by

$$K = -K_\alpha^{-1} \bar{B}^\top P_1 \bar{A}, \tag{11}$$

$$L = \bar{A} P_3 \bar{C}^\top L_\alpha^{-1}. \tag{12}$$

Finally, the optimal linear compensator is

$$u_k = K\hat{x}_k, \quad \text{and} \tag{13}$$

$$\hat{x}_{k+1} = (\bar{A} + \bar{B}K)\hat{x}_k + L(y_k - \bar{C}\hat{x}_k), = (\bar{A} + \bar{B}K - L\bar{C})\hat{x}_k + LC_k x_k + L v_k \tag{14}$$

It is necessary to account for the multiplicative noise to achieve the minimum quadratic cost; furthermore, it is straightforward to find systems in (1) and (2) which are *mean-square unstable* when controlled by (multiplicative-noise-ignorant) LQG, meaning that it is necessary to account for multiplicative noise to achieve mean-square stability.

B. Residual From Multiplicative Noise LQG

We define the estimation error as $e_k = x_k - \hat{x}_k$. Then the estimation error evolves as follows

$$e_{k+1} = (\bar{A} - \hat{B}_k K - L\bar{C})e_k + (\hat{A}_k + \hat{B}_k K - L\hat{C}_k)x_k + w_k - L v_k. \tag{15}$$

It is evident from above that estimation error is a function of the multiplicative noise terms. We now elaborate how to obtain the residual signal required for anomaly detection. Define the residual $r_k \in \mathbb{R}^p$ as

$$r_k = y_k - \bar{C}\hat{x}_k = \bar{C}e_k + \hat{C}_k x_k + v_k \quad \text{and} \tag{16}$$

$$\mathbb{E}[r_k] = \mathbb{E}[\bar{C}e_k + \hat{C}_k x_k + v_k] = \bar{C}\mathbb{E}[e_k]. \tag{17}$$

Then, r_k is not necessarily Gaussian due to the multiplicative noise and has mean $\mathbb{E}[r_k] = \bar{C}\mathbb{E}[e_k]$ (it becomes zero mean $\forall k \geq 0$ if $e_0 = 0$) with raw second moment matrix whose vectorized form is given by

$$R_k = (\bar{C} \otimes \bar{C})E_k + \mathbb{E}\left[\hat{C}_k \otimes \hat{C}_k\right]X_k + \text{vec}(\Sigma_v). \tag{18}$$

To compute the steady state raw second moments of the residual r_k, we define

$$E_k = \text{vec}\left(\mathbb{E}[e_k e_k^\top]\right), \quad X_k = \text{vec}\left(\mathbb{E}[x_k x_k^\top]\right),$$

$$\tilde{X}_k = \text{vec}\left(\mathbb{E}\left[x_k \hat{x}_k^\top\right]\right), \quad \check{X}_k = \text{vec}\left(\mathbb{E}\left[\hat{x}_k x_k^\top\right]\right),$$

$$\hat{X}_k = \text{vec}\left(\mathbb{E}\left[\hat{x}_k \hat{x}_k^\top\right]\right), \quad R_k = \text{vec}\left(\mathbb{E}[r_k r_k^\top]\right)$$

$$\mathcal{X}_k := \begin{bmatrix} X_k^\top & \tilde{X}_k^\top & \check{X}_k^\top & \hat{X}_k^\top \end{bmatrix}^\top, \quad \mathcal{V} := \begin{bmatrix} \text{vec}(\Sigma_w) \\ \text{vec}(\Sigma_v) \end{bmatrix},$$

$$\Sigma_A' = \mathbb{E}\left[\hat{A}_k \otimes \hat{A}_k\right] = \sum_{i=1}^{n_a} \sigma_{a,i}^2 (\mathcal{A}_i \otimes \mathcal{A}_i),$$

$$H = \begin{bmatrix} \bar{A}\otimes\bar{A}+\Sigma'_A & (\bar{B}K)\otimes\bar{A} & \bar{A}\otimes(\bar{B}K) & \left(\bar{B}\otimes\bar{B}+\Sigma'_B\right)(K\otimes K) \\ (L\bar{C})\otimes\bar{A} & (\bar{A}+\bar{B}K-L\bar{C})\otimes\bar{A} & (L\bar{C})\otimes(\bar{B}K) & (\bar{A}+\bar{B}K-L\bar{C})\otimes(\bar{B}K) \\ \bar{A}\otimes(L\bar{C}) & (\bar{B}K)\otimes(L\bar{C}) & \bar{A}\otimes(\bar{A}+\bar{B}K-L\bar{C}) & (\bar{B}K)\otimes(\bar{A}+\bar{B}K-L\bar{C}) \\ (L\otimes L)(\bar{C}\otimes\bar{C}+\Sigma'_C) & (\bar{A}+\bar{B}K-L\bar{C})\otimes(L\bar{C}) & (L\bar{C})\otimes(\bar{A}+\bar{B}K-L\bar{C}) & (\bar{A}+\bar{B}K-L\bar{C})\otimes(\bar{A}+\bar{B}K-L\bar{C}) \end{bmatrix}.$$

Fig. 1. The Matrix H in (19) with terms containing the second moments of entries of the vector $\mathcal{X}_k$.

$$\Sigma'_B = \mathbb{E}\Big[\hat{B}_k\otimes\hat{B}_k\Big] = \sum_{j=1}^{n_b}\sigma^2_{b,j}(\mathcal{B}_j\otimes\mathcal{B}_j),$$
$$\Sigma'_C = \mathbb{E}\Big[\hat{C}_k\otimes\hat{C}_k\Big] = \sum_{l=1}^{n_c}\sigma^2_{c,l}(\mathcal{C}_l\otimes\mathcal{C}_l).$$

Then, it is straight forward to see that $\mathcal{X}_k$ evolves as follows

$$\mathcal{X}_{k+1} = H\mathcal{X}_k + \underbrace{\begin{bmatrix} I_n\otimes I_n & 0_{n^2\times 1} \\ 0_{n^2\times n^2} & 0_{n^2\times 1} \\ 0_{n^2\times n^2} & 0_{n^2\times 1} \\ 0_{n^2\times n^2} & L\otimes L \end{bmatrix}}_{:=\Phi}\mathcal{V}, \tag{19}$$

where the matrix H in (19) gathers all the resulting coefficients obtained while expanding the entries of the vector $\mathcal{X}_k$. The algebra resulting in the following expression of H is available in the Appendix of [19]. Since the optimal gain matrices K, L achieve mean-square compensation of the system (1) and (2), the covariance of the estimation error will have a steady state value. Since by assumption, $\bar{A}-L\bar{C}$ is Schur stable, we see that $\mathbb{E}[e_k]\to 0$ as $k\to\infty$ regardless of the initial state-residual e_0 which in turn results in $\mathbb{E}[r_k]\to 0$ as $k\to\infty$. That is, $\mathbb{E}[e_\infty]=0 \implies \mathbb{E}[r_\infty]=0$ and subsequently in steady state,

$$\mathcal{X}_\infty = H\mathcal{X}_\infty + \Phi\mathcal{V}. \tag{20}$$
$$\iff \mathcal{X}_\infty = (I_{4n^2}-H)^{-1}\Phi\mathcal{V}. \tag{21}$$

This amounts to solving a (generalized) Lyapunov equation. Such an equation can be solved more efficiently by specialized solvers which do not require the inverse to be computed explicitly; for simplicity we present the equation and its solution in this form. However, the Schur stability of the matrix H subject to the mean-square compensation achieved by the matrices (K, L) determines whether the resulting $\mathcal{X}_\infty$ (which exists no matter whatever approach is used to compute it) can be employed to compute the steady state residual moments. For instance, in a strong multiplicative noise setting, the matrix H defined using (K, L) matrices that do not achieve mean-square compensation will *not* be Schur stable and the resulting $\mathcal{X}_\infty$ cannot be used meaning that steady state Σ_r does not exist. Having obtained a valid $\mathcal{X}_\infty$, the steady state second moments of the state- and output-residuals can then be computed as

$$E_\infty = X_\infty - \tilde{X}_\infty - \check{X}_\infty + \hat{X}_\infty, \quad \text{and} \tag{22}$$
$$R_\infty = (\bar{C}\otimes\bar{C})E_\infty + \Sigma'_C X_\infty + \mathrm{vec}(\Sigma_v). \tag{23}$$

Finally, using the matrix reshaping operator $\mathrm{mat}(\cdot)$, we retrieve the steady state Σ_r as follows

$$\Sigma_{x_\infty} = \mathrm{mat}(E_\infty, n, n), \quad \text{and} \tag{24}$$
$$\Sigma_r = \mathrm{mat}(R_\infty, p, p). \tag{25}$$

IV. Anomaly Detector Design With Residual From MLQG Compensation

We now present how to analyze the residual obtained from the MLQG compensator and elaborate the procedure to construct the corresponding anomaly detector threshold in this section. Note that the covariance of the residual computed through (18) is a function of covariance matrices of both the additive and multiplicative noises. This is in sharp contrast to the case in [7], [8] where the residual covariance was just a function of the additive noise covariance. Further, to account for the changes in the covariance of the residual, we form a quadratic distance measure as

$$q_k = r_k^\top \Sigma_r^{-1} r_k. \tag{26}$$

It is then straightforward to see that

$$\begin{aligned}\mathbb{E}[q_k] &= \mathbb{E}[r_k^\top\Sigma_r^{-1}r_k] \\ &= \mathbf{Tr}(\Sigma_r^{-1}\Sigma_r) + (\bar{C}\mathbb{E}[e_k])^\top\Sigma_r^{-1}(\bar{C}\mathbb{E}[e_k]) \\ &= p + (\mathbb{E}[e_k])^\top\bar{C}^\top\Sigma_r^{-1}\bar{C}\mathbb{E}[e_k].\end{aligned} \tag{27}$$

This implies that (27) is applicable only when mean-square compensation is achieved through properly designed (K, L) matrix pair as the steady state Σ_r is guaranteed to exist in that case. Then, for a given q_k from (26) and a threshold $\alpha\in\mathbb{R}_{>0}$ corresponding to a desired false alarm rate $\mathcal{F}$, the anomaly detector can be designed such that alarm time(s) $k^\star\in\mathbb{N}$ are produced according to the following rules

$$\begin{cases} q_k \le \alpha, & \text{no alarm,} \\ q_k > \alpha, & \text{alarm: } k^\star = k. \end{cases} \tag{28}$$

If $\mathbb{P}_{r_k}$ was Gaussian, then q_k would follow the chi-squared distribution, meaning that for a given tail probability defined using $\mathcal{F}$, the chi-squared detector described as in [2] can be used to obtain the required detector threshold. However, in our setting due to the multiplicative noises, $\mathbb{P}_{r_k}$ is *non-Gaussian* and thereby the chi-squared detector is *not* appropriate. We instead utilize a moment-based approach for constructing the threshold. We propose to use the higher-order moment based anomaly detector design proposed in [7] to design the detector threshold in this setting. The residual q_k is collected for a sufficiently long period of time to form the s-moments based ambiguity set $\mathcal{P}_q^s := \{\mathbb{P}_q \mid \mathbb{E}[q_k^s] = M_q^s\}$. The optimal threshold $\alpha^\star_{q,s}$[3] satisfying

$$\sup_{\mathbb{P}_q\in\mathcal{P}_q^s}\mathbb{P}_q\Big[q_k > \alpha^\star_{q,s}\Big] \le \mathcal{F}, \tag{29}$$

can then be obtained by directly invoking [7, Th. 4] corresponding to a given desired false alarm rate $\mathcal{F}$.

[3]The two subscripts q, s in $\alpha^\star_{q,s}$ denote the random variable and the number of moments considered respectively.

V. Numerical Results

We consider an inverted pendulum with a torque-producing actuator whose dynamics have been linearized about the vertical equilibrium. That is, the pendulum of mass m is suspended by a mass-less rod of length l and the angle θ is measured from the downward vertical with positive counter clockwise direction. The corresponding nonlinear differential equation of the pendulum mass is

$$\ddot{\theta} = m_c \sin(\theta) + \tau, \tag{30}$$

where $m_c = -\frac{g}{l}$ denotes the uncertain mass constant. Let us denote the state vector by $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} \theta & \dot{\theta} \end{bmatrix}$ and the torque input by $u = \tau$. Then, the corresponding discrete time dynamics obtained through the forward Euler discretization of the linearized dynamics of (30) around the equilibrium point $\tilde{x} = (\pi, 0)$ with step size Δt is

$$x_{k+1} = \begin{bmatrix} 1 & \Delta t \\ m_c \Delta t & 1 \end{bmatrix} x_k + \begin{bmatrix} 0 \\ \Delta t \end{bmatrix} u_k + w_k. \tag{31}$$

Uncertainty on the mass constant m_c corresponds to uncertainty on the matrix A. We consider an example where the true mass constant is $m_c = 10$, but the nominal model underestimates it as $m_c = 5$. We take a step size $\Delta t = 0.1$. At discrete time instances, the sensor returns a noisy measurement of the angular position of pendulum. Hence the corresponding linearized noisy output model is,

$$y = \theta + v_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k + v_k. \tag{32}$$

Both w_k and v_k are sampled from the multivariate Laplacian (which has heavier tails than Gaussian with same mean and covariance) with zero-mean and covariance $\Sigma_w = 2I_n$, $\Sigma_v = 2I_p$ respectively. The state and control penalty matrices are $Q = I_n, R = I_m$ respectively. The multiplicative noise was considered to exist both in the A and C matrices, with the direction matrices being $\mathcal{A}_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ and $\mathcal{C}_1 = \begin{bmatrix} 0.1 & 0 \end{bmatrix}$ with the multiplicative noise variances $\gamma_{k,1} \sim \mathcal{N}(0, \sigma_{a,1}^2)$, $\kappa_{k,1} \sim \mathcal{N}(0, \sigma_{c,1}^2)$ respectively. The non-Gaussian additive primitive noises w_k, v_k along with these multiplicative noises render the traditional chi-squared detector to be ineffective as the system states will evolve to be *non-Gaussian* for all $t > 0$. Through simulation, we collected the quadratic distance measure q_k data for $T = 10^7$ time steps for the above system with multiplicative noises under two different settings namely, 1) using the standard LQG, and 2) using multiplicative noise-driven LQG compensators. The q_k data was then used to tune the anomaly detector for a desired false alarm rate of $\mathcal{F} = 5\%$ using [7, Th. 4] with $s = 4$ moments in (29) and along with a bisection tolerance of $\epsilon = 10^{-4}$. The resulting moment bound problem was solved using the SOSToolbox on MATLAB with the SeDuMi solver. The code is made publicly available at https://github.com/TSummersLab/AnomalyDetectionMultiplicativeNoise

A. LQG & MLQG With Low Multiplicative Noises

When the system was simulated with low multiplicative noise variances $\sigma_{a,1}^2 = \sigma_{c,1}^2 \leq 0.10$, the resulting (K, L) matrix pair from both the LQG and the MLQG compensators had similar values and the anomaly detectors from both compensators had similar good performances. However, the performance of MLQG started getting better with $\sigma_{a,1}^2 = \sigma_{c,1}^2 > 0.10$ and the results with $\sigma_{a,1}^2 = \sigma_{c,1}^2 = 0.06$ are shown in Figure 2. The histograms of the q_k data using the MLQG and LQG estimators are shown in red and cyan colors respectively. The mean-square compensation of the MLQG compensator was verified via the convergence of the coupled Riccati equations and subsequently the corresponding collected q_k data resulted in an optimal detector threshold $\alpha_{q,4}^\star = 8.247$ with false alarm rate being 0.89%. Similarly, when the q_k data collected from the standard LQG was evaluated against a similarly computed threshold $\alpha_{q,4}^\star = 8.422$, it resulted in 0.86% false alarms. Though both MLQG and LQG achieve mean-square compensation at a lower noise setting, the MLQG results in a tighter threshold than the LQG. Further, the resulting H matrix from LQG compensator *ceased* to be Schur stable for $\sigma_{a,1}^2 = \sigma_{c,1}^2 > 0.11$ agreeing with results in Table I. Supposedly, if we used the unstable H matrix in the LQG case, it resulted in $\hat{\mathbb{E}}[q_k] \to \infty$ when the variances became stronger and thereby restricted us from using even the simplest Markov bound in this case to obtain the detector threshold.

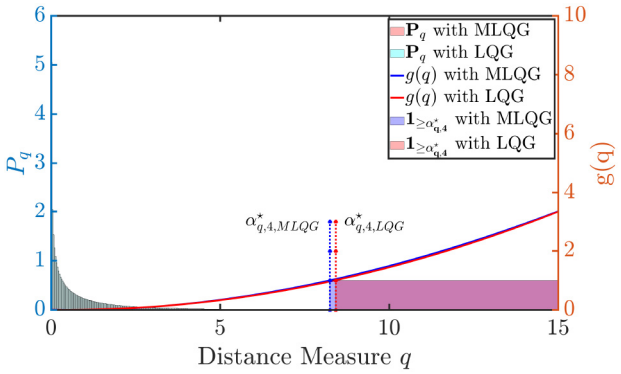

Fig. 2. **Detector Threshold With Multiplicative Noise:** The histograms of the q_k using the MLQG and LQG estimators with $\sigma_{a,1}^2 = \sigma_{c,1}^2 = 0.06$ are shown in red and cyan colors respectively. The moment based polynomial $g(q)$ shown in blue and red curves bound their indicator functions in shaded blue and red respectively. Though both MLQG and LQG achieve mean-square compensation, the MLQG results in a tighter threshold than the LQG.

TABLE I
Effect of Varying the Low Multiplicative Noise Variances ($\sigma_{a,1}^2 = \sigma_{c,1}^2$) on the Resulting $\lambda_{\max}(H)$ Corresponding to LQG and MLQG Compensators Are Shown Here

Method	$\sigma_{a,1}^2 = \sigma_{c,1}^2$				
	0.02	0.04	0.06	0.08	0.10
$\lambda_{\max}(H)$ LQG	0.9105	0.9414	0.9625	0.9789	0.9926
$\lambda_{\max}(H)$ MLQG	0.8908	0.9071	0.9159	0.9217	0.9259

B. Effect of Multiplicative Noise Variance on the Worst Case False Alarm Rate

Here, we show how the variances $\sigma_{a,1}^2, \sigma_{c,1}^2$ of the multiplicative noises $\gamma_{k,1}, \kappa_{k,1}$ respectively affect the resulting anomaly detector's worst case false alarm rate. Starting from $\sigma_{a,1}^2 = \sigma_{c,1}^2 = 0.15$, we simulated the system by increasing the variances and the results are in Table II. It

TABLE II

EFFECT OF VARYING THE MULTIPLICATIVE NOISE VARIANCES ($\sigma^2_{a,1} = \sigma^2_{c,1}$) ON THE RESULTING $\lambda_{\max}(H)$, STEADY STATE Σ_r, OPTIMAL THRESHOLD $\alpha^\star_{q,4}$, SAMPLE BASED MEAN $\hat{\mathbb{E}}[q_k]$ AND THE WORST CASE FALSE ALARM RATES $\mathcal{F}_{worse}$ FROM THE MLQG COMPENSATOR ARE SHOWN HERE. IT IS EVIDENT THAT THE MLQG IS CAPABLE OF MEAN-SQUARE COMPENSATING THE SYSTEM EVEN WITH INCREASING MULTIPLICATIVE NOISE VARIANCES UP TO A LIMIT

$\sigma^2_{a,1} = \sigma^2_{c,1}$	$\lambda_{\max}(H)$	Σ_r	$\hat{\mathbb{E}}[q_k]$	$\mathcal{F}_{worse}(\%)$	$\alpha^\star_{q,4}$
0.15	0.9329	6.54	1.000	0.88	8.31
0.20	0.9372	6.73	1.000	0.87	8.37
0.25	0.9403	6.92	1.000	0.79	8.67
0.30	0.9426	7.10	1.000	0.74	8.91
⋮	⋮	⋮	⋮	⋮	⋮
3.77	N/A	N/A	N/A	N/A	N/A

is evident that MLQG compensator was capable of mean-square compensate the system with increasing covariances by resulting in finite mean (equal to 1 and thereby agreeing with (27)). Starting from $\sigma^2_{a,1} = \sigma^2_{c,1} \geq 0.45$, numerical issues started accompanying the threshold calculations due to exploding values of the moments (can be addressed using orthogonal basis such as the Legendre polynomial basis to provide numerical stability). Specifically, when the variances were increased beyond $\sigma^2_{a,1} = \sigma^2_{c,1} \geq 3.77$, the coupled Riccatti equations corresponding to the MLQG stopped converging as mean-square compensation was lost for such higher variance multiplicative noises. The effect of increasing variance also affected the resulting false alarm rates when the residuals from the MLQG compensator was compared against its respective threshold. The resulting optimal threshold $\alpha^\star_{q,4}$ increased when the multiplicative noise variances increased. For this reason, in this problem setting the false alarm rate of MLQG happened to decrease with increased multiplicative noise variance; there is a nontrivial relation between the multiplicative noise variances and the threshold designed by the detection scheme, which depends, e.g., on the coupled Riccati equation solution. As shown in Table II, the MLQG with finite set of $s = 4$ empirical moments starting from $\hat{\mathbb{E}}[q_k]$ guaranteed that the resulting worst case false alarm rate are always upper bounded by the desired value of $\mathcal{F} = 5\%$.

VI. CONCLUSION

An extension of the state-of-the-art anomaly detection algorithms for CPS with modeling errors via the multiplicative noise framework was discussed in this letter. The multiplicative noise-driven LQG being a robust state estimator was used to hedge against the model risk to construct the state estimate. The proposed method was demonstrated using a numerical simulation. Future work seeks to investigate the setting where the multiplicative noise distributions are unknown and to obtain online estimates of the system dynamics through system identification technique combined with the above compensator for implementing data-driven distributionally robust anomaly detection for vulnerable CPS.

REFERENCES

[1] J. Giraldo *et al.*, "A survey of physics-based attack detection in cyber-physical systems," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–36, Jul. 2018.

[2] N. Hashemi and J. Ruths, "Generalized chi-squared detector for LTI systems with non-Gaussian noise," in *Proc. Amer. Control Conf. (ACC)*, Philadelphia, PA, USA, 2019, pp. 404–410.

[3] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[4] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Oper. Res.*, vol. 58, pp. 902–917, Apr. 2010.

[5] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations," *Math. Program.*, vol. 171, nos. 1–2, pp. 115–166, 2018.

[6] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Oper. Res.*, vol. 62, no. 6, pp. 1358–1376, 2014.

[7] V. Renganathan, N. Hashemi, J. Ruths, and T. H. Summers, "Higher-order moment-based anomaly detection," *IEEE Contr. Syst. Lett.*, vol. 6, pp. 211–216, 2022, doi: 10.1109/LCSYS.2021.3058269.

[8] V. Renganathan, N. Hashemi, J. Ruths, and T. H. Summers, "Distributionally robust tuning of anomaly detectors in cyber-physical systems with stealthy attacks," in *Proc. Amer. Control Conf. (ACC)*, Denver, CO, USA, 2020, pp. 1247–1252.

[9] D. Li and S. Martínez, "High-confidence attack detection via Wasserstein-metric computations," *IEEE Contr. Syst. Lett.*, vol. 5, no. 2, pp. 379–384, Apr. 2021.

[10] W. Li, E. Todorov, and R. E. Skelton, "Estimation and control of systems with multiplicative noise via linear matrix inequalities," in *Proc. Amer. Control Conf.*, Portland, OR, USA, 2005, pp. 1811–1816.

[11] B. J. Gravell, P. M. Esfahani, and T. H. Summers, "Robust control design for linear systems via multiplicative noise," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 7392–7399, 2020.

[12] B. Gravell, P. M. Esfahani, and T. Summers, "Learning optimal controllers for linear systems with multiplicative noise via policy gradient," *IEEE Trans. Autom. Control*, vol. 66, no. 11, pp. 5283–5298, Nov. 2021.

[13] B. C. Levy and R. Nikoukhah, "Robust state space filtering under incremental model perturbations subject to a relative entropy tolerance," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 682–695, Mar. 2013.

[14] M. Zorzi, "Robust Kalman filtering under model perturbations," *IEEE Trans. Autom. Control*, vol. 62, no. 6, pp. 2902–2907, Jun. 2017.

[15] S. S. Abadeh, V. A. Nguyen, D. Kuhn, and P. M. M. Esfahani, "Wasserstein distributionally robust Kalman filtering," in *Advances in Neural Information Processing Systems*, vol. 31. Red Hook, NY, USA: Curran Assoc., Inc., 2018.

[16] R. E. Kalman *et al.*, "Contributions to the theory of optimal control," *Boletin de la Sociedad Matematica Mexicana*, vol. 5, no. 2, pp. 102–119, 1960.

[17] W. L. De Koning, "Compensatability and optimal compensation of systems with white parameters," *IEEE Trans. Autom. Control*, vol. 37, no. 5, pp. 579–588, May 1992.

[18] L. El Ghaoui, "State-feedback control of systems with multiplicative noise via linear matrix inequalities," *Syst. Control Lett.*, vol. 24, no. 3, pp. 223–228, 1995.

[19] V. Renganathan, B. J. Gravell, J. Ruths, and T. H. Summers, "Anomaly detection under multiplicative noise model uncertainty," 2021, *arXiv:2103.15228*.