A Dynamical System Pair with Identical First Two Moments But Different Probability Densities

Abhishek Halder, Kooktae Lee, and Raktim Bhattacharya

Abstract— **Often in the literature, stochastic dynamical systems are approximated by moment closure techniques, closure in second moment being common practice. This refers to truncating the statistics generated by time varying probability density functions which evolve under the action of the trajectory-level dynamics. Although it is known that such moment closure approximations may lead to incorrect inferences, explicit examples at the dynamical systems level, are rare in the literature. In this paper, using optimal transport theory, we construct two dynamical systems such that starting from the *same initial condition ensemble*, their first two moments match *at all times*, but the underlying probability densities do not. This example serves as a motivation to consider the entire joint probability density function, as opposed to first few moments, for approximating stochastic systems in general, and stochastic jump linear systems in particular.**

I. Introduction

In recent times, stochastic jump systems [1] have emerged as a powerful modeling framework for a variety of applications like communication networks [2], [3], nonlinear optics [4], multi-target tracking [5], failure prone manufacturing systems [6], and flight controllers subject to electromagnetic disturbances [7]. These applications have the commonality that the system dynamics undergo abrupt variations between a finite number of modes. Although considerable progress have been made in the analysis of such systems, often the results have significant computational complexities depending on the nature of stochastic jumps and modal dynamics. This is particularly true for problems like uncertainty propagation and estimation [5], [8], [9], where one has additional probabilistic uncertainties (e.g. initial condition, parametric) than the occurence of random jumps. An alternative then, is to find a simpler abstraction of the system dynamics, upon which one can perform the analysis and the results would approximate those of the original stochastic jump system.

A. Related Work

Approximation of stochastic jump systems have been studied in literature from the perspective of approximate bisimulation [10], $\mathcal{H}_\infty$ model reduction [11], and balanced truncation [12]. These ideas quantify *trajectory level closeness* between the outputs of the original and approximated system. From a statistical perspective, a natural approximate model would be one that approximates the time varying joint output probability density function (PDF) generated by the original stochastic jump system. This idea of *density level closeness* appeared recently [13]–[15] in the context of model validation. In practice, moment closure approximations are widely used [16]–[18], where one derives a model that matches first few statistical moments of the original system. Since closure approximations truncate the infinite dimensional moment dynamics, they can be thought of as intermediate between trajectory level and density level closeness. In particular, Gaussian moment closure, where all except the first two moments are ignored, is often used in the literature [16].

If the original dynamics generates non-Gaussian joint output PDFs, then Gaussian moment closure, for example, would result approximation errors. Since the joint PDF of the original system changes over time, so does the approximation error. Hence, it is difficult to gauge in general, whether moment closure approximation error would be significant or not, for most of the times. This, we believe, has partly contributed to the popularity of the closure techniques, since it may happen that the original system has a fast time scale dynamics that moves the non-Gaussian transient PDFs quickly to a Gaussian.

B. Contributions of This Paper

The purpose of this paper is to construct an approximation for a given *stochastic jump linear system* (SJLS) such that starting from the *same joint PDF over the space of initial conditions*, the first two statistical moments generated by the approximate and original dynamics match *at all times*, however the joint PDFs do not. Such explicit example, to the best of our knowledge, has been lacking in the literature.

A secondary contribution of this paper is the use of optimal transport to construct an *affine time varying* (ATV) approximation for the SJLS, while matching the first two moments. Unlike existing methods [10]–[12], we do not make any *a priori* assumption on the structure of the approximate dynamical system. While most previous works assume the approximation to be structurally same (but reduced order) as that of the original system, our results provide *non-jump approximation for jump linear system.*

C. Structure of the Paper

This paper is organized as follows. In Section II, we introduce discrete-time stochastic jump linear systems and for Gaussian or MoG initial PDF, derive its joint state PDF evolution in closed form. Then we derive the mean and covariance evolution for the same. To demonstrate that the choice of approximating model structure is non-trivial, in

Abhishek Halder, Kooktae Lee, and Raktim Bhattacharya are with the Department of Aerospace Engineering, Texas A&M University, College Station, TX 77843, USA, {ahalder,animodor,raktim}@tamu.edu

Section III, we show that it is impossible to realize the mean-covariance sequences generated by the stochastic jump linear system via a linear time invariant dynamics, unless the jump process is i.i.d. In Section IV, we briefly review the optimal transport ideas. Section V combines the results of Section II and the optimal transport ideas from Section IV, to derive a constructive Gaussian moment closure approximation for stochastic jump linear system. A numerical example is worked out in Section VI. Section VII concludes the paper.

D. Notations

Most notations are standard. The set of natural numbers is denoted as $\mathbb{N}$, and $\mathbb{N}_0 \triangleq \mathbb{N} \cup \{0\}$. The symbol Symm_+ denotes the space of symmetric positive semi-definite matrices. The notation $x \sim \rho$ means that the random vector x has the joint PDF $\rho(x)$. We will use the abbreviation "MoG" to mean mixture of Gaussians, and "G" to mean Gaussians. For example, MoG(k) will denote the mixture of Gaussian PDF at time k. Furthermore, $\sharp$ denotes the push-forward for a random vector, and the superscript $\star$ refers to optimality, unless defined otherwise. The notation $\mathcal{N}(\mu, \Sigma)$ denotes Gaussian PDF with mean μ and covariance Σ.

II. Stochastic Jump Linear Systems

A. Preliminaries

Definition 1: (**Jump linear system**) A discrete-time jump linear system (JLS) with $m \in \mathbb{N}$ modes, is given by

$$x(k+1) \quad = \quad A_{\sigma_k} x(k), \quad x \in \mathbb{R}^n, \qquad (1)$$

where the discrete time index $k \in \mathbb{N}_0$, the set of non-negative integers. The symbol $\{\sigma_k\}$ denotes the switching sequence of the jump system, i.e. $\sigma_k : \mathbb{N}_0 \mapsto \{1, 2, \ldots, m\}$. Thus, the JLS (1) is characterized by (i) a set of m matrices $\{A_i\}_{i=1}^m$, and (ii) a switching sequence $\{\sigma_k\}_{k\in\mathbb{N}_0}$.

Remark 1: (**Stochastic jump linear system**) The switching sequence $\{\sigma_k\}$ may be generated through a deterministic policy, or as a sample path realization of a discrete time stochastic process over the finite set $\{1, 2, \ldots, m\}$. In the latter case, (1) is referred as a *stochastic jump linear system* (SJLS), and is characterized by the set of modal matrices $\{A_i\}_{i=1}^m$, and the sequence of occupation probability vectors $\{\pi(k)\}_{k\in\mathbb{N}_0} \triangleq \{\pi_1(k), \pi_2(k), \ldots, \pi_m(k)\}_{k\in\mathbb{N}_0}$ governing the stochastic switching sequence $\{\sigma_k\}_{k\in\mathbb{N}_0}$. In this paper, we will consider approximating discrete time SJLS, given by the tuple $(\{A_i\}_{i=1}^m, \{\pi(k)\}_{k\in\mathbb{N}_0})$.

B. Uncertainty Propagation in SJLS

Next, we consider the evolution of initial condition uncertainties through dynamics (1). For clarity of exposition, we only consider the case where the joint PDF over initial conditions is Gaussian, and the output vector $y(k) \in \mathbb{R}^n$ for (1), is identical to the state $x(k)$. It will be apparent from the sequel that our results can be generalized to the case when both these assumptions are violated, e.g. when the initial PDF is a mixture of Gaussian, instead of Gaussian; and when $y(k) = C_{\sigma_k} x(k)$, $y \in \mathbb{R}^{n_o}$, $n_o \neq n$. For notational ease, we will not consider these generalizations in this paper.

Lemma 1: Given m absolutely continuous random vectors $X_1, \ldots, X_m$, with respective CDF $F_j(x)$, and PDF $\varsigma_j(x)$, where $j = 1, 2, \ldots, m$, and $x \in \mathbb{R}^n$, let $X \triangleq X_j$ with probability $\alpha_j \in [0,1]$, $\sum_{i=1}^{m} \alpha_j = 1$. Then, the CDF and PDF of the n-dimensional random vector X are given by

$$F(x) = \sum_{j=1}^{m} \alpha_j F_j(x), \quad \varsigma(x) = \sum_{j=1}^{m} \alpha_j \varsigma_j(x). \qquad (2)$$

Proof: $F(x) \triangleq \mathbb{P}(X \leq x) = \sum_{j=1}^{m} \mathbb{P}(X = X_j)\, \mathbb{P}(X_j \leq x)$ $= \sum_{j=1}^{m} \alpha_j F_j(x)$, where we have used the law of total probability. Since each X_j and hence X, is absolutely continuous, we have $\varsigma(x) = \sum_{j=1}^{m} \alpha_j \varsigma_j(x)$. ■

A consequence of Lemma 1 is that the joint state and output PDF of any stochastic jump system, not necessarily linear, is of mixture type, namely a convex sum of component PDFs. Next, we provide a closed form formula of the joint state PDF evolution for SJLS (1), under the assumption that the initial joint PDF is Gaussian.

Theorem 1: (**SJLS joint state PDF at time k**) Consider a discrete-time SJLS $(\{A_i\}_{i=1}^m, \{\pi(k)\}_{k\in\mathbb{N}_0})$ with the initial Gaussian joint state PDF $\varsigma_0 = \mathcal{N}(\mu_0, \Sigma_0)$, where $\mu_0 \in \mathbb{R}^n$, $\Sigma_0 \in \text{Symm}_+$. Then, the joint state PDF at time k, denoted by $\varsigma(k)$, is given by

$$\varsigma(k) = \sum_{j_k=1}^{m} \sum_{j_{k-1}=1}^{m} \cdots \sum_{j_1=1}^{m} \left(\prod_{r=1}^{k} \pi_{j_r}(r) \right) \mathcal{N}\left(A_{j_k}^{*} \mu_0, A_{j_k}^{*} \Sigma_0 A_{j_k}^{*\top}\right), \qquad (3)$$

where $A_{j_k}^{*} \triangleq \prod_{r=k}^{1} A_{j_r} = A_{j_k} A_{j_{k-1}} \ldots A_{j_2} A_{j_1}$.

Proof: Starting from ς_0 at $k = 0$, the *modal PDF* at time $k = 1$, becomes

$$\varsigma_j(1) = \mathcal{N}\left(A_j \mu_0, A_j \Sigma_0 A_j^{\top}\right), \; j = 1, \cdots, m, \qquad (4)$$

From Lemma 1, it follows that the *state PDF* at $k = 1$, is

$$\varsigma(1) = \sum_{j_1=1}^{m} \pi_{j_1}(1)\, \mathcal{N}\left(A_{j_1} \mu_0, A_{j_1} \Sigma_0 A_{j_1}^{\top}\right), \qquad (5)$$

where $\pi_{j_1}(1)$ is the occupation probability for mode j_1 at time $k = 1$. Notice that (5) is an MoG with m component Gaussians. Next, we utilize the fact that linear transformation of an MoG is an equal component MoG with linearly transformed component means and congruently transformed component covariances (see Theorem 6 and Corollary 7 in [19]). This results the modal PDF at $k = 2$ as

$$\varsigma_j(2) = \sum_{j_1=1}^{m} \pi_{j_1}(1)\, \mathcal{N}\big((A_j A_{j_1})\mu_0, (A_j A_{j_1})\Sigma_0 (A_j A_{j_1})^{\top}\big), \qquad (6)$$

for $j=1,\ldots,m$; and consequently

$$\varsigma(2)=\sum_{j_2=1}^{m}\sum_{j_1=1}^{m}\pi_{j_2}(2)\pi_{j_1}(1)\,\mathcal{N}\big((A_{j_2}A_{j_1})\mu_0,\\ (A_{j_2}A_{j_1})\Sigma_0(A_{j_2}A_{j_1})^\top\big). \quad (7)$$

Continuing with this recursion till time k, we arrive at (3), which is an MoG with m^k Gaussian components. ■

Remark 2: From the above proof, it follows that if the initial PDF ς_0, instead of being joint Gaussian, were an m_0 component MoG given by $\varsigma_0=\sum\limits_{j_0=1}^{m_0}\alpha_{j_0}\,\mathcal{N}(\mu_{j_0},\Sigma_{j_0})$, $\sum\limits_{j_0=1}^{m_0}\alpha_{j_0}=1$, then $\varsigma(k)$ would be an MoG with m_0m^k components, given by

$$\varsigma(k)=\sum_{j_k=1}^{m}\sum_{j_{k-1}=1}^{m}\cdots\sum_{j_1=1}^{m}\sum_{j_0=1}^{m_0}\left(\prod_{r=1}^{k}\pi_{j_r}(r)\right)\\ \alpha_{j_0}\mathcal{N}\left(A^*_{j_k}\mu_{j_0},A^*_{j_k}\Sigma_{j_0}A^{*\top}_{j_k}\right). \quad (8)$$

Next, we compute the mean and covariance of *any mixture PDF*, not necessarily MoG, in terms of the means and covariances of its component PDFs.

Lemma 2: (**Mean-covariance of mixture PDF**) Consider any q-component mixture PDF $\varsigma(x)=\sum\limits_{j=1}^{q}\beta_j\varsigma_j(x)$, with $\sum\limits_{j=1}^{q}\beta_j=1$, that has component mean-covariance pairs (μ_j,Σ_j), $j=1,\ldots,q$. Then, the mean-covariance pair $(\mu_{\text{mix}},\Sigma_{\text{mix}})$ for the mixture PDF $\varsigma(x)$, is given by

$$\mu_{\text{mix}}=\sum_{j=1}^{q}\beta_j\mu_j, \quad (9)$$

$$\Sigma_{\text{mix}}=\sum_{j=1}^{q}\beta_j\left(\Sigma_j+(\mu_j-\mu_{\text{mix}})(\mu_j-\mu_{\text{mix}})^\top\right). \quad (10)$$

Proof: By definition, mean vector of the mixture PDF is

$$\mu_{\text{mix}}\triangleq\int_{\mathbb{R}^n}x\varsigma(x)dx=\sum_{j=1}^{q}\beta_j\int_{\mathbb{R}^n}x\varsigma_j(x)dx=\sum_{j=1}^{q}\beta_j\mu_j.$$

Next, covariance matrix of the mixture PDF is

$$\Sigma_{\text{mix}}\triangleq\mathbb{E}\left[(x-\mu_{\text{mix}})(x-\mu_{\text{mix}})^\top\right]=\mathbb{E}\left[xx^\top\right]-\mu_{\text{mix}}\mu_{\text{mix}}^\top\\ =\sum_{j=1}^{q}\beta_j\int_{\mathbb{R}^n}(x-\mu_{\text{mix}}+\mu_{\text{mix}})(x-\mu_{\text{mix}}+\mu_{\text{mix}})^\top\varsigma_j(x)\,dx\\ -\mu_{\text{mix}}\mu_{\text{mix}}^\top\\ =\sum_{j=1}^{q}\beta_j\left(\Sigma_j+(\mu_j-\mu_{\text{mix}})(\mu_j-\mu_{\text{mix}})^\top\right).$$

■

Corollary 2: (**SJLS mean-covariance at time** k) The time evolution of the mean vector $\mu(k)$ for a discrete-time SJLS $(\{A_i\}_{i=1}^m,\{\pi(k)\}_{k\in\mathbb{N}_0})$ with initial Gaussian joint state PDF $\varsigma_0=\mathcal{N}(\mu_0,\Sigma_0)$, is given by

$$\mu(k)=\sum_{j_k=1}^{m}\sum_{j_{k-1}=1}^{m}\cdots\sum_{j_1=1}^{m}\left(\prod_{r=1}^{k}\pi_{j_r}(r)\right)\underbrace{A^*_{j_k}\mu_0}_{\text{component mean}}, \quad (11)$$

and the evolution of the covariance matrix $\Sigma(k)$ is given by

$$\Sigma(k)=\sum_{j_k=1}^{m}\sum_{j_{k-1}=1}^{m}\cdots\sum_{j_1=1}^{m}\left(\prod_{r=1}^{k}\pi_{j_r}(r)\right)\Bigg(\underbrace{A^*_{j_k}\Sigma_0A^{*\top}_{j_k}}_{\text{component covariance}}\\ +\left(A^*_{j_k}\mu_0-\mu(k)\right)\left(A^*_{j_k}\mu_0-\mu(k)\right)^\top\Bigg). \quad (12)$$

Proof: The proof follows by combining Theorem 1 with Lemma 2. ■

III. Limitation of LTI Model Structure for SJLS Statistics Realization

Given the SJLS $(\{A_i\}_{i=1}^m,\{\pi(k)\}_{k\in\mathbb{N}_0})$, consider a candidate linear time invariant (LTI) approximation

$$\widehat{x}(k+1)=\widehat{A}\,\widehat{x}(k). \quad (13)$$

Starting from $\varsigma_0=\mathcal{N}(\mu_0,\Sigma_0)$, we would like to investigate if an LTI map (13) exists that can realize the SJLS mean-covariance sequences given by (11) and (12).

Theorem 3: (**Generic SJLS statistics are not LTI realizable**) Starting from an initial joint state PDF $\varsigma_0=\mathcal{N}(\mu_0,\Sigma_0)$, the mean-covariance sequences generated by a discrete-time SJLS $(\{A_i\}_{i=1}^m,\{\pi(k)\}_{k\in\mathbb{N}_0})$, is LTI realizable iff $\pi(k)$ is independent of k, for all $k\in\mathbb{N}_0$.

Proof: If possible, suppose there exists an LTI map (13), or equivalently a real square matrix $\widehat{A}$, that realizes the SJLS mean-covariance sequences given by (11) and (12). Consider an arbitrary time interval $\Delta t_k\triangleq[t_k,t_{k+1})$ with $k\in\mathbb{N}_0$, over which we claim $\widehat{A}:(\mu(k),\Sigma(k))\mapsto(\mu(k+1),\Sigma(k+1))$. However, being a linear transformation, $\widehat{A}$ needs to satisfy $\mu(k+1)=\widehat{A}\mu(k)$, and $\Sigma(k+1)=\widehat{A}\Sigma(k)\widehat{A}^\top$. Since $A^*_{j_{k+1}}=A_{j_{k+1}}A^*_{j_k}$, and $\prod\limits_{r=1}^{k+1}\pi_{j_r}(r)=\pi_{j_{k+1}}(k+1)\prod\limits_{r=1}^{k}\pi_{j_r}(r)$, hence (11)-(12) yield

$$\mu(k+1)=\sum_{j_{k+1}=1}^{m}\pi_{j_{k+1}}(k+1)A_{j_{k+1}}\sum_{j_k=1}^{m}\cdots\sum_{j_1=1}^{m}\left(\prod_{r=1}^{k}\pi_{j_r}(r)\right)\\ A^*_{j_k}\mu_0, \quad (14)$$

$$\Sigma(k+1)=\left(\sum_{j_{k+1}=1}^{m}\pi_{j_{k+1}}(k+1)A_{j_{k+1}}\right)\Bigg[\sum_{j_k=1}^{m}\sum_{j_{k-1}=1}^{m}\cdots\sum_{j_1=1}^{m}\\ \left(\prod_{r=1}^{k}\pi_{j_r}(r)\right)\Big(A^*_{j_k}\Sigma_0A^{*\top}_{j_k}+\left(A^*_{j_k}\mu_0-\mu(k)\right)\\ \left(A^*_{j_k}\mu_0-\mu(k)\right)^\top\Big)\Bigg]\left(\sum_{j_{k+1}=1}^{m}\pi_{j_{k+1}}(k+1)A_{j_{k+1}}\right)^\top. \quad (15)$$

For non-trivial case $\mu_0 \neq 0$, requiring $\mu(k+1) = \widehat{A}\mu(k)$ from (11) and (14), we arrive at the matrix equation

$$\sum_{j_{k+1}=1}^{m} \pi_{j_{k+1}}(k+1) A_{j_{k+1}} = \widehat{A}, \tag{16}$$

which we recover again by requiring $\Sigma(k+1) = \widehat{A}\Sigma(k)\widehat{A}^\top$ from (12) and (15), for non-trivial case $\Sigma_0 \neq 0$. Notice that for general SJLS $(\{A_i\}_{i=1}^m, \{\pi(k)\}_{k\in\mathbb{N}_0})$, no constant matrix $\widehat{A}$ can satisfy (16) unless $\pi(k)$ is independent of k. Since our choice of interval Δt_k was arbitrary, this conclusion holds for each $k \in \mathbb{N}_0$, that is, (16) holds iff $\{\pi(k)\}_{k\in\mathbb{N}_0}$ is a sequence of constant occupation probability vectors. ■

Corollary 4: (**When average dynamics realizes mean-coavraince**) Mean-covariance sequence generated by an i.i.d. JLS $(\{A_i\}_{i=1}^m, \{\pi(k) = \pi\}_{k\in\mathbb{N}_0})$ can be realized by an LTI map (13) with $\widehat{A} = \sum_{i=1}^{m} \pi_i A_i$. The same for a generic SJLS can be realized by a linear time varying (LTV) map $\widehat{x}(k+1) = \widetilde{A}_k \widehat{x}(k)$ with $\widetilde{A}_k = \sum_{i=1}^{m} \pi_i(k) A_i$, namely by selecting the "instantaneous average dynamics" from the convex polytope of the modal matrices.

Remark 3: (**SJLS mean-covariance recursion**) Comparing (11) with (14), and (12) with (15), we get the recursions

$$\mu(k+1) = \sum_{i=1}^{m} \pi_i(k+1) A_i \mu(k), \tag{17}$$

$$\Sigma(k+1) = \left(\sum_{i=1}^{m} \pi_i(k+1) A_i\right) \Sigma(k) \left(\sum_{i=1}^{m} \pi_i(k+1) A_i\right)^\top \tag{18}$$

IV. Monge-Kantorovich Optimal Transport

A. Background

In this subsection, we provide some background on Monge-Kantorovich optimal transport [20] that will be useful for approximating the SJLS in first two statistical moments. One key aspect of optimal transport theory is the definition of a distance, called *Wasserstein distance*, between two *given* PDFs ρ and $\widehat{\rho}$, that measures the *shape difference* between them by quantifying the minimum amount of work needed to morph one PDF to the other.

Definition 2: (**Wasserstein distance**) The L_2 Wasserstein distance of order 2 (hereafter referred simply as *Wasserstein distance* W), between two d-dimensional random vectors $y \sim \rho$, and $\widehat{y} \sim \widehat{\rho}$, is defined as

$$W(\rho, \widehat{\rho}) \triangleq \left(\inf_{\varrho \in \mathcal{P}_2(\rho,\widehat{\rho})} \mathbb{E}\left[\| y - \widehat{y} \|^2_{\ell_2(\mathbb{R}^d)} \right] \right)^{\frac{1}{2}}, \tag{19}$$

where the $\mathbb{E}[\cdot]$ is taken with respect to the joint PDF $\varrho(y, \widehat{y})$ that makes the cost function achieve the infimum. The symbol $\mathcal{P}_2(\rho, \widehat{\rho})$ denotes the set of all joint PDFs supported over $\mathbb{R}^{2d}$, having finite second moments, whose first marginal is ρ, and second marginal is $\widehat{\rho}$.

Remark 4: It can be shown [21] that W defines a metric on the manifold of PDFs, and remains well defined between the *distributions* even though the random vectors y and $\widehat{y}$ are not absolutely continuous (i.e. ρ and $\widehat{\rho}$ don't exist).

B. Optimal Transport Map a.k.a. Brenier Map

Definition 3: (**Optimal transport map**) The optimal transport map $\beta : \mathbb{R}^d \mapsto \mathbb{R}^d$ associated with (19), that satisfies $y = \beta(\widehat{y})$, is defined as

$$\beta^\star \triangleq \underset{\beta(\cdot)}{\text{arginf}} \int_{\mathbb{R}^d} \| \beta(\widehat{y}) - \widehat{y} \|^2_{\ell_2(\mathbb{R}^d)} \ \widehat{\rho}(\widehat{y}) \ d\widehat{y}, \quad \text{subject to } \rho = \beta \sharp \widehat{\rho}. \tag{20}$$

Remark 5: In (20), the optimization takes place over all push-forward maps with the specified PDFs as boundary conditions. Since there are infinite ways to morph a PDF to another, finding a push-forward or transport map $\beta(\cdot)$ is underdetermined unless we require that the push-forward is optimal in some sense. Thus, (20) finds that particular push-forward which entails minimum amount of work. Notice that the infimum value for (20) is W^2, given by (19).

Remark 6: It is known [22] that the existence and uniqueness of $\beta^\star(\cdot)$ are guaranteed. The optimal transport map $\beta^\star$ is also known as the *Brenier map*.

Theorem 5: (**Brenier map for Gaussian to Gaussian transport**) [23], [24] The optimal transport map $\beta^\star$ between two Gaussian random vectors $y \sim \mathcal{N}(\nu, S)$, and $\widehat{y} \sim \mathcal{N}\left(\widehat{\nu}, \widehat{S}\right)$, is an affine transformation $y = \Gamma\widehat{y} + \gamma$, where

$$\Gamma = \sqrt{S}\left(\sqrt{S}\,\widehat{S}\,\sqrt{S}\right)^{-\frac{1}{2}} \sqrt{S}, \tag{21}$$

$$\gamma = \nu - \widehat{\nu}. \tag{22}$$

Remark 7: Theorem 5 implies that the minimum effort way to morph a Gaussian PDF to another, is via a translation, rotation and scaling. This is intuitive if we think about the geometric way to morph a given ellipsoid to another such as to entail minimum work.

V. Gaussian Moment Closure for Stochastic Jump Linear Systems via Optimal Transport

Now we will derive a Gaussian moment closure approximation for a discrete-time SJLS $(\{A_i\}_{i=1}^m, \{\pi(k)\}_{k\in\mathbb{N}_0})$ with the initial Gaussian PDF $\varsigma_0 = \mathcal{N}(\mu_0, \Sigma_0)$. From Theorem 1, we know that at time k, the SJLS joint state PDF $\varsigma(k)$ is an MoG with m^k component Gaussians. To perform Gaussian moment closure, we seek an *approximate model* of the SJLS that will generate state PDF $\widehat{\varsigma}(k) = \mathcal{N}\left(\widehat{\mu}(k), \widehat{\Sigma}(k)\right)$ such that $\widehat{\mu}(k) = \mu(k)$, and $\widehat{\Sigma}(k) = \Sigma(k)$, where $\mu(k)$ and $\Sigma(k)$ are the mean and covariance of the SJLS at time k, and are given by Corollary 2. Hence, over any time interval $[t_k, t_{k+1})$, $k \in \mathbb{N}_0$, the approximated dynamical system is required to transport between specified Gaussian random vectors $\widehat{x}(k) \sim \mathcal{N}\left(\widehat{\mu}(k), \widehat{\Sigma}(k)\right)$ and $\widehat{x}(k+1) \sim \mathcal{N}\left(\widehat{\mu}(k+1), \widehat{\Sigma}(k+1)\right)$, where $\widehat{x} \in \mathbb{R}^n$ denotes the state vector of the approximated dynamics. In other words, we seek the transport map β_k such that $\widehat{x}(k+1) = \beta_k(\widehat{x}(k))$.

Theorem 6: (**Gaussian moment closure approximation for SJLS guaranteeing optimal transport is an ATV system**) For the Gaussian moment closure approximation of a discrete-time SJLS $(\{A_i\}_{i=1}^m, \{\pi(k)\}_{k\in\mathbb{N}_0})$ with initial

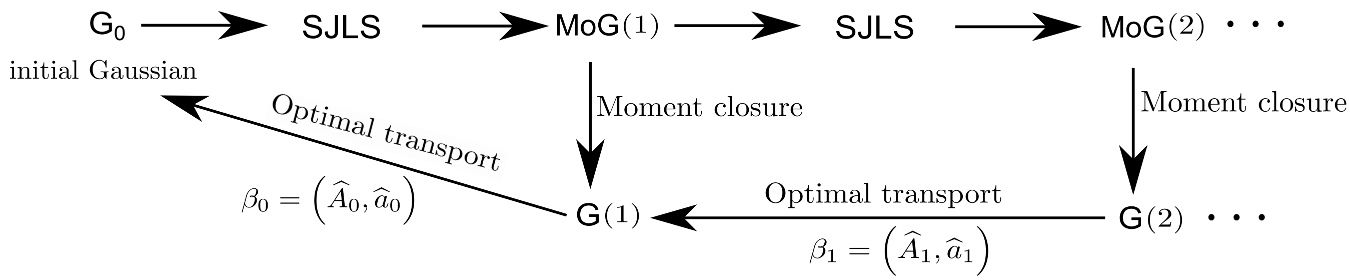

Fig. 1. The schematic of the proposed Gaussian moment closure approximation for SJLS.

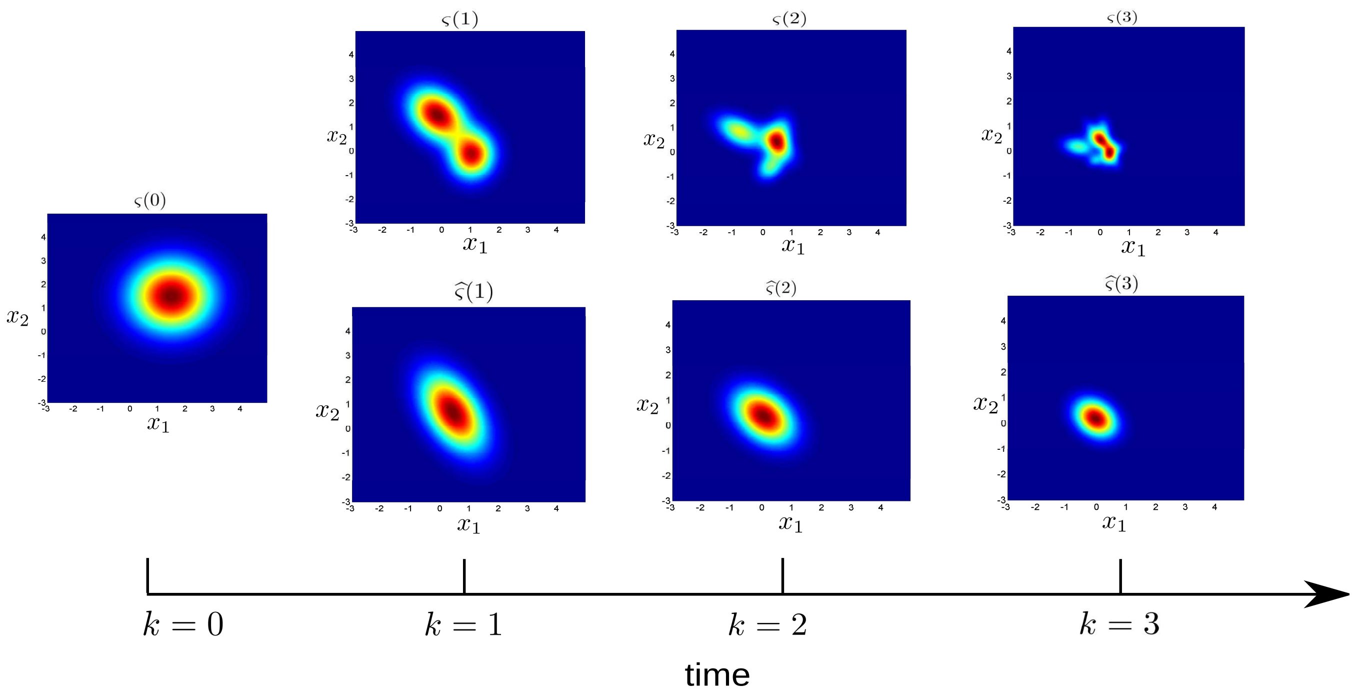

Fig. 2. Comparison of MJLS state PDF $\varsigma(k)$ with Gaussian moment closure PDF $\widehat{\varsigma}(k)$ from ATV approximated dynamics. Notice that $\varsigma(1)$, $\varsigma(2)$ and $\varsigma(3)$ are MoG PDFs with 2, 4 and 8 Gaussian components, respectively. On the other hand, $\widehat{\varsigma}(1)$, $\widehat{\varsigma}(2)$ and $\widehat{\varsigma}(3)$ are all Gaussian PDFs, whose mean vector and covariance matrix coincide with their respective MoG counterparts $\varsigma(1)$, $\varsigma(2)$ and $\varsigma(3)$, respectively.

Gaussian PDF $\varsigma_0 = \mathcal{N}(\mu_0, \Sigma_0)$, the optimal transport map $\widehat{x}(k+1) = \beta_k^{\star}(\widehat{x}(k))$ that solves

$$\underset{\beta(\cdot)}{\text{arginf}}\; \mathbb{E}\left[\| \,\beta\left(\widehat{x}(k)\right) - \widehat{x}(k)\, \|^2_{\ell_2(\mathbb{R}^n)}\right], \tag{23}$$

subject to $\mathcal{N}\left(\widehat{\mu}(k+1), \widehat{\Sigma}(k+1)\right) = \beta \sharp \mathcal{N}\left(\widehat{\mu}(k), \widehat{\Sigma}(k)\right)$, is given by the ATV dynamics

$$\widehat{x}(k+1) = \widehat{A}_k \widehat{x}(k) \;+\; \widehat{a}_k, \tag{24}$$

where

$$\widehat{A}_k \;=\; \sqrt{\Sigma_{k+1}}\left(\sqrt{\Sigma_{k+1}}\,\Sigma_k\,\sqrt{\Sigma_{k+1}}\right)^{-\frac{1}{2}}\sqrt{\Sigma_{k+1}}, \tag{25}$$

$$\widehat{a}_k \;=\; \mu_{k+1} - \mu_k, \tag{26}$$

and $\mu(k)$ and $\Sigma(k)$ are given by (11) and (12). In other words, $\beta_k^{\star}$ is characterized by the matrix-vector pair $\left(\widehat{A}_k, \widehat{a}_k\right)$.

Proof: This is a direct consequence of Theorem 5. ■

Remark 8: Theorem 6 implies that the discrete-time SJLS $\left(\{A_i\}_{i=1}^m, \{\pi(k)\}_{k\in\mathbb{N}_0}\right)$ and its ATV approximation (24), with initial Gaussian PDF $\varsigma_0 = \mathcal{N}(\mu_0, \Sigma_0)$, generate identical mean and covariances, by construction, i.e. $\widehat{\mu}(k) = \mu(k)$, and $\widehat{\Sigma}(k) = \Sigma(k)$. Yet, their PDFs are not same since $\varsigma(k)$ is an MoG with m^k component Gaussians; and $\widehat{\varsigma}(k)$ is a Gaussian with its mean and covariance same as that of the SJLS MoG at that time.

Remark 9: The algorithmic construction of the ATV moment closure approximation is depicted in Fig. 1.

VI. Numerical Example

To illustrate the ideas presented so far, consider Gaussian moment closure approximation for a discrete-time Markov jump linear system (MJLS) with 2 modes, given by $\left(\{A_i\}_{i=1}^2, \{\pi(k)\}_{k\in\mathbb{N}_0}\right)$ where the modal matrices are

$$A_1 = \begin{bmatrix} 0.4 & -0.5 \\ 0.4 & 0.6 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.3 & 0.4 \\ -0.5 & 0.4 \end{bmatrix}, \tag{27}$$

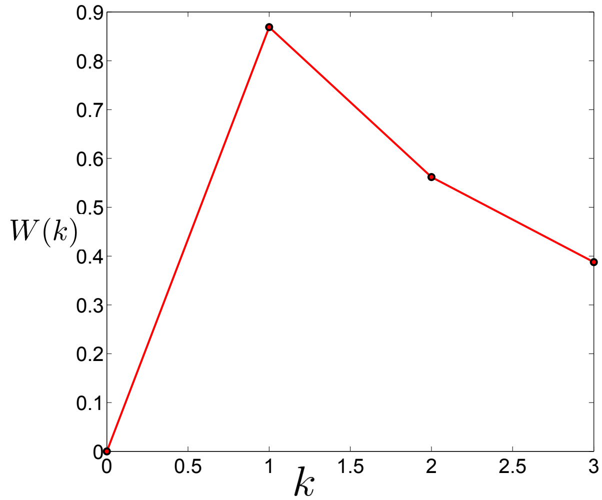

Fig. 3. Time history of the Wasserstein distance W between $\varsigma(k)$ and $\widehat{\varsigma}(k)$, computed by solving an LP at each fixed k. We refer the readers to [13], [14] for details of this computation.

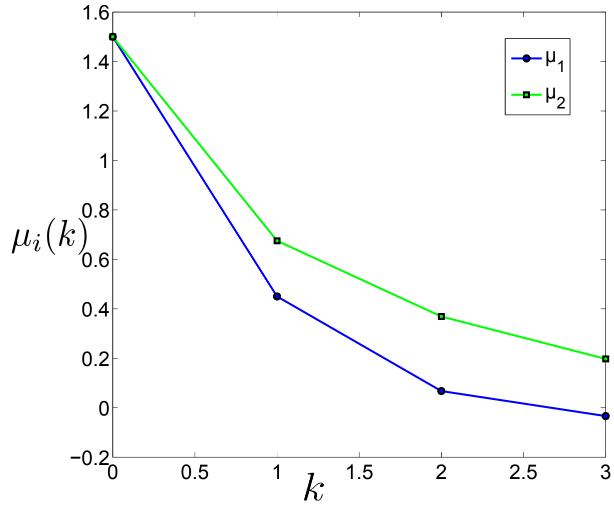

Fig. 4. Time history of the components of the matched mean vector for the MJLS and its Gaussian moment closure approximation.

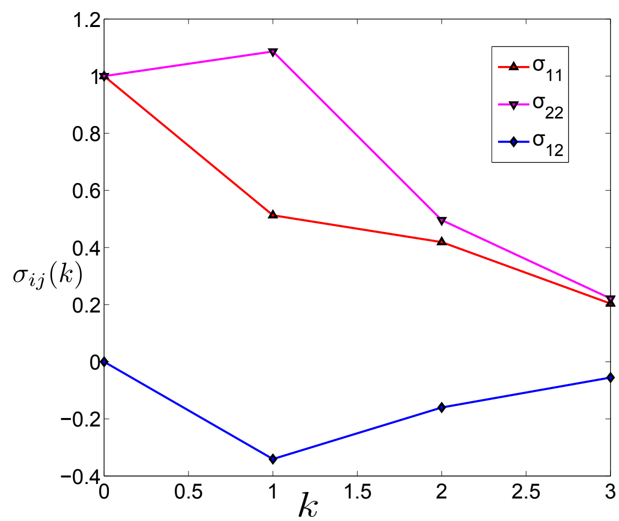

Fig. 5. Time history of the components of the matched covariance matrix for the MJLS and its Gaussian moment closure approximation.

and the occupation probability vector $\pi(k)$ comes from a discrete-time Markov chain $\pi(k+1) = \pi(k)P$, where the transition probability matrix $P = \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix}$. For these choice of parameters, it can be verified [25] that the MJLS is mean-square stable. Taking the initial joint PDF $\varsigma_0 \triangleq \varsigma(0) = \mathcal{N}(\mu_0, \Sigma_0)$ with $\mu_0 = \{1.5, 1.5\}^\top$ and $\Sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and the initial occupation probability vector to be $\pi(0) = \{0.5, 0.5\}$, we plot the MJLS state PDF $\varsigma(k)$ and Gaussian moment closure PDF $\widehat{\varsigma}(k)$ from ATV approximated dynamics, in Fig. 2. From this figure, it is evident that the PDFs $\varsigma(k)$ and $\widehat{\varsigma}(k)$ may have significant mismatch in shapes representing differing trajectory concentrations, although their first two moments match.

In Fig. 3, we plot the Wasserstein distance W between $\varsigma(k)$ and $\widehat{\varsigma}(k)$. This plot provides the quantitative evidence of joint PDF mismatch between the true MJLS and its Gaussian moment closure approximation. Time histories of the components of the *matched* mean vector and covariance matrix are plotted in Fig. 4 and 5, respectively.

VII. Conclusions

In this paper, we considered Gaussian moment closure approximation for discrete-time stochastic jump linear systems. Using optimal transport ideas, we have constructed an affine time-varying dynamical system that matches the time-varying mean and covariance of the stochastic jump linear system. One contribution of this paper is to construct a non-jump moment closure approximation for a stochastic jump system, which is new compared to existing literature. Another contribution is to demonstrate that for stochastic jump linear systems, matching first two moments need not match PDF shapes. The example provided in this paper, is hoped to have pedagogical value from this perspective.

References

[1] M. Mariton, *Jump Linear Systems in Automatic Control*. CRC Press; 1990.

[2] L. Xiao, A. Hassibi, and J.P. How, "Control with Random Communication Delays via A Discrete-time Jump System Approach". *Proceedings of the 2000 American Control Conference*, Vol. 3, pp. 2199–2204, 2000.

[3] J.P. Hespanha, *Stochastic Hybrid Systems: Application to Communication Networks*. Hybrid systems: Computation and Control, Lecture Notes in Computer Science, Springer; pp. 387–401, 2004.

[4] A.G. Kofman, R. Zaibel, A.M. Levine, and Y. Prior, "Non-Markovian Stochastic Jump Processes in Nonlinear Optics". *Physical Review Letters*, Vol. 61, No. 3, pp. 251–254, 1988.

[5] Y. Bar-Shalom, and X.R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS; 1995.

[6] R. Akella, and P.R. Kumar, "Optimal Control of Production Rate in A Failure Prone Manufacturing System". *IEEE Transactions on Automatic Control*, Vol. 31, No. 2, pp. 116–126, 1986.

[7] W.S. Gray, and O. Gonzalez, "Modeling Electromagnetic Disturbances in Closed-loop Computer Controlled Flight Systems". *Proceedings of the 1998 American Control Conference*, Vol. 1, pp. 359–364, 1998.

[8] G. Ackerson, and K.S. Fu, "On State Estimation in Switching Environments". *IEEE Transactions on Automatic Control*, Vol. 15, No. 1, pp. 10–17, 1970.

[9] H.A.P. Blom, and Y. Bar-Shalom, "The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients". *IEEE Transactions on Automatic Control*, Vol. 33, No. 8, pp. 780–783, 1988.

[10] A.A. Julius, A. Girard, and G.J. Pappas, "Approximate Bisimulation for A Class of Stochastic Hybrid Systems". *Proceedings of the 2006 American Control Conference*, pp. 4724–4729, 2006.

[11] L. Zhang, B. Huang, and L. Zames, "$\mathcal{H}_\infty$ Model Reduction of Markovian Jump Linear Systems". *Systems & Control Letters*, Vol. 50, No. 2, pp. 103–118, 2003.

[12] G. Kotsalis, and A. Rantzer, "Balanced Truncation for Discrete Time Markov Jump Linear Systems". *IEEE Transactions on Automatic Control*, Vol. 55, No. 11, pp. 2606-2611, 2010.

[13] A. Halder, and R. Bhattacharya, "Model Validation: A Probabilistic Formulation", *50th IEEE Conference on Decision and Control and European Control Conference*, pp. 1692–1697, 2011.

[14] A. Halder, and R. Bhattacharya, "Further Results on Probabilistic Model Validation in Wasserstein Metric", *51st IEEE Conference on Decision and Control*, pp. 5542–5547, 2012.

[15] A. Halder, and R. Bhattacharya, "Frequency Domain Model Validation in Wasserstein Metric", *Proceedings of the 2013 American Control Conference*, pp. 5845–5850, 2013.

[16] P. Whittle, "On the Use of the Normal Approximation in the Treatment of Stochastic Processes". *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 268–281, 1957.

[17] I. Nåsell, "An Extension of the Moment Closure Method". *Theoretical Population Biology*, Vol. 64, No. 2, pp. 233–239, 2003.

[18] A. Singh, and J.P. Hespanha, "Moment Closure Techniques for Stochastic Models in Population Biology", *American Control Conference*, 2006.

[19] S. Ali-Löytty, "On the Convergence of the Gaussian Mixture Filter". *Tampereen teknillinen yliopisto*, Matematiikan laitos. Tutkimusraportti, 89, 2008.

[20] C. Villani, *Topics in Optimal Transportation*. Graduate Studies in Mathematics, First ed., American Mathematical Society; 2003.

[21] S.T. Rachev, *Probability Metrics and the Stability of Stochastic Models*. John Wiley, First ed., 1991.

[22] Y. Brenier, "Polar Factorization and Monotone Rearrangement of Vector-valued Functions". *Communications on Pure and Applied Mathematics*, Vol. 44, No. 4, pp. 375–417, 1991.

[23] I. Olkin, and F. Pukelsheim, "The Distance Between Two Random Vectors with Given Dispersion Matrices". *Linear Algebra and its Applications*, Vol. 48, pp. 257–263, 1982.

[24] M. Knott, and C.S. Smith, "On the Optimal Mapping of Distributions". *Journal of Optimization Theory and Applications*, Vol. 43, No. 1, pp. 39–49, 1984.

[25] O.D.V. Costa, M.D. Fragoso, and R.P. Marques, *Discrete-time Markov Jump Linear Systems*. Springer; 2006.