Estimation of Infinitesimal Generators for Unknown Stochastic Hybrid Systems via Sampling: A Formal Approach

Ameneh Nejati, *Student Member, IEEE*, Abolfazl Lavaei, Sadegh Soudjani, *Members, IEEE*, and Majid Zamani, *Senior Member, IEEE*

***Abstract*—In this work, we develop a data-driven framework with formal confidence bounds for the estimation of infinitesimal generators for continuous-time stochastic hybrid systems with unknown dynamics. The proposed approximation scheme employs both time discretization and sampling from the solution process, and estimates the infinitesimal generator of the solution process via a set of data collected from trajectories of systems. We assume some mild continuity assumptions on the dynamics of the system and quantify the closeness between the infinitesimal generator and its approximation while ensuring an a-priori guaranteed confidence bound. To provide a reasonable closeness precision, we discuss significant roles of both time discretization and number of data in our approximation scheme. In particular, for a fixed number of data, variance of the estimation converges to infinity when the time discretization goes to zero. The proposed approximation framework guides us how to jointly select a suitable data size and a time discretization parameter to cope with this counter-intuitive behavior. We demonstrate the effectiveness of our proposed results by applying them to a *nonlinear* jet engine compressor with *unknown* dynamics.**

***Index Terms*—Data-Driven Estimation, Infinitesimal Generators, Unknown Stochastic Hybrid Systems.**

I. Introduction

INFINITESIMAL generator of a continuous-time stochastic process is a partial differential operator that encodes large amounts of information about the stochastic process. In particular, infinitesimal generator plays a significant role in the analysis of continuous-time stochastic systems including (i) stability verification and controller synthesis via (control) Lyapunov functions (*e.g.,* [1]); (ii) input-to-state stability (ISS) property of continuous-time stochastic systems (*e.g.,* [2]); (iii) establishing similarity relations between two continuous-time stochastic systems via stochastic simulation functions (*e.g.,* [3], [4]); (iv) incremental stability of continuous-time stochastic control systems (*e.g.,* [5]), and (v) safety verification and controller synthesis of continuous-time stochastic systems via barrier certificates (*e.g.,* [6], [7], [8]), to name a few. Hence, computing the infinitesimal generator is a crucial step in developing an analysis framework for continuous-time stochastic systems.

This work was supported in part by the NSF under Grant CNS-2145184 and by the UKRI EPSRC New Investigator Award CodeCPS (EP/V043676/1).
A. Nejati is with Department of Electrical Engineering, Technical University of Munich, Germany. A. Lavaei is with Institute for Dynamic Systems and Control, ETH Zurich, Switzerland. S. Soudjani is with School of Computing, Newcastle University, United Kingdom. Majid Zamani is with Department of Computer Science, University of Colorado Boulder, USA and LMU Munich, Germany. Email: amy.nejati@tum.de, alavaei@ethz.ch, sadegh.soudjani@newcastle.ac.uk, majid.zamani@colorado.edu.

In general, closed-form models for many (continuous-time) physical systems are either not available or too complex to be of any practical use. Hence, one cannot utilize model-based techniques to analyze many real-world applications. Although there have been some results on model identification techniques (a.k.a., *indirect* data-driven approaches) to first learn approximate models and then employ model-based techniques over them, (see *e.g.,* [9, and references herein]), acquiring an accurate model for complex systems (if not impossible) can be very complex, time-consuming, and expensive. These difficulties motivate a need to develop *direct data-driven approaches* to bypass the model identification phase and directly employ system measurements for the analysis.

Contributions. Our main contribution is to develop a data-driven scheme for the formal estimation of infinitesimal generators of continuous-time stochastic *hybrid* systems with unknown dynamics. In our proposed setting, we first approximate the infinitesimal generator of the stochastic system via a set of data collected from its trajectories. We then provide a formal framework to compute the probabilistic error between the approximated infinitesimal generator and the exact one corresponding to unknown dynamics with an a-priori guaranteed confidence bound. We discuss that both the sampling time and the number of data are essential to provide a reasonable closeness precision. We illustrate our data-driven results over a jet engine compressor with unknown dynamics.

Related Works. A limited subset of the provided results in this work has been presented in [10]. Our approach here differs from the one in [10] in three main directions. First, we enlarge underlying dynamics to a class of stochastic *hybrid* systems by adding Poisson processes to the dynamics, while the results in [10] only deal with stochastic systems with Brownian motions. The class of hybrid systems studied in our work has been widely used in the literature (see [1], [11], [12]). Second, we propose a data-driven scheme for the estimation of infinitesimal generators of stochastic hybrid systems with *control inputs*, while the results in [10] only deal with stochastic *autonomous* systems. Third, we apply our results to a *nonlinear* jet engine to show the applicability of our techniques to stochastic hybrid systems with nonlinear dynamics. In addition, we provide proofs of all statements here, some of which were omitted in [10].

It is worth mentioning that although the proposed results in [13] also estimate the infinitesimal generator of stochastic processes, they are based on the assumption of knowing the precise model of the system. To the best of our knowledge, our work is the first to propose a data-driven framework for estimating the infinitesimal generators of stochastic *hybrid* systems with *unknown dynamics* while providing error bounds.

II. Continuous-Time Stochastic Hybrid Systems

A. Notation and Preliminaries

We denote the set of nonnegative and positive integers by $\mathbb{N} := \{0, 1, 2, \ldots\}$ and $\mathbb{N}_{\geq 1} := \{1, 2, 3, \ldots\}$, respectively. Symbols $\mathbb{R}$, $\mathbb{R}_{>0}$, and $\mathbb{R}_{\geq 0}$ denote the set of real, positive and nonnegative real numbers, respectively. We employ $x = [x_1; \ldots; x_N]$ to denote the corresponding vector of dimension $\sum_i n_i$, given N vectors $x_i \in \mathbb{R}^{n_i}$, $n_i \in \mathbb{N}_{\geq 1}$, and $i \in \{1, \ldots, N\}$. Given a matrix $A \in \mathbb{R}^{N \times N}$ with diagonal entries $a_1, \ldots, a_N$, we define $\mathsf{Tr}(A) = \sum_{i=1}^{N} a_i$. Given any $a \in \mathbb{R}$, $|a|$ denotes the absolute value of a. We denote by $\|x\|$ the 2-norm of any row or column vector x. We also denote by $\|A\|_F := \sqrt{\mathsf{Tr}(A^T A)}$ the Frobenius norm of any matrix $A \in \mathbb{R}^{m \times n}$.

We consider a probability space $(\Omega, \mathcal{F}_\Omega, \mathbb{P}_\Omega)$, where Ω is the sample space, $\mathcal{F}_\Omega$ is a sigma-algebra on Ω, and $\mathbb{P}_\Omega$ is a probability measure. We assume that triple $(\Omega, \mathcal{F}_\Omega, \mathbb{P}_\Omega)$ denotes a probability space endowed with a filtration $\mathbb{F} = (\mathcal{F}_s)_{s \geq 0}$ satisfying the usual conditions of completeness and right continuity. Let $(\mathbb{W}_s)_{s \geq 0}$ be a b-dimensional $\mathbb{F}$-Brownian motion, and $(\mathbb{P}_s)_{s \geq 0}$ be an r-dimensional $\mathbb{F}$-Poisson process. The Poisson process $\mathbb{P}_s = [\mathbb{P}_s^1; \cdots; \mathbb{P}_s^{\mathsf{r}}]$ models r events whose occurrences are assumed to be independent of each other.

B. Continuous-Time Stochastic Hybrid Systems

Definition 2.1: A continuous-time stochastic hybrid system (ct-SHS) in this work is characterized by the tuple

$$\Sigma = (X, U, \mathcal{U}, f, \sigma, \rho), \tag{1}$$

where:

- $X \subseteq \mathbb{R}^n$ is the state set of the system;
- $U \subseteq \mathbb{R}^m$ is the input set of the system;
- $\mathcal{U}$ is a subset of sets of $\mathbb{F}$-progressively measurable processes taking values in $\mathbb{R}^m$;
- $f : X \times U \to \mathbb{R}^n$ is the drift term;
- $\sigma : \mathbb{R}^n \to \mathbb{R}^{n \times \mathsf{b}}$ is the diffusion term;
- $\rho : \mathbb{R}^n \to \mathbb{R}^{n \times \mathsf{r}}$ is the reset term.

A continuous-time stochastic hybrid system Σ satisfies

$$\Sigma \colon \mathsf{d}x(t) = f(x(t), u(t))\mathsf{d}t + \sigma(x(t))\mathsf{d}\mathbb{W}_t + \rho(x(t))\mathsf{d}\mathbb{P}_t, \tag{2}$$

$\mathbb{P}$-almost surely ($\mathbb{P}$-a.s.) for any $u \in \mathcal{U}$, where the stochastic process $x : \Omega \times \mathbb{R}_{\geq 0} \to X$ is called the *solution process* of Σ. Here, we assume that Poisson processes $\mathbb{P}_s^z$, for any $z \in \{1, \ldots, \mathsf{r}\}$, have rates λ_z.

To ensure the existence, uniqueness, and strong Markov property of the solution process [14], we assume that the drift, diffusion, and reset terms are all globally Lipschitz continuous (cf. Assumption 1). We perform our analysis over X and U which are assumed to be compact subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. This is motivated by boundedness assumptions required for our theoretical results (cf. Assumption 2).

In the sequel, we call $f(x, u)$ as *infinitesimal mean* and define $\mathsf{c}(x) := \sigma(x)\sigma(x)^T$ as the *infinitesimal covariance*. We now formally present the infinitesimal generator of the stochastic process in the following definition [15].

Definition 2.2: The *infinitesimal generator* $\mathcal{L}$ of the process $x(t)$ acting on a twice continuously-differentiable function $\mathcal{V} : X \to \mathbb{R}$ is defined as

$$\begin{aligned}\mathcal{L}\mathcal{V}(x) = \; & \partial_x \mathcal{V}(x) f(x, u) + \frac{1}{2}\mathsf{Tr}(\mathsf{c}(x)\partial_{x,x}\mathcal{V}(x)) \\ & + \sum_{j=1}^{\mathsf{r}} \lambda_j \big(\mathcal{V}(x + \rho(x)\mathsf{e}_j^{\mathsf{r}}) - \mathcal{V}(x)\big),\end{aligned} \tag{3}$$

where $\partial_x \mathcal{V}(x) = \big[\frac{\partial \mathcal{V}(x)}{\partial x_i}\big]_i$ is a row vector, $\partial_{x,x}\mathcal{V}(x) = \big[\frac{\partial^2 \mathcal{V}(x)}{\partial x_i \partial x_j}\big]_{i,j}$, λ_j is the rate of Poisson process, and $\mathsf{e}_j^{\mathsf{r}}$ is an r-dimensional vector with 1 on the j-th entry and 0 elsewhere.

Infinitesimal generator of a stochastic process can be leveraged to compute the expected value of any function of the solution process (*i.e.,* $\mathcal{V}(x(\tau))$) via Dynkin's formula [16] as:

$$\mathbb{E}_x\Big[\mathcal{V}(x(\tau))\Big] = \mathcal{V}(x(0)) + \mathbb{E}_x\Big[\int_0^\tau \mathcal{L}\mathcal{V}(x(t))\mathsf{d}t\Big], \tag{4}$$

for all $x(0) \in X$, where $\mathbb{E}_x$ is the expected value conditioned on $x(0)$.

In this work, we assume drift, diffusion, and reset terms f, σ, ρ in (1) are all *unknown*. In order to provide a formal framework for the estimation of the infinitesimal generator in (3), we first approximate the infinitesimal generator $\mathcal{L}$ by

$$\widehat{\mathcal{L}}_1\mathcal{V}(x) := \frac{\mathbb{E}_x\Big[\mathcal{V}(x_\tau)\Big] - \mathcal{V}(x)}{\tau}, \quad \forall x \in X, \tag{5}$$

where x_τ denotes the value of the solution process at time τ starting from an initial condition x.

Since there is no closed-form solution for the expected value in (5), one cannot directly utilize (5) as the approximation of the infinitesimal generator. Let $(x_\tau^i)_{i=1}^N$ be N independent and identically distributed (i.i.d.) sampled data by extracting N solution processes $x_\tau^i, i \in \{1, \ldots, N\}$, at time τ from the same initial condition under N different independent noise realizations. We now employ an empirical approximation of the expected value and propose another layer of approximation for the infinitesimal generator $\mathcal{L}$ as

$$\widehat{\mathcal{L}}_2\mathcal{V}(x) := \frac{\frac{1}{N}\sum_{i=1}^N \mathcal{V}(x_\tau^i) - \mathcal{V}(x)}{\tau}, \quad \forall x \in X. \tag{6}$$

We now formalize the main problem that we aim to solve.

Problem 2.3: Provide a formal framework to quantify $\delta \in \mathbb{R}_{\geq 0}$ as the distance between the infinitesimal generator of stochastic process in (3) and its data-driven approximation in (6) with a given a-priori confidence $\beta \in (0, 1]$ as

$$\mathbb{P}\Big\{|\widehat{\mathcal{L}}_2\mathcal{V}(x) - \mathcal{L}\mathcal{V}(x)| \leq \delta\Big\} \geq 1 - \beta, \quad \forall x \in X. \tag{7}$$

Remark 2.4: Note that the empirical approximation in (6) can be utilized for scenarios in which the infinitesimal generator needs to be computed for finitely-many initial conditions. Examples of such scenarios include safety verification and synthesis of stochastic hybrid systems similar to [17] or construction of finite Markov decision processes and establishing similarity relations between two stochastic systems via stochastic simulation functions as in [4], where dynamics of underlying systems are *unknown*.

It should be noted that the confidence $1-\beta$ in (7) is due to the data-driven nature of our proposed estimation procedure. This type of guarantee is very similar to the one provided by Chernoff bound in statistical model checking (see [18, Section 9]).

III. Data-Driven Framework

In our proposed setting, we first quantify the formal closeness between $\mathcal{L}\mathcal{V}(x)$ and its first approximation $\widehat{\mathcal{L}}_1\mathcal{V}(x)$ as in (5). We then quantify the distance between $\widehat{\mathcal{L}}_1\mathcal{V}(x)$ and its empirical approximation $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ as in (6). We finally propose our solution for the closeness quantification between $\mathcal{L}\mathcal{V}(x)$ and $\widehat{\mathcal{L}}_2\mathcal{V}(x)$. To do so, we first need to raise the following two assumptions.

Assumption 1: Suppose $f, \sigma, \rho, u(t), \mathsf{c}(x), \mathcal{V}(x), \partial_x\mathcal{V}(x)$, and $\partial_{x,x}\mathcal{V}(x)$ are all Lipschitz continuous with, respectively, Lipschitz constants $\mathscr{L}_f, \mathscr{L}_u, \mathscr{L}_\sigma, \mathscr{L}_\rho, \bar{\mathscr{L}}_u, \mathscr{L}_\mathsf{c}, \mathscr{L}_\mathcal{V}, \mathscr{L}_{\mathcal{V}_1}, \mathscr{L}_{\mathcal{V}_2} \in \mathbb{R}_{\geq 0}$ as the following, $\forall x, x' \in X$, $\forall u, u' \in U$, $\forall t, t' \in \mathbb{R}_{\geq 0}$:

$$\|f(x,u)-f(x',u')\| \leq \mathscr{L}_f\|x-x'\|+\mathscr{L}_u\|u-u'\|,$$
$$\|\sigma(x)-\sigma(x')\|_F\leq\mathscr{L}_\sigma\|x-x'\|, \|\rho(x)-\rho(x')\|_F\leq\mathscr{L}_\rho\|x-x'\|,$$
$$\|u(t)-u(t')\|\leq\bar{\mathscr{L}}_u|t-t'|, \|\mathsf{c}(x)-\mathsf{c}(x')\|_F \leq \mathscr{L}_\mathsf{c}\|x-x'\|,$$
$$|\mathcal{V}(x)-\mathcal{V}(x')| \leq \mathscr{L}_\mathcal{V}\|x-x'\|,$$
$$\|\partial_x\mathcal{V}(x)-\partial_{x'}\mathcal{V}(x')\| \leq \mathscr{L}_{\mathcal{V}_1}\|x-x'\|,$$
$$\|\partial_{x,x}\mathcal{V}(x)-\partial_{x',x'}\mathcal{V}(x')\|_F \leq \mathscr{L}_{\mathcal{V}_2}\|x-x'\|.$$

Assumption 2: Suppose $f(x)$, $\mathsf{c}(x)$, $\sigma(x)$, $\rho(x)$, $\mathcal{V}(x)$, $\partial_x\mathcal{V}(x)$, and $\partial_{x,x}\mathcal{V}(x)$ are all bounded with constants $\mathcal{B}_f$, $\mathcal{B}_\mathsf{c}, \mathcal{B}_\sigma, \mathcal{B}_\rho, \mathcal{B}_\mathcal{V}, \mathcal{B}_{\mathcal{V}_1}, \mathcal{B}_{\mathcal{V}_2} \in \mathbb{R}_{\geq 0}$ as, $\forall x \in X, \forall u \in U$:

$$\|f(x,u)\|\leq\mathcal{B}_f, \|\mathsf{c}(x)\|_F\leq\mathcal{B}_\mathsf{c}, \|\sigma(x)\|_F\leq\mathcal{B}_\sigma, \|\rho(x)\|_F\leq\mathcal{B}_\rho,$$
$$|\mathcal{V}(x)| \leq \mathcal{B}_\mathcal{V},\ \|\partial_x\mathcal{V}(x)\| \leq \mathcal{B}_{\mathcal{V}_1},\ \|\partial_{x,x}\mathcal{V}(x)\|_F \leq \mathcal{B}_{\mathcal{V}_2}.$$

By leveraging Assumptions 1-2, we propose next result showing that $\mathcal{L}\mathcal{V}(x)$ is also Lipschitz continuous.

Theorem 3.1: Under Assumptions 1-2, $\mathcal{L}\mathcal{V}(x)$ is Lipschitz continuous with Lipschitz constants $\mathscr{L}_1, \mathscr{L}_2\in\mathbb{R}_{\geq 0}$:

$$|\mathcal{L}\mathcal{V}(x)-\mathcal{L}\mathcal{V}(x')| \leq \mathscr{L}_1\|x-x'\|+\mathscr{L}_2\|u-u'\|,$$

for all $x, x' \in X$ and all $u, u' \in U$, where

$$\mathscr{L}_1 = \mathcal{B}_{\mathcal{V}_1}\mathscr{L}_f + \mathcal{B}_f\mathscr{L}_{\mathcal{V}_1} + \frac{1}{2}(\mathscr{L}_\mathsf{c}\mathcal{B}_{\mathcal{V}_2} + \mathcal{B}_\mathsf{c}\mathscr{L}_{\mathcal{V}_2}) + \sum_{j=1}^{\mathsf{r}}\lambda_j(2\mathscr{L}_\mathcal{V}+\mathscr{L}_\mathcal{V}\mathscr{L}_\rho), \quad \mathscr{L}_2=\mathcal{B}_{\mathcal{V}_1}\mathscr{L}_u.$$

Proof: Using the definition of $\mathcal{L}\mathcal{V}(x)$ in (3), we have

$$\begin{aligned}|\mathcal{L}\mathcal{V}(x)-\mathcal{L}\mathcal{V}(x')| &\leq |\partial_x\mathcal{V}(x)f(x,u)-\partial_{x'}\mathcal{V}(x')f(x',u')|\\ &+|\frac{1}{2}\mathsf{Tr}\big(\mathsf{c}(x)\partial_{x,x}\mathcal{V}(x)-\mathsf{c}(x')\partial_{x',x'}\mathcal{V}(x')\big)|\\ &+|\sum_{j=1}^{\mathsf{r}}\lambda_j\big(\mathcal{V}(x+\rho(x)\mathsf{e}_j^\mathsf{r})-(\mathcal{V}(x'+\rho(x')\mathsf{e}_j^\mathsf{r})\big)|\\ &+|\sum_{j=1}^{\mathsf{r}}\lambda_j(\mathcal{V}(x)-\mathcal{V}(x'))|. \end{aligned} \tag{8}$$

Using the following inequality

$$|\mathbb{A}^T\mathbb{B}-\mathbb{C}^T\mathbb{D}| \leq \|\mathbb{A}\|\|\mathbb{B}-\mathbb{D}\|+\|\mathbb{D}\|\|\mathbb{A}-\mathbb{C}\|,$$

for all $\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D} \in \mathbb{R}^n$, and Assumptions 1-2, the first term in the right-hand side of (8) is upper bounded by

$$\begin{aligned}&\|\partial_x\mathcal{V}(x)\|\|f(x,u)-f(x',u')\|+\|f(x',u')\|\|\partial_x\mathcal{V}(x)-\partial_{x'}\mathcal{V}(x')\|\\ &\leq \mathcal{B}_{\mathcal{V}_1}(\mathscr{L}_f\|x-x'\|+\mathscr{L}_u\|u-u'\|)+\mathcal{B}_f\mathscr{L}_{\mathcal{V}_1}\|x-x'\|.\end{aligned}$$

Using the notation $\mathbb{A}\circ\mathbb{B}$ as the Hadamard product of two matrices $\mathbb{A}, \mathbb{B}$, the second term in the right-hand side of (8) is upper bounded as

$$\begin{aligned}&\frac{1}{2}\Big|\sum_{i,j}[\mathsf{c}(x)\circ\partial_{x,x}\mathcal{V}(x)]_{i,j}-[\mathsf{c}(x')\circ\partial_{x',x'}\mathcal{V}(x')]_{i,j}\Big|\\ &\leq\frac{1}{2}\sum_{i,j}\Big|[(\mathsf{c}(x)-\mathsf{c}(x'))\circ\partial_{x,x}\mathcal{V}(x)]_{i,j}\Big|\\ &\quad+\frac{1}{2}\sum_{i,j}\Big|[\mathsf{c}(x')\circ(\partial_{x,x}\mathcal{V}(x)-\partial_{x',x'}\mathcal{V}(x'))]_{i,j}\Big|\\ &\leq\frac{1}{2}\Big[\sum_{i,j}\Big|[\mathsf{c}(x)-\mathsf{c}(x')]_{i,j}^2\sum_{i,j}[\partial_{x,x}\mathcal{V}(x)]_{i,j}^2\Big]^{\frac{1}{2}}\\ &\quad+\frac{1}{2}\Big[\sum_{i,j}[\mathsf{c}(x')]_{i,j}^2\sum_{i,j}[\partial_{x,x}\mathcal{V}(x)-\partial_{x',x'}\mathcal{V}(x')]_{i,j}^2\Big]^{\frac{1}{2}}\\ &=\frac{1}{2}\|\mathsf{c}(x)-\mathsf{c}(x')\|_F\|\partial_{x,x}\mathcal{V}(x)\|_F+\frac{1}{2}\|\mathsf{c}(x')\|_F\|\partial_{x,x}\mathcal{V}(x)\\ &\quad-\partial_{x',x'}\mathcal{V}(x'))\|_F\leq\frac{1}{2}(\mathscr{L}_\mathsf{c}\mathcal{B}_{\mathcal{V}_2}+\mathcal{B}_\mathsf{c}\mathscr{L}_{\mathcal{V}_2})\|x-x'\|.\end{aligned}$$

Since $\mathcal{V}(x)$ is Lipschitz continuous according to Assumption 1, two last terms in the right-hand side of (8) are upper bounded as

$$\begin{aligned}&\sum_{j=1}^{\mathsf{r}}\lambda_j\big(\mathscr{L}_\mathcal{V}\|x-x'+\rho(x)\mathsf{e}_j^\mathsf{r}-\rho(x')\mathsf{e}_j^\mathsf{r}\|+\mathscr{L}_\mathcal{V}\|x-x'\|\big)\\ &\leq\sum_{j=1}^{\mathsf{r}}\lambda_j\big(\mathscr{L}_\mathcal{V}(\|x-x'\|+\|\rho(x)-\rho(x')\|_F\|\mathsf{e}_j^\mathsf{r}\|)+\mathscr{L}_\mathcal{V}\|x-x'\|\big)\\ &\leq\sum_{j=1}^{\mathsf{r}}\lambda_j\big(\mathscr{L}_\mathcal{V}(\|x-x'\|+\mathscr{L}_\rho\|x-x'\|)+\mathscr{L}_\mathcal{V}\|x-x'\|\big)\\ &=\sum_{j=1}^{\mathsf{r}}\lambda_j\big(2\mathscr{L}_\mathcal{V}+\mathscr{L}_\mathcal{V}\mathscr{L}_\rho\big)\|x-x'\|.\end{aligned}$$

Combining the three upper bounds completes the proof. ■

Now as the first step, we formally quantify the closeness between $\mathcal{L}\mathcal{V}(x)$ and its first approximation $\widehat{\mathcal{L}}_1\mathcal{V}(x)$ in the following theorem.

Theorem 3.2: Under Assumptions 1-2 and Theorem 3.1, one has

$$|\widehat{\mathcal{L}}_1\mathcal{V}(x)-\mathcal{L}\mathcal{V}(x)| \leq \delta_1, \quad \forall x \in X,$$

where:

$$\delta_1 := \mathscr{L}_1\big(\frac{1}{2}(\mathcal{B}_f+\mathcal{B}_\rho\sum_{j=1}^{\mathsf{r}}\lambda_j)\tau+\frac{2}{3}\mathcal{B}_\sigma\sqrt{\tau}\big)+\frac{\tau}{2}\mathscr{L}_2\bar{\mathscr{L}}_u. \tag{9}$$

Proof: Using Dynkin's formula in (4) and by considering the definition of $\widehat{\mathcal{L}}_1\mathcal{V}(x)$ in (5), one has

$$\widehat{\mathcal{L}}_1\mathcal{V}(x)=\mathbb{E}\Big[\frac{1}{\tau}\int_0^\tau \mathcal{L}\mathcal{V}(x(t))\mathsf{d}t\Big],$$

where $x:=x(0)$. By subtracting $\mathcal{L}\mathcal{V}(x)$ from two sides:

$$\widehat{\mathcal{L}}_1\mathcal{V}(x)-\mathcal{L}\mathcal{V}(x)=\mathbb{E}\Big[\frac{1}{\tau}\int_0^\tau \big(\mathcal{L}\mathcal{V}(x(t))-\mathcal{L}\mathcal{V}(x)\big)\mathsf{d}t\Big].$$

Consequently,

$$|\widehat{\mathcal{L}}_1\mathcal{V}(x)-\mathcal{L}\mathcal{V}(x)|\leq\frac{1}{\tau}\int_0^\tau\mathbb{E}\Big[|\mathcal{L}\mathcal{V}(x(t))-\mathcal{L}\mathcal{V}(x)|\Big]\mathsf{d}t.$$

By employing Theorem 3.1, one has

$$\begin{aligned}&|\widehat{\mathcal{L}}_1\mathcal{V}(x)-\mathcal{L}\mathcal{V}(x)|\\&\leq\frac{1}{\tau}\int_0^\tau\mathbb{E}\Big[\mathscr{L}_1\|x(t)-x\|+\mathscr{L}_2\|u(t)-u\|\Big]\mathsf{d}t,\end{aligned}$$

with $u:=u(0)$. Since $\|u(t)-u(t')\|\leq\bar{\mathscr{L}}_u|t-t'|$:

$$|\widehat{\mathcal{L}}_1\mathcal{V}(x)-\mathcal{L}\mathcal{V}(x)|\leq\frac{\mathscr{L}_1}{\tau}\int_0^\tau\mathbb{E}\Big[\|x(t)-x\|\Big]\mathsf{d}t+\frac{\tau}{2}\mathscr{L}_2\bar{\mathscr{L}}_u. \tag{10}$$

Now we aim at finding an upper bound for $\mathbb{E}\Big[\|x(t)-x\|\Big]$. Under the continuity property of the solution process of the system, we have

$$x(t)=x+\int_0^t f(x(s),u(s))\mathsf{d}s+\int_0^t\sigma(x(s))\mathsf{d}\mathbb{W}_s+\sum_{j=1}^{\mathsf{r}}\sum_{i=1}^{\mathbb{P}_t^j}\rho(x_{s_i})\mathsf{e}_j^{\mathsf{r}}, \tag{11}$$

where $\mathbb{P}_t=[\mathbb{P}_t^1;\cdots;\mathbb{P}_t^{\mathsf{r}}]$ is the Poisson process with r events and x_{s_i} is the solution process of the system that jumps at times s_i. Then, one obtains

$$\begin{aligned}&\mathbb{E}\Big[\|x(t)-x\|\Big]\\&=\mathbb{E}\Big[\|\int_0^t f(x(s),u(s))\mathsf{d}s+\int_0^t\sigma(x(s))\mathsf{d}\mathbb{W}_s+\sum_{j=1}^{\mathsf{r}}\sum_{i=1}^{\mathbb{P}_t^j}\rho(x_{s_i})\mathsf{e}_j^{\mathsf{r}}\|\Big]\\&\leq\mathbb{E}\Big[\|\int_0^t f(x(s),u(s))\mathsf{d}s\|+\|\int_0^t\sigma(x(s))\mathsf{d}\mathbb{W}_s\|+|\sum_{j=1}^{\mathsf{r}}\sum_{i=1}^{\mathbb{P}_t^j}\rho(x_{s_i})\mathsf{e}_j^{\mathsf{r}}|\Big].\end{aligned}$$

According to Jensen's inequality, for any vector $a\in\mathbb{R}^n$, $\mathbb{E}[\|a\|]\leq\sqrt{\mathbb{E}[a^Ta]}$. Then,

$$\begin{aligned}\mathbb{E}\Big[\|x(t)-x\|\Big]\leq&\Big[\mathbb{E}\Big[\int_0^t f(x(s),u(s))^T\mathsf{d}s\int_0^t f(x(s),u(s))\mathsf{d}s\Big]\Big]^{\frac{1}{2}}\\&+\Big[\mathbb{E}\Big[\int_0^t\sigma(x(s))^T\mathsf{d}\mathbb{W}_s^T\int_0^t\sigma(x(s))\mathsf{d}\mathbb{W}_s\Big]\Big]^{\frac{1}{2}}\\&+\mathbb{E}\Big[\sum_{j=1}^{\mathsf{r}}\sum_{i=1}^{\mathbb{P}_t^j}\|\rho(x_{s_i})\|\|\mathsf{e}_j^{\mathsf{r}}\|\Big].\end{aligned} \tag{12}$$

Under Assumption 2, the first term in the right-hand side of (12) is upper bounded by

$$\Big[\mathbb{E}\Big[\int_0^t\int_0^t\|f(x(s_1),u(s_1))\|\|f(x(s_2),u(s_2))\|\mathsf{d}s_1\mathsf{d}s_2\Big]\Big]^{\frac{1}{2}}\leq\mathcal{B}_f t. \tag{13}$$

In addition, using the multivariate version of the Itô isometry property [15] and Assumption 2, one can bound the second term in the right-hand side of (12) as

$$\Big[\int_0^t\mathbb{E}\Big[\|\sigma(x(s))\|_F^2\Big]\mathsf{d}s\Big]^{\frac{1}{2}}\leq\mathcal{B}_\sigma\sqrt{t}. \tag{14}$$

Moreover,

$$\begin{aligned}\mathbb{E}\Big[\sum_{j=1}^{\mathsf{r}}\sum_{i=1}^{\mathbb{P}_t^j}\|\rho(x_{s_i})\|\|\mathsf{e}_j^{\mathsf{r}}\|\Big]&\leq\mathbb{E}\Big[\sum_{j=1}^{\mathsf{r}}\sum_{i=1}^{\mathbb{P}_t^j}\mathcal{B}_\rho\Big]\\&=\mathcal{B}_\rho\sum_{j=1}^{\mathsf{r}}\mathbb{E}\big[\mathbb{P}_t^j\big]\leq\mathcal{B}_\rho\sum_{j=1}^{\mathsf{r}}\lambda_j t.\end{aligned} \tag{15}$$

By substituting (13)-(15) in (12), one has

$$\mathbb{E}\Big[\|x(t)-x\|\Big]\leq\mathcal{B}_f t+\mathcal{B}_\sigma\sqrt{t}+\mathcal{B}_\rho\sum_{j=1}^{\mathsf{r}}\lambda_j t. \tag{16}$$

Consequently, by substituting (16) in (10), one has

$$\begin{aligned}&|\widehat{\mathcal{L}}_1\mathcal{V}(x)-\mathcal{L}\mathcal{V}(x)|\\&\leq\frac{\mathscr{L}_1}{\tau}\int_0^\tau(\mathcal{B}_f t+\mathcal{B}_\sigma\sqrt{t}+\mathcal{B}_\rho\sum_{j=1}^{\mathsf{r}}\lambda_j t)\mathsf{d}t+\frac{\tau}{2}\mathscr{L}_2\bar{\mathscr{L}}_u\\&=\mathscr{L}_1\big(\frac{1}{2}(\mathcal{B}_f+\mathcal{B}_\rho\sum_{j=1}^{\mathsf{r}}\lambda_j)\tau+\frac{2}{3}\mathcal{B}_\sigma\sqrt{\tau}\big)+\frac{\tau}{2}\mathscr{L}_2\bar{\mathscr{L}}_u,\end{aligned}$$

which completes the proof. ■

Remark 3.3: If input signal u is piece-wise constant of duration τ instead of being Lipschitz continuous, the error term contributed by u in our setting will be zero. Accordingly, the bound δ_1 in (9) is reduced to $\delta_1:=\mathscr{L}_1\big(\frac{1}{2}(\mathcal{B}_f+\mathcal{B}_\rho\sum_{j=1}^{\mathsf{r}}\lambda_j)\tau+\frac{2}{3}\mathcal{B}_\sigma\sqrt{\tau}\big)$.

As the second step, we now quantify the closeness between $\widehat{\mathcal{L}}_1\mathcal{V}(x)$ and $\widehat{\mathcal{L}}_2\mathcal{V}(x)$. To do so, we first formulate a bound on the variance of $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ in the next theorem.

Theorem 3.4: Under Assumptions 1-2 and Theorem 3.1, the variance of $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ in (6) is bounded by

$$\mathsf{Var}(\widehat{\mathcal{L}}_2\mathcal{V}(x))\leq\frac{1}{N}\Big[\frac{\gamma_1}{\tau}+\frac{\gamma_2}{\sqrt{\tau}}+\gamma_3\Big], \tag{17}$$

for some $\gamma_1,\gamma_2,\gamma_3\in\mathbb{R}_{\geq0}$.

Proof: Since $(x_\tau^i)_{i=1}^N$ is N i.i.d. sampled data by extracting N solution processes $x_\tau^i, i\in\{1,\ldots,N\}$, at time τ from the same initial condition under N different independent noise realizations, we compute the variance of empirical mean as

$$\begin{aligned}&\mathsf{Var}(\widehat{\mathcal{L}}_2\mathcal{V}(x))=\frac{1}{\tau^2N}\mathsf{Var}(\mathcal{V}(x_\tau^i))\\&=\frac{1}{\tau^2N}\Big[\mathbb{E}[\mathcal{V}(x_\tau^i)^2]-\mathbb{E}[\mathcal{V}(x_\tau^i)]^2\Big]\\&=\frac{1}{\tau^2N}\Big[\mathbb{E}[\mathcal{V}(x_\tau^i)^2]-\mathcal{V}(x)^2-\mathbb{E}[\mathcal{V}(x_\tau^i)]^2+\mathcal{V}(x)^2\Big]\\&=\frac{1}{\tau N}\Big[\frac{\mathbb{E}[\mathcal{V}(x_\tau^i)^2]-\mathcal{V}(x)^2}{\tau}\Big]\\&\quad-\frac{1}{\tau N}\frac{[\mathbb{E}[\mathcal{V}(x_\tau^i)]-\mathcal{V}(x)][\mathbb{E}[\mathcal{V}(x_\tau^i)]+\mathcal{V}(x)]}{\tau}\\&=\frac{1}{\tau N}\widehat{\mathcal{L}}_1\mathcal{V}(x)^2-\frac{1}{\tau N}\widehat{\mathcal{L}}_1\mathcal{V}(x)(\mathbb{E}[\mathcal{V}(x_\tau^i)]+\mathcal{V}(x)).\end{aligned}$$

Similar to (9), one can also quantify the distance between $\widehat{\mathcal{L}}_1\mathcal{V}(x)^2$ and $\mathcal{L}\mathcal{V}(x)^2$ as $|\widehat{\mathcal{L}}_1\mathcal{V}(x)^2 - \mathcal{L}\mathcal{V}(x)^2| \leq \bar{\delta}_1$, where

$$\bar{\delta}_1 := \bar{\mathscr{L}}_1\big(\frac{1}{2}(\mathcal{B}_f + \mathcal{B}_\rho\sum_{j=1}^{\mathsf{r}}\lambda_j)\tau + \frac{2}{3}\mathcal{B}_\sigma\sqrt{\tau}\big) + \frac{\tau}{2}\bar{\mathscr{L}}_2\bar{\mathscr{L}}_u,$$

with

$$\bar{\mathscr{L}}_1 = \bar{\mathcal{B}}_{\mathcal{V}_1}\mathscr{L}_f + \mathcal{B}_f\bar{\mathscr{L}}_{\mathcal{V}_1} + \frac{1}{2}(\mathscr{L}_{\mathsf{c}}\bar{\mathcal{B}}_{\mathcal{V}_2} + \mathcal{B}_{\mathsf{c}}\bar{\mathscr{L}}_{\mathcal{V}_2}) + \sum_{j=1}^{\mathsf{r}}\lambda_j(2\bar{\mathscr{L}}_{\mathcal{V}} + \bar{\mathscr{L}}_{\mathcal{V}}\mathscr{L}_\rho), \quad \bar{\mathscr{L}}_2 = \bar{\mathcal{B}}_{\mathcal{V}_1}\mathscr{L}_u,$$

where $\bar{\mathcal{B}}_{\mathcal{V}_1}, \bar{\mathcal{B}}_{\mathcal{V}_2}, \bar{\mathscr{L}}_{\mathcal{V}}, \bar{\mathscr{L}}_{\mathcal{V}_1}, \bar{\mathscr{L}}_{\mathcal{V}_2}$ are constants similar to the ones in Assumptions 1-2 but for $\mathcal{V}(x)^2$. These constants can be readily obtained using $\mathcal{B}_{\mathcal{V}_1}, \mathcal{B}_{\mathcal{V}_2}, \mathscr{L}_{\mathcal{V}}, \mathscr{L}_{\mathcal{V}_1}, \mathscr{L}_{\mathcal{V}_2}$. Then,

$$\mathrm{Var}(\widehat{\mathcal{L}}_2\mathcal{V}(x)) \leq \frac{1}{\tau N}\big((\mathcal{L}\mathcal{V}(x)^2 + \bar{\delta}_1) - (\mathcal{L}\mathcal{V}(x) - \delta_1)2\mathcal{B}_{\mathcal{V}}\big).$$

Accordingly, one has

$$\mathrm{Var}(\widehat{\mathcal{L}}_2\mathcal{V}(x)) \leq \frac{\gamma_1 + \gamma_2\sqrt{\tau} + \gamma_3\tau}{\tau N} = \frac{1}{N}[\frac{\gamma_1}{\tau} + \frac{\gamma_2}{\sqrt{\tau}} + \gamma_3],$$

with γ_1 satisfying $|\mathcal{L}\mathcal{V}(x)^2 - 2\mathcal{B}_{\mathcal{V}}\mathcal{L}\mathcal{V}(x)| \leq \gamma_1, \forall x \in X$, and

$$\gamma_2 := \frac{2}{3}\mathcal{B}_\sigma(\bar{\mathscr{L}}_1 + 2\mathcal{B}_{\mathcal{V}}\mathscr{L}_1),$$

$$\gamma_3 := \frac{1}{2}((\mathcal{B}_f + \mathcal{B}_\rho\sum_{j=1}^{\mathsf{r}}\lambda_j)(\bar{\mathscr{L}}_1 + 2\mathcal{B}_{\mathcal{V}}\mathscr{L}_1) + (\bar{\mathscr{L}}_2\bar{\mathscr{L}}_u + 2\mathcal{B}_{\mathcal{V}}\mathscr{L}_2\bar{\mathscr{L}}_u)).$$

Note that γ_1 can be computed using parameters of Assumptions 1-2, and this completes the proof. ∎

In the next theorem, we employ Chebyshev's inequality [19] and quantify the mismatch between approximated values of the infinitesimal generator in (5) and (6) by providing an a-priori confidence bound.

Theorem 3.5: Let $\widehat{\mathcal{L}}_1\mathcal{V}(x)$ and $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ be approximations of the infinitesimal generator $\mathcal{L}V(x)$ based on the expected value and empirical approximation as in (5) and (6), respectively. For any $\beta \in (0, 1]$, we have

$$\mathbb{P}\Big\{|\widehat{\mathcal{L}}_1\mathcal{V}(x) - \widehat{\mathcal{L}}_2\mathcal{V}(x)| \leq \delta_2\Big\} \geq 1 - \beta, \quad \forall x \in X,$$

with

$$\delta_2 := \big[\frac{1}{\beta N}\big(\frac{\gamma_1}{\tau} + \frac{\gamma_2}{\sqrt{\tau}} + \gamma_3\big)\big]^{\frac{1}{2}}. \tag{18}$$

Proof: We know that $\mathbb{E}\big[\widehat{\mathcal{L}}_2\mathcal{V}(x)\big] = \widehat{\mathcal{L}}_1\mathcal{V}(x)$. According to Chebyshev's inequality [19], one has

$$\mathbb{P}\Big\{|\widehat{\mathcal{L}}_1\mathcal{V}(x) - \widehat{\mathcal{L}}_2\mathcal{V}(x)| \leq \delta_2\Big\} = \mathbb{P}\Big\{|\mathbb{E}\big[\widehat{\mathcal{L}}_2\mathcal{V}(x)\big] - \widehat{\mathcal{L}}_2\mathcal{V}(x)| \leq \delta_2\Big\} \geq 1 - \frac{\sigma^2}{\delta_2^2},$$

for any $\delta_2 \in \mathbb{R}_{>0}$, where σ^2 is the variance of $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ which can be computed using Theorem 3.4:

$$\sigma^2 := \mathrm{Var}\big[\frac{1}{N}\sum_{i=1}^{N}\mathcal{V}(x_\tau^i)\big] \leq \frac{1}{N}\big[\frac{\gamma_1}{\tau} + \frac{\gamma_2}{\sqrt{\tau}} + \gamma_3\big].$$

Considering $\beta = \sigma^2/\delta_2^2$ gives the expression (18) for δ_2 as a function of β, which completes the proof. ∎

By leveraging Theorems 3.2 and 3.5, we now propose the next theorem as our solution to Problem 2.3 for the formal quantification of the closeness between the infinitesimal generator $\mathcal{L}\mathcal{V}(x)$ and its data-driven approximation $\widehat{\mathcal{L}}_2\mathcal{V}(x)$.

Theorem 3.6: Let $\mathcal{L}\mathcal{V}(x(t))$ be the infinitesimal generator of the stochastic process $x(t)$ and $\widehat{\mathcal{L}}_2\mathcal{V}(x(t))$ be its approximation via the empirical mean as in (6). By employing the results of Theorems 3.2 and 3.5, one has

$$\mathbb{P}\Big\{|\widehat{\mathcal{L}}_2\mathcal{V}(x) - \mathcal{L}\mathcal{V}(x)| \leq \delta\Big\} \geq 1 - \beta, \quad \forall x \in X,$$

for any $\beta \in (0, 1]$ with $\delta = \delta_1 + \delta_2$, where δ_1 and δ_2 are defined in (9) and (18), respectively.

Proof: By defining events

$$\mathcal{A}_1 = \{|\widehat{\mathcal{L}}_1\mathcal{V}(x) - \mathcal{L}\mathcal{V}(x)| \leq \delta_1\}, \mathcal{A}_2 = \{|\widehat{\mathcal{L}}_1\mathcal{V}(x) - \widehat{\mathcal{L}}_2\mathcal{V}(x)| \leq \delta_2\},$$
$$\mathcal{A}_3 = \{|\widehat{\mathcal{L}}_2\mathcal{V}(x) - \mathcal{L}\mathcal{V}(x)| \leq \delta\},$$

one has $\mathbb{P}\{\bar{\mathcal{A}}_1\} = 0$ since $\mathcal{A}_1$ is a deterministic inequality and holds true and $\mathbb{P}\{\bar{\mathcal{A}}_2\} \leq \beta$, where $\bar{\mathcal{A}}_1$ and $\bar{\mathcal{A}}_2$ are the complement of $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. We are interested in computing the concurrent occurrence of events $\mathcal{A}_1$ and $\mathcal{A}_2$:

$$\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) = 1 - \mathbb{P}\big(\bar{\mathcal{A}}_1 \cup \bar{\mathcal{A}}_2\big).$$

Since $\mathbb{P}\big(\bar{\mathcal{A}}_1 \cup \bar{\mathcal{A}}_2\big) \leq \mathbb{P}(\bar{\mathcal{A}}_1) + \mathbb{P}(\bar{\mathcal{A}}_2)$, we have

$$\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) \geq 1 - \mathbb{P}\big(\bar{\mathcal{A}}_1\big) - \mathbb{P}\big(\bar{\mathcal{A}}_2\big) \geq 1 - \beta. \tag{19}$$

Due to the triangle inequality, $\mathcal{A}_1 \cap \mathcal{A}_2 \subseteq \mathcal{A}_3$, and accordingly, $\mathbb{P}\big(\mathcal{A}_1 \cap \mathcal{A}_2\big) \leq \mathbb{P}\big(\mathcal{A}_3\big)$. By employing (19), one has $\mathbb{P}\big(\mathcal{A}_3\big) \geq 1 - \beta$, which completes the proof. ∎

Remark 3.7: The variance of the empirical mean in (17) has an inverse relation with both the sampling time and the number of data. On the other hand, $\widehat{\mathcal{L}}_1\mathcal{V}(x)$ converges to $\mathcal{L}\mathcal{V}(x)$ if τ goes to zero. This means the overall closeness δ between the infinitesimal generator $\mathcal{L}\mathcal{V}(x)$ and its approximation $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ is improved by increasing the number of data N (which is only appears in δ_2). However, since τ appears in both δ_1 and δ_2, decreasing τ does not necessarily improve δ and its optimal value should be computed to reach the least error (cf. Fig. 1 Top).

IV. Case Study: Jet Engine Compressor

To demonstrate the effectiveness of the proposed results, we apply our data-driven approaches to a *nonlinear* jet engine compressor [20]:

$$\Sigma: \begin{bmatrix}\mathsf{d}x_1(t)\\ \mathsf{d}x_2(t)\end{bmatrix} = \begin{bmatrix}-x_2(t) - \frac{3}{2}x_1^2(t) - \frac{1}{2}x_1^3(t)\\ x_1(t) - u(t)\end{bmatrix}\mathsf{d}t + \begin{bmatrix}0.1\mathsf{d}\mathbb{W}_t\\ 0.1\mathsf{d}\mathbb{W}_t\end{bmatrix} + \begin{bmatrix}0.1\mathsf{d}\mathbb{P}_t\\ 0.1\mathsf{d}\mathbb{P}_t\end{bmatrix},$$

where $x_1 = \Phi - 1 \in [0, 1], x_2 = \Psi - \Lambda - 2 \in [0, 1]$, with Φ, Ψ, Λ being, respectively, the mass flow, the pressure rise, and a constant. We assume that the model is unknown to us. In addition, the controller is also unknown and we only have its Lipschitz constant as $\bar{\mathscr{L}}_u = 1.12$.

We fix $\mathcal{V}(x) = 0.01x_1^2 + 0.02x_1x_2 + 0.01x_2^2$, and compute parameters of Assumptions 1-2. Then one has

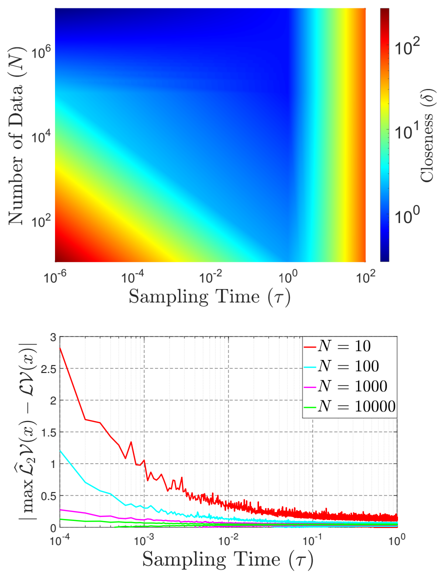

Fig. 1. Top: Closeness δ, represented by 'colour bar', based on different ranges of the sampling time τ and number of data N. As it can be observed, for a fixed number of N, the total error δ first decreases for $\tau \in [10^{-6}, 10^0]$ and again increases for $\tau \in [10^0, 10^2]$ (cf. Remark 3.7). Bottom: Difference between the exact analytical $\mathcal{L}\mathcal{V}(x)$ and its approximation $\widehat{\mathcal{L}}_2\mathcal{V}(x)$.

$\mathscr{L}_{\mathcal{V}} = 0.056, \mathscr{L}_{\mathcal{V}_1} = 0.04, \mathscr{L}_{\mathcal{V}_2} = 0, \mathcal{B}_{\mathcal{V}} = 0.04, \mathcal{B}_{\mathcal{V}_1} = 0.056, \mathcal{B}_{\mathcal{V}_2} = 0.04$. We also assume that $\mathscr{L}_f = 4.7, \mathscr{L}_u = 1, \mathscr{L}_c = 0, \mathcal{B}_f = 3.08, \mathcal{B}_c = 0.014, \mathcal{B}_\sigma = 0.14$, and $\mathcal{B}_\rho = 0.14$. We fix $\tau = 0.01$. Then according to Theorem 3.2, one can guarantee that the closeness between $\mathcal{L}\mathcal{V}(x)$ and its first approximation $\widehat{\mathcal{L}}_1\mathcal{V}(x)$ can be bounded by $\delta_1 = 0.01$, *i.e.,*

$$|\widehat{\mathcal{L}}_1\mathcal{V}(x) - \mathcal{L}\mathcal{V}(x)| \leq 0.01, \quad \forall x \in X.$$

We now proceed with computing an upper bound for the variance of $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ according to Theorem 3.4. By selecting $N = 10^5$, one has $\mathrm{Var}(\widehat{\mathcal{L}}_2\mathcal{V}(x)) \leq 9.5 \times 10^{-6}$. Now according to Theorem 3.5, by taking $\beta = 0.01$, we compute the closeness between $\widehat{\mathcal{L}}_1\mathcal{V}(x)$ and $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ as $\delta_2 = 0.03$ with a confidence of at least 99%, *i.e.,*

$$\mathbb{P}\Big\{|\widehat{\mathcal{L}}_1\mathcal{V}(x) - \widehat{\mathcal{L}}_2\mathcal{V}(x)| \leq 0.03\Big\} \geq 0.99, \quad \forall x \in X.$$

According to Theorem 3.6, we formally quantify the closeness between the infinitesimal generator $\mathcal{L}\mathcal{V}(x)$ and its approximation via the empirical mean $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ as $\delta = 0.04$ with a confidence of at least 99%, *i.e.,*

$$\mathbb{P}\Big\{|\widehat{\mathcal{L}}_2\mathcal{V}(x) - \mathcal{L}\mathcal{V}(x)| \leq 0.04\Big\} \geq 0.99, \quad \forall x \in X.$$

Simulation Results. We fix $\beta = 0.01$ (*i.e.,* confidence is 99%) and plot the closeness δ based on different ranges of the sampling time τ and number of data N in Fig. 1 (Top). As can be observed, the closeness δ between the infinitesimal generator $\mathcal{L}\mathcal{V}(x)$ and its approximation $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ is improved by increasing the number of data N. However, since $\delta = \delta_1 + \delta_2$ and the sampling time τ appears in both δ_1 in (9) and δ_2 in (18), the closeness δ is not monotonic for a given confidence $1 - \beta$. We now assume that we know the model and compute $\mathcal{L}\mathcal{V}(x)$ via (3) which is clearly independent of the sampling time. We compute $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ based on (6). In Fig. 1 (Bottom), we plot the difference between the exact $\mathcal{L}\mathcal{V}(x)$ and its approximation $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ for the initial condition $x_1(0) = 0.1, x_2(0) = 0.2$ but for different ranges of the sampling time. We compute $\widehat{\mathcal{L}}_2\mathcal{V}(x)$ 500 times with different numbers of data and plot only the maximum of computed values. As can be observed, for the small sampling time (*e.g.,* 10^{-4}), the number of data should be large enough such that $N\tau$ remains large enough, and accordingly, one can provide a reasonable closeness precision between $\mathcal{L}\mathcal{V}(x)$ and $\widehat{\mathcal{L}}_2\mathcal{V}(x)$.

References

[1] A. R. Teel, A. Subbaraman, and A. Sferlazza, "Stability analysis for stochastic hybrid systems: A survey," *Automatica*, vol. 50, no. 10, pp. 2435–2456, 2014.

[2] P. Zhao, Y. Kang, and D. Zhai, "On input-to-state stability of stochastic nonlinear systems with markovian jumping parameters," *International journal of control*, vol. 85, no. 4, pp. 343–349, 2012.

[3] A. A. Julius and G. J. Pappas, "Approximations of stochastic hybrid systems," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1193–1203, 2009.

[4] A. Nejati, S. Soudjani, and M. Zamani, "Compositional abstraction-based synthesis for continuous-time stochastic hybrid systems," *European Journal of Control*, vol. 57, pp. 82–94, 2021.

[5] M. Zamani, P. Mohajerin Esfahani, R. Majumdar, A. Abate, and J. Lygeros, "Symbolic control of stochastic systems via approximately bisimilar finite abstractions," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3135–3150, 2014.

[6] S. Prajna, A. Jadbabaie, and G. J. Pappas, "A framework for worst-case and stochastic safety verification using barrier certificates," *IEEE Transactions on Automatic Control*, vol. 52, no. 8, pp. 1415–1428, 2007.

[7] C. Santoyo, M. Dutreix, and S. Coogan, "Verification and control for finite-time safety of stochastic systems via barrier functions," in *Proceedings of the IEEE Conference on Control Technology and Applications (CCTA)*, 2019, pp. 712–717.

[8] A. Nejati, S. Soudjani, and M. Zamani, "Compositional construction of control barrier functions for networks of continuous-time stochastic systems," *Proceedings of the 21st IFAC World Congress*, vol. 53, no. 2, pp. 1856–1861, 2020.

[9] Z. Hou and Z. Wang, "From model-based control to data-driven control: Survey, classification and perspective," *Information Sciences*, vol. 235, pp. 3–35, 2013.

[10] A. Nejati, A. Lavaei, S. Soudjani, and M. Zamani, "Data-driven estimation of infinitesimal generators of stochastic systems," *Proceedings of the 7th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS)*, vol. 54, no. 5, pp. 277–282, 2021.

[11] H. A. Blom, J. Lygeros, M. Everdij, S. Loizou, and K. Kyriakopoulos, *Stochastic hybrid systems: theory and safety critical applications*. Springer, 2006, vol. 337.

[12] C. G. Cassandras and J. Lygeros, *Stochastic hybrid systems*. CRC Press, 2006.

[13] L. Donati, M. Heida, B. G. Keller, and M. Weber, "Estimation of the infinitesimal generator by square-root approximation," *Journal of Physics: Condensed Matter*, vol. 30, no. 42, p. 425201, 2018.

[14] B. K. Øksendal and A. Sulem, *Applied stochastic control of jump diffusions*. Springer, 2005, vol. 498.

[15] B. Oksendal, *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[16] E. B. Dynkin, "Markov processes." Springer, 1965, pp. 77–104.

[17] A. Salamati, A. Lavaei, S. Soudjani, and M. Zamani, "Data-driven verification and synthesis of stochastic systems via barrier certificates," *submitted for publication, arXiv:2111.10330*, 2021.

[18] A. Lavaei, S. Soudjani, A. Abate, and M. Zamani, "Automated verification and synthesis of stochastic hybrid systems: A survey," *Automatica, to appear*, 2022.

[19] J. G. Saw, M. C. Yang, and T. C. Mo, "Chebyshev inequality with estimated mean and variance," *The American Statistician*, vol. 38, no. 2, pp. 130–132, 1984.

[20] A. Anta and P. Tabuada, "To sample or not to sample: Self-triggered control for nonlinear systems," *IEEE Transactions on automatic control*, vol. 55, no. 9, pp. 2030–2042, 2010.