Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Controllability Gramian of Nonlinear Gaussian Process State Space Models with Application to Model Sparsification ⋆

Kenji Kashima * **Misaki Imai** *

* *K. Kashima and M. Imai are with the Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan.*
kk@i.kyoto-u.ac.jp, imai.misaki@bode.amp.i.kyoto-u.ac.jp

Abstract: For linear control systems, the so-called controllability Gramian has played an important role to quantify how effectively the dynamical states can be driven to a target one by a suitable driving input. On the other hand, thanks to the availability of Big Data, the Gaussian process state space model, a data-driven probabilistic modeling framework, has attracted much attention in recent years. In this paper, we newly introduce the concept of the controllability Gramian for nonlinear dynamics represented by the Gaussian process state space model, aiming at better understanding of this new modeling framework. Then, its effective calculation method and application to model sparsification are investigated.

Keywords: Gaussian process state space model, Controllability Gramian, Machine learning

1. INTRODUCTION

In system and control theory for linear systems, fundamental mathematical structure such as controllability, observability, optimal control, robustness has been well understood. These concepts have played a crucial role in control system design and analysis; see e.g., Zhou et al. (1996). Furthermore, various results have been successfully extended to nonlinear control systems to some extent, which also have been utilized to construct nonlinear control systems; see Khalil (2002). On the other hand, based on recent availability of Big Data and remarkable advance of machine learning, data-driven modeling has attracted much attention in controls community. Such a machine-learning assisted modeling is easier to implement than the first principles modeling and has higher describing ability in general. However, the obtained models are mainly used for simulation or simple error evaluation, but not for understanding fundamental mathematical structure of the modeled dynamics, which shows a clear contrast to the conventional system and control theory.

The ultimate goal of our study is to bridge the gap between these two complementary disciplines. Toward this end, in this paper, we focus on the *Gaussian Process State Space Models* (abbr. GPSSM), which is a typical data-driven modeling framework Eleftheriadis et al. (2017); McHutchon (2014). Then, we attempt to extend the concept of the controllability Gramian, one of the most classical tools in systems control theory, to GPSSM. GPSSM can be viewed as a probabilistic modeling method by local interpolation of the vector field of dynamics. In comparison to other similar methods such as Just-in-Time modeling (Cleveland and Devlin (1988); Zheng and Kimura (2008); Stenman et al. (1996)), the Gaussian process approximation has several mathematically interesting properties; Rasmussen and Williams (2006). In particular, in this paper, *moment matching* technique of the Gaussian process approximation is utilized for effective calculation of the controllability Gramian that we newly introduce.

The organization of this paper is as follows: In Section 2, we briefly overview several definitions and fundamental results on the controllability Gramian. Then, we newly introduce controllability Gramian for GPSSM, and investigate its practical calculation procedure in Section 3. In Section 4, we apply the newly introduced controllability Gramian to model sparsification. Some concluding remarks are given in Section 5.

Notations: The determinant of a square matrix $\boldsymbol{X}$ is denoted by $|\boldsymbol{X}|$. The matrix exponential of a square matrix $\boldsymbol{X}$ is represented by $\mathrm{e}^{\boldsymbol{X}}$. The probability density of the multidimensional (denoted by k) normal distribution $\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ with mean vector $\boldsymbol{\mu} \in \mathbb{R}^k$ and (positive definite) covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{k\times k}$ is defined by

$$\frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

The variable $\boldsymbol{x}$ is omitted if it is clear from the context.

2. PRELIMINARY: QUANTITATIVE CONTROLLABILITY ANALYSIS OF NONLINEAR SYSTEMS

2.1 Controllability Gramian of linear systems

Let us consider the continuous-time linear time-invariant system

$$\frac{d}{dt}\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}u(t), \quad \boldsymbol{x}(0) = \boldsymbol{0} \tag{1}$$

⋆ This work was supported in part by JSPS KAKENHI under Grant JP18H01461.

where $\boldsymbol{x}(t)$ is the n-dimensional state, $u(t)$ is the scalar input, and $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^n$. Then, (finite-time) controllability Gramian $\mathbf{G}_\tau$ is defined by

$$\mathbf{G}_\tau := \int_0^\tau \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{B}\boldsymbol{B}^\top \mathrm{e}^{\boldsymbol{A}^\top t}dt. \tag{2}$$

This positive semi-definite matrix contains rich information about controllability of dynamical systems. For example, $\mathbf{G}_\tau$ is nonsingular if and only if the system is controllable in that all the state transfer (i.e., any initial and target states pair) is realizable by a suitable finite energy input.

Beyond this binary (controllable or not) property, for quantitative analysis of controllability, let us introduce the controllability function $L^\tau(\tilde{\boldsymbol{x}})$ as the minimum input energy required to drive the state to a target one $\tilde{\boldsymbol{x}}$ at time τ, i.e.,

$$L^\tau(\tilde{\boldsymbol{x}}) := \inf_u \int_0^\tau \frac{1}{2}u(t)^2 dt \quad \text{s.t. } \boldsymbol{x}(\tau) = \tilde{\boldsymbol{x}}. \tag{3}$$

It is well-known that, when the matrix in (2) is nonsingular, this controllability function satisfies

$$L^\tau(\tilde{\boldsymbol{x}}) = \frac{1}{2}\tilde{\boldsymbol{x}}^\top \mathbf{G}_\tau^{-1}\tilde{\boldsymbol{x}}. \tag{4}$$

This equation indicates that states on which the controllability function is large (i.e., states that are not reachable by small energy input) are characterized by eigenspace corresponding to small eigenvalue of $\mathbf{G}_\tau$. Elimination of such a subspace is the central idea in control theoretic model reduction methods and their error analysis; see Antoulas (2005) for details.

2.2 Controllability function and noise response data

Next, we move to input-affine nonlinear systems. To the best of our knowledge, extension of the definition in (2) is not straightforward; see Lall et al. (2002) for a practically useful variant. On the other hand, the definition (3) of the controllability function does not depend on the linearity of the dynamics, although it is challenging to compute analytically or even numerically because it is no longer a quadratic function. In such a situation, Kashima (2016) utilized (4) as a definition to introduce a controllability Gramian for nonlinear systems. To see this, consider the input-affine nonlinear system having noise term (Kappen (2005))

$$\begin{aligned}&\frac{d}{dt}\boldsymbol{x}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t)\big) + \boldsymbol{g}\big(\boldsymbol{x}(t)\big)\big(u(t) + \sqrt{T}\xi(t)\big) \\ &\boldsymbol{x}(0) = \mathbf{0}\end{aligned} \tag{5}$$

where $\xi(t)$ is normalized white noise [1] and $T > 0$. Then, define the stochastic controllability function

$$\mathcal{L}^\tau(\tilde{\boldsymbol{x}}) := \inf_u \mathbb{E}_\xi\left[\int_0^\tau \frac{1}{2}\|u(t)\|^2 dt + \Phi(\boldsymbol{x}(\tau) - \tilde{\boldsymbol{x}})\right] \tag{6}$$

$$\Phi(\mathbf{0}) = 0, \quad \mathrm{e}^{-\Phi(x)} \propto \delta(x) \tag{7}$$

where $\mathbb{E}_\xi[\cdot]$ represents expectation with respect to the noise [2]. Furthermore, by using Gibbs distribution, define

$$\boldsymbol{G}(\mathcal{L}) := \int \phi_{\mathcal{L}}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^\top d\tilde{\boldsymbol{x}}, \tag{8}$$

$$\phi_{\mathcal{L}}(\tilde{\boldsymbol{x}}) := \frac{\mathrm{e}^{-\mathcal{L}(\tilde{x})}}{\int \mathrm{e}^{-\mathcal{L}(\tilde{x})}d\tilde{\boldsymbol{x}}}.$$

Then, it is verified that, for linear system (1), $\boldsymbol{G}(\mathcal{L}^\tau)$ coincides with $\boldsymbol{G}_\tau$ in (2). In view of this, we refer to $\boldsymbol{G}(\mathcal{L}^\tau)$ as *controllability Gibbs Gramian* of system (5). While $\boldsymbol{G}(\mathcal{L}^\tau)$ is still difficult to compute directly, the following proposition suggests another way of computing this matrix:

Proposition 2.1. Let $\bar{\boldsymbol{x}}(t)$ be the solution of (5) with $u(t) = 0$. Then,

$$\mathcal{G}_\tau := \mathbb{E}_\xi[\bar{\boldsymbol{x}}(\tau)\bar{\boldsymbol{x}}(\tau)^\top] \tag{9}$$

satisfies

$$\boldsymbol{G}(\mathcal{L}^\tau/T) = \mathcal{G}_\tau. \tag{10}$$

This proposition states that $\mathcal{G}_\tau$ can be computed by the Monte Carlo simulation based on noise response data of (5).

3. CONTROLLABILITY OF GAUSSIAN PROCESS STATE SPACE MODELS

3.1 Definition

In this section, we investigate GPSSM, which is a dynamical system whose vector fields are probabilistically modeled by the Gaussian process approximation (Wang et al. (2006)). Suppose that for $\{\boldsymbol{x}_i\}_{i=1\dots N}$, noisy observation $\{\boldsymbol{f}_i\}_{i=1\dots N}$ of $\boldsymbol{f}(\boldsymbol{x}_i)$ are available. In view of the observation based on (5) in the previous section, let us consider the following discrete-time GPSSM with noise term:

$$\begin{aligned}&\boldsymbol{x}_{s+1} = \boldsymbol{f}(\boldsymbol{x}_s) + \boldsymbol{g}(\boldsymbol{x}_s)(u_s + \sqrt{T}\xi_s), \\ &\boldsymbol{x}_0 = \mathbf{0}\end{aligned} \tag{11}$$

where ξ_s is independent and identically distributed (i.i.d.) normalized Gaussian random variables and $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^n$ is a known function. The probability distribution of a-th component $f_a(\boldsymbol{x})$ of $\boldsymbol{f}(\boldsymbol{x})$ is a Gaussian process whose *a posteriori* probability distribution is given by

$$f_a(\boldsymbol{x}) \sim \mathcal{N}\left(\boldsymbol{k}_a^\top \boldsymbol{\beta}_a, k_a - \boldsymbol{k}_a^\top \boldsymbol{K}_a^{-1}\boldsymbol{k}_a\right).$$

with

$$\begin{aligned}\boldsymbol{K}_a &:= \{k_a(\boldsymbol{x}_i, \boldsymbol{x}_j) + \delta_{ij}\bar{\sigma}_a^2\}_{j=1\dots N}^{i=1\dots N} \\ \boldsymbol{k}_a &:= \{k_a(\boldsymbol{x}, \boldsymbol{x}_i)\}_{i=1\dots N} \\ k_a &:= k_a(\boldsymbol{x}, \boldsymbol{x}) \\ \boldsymbol{f}_a &:= \{f_a(\boldsymbol{x}_i)\}_{i=1\dots N} \\ \boldsymbol{\beta}_a &:= \boldsymbol{K}_a^{-1}\boldsymbol{f}_a\end{aligned}$$

where the kernel function is defined by

$$k_a(\boldsymbol{x}, \boldsymbol{y}) = \sigma_a^2 \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{y})^\top \boldsymbol{\Lambda}_a^{-1}(\boldsymbol{x} - \boldsymbol{y})\right)$$

with $\sigma_a > 0$ and diagonal matrix $\boldsymbol{\Lambda}_a$; see Fig. 1.

For this system, having Proposition 2.1 in our mind, let us define its controllability Gramian as follows:

[1] More mathematically rigorous formulation requires stochastic differential equations.

[2] Intuitively, Eq. (7) represents the terminal constraint $x(\tau) = \tilde{x}$ as a terminal cost; see Kashima (2016).

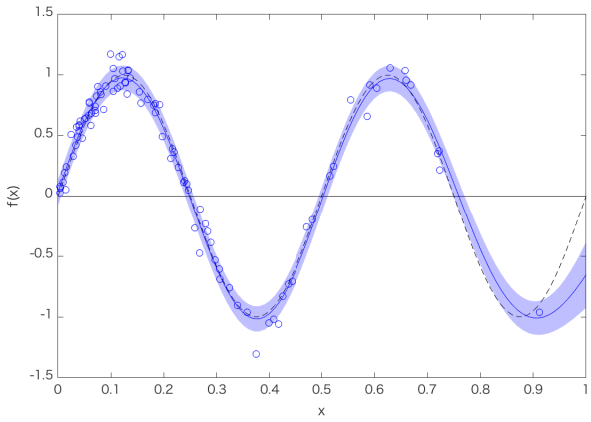


Fig. 1. Gaussian process approximation (dashed: function to be estimated, circle: noisy observation, solid: mean of *a posteriori* distribution, envelope: 1σ-confidence interval)

Definition 3.1. Let $\bar{\boldsymbol{x}}_s$ be the solution for (11) with $u_s = 0$. Then, the controllability Gramian of the Gaussian Process State Space Model (11) is defined by

$$\mathcal{G}_s := \mathbb{E}_{\xi,\boldsymbol{f}}[\bar{\boldsymbol{x}}_s \bar{\boldsymbol{x}}_s^\top], \tag{12}$$

where $\mathbb{E}_{\xi,\boldsymbol{f}}[\cdot]$ represents the expectation with respect to the noise process ξ and the probabilistic uncertainty of the vector field $\boldsymbol{f}$.

Remark 1. Although the input-affine structure is assumed only because the theoretical result has been obtained for this class. Actually, the GPSSM does not need to be in this form.

3.2 Approximate computation via moment matching

Although it may seem possible to compute (12) via Monte Carlo simulation as in the case of Proposition 2.1, it is not realistic due to the need for random sampling of the vector field $\boldsymbol{f}(\cdot)$. To circumvent this difficulty, we propose to utilize moment matching; Girard et al. (2003). The moment matching is a method to perform Gaussian process prediction with uncertain (normal distribution) input data. It should be noted that this moment matching often has enough accuracy for control and have been utilized for reinforcement learning; see e.g., Deisenroth and Rasmussen (2011) for PILCO.

The detailed formula is given as follows: Assume that the probability distribution of $\bar{\boldsymbol{x}}$ at time k can be approximated by a normal distribution

$$\mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$$

with a suitable $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$. Then, we have

$$\begin{aligned}\boldsymbol{\mu}_{s+1} &= \mathbb{E}\left[\boldsymbol{f}(\bar{\boldsymbol{x}}_s)\right] \\ \boldsymbol{\Sigma}_{s+1} &= \mathrm{var}\left[\boldsymbol{f}(\bar{\boldsymbol{x}}_s)\right] + T\ \mathbb{E}\left[\boldsymbol{g}(\bar{\boldsymbol{x}}_s)\boldsymbol{g}(\bar{\boldsymbol{x}}_s)^\top\right].\end{aligned}$$

In order to compute this update in the right hand side, let us define $\boldsymbol{q}, \boldsymbol{Q}$ in an elementwise manner as follows:

$$\begin{aligned}\boldsymbol{q}_a &:= \mathbb{E}[\boldsymbol{k}_a] \\ (\boldsymbol{q}_a)_i &= \sigma_a^2 \left|\boldsymbol{\Sigma}_s \boldsymbol{\Lambda}_a^{-1} + \boldsymbol{I}\right|^{-1/2} \mathcal{G}(\boldsymbol{x}_i; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s + \boldsymbol{\Lambda}_a) \\ \boldsymbol{Q}_{ab} &:= \mathbb{E}[\boldsymbol{k}_a \boldsymbol{k}_b^\top] \\ (\boldsymbol{Q}_{ab})_{ij} &= \left|\boldsymbol{\Sigma}_s(\boldsymbol{\Lambda}_a^{-1} + \boldsymbol{\Lambda}_b^{-1}) + \boldsymbol{I}\right|^{-1/2} \times \\ &\qquad k_a(\boldsymbol{x}_i, \boldsymbol{\mu}_s) k_b(\boldsymbol{x}_j, \boldsymbol{\mu}_s) \exp\left(\frac{1}{2}\boldsymbol{z}_{ij}^\top \boldsymbol{T}^{-1} \boldsymbol{z}_{ij}\right) \\ \mathcal{G}(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{\Sigma}) &:= \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{y})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{y})\right) \\ \boldsymbol{z}_{ij} &:= \boldsymbol{\Lambda}_a^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_s) + \boldsymbol{\Lambda}_b^{-1}(\boldsymbol{x}_j - \boldsymbol{\mu}_s) \\ \boldsymbol{T} &:= \boldsymbol{\Lambda}_a^{-1} + \boldsymbol{\Lambda}_b^{-1} + \boldsymbol{\Sigma}_s^{-1}\end{aligned}$$

for $1 \le a, b \le n$, $1 \le i, j \le N$, Then, we obtain

$$\begin{aligned}\mathbb{E}\left[f_a(\bar{\boldsymbol{x}})\right] &= \boldsymbol{\beta}_a^\top \boldsymbol{q}_a \\ \mathrm{var}[f_a(\bar{\boldsymbol{x}})] &= k_a - \mathrm{Tr}[\boldsymbol{K}_a^{-1}\boldsymbol{Q}_{aa}] \\ &\quad + \boldsymbol{\beta}_a^\top(\boldsymbol{Q}_{aa} - \boldsymbol{q}_a \boldsymbol{q}_a^\top)\boldsymbol{\beta}_a \\ \mathrm{cov}[f_a(\bar{\boldsymbol{x}}), f_b(\bar{\boldsymbol{x}})] &= \boldsymbol{\beta}_a^\top(\boldsymbol{Q}_{ab} - \boldsymbol{q}_a \boldsymbol{q}_b^\top)\boldsymbol{\beta}_b\end{aligned}$$

Based on this recurrent formula, we can approximately compute the controllablity Gramian of GPSSM directly from available data $\{\boldsymbol{x}_i\}$ and $\{\boldsymbol{f}_i\}$.

The definitions and characteristics of several controllability Gramian are summarized in Table 1.

3.3 Numerical example

In order to perform comparison with the true value of the controllability Gramian of the *real* dynamics, let the system to be modeled be the discrete-time linear system such as

$$\boldsymbol{x}_{s+1} = \boldsymbol{A}\boldsymbol{x}_s + \boldsymbol{B}(\boldsymbol{u}_s + \xi_s), \quad \boldsymbol{x}_0 = \boldsymbol{0} \tag{13}$$

with

$$\boldsymbol{A} = \begin{bmatrix} 0.86 & 0.06 \\ 0.08 & 0.94 \end{bmatrix},$$
$$\boldsymbol{B} = \begin{bmatrix} 0.01 \\ 0.01 \end{bmatrix}.$$

Suppose that the matrix $\boldsymbol{B}$ is known and that the vector field is estimated through the Gaussian process approximation by $\{\boldsymbol{x}_i\}$ and $\boldsymbol{f}_i := A\boldsymbol{x}_i + \epsilon_i$ with i.i.d. observation noise $\epsilon_i \in \mathbb{R}^2$ generated by

$$\epsilon_i \sim \mathcal{N}(\boldsymbol{0}, 0.01^2 \boldsymbol{I}).$$

In this example, we generated $\boldsymbol{x}_i$ according to the uniform distribution over

$$S := [-0.5, 0.5]^2. \tag{14}$$

The parameter in the kernel function is $\bar{\sigma}^2 = 10^{-4}, \sigma_a^2 = 10$ and $\Lambda_a = 10 \cdot I$.

For comparison purpose, we calculate the conventional controllability Gramian for linear systems:

$$\mathbf{G}_\tau := \sum_{s=0}^{\tau} \boldsymbol{A}^s \boldsymbol{B}\boldsymbol{B}^\top (\boldsymbol{A}^\top)^s. \tag{15}$$

The controllability Gramian $\mathcal{G}_\tau$ defined in Definition 3.1 and approximately computed via the recurrent formula

Table 1. Comparison of definitions and computation of controllability Gramians

System description	Modeling from data	Definition	Calculation
Linear system (1)	linear regression	(2)	Lyapunov equation
Nonlinear system (5)	nonlinear regression	(8)	Monte Carlo simulation of noise response data and principal component analysis (Proposition 2.1)
Gaussian process state space models (11)	kernel method	Definition 3.1	Analytic recurrent formula for approximated calculation (Section 3.2)

based on moment matching and $\mathbf{G}_\tau$ in (15) are shown in Fig. 2 for

$$N = 20, 100, 500.$$

This results, as expected, indicates that the gap between the calculated controllability Gramian $\mathcal{G}_\tau$ and $\mathbf{G}_\tau$ becomes smaller as N increases, which should be a consequence of the fact that the obtained GPSSM becomes closer to (13) in a *probalistic* sense.

4. APPLICATION TO MODEL SPARSIFICATION

Sparse Gaussian process is an approximation method to reduce the computation complexity of prediction based on the Gaussian process (Snelson and Ghahramani (2006)). In this section, we propose to determine parameters required for the sparsification based on the controllability Gramian. For this purpose, we here provide a brief review of standard idea of sparse Gaussian process. For the computation of the Gaussian process, the inverse of the kernel matrix, which is $O(N^3)$ with the data size N, is needed. The motivation of the sparsification is to reduce this operation with reasonable approximation error.

Suppose that we attempt to calculate the *a posteriori* probability distribution of the output y_* for a new input x_* by using N sets of available data $\boldsymbol{x} = \{x_i\}_{i=1\dots N}, \boldsymbol{y} = \{y_i\}_{i=1\dots N}$. In order to reduce the computation complexity, an *inducing variable* (or pseudo input) $\hat{\boldsymbol{x}} = \{\hat{x}_i\}_{i=1\dots M}$ is introduced. Given the output $\hat{\boldsymbol{y}}$ for the inducing variable, assume that $\boldsymbol{y}, y_*$ is conditionally independent. Then, the probability distribution of y_* is given by

$$y_* \mid \boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{k}_{M*}^\top \boldsymbol{Q}^{-1}\boldsymbol{K}_{NM}\boldsymbol{\Lambda}^{-1}\boldsymbol{y}, \; k_{**} - \boldsymbol{k}_{M*}^\top(\boldsymbol{K}_{MM}^{-1} - \boldsymbol{Q}^{-1})\boldsymbol{k}_{M*}\right).$$

Here, we define

$$\begin{aligned}
\boldsymbol{K}_{NN} &:= \{k(x_i, x_j) + \delta_{ij}\sigma^2\}_{j=1\dots N}^{i=1\dots N} \\
\boldsymbol{K}_{NM} &:= \{k(x_i, \hat{x}_j)\}_{j=1\dots M}^{i=1\dots N} \\
\boldsymbol{K}_{MM} &:= \{k(\hat{x}_i, \hat{x}_j)\}_{j=1\dots M}^{i=1\dots M} \\
\boldsymbol{k}_{M*} &:= \{k(\hat{x}_i, x_*)\}_{i=1\dots M} \\
k_{**} &:= k(x_*, x_*) \\
\boldsymbol{\Lambda} &:= \mathrm{diag}[\boldsymbol{K}_{NN} - \boldsymbol{K}_{NM}\boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM}^\top] \\
\boldsymbol{Q} &:= \boldsymbol{K}_{MM} + \boldsymbol{K}_{NM}^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{K}_{NM}
\end{aligned}$$

with kernel function $k(\cdot,\cdot)$. This calculation is $O(M^2N)$, which significantly decreases the computation burden if $M \ll N$. Note that the choice of the inducing variable $\hat{\boldsymbol{x}}$ largely affects the approximation accuracy. It is empirically known that the sparse Gaussian process provides better approximation around inducing variables. This motivates us to take inducing variables in "important" region.

As in the case of model reduction of linear systems based on the controllability Gramian, let us assume that states that are reachable by a less energy input are more important. In other words, it seems reasonable that inducing variables are taken in accordance with the probability distribution given by the controllability function.

4.1 Numerical simulation (continued)

Let us revisit the GPSSM constructed in Section 3.3 with $N = 500$. Then, we generate two sparse Gaussian processes with $M = 50$; inducing variables $\hat{\boldsymbol{x}}_i$ are generated randomly with

- the uniform distribution over S in (14),
- the probability density function given by the following Gaussian Mixture:

$$\frac{1}{100}\sum_{k=0}^{99}\mathcal{N}(\boldsymbol{x}; \mathbf{0}, \mathcal{G}_\tau). \tag{16}$$

Figs. 3 and 4 show the probability density and generated inducing variables of the proposed method.

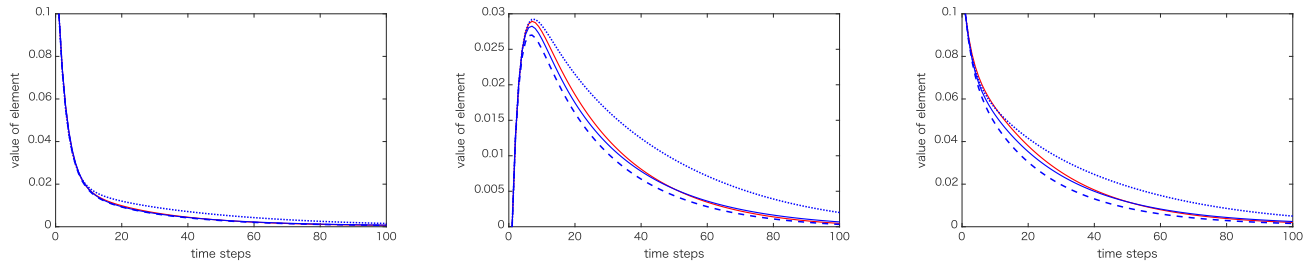
In order to compare the resulting accuracy, let $\boldsymbol{\mu_f}(\boldsymbol{x})$ and $\boldsymbol{\Sigma_f}(\boldsymbol{x})$ be the mean vector and covariance matrix of the *a posteriori* distribution obtained by the original and sparsified Gaussian processes of $\boldsymbol{f}$. In other words, the likelihood function of the vector field at $\boldsymbol{x}$ is given by

$$\mathcal{N}(\boldsymbol{f}; \boldsymbol{\mu_f}(\boldsymbol{x}), \boldsymbol{\Sigma_f}(\boldsymbol{x})). \tag{17}$$

Figs. 5 to 7 show the likelihood of the *real* dynamics, i.e., $\mathcal{N}(\boldsymbol{Ax}; \boldsymbol{\mu_f}(\boldsymbol{x}), \boldsymbol{\Sigma_f}(\boldsymbol{x}))$. Roughly speaking, larger likelihood implies that the obtained model generates the real dynamics with higher probability. Since M is much smaller than N, it is inevitable that the accuracy deteriorates upon the sparsification. However, the proposed method provides a model whose accuracy on the effectively reachable region retains relatively high in comparison to the sparsification by the uniform distribution, compromising the accuracy on other domain.

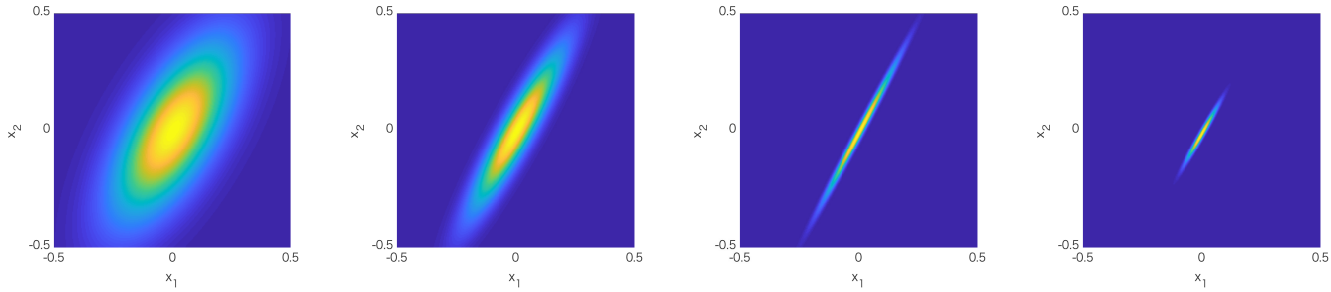
5. CONCLUSION

In this paper, following the discussion in Kashima (2016), we newly introduced a novel controllability Gramian, and proposed its approximate computation method. The characteristic of several definitions of controllability Gramian are summarized in Table 1. At the moment, we have not yet obtained rigorous theoretical underpinnings for Definition 3.1; see also Todorov (2009) for discrete-time linearly solvable Markov decision processes. It is remarkable feature of Gaussian process state space models that the pro-



(a) $(1,1)$-element (b) $(1,2)$-element (c) $(2,2)$-element

Fig. 2. Comparison of the controllability Gramians; horizontal axis: τ, red (solid): $\mathbf{G}_\tau$, blue: $\mathcal{G}_\tau$ for $N = 20$ (dotted), 100 (dashed), 500 (solid).



(a) $\tau = 5$ (b) $\tau = 10$ (c) $\tau = 20$ (d) $\tau = 100$

Fig. 3. Effectively reachable region estimated by the cotrollability Gramian: Heat map of $\mathcal{N}(\boldsymbol{x}; \mathbf{0}, \mathcal{G}_\tau)$ for $\tau = 5, 10, 20, 100$.

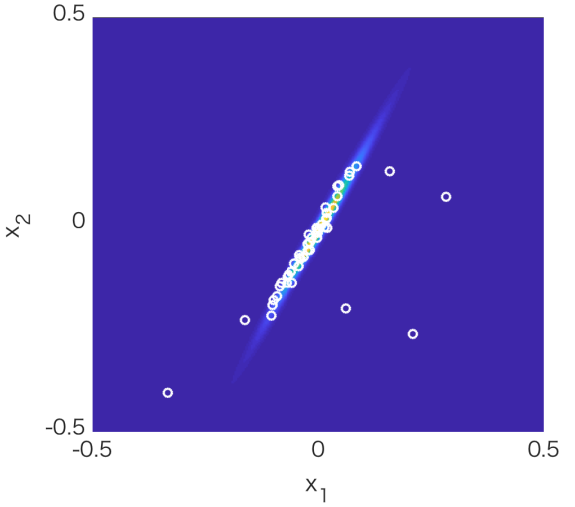


Fig. 4. Inducing variables of the proposed method: scattering plot with heat map of the probability density function in (16).

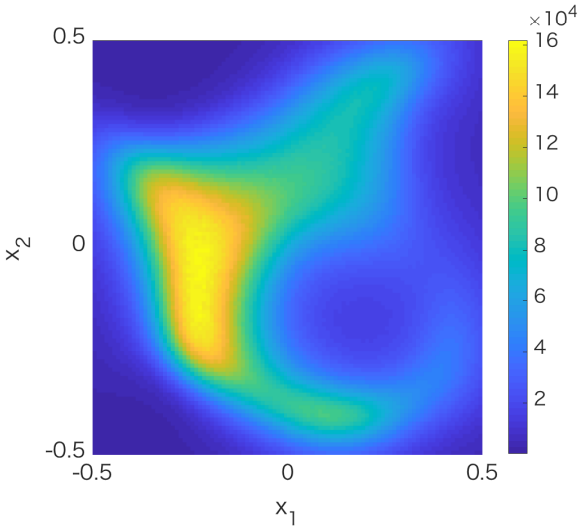


Fig. 5. Heatmap of likelihood of real dynamics: original GP.

posed method is computationally efficient and requires no Monte Carlo sampling despite the dynamics' nonlinearity.

In linear control system theory, controllability (and observability) plays a crucial role for model reduction. In Section 4, the newly introduced controllability Gramian was utilized for model sparsification. The numerical results performed suggests this is promising direction. Theoretical justification of the usefulness of the controllability Gramian-based sparsification, including the effect of the moment matching approximation and overconfidence issue (Quiñonero-Candela et al. (2007)) and introduction of observability Gramian are currently under investigation.

REFERENCES

Antoulas, A.C. (2005). *Approximation of Large-Scale Dynamical Systems*. SIAM.

Cleveland, W.S. and Devlin, S.J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403), 596–610. doi:10.1080/01621459.1988.10478639.

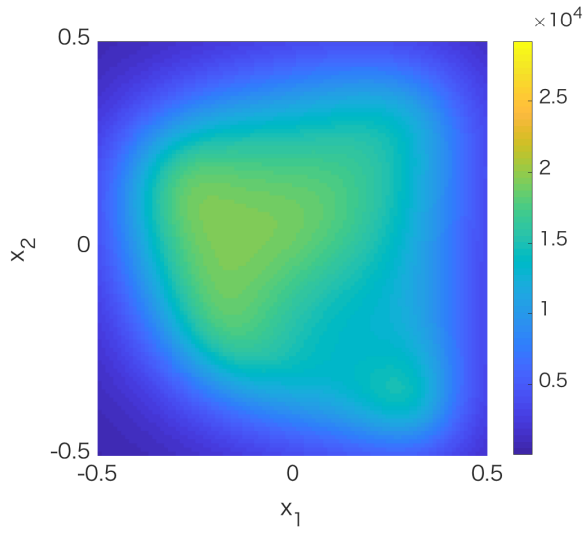


Fig. 6. Heatmap of likelihood of real dynamics: sparsified GP with the uniform distribution.

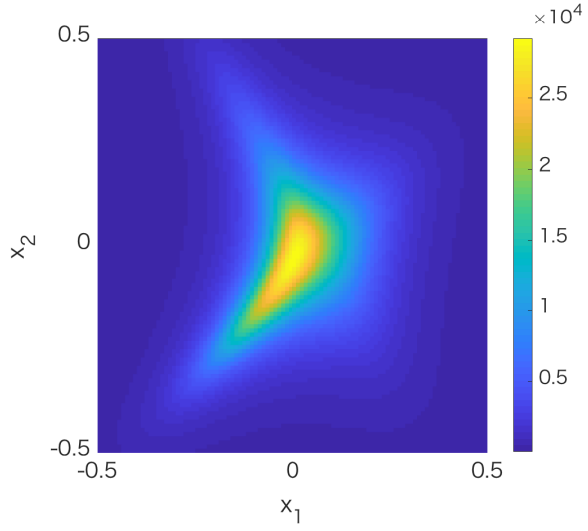


Fig. 7. Heatmap of likelihood of real dynamics: sparsified GP with the proposed distribution.

Deisenroth, M.P. and Rasmussen, C.E. (2011). PILCO: A model-based and data-efficient approach to policy search. *Proceedings of the International Conference on Machine Learning*, 465–472.

Eleftheriadis, S., Nicholson, T.F., Deisenroth, M.P., and Hensman, J. (2017). Identification of Gaussian process state space models. *Advances in Neural Information Processing Systems*, 2017-Decem(2014), 5310–5320.

Girard, A., Rasmussen, C.E., Candela, J.Q., and Murray-Smith, R. (2003). Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting. *Advances in Neural Information Processing Systems*, (December 2002), 9–14.

Kappen, H.J. (2005). Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11), P11011–P11011. doi:10.1088/1742-5468/2005/11/P11011.

Kashima, K. (2016). Noise response data reveal novel controllability Gramian for nonlinear network dynamics. *Scientific Reports*, 6, 27300. doi:10.1038/srep27300.

Khalil, H.K. (2002). *Nonlinear Systems*. Prentice Hall, 3rd edition.

Lall, S., Marsden, J.E., and Glavaški, S. (2002). A subspace approach to balanced truncation for model reduction of nonlinear control systems. *International Journal of Robust and Nonlinear Control*, 12(September 2000), 519–535. doi:10.1002/rnc.657.

McHutchon, A. (2014). *Nonlinear Modelling and Control using Gaussian Processes*. Ph.D. thesis, University of Cambridge.

Quiñonero-Candela, J., Rasmussen, C.E., and Williams, C.K.I. (2007). Approximation Methods for Gaussian Process Regression. In *Large-Scale Kernel Machines*, 203–223. The MIT Press. doi:10.7551/mitpress/7496.003.0011.

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*, volume 14. doi:10.1142/S0129065704001899.

Snelson, E. and Ghahramani, Z. (2006). Sparse Gaussian processes using pseudo-inputs. In Y. Weiss, B. Schölkopf, and J.C. Platt (eds.), *Advances in Neural Information Processing Systems 18*, 1257–1264. MIT Press.

Stenman, A., Gustafsson, F., and Ljung, L. (1996). Just in time models for dynamical systems. In *Proceedings of 35th IEEE Conference on Decision and Control*, volume 1, 1115–1120. IEEE. doi:10.1109/CDC.1996.574658.

Todorov, E. (2009). Efficient computation of optimal actions. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28), 11478–11483. doi:10.1073/pnas.0710743106.

Wang, J., Hertzmann, A., and Blei, D.M. (2006). Gaussian process dynamical models. In Y. Weiss, P.B. Schölkopf, and J.C. Platt (eds.), *Advances in Neural Information Processing Systems 18*, 1441–1448. MIT Press.

Zheng, Q. and Kimura, H. (2008). Just-in-Time modeling for function prediction and its applications. *Asian Journal of Control*, 3(1), 35–44. doi:10.1111/j.1934-6093.2001.tb00040.x.

Zhou, K., Doyle, J.C., and Glover, K.K. (1996). *Robust and Optimal Control*. Prentice Hall.