# Physically Consistent Modeling & Identification of Nonlinear Friction with Dissipative Gaussian Processes

**Rui Dai**[†,*]                                                                RUI.DAI@IIT.IT

**Giulio Evangelisti**[*]                                           GIULIO.EVANGELISTI@TUM.DE

**Sandra Hirche**[*]                                                         HIRCHE@TUM.DE

[†]*Humanoids and Human-Centered Mechatronics Research Line, Istituto Italiano di Tecnologia (IIT), 16163 Genova, Italy.*

[*]*Chair of Information-oriented Control, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany*

**Editors:** A. Abate, K. Margellos, A. Papachristodoulou

## Abstract

Friction modeling has always been a challenging problem due to the complexity of real physical systems. Although a few state-of-the-art structured data-driven methods show their efficiency in nonlinear system modeling, deterministic passivity as one of the significant characteristics of friction is rarely considered in these methods. To address this issue, we propose a Gaussian Process based model that preserves the inherent structural properties such as passivity. A matrix-vector physical structure is considered in our approaches to ensure physical consistency, in particular, enabling a guarantee of positive semi-definiteness of the damping matrix. An aircraft benchmark simulation is employed to demonstrate the efficacy of our methodology. Estimation accuracy and data efficiency are increased substantially by considering and enforcing more structured physical knowledge. Also, the fulfillment of the dissipative nature of the aerodynamics is validated numerically.

**Keywords:** Friction Identification, Gaussian Process, Passivity, Dissipativity.

## 1. Introduction

Passive systems are a class of dynamical systems that can only increase stored energy when there is external energy inflow. This property holds practical significance in Euler-Lagrangian systems, including robots, aircraft, vehicles, etc. For mechanical systems, friction is a direct source of dissipation. Modeling with a guarantee of its passivity can help ensure the system's stability by the passivity theorem (Khalil, 2002). In addition, its accurate modeling is also beneficial in improving control accuracy by friction compensation (Olsson et al., 1998).

For decades, researchers have constructed many analytical models of friction that can characterize specific characteristics, from static to dynamic (Bliman and Sorine, 1995; De Wit et al., 1995; Lampaert et al., 2002; Feng et al., 2019). However, the characteristics are usually nonlinear, and the behavior of friction may vary significantly under different conditions, such as hysteretic effects (Wojewoda et al., 2008). There is no universal physical model that can be applied to all situations. As a promising alternative, several intelligent model-free computing approaches can capture the complexity and nonlinear properties of a system by analyzing observations from the environments (Huang et al., 2019; Peng et al., 2022; Auriol et al., 2022; Dong et al., 2021). However, modeling the dynamics from data is often challenging because of the high dimensionality of the real-world systems (Geist and Trimpe, 2021). In addition, data collection on physical systems is expensive and time-consuming. To address this issue, recent efforts strive to integrate physical structure as prior knowledge with data-driven modeling to improve learning efficiency.

## 1.1. Related Work

The structured models inherit the potential to combine the advantages of both analytical and data-driven modeling while avoiding their limitations when used individually (Geist and Trimpe, 2021; Nelles, 2020). Rigid-body dynamics represent a valuable source of prior structural knowledge for mechanical systems (Hwangbo et al., 2019). Nguyen-Tuong and Peters (2010) introduce two semi-parametric regression approaches, where the physical model knowledge can be incorporated into either the mean function or the kernel of a nonparametric Gaussian Process Regression (GPR). Their findings demonstrate that semiparametric models, trained with rigid body dynamics as a prior, outperform standard rigid body dynamics models when applied to real-world data. Physical constraints are integrated into the learning process by introducing structure to the learned parameters in Sutanto et al. (2020). This framework enables the acquisition of physically plausible dynamics through gradient descent, enhancing both training speed and the generalization of learned dynamic models. In addition, structured learning frameworks can guarantee physical properties such as energy conservation of Lagrangian systems (Evangelisti and Hirche, 2022). This Lagrangian-GP-based learning framework is extended to include dissipative systems in Evangelisti and Hirche (2024) but a passivity preserving guarantee is not given. In the literature, passivity as a significant phenomenon in mechanical systems is rarely considered. Koch et al. (2020) verifies dissipativity of linear time-invariant systems from noisy data with guaranteed robustness and the sector bounds which can probabilistically show the passivity is derived in Fiedler et al. (2021). However, to the best knowledge of the authors, deterministic guarantee of passivity from data for nonlinear systems has not been investigated.

## 1.2. Contribution and Structure

The main contribution of our work is proposing a physically structured Gaussian Process that preserves passivity for friction modelling and identification. We consider a matrix-vector structure for the prior mean and kernels in GP, with a full and diagonal damping matrix, respectively. We investigate their passivity guarantee of the damping estimates by showing the positive semi-definiteness of the damping matrix.

After introducing the formulation of our problem in Section 2, we review the basic GP framework in Section 3. In Section 4, structured dissipative GP that provide passivity guarantees are proposed. We evaluate our approach and discuss the results in Section 5 and conclude in Section 6.

## 1.3. Notation

We use bold lower/upper case characters to denote vectors/matrices, respectively. $\boldsymbol{I}$ denotes the identity. $[\boldsymbol{x}_i]_{i=1}^{\mathcal{D}} \in \mathbb{R}^{N\mathcal{D}}$ denotes a stacked vector of $\mathcal{D}$ indexed vectors $\boldsymbol{x}_i \in \mathbb{R}^N$ and $[\boldsymbol{X}_i]_{i=1}^{\mathcal{D}} \in \mathbb{R}^{N\mathcal{D} \times N}$ a stacked matrix of $\mathcal{D}$ indexed matrices $\boldsymbol{X}_i \in \mathbb{R}^{N \times N}$. The direct sum is denoted as $\oplus$. We use $\succeq$ and $\preceq$ to denote matrix inequality.

## 2. Problem Formulation and Passivity

Our work considers Euler-Lagrange systems (Ortega et al., 1998) whose equations of motion are given by

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) + \boldsymbol{D}(\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} = \boldsymbol{\tau}, \tag{1}$$

2

where $q \in \mathbb{R}^N$ is the generalized coordinate, $M(q) \in \mathbb{R}^{N \times N}$ the positive definite inertia matrix, $C(q, \dot{q}) \in \mathbb{R}^{N \times N}$ the Coriolis matrix, $g(q) \in \mathbb{R}^N$ the generalized potential forces and $D(\dot{q}) \in \mathbb{R}^{N \times N}$ the damping matrix. Note that we neglect the positional dependence of the damping matrix to prevent notational complexity. $\tau \in \mathbb{R}^N$ is the input actuation applied to the system. This formulation can be applied to a large varieties of rigid body systems including robots, aircraft, etc. According to the definition of passivity in Willems (1972), passivity of this system can be guaranteed if the damping matrix $D(\dot{q})$ is positive semidefinite, i.e. $\dot{q}^\top D(\dot{q})\dot{q} \geq 0$.

## 3. Gaussian Process Framework

We first review the Gaussian Process regression framework based on Rasmussen and Williams (2006) in this section. A Gaussian process is a collection of random variables and any finite number of which have a joint Gaussian distribution. It can be used to describe a distribution over functions and consider inference in a function space.

Considering $N$-dimensional inputs $x, x' \in \mathcal{X}$ in a continuous domain $\mathcal{X} \subseteq \mathbb{R}^N$, and a real process $f(x) \in \mathbb{R}$, a GP is completely specified by its mean $m(x) = \mathbb{E}[f(x)]$ and covariance function $k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$, written as $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$. The mean function provides the prior knowledge of the unknown function, while the kernel determines higher-level functional properties such as smoothness. It specifies the similarity between two values of the function evaluated on two specific inputs.

The GP inherits the Gaussian distribution properties and we can infer the function from the given observations. Considering $\mathcal{D}$ finite noisy observations $y = [y_d]_{d=1}^{\mathcal{D}}$ with $y_d = f(x_d) + \varepsilon$, where $\varepsilon$ is an additive i.i.d. Gaussian noise process with variance $\sigma_\varepsilon^2$, then the covariance of the noisy observations is $\mathrm{cov}(y) = K(X, X) + \Sigma = K(X, X) + \sigma_\varepsilon^2 I$, where $K(X, X) \in \mathbb{R}^{\mathcal{D} \times \mathcal{D}}$ denotes the covariance evaluated at all pairs of training points. The joint Gaussian distribution of the observations and the test points can be written as

$$\begin{bmatrix} y \\ f(X_*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma_\varepsilon^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right), \tag{2}$$

where $m(\cdot)$ denotes the prior mean of the function, $K(X, X_*)$ denotes the covariance evaluated at all pairs of training data and test data, $K(X_*, X_*)$ denotes the covariance evaluated at all pairs of test data and $K(X_*, X) = K^\top(X, X_*)$. The posterior conditional distribution based on the prior distribution on the observations can be obtained by $f(X_*) \mid X, y, X_* \sim \mathcal{N}(\overline{f}(X_*), \mathrm{cov}(f(X_*)))$, where $\overline{f}(X_*) = m(X_*) + K(X_*, X)[K(X, X) + \sigma_\varepsilon^2 I]^{-1}(y - m(X))$, $\mathrm{cov}(f(X_*)) = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_\varepsilon^2 I]^{-1} K(X, X_*)$.

## 4. Dissipative Gaussian Processes

After introducing the passive system and the basic GP framework, we move on to our contribution of a physically consistent GP enforcing a damping torque with a matrix-vector form. Passivity is guaranteed analytically via a sufficient condition for ensuring positive semidefiniteness of the damping matrix. For this, we analyze the posterior mean estimates of a full, and then diagonal, damping matrix.

## 4.1. Matrix-vector Structured Gaussian Processes for Full Damping Matrices

In this section, we first investigate the general case of a full damping matrix.

### 4.1.1. GP STRUCTURE AND FORMULATION

For the unknown damping term in the physical dynamic system (1), we assume the independence of the components in the damping matrix.

**Assumption 1** *Consider a damping torque vector in the form $\boldsymbol{\tau}^d(\dot{\boldsymbol{q}}) = \boldsymbol{D}(\dot{\boldsymbol{q}})\dot{\boldsymbol{q}}$, where $\dot{\boldsymbol{q}} \in \mathbb{R}^N$ is the generalized velocity. The elements $d_{mn}(\dot{\boldsymbol{q}})$ of the damping matrix $\boldsymbol{D}(\dot{\boldsymbol{q}}) = [d_{mn}(\dot{\boldsymbol{q}})]_{m,n=1}^N \in \mathbb{R}^{N \times N}$ are stochastically independent, i.e., $\mathbb{E}[d_{mn}d_{ij}] = \mathbb{E}[d_{mn}]\mathbb{E}[d_{ij}]$ for $m, n \neq i, j$.*

This assumption is reasonable because we can choose a suitable input coordinate such that each dimension is independent and the torque is in its span. We model each element $d_{mn}(\dot{\boldsymbol{q}})$ independently

$$d_{mn}(\dot{\boldsymbol{q}}) \sim \mathcal{GP}\left(m_{d_{mn}}(\dot{\boldsymbol{q}}), k_{d_{mn}}\left(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}'\right)\right). \tag{3}$$

**Lemma 1** *Under Ass. 1 and the independent GP models (3), denoting $\boldsymbol{k}_{d_n} = [k_{d_{mn}}\left(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}'\right)]_{m=1}^N$ as the n-th column of the matrix kernel associated with $\boldsymbol{D}(\dot{\boldsymbol{q}})$, the covariance of the damping torque $\boldsymbol{\tau}^d(\dot{\boldsymbol{q}})$ is*

$$\boldsymbol{K}_{\boldsymbol{\tau}}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}') = \mathrm{cov}(\boldsymbol{\tau}^d(\dot{\boldsymbol{q}})) = \sum_{n=1}^N \dot{q}_n \dot{q}_n' \mathrm{diag}(\boldsymbol{k}_{d_n}) \tag{4}$$

**Proof** Define a mean matrix $\boldsymbol{M_D}(\dot{\boldsymbol{q}}) = [m_{d_{mn}}(\dot{\boldsymbol{q}})]_{m,n=1}^N$ corresponding to $\boldsymbol{D}(\dot{\boldsymbol{q}})$, then

$$\begin{aligned}
\mathrm{cov}(\boldsymbol{\tau}^d(\dot{\boldsymbol{q}})) &= \mathbb{E}\left[\left(\boldsymbol{D}(\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} - \boldsymbol{M_D}(\dot{\boldsymbol{q}})\dot{\boldsymbol{q}}\right)\left(\boldsymbol{D}(\dot{\boldsymbol{q}}')\dot{\boldsymbol{q}}' - \boldsymbol{M_D}(\dot{\boldsymbol{q}}')\dot{\boldsymbol{q}}'\right)^\top\right] \\
&= \mathbb{E}\left[\left(\sum_{n=1}^N \dot{q}_n \left(\boldsymbol{d}_{\cdot n}(\dot{\boldsymbol{q}}) - \boldsymbol{m}_{\boldsymbol{d}_{\cdot n}}(\dot{\boldsymbol{q}})\right)\right)\left(\sum_{m=1}^N \dot{q}_m' \left(\boldsymbol{d}_{\cdot m}(\dot{\boldsymbol{q}}) - \boldsymbol{m}_{\boldsymbol{d}_{\cdot m}}(\dot{\boldsymbol{q}})\right)\right)^\top\right] \\
&= \sum_{n=1}^N \sum_{m=1}^N \dot{q}_n \dot{q}_m' \mathbb{E}\left[\left(\boldsymbol{d}_{\cdot n}(\dot{\boldsymbol{q}}) - \boldsymbol{m}_{\boldsymbol{d}_{\cdot n}}(\dot{\boldsymbol{q}})\right)\left(\boldsymbol{d}_{\cdot m}(\dot{\boldsymbol{q}}) - \boldsymbol{m}_{\boldsymbol{d}_{\cdot m}}(\dot{\boldsymbol{q}})\right)^\top\right] \\
&= \sum_{n=1}^N \dot{q}_n \dot{q}_n' \mathbb{E}\left[\left(\boldsymbol{d}_{\cdot n}(\dot{\boldsymbol{q}}) - \boldsymbol{m}_{\boldsymbol{d}_{\cdot n}}(\dot{\boldsymbol{q}})\right)\left(\boldsymbol{d}_{\cdot n}(\dot{\boldsymbol{q}}) - \boldsymbol{m}_{\boldsymbol{d}_{\cdot n}}(\dot{\boldsymbol{q}})\right)^\top\right] \\
&= \sum_{n=1}^N \dot{q}_n \dot{q}_n' \mathrm{diag}(\boldsymbol{k}_{d_n})
\end{aligned} \tag{5}$$

∎

Let us model the torques as a GP

$$\boldsymbol{\tau}^d(\dot{\boldsymbol{q}}) \sim \mathcal{GP}\left(\boldsymbol{m}_{\boldsymbol{\tau}}(\dot{\boldsymbol{q}}), \boldsymbol{K}_{\boldsymbol{\tau}}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}')\right), \tag{6}$$

with the structured prior mean $\boldsymbol{m}_{\boldsymbol{\tau}}(\dot{\boldsymbol{q}}) = \boldsymbol{M_D}(\dot{\boldsymbol{q}})\dot{\boldsymbol{q}}$ and kernel (4). Assume access to a number of $\mathcal{D}$ noise-free generalized velocity measurements $\dot{\boldsymbol{Q}} = [\dot{\boldsymbol{q}}_i^\top]_{i=1}^{\mathcal{D}}$ and noise-corrupted output measurements $\boldsymbol{Y} = [\boldsymbol{y}_i^\top]_{i=1}^{\mathcal{D}}$, where $\boldsymbol{y}_i = \boldsymbol{\tau}_i^d(\dot{\boldsymbol{q}}_i) + \boldsymbol{\varepsilon}_i$, $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I})$. Introducing a stacked vector of

outputs $\boldsymbol{y} = \operatorname{vec}(\boldsymbol{Y}^\top)$ with mean $\boldsymbol{m_y} = [\boldsymbol{m_\tau}(\dot{\boldsymbol{q}}_i)]_{i=1}^{\mathcal{D}}$, then the joint distribution of observations and desired estimate with velocity $\dot{\boldsymbol{q}}_*$ is given by

$$\left[ \begin{array}{c} \boldsymbol{y} \\ \boldsymbol{\tau}^d(\dot{\boldsymbol{q}}_*) \end{array} \right] \sim \mathcal{N} \left( \left[ \begin{array}{c} \boldsymbol{m_y} \\ \boldsymbol{m_\tau}(\dot{\boldsymbol{q}}_*) \end{array} \right], \left[ \begin{array}{cc} \boldsymbol{K_y}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}}) & \boldsymbol{K_\tau}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}}_*) \\ \boldsymbol{K_\tau^\top}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}}_*) & \boldsymbol{K_\tau}(\dot{\boldsymbol{q}}_*, \dot{\boldsymbol{q}}_*), \end{array} \right] \right), \tag{7}$$

where

$$\boldsymbol{K_y}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}}) = [\boldsymbol{K_\tau}(\dot{\boldsymbol{q}}_i, \dot{\boldsymbol{q}}_j)]_{i,j=1}^{\mathcal{D}} + \sigma_\varepsilon^2 \boldsymbol{I} \tag{8a}$$

$$\boldsymbol{K_\tau}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}}_*) = [\boldsymbol{K_\tau}(\dot{\boldsymbol{q}}_i, \dot{\boldsymbol{q}}_*)]_{i=1}^{\mathcal{D}}. \tag{8b}$$

### 4.1.2. POSITIVE SEMIDEFINITENESS

Before we show the guarantee for passivity of this specific structured GP, two minor restrictions on the used kernel functions and the structure of the prior mean need to be imposed.

**Assumption 2** *Each kernel component is bounded by $\sigma_{f_{mn}}^2$, respectively, i.e. $|k_{d_{mn}}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}')| \leq \sigma_{f_{mn}}^2$.*

**Assumption 3** *The prior mean matrix $\boldsymbol{M_D}(\dot{\boldsymbol{q}})$ is diagonal and positive semi-definite, i.e. $\boldsymbol{m_\tau}(\dot{\boldsymbol{q}}) = \operatorname{diag}(\boldsymbol{m_d}(\dot{\boldsymbol{q}}))\dot{\boldsymbol{q}}$ with non-negative $m_{d_m}(\dot{\boldsymbol{q}}) \geq 0$.*

Note that Assumption 2 includes a wide variety of covariance functions including the commonly used squared exponential (SE) kernel and Assumption 3 can be easily constructed by the prior mean. Then, according to the conditional distribution of the mean estimate in Section 3, we obtain

$$\hat{\boldsymbol{\tau}}^d(\dot{\boldsymbol{q}}) = \operatorname{diag}(\boldsymbol{m_d}(\dot{\boldsymbol{q}}))\dot{\boldsymbol{q}} + \sum_{n=1}^{N} \dot{q}_n \boldsymbol{K_n^\top}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}}) \tilde{\boldsymbol{Q}}_n \boldsymbol{K_y^{-1}}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})(\boldsymbol{y} - \boldsymbol{m_y}), \tag{9}$$

where $\boldsymbol{K_n}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}}) = [\boldsymbol{K_n}(\dot{\boldsymbol{q}}_i, \dot{\boldsymbol{q}})]_{i=1}^{D}$ with $\boldsymbol{K_n}(\dot{\boldsymbol{q}}_i, \dot{\boldsymbol{q}}) = \operatorname{diag}(\boldsymbol{k}_{d_n}(\dot{\boldsymbol{q}}_i, \dot{\boldsymbol{q}}))$. $\tilde{\boldsymbol{Q}}_n = \oplus_{i=1}^{\mathcal{D}}(\tilde{q}_{n,i})\boldsymbol{I}$ with $\boldsymbol{I} \in \mathbb{R}^{N \times N}$, and $\tilde{q}_{n,i}$ is the $(i, n)$-th element in $\dot{\boldsymbol{Q}}$. We also denote $\Delta \boldsymbol{y} = \boldsymbol{y} - \boldsymbol{m_y}$ and $\dot{\tilde{\boldsymbol{q}}} = \operatorname{vec}(\dot{\boldsymbol{Q}}^\top)$.

**Theorem 1** *Consider the damping torque estimate (9) under Ass. 2 and 3. Positive semidefiniteness of the matrix estimate $\hat{\boldsymbol{D}}(\dot{\boldsymbol{q}})$ is guaranteed deterministically if*

$$\boldsymbol{\Sigma}_f \preceq \frac{\sigma_\varepsilon^2}{\sqrt{\mathcal{D}}\|\dot{\tilde{\boldsymbol{q}}}\|_\infty \|\Delta \boldsymbol{y}\|} \operatorname{diag}(\boldsymbol{m_d}(\dot{\boldsymbol{q}})) \tag{10}$$

*holds $\forall \dot{\boldsymbol{q}} \in \mathbb{R}^N$ with the hypervariance matrix $\boldsymbol{\Sigma}_f = [\,|\sigma_{f_{mn}}|\,]_{m,n=1}^N$.*

**Proof** According to (9), we can compute the dissipated power as

$$\dot{\boldsymbol{q}}^\top \hat{\boldsymbol{\tau}}^d(\dot{\boldsymbol{q}}) = \dot{\boldsymbol{q}}^\top \operatorname{diag}(\boldsymbol{m_d}(\dot{\boldsymbol{q}}))\dot{\boldsymbol{q}} + \sum_{m=1}^{N} \sum_{n=1}^{N} \dot{q}_m \dot{q}_n \underbrace{\boldsymbol{k}_{d_{mn}}^\top(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}}) \tilde{\boldsymbol{Q}}_n}_{=\tilde{\boldsymbol{k}}_{mn}^\top} \underbrace{\boldsymbol{K_y^{-1}}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}}) \Delta \boldsymbol{y}}_{=\Delta \boldsymbol{x}}. \tag{11}$$

with $\boldsymbol{k}_{d_{mn}}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})$ the m-th column of $\boldsymbol{K_n}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})$. Under Assumption 2, we can obtain

$$\|\tilde{\boldsymbol{k}}_{mn}\| = \|\boldsymbol{k}_{d_{mn}}^\top(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}}) \tilde{\boldsymbol{Q}}_n\| \leq \|\boldsymbol{k}_{d_{mn}}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})\| \|\tilde{\boldsymbol{Q}}_n\| \leq \sqrt{\mathcal{D}} |\sigma_{f_{mn}}| \|\dot{\tilde{\boldsymbol{q}}}\|_\infty. \tag{12}$$

Through the matrix norm inequality (Malek-Shahmirzadi, 1983),

$$\|\Delta \boldsymbol{x}\| = \|\boldsymbol{K}_{\boldsymbol{y}}^{-1}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})\Delta \boldsymbol{y}\| \leq \|\boldsymbol{K}_{\boldsymbol{y}}^{-1}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})\|\|\Delta \boldsymbol{y}\| \leq \sigma_{\max}\left(\boldsymbol{K}_{\boldsymbol{y}}^{-1}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})\right)\|\Delta \boldsymbol{y}\| \quad (13)$$

where $\sigma_{\max}$ denotes the maximum singular value of a matrix. Since $\boldsymbol{K}_{\boldsymbol{y}}^{-1}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})$ is positive definite and symmetric, its eigenvalues are equal to its singular values. Based on Weyl's inequality (Franklin, 2012) and positive semidefiniteness of the Gram matrix,

$$\lambda_{\min}\left(\boldsymbol{K}_{\boldsymbol{y}}\right) = \lambda_{\min}\left(\boldsymbol{K}_{\boldsymbol{\tau}} + \sigma_{\varepsilon}^2 \boldsymbol{I}\right) \geq \lambda_{\min}\left(\boldsymbol{K}_{\boldsymbol{\tau}}\right) + \lambda_{\min}\left(\sigma_{\varepsilon}^2 \boldsymbol{I}\right) \geq \lambda_{\min}\left(\sigma_{\varepsilon}^2 \boldsymbol{I}\right) = \sigma_{\varepsilon}^2. \quad (14)$$

We neglect the variable dependency here for notation simplicity. Since $\lambda_{\max}\left(\boldsymbol{A}^{-1}\right) = \frac{1}{\lambda_{\min}(\boldsymbol{A})}$ for regular $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ then

$$\|\Delta \boldsymbol{x}\| \leq \frac{1}{\lambda_{\min}\left(\boldsymbol{K}_{\boldsymbol{y}}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})\right)} \|\Delta \boldsymbol{y}\| \leq \frac{\|\Delta \boldsymbol{y}\|}{\sigma_{\varepsilon}^2}$$
$$\tilde{\boldsymbol{k}}_{mn}^{\top}\Delta \boldsymbol{x} \leq |\tilde{\boldsymbol{k}}_{mn}^{\top}\Delta \boldsymbol{x}| \leq \|\tilde{\boldsymbol{k}}_{mn}\|\|\Delta \boldsymbol{x}\| \leq \sqrt{\mathcal{D}}\|\tilde{\dot{\boldsymbol{q}}}\|_{\infty} \|\Delta \boldsymbol{y}\| \frac{|\sigma_{f_m}|}{\sigma_{\varepsilon}^2}. \quad (15)$$

Next, we derive from $\dot{\boldsymbol{q}}^{\top}\hat{\boldsymbol{\tau}}^d \geq \boldsymbol{0}$ the inequality $\boldsymbol{M_D} \succeq [\tilde{\boldsymbol{k}}_{mn}^{\top}\Delta \boldsymbol{x}]_{mn}$. Broken up further, this gives

$$\dot{\boldsymbol{q}}^{\top}\hat{\boldsymbol{\tau}}^d(\dot{\boldsymbol{q}}) \geq \dot{\boldsymbol{q}}^{\top}\mathrm{diag}(\boldsymbol{m_d}(\dot{\boldsymbol{q}}))\dot{\boldsymbol{q}} - \sum_{m=1}^{N}\sum_{n=1}^{N}|\dot{q}_m \dot{q}_n \tilde{\boldsymbol{k}}_{mn}^{\top}\Delta \boldsymbol{x}|$$
$$\geq \dot{\boldsymbol{q}}^{\top}\mathrm{diag}(\boldsymbol{m_d}(\dot{\boldsymbol{q}}))\dot{\boldsymbol{q}} - \sum_{m=1}^{N}\sum_{n=1}^{N}|\dot{q}_m||\dot{q}_n|\|\tilde{\boldsymbol{k}}_{mn}\|\|\Delta \boldsymbol{x}\|$$
$$\geq \dot{\boldsymbol{q}}^{\top}\mathrm{diag}(\boldsymbol{m_d}(\dot{\boldsymbol{q}}))\dot{\boldsymbol{q}} - \sqrt{\mathcal{D}}\|\tilde{\dot{\boldsymbol{q}}}\|_{\infty}\|\Delta \boldsymbol{y}\|\frac{1}{\sigma_{\varepsilon}^2}\sum_{m=1}^{N}\sum_{n=1}^{N}|\dot{q}_m||\dot{q}_n||\sigma_{f_{mn}}|$$
$$= |\dot{\boldsymbol{q}}|^{\top}\left(\mathrm{diag}(\boldsymbol{m_d}(\dot{\boldsymbol{q}})) - \frac{\sqrt{\mathcal{D}}\|\tilde{\dot{\boldsymbol{q}}}\|_{\infty}\|\Delta \boldsymbol{y}\|}{\sigma_{\varepsilon}^2}\boldsymbol{\Sigma}_f\right)|\dot{\boldsymbol{q}}| \geq 0 \quad (16)$$

Thus, passivity can be guaranteed by enforcing the matrix inequality (10), proving the result. ∎

Theorem 1 asserts that, when provided with observations of velocities and torques, ensuring guaranteed passivity involves selecting a basic kernel component, such as the SE kernel, whose bound can be regulated through hyperparameters $\sigma_{f_{mn}}$. This guarantee can be achieved by opting for either a sufficiently large prior mean or noise variance. However, the bound of (10) becomes tighter for increasing data set sizes. A specific conclusion can also be drawn when we contemplate a diagonal damping matrix.

## 4.2. Matrix-vector Structured Gaussian Processes for Diagonal Damping Matrices

### 4.2.1. GP STRUCTURE AND FORMULATION

In this section, we consider a specific restriction on the damping matrix where $\boldsymbol{D}(\dot{\boldsymbol{q}}) \in \mathbb{R}^{N \times N}$ is a diagonal damping matrix which can be written as $\boldsymbol{D}(\dot{\boldsymbol{q}}) = \mathrm{diag}(d_1(\dot{\boldsymbol{q}}), d_2(\dot{\boldsymbol{q}}), \ldots, d_N(\dot{\boldsymbol{q}}))$. Defining $\boldsymbol{d}(\dot{\boldsymbol{q}}) = [d_1(\dot{\boldsymbol{q}}), d_2(\dot{\boldsymbol{q}}), \ldots, d_N(\dot{\boldsymbol{q}})]^{\top}$, a vector-valued damping function can be modeled as

$$\boldsymbol{d}(\dot{\boldsymbol{q}}) \sim \mathcal{GP}\left(\boldsymbol{M_D}(\dot{\boldsymbol{q}}), \boldsymbol{K_d}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}')\right), \quad (17)$$

where $\boldsymbol{M_D}(\dot{\boldsymbol{q}}) = [m_{d_1}(\dot{\boldsymbol{q}}), \dots, m_{d_N}(\dot{\boldsymbol{q}})]^\top$ and $\boldsymbol{K_d}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}') = \mathrm{diag}(k_{d_1}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}'), \dots k_{d_N}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}'))$. We can write the dissipative torque as

$$\boldsymbol{\tau}^d(\dot{\boldsymbol{q}}) = \boldsymbol{D}(\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} = \mathrm{diag}(\boldsymbol{d}(\dot{\boldsymbol{q}}))\dot{\boldsymbol{q}} = \mathrm{diag}(\dot{\boldsymbol{q}})\boldsymbol{d}(\dot{\boldsymbol{q}}), \tag{18}$$

Defining a mean matrix $\boldsymbol{M_D}(\dot{\boldsymbol{q}}) = \mathrm{diag}(\boldsymbol{M_D}(\dot{\boldsymbol{q}}))$, the covariance of the damping torque $\boldsymbol{\tau}^d(\dot{\boldsymbol{q}})$ is

$$\begin{aligned}
\mathrm{cov}(\boldsymbol{\tau}^d(\dot{\boldsymbol{q}})) &= \mathbb{E}[(\boldsymbol{D}(\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} - \boldsymbol{M_D}(\dot{\boldsymbol{q}})\dot{\boldsymbol{q}})(\boldsymbol{D}(\dot{\boldsymbol{q}}')\dot{\boldsymbol{q}}' - \boldsymbol{M_D}(\dot{\boldsymbol{q}}')\dot{\boldsymbol{q}}')^\top] \\
&= \mathbb{E}[\mathrm{diag}(\dot{\boldsymbol{q}})(\boldsymbol{d}(\dot{\boldsymbol{q}}) - \boldsymbol{M_D}(\dot{\boldsymbol{q}}))(\boldsymbol{d}(\dot{\boldsymbol{q}}') - \boldsymbol{M_D}(\dot{\boldsymbol{q}}'))^\top \mathrm{diag}(\dot{\boldsymbol{q}}')^\top] \\
&= \mathrm{diag}(\dot{\boldsymbol{q}})\mathbb{E}[(\boldsymbol{d}(\dot{\boldsymbol{q}}) - \boldsymbol{M_D}(\dot{\boldsymbol{q}}))(\boldsymbol{d}(\dot{\boldsymbol{q}}') - \boldsymbol{M_D}(\dot{\boldsymbol{q}}'))^\top]\mathrm{diag}(\dot{\boldsymbol{q}}') \\
&= \mathrm{diag}(\dot{\boldsymbol{q}})\boldsymbol{K_d}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}')\mathrm{diag}(\dot{\boldsymbol{q}}')
\end{aligned} \tag{19}$$

The torque GP model and joint distribution are given analogously by (6)–(7) with different kernel

$$\boldsymbol{K_\tau}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}') = \mathrm{diag}(\dot{\boldsymbol{q}})\boldsymbol{K_d}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}')\mathrm{diag}(\dot{\boldsymbol{q}}'). \tag{20}$$

### 4.2.2. POSITIVE SEMIDEFINITENESS

Having introduced this diagonal structured GP formulation, we investigate its physical properties as positive semidefiniteness of damping matrix to guarantee passivity. Before that, We also first impose a trivial structural restriction on the used kernel functions.

**Assumption 4** *Each kernel component $k_{dn}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}')$ is bounded by $\sigma_{f_n}^2$, i.e. $|k_{d_n}(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}')| \le \sigma_{f_n}^2$.*

With the diagonal form of $\boldsymbol{D}(\dot{\boldsymbol{q}})$, the torque components are independent, so w.l.o.g. by marginalization, we can analyze one component of the torque estimation to show its property, which is

$$\hat{\tau}_n^d(\dot{\boldsymbol{q}}) = m_{d_n}(\dot{\boldsymbol{q}})\dot{q}_n + \dot{q}_n \boldsymbol{k}_{d_n}^\top(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})\mathrm{diag}(\tilde{\dot{\boldsymbol{q}}})\boldsymbol{K_y}^{-1}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})(\boldsymbol{y} - \boldsymbol{m_y}) \tag{21}$$

where $\boldsymbol{k}_{\tau_n}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})$ is the n-th colunm in $\boldsymbol{K_\tau}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})$, and $\boldsymbol{k}_{d_n}^\top(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})$ is the n-th colunm in $\boldsymbol{K_d}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})$.

**Corollary 1** *Consider the posterior estimate $\hat{\boldsymbol{D}}(\dot{\boldsymbol{q}})$ with kernels components under Ass. 4 and the damping torque estimate (21) for each torque component. Positive semidefiniteness of the matrix estimate $\hat{\boldsymbol{D}}(\dot{\boldsymbol{q}})$ is guaranteed deterministically if $\forall n = 1, \dots, N$,*

$$|\sigma_{f_n}| \le \frac{\sigma_\varepsilon^2}{\sqrt{D}\|\tilde{\dot{\boldsymbol{q}}}\|_\infty \|\Delta \boldsymbol{y}\|} m_{d_n}(\dot{\boldsymbol{q}}) \tag{22}$$

**Proof** According to (21), the dissipated power for each dimension is

$$\dot{q}_n \hat{\tau}_n^d(\dot{\boldsymbol{q}}) = \dot{q}_n^2(m_{d_n}(\dot{\boldsymbol{q}}) + \underbrace{\boldsymbol{k}_{d_n}^\top(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})\mathrm{diag}(\tilde{\dot{\boldsymbol{q}}})}_{\boldsymbol{k}_n^\top}\underbrace{\boldsymbol{K_y}^{-1}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})\Delta \boldsymbol{y}}_{\Delta \boldsymbol{x}}). \tag{23}$$

Similar as the proof in Theorem 1, We can obtain

$$\|\boldsymbol{k}_n\| = \|\boldsymbol{k}_{d_n}^\top(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})\mathrm{diag}(\tilde{\dot{\boldsymbol{q}}})\| \le \|\boldsymbol{k}_{d_n}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{q}})\|\|\mathrm{diag}(\tilde{\dot{\boldsymbol{q}}})\| \le \sqrt{\mathcal{D}}|\sigma_{f_n}|\|\tilde{\dot{\boldsymbol{q}}}\|_\infty \tag{24}$$

$$\|\Delta\boldsymbol{x}\| = \|\boldsymbol{K_y}^{-1}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})\Delta\boldsymbol{y}\| \leq \|\boldsymbol{K_y}^{-1}(\dot{\boldsymbol{Q}}, \dot{\boldsymbol{Q}})\|\|\Delta\boldsymbol{y}\| \leq \frac{\|\Delta\boldsymbol{y}\|}{\sigma_{\varepsilon}^2} \tag{25}$$

then

$$\boldsymbol{k}_n^\top \Delta\boldsymbol{x} \geq -|\boldsymbol{k}_n^\top\Delta\boldsymbol{x}| \geq -\|\boldsymbol{k}_n\|\|\Delta\boldsymbol{x}\| \geq -\sqrt{D}\|\tilde{\dot{\boldsymbol{q}}}\|_\infty\|\Delta\boldsymbol{y}\|\frac{|\sigma_{f_n}|}{\sigma_{\varepsilon}^2} \tag{26}$$

With the condition of (22), $\boldsymbol{k}_n^\top\Delta\boldsymbol{x} \geq -m_{d_n}(\dot{\boldsymbol{q}})$ and the dissipated power

$$\dot{q}_n\hat{\tau}_n^d(\dot{\boldsymbol{q}}) = \dot{q}_n^2\left(m_{d_n}(\dot{\boldsymbol{q}}) + \boldsymbol{k}_n^\top\Delta\boldsymbol{x}\right) \geq 0. \tag{27}$$

Because of the diagonal structure of $\hat{\boldsymbol{D}}(\dot{\boldsymbol{q}})$, its positive semidefiniteness can be directly derived. ∎

## 5. Simulation and Results

We evaluate our methods by estimating the unknown aerodynamics based on an aircraft benchmark by ONERA & AIRBUS, which simulates the aircraft landing period in full configuration from 1000ft above the runway until touch-down (Biannic and Boada-Bauxell, 2016). We determine the domain of the air velocities by setting up the three-dimensional wind inputs that change periodically within $[-25, 25]m/s$ with different frequencies $[0, 1, 0.2, 0.3]$ and phases $[0, 2, 3]$, and obtain the air velocities in an angle-based coordinate which consists of the angle of attack $\alpha$, the side-slip angle $\beta$ and the norm of the airspeed $v_a$. The generalized velocity is set to $\dot{\boldsymbol{q}} = [\alpha, \beta, v_a]^\top$. We also use the benchmark model to collect the corresponding aerodynamic forces $\boldsymbol{\tau}^d$, namely drag, lateral and lift. For notation simplicity, we use ARD-GP to denote the standard GP methods with SE kernel based on automatic relevance determination (ARD) (Rasmussen and Williams, 2006), Diag-D-GP to denote the GP methods with the by our proposed diagonal form damping matrix, and Full-D-GP to denote the methods with the full damping matrix. To make comparison more in consistent and the optimization process simpler, we choose a fixed noise variance $\sigma_\varepsilon = 100$ for all unconstrained method and choose $\sigma_\varepsilon = 10^8$ to fulfill the constraint. Note that the aerodynamic forces are in the order of $10^6$, so this large noise variance is still reasonable in practice. We first optimize all the other hyperparameters of ARD-GP by minimizing the mean squared error on a validation set. Then we choose the same length-scale $\boldsymbol{\ell} = [18, 18, 0.2]^\top$ and optimize the hypervariances for all the other methods. We first compare the estimation performance of each method and then show the guarantee of passivity numerically with our constrained GP.

### 5.1. Tracking Performance and Data Efficiency

The training results are evaluated in two types. One is the aerodynamics tracking error based on a trajectory along the landing period, as shown in Fig. 1(a). We sample 50 data points from the trajectory with equivalent interval as the training set and 100 points as validation set. The relative error is the difference between the estimation and ground truth normalized with the mass of the aircraft. Our proposed Full-D-GP performs the best compared with the others, which obtain a relative error with almost 0 for all three dimensions. The statistical values of the lateral estimation error as an example for different methods are listed in Table 1. Note that with the assumption of a diagonal damping matrix, we lose the dependencies between each dimension. The relative error of drag in Diag-D-GP is larger than the others because the drag should also contain components based on the

last velocity dimension, norm of the airspeed. The other evaluation type is in a volume space in the domain of the three-dimensional inputs. In this case, the inputs can be chosen as any combination of the values for each dimension in the limited domain. We evaluate the normalized mean squared error (NMSE) in the volume space with increased training data by $\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 / \sum_{i=1}^{N}(y_i - \bar{y})^2$, where $\hat{y}_i$ and $y_i$ are the estimation and ground truth of the i-th data, respectively, and $\bar{y}$ is the mean of the data. The results are shown in Fig 1(b). Full-D-GP achieves the highest training efficiency for all dimensions and converges to a lower NMSE, which follows the same phenomena as training for a trajectory. The NMSE estimated by the constrained Full-D-GP remains at a certain level with the increase of the data set because of the large noise variance. The simulation results show a significant improvement in data efficiency of training when we consider more physical structure as a prior. By involving our proposed constraint, the Full-D-GP maintains efficiency and reasonable performance.
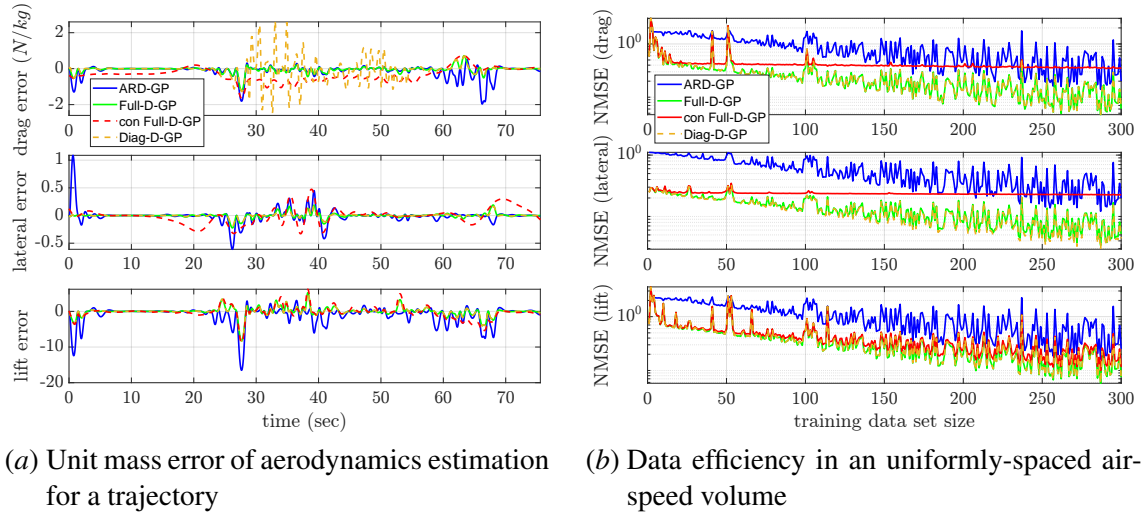


(a) Unit mass error of aerodynamics estimation for a trajectory

(b) Data efficiency in an uniformly-spaced airspeed volume

Figure 1: Training results based on trajectory and volume space

|  | Full-D-GP | Diag-D-GP | ARD-GP |
|---|---|---|---|
| mean | $-7.83 \times 10^{-4}$ | $-5.73 \times 10^{-4}$ | -0.0052 |
| variance | 0.0019 | 0.0012 | 0.175 |

Table 1: Statistical values of the relative lateral force estimation error along a landing trajectory

## 5.2. Passivity Characteristic

The passivity can be preserved if the hyperparameters are chosen with the conditions we derive in Section 4. By involving these conditions as constraints during optimization, we can show the passivity of the estimation numerically by presenting the dissipated power. Consider the sufficient condition (10) for full structured kernel, we determine the mean $m_{d_n}(\dot{q})$ by solving a least squared problem to minimize $\|\tau_n^d - m_{d_n}(\dot{q})\dot{q}_n\|$ where $\tau_n^d$ and $\dot{q}_n$ are the collected forces and velocities of n-th dimension, respectively. $\mathcal{D}, \|\ddot{q}\|_\infty$ and $\|\Delta y\|$ can be directly obtained when the data is given. The dissipated power $P = \dot{q}^\top \tau^d$ should always be non-negative for a passive system, whose distribution based on our test data collected from the aircraft benchmark and the corresponding

estimation based on the hyperparameters with and without constraints are shown in Fig. 2. The collected data is noisy but most of the points are positive, so the estimation based on non-constrained GP shifts the distribution to a positive direction but still cannot preserve passivity. The passivity is deterministically guaranteed with constrained hyperparameters.
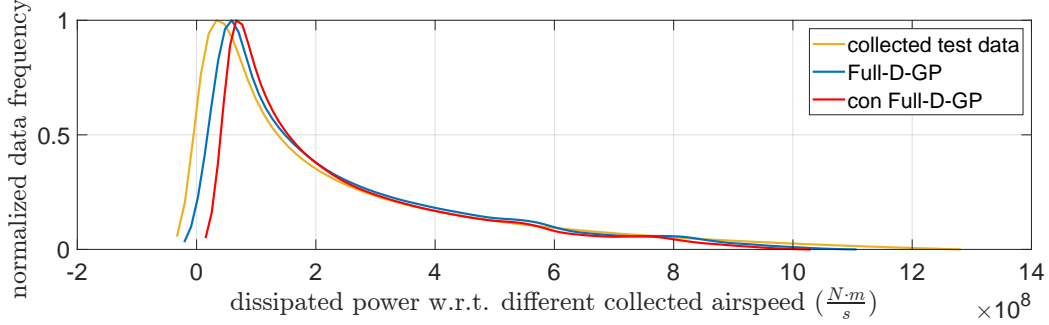


Figure 2: Dissipated power distribution from the collected test data and the corresponding estimation based on constrained and unconstrained methods

## 6. Conclusion

Our approach towards structural learning methods for friction modeling and identification integrates physical structure as prior knowledge with Gaussian Processes. Considering the mechanical system with the velocity-dependent dissipative term, we introduce a matrix-vector structure for prior mean and kernels, with a full and diagonal damping matrix, respectively. We investigate their passivity guarantee of the damping estimates by showing the positive semidefiniteness of the damping matrix. Constrained Gaussian Process algorithms are proposed to guarantee passivity during training by optimizing hyperparameters with constraints. An aircraft benchmark simulation is employed to demonstrate the efficacy of our methodology. By considering more information in the structure, higher data efficiency and estimate accuracy can be obtained. In addition, the passivity characteristic of the estimated aerodynamic forces w.r.t. the generalized velocities is validated numerically.

## Acknowledgments

## References

J. Auriol, R. Shor, S. Niculescu, and N. Kazemi. Estimating drill string friction with model-based and data-driven methods. In *2022 American Control Conference (ACC)*, pages 3464–3469, 2022.

J-M. Biannic and J. Boada-Bauxell. A Civilian Aircraft Landing Challenge. On-line available from the aerospace benchmark section of the SMAC Toolbox, http://w3.onera.fr/smac/, 2016.

P. Bliman and M. Sorine. Easy-to-use realistic dry friction models for automatic control. In *Proceedings of 3rd European control conference*, volume 113, pages 267–272, 1995.

C. C. De Wit, H. Olsson, K. J. Åström, and P. Lischinsky. A new model for control of systems with friction. *IEEE Transactions on automatic control*, 40(3):419–425, 1995.

A. Dong, Z. Du, and Z. Yan. Friction modeling and compensation for haptic master manipulator based on deep gaussian process. *Mechanism and Machine Theory*, 166:104480, 2021.

G. Evangelisti and S. Hirche. Physically consistent learning of conservative lagrangian systems with gaussian processes. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 4078–4085, 2022.

G. Evangelisti and S. Hirche. Data-driven momentum observers with physically consistent gaussian processes. *IEEE Transactions on Robotics*, 40:1938–1951, 2024. doi: 10.1109/TRO.2024. 3366818.

H. Feng, W. Qiao, C. Yin, H. Yu, and D. Cao. Identification and compensation of non-linear friction for a electro-hydraulic system. *Mechanism and Machine Theory*, 141:1–13, 2019.

C. Fiedler, C. W. Scherer, and S. Trimpe. Learning-enhanced robust controller synthesis with rigorous statistical and control-theoretic guarantees. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 5122–5129, 2021.

J. N. Franklin. *Matrix Theory*. Dover Books on Mathematics. Dover Publications, 2012.

A. R. Geist and S. Trimpe. Structured learning of rigid-body dynamics: A survey and unified view from a robotics perspective. *GAMM-Mitteilungen*, 44(2):e202100009, 2021.

S. Huang, W. Liang, and K. K. Tan. Intelligent friction compensation: A review. *IEEE/ASME Transactions on Mechatronics*, 24(4):1763–1774, 2019.

J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter. Learning agile and dynamic motor skills for legged robots. *Science Robotics*, 4(26), 2019.

H.K. Khalil. *Nonlinear Systems*. Pearson Education. Prentice Hall, 2002.

A. Koch, J. Berberich, and F. Allgöwer. Verifying dissipativity properties from noise-corrupted input-state data. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 616–621, 2020.

V. Lampaert, J. Swevers, and F. Al-Bender. Modification of the leuven integrated friction model structure. *IEEE Transactions on Automatic Control*, 47(4):683–687, 2002.

M. Malek-Shahmirzadi. A characterization of certain classes of matrix norms. *Linear and Multilinear Algebra*, 13(2):97–99, 1983.

Oliver Nelles. *Nonlinear System Identification*. Springer International Publishing, 2020.

D. Nguyen-Tuong and J. Peters. Using model knowledge for learning inverse dynamics. In *2010 IEEE International Conference on Robotics and Automation*, pages 2677–2682, 2010.

H. Olsson, K.J. Åström, C. Canudas de Wit, M. Gäfvert, and P. Lischinsky. Friction models and friction compensation. *European Journal of Control*, 4(3):176–195, 1998.

R. Ortega, A. Loría, P. J. Nicklasson, and H. Sira-Ramírez. *Passivity-based Control of Euler-Lagrange Systems*. Springer London, 1998.

H. Peng, N. Song, F. Li, and S. Tang. A Mechanistic-Based Data-Driven Approach for General Friction Modeling in Complex Mechanical System. *Journal of Applied Mechanics*, 89(7), 05 2022.

C.E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006.

G. Sutanto, A. S. Wang, Y. Lin, M. Mukadam, G. S. Sukhatme, A. Rai, and F. Meier. Encoding physical constraints in differentiable newton-euler algorithm. In *2nd Conference on Learning for Dynamics and Control*, pages 804–813. PMLR, 2020.

J. C. Willems. Dissipative dynamical systems part I: General theory. *Archive for Rational Mechanics and Analysis*, 45(5):321–351, 1972.

J. Wojewoda, A. Stefański, M. Wiercigroch, and T. Kapitaniak. Hysteretic effects of dry friction: modelling and experimental studies. *Phil. Trans. R. Soc. A.*, 366:747–765, 2008.