Physically Consistent Learning of Conservative Lagrangian Systems with Gaussian Processes

Giulio Evangelisti and Sandra Hirche

Abstract— This paper proposes a physically consistent Gaussian Process (GP) enabling the data-driven modelling of uncertain Lagrangian systems. The function space is tailored according to the energy components of the Lagrangian and the differential equation structure, analytically guaranteeing properties such as energy conservation and quadratic form. The novel formulation of Cholesky decomposed matrix kernels allow the probabilistic preservation of positive definiteness. Only differential input-to-output measurements of the function map are required while Gaussian noise is permitted in torques, velocities, and accelerations. We demonstrate the effectiveness of the approach in numerical simulation.

I. INTRODUCTION

Euler-Lagrange (EL) systems represent a wide variety of physical systems and are the foundation of a large class of theoretical concepts in control, such as stability, passivity, and in feedback law design. The application of Gaussian Processes (GPs) in learning-based control holds great promise for the control of uncertain systems, improving important aspects like performance [1], safety [2], and data-efficiency [3]. Yet, in general GP regression does not account for physically consistency, hampering the provision of safety guarantees when used in model-based control approaches.

We address this issue by formulating GPs consistent with the relevant physical laws and mathematical models, thus implicitly performing a trade-off between the flexibility of learning and physics-imposed symmetries. Traditionally, parametric techniques [4] have been applied to the identification of EL systems, as reviewed in [5]. However, mostly assuming linearity in the parameters, these methods eventually become infeasible with increasingly nonlinear, and coupled, dynamics, as for example in soft robots [6], aircrafts [7], and marine vehicles [8].

Based on Bayesian principles, GPs allow the inclusion of analytical prior knowledge, thus forming a bridge to parametric model-based approaches. The authors of [9] propose the regression of Hamiltonian systems with GPs, preserving the symplectic structure, but without any guarantees and requiring the availability of noiseless impulse measurements. A Lagrangian kernel is derived in [10] within the framework of Reproducing Kernel Hilbert Spaces (RKHS). While analogies to our stochastic GP-based approach naturally exist, the work lacks a further specification of the associated function

Both authors are members of the Chair of Information-oriented Control, Department of Electrical and Computer Engineering, Technical University of Munich, D-80333 Munich, Germany [giulio.evangelisti, hirche]@tum.de

space, prohibiting a comprehensive analysis of control relevant properties closely intertwined with physical consistency. In [11], the EL residual dynamics are compensated by employing independent GPs per degree of freedom, neglecting correlations and structure. This leads to an implicit differential equation system, the necessity of noiseless acceleration measurements and the curse of dimensionality. Guaranteeing physical consistency of data-driven methods still represents a major open problem.

The main contribution of this work is the derivation of a physically consistent Lagrangian-Gaussian Process (L-GP) constructed by linear operators, embedding the differential equation structure into the model. Combined with quadratic kernels, we extend the GP concept from scalar function to symmetric matrix spaces. Based on a Cholesky decomposition of the underlying matrix kernel, we are able to provide a probabilistic guarantee for the positive definiteness of the estimates. Further physical consistencies of the L-GP analytically preserving equilibria, passivity, and conservatism are formally shown.

After defining the problem setting in Section II, we review the framework of GPs in Section III and describe our proposed unification with EL systems in Section IV. Section V presents our main results. Numerical illustrations validate theory and effectiveness of the approach in Section VI.

II. PROBLEM FORMULATION

In this paper, we consider the class of conservative EL systems whose equations of motion are given by[1]

$$\frac{d}{dt}\left(\nabla_{\dot{q}}L\right) - \nabla_{q}L = \boldsymbol{\tau} \tag{1}$$

based on the set of generalized coordinates $\boldsymbol{q} \in \mathbb{R}^N$ and the unknown Lagrangian L. The generalized forces $\boldsymbol{\tau} \in \mathbb{R}^N$ are the measurable, possibly underactuated, inputs to the system.

Assumption 1: The unknown Lagrangian is autonomous with $L \equiv L(\boldsymbol{q}, \dot{\boldsymbol{q}}) = T(\boldsymbol{q}, \dot{\boldsymbol{q}}) - V(\boldsymbol{q})$ and composed of the

---

[1]**Notation:** Vectors $\boldsymbol{a}$ and matrices $\boldsymbol{A}$ are denoted with bold lower and upper case characters, respectively. The nabla operator ∇ multiplied with a differentiable scalar field $f(\boldsymbol{x})$ gives its gradient w.r.t. $\boldsymbol{x} \in \mathcal{X}$. The partial gradient w.r.t. a subset of variables $\boldsymbol{y} \in \mathcal{Y}$, $\mathcal{Y} \subset \mathcal{X}$, is described by $\nabla_{\boldsymbol{y}} f$. $\boldsymbol{I}$ denotes the identity, $\boldsymbol{0}$ the zero and $\boldsymbol{1}$ the ones matrix. A collection of D indexed vectors $\boldsymbol{x}_i \in \mathbb{R}^N$ is stacked row-wise in the matrix $\boldsymbol{X} = [\boldsymbol{x}_i^\top]_{i=1}^D \in \mathbb{R}^{D \times N}$ such that x_{ij} refers to the j-th component of $\boldsymbol{x}_i$. For notational convenience, we omit the indexing in the matrix expressions writing $[x_{ij}] = \boldsymbol{X} = [\boldsymbol{x}_i^\top] = [\boldsymbol{y}_j] = \boldsymbol{Y}^\top = [y_{ij}]$. $\mathrm{E}[\cdot]$, $\mathrm{Var}[\cdot]$ and $\mathrm{Cov}[\cdot]$ denote the expectation, variance and covariance operators w.r.t. random variables. The Hadamard product and square are denoted by $\boldsymbol{x} \circ \boldsymbol{y}$ and $\boldsymbol{x}^{\circ 2} = \boldsymbol{x} \circ \boldsymbol{x}$, respectively, the Kronecker product with $\otimes$, and the direct sum with $\oplus$. With $\lambda_k(\cdot)$ we denote the k-th largest eigenvalue corresponding to the components of the decreasingly ordered eigenvalue vector $\boldsymbol{\lambda}^{\downarrow}(\cdot)$.

difference between unknown kinetic energy $T \colon \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ and unknown potential energy $V \colon \mathbb{R}^N \to \mathbb{R}$.

Assumption 2: The kinetic energy is of the form $T(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \frac{1}{2}\dot{\boldsymbol{q}}^T \boldsymbol{M}(\boldsymbol{q})\dot{\boldsymbol{q}}$ where $\boldsymbol{M} \colon \mathbb{R}^N \to \mathbb{R}^{N \times N}$ is the unknown, (symmetric) positive definite, inertia matrix.

Assumption 3: The potential energy is decomposable into the sum $V(\boldsymbol{q}) = G(\boldsymbol{q}) + U(\boldsymbol{q})$ of gravitational and elastic potential energies G and U, respectively, where the former has an equilibrium, $G(\boldsymbol{0}) = 0$, $\nabla_{\boldsymbol{q}} G(\boldsymbol{0}) = \boldsymbol{0}$, and the latter is given by $U(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T \boldsymbol{S}(\boldsymbol{q})\boldsymbol{q}$ with unknown, (symmetric) positive definite, stiffness matrix $\boldsymbol{S} \colon \mathbb{R}^N \to \mathbb{R}^{N \times N}$.

Assumptions 1–3 essentially describe the considered system class and represent its physical properties. Assumption 1 restricts the systems described by (1) to be only implicitly dependent on time and free from disturbances caused by other external forces such as dissipation. Extending to handle dissipative and time-variant systems is possible and topic of current research. Assumption 2 is a valid approach when dealing, e.g., with classical mechanical rigid body systems, while Assumption 3 constrains chosen coordinates to have an equilibrium at the origin.

Exploiting Assumptions 1–3 and applying the chain rule to (1), we obtain the well-known matrix-vector expression

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) = \boldsymbol{\tau}, \qquad (2)$$

introducing the (non-unique) generalized Coriolis matrix $\boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}}) \coloneqq \frac{1}{2}[\nabla_{\dot{\boldsymbol{q}}}\nabla_{\boldsymbol{q}}^\top T + \dot{\boldsymbol{M}}(\boldsymbol{q}) - \nabla_{\boldsymbol{q}}\nabla_{\dot{\boldsymbol{q}}}^\top T]$ as in [12] and the vector of generalized potential forces $\boldsymbol{g}(\boldsymbol{q}) \coloneqq \nabla_{\boldsymbol{q}} V$, having exploited $\nabla_{\dot{\boldsymbol{q}}}\nabla_{\dot{\boldsymbol{q}}}^\top T = \boldsymbol{M}(\boldsymbol{q})$, $\nabla_{\boldsymbol{q}} T = \frac{1}{2}(\nabla_{\boldsymbol{q}}\nabla_{\dot{\boldsymbol{q}}}^\top T)\dot{\boldsymbol{q}}$ as well as $\left(\nabla_{\dot{\boldsymbol{q}}}\nabla_{\boldsymbol{q}}^\top T\right)\dot{\boldsymbol{q}} = \dot{\boldsymbol{M}}(\boldsymbol{q})\dot{\boldsymbol{q}}$. Note that there are other ways to define the matrix $\boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})$. This particular choice, however, has the unique property that $\dot{\boldsymbol{M}} - 2\boldsymbol{C}$ is skew-symmetric.

Our objective is to approximate the unknown Lagrangian function $L(\boldsymbol{q}, \dot{\boldsymbol{q}})$ by a data-based estimate $\hat{L}(\boldsymbol{q}, \dot{\boldsymbol{q}})$ which is physically consistent, i.e., maximally aligned with the existing physical knowledge, in our case Assumptions 1–3. We assume access to noisefree positional observations $\boldsymbol{q}_i$, $i = 1, \ldots, D$ with $D \in \mathbb{N}$, and to potentially noisy velocity, acceleration and torque measurements

$$\dot{\boldsymbol{x}}_i = \begin{bmatrix} \dot{\boldsymbol{q}}_i \\ \ddot{\boldsymbol{q}}_i \end{bmatrix} + \begin{bmatrix} \boldsymbol{\omega}_i \\ \boldsymbol{\alpha}_i \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{\omega}_i \\ \boldsymbol{\alpha}_i \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\omega}_i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i} \end{bmatrix}\right), \quad \text{(3a)}$$

$$\boldsymbol{y}_i = \boldsymbol{\tau}_i(\boldsymbol{q}_i, \dot{\boldsymbol{q}}_i, \ddot{\boldsymbol{q}}_i) + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_i}). \quad \text{(3b)}$$

The noise processes $\{\boldsymbol{\omega}_i\}$, $\{\boldsymbol{\alpha}_i\}$ and $\{\boldsymbol{\varepsilon}_i\}$ are white, zero-mean, uncorrelated, and have known covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\omega}_i}$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_i}$, respectively. After collecting all D observations at positions $\boldsymbol{Q} = [\boldsymbol{q}_i^\top]$ with analogous matrices $\dot{\boldsymbol{X}} = [\dot{\boldsymbol{x}}_i^\top]$ and $\boldsymbol{Y} = [\boldsymbol{y}_i^\top]$, we obtain the training data set

$$\mathcal{D} = \{\boldsymbol{Q}, \dot{\boldsymbol{X}}, \boldsymbol{Y}\}.$$

Remark 1: Extending to noisy positional inputs is possible via certain GP-based methods as in, e.g., [13], [14], yet they are mostly based on first-order Taylor approximations.

III. GAUSSIAN PROCESS FRAMEWORK

This section gives a brief overview of the mathematical framework of GPs based on [15]. For a more complete introduction, the reader is referred to the literature [16]–[18].

A. Inference with GPs

For $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ in a continuous domain $\mathcal{X} \subseteq \mathbb{R}^{2N}$, a GP with mean $m(\boldsymbol{x})$ and covariance $k(\boldsymbol{x}, \boldsymbol{x}')$, denoted by

$$f(\boldsymbol{x}) \sim \mathcal{GP}\left(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right) \qquad (4)$$

and with scalar $f(\boldsymbol{x}) \in \mathbb{R}$, is a stochastic process extending the Gaussian probability distribution from random variables to functions. As such, it inherits the convenient mathematical properties of the Normal distribution. In particular, any finite marginalization $\boldsymbol{z} = [z_i]$ of function observations $z_i = f(\boldsymbol{x_i}) + \zeta_i$, corrupted by the zero-mean white-noise process $\zeta_i \in \mathbb{R}$ and corresponding to the finite subset of function evaluations at $\{\boldsymbol{x}_i\} \subset \mathcal{X}$, is jointly Gaussian distributed.

A GP is fully characterized by its mean and kernel function $m(\boldsymbol{x}) = \mathrm{E}[f(\boldsymbol{x})]$ and $k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}[f(\boldsymbol{x}), f(\boldsymbol{x}')]$, respectively. The former allows the analytical inclusion of prior knowledge about the unknown function, whereas the latter determines higher level functional properties such as smoothness. Estimation of the covariance describing hyperparameters is mostly done via optimization of the marginal likelihood to maximize the probability of observing the measured outputs.

For regression, GPs exploit the joint Gaussian distribution of observations $\boldsymbol{z}$ at measurement locations $\boldsymbol{X} = [\boldsymbol{x}_i^\top]$ and prediction $f(\boldsymbol{x})$ w.r.t. the test input $\boldsymbol{x} \in \mathcal{X}$ given by

$$\begin{bmatrix} f(\boldsymbol{x}) \\ \boldsymbol{z} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\boldsymbol{x}) \\ \boldsymbol{m} \end{bmatrix}, \begin{bmatrix} k(\boldsymbol{x}, \boldsymbol{x}) & \boldsymbol{k}^\top(\boldsymbol{x}) \\ \boldsymbol{k}(\boldsymbol{x}) & \boldsymbol{K} + \boldsymbol{\Sigma} \end{bmatrix}\right), \quad (5)$$

where we have introduced the mean $\boldsymbol{m} = [m(\boldsymbol{x}_i)]$ and covariance $\boldsymbol{k}(\boldsymbol{x}) = [k(\boldsymbol{x}_i, \boldsymbol{x})]$ vectors, the Gram matrix $\boldsymbol{K} = [k(\boldsymbol{x}_i, \boldsymbol{x}_j)]$ as well as the noise covariance matrix $\boldsymbol{\Sigma} = \mathrm{Var}[\boldsymbol{\zeta}]$ with $\boldsymbol{\zeta} = [\zeta_i]$. Here, we have dropped the dependencies of $\boldsymbol{m}(\boldsymbol{X})$, $\boldsymbol{k}(\boldsymbol{X}, \boldsymbol{x})$ and $\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X})$ on $\boldsymbol{X}$ for notational simplicity. Conditioning on the observations yields a conditional Gaussian distribution with posterior mean $\hat{f}(\boldsymbol{x}) \equiv \mathrm{E}[f(\boldsymbol{x})|\boldsymbol{z}]$ given analytically by

$$\hat{f}(\boldsymbol{x}) = m(\boldsymbol{x}) + \boldsymbol{k}^\top(\boldsymbol{x})\left(\boldsymbol{K} + \boldsymbol{\Sigma}\right)^{-1}\left(\boldsymbol{z} - \boldsymbol{m}\right). \quad (6)$$

B. Linear Operators and GPs

Derivatives and integrals of GPs remain GPs, since these operations are linear. Applying a linear transformation operator $\boldsymbol{\mathcal{T}}_{\boldsymbol{x}}$ to the GP of (4) leads to a new, possibly multidimensional, GP in the form of

$$\boldsymbol{\mathcal{T}}_{\boldsymbol{x}} f(\boldsymbol{x}) \sim \mathcal{GP}\left(\boldsymbol{m}_{\boldsymbol{\mathcal{T}}}(\boldsymbol{x}), \boldsymbol{K}_{\boldsymbol{\mathcal{T}}}(\boldsymbol{x}, \boldsymbol{x}')\right) \qquad (7)$$

with transformed mean $\boldsymbol{m}_{\boldsymbol{\mathcal{T}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{T}}_{\boldsymbol{x}} m(\boldsymbol{x})$ and kernel

$$\boldsymbol{K}_{\boldsymbol{\mathcal{T}}}(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{\mathcal{T}}_{\boldsymbol{x}} k(\boldsymbol{x}, \boldsymbol{x}') \boldsymbol{\mathcal{T}}_{\boldsymbol{x}'}^\top. \qquad (8)$$

Here, $\boldsymbol{\mathcal{T}}_{\boldsymbol{x}'}^\top$ is applied from the right [9] as is necessary when considering matrix operators.

IV. REGRESSION OF LAGRANGIAN SYSTEMS

Our key idea is to consistently model the unknown Lagrangian using a GP. We propose a structural approach for the application of GP regression (4)–(6) to EL systems (1)–(2): deriving specific GPs and kernels, we ensure consistency

with physical requirements, as in Assumptions 1–3, and embed the structure of the differential equations using Nabla operators ∇, thereby inducing multidimensional L-GPs according to (7)–(8) based on the linearity of differentiation.

A. Energy Structuring

Firstly, we model the Lagrangian as a composite GP $L(\boldsymbol{q}, \dot{\boldsymbol{q}}) = T(\boldsymbol{q}, \dot{\boldsymbol{q}}) - V(\boldsymbol{q})$ writing

$$L(\boldsymbol{q}, \dot{\boldsymbol{q}}) \sim \mathcal{GP}\left(m_L(\boldsymbol{q}, \dot{\boldsymbol{q}}), k_L(\boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{q}', \dot{\boldsymbol{q}}')\right) \tag{9}$$

with the subjacent kinetic and potential energy GPs

$$T(\boldsymbol{q}, \dot{\boldsymbol{q}}) \sim \mathcal{GP}\left(m_T(\boldsymbol{q}, \dot{\boldsymbol{q}}), k_T(\boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{q}', \dot{\boldsymbol{q}}')\right), \tag{10a}$$
$$V(\boldsymbol{q}) \sim \mathcal{GP}\left(m_V(\boldsymbol{q}), k_V(\boldsymbol{q}, \boldsymbol{q}')\right). \tag{10b}$$

Here, the means $m_L = m_T - m_V$ and covariances $k_L = k_T + k_V$ specify the L-GP (9) in a structurally consistent manner. In particular, the potential energy GP (10b) depends only on the position $\boldsymbol{q}$ by construction. It is split up further into the sum $V(\boldsymbol{q}) = G(\boldsymbol{q}) + U(\boldsymbol{q})$ of independent gravitational and elastic potential GPs

$$G(\boldsymbol{q}) \sim \mathcal{GP}\left(m_G(\boldsymbol{q}), k_G(\boldsymbol{q}, \boldsymbol{q}')\right), \tag{11a}$$
$$U(\boldsymbol{q}) \sim \mathcal{GP}\left(m_U(\boldsymbol{q}), k_U(\boldsymbol{q}, \boldsymbol{q}')\right). \tag{11b}$$

B. Quadratic and Positive Definite Structures

Consistency of kinetic and elastic GPs is ensured by according design of their mean and covariance functions. If available, we impose on the according priors the same quadratic structure based on, e.g., for the kinetic energy, the (symmetric) positive definite a-priori matrix $\boldsymbol{M}_0(\boldsymbol{q}) \succ \boldsymbol{0}$, $\forall \boldsymbol{q} \in \mathbb{R}^N$, as described in the following.

Assumption 4: The priors $m_{f_k} \in \{m_U, m_T\}$ are in the same form as the unknown energy functions

$$f_k(\boldsymbol{q}) \coloneqq \frac{1}{2} \overset{(k)\top}{\boldsymbol{q}} \boldsymbol{H}_k(\boldsymbol{q}) \overset{(k)}{\boldsymbol{q}} \tag{12}$$

for $k \in \{0, 1\}$, where $m_{f_0} \equiv m_U$ with $\boldsymbol{H}_{00} \equiv \boldsymbol{S}_0$ and $m_{f_1} \equiv m_T$ with $\boldsymbol{H}_{10} \equiv \boldsymbol{M}_0$. In addition, the gravitational prior fulfills the equilibrium condition such that $m_G(\boldsymbol{0}) = 0$ and $\nabla_{\boldsymbol{q}} m_G(\boldsymbol{0}) = \boldsymbol{0}$ hold.

Having specified the means in (10a) and (11b), we now move on to the kernels. Consider the quadratic functional

$$\kappa_k(\boldsymbol{q}, \boldsymbol{q}') = \frac{1}{4}\left(\overset{(k)}{\boldsymbol{q}} \circ \overset{(k)'}{\boldsymbol{q}}\right)^\top \boldsymbol{\Theta}_k(\boldsymbol{q}, \boldsymbol{q}')\left(\overset{(k)}{\boldsymbol{q}} \circ \overset{(k)'}{\boldsymbol{q}}\right) \tag{13}$$

with differential-operational index $k \in \{0, 1\}$ and the (symmetric) positive definite Cholesky decomposed matrix kernel

$$\boldsymbol{\Theta}_k(\boldsymbol{q}, \boldsymbol{q}') = \boldsymbol{R}_k^\top(\boldsymbol{q}, \boldsymbol{q}') \boldsymbol{R}_k(\boldsymbol{q}, \boldsymbol{q}'). \tag{14}$$

The Cholesky factor $\boldsymbol{R}_k$ is an upper-right triangular matrix

$$\boldsymbol{R}_k(\boldsymbol{q}, \boldsymbol{q}') = \begin{bmatrix} r_{k11}(\boldsymbol{q}, \boldsymbol{q}') & \cdots & r_{k1N}(\boldsymbol{q}, \boldsymbol{q}') \\ & \ddots & \vdots \\ \boldsymbol{0} & & r_{kNN}(\boldsymbol{q}, \boldsymbol{q}') \end{bmatrix} \tag{15}$$

consisting for $0 \le n \le m \le N$ of kernels $r_{knm}(\boldsymbol{q}, \boldsymbol{q}')$. We now use (13) to set the covariances in (10a) and (11b) to:

$$k_U \equiv k_{f_0} \coloneqq \kappa_0, \tag{16a}$$
$$k_T \equiv k_{f_1} \coloneqq \kappa_1. \tag{16b}$$

Thus, we implicitly model the stiffness and inertia matrices denoted in accordance with (12) by $\boldsymbol{H}_k \coloneqq [h_{knm}]$ as symmetric, matrix-valued, GPs with entries

$$h_{knm}(\boldsymbol{q}) \sim \mathcal{GP}\left(\eta_{knm}(\boldsymbol{q}), \theta_{knm}(\boldsymbol{q}, \boldsymbol{q}')\right),$$

where $h_{knm} = h_{kmn}$, and analogously constructing $\boldsymbol{H}_{k0} \coloneqq [\eta_{knm}]$. The kernels $\theta_{knm} = \boldsymbol{\rho}_{kn}^\top \boldsymbol{\rho}_{km}$ stem from the inner product of the non-zero columns from (15), i.e, $\boldsymbol{\rho}_{km} \coloneqq [r_{klm}] \in \mathbb{R}^P$ for $P = \min\{n, m\}$. Note that, due to the nonlinearity of multiplication, the r_{klm} may not be associated any further with underlying GPs, or more specifically, Gaussian Random Variables (GRVs). However, the covariances θ_{knm} remain valid kernel functions according to [15].

C. Differential Equation Embedding

Let us now use the chain rule to expand $\boldsymbol{\mathcal{L}_q} \equiv \frac{d}{dt}\nabla_{\dot{\boldsymbol{q}}} - \nabla_{\boldsymbol{q}}$ from (1) by writing

$$\boldsymbol{\mathcal{L}_q} = \left(\nabla_{\dot{\boldsymbol{q}}}^\top \ddot{\boldsymbol{q}} + \nabla_{\boldsymbol{q}}^\top \dot{\boldsymbol{q}}\right)\nabla_{\dot{\boldsymbol{q}}} - \nabla_{\boldsymbol{q}} \coloneqq \boldsymbol{\mathcal{L}}_{\boldsymbol{q}, \dot{\boldsymbol{q}}}(\dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}}). \tag{17}$$

Applying the Lagrangian-differential vector operator (17) to the GP from (9), we obtain for the torques (1)–(2) a vector-valued GP $\boldsymbol{\tau} \equiv \boldsymbol{\tau}(\boldsymbol{q}, \boldsymbol{q}'; \dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}}, \dot{\boldsymbol{q}}', \ddot{\boldsymbol{q}}') \in \mathbb{R}^N$ denoted by

$$\boldsymbol{\tau} \sim \mathcal{GP}\left(\boldsymbol{m}_{\boldsymbol{\tau}}(\boldsymbol{q}; \dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}}), \boldsymbol{K}_{\boldsymbol{\tau}}(\boldsymbol{q}, \boldsymbol{q}'; \dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}}, \dot{\boldsymbol{q}}', \ddot{\boldsymbol{q}}')\right) \tag{18}$$

with mean vector $\boldsymbol{m}_{\boldsymbol{\tau}} = \boldsymbol{\mathcal{L}_q} m_L$ and kernel matrix $\boldsymbol{K}_{\boldsymbol{\tau}} = \boldsymbol{\mathcal{L}_q}\boldsymbol{\mathcal{L}}_{\boldsymbol{q}'}^\top k_L$. As indicated in (18) via the semicolon, $\dot{\boldsymbol{q}}$ and $\ddot{\boldsymbol{q}}$ have the role of regressors, whereas $\boldsymbol{q}$ remains a conventional input.

D. Joint GP Distribution

Finally, we assert a joint GP for the energies (10a)–(10b) and torques (18). Including for $\boldsymbol{\gamma}^\top(\boldsymbol{q}) \coloneqq [V(\boldsymbol{q}), \nabla_{\boldsymbol{q}}^\top V(\boldsymbol{q})]$ the equilibrium condition

$$\boldsymbol{\gamma}_0 \coloneqq \boldsymbol{\gamma}(\boldsymbol{0}) = \boldsymbol{0}$$

in alignment with Assumption 3, we write

$$\begin{bmatrix} T \\ V \\ \boldsymbol{\gamma}_0 \\ \boldsymbol{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_T \\ m_V \\ \boldsymbol{0} \\ \boldsymbol{m}_{\boldsymbol{y}} \end{bmatrix}, \begin{bmatrix} \sigma_T^2 & 0 & \boldsymbol{0}^\top & \boldsymbol{k}_{\boldsymbol{y}T}^\top \\ 0 & \sigma_V^2 & \boldsymbol{k}_{\boldsymbol{0}V}^\top & \boldsymbol{k}_{\boldsymbol{y}V}^\top \\ \boldsymbol{0} & \boldsymbol{k}_{\boldsymbol{0}V} & \boldsymbol{K}_{\boldsymbol{0}} & \boldsymbol{K}_{\boldsymbol{y}\boldsymbol{0}}^\top \\ \boldsymbol{k}_{\boldsymbol{y}T} & \boldsymbol{k}_{\boldsymbol{y}V} & \boldsymbol{K}_{\boldsymbol{y}\boldsymbol{0}} & \boldsymbol{K}_{\boldsymbol{y}} \end{bmatrix}\right) \tag{19}$$

dropping dependencies on all variables for notational brevity. Also, we have introduced: the stacked vector of outputs $\boldsymbol{y} = \mathrm{vec}(\boldsymbol{Y}^\top)$ with mean $\boldsymbol{m}_{\boldsymbol{y}} = [\boldsymbol{m}_{\boldsymbol{\tau}}(\boldsymbol{q}_i; \dot{\boldsymbol{q}}_i, \ddot{\boldsymbol{q}}_i)]$ and covariance

$$\boldsymbol{K}_{\boldsymbol{y}} = \left[\boldsymbol{K}_{\boldsymbol{\tau}}(\boldsymbol{q}_i, \boldsymbol{q}_j; \dot{\boldsymbol{q}}_i, \ddot{\boldsymbol{q}}_i, \dot{\boldsymbol{q}}_j, \ddot{\boldsymbol{q}}_j)\right] + \oplus_i \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_i}, \tag{20}$$

the energy variances $\sigma_T^2(\boldsymbol{q}, \dot{\boldsymbol{q}}) = k_T(\boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{q}, \dot{\boldsymbol{q}})$ and $\sigma_V^2(\boldsymbol{q}) = k_V(\boldsymbol{q}, \boldsymbol{q})$, the equilibrium variance

$$\boldsymbol{K}_{\boldsymbol{0}} = \begin{bmatrix} k_V(\boldsymbol{0}, \boldsymbol{0}) & \nabla_{\boldsymbol{q}'}^\top k_V(\boldsymbol{0}, \boldsymbol{0}) \\ \nabla_{\boldsymbol{q}} k_V(\boldsymbol{0}, \boldsymbol{0}) & \nabla_{\boldsymbol{q}}\nabla_{\boldsymbol{q}'}^\top k_V(\boldsymbol{0}, \boldsymbol{0}) \end{bmatrix}$$

and covariance $K_{y0} = [k_{yV}(0), \nabla_q^\top k_{yV}(0)]$, as well as the remaining equilibrium-potential $k_{0V}(q)$ and Lagrangian-differential energy $k_{yT}(q, \dot{q})$, $k_{yV}(q)$ covariances

$$k_{0V}(q) = \begin{bmatrix} k_V(0, q) \\ \nabla_{q_i} k_V(0, q) \end{bmatrix}, \tag{21a}$$

$$k_{yT}(q, \dot{q}) = [\mathcal{L}_{q_i, \dot{q}_i}(\dot{q}_i, \ddot{q}_i) k_T(q_i, \dot{q}_i, q, \dot{q})], \tag{21b}$$

$$k_{yV}(q) = [\nabla_{q_i} k_V(q_i, q)]. \tag{21c}$$

The fully formulated L-GP (19) is the first part of our contribution. Before moving on to analyze its physical consistency, we propose a method for partial input noise compensation.

E. Nonlinear Noise Compensation

The regressor structure stemming from the application of (17) to (16b) allows for the compensation of noise in the differential inputs $\dot{q}_i$ and $\ddot{q}_i$. Therefore, as a last step, we combine (2) and (18) in order to transform (3b) into

$$y_i = \tau(q_i; \dot{x}_i) + \tilde{\varepsilon}_i$$

with the transformed composite noise variable

$$\tilde{\varepsilon}_i = -(\nabla_{\dot{q}_i} \nabla_{\dot{q}_i}^\top T) \alpha_i - (\nabla_{\dot{q}_i} \nabla_{q_i}^\top T) \omega_i + \varepsilon_i.$$

Approximating the product of two GRVs to remain Gaussian, we obtain a heteroscedastic (state-dependent) noise model

$$\tilde{\varepsilon}_i \sim \mathcal{N}(0, \Sigma_{\tilde{\varepsilon}_i}(q_i, \dot{q}_i)), \tag{22}$$

where $\Sigma_{\tilde{\varepsilon}_i} = \Sigma_{\tilde{\alpha}_i}(q_i) + \Sigma_{\tilde{\omega}_i}(q_i, \dot{q}_i) + \Sigma_{\varepsilon_i}$. After extensive computations, it can be verified that the transformed $\Sigma_{\tilde{\alpha}_i} := \text{Var}[(\nabla_{\dot{q}_i} \nabla_{\dot{q}_i}^\top T) \alpha_i]$ and $\Sigma_{\tilde{\omega}_i} := \text{Var}[(\nabla_{\dot{q}_i} \nabla_{q_i}^\top T) \omega_i]$, exploiting stochastic independencies, are given by

$$\Sigma_{\tilde{\alpha}_i} = M_0 \Sigma_{\alpha_i} M_0 + (\mathbf{1} - I) \circ \Sigma_{\alpha_i} \circ \Theta_{1ii} + \text{diag}(\Theta_{1ii} \sigma_{\alpha_i})$$

$$\Sigma_{\tilde{\omega}_i} = C_0 \Sigma_{\omega_i} C_0^\top + \text{diag}(\Gamma(q_i, \dot{q}_i) \sigma_{\omega_i}) \tag{23}$$

with the inertial variance $\Theta_{1ii} := \Theta_1(q_i, q_i)$, the Coriolis mean $C_0(q_i, \dot{q}_i) = [M_0(q_i) \dot{q}_i] \nabla_{q_i}^\top$ and variance $\Gamma(q_i, \dot{q}_i) = [\Theta_1(q_i, q_i) \dot{q}_i^{\circ 2}] (\nabla_{q_i}^{\circ 2})^\top$ matrices as well as the main diagonals vectors σ_{α_i} and σ_{ω_i} of Σ_{α_i} and Σ_{ω_i}, respectively.

Proposition 1: The distribution (22) with state-dependent variances (23) is a stochastically consistent approximation which becomes exact for the limit case $\Sigma_{\alpha_i}, \Sigma_{\omega_i} \to 0$, or the case of certainly known $T = m_T$ with $k_T \equiv 0$.

Proof: The limit case $\Sigma_{\alpha_i}, \Sigma_{\omega_i} \to 0$ leads to noisefree inputs with $\tilde{\varepsilon}_i = \varepsilon_i \sim \mathcal{N}(0, \Sigma_{\varepsilon_i})$. The case of known T with certainty $k_T \equiv 0$ leads to $M_0 \equiv M$ and $\Sigma_{\tilde{\alpha}_i} = M_0 \Sigma_{\alpha_i} M_0^\top$, respectively. Since the transformation $\tilde{\alpha}_i = -M\alpha$ is linear, $\tilde{\alpha}_i \sim \mathcal{N}(0, \Sigma_{\tilde{\alpha}_i})$ holds exactly. Proceeding analgously for $\tilde{\omega}_i = -(\nabla_{\dot{q}_i} \nabla_{q_i}^\top T) \omega$, we arrive at (22). ∎

By casting from the temporal into the spatial domain, (22) unifies the L-GP framework with techniques from nonlinear Kalman filtering [19]. For effective noise compensation in the inputs $\dot{q}_i$ and $\ddot{q}_i$, we replace the block matrices Σ_{ε_i} in (20) by $\Sigma_{\tilde{\varepsilon}_i}(q_i, \xi_i)$ from (22), where $\xi_i := [\dot{x}_{in}]$. Also, we compute the torque covariance matrix (20) and the Lagrangian-differential kinetic covariance vector (21b) by setting $\dot{x}_i^\top = [\dot{q}_i^\top, \ddot{q}_i^\top], \forall i, j \in \{1, \dots, D\}$.

V. INVARIANCE PROPERTIES

Having introduced the novel framework of L-GPs, we now investigate their properties and derive certain guarantees.

A. Quadratic Form

Employing the kernel structure (13), quadratic form of the energy GPs (10a) and (11b) can be guaranteed analytically. In order to specify our result in Lemma 1, we require the following definition: for $\delta x_i \in \mathbb{R}^N$ and fixed q, the directional derivative of $\Theta_k(q_i, q)$ along δx_i at q_i gives

$$\Phi_k(q_i, \delta x_i, q) := [\delta x_i^\top \nabla_{q_i} \theta_{knm}(q_i, q)] \tag{24}$$

The matrix Φ_k is the directional matrix-derivative of Θ_k from (14) moving through position q_i with velocity δx_i. In particular, note that

$$\Phi_k(q_i(\tau), \dot{q}_i(\tau), q) = \dot{\Theta}_k(q_i(\tau), q)$$

holds for fixed q by deriving w.r.t. the virtual time τ.

Lemma 1: The mean estimates $\hat{f}_k(q) \equiv \text{E}[f_k(q)|y, \gamma_0]$ of kinetic and elastic GPs (10a) and (11b) in the joint model (19) are analytically guaranteed to be of quadratic form:

$$\hat{f}_k(q) = \frac{1}{2} \overset{(k)\top}{q} \hat{H}_k(q) \overset{(k)}{q}. \tag{25}$$

The posterior matrix estimates $\hat{H}_k$ are decomposed as

$$\hat{H}_k(q) = H_{k0}(q) + \frac{1}{2} \sum_{i=1}^{D} N_{ki}(q) + N_{ki}^\top(q), \tag{26}$$

where the basis matrices N_{ki} are given by

$$N_{ki}(q) = \frac{\partial^k}{\partial \tau^k} \left[\delta x_i \overset{(k)\top}{q_i}(\tau) \circ \Theta_k(q_i(\tau), q) \right] + \frac{1}{2} (-1)^k \overset{(k)(k)\top}{q_i q_i} \circ \Phi_k(q_i, \delta x_i, q) \tag{27}$$

with the subvectors $\delta x_i = [\Delta x_n]_{n=1+(i-1)N}^{iN}$ of the transformed innovation $\Delta x = K_D^{-1} \Delta y$, the Kalman-like kernel gain matrix $K_D = K_y - K_{y0} K_0^{-1} K_{y0}^\top$, and the innovation difference $\Delta y = y - m_y$.

Proof: We set $G \equiv 0$ w.l.o.g., leading to $V = U \equiv f_0$ and $k_{0V} = 0$ due to (16a) and (13). Marginalizing (19) over the complementary $f_{\bar{k}}$ for $\{\bar{k}\} = \{0, 1\} \setminus \{k\}$ and conditioning on y and γ_0 yields another Gaussian distribution with mean

$$\hat{f}_k(q) = m_{f_k}(q) + \begin{bmatrix} 0^\top & k_{yf_k}^\top(q) \end{bmatrix} \begin{bmatrix} K_0 & K_{y0}^\top \\ K_{y0} & K_y \end{bmatrix}^{-1} \begin{bmatrix} \gamma_0 \\ \Delta y \end{bmatrix}.$$

Then, applying standard inversion formulas for partitioned matrices together with $\gamma_0 = 0$, we can write

$$\hat{f}_k(q) = m_{f_k}(q) + k_{yf_k}^\top(q)(K_y - K_{y0} K_0^{-1} K_{y0}^\top)^{-1} \Delta y. \tag{28}$$

Defining $l_{ki}(q) \equiv \mathcal{L}_{q_i} \kappa_k(q_i, q)$ as the application of (17) to (13) such that $k_{yf_k} = [l_{ki}]$, it can be verified that

$$l_{ki}(q) = \frac{1}{2} \frac{\partial^k}{\partial \tau^k} \left[\overset{(k)}{q}(t) \overset{(k)\top}{q_i}(\tau) \circ \Theta_k(q_i(\tau), q(t)) \right] \overset{(k)}{q}(t) + \frac{1}{4} (-1)^k \sum_n \sum_m \overset{(k)}{q_{in}} \overset{(k)}{q_{im}} \overset{(k)(k)}{q_n q_m} \nabla_{q_i} \theta_{knm}(q_i, q)$$

holds after some differential vector-algebraic computations, where we utilize among other intermediate steps that

$$\nabla_{\dot{\boldsymbol{q}}_i} \nabla_{\dot{\boldsymbol{q}}_i}^\top \kappa_k(\boldsymbol{q}_i, \boldsymbol{q}) = \begin{cases} 0\,, & k = 0 \\ \frac{1}{2}\dot{\boldsymbol{q}}\dot{\boldsymbol{q}}^\top \circ \boldsymbol{\Phi}_k(\boldsymbol{q}_i, \dot{\boldsymbol{q}}_i, \boldsymbol{q})\,, & k = 1 \end{cases}.$$

The full derivation is omitted due to space limitations. Further exploiting the commutativity of the Hadamard product and the definition of the projected gradient matrix $\boldsymbol{\Phi}_k$ in (24), we follow that $\delta \boldsymbol{x}_i^\top \boldsymbol{l}_{ki}(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^{(k)\top} \boldsymbol{N}_{ki}(\boldsymbol{q})\boldsymbol{q}^{(k)}$ with $\boldsymbol{N}_{ki}$ as in (27). Then, eliminating the skew-symmetric part of $\boldsymbol{N}_{ki}$ combined with the quadratic priors m_{f_k} from Section IV-B, we reformulate (28) using $\boldsymbol{k}_{\boldsymbol{y}f_k}^\top \boldsymbol{K}_D^{-1}\Delta\boldsymbol{y} = \sum_i \delta \boldsymbol{x}_i^\top \boldsymbol{l}_{ki}$ and obtain (25)–(26). ∎

Essentially, Lemma 1 follows from the quadratic kernel structure (13). The summation (26) represents an extension of matrix decompositions from vector to matrix spaces which are aligned with the Lagrangian structure (1)–(2).

B. Positive Definiteness

We now move on to investigate positive definiteness of the matrix estimates $\hat{\boldsymbol{H}}_k$. Before that, however, we impose a minor structural restriction on the used kernel functions.

Assumption 5: The Cholesky matrix kernel components r_{knm} from (15) and the gravitational covariance k_G from (11a) belong to the class of metric kernels $\mathcal{M}$ given by

$$\mathcal{M} = \left\{ \boldsymbol{\Lambda} \succ \boldsymbol{0} \,\middle|\, k_{\mathcal{M}}(|\boldsymbol{d}|) = \sigma^2 \exp\left(-1/2\boldsymbol{d}^\top \boldsymbol{\Lambda}\boldsymbol{d}\right) \right\} \quad (29)$$

parametrized by $\sigma_{knm}, \boldsymbol{\Lambda}_{knm}$ and $\sigma_G, \boldsymbol{\Lambda}_G$, respectively.

Note that class $\mathcal{M}$ in Assumption 5 includes a wide variety of covariance functions including the commonly used squared exponential kernel [15]. Therein, kernels $k(\boldsymbol{q}, \boldsymbol{q}') \equiv k_{\mathcal{M}}(|\boldsymbol{d}|)$ are Gaussian radial basis functions. They depend only on the squared length of the difference

$$\boldsymbol{d} \coloneqq \boldsymbol{q} - \boldsymbol{q}' \quad (30)$$

in the Riemann space [20, p. 243] defined by metric $\boldsymbol{\Lambda} \succ \boldsymbol{0}$. The special case of variance $\sigma^2 = (2\pi)^{-N/2}|\det \boldsymbol{\Lambda}^{1/2}|$ leads to i.i.d. $\boldsymbol{d} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}^{-1})$.

We can now explicitly state the components r_{knm} of the Cholesky factor $\boldsymbol{R}_k$ from (15). Introducing for $0 \leq n \leq m \leq N$ the functions $\tilde{r}_{knm} \in \mathcal{M}$ from Assumption 5 with identical Riemannian hypermetrics $\boldsymbol{\Lambda}_{knm} = \boldsymbol{\Lambda}_k$, we set

$$r_{knm} \equiv \tilde{r}_{knm}\,, \quad \boldsymbol{\Sigma}_{f_k} \coloneqq [\sigma_{knm}^2]\,, \quad \boldsymbol{\Sigma}_{\boldsymbol{d}_k} \coloneqq \boldsymbol{\Lambda}_k^{-1}\,. \quad (31)$$

Note that the hypervariance $\boldsymbol{\Sigma}_{f_k}$ is upper-triangular.

We now make a stochastic analysis of the data set $\mathcal{D}$ gathered via (3a)–(3b), enabling the probabilistic investigation of positive definiteness preservation. Based on the product $\boldsymbol{p}_i^{(n)}$ and difference $\boldsymbol{d}_j$ vectors

$$\overset{(n)}{\boldsymbol{p}_i} \coloneqq \delta \boldsymbol{x}_i \circ \overset{(n)}{\boldsymbol{q}_i}\,, \quad \forall n \in \{k, 2k\}\,, \quad (32a)$$
$$\boldsymbol{d}_j \coloneqq \boldsymbol{q} - \boldsymbol{q}_j\,, \quad \forall j = 1, \ldots, D\,, \quad (32b)$$

we define for $\boldsymbol{\mu}_k, \boldsymbol{\nu}_k \in \mathbb{R}^D$ the composite random variable

$$\boldsymbol{\beta}_k \coloneqq \boldsymbol{\mu}_k + \boldsymbol{\nu}_k \quad (33)$$

for those $i \in \{1, \ldots, D\}$ which satisfy component-wise $\forall n \in \{k, 2k\}$, $k \in \{0, 1\}$, the condition $\boldsymbol{p}_i^{(n)} > \boldsymbol{0}$, by

$$\mu_{ki} = \Upsilon_k\left(\overset{(2k)}{\boldsymbol{p}_i}\right) \det \operatorname{diag}\left(\boldsymbol{\varrho}_k^\downarrow \circ \overset{(2k)\uparrow}{\boldsymbol{p}_i}\right)$$
$$+ 2k\dot{\vartheta}_{ki} \begin{cases} \Upsilon_k\left(\overset{(k)}{\boldsymbol{p}_i}\right) \det \operatorname{diag}\left(\boldsymbol{\varrho}_k^\downarrow \circ \overset{(k)\uparrow}{\boldsymbol{p}_i}\right) & \dot{\vartheta}_{ki} \geq 0 \\ \varrho_{k1} \max_n \overset{(k)}{p_{in}} & \dot{\vartheta}_{ki} < 0 \end{cases}$$
$$\nu_{ki} = \varphi_{ki} \begin{cases} \Upsilon_k\left(\overset{(k)\circ 2}{\boldsymbol{q}_i}\right) \det \operatorname{diag}\left(\boldsymbol{\varrho}_k^\downarrow \circ \overset{(k)\circ 2\uparrow}{\boldsymbol{q}_i}\right) & \varphi_{ki} \geq 0 \\ \varrho_{k1} \max_n \overset{(k)_2}{q_{in}} & \varphi_{ki} < 0 \end{cases}$$

with constant, decreasingly ordered, radial vector $\boldsymbol{\varrho}_k^\downarrow \in \mathbb{R}^N$, $\boldsymbol{\varrho}_k^\downarrow > \boldsymbol{0}$. Here, the projection φ_{ki}, the differential angle $\dot{\vartheta}_{ki}$ and the normalizing function $\Upsilon_k(\boldsymbol{p}_i)$ are given by

$$\varphi_{ki}(\delta\boldsymbol{x}_i) = (-1)^k \boldsymbol{d}_i^\top \boldsymbol{\Lambda}_k \delta \boldsymbol{x}_i\,, \quad (34a)$$
$$\dot{\vartheta}_{ki}(\boldsymbol{q}_i(\tau)) = \varphi_{ki}(\dot{\boldsymbol{q}}_i(\tau))\,, \quad (34b)$$
$$\Upsilon_k(\boldsymbol{p}_i) = 1/\Pi_{j=1}^{N-1}\lambda_j(\boldsymbol{\Sigma}_{f_k}^2 \operatorname{diag}(\boldsymbol{p}_i))\,. \quad (34c)$$

If $\boldsymbol{p}_i^{(n)} > \boldsymbol{0}$ does not hold $\forall n \in \{k, 2k\}$, we define

$$\beta_{ki} \coloneqq \exp(\|\boldsymbol{d}_i\|_{\boldsymbol{\Lambda}_k}^2)\lambda_N(\hat{\boldsymbol{H}}_{ki})\,, \quad \overset{(n)}{\boldsymbol{p}_i} \not> \boldsymbol{0}\,, \quad (35)$$

where $\hat{\boldsymbol{H}}_{ki} = \frac{1}{2}(\boldsymbol{N}_{ki} + \boldsymbol{N}_{ki}^\top)$ from (26). The random variable $\boldsymbol{\beta}_k$ from (33)–(35) is the result of nonlinearly transformed GRVs stemming from the data set $\mathcal{D}$. We now formulate the main result.

Theorem 1: Consider the posterior, generalized, matrix-valued L-GP estimate $\hat{\boldsymbol{H}}_k(\boldsymbol{q})$ from (26) with metric covariances $r_{knm} \in \mathcal{M}$ forming the Cholesky kernel $\boldsymbol{\Theta}_k(\boldsymbol{q}_i, \boldsymbol{q})$ from (14) identically specified by Riemannian hypermetric $\boldsymbol{\Sigma}_{\boldsymbol{d}_k}^{-1}$ via Assumption 5 and (31). Positive definiteness of the generalized matrix estimate is guaranteed with probability

$$\Pr\{\hat{\boldsymbol{H}}_k(\boldsymbol{q}) \succ \boldsymbol{0}\} = 1 - \delta_k(\boldsymbol{q})\,, \quad (36)$$

where the physical inconsistency measure $\delta_k(\boldsymbol{q})$ is upper bounded according to

$$\delta_k(\boldsymbol{q}) \leq \Pr\left\{\lambda_N(\boldsymbol{H}_{k0}) + \exp\left[-\operatorname{vec}(\boldsymbol{\Sigma}_{\boldsymbol{d}_k}^{-1})^\top \boldsymbol{D}(\boldsymbol{q})\right]\boldsymbol{\beta}_k \leq 0\right\}$$

based on the column-wise Kronecker-constructed squared distance matrix

$$\boldsymbol{D}(\boldsymbol{q}) = [\boldsymbol{d}_j(\boldsymbol{q}) \otimes \boldsymbol{d}_j(\boldsymbol{q})]$$

which depends only on the absolute values of the positional joint distances $|\boldsymbol{d}_j(\boldsymbol{q})| = |\boldsymbol{q} - \boldsymbol{q}_j|$.

Proof: Firstly, we express the basis matrices (27) from Lemma 1 writing $\boldsymbol{\Theta}_k \equiv \tilde{\boldsymbol{\Theta}}_k(\boldsymbol{d})$ based on Assumption 5 and (31). Proceeding analogously via $\boldsymbol{\Phi}_k \equiv \tilde{\boldsymbol{\Phi}}_k(\boldsymbol{d}, \delta\boldsymbol{x}_i)$, we follow that

$$\tilde{\boldsymbol{\Phi}}_k(\boldsymbol{d}_i, \delta\boldsymbol{x}_i) = (-1)^k 2\varphi_{ki}(\delta\boldsymbol{x}_i)\tilde{\boldsymbol{\Theta}}_k(\boldsymbol{d}_i)$$

with the projection φ_{ki} and the difference vector $\boldsymbol{d}_i$ as in (34a) and (32b), respectively. Then, we can write for each matrix summand $\boldsymbol{N}_{ki}$ in (27) that

$$\boldsymbol{N}_{ki} = \bar{\boldsymbol{N}}_{ki} \circ \tilde{\boldsymbol{\Theta}}_k(|\boldsymbol{d}_i|)\,,$$

where

$$\bar{\boldsymbol{N}}_{ki} = \sum_{n=0}^{k} \left(2\dot{\vartheta}_{ki}(\boldsymbol{q}_i)\right)^{k-n} \delta\boldsymbol{x}_i^{(n+k)\top}\boldsymbol{q}_i^\top + \varphi_{ki}(\delta\boldsymbol{x}_i)^{(k)}\boldsymbol{q}_i^{(k)\top} \quad (37)$$

dropping dependencies on $\boldsymbol{q}$ for the sake of notational brevity. Consider now the i-th summand $\hat{\boldsymbol{H}}_{ki} = \frac{1}{2}(\boldsymbol{N}_{ki} + \boldsymbol{N}_{ki}^\top)$ from (26). Exploiting

$$\tilde{\boldsymbol{\Theta}}_k(|\boldsymbol{d}_i|) = \exp(-\|\boldsymbol{d}_i\|_{\boldsymbol{\Lambda}_k}^2)\boldsymbol{\Sigma}_{f_k}^2\,,$$

where $\boldsymbol{\Sigma}_{f_k}^2 = \boldsymbol{\Sigma}_{f_k}^\top\boldsymbol{\Sigma}_{f_k}$, and its eigenvalue decomposition

$$\boldsymbol{\Sigma}_{f_k}^2 = \boldsymbol{W}_k\text{diag}(\boldsymbol{\varrho}_k)\boldsymbol{W}_k^\top$$

with decreasingly ordered $\boldsymbol{\varrho}_k = \boldsymbol{\lambda}^\downarrow(\boldsymbol{\Sigma}_{f_k}^2) > \boldsymbol{0}$, we now leverage the eigenvalue behavior of the sum of Hermitian matrices [21] in combination with Weyl's inequality [22, Theorem III.2.1]. Thus, with the Hermitian summation $\hat{\boldsymbol{H}}_k = \sum_i \hat{\boldsymbol{H}}_{ki}$ as in (26), we derive

$$\lambda_N(\hat{\boldsymbol{H}}_k) \geq \sum_{i=1}^{D} \lambda_N(\hat{\boldsymbol{H}}_{ki})\,.$$

Then, utilizing the product constructions (37), we obtain

$$\lambda_N(\hat{\boldsymbol{H}}_k) \geq \lambda_N(\boldsymbol{H}_{k0}) + \exp\left[-\text{vec}(\boldsymbol{\Lambda}_k)^\top\boldsymbol{D}\right]\boldsymbol{\beta}_k \quad (38)$$

by successively exploiting the invariance of eigenvalues w.r.t. similarity transforms and by using $\|\boldsymbol{d}_i\|_{\boldsymbol{\Lambda}_k}^2 = (\boldsymbol{d}_i \otimes \boldsymbol{d}_i)^\top\text{vec}(\boldsymbol{\Lambda}_k)$ as well as the definition of the random vector $\boldsymbol{\beta}_k \in \mathbb{R}^D$ as given component-wise by (33)–(35). Taking the complement of the set in which the lower bound (38) is positive, we arrive at (36) and finish the proof. ∎

Note that, for a given data set $\mathcal{D}$, the upper bound on δ_k in Theorem 1 can be used to deterministically estimate the domain in which positive definiteness is fulfilled. A probabilistic computation can be made by means of the distribution of products and quotients of continuous, nonstandardized, independent Normal random variables [23].

C. Equilibria

Despite the point-wise inclusion of $\boldsymbol{\gamma}_0 = \boldsymbol{0}$ in the distribution (19) being stochastic, we can provide an analytical guarantee for the preservation of the equilibrium at the origin.

Theorem 2: The L-GP-based potential mean estimate $\hat{V}(\boldsymbol{q}) \equiv \text{E}[V(\boldsymbol{q})|\boldsymbol{y},\boldsymbol{\gamma}_0]$ is guaranteed to have an equilibrium:

$$\hat{V}(\boldsymbol{0}) = 0\,, \quad \nabla_{\boldsymbol{q}}\hat{V}(\boldsymbol{0}) = \boldsymbol{0}\,. \quad (39)$$

Proof: Analogously to the derivation of (28), we now marginalize (19) over T instead of V, and condition on $\boldsymbol{y}$ as well as $\boldsymbol{\gamma}_0$. This leads to $\hat{V}(\boldsymbol{q}) \equiv \text{E}[V(\boldsymbol{q})|\boldsymbol{y},\boldsymbol{\gamma}_0]$ with

$$\hat{V}(\boldsymbol{q}) = m_V(\boldsymbol{q}) + \left(\boldsymbol{k}_{\boldsymbol{y}V}^\top - \boldsymbol{k}_{0V}^\top\boldsymbol{K}_0^{-1}\boldsymbol{K}_{\boldsymbol{y}0}^\top\right)\boldsymbol{K}_D^{-1}\Delta\boldsymbol{y}\,. \quad (40)$$

Based on the quadratic form of $\hat{U}$ proven in Lemma 1, we set $U \equiv 0$ w.l.o.g. and obtain $V = G$. Exploiting the structure (29) of the gravitational kernel k_G according to Assumption 5 then leads to $\boldsymbol{k}_{0V}^\top(\boldsymbol{0}) = [\sigma_G^2, \boldsymbol{0}^\top]$, block-diagonal $\boldsymbol{K}_0 = \sigma_G^2\text{diag}(1,\boldsymbol{\Lambda}_G)$ and thus $\boldsymbol{k}_{0V}^\top(\boldsymbol{0})\boldsymbol{K}_0^{-1} = [1, \boldsymbol{0}^\top]$. Therefore,

$$\boldsymbol{k}_{\boldsymbol{y}V}^\top(\boldsymbol{0}) = \boldsymbol{k}_{0V}^\top(\boldsymbol{0})\boldsymbol{K}_0^{-1}\boldsymbol{K}_{\boldsymbol{y}0}^\top\,. \quad (41)$$

Plugging (41) into (40) and using $m_G(\boldsymbol{0}) = 0$ due to Assumption 3, we follow $\hat{V}(\boldsymbol{0}) = 0$ holds. Similarly applying the ∇ operator to (40), we compute $\nabla_{\boldsymbol{q}}\boldsymbol{k}_{\boldsymbol{y}V}^\top(\boldsymbol{0}) = \nabla_{\boldsymbol{q}}\boldsymbol{k}_{0V}^\top(\boldsymbol{0})\boldsymbol{K}_0^{-1}\boldsymbol{K}_{\boldsymbol{y}0}^\top$. Thus we can conclude that $\nabla_{\boldsymbol{q}}\hat{V}(\boldsymbol{0}) = \boldsymbol{0}$ holds, again for $\nabla_{\boldsymbol{q}}m_G(\boldsymbol{0}) = \boldsymbol{0}$ as in Assumption 3. ∎

Corollary 1: The estimative Lagrangian $\hat{L} \equiv \text{E}[L|\boldsymbol{y},\boldsymbol{\gamma}_0]$ based on (19) is analytically guaranteed to have a stationary point at $\boldsymbol{q} = \dot{\boldsymbol{q}} = \boldsymbol{0}$, i.e., $\nabla_{\boldsymbol{q},\dot{\boldsymbol{q}}}\hat{L}(\boldsymbol{0},\boldsymbol{0}) = \boldsymbol{0}$, with $\hat{L}(\boldsymbol{0},\boldsymbol{0}) = 0$.

Proof: Since $\hat{L} = \hat{T} - \hat{G} - \hat{U}$, this directly follows from the preservation of quadratic kinetic and elastic forms proven in Lemma 1 combined with the guaranteed gravitational equilibrium from Theorem 2, i.e., $\hat{G}(\boldsymbol{0}) = 0$. ∎

D. Energy Conservation

Lastly, the L-GP is also equivalent to a dynamic system. We investigate this intuition in the following.

Proposition 2: Upon explicit inclusion of a test torque $\boldsymbol{\tau}(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}})$ in the form of (18) in (19), the conditional expectation $\hat{\boldsymbol{\tau}}(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}}) \equiv \text{E}[\boldsymbol{\tau}(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}})|\boldsymbol{y},\boldsymbol{\gamma}_0]$ is given by

$$\hat{\boldsymbol{\tau}}(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}}) = \hat{\boldsymbol{M}}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \hat{\boldsymbol{C}}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \hat{\boldsymbol{g}}(\boldsymbol{q}) \quad (42)$$

with $\hat{\boldsymbol{C}} = \frac{1}{2}[\nabla_{\dot{\boldsymbol{q}}}\nabla_{\boldsymbol{q}}^\top\hat{T} + \dot{\hat{\boldsymbol{M}}} - \nabla_{\boldsymbol{q}}\nabla_{\dot{\boldsymbol{q}}}^\top\hat{T}]$, $\hat{\boldsymbol{M}}$ from (26) for $k = 1$, and $\hat{\boldsymbol{g}} = \nabla_{\boldsymbol{q}}\hat{V}$ from (40) and (26) for $k = 0$.

Proof: Follows as a direct consequence of the kernel construction via the Lagrangian-differential operator (17) and the equivalent representation of (1) by (2). Thus, we include (18) in (19) and get

$$\text{E}[\boldsymbol{\tau}(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}})|\boldsymbol{y},\boldsymbol{\gamma}_0] \equiv \mathcal{L}_{\boldsymbol{q},\dot{\boldsymbol{q}}}\hat{L}(\boldsymbol{q},\dot{\boldsymbol{q}}) = \hat{\boldsymbol{\tau}}(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}})$$

after marginalization and conditioning, proving the result. ∎

We now show that the equivalent L-GP dynamics (42) preserve the physical property of energy conservation as a last result. For notational consistency, we use the (estimative) state vector $\hat{\boldsymbol{x}}^\top = [\boldsymbol{q}^\top, \dot{\boldsymbol{q}}^\top]$ in the following.

Corollary 2: Consider the system $\dot{\hat{\boldsymbol{x}}} = \hat{\boldsymbol{\phi}}(\hat{\boldsymbol{x}},\hat{\boldsymbol{u}})$ defined $\forall\hat{\boldsymbol{x}} \in \boldsymbol{\Xi}$ in the compact domain $\boldsymbol{\Xi} \subseteq \mathcal{X} \subseteq \mathbb{R}^{2N}$ with

$$\boldsymbol{\Xi} = \{\boldsymbol{0}\} \cup \{\hat{\boldsymbol{M}}(\hat{\boldsymbol{x}}_1) \succ \boldsymbol{0} \wedge \hat{V}(\hat{\boldsymbol{x}}_1) \geq -1/2\lambda_N(\hat{\boldsymbol{M}})\|\hat{\boldsymbol{x}}_2\|^2\}$$

spanned $\forall\hat{\boldsymbol{u}} \in \mathbb{R}^N$ by the GP (18) with estimate (42), where

$$\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{x}},\hat{\boldsymbol{u}}) = \begin{bmatrix} \hat{\boldsymbol{x}}_2 \\ \hat{\boldsymbol{M}}^{-1}(\hat{\boldsymbol{x}}_1)(\hat{\boldsymbol{u}} - \hat{\boldsymbol{C}}(\hat{\boldsymbol{x}}_1,\hat{\boldsymbol{x}}_2)\hat{\boldsymbol{x}}_2 - \hat{\boldsymbol{g}}(\hat{\boldsymbol{x}}_1)) \end{bmatrix}\,. \quad (43)$$

The dynamic system (43) with state vector $\hat{\boldsymbol{x}}^\top = [\hat{\boldsymbol{x}}_1^\top, \hat{\boldsymbol{x}}_2^\top]$ and input $\hat{\boldsymbol{u}} = \hat{\boldsymbol{\tau}}$ is passive and, moreover, lossless, with respect to the energy storage function $\hat{E} = \hat{T} + \hat{V}$, where

$$\hat{E}(\hat{\boldsymbol{x}}) = 1/2\hat{\boldsymbol{x}}_2^\top\hat{\boldsymbol{M}}(\hat{\boldsymbol{x}}_1)\hat{\boldsymbol{x}}_2 + \hat{V}(\hat{\boldsymbol{x}}_1)\,, \quad (44)$$

based on (25), (40), and the system output mapping $\hat{\boldsymbol{y}} = \hat{\boldsymbol{x}}_2$.

Proof: From the definition of $\boldsymbol{\Xi}$, we can follow that (43) is locally Lipschitz in the domain $\boldsymbol{\Xi} \times \mathbb{R}^N$. Also, the storage function (44) is positive semidefinite, since $\hat{E}(\boldsymbol{0}) = 0$ due to Corollary 1 and $\lambda_N(\hat{\boldsymbol{M}})\|\hat{\boldsymbol{x}}_2\|^2 \leq \hat{\boldsymbol{x}}_2^\top\hat{\boldsymbol{M}}\hat{\boldsymbol{x}}_2$ for $\hat{\boldsymbol{M}} \succ \boldsymbol{0}$ ensures $\hat{E}(\hat{\boldsymbol{x}}) \geq 0$, $\forall\boldsymbol{x} \in \boldsymbol{\Xi}$. The output function $\hat{\boldsymbol{y}} = \hat{\boldsymbol{x}}_2$ is continuous and for the dynamics $\hat{\boldsymbol{\phi}}(\boldsymbol{0},\boldsymbol{0}) = \boldsymbol{0}$ holds. Taking

the time derivative of (44) along the trajectories (43), we obtain the power fed into the system by input $\hat{\boldsymbol{u}}$ computing

$$\dot{E} = \hat{\boldsymbol{x}}_2^\top \left(\hat{\boldsymbol{M}}(\hat{\boldsymbol{x}}_1)\dot{\hat{\boldsymbol{x}}}_2 + \frac{1}{2}\dot{\hat{\boldsymbol{M}}}(\hat{\boldsymbol{x}}_1)\hat{\boldsymbol{x}}_2 + \nabla_{\hat{\boldsymbol{x}}_1}\hat{V}(\hat{\boldsymbol{x}}_1) \right)$$
$$= \hat{\boldsymbol{x}}_2^\top \left(\hat{\boldsymbol{M}}(\hat{\boldsymbol{x}}_1)\dot{\hat{\boldsymbol{x}}}_2 + \hat{C}(\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2)\hat{\boldsymbol{x}}_2 + \hat{\boldsymbol{g}}(\hat{\boldsymbol{x}}_1) \right)$$
$$= \hat{\boldsymbol{y}}^\top \hat{\boldsymbol{u}},$$

where we have exploited the symmetry of $\hat{\boldsymbol{M}} = \hat{\boldsymbol{M}}^\top$, the identity $\dot{\hat{\boldsymbol{M}}} = \hat{C} + \hat{C}^\top$ and the skew-symmetry of $\dot{\hat{\boldsymbol{M}}} - 2\hat{C}^\top$. Thus, according to [24], the system (43) is lossless. ∎

Note that from the passivity of system (43) we can also directly follow stability of the origin of the unforced system $\dot{\hat{\boldsymbol{x}}} = \hat{\boldsymbol{\phi}}(\hat{\boldsymbol{x}}, \boldsymbol{0})$ by taking (44) as a Lyapunov function [24].

VI. NUMERICAL ILLUSTRATIONS

In this section, we now demonstrate the efficacy of our approach and validate the derived theoretical results, choosing a simple example for the sake of easy interpretation.

A. Setup

We benchmark our method on the two-link robotic manipulator from [12, p. 164]. Gravity $g = 10$ ms^{-2} acts along the positive x-axis, leading to an equilibrium at the origin $\boldsymbol{q} = \boldsymbol{0}$. The links have unit masses $m_n = 1$ kg and unit lengths $l_n = 1$ m for $n \in \{1, 2\}$. Estimates are available but erroneous: $\hat{m}_n = (1 + \Delta_n)m_n$, $\hat{l}_n = (1 + \Delta_n)l_n$, where $\Delta_n = (-1)^n/2$.

For all of the numerical simulations, we use only 25 training pairs equally distanced on the domain $\boldsymbol{q} \in [-1, 1]^2$ with fixed $\dot{q}_n = (-1)^n$ and $\ddot{\boldsymbol{q}} = \boldsymbol{1}$. The torque measurement noise has covariance $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_i} = \sigma_{\boldsymbol{\varepsilon}}^2 \boldsymbol{I}$, $\sigma_{\boldsymbol{\varepsilon}} = 0.1$ Nm, while the differential process noise corrupting the accelerations has covariance $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i} = \sigma_{\alpha}^2 \boldsymbol{I}$, $\sigma_{\alpha} = \pi/180$ rad/s^2. Positions and velocities are kept noisefree, enabling explicit numerical integration with ode45. We reduce the kinetic mass inertia hypermetrics $\boldsymbol{\Lambda}_{knm} = \boldsymbol{\Lambda}_k$, $\forall n, m \in \{1, 2\}$, to the constant Euclidian form $\boldsymbol{\Lambda}_k^{-1} = \sigma_{\boldsymbol{d}_T}^2 \boldsymbol{I}$, where $k = 1$, and fix $\sigma_{\boldsymbol{d}_T} = 10^2$. The same is done for the gravitational distance covariance keeping $\boldsymbol{\Sigma}_{\boldsymbol{d}_G}^{1/2} = \operatorname{diag}([1.6, 2.7])$. The remaining radial gravitational hypervariance σ_G^2 along with the Cholesky-kinetic upper-triangular hypercovariance $\boldsymbol{\Sigma}_{f_k}$, where again $k = 1$, are optimized via the log-likelihood.

B. Results

As a first experiment, we demonstrate the closed-loop applicability of the L-GP model. For this, we compare a standard PD tracking controller [12, p. 194] with its L-GP-based version as shown in Fig. 1. Starting from the equilibrium, the controllers with gains $\boldsymbol{K}_p = \boldsymbol{K}_d = 10\boldsymbol{I}$ have the task of following a sinusoidal reference trajectory $\boldsymbol{q}_d(t) = a_d \sin t \boldsymbol{1}$ with amplitude $a_d = \pi/2$. The simulation shows that the L-GP-based version considerably improves the accuracy, demonstrating reliable performance despite the suboptimal distribution of the data points, independent from the reference trajectory, requiring substantial extrapolation.
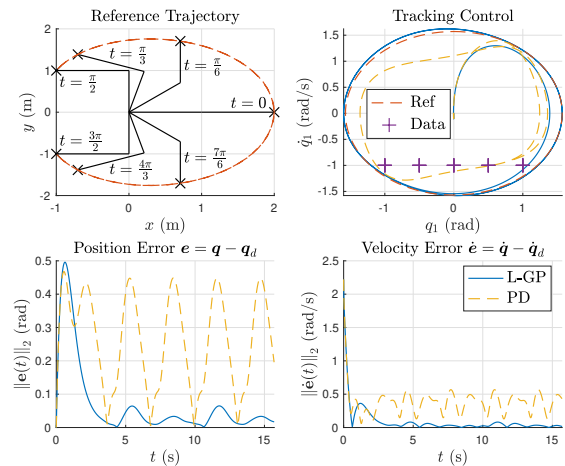

Fig. 1. Tracking performance of the standard PD controller (dashed yellow line) and its L-GP augmented version (solid blue line). The top left plot shows the reference trajectory (dashed red line) w.r.t. the end-effector in its cartesian work space, the top right the closed-loop behavior in the state space of the first joint along with the training data points (purple crosses). The bottom two subfigures indicate the Euclidian norms of position and velocity tracking errors over time.
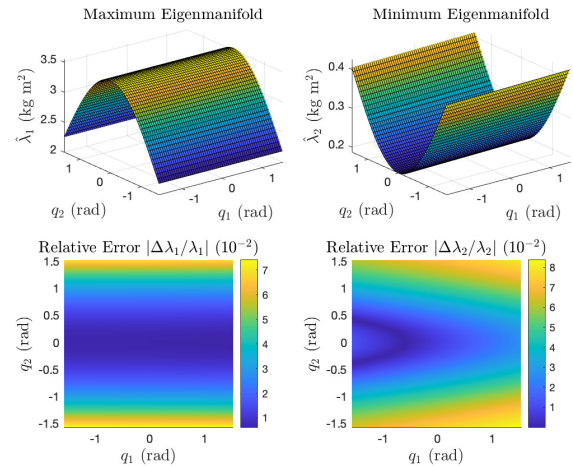

Fig. 2. Eigenmanifolds of the generalized inertia matrix of the two-link robot over the domain $\boldsymbol{q} \in [-\pi/2, \pi/2]^2$. The top left and right surface plots show the maximum and minimum estimative eigenvalue manifolds of the L-GP-based estimate $\hat{\boldsymbol{M}}(\boldsymbol{q})$, respectively, while the bottom two subfigures are heat maps indicating the relative approximation errors in percent.

Next, we validate the physical consistency of the L-GP. Therefore, using the same parametrization as in the previous experiment, we evaluate the mass inertia estimate over the joint domain $\boldsymbol{q} \in [-a_d, a_d]^2$ and investigate its eigenmanifolds, as shown in Fig. 2. Clearly, the L-GP accurately approximates the positive definite function space despite only being a subcomponent of the input-output relation. The consistency of the gravitational potential estimate with equilibrium is validated in Fig. 3, along with the conservatism of the equivalent L-GP-based dynamics simulated for different initial conditions $\hat{\boldsymbol{x}}(0) = [a_0 \boldsymbol{1}^\top, \boldsymbol{0}^\top]^\top$ with displacement amplitudes $a_0 = 0.1$ (red line), $a_0 = 0.5$ (yellow line)
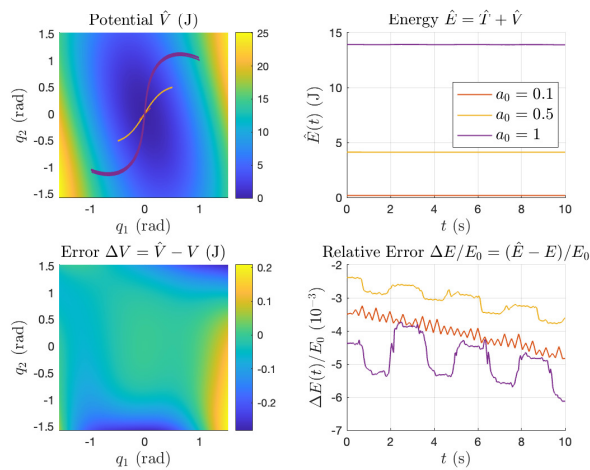
Fig. 3. Potential energy estimate and free trajectories of the L-GP equivalent dynamic system (top left plot) with initial conditions $\boldsymbol{q}(0) = a_0\mathbf{1}$ and $\dot{\boldsymbol{q}}(0) = \mathbf{0}$ for $a_0 = 0.1$ (red line), $a_0 = 0.5$ (yellow line) and $a_0 = 1$ (purple line). The constant time evolution of the energy estimates are shown in the top right plot for the same trajectories. The bottom left plot visualizes the signed potential approximation error, the bottom right one the signed relative energy error in permille.

and $a_0 = 1$ (purple line). The signed potential and energy approximation errors show their properties of local positive definiteness and passivity, respectively.

VII. CONCLUSION

This paper presented an approach for the identification of uncertain Lagrangian systems exploiting prior physical knowledge by kernel construction. Physical consistency of the data-driven method in terms of guarantuees for the fulfillment of certain properties was proven rigorously and validated in numerical simulation, along with the effective applicability of the model to an exemplary tracking control problem. Future research will focus on the extension of the approach to handle dissipative and time-variant systems. Another promising direction is the application of the Lagrangian variance estimate for uncertainty quantification.

VIII. ACKNOWLEDGMENTS

REFERENCES

[1] A. Gahlawat, P. Zhao, A. Patterson, N. Hovakimyan, and E. Theodorou, "L1-gp: L1 adaptive control with bayesian learning," in *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, vol. 120 of *Proceedings of Machine Learning Research*, pp. 826–837, PMLR, 10–11 Jun 2020.

[2] F. Berkenkamp, R. Moriconi, A. P. Schoellig, and A. Krause, "Safe learning of regions of attraction for uncertain, nonlinear systems with gaussian processes," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 4661–4666, 2016.

[3] M. P. Deisenroth, D. Fox, and C. E. Rasmussen, "Gaussian processes for data-efficient learning in robotics and control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 2, pp. 408–423, 2015.

[4] L. Ljung, *System Identification: Theory for the User*. PTR Prentice Hall Information and System Sciences, Pearson, 2nd ed., 1998.

[5] Q. Leboutet, J. Roux, A. Janot, J. R. Guadarrama-Olvera, and G. Cheng, "Inertial parameter identification in robotics: A survey," *Applied Sciences*, vol. 11, no. 9, 2021.

[6] A. A. Amiri Moghadam, K. Torabi, A. Kaynak, M. N. H. Zainal Alam, A. Kouzani, and B. Mosadegh, "Control-oriented modeling of a polymeric soft robot," *Soft Robotics*, vol. 3, no. 2, pp. 82–97, 2016.

[7] R. E. Day, "Coupling dynamics in aircraft: A historical perspective," special publication, NASA Dryden Flight Research Center, Edwards, CA, 1997.

[8] T. I. Fossen and O.-E. Fjellstad, "Nonlinear modelling of marine vehicles in 6 degrees of freedom," *Mathematical Modelling of Systems*, vol. 1, no. 1, pp. 17–27, 1995.

[9] K. Rath, C. G. Albert, B. Bischl, and U. von Toussaint, "Symplectic gaussian process regression of maps in hamiltonian systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 5, p. 053121, 2021.

[10] C.-A. Cheng and H.-P. Huang, "Learn the lagrangian: A vector-valued rkhs approach to identifying lagrangian systems," *IEEE Transactions on Cybernetics*, vol. 46, no. 12, pp. 3247–3258, 2016.

[11] T. Beckers, D. Kulić, and S. Hirche, "Stable gaussian process based tracking control of euler–lagrange systems," *Automatica*, vol. 103, pp. 390–397, 2019.

[12] R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

[13] A. Mchutchon and C. Rasmussen, "Gaussian process training with input noise," in *Advances in Neural Information Processing Systems*, vol. 24, Curran Associates, Inc., 2011.

[14] J. E. Johnson, V. Laparra, and G. Camps-Valls, "Accounting for input noise in gaussian process parameter retrieval," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 3, pp. 391–395, 2020.

[15] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[16] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Inferring solutions of differential equations using noisy multi-fidelity data," *Journal of Computational Physics*, vol. 335, pp. 736–746, 2017.

[17] E. Solak, R. Murray-smith, W. Leithead, D. Leith, and C. Rasmussen, "Derivative observations in gaussian process models of dynamic systems," in *Advances in Neural Information Processing Systems*, vol. 15, MIT Press, 2002.

[18] M. A. Álvarez, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: A review," *Foundations and Trends® in Machine Learning*, vol. 4, no. 3, pp. 195–266, 2012.

[19] S. Reece and S. Roberts, "An introduction to gaussian processes for the kalman filter expert," in *2010 13th International Conference on Information Fusion*, pp. 1–9, 2010.

[20] D. Lovelock and H. Rund, *Tensors, Differential Forms, and Variational Principles*. Dover Publications, 1989.

[21] A. Knutson and T. Tao, "Honeycombs and sums of hermitian matrices," *Notices Of The American Mathematical Society*, vol. 48, no. 2, pp. 175–186, 2001.

[22] R. Bhatia, *Matrix Analysis*. Graduate Texts in Mathematics, Springer, New York, NY, 1 ed., 1997.

[23] M. D. Springer, *The Algebra of Random Variables*. John Wiley & Sons, Inc., 1979.

[24] H. K. Khalil, *Nonlinear Systems*. Prentice Hall, third ed., 2002.