

Physically Consistent Gaussian Processes for Learning-based Control of Euler-Lagrange Systems

Giulio Evangelisti
Sandra Hirche

Chair of Information-oriented Control
Technical University of Munich

Workshop 5: Physics-informed Machine Learning for Modeling, Control, and Optimization
2024 American Control Conference (ACC), 07/09/2024

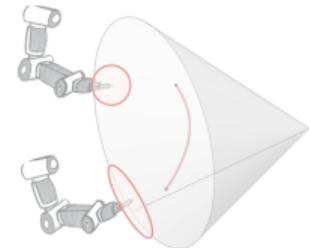
Motivation

Fundamental Design Procedure of Passivity-based Control [Ortega+ 1998]

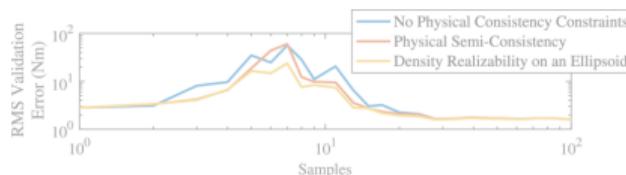
Practically meaningful nonlinear control theory should start with a practically meaningful system class & account for its physical structure \Rightarrow Euler-Lagrange systems

Benefits of physical structuring:

- Reliability: performance depends on accuracy & consistency
 - improved generalization via physical integrity
 - physical constraints such as $M \succ 0$ for safety & stability



[Jaquier+ 2022]



[Wensing+ 2018]

- Data efficiency: in engineering practice often only limited trajectory data available

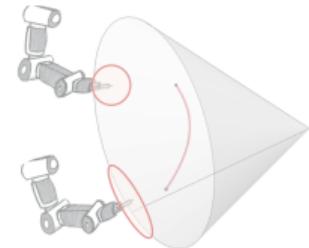
Motivation

Fundamental Design Procedure of Passivity-based Control [Ortega+ 1998]

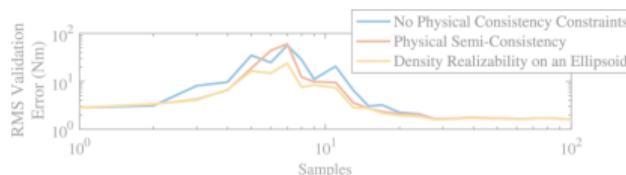
Practically meaningful nonlinear control theory should start with a practically meaningful system class & account for its physical structure \Rightarrow Euler-Lagrange systems

Benefits of physical structuring:

- Reliability: performance depends on accuracy & consistency
 - improved generalization via physical integrity
 - physical constraints such as $M \succ 0$ for safety & stability



[Jaquier+ 2022]



[Wensing+ 2018]

- Data efficiency: in engineering practice often only limited trajectory data available

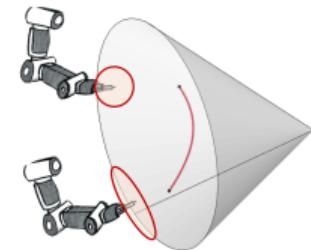
Motivation

Fundamental Design Procedure of Passivity-based Control [Ortega+ 1998]

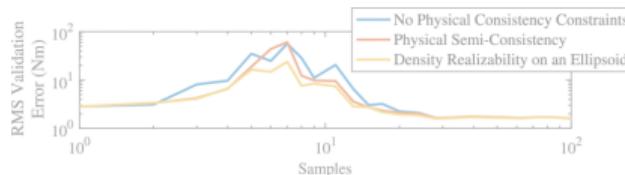
Practically meaningful nonlinear control theory should start with a practically meaningful system class & account for its physical structure \Rightarrow Euler-Lagrange systems

Benefits of physical structuring:

- Reliability: performance depends on accuracy & consistency
 - improved generalization via physical integrity
 - physical constraints such as $M \succ 0$ for safety & stability



[Jaquier+ 2022]



[Wensing+ 2018]

- Data efficiency: in engineering practice often only limited trajectory data available

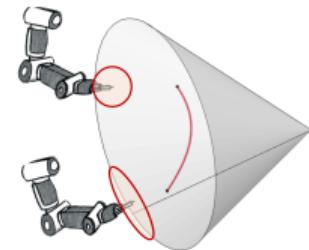
Motivation

Fundamental Design Procedure of Passivity-based Control [Ortega+ 1998]

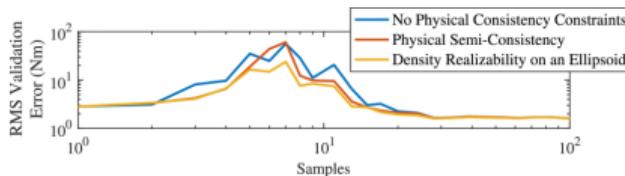
Practically meaningful nonlinear control theory should start with a practically meaningful system class & account for its physical structure \Rightarrow Euler-Lagrange systems

Benefits of physical structuring:

- Reliability: performance depends on accuracy & consistency
 - improved generalization via physical integrity
 - physical constraints such as $M \succ 0$ for safety & stability



[Jaquier+ 2022]



[Wensing+ 2018]

- Data efficiency: in engineering practice often only limited trajectory data available

Motivation

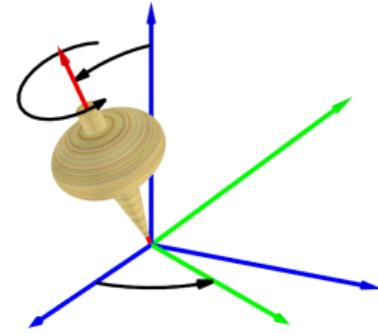
Classical Rigid Body Dynamics

Linearity in the parameters [Spong+ 2020]:

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) =: \boldsymbol{Y}(\boldsymbol{q}, \dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}})\boldsymbol{\theta} = \boldsymbol{\tau}$$

with

- inertial parameter vector $\boldsymbol{\theta} \in \mathbb{R}^P$
- regressor matrix $\boldsymbol{Y}(\boldsymbol{q}, \dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}}) \in \mathbb{R}^{N \times P}$
- $P \leq 10N$ (masses, COG coordinates, MOI tensors)



Creative Commons

Assumptions: rigid links, known inertia distribution, etc.

Motivation

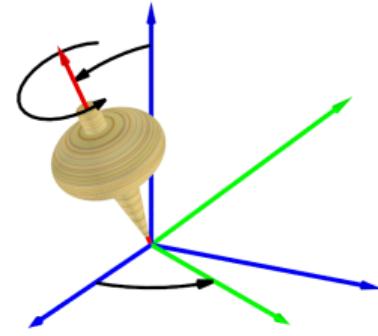
Classical Rigid Body Dynamics

Linearity in the parameters [Spong+ 2020]:

$$M(\boldsymbol{q})\ddot{\boldsymbol{q}} + C(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) =: \boldsymbol{Y}(\boldsymbol{q}, \dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}})\boldsymbol{\theta} = \boldsymbol{\tau}$$

with

- inertial parameter vector $\boldsymbol{\theta} \in \mathbb{R}^P$
- regressor matrix $\boldsymbol{Y}(\boldsymbol{q}, \dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}}) \in \mathbb{R}^{N \times P}$
- $P \leq 10N$ (masses, COG coordinates, MOI tensors)



Creative Commons

Assumptions: rigid links, known inertia distribution, etc.

Motivation

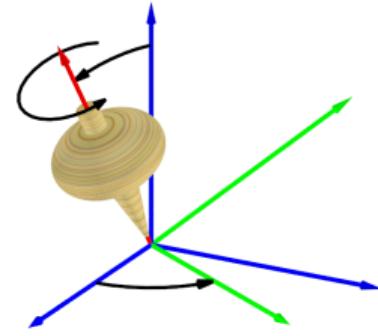
Classical Rigid Body Dynamics

Linearity in the parameters [Spong+ 2020]:

$$M(\boldsymbol{q})\ddot{\boldsymbol{q}} + C(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) =: \boldsymbol{Y}(\boldsymbol{q}, \dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}})\boldsymbol{\theta} = \boldsymbol{\tau}$$

with

- inertial parameter vector $\boldsymbol{\theta} \in \mathbb{R}^P$
- regressor matrix $\boldsymbol{Y}(\boldsymbol{q}, \dot{\boldsymbol{q}}, \ddot{\boldsymbol{q}}) \in \mathbb{R}^{N \times P}$
- $P \leq 10N$ (masses, COG coordinates, MOI tensors)

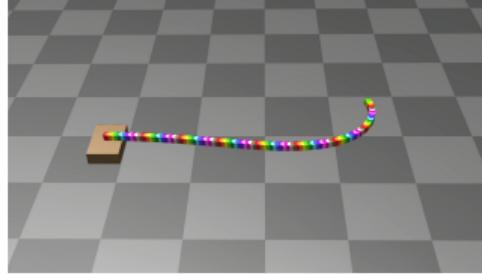


Creative Commons

Assumptions: rigid links, known inertia distribution, etc.

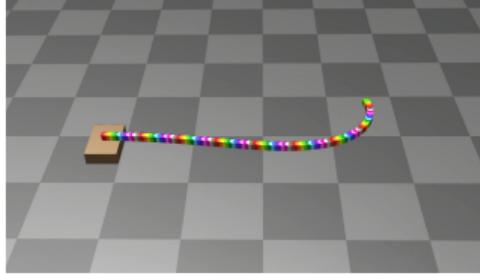
Motivation

Soft robots



Motivation

Soft robots

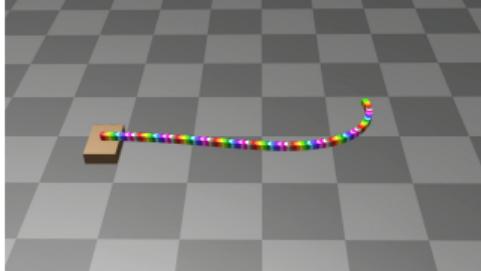


Robotic mechanical work



Motivation

Soft robots



Robotic mechanical work

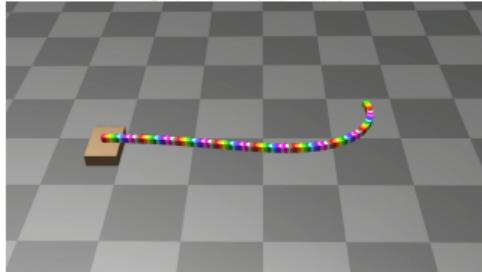


Robotic rehabilitation



Motivation

Soft robots



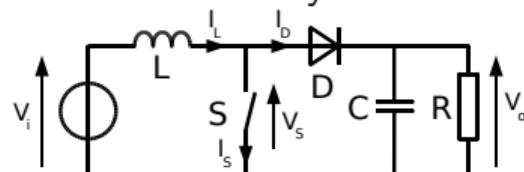
Robotic mechanical work



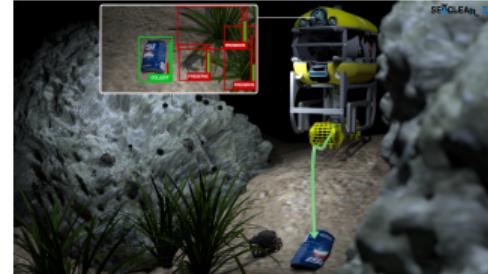
Robotic rehabilitation



Electrical systems



Underwater robots



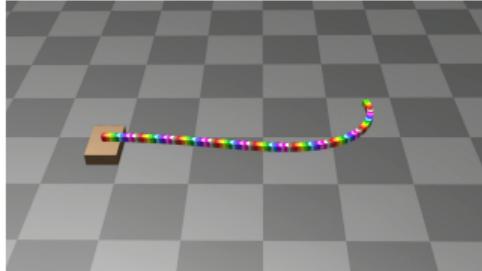
Aerodynamics and -elastics



Parametric modeling time-consuming, complex, unsatisfactory or even infeasible

Motivation

Soft robots



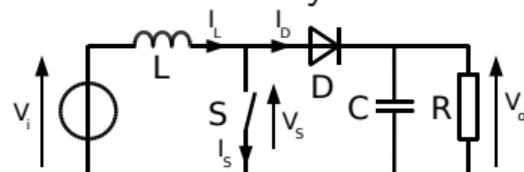
Robotic mechanical work



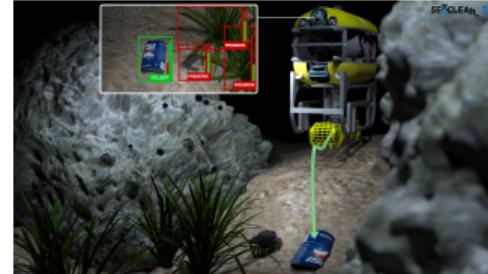
Robotic rehabilitation



Electrical systems



Underwater robots



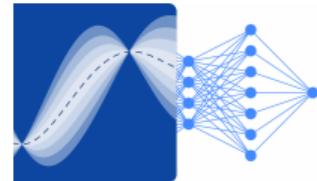
Aerodynamics and -elastics



Parametric modeling time-consuming, complex, unsatisfactory or even infeasible

Related Work

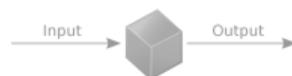
- Deep Lagrangian Networks [Lutter+ 2019], Hamiltonian Neural Networks [Greydanus+ 2019]: ~~uncertainty quantification~~



- Dependencies on nonmeasurable momentum [Greydanus+ 2019; Rath+ 2021] or exactly known inertia matrix [Rath+ 2022; Geist+ 2020]



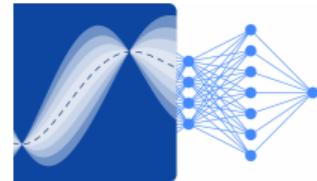
- Missing guarantees or physical structure of other Lagrangian modeling approaches [Cheng+ 2016; Ober-Blöbaum+ 2023]



- Rigidity of parametric approaches vs. too much flexibility (physical inconsistencies) of black-box-model learning

Related Work

- Deep Lagrangian Networks [Lutter+ 2019], Hamiltonian Neural Networks [Greydanus+ 2019]: ~~uncertainty quantification~~



- Dependencies on nonmeasurable momentum [Greydanus+ 2019; Rath+ 2021] or exactly known inertia matrix [Rath+ 2022; Geist+ 2020]

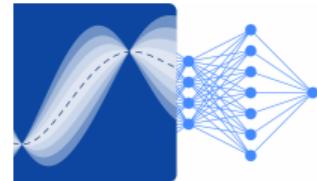
- Missing guarantees or physical structure of other Lagrangian modeling approaches [Cheng+ 2016; Ober-Blöbaum+ 2023]



- Rigidity of parametric approaches vs. too much flexibility (physical inconsistencies) of black-box-model learning

Related Work

- Deep Lagrangian Networks [Lutter+ 2019], Hamiltonian Neural Networks [Greydanus+ 2019]: ~~uncertainty quantification~~



- Dependencies on nonmeasurable momentum [Greydanus+ 2019; Rath+ 2021] or exactly known inertia matrix [Rath+ 2022; Geist+ 2020]



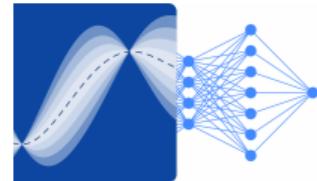
- Missing guarantees or physical structure of other Lagrangian modeling approaches [Cheng+ 2016; Ober-Blöbaum+ 2023]



- Rigidity of parametric approaches vs. too much flexibility (physical inconsistencies) of black-box-model learning

Related Work

- Deep Lagrangian Networks [Lutter+ 2019], Hamiltonian Neural Networks [Greydanus+ 2019]: ~~uncertainty quantification~~



- Dependencies on nonmeasurable momentum [Greydanus+ 2019; Rath+ 2021] or exactly known inertia matrix [Rath+ 2022; Geist+ 2020]
- Missing guarantees or physical structure of other Lagrangian modeling approaches [Cheng+ 2016; Ober-Blöbaum+ 2023]



- Rigidity of parametric approaches vs. too much flexibility (physical inconsistencies) of black-box-model learning

Lagrangian Mechanics

Conservative Euler-Lagrange (EL) Systems [Featherstone 2007; Murray+ 1994]

Equations of Motion:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = \tau$$

Equivalent Rigid Body Dynamics:

$$M(q)\ddot{q} + \underbrace{C(q, \dot{q})}_{=\frac{1}{2}(\frac{\partial^2 T}{\partial \dot{q} \partial q^\top} + \dot{M} - \frac{\partial^2 T}{\partial q \partial \dot{q}^\top})} \dot{q} + \underbrace{g(q)}_{=\frac{\partial V}{\partial q}} = \tau$$

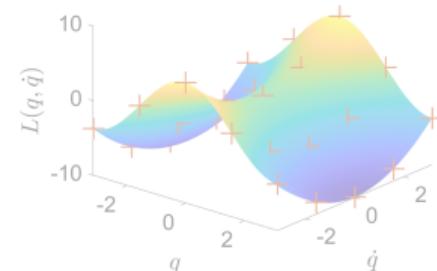
with (unknown) Lagrangian $L \equiv L(q, \dot{q}) = T(q, \dot{q}) - V(q) = \frac{1}{2}\dot{q}^\top M(q)\dot{q} - V(q)$

Objective:

Data-driven physically consistent approximation of L

$$\underline{L} = \frac{1}{2} \sum_i \sum_j \dot{q}_i \dot{q}_j m_{ij} - q_i q_j k_{ij} - g \sum_i m_i h_i \quad \Rightarrow$$

$$\underline{M} = \sum_i J_i^\top \mathcal{M}_i J_i$$



Lagrangian Mechanics

Conservative Euler-Lagrange (EL) Systems [Featherstone 2007; Murray+ 1994]

Equations of Motion:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = \tau$$

Equivalent Rigid Body Dynamics:

$$M(q)\ddot{q} + \underbrace{C(q, \dot{q})}_{= \frac{1}{2}(\frac{\partial^2 T}{\partial \dot{q} \partial q^\top} + \dot{M} - \frac{\partial^2 T}{\partial q \partial \dot{q}^\top})} \dot{q} + \underbrace{g(q)}_{= \frac{\partial V}{\partial q}} = \tau$$

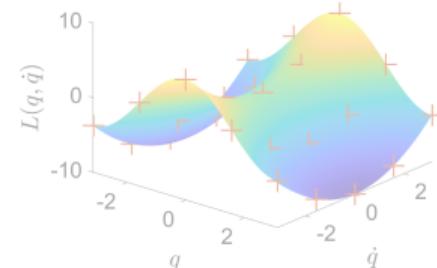
with (unknown) Lagrangian $L \equiv L(q, \dot{q}) = T(q, \dot{q}) - V(q) = \frac{1}{2}\dot{q}^\top M(q)\dot{q} - V(q)$

Objective:

Data-driven **physically consistent** approximation of L

$$\underline{L} = \frac{1}{2} \sum_i \sum_j \dot{q}_i \dot{q}_j m_{ij} - q_i q_j k_{ij} - g \sum_i m_i h_i \quad \Rightarrow$$

$$\underline{M} = \sum_i J_i^\top \mathcal{M}_i J_i$$



Lagrangian Mechanics

Conservative Euler-Lagrange (EL) Systems [Featherstone 2007; Murray+ 1994]

Equations of Motion:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = \tau$$

Equivalent Rigid Body Dynamics:

$$M(q)\ddot{q} + \underbrace{C(q, \dot{q})}_{=\frac{1}{2}(\frac{\partial^2 T}{\partial \dot{q} \partial q^\top} + \dot{M} - \frac{\partial^2 T}{\partial q \partial \dot{q}^\top})} \dot{q} + \underbrace{g(q)}_{=\frac{\partial V}{\partial q}} = \tau$$

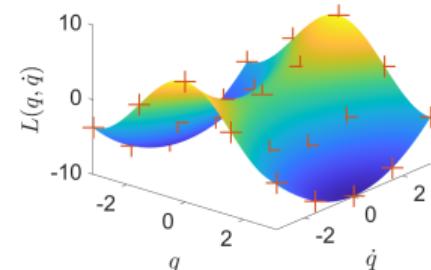
with (unknown) Lagrangian $L \equiv L(q, \dot{q}) = T(q, \dot{q}) - V(q) = \frac{1}{2}\dot{q}^\top M(q)\dot{q} - V(q)$

Objective:

Data-driven **physically consistent** approximation of L

$$\cancel{L = \frac{1}{2} \sum_i \sum_j \dot{q}_i \dot{q}_j m_{ij} - q_i q_j k_{ij} - g \sum_i m_i h_i} \quad \Rightarrow$$

$$\cancel{M = \sum_i J_i^\top \mathcal{M}_i J_i}$$



Lagrangian Mechanics

Dissipative Euler-Lagrange (EL) Systems [Featherstone 2007; Murray+ 1994]

Equations of Motion:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = \tau - \tau_f$$

Equivalent Rigid Body Dynamics:

$$M(\boldsymbol{q})\ddot{\boldsymbol{q}} + \underbrace{C(\boldsymbol{q}, \dot{\boldsymbol{q}})}_{=\frac{1}{2}\left(\frac{\partial^2 T}{\partial \dot{\boldsymbol{q}} \partial \dot{\boldsymbol{q}}^\top} + \dot{M} - \frac{\partial^2 T}{\partial \boldsymbol{q} \partial \dot{\boldsymbol{q}}^\top}\right)} \dot{\boldsymbol{q}} + \underbrace{\boldsymbol{g}(\boldsymbol{q})}_{=\frac{\partial V}{\partial \boldsymbol{q}}} = \tau - \underbrace{\boldsymbol{D}(\dot{\boldsymbol{q}})\dot{\boldsymbol{q}}}_{\text{dissipation}}$$

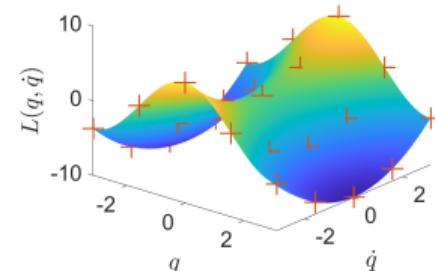
with (unknown) Lagrangian $L \equiv L(\boldsymbol{q}, \dot{\boldsymbol{q}}) = T(\boldsymbol{q}, \dot{\boldsymbol{q}}) - V(\boldsymbol{q}) = \frac{1}{2}\dot{\boldsymbol{q}}^\top M(\boldsymbol{q})\dot{\boldsymbol{q}} - V(\boldsymbol{q})$

Objective:

Data-driven **physically consistent** approximation of L

$$\cancel{L = \frac{1}{2} \sum_i \sum_j \dot{q}_i \dot{q}_j m_{ij} - q_i q_j k_{ij} - g \sum_i m_i h_i} \Rightarrow$$

$$\cancel{\boldsymbol{M} = \sum_i \boldsymbol{J}_i^\top \boldsymbol{\mathcal{M}}_i \boldsymbol{J}_i}$$



Physically Consistent Modeling

Definition: Physical Consistency

A model is **physically consistent** if constrained to respect the physical laws relevant to its modeling level.



- Euler-Lagrange systems: Principle of Least Action, conservative forces $\mathbf{g} = \frac{\partial V}{\partial \dot{\mathbf{q}}}$
- Rigid body systems: parametric pseudo-inertias $\begin{bmatrix} \Sigma_i & \mathbf{h}_i \\ \mathbf{h}_i & m_i \end{bmatrix} \succ 0$
- Classification into deterministic/analytical and probabilistic, or full and semi-consistency, possible

Physically Consistent Modeling

Definition: Physical Consistency

A model is **physically consistent** if constrained to respect the physical laws relevant to its modeling level.



- Euler-Lagrange systems: Principle of Least Action, conservative forces $\mathbf{g} = \frac{\partial V}{\partial \dot{\mathbf{q}}}$
- Rigid body systems: parametric pseudo-inertias $\begin{bmatrix} \Sigma_i & h_i \\ h_i & m_i \end{bmatrix} \succ 0$
- Classification into deterministic/analytical and probabilistic, or full and semi-consistency, possible

Physically Consistent Modeling

Definition: Physical Consistency

A model is **physically consistent** if constrained to respect the physical laws relevant to its modeling level.



- Euler-Lagrange systems: Principle of Least Action, conservative forces $\mathbf{g} = \frac{\partial V}{\partial \dot{\mathbf{q}}}$
- Rigid body systems: parametric pseudo-inertias $\begin{bmatrix} \Sigma_i & \mathbf{h}_i \\ \mathbf{h}_i & m_i \end{bmatrix} \succ 0$
- Classification into deterministic/analytical and probabilistic, or full and semi-consistency, possible

Physically Consistent Modeling

Definition: Physical Consistency

A model is **physically consistent** if constrained to respect the physical laws relevant to its modeling level.



- Euler-Lagrange systems: Principle of Least Action, conservative forces $\mathbf{g} = \frac{\partial V}{\partial \dot{\mathbf{q}}}$
- Rigid body systems: parametric pseudo-inertias $\begin{bmatrix} \Sigma_i & \mathbf{h}_i \\ \mathbf{h}_i & m_i \end{bmatrix} \succ 0$
- Classification into deterministic/analytical and probabilistic, or full and semi-consistency, possible

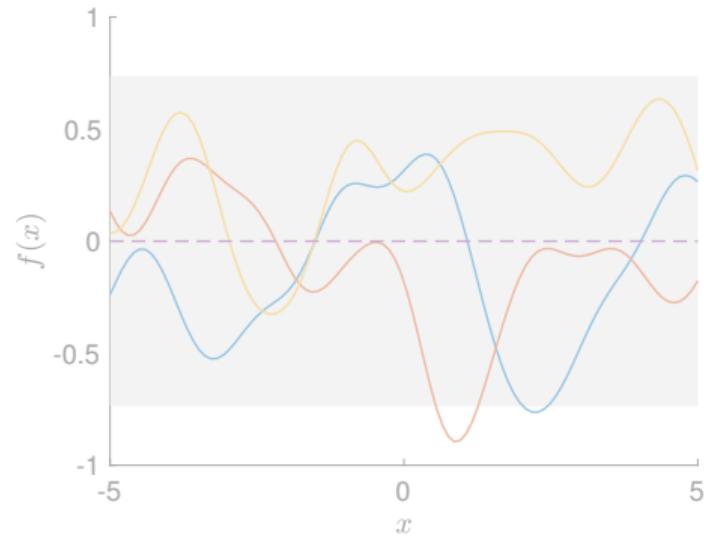
Background: Gaussian Processes

Gaussian Process (GP) [Rasmussen+ 2006]

For \mathbf{x}, \mathbf{x}' in $\mathcal{X} \subset \mathbb{R}^M$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

with mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$



Joint Gaussian distribution:

$$\begin{bmatrix} f(\mathbf{x}) \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}) \\ \mathbf{m} \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}, \mathbf{x}) & \mathbf{k}^\top(\mathbf{x}) \\ \mathbf{k}(\mathbf{x}) & \mathbf{K} + \Sigma_\epsilon \end{bmatrix}\right)$$

with noisy observations $y_i = f(x_i) + \varepsilon_i$, zero-mean white noise $\varepsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_\epsilon)$

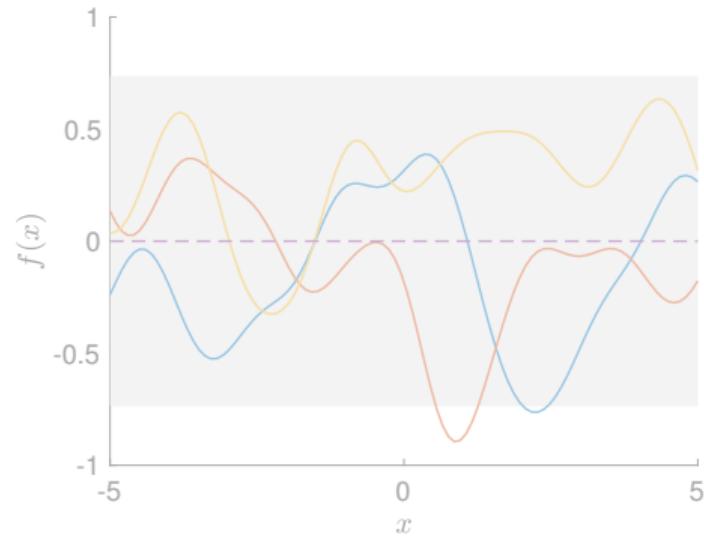
Background: Gaussian Processes

Gaussian Process (GP) [Rasmussen+ 2006]

For \mathbf{x}, \mathbf{x}' in $\mathcal{X} \subset \mathbb{R}^M$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

with mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$



Joint Gaussian distribution:

$$\begin{bmatrix} f(\mathbf{x}) \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}) \\ \mathbf{m} \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}, \mathbf{x}) & \mathbf{k}^\top(\mathbf{x}) \\ \mathbf{k}(\mathbf{x}) & \mathbf{K} + \Sigma_\epsilon \end{bmatrix}\right)$$

with noisy observations $y_i = f(\mathbf{x}_i) + \varepsilon_i$, zero-mean white noise $\varepsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_\epsilon)$

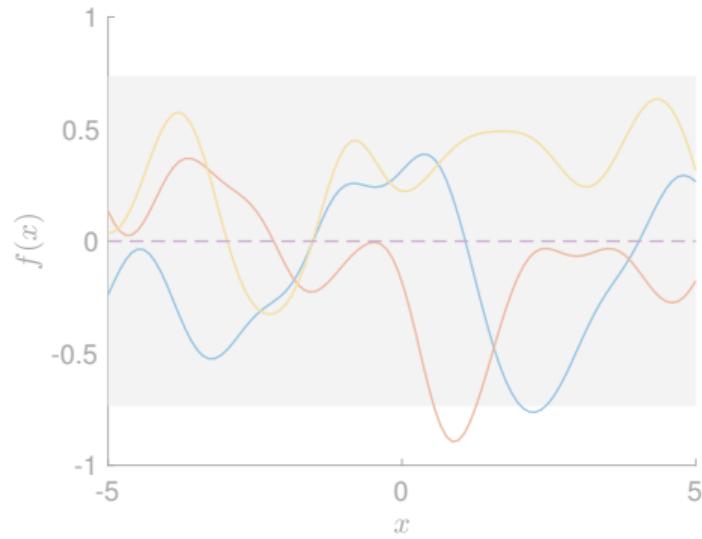
Background: Gaussian Processes

Gaussian Process (GP) [Rasmussen+ 2006]

For \mathbf{x}, \mathbf{x}' in $\mathcal{X} \subset \mathbb{R}^M$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

with mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$



Regression

Conditioning leads to posterior $f(\mathbf{x})|\mathbf{y} \sim \mathcal{N}(\mu_f(\mathbf{x}), \sigma_f(\mathbf{x}))$ with

$$\mu_f(\mathbf{x}) \equiv \mathbb{E}[f(\mathbf{x})|\mathbf{y}] = m(\mathbf{x}) + \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}(\mathbf{y} - \mathbf{m})$$

$$\sigma_f^2(\mathbf{x}) \equiv \text{Var}[f(\mathbf{x})|\mathbf{y}] = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{k}(\mathbf{x})$$

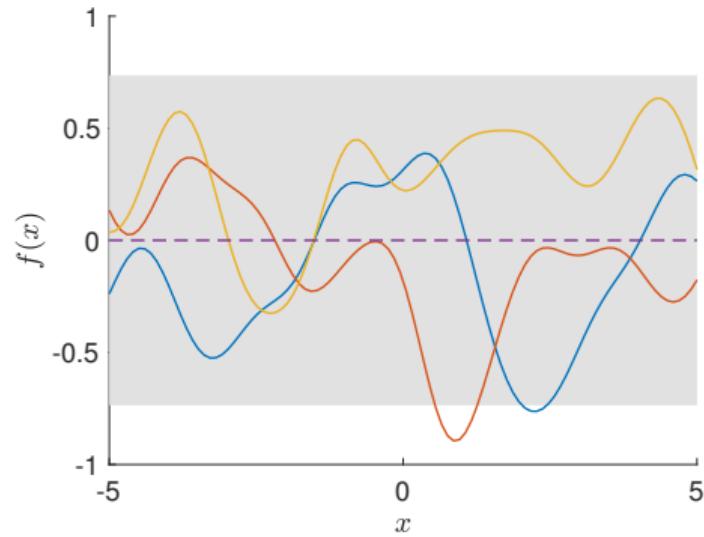
Background: Gaussian Processes

Gaussian Process (GP) [Rasmussen+ 2006]

For \mathbf{x}, \mathbf{x}' in $\mathcal{X} \subset \mathbb{R}^M$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

with mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$



Regression

Conditioning leads to posterior $f(\mathbf{x})|\mathbf{y} \sim \mathcal{N}(\mu_f(\mathbf{x}), \sigma_f(\mathbf{x}))$ with

$$\mu_f(\mathbf{x}) \equiv \text{E}[f(\mathbf{x})|\mathbf{y}] = m(\mathbf{x}) + \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}(\mathbf{y} - \mathbf{m})$$

$$\sigma_f^2(\mathbf{x}) \equiv \text{Var}[f(\mathbf{x})|\mathbf{y}] = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{k}(\mathbf{x})$$

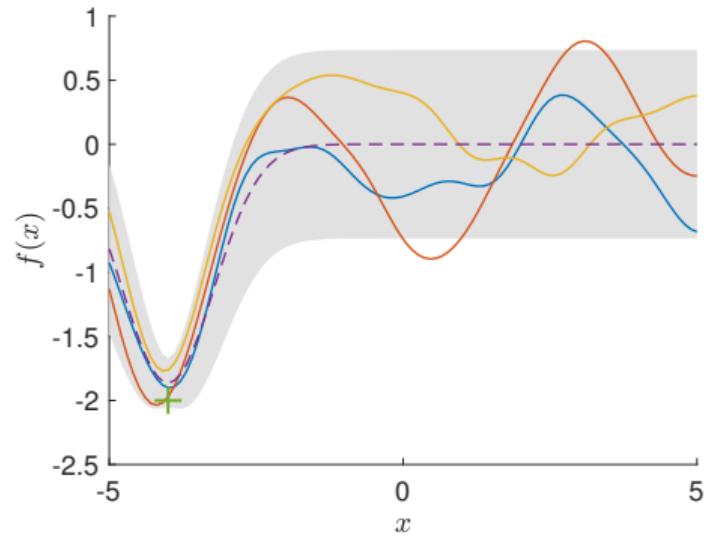
Background: Gaussian Processes

Gaussian Process (GP) [Rasmussen+ 2006]

For \mathbf{x}, \mathbf{x}' in $\mathcal{X} \subset \mathbb{R}^M$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

with mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$



Regression

Conditioning leads to posterior $f(\mathbf{x})|\mathbf{y} \sim \mathcal{N}(\mu_f(\mathbf{x}), \sigma_f(\mathbf{x}))$ with

$$\mu_f(\mathbf{x}) \equiv \text{E}[f(\mathbf{x})|\mathbf{y}] = m(\mathbf{x}) + \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}(\mathbf{y} - \mathbf{m})$$

$$\sigma_f^2(\mathbf{x}) \equiv \text{Var}[f(\mathbf{x})|\mathbf{y}] = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{k}(\mathbf{x})$$

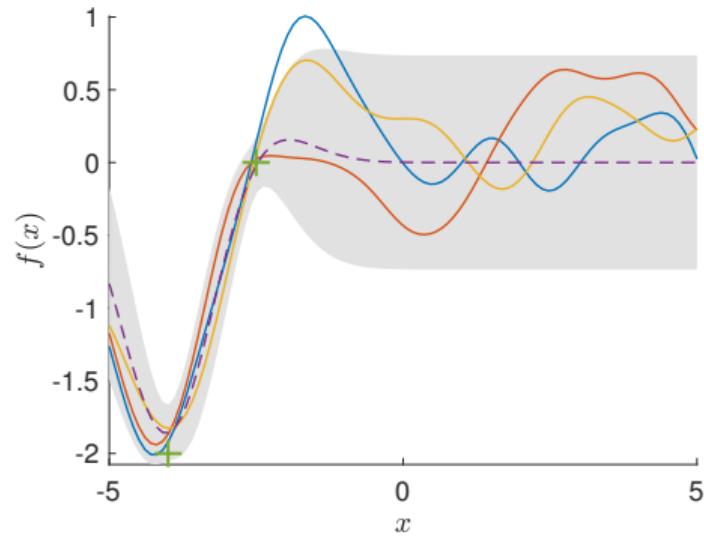
Background: Gaussian Processes

Gaussian Process (GP) [Rasmussen+ 2006]

For \mathbf{x}, \mathbf{x}' in $\mathcal{X} \subset \mathbb{R}^M$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

with mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$



Regression

Conditioning leads to posterior $f(\mathbf{x})|\mathbf{y} \sim \mathcal{N}(\mu_f(\mathbf{x}), \sigma_f(\mathbf{x}))$ with

$$\mu_f(\mathbf{x}) \equiv \text{E}[f(\mathbf{x})|\mathbf{y}] = m(\mathbf{x}) + \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}(\mathbf{y} - \mathbf{m})$$

$$\sigma_f^2(\mathbf{x}) \equiv \text{Var}[f(\mathbf{x})|\mathbf{y}] = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{k}(\mathbf{x})$$

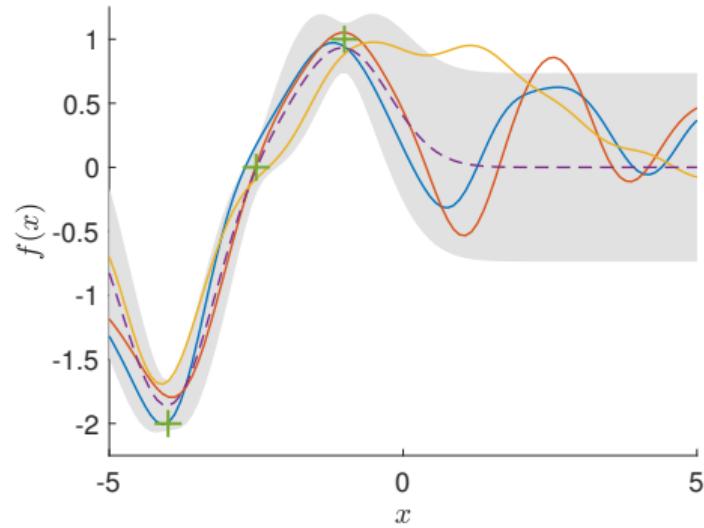
Background: Gaussian Processes

Gaussian Process (GP) [Rasmussen+ 2006]

For \mathbf{x}, \mathbf{x}' in $\mathcal{X} \subset \mathbb{R}^M$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

with mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$



Regression

Conditioning leads to posterior $f(\mathbf{x})|\mathbf{y} \sim \mathcal{N}(\mu_f(\mathbf{x}), \sigma_f(\mathbf{x}))$ with

$$\mu_f(\mathbf{x}) \equiv \text{E}[f(\mathbf{x})|\mathbf{y}] = m(\mathbf{x}) + \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}(\mathbf{y} - \mathbf{m})$$

$$\sigma_f^2(\mathbf{x}) \equiv \text{Var}[f(\mathbf{x})|\mathbf{y}] = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{k}(\mathbf{x})$$

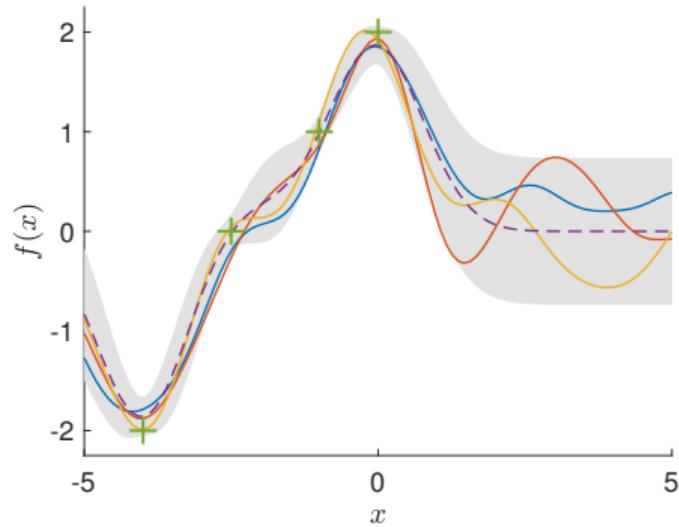
Background: Gaussian Processes

Gaussian Process (GP) [Rasmussen+ 2006]

For \mathbf{x}, \mathbf{x}' in $\mathcal{X} \subset \mathbb{R}^M$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

with mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$



Regression

Conditioning leads to posterior $f(\mathbf{x})|\mathbf{y} \sim \mathcal{N}(\mu_f(\mathbf{x}), \sigma_f(\mathbf{x}))$ with

$$\mu_f(\mathbf{x}) \equiv \text{E}[f(\mathbf{x})|\mathbf{y}] = m(\mathbf{x}) + \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}(\mathbf{y} - \mathbf{m})$$

$$\sigma_f^2(\mathbf{x}) \equiv \text{Var}[f(\mathbf{x})|\mathbf{y}] = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{k}(\mathbf{x})$$

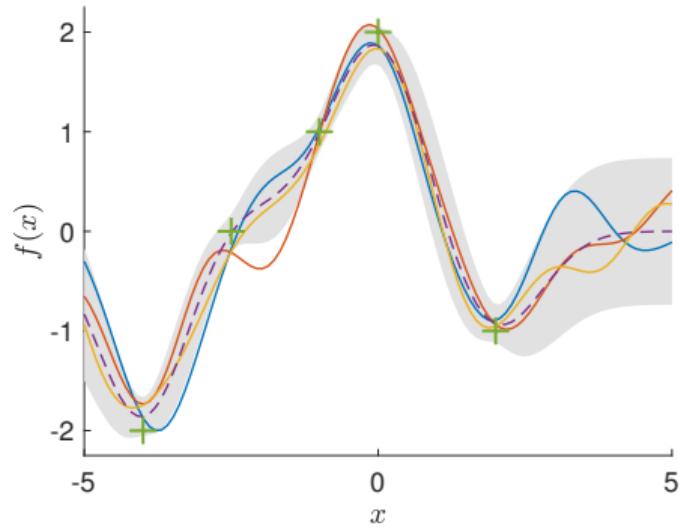
Background: Gaussian Processes

Gaussian Process (GP) [Rasmussen+ 2006]

For \mathbf{x}, \mathbf{x}' in $\mathcal{X} \subset \mathbb{R}^M$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

with mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$



Regression

Conditioning leads to posterior $f(\mathbf{x})|\mathbf{y} \sim \mathcal{N}(\mu_f(\mathbf{x}), \sigma_f(\mathbf{x}))$ with

$$\mu_f(\mathbf{x}) \equiv \text{E}[f(\mathbf{x})|\mathbf{y}] = m(\mathbf{x}) + \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}(\mathbf{y} - \mathbf{m})$$

$$\sigma_f^2(\mathbf{x}) \equiv \text{Var}[f(\mathbf{x})|\mathbf{y}] = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x})(\mathbf{K} + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{k}(\mathbf{x})$$

Background: Multidimensional GPs

Multidimensional GP [Álvarez+ 2012]

For x, x' in $\mathcal{X} \subset \mathbb{R}^M$, vector-valued $f: \mathbb{R}^M \rightarrow \mathbb{R}^N$:

$$f(x) \sim \mathcal{GP}(\mathbf{m}(x), \mathbf{K}(x, x'))$$

with mean $\mathbf{m}(x) \in \mathbb{R}^N$ and kernel $\mathbf{0} \preceq \mathbf{K}(x, x') \in \mathbb{R}^{N \times N}$

Joint Distribution:

For noisy observations $y_i = f(x_i) + \epsilon_i$, white noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_\epsilon)$:

$$\begin{bmatrix} f(x) \\ \text{vec}(\mathbf{Y}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(x) \\ \text{vec}(\mathbf{m}(\mathbf{X})) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(x, x) & \mathbf{K}(x, \mathbf{X}) \\ \mathbf{K}(\mathbf{X}, x) & \mathbf{K}(\mathbf{X}, \mathbf{X}) + \Sigma_\epsilon \end{bmatrix}\right)$$

with D observations $\mathbf{Y} \in \mathbb{R}^{N \times D}$ at locations $\mathbf{X} \in \mathbb{R}^{M \times D}$

Background: Multidimensional GPs

Multidimensional GP [Álvarez+ 2012]

For x, x' in $\mathcal{X} \subset \mathbb{R}^M$, vector-valued $f: \mathbb{R}^M \rightarrow \mathbb{R}^N$:

$$f(x) \sim \mathcal{GP}(\mathbf{m}(x), \mathbf{K}(x, x'))$$

with mean $\mathbf{m}(x) \in \mathbb{R}^N$ and kernel $\mathbf{0} \preceq \mathbf{K}(x, x') \in \mathbb{R}^{N \times N}$

Joint Distribution:

For noisy observations $y_i = f(x_i) + \epsilon_i$, white noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_\epsilon)$:

$$\begin{bmatrix} f(x) \\ \text{vec}(\mathbf{Y}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(x) \\ \text{vec}(\mathbf{m}(\mathbf{X})) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(x, x) & \mathbf{K}(x, \mathbf{X}) \\ \mathbf{K}(\mathbf{X}, x) & \mathbf{K}(\mathbf{X}, \mathbf{X}) + \Sigma_\epsilon \end{bmatrix}\right)$$

with D observations $\mathbf{Y} \in \mathbb{R}^{N \times D}$ at locations $\mathbf{X} \in \mathbb{R}^{M \times D}$

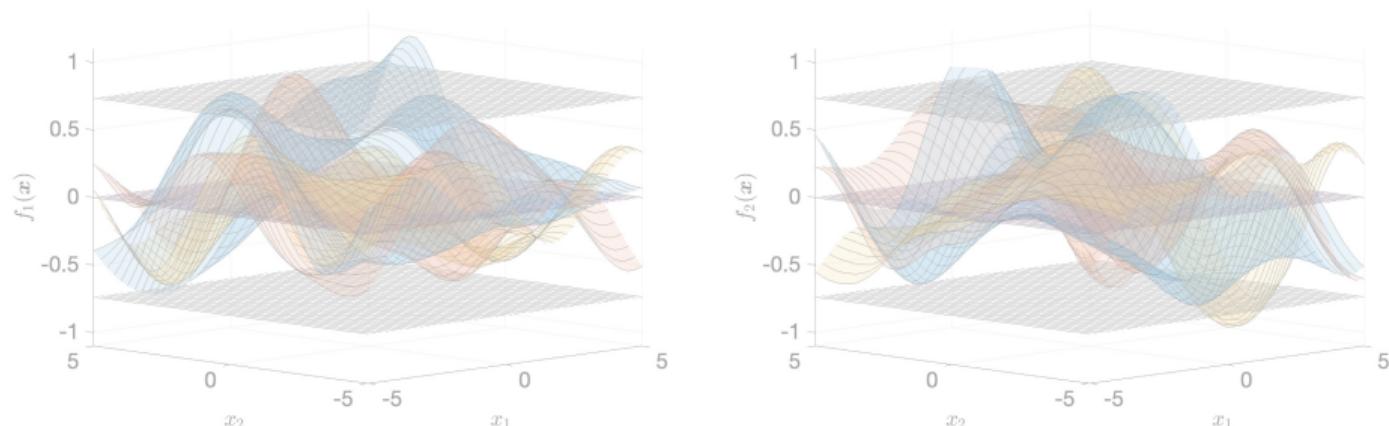
Background: Multidimensional GPs

Regression

Conditioning leads to multidimensional $\mathbf{f}(\mathbf{x}) | \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}))$ with

$$\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \text{vec}(\mathbf{Y} - \mathbf{m}(\mathbf{X}))$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}) = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x})$$



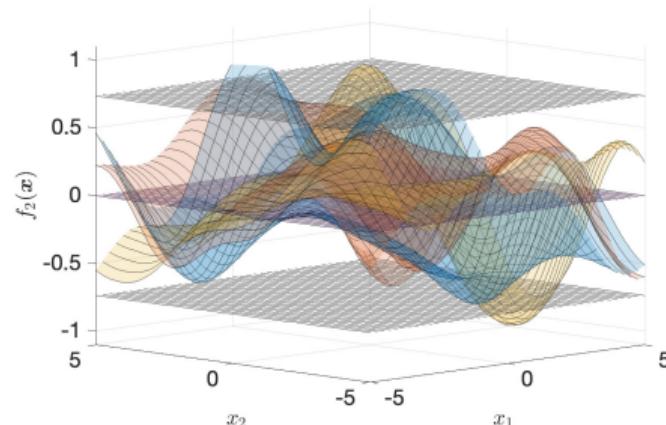
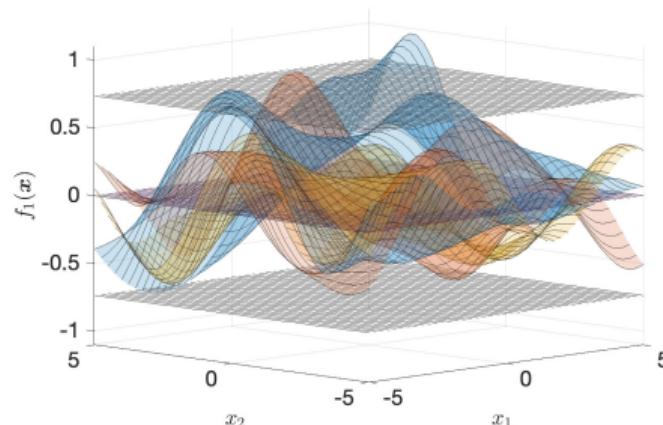
Background: Multidimensional GPs

Regression

Conditioning leads to multidimensional $\mathbf{f}(\mathbf{x}) | \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}))$ with

$$\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \text{vec}(\mathbf{Y} - \mathbf{m}(\mathbf{X}))$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}) = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x})$$



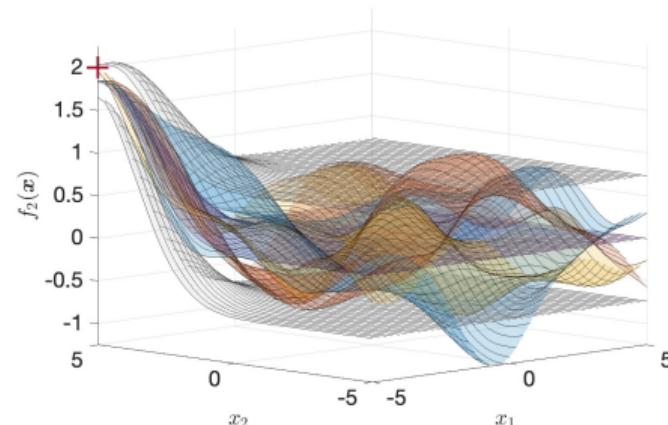
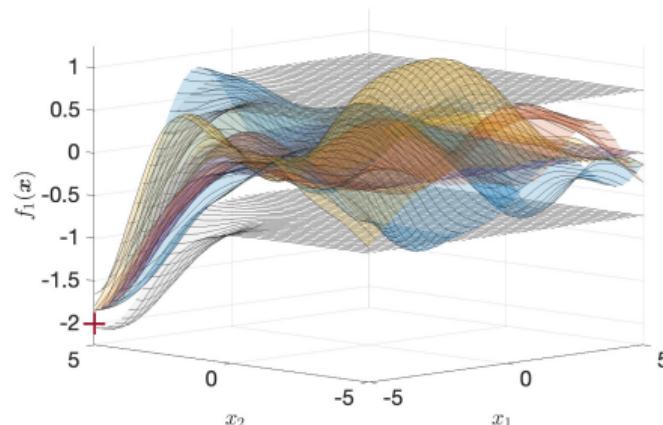
Background: Multidimensional GPs

Regression

Conditioning leads to multidimensional $\mathbf{f}(\mathbf{x}) | \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}))$ with

$$\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \text{vec}(\mathbf{Y} - \mathbf{m}(\mathbf{X}))$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}) = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x})$$



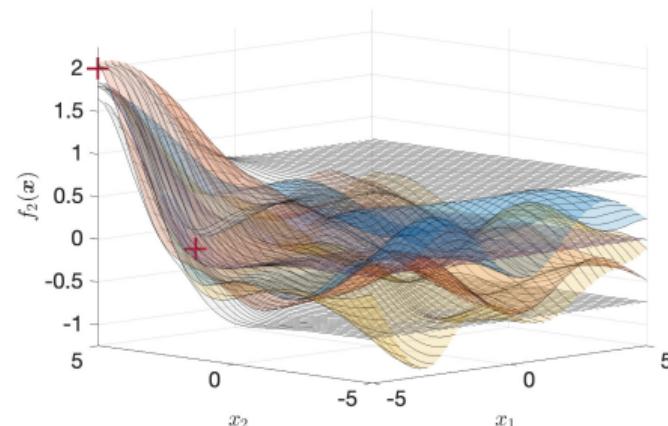
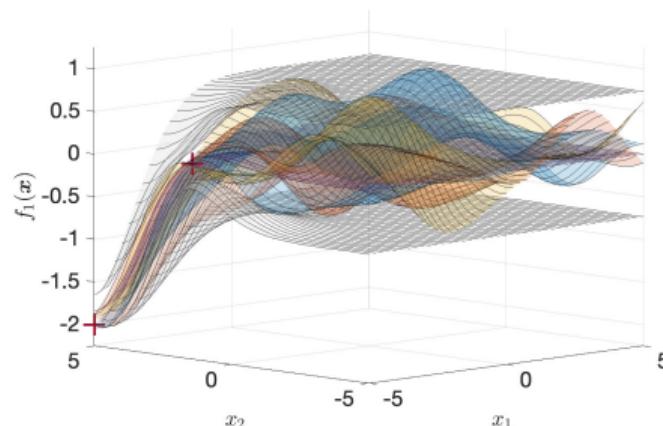
Background: Multidimensional GPs

Regression

Conditioning leads to multidimensional $\mathbf{f}(\mathbf{x}) | \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}))$ with

$$\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\epsilon})^{-1} \text{vec}(\mathbf{Y} - \mathbf{m}(\mathbf{X}))$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}) = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\epsilon})^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x})$$



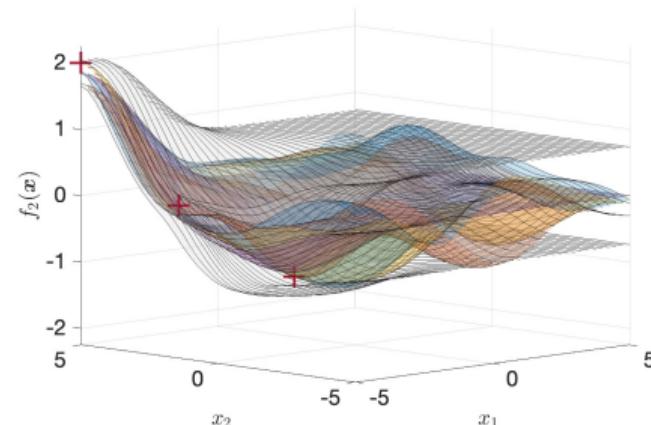
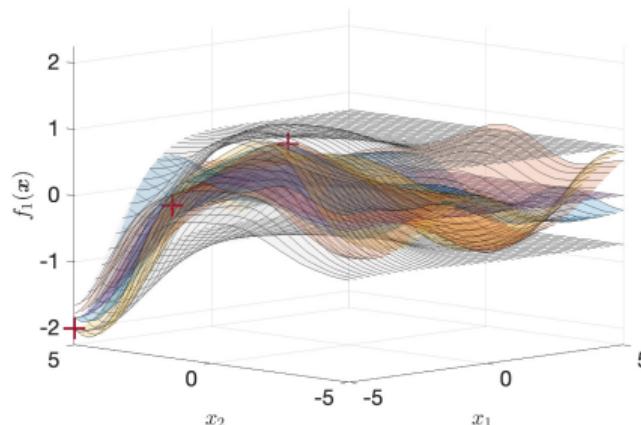
Background: Multidimensional GPs

Regression

Conditioning leads to multidimensional $\mathbf{f}(\mathbf{x}) | \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}))$ with

$$\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \text{vec}(\mathbf{Y} - \mathbf{m}(\mathbf{X}))$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}) = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x})$$



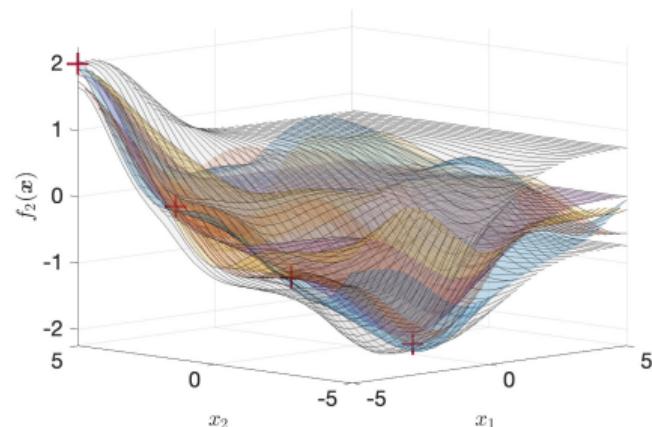
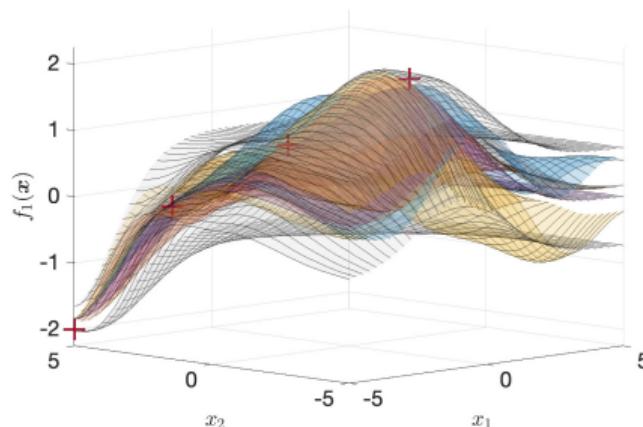
Background: Multidimensional GPs

Regression

Conditioning leads to multidimensional $\mathbf{f}(\mathbf{x}) | \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}))$ with

$$\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \text{vec}(\mathbf{Y} - \mathbf{m}(\mathbf{X}))$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}) = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x})$$



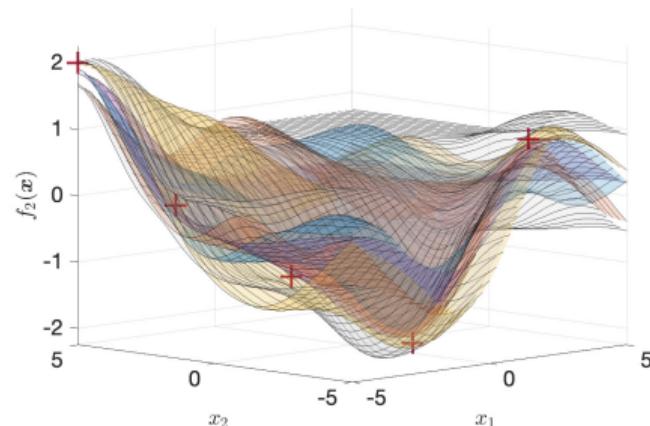
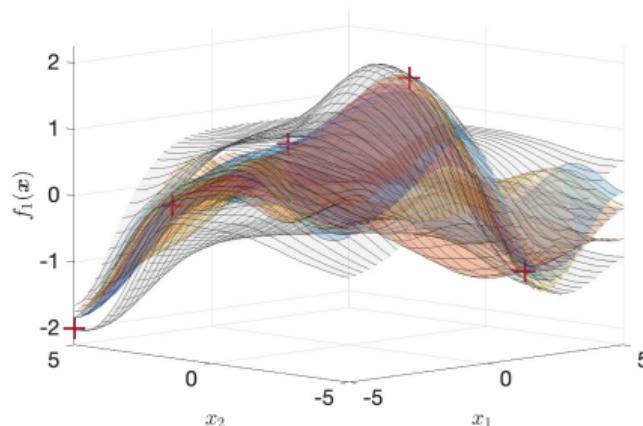
Background: Multidimensional GPs

Regression

Conditioning leads to multidimensional $\mathbf{f}(\mathbf{x}) | \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}))$ with

$$\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \text{vec}(\mathbf{Y} - \mathbf{m}(\mathbf{X}))$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}) = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x})$$



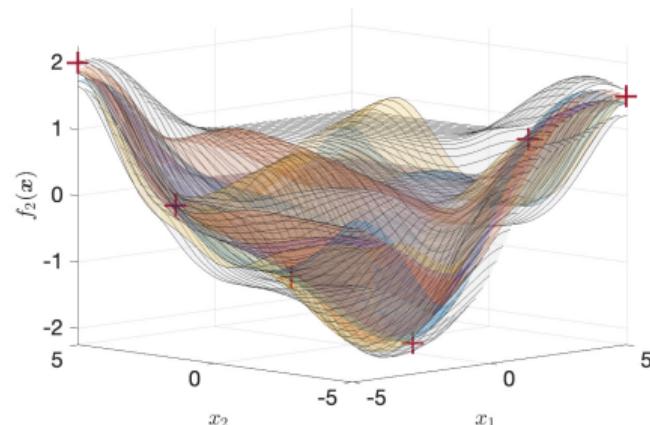
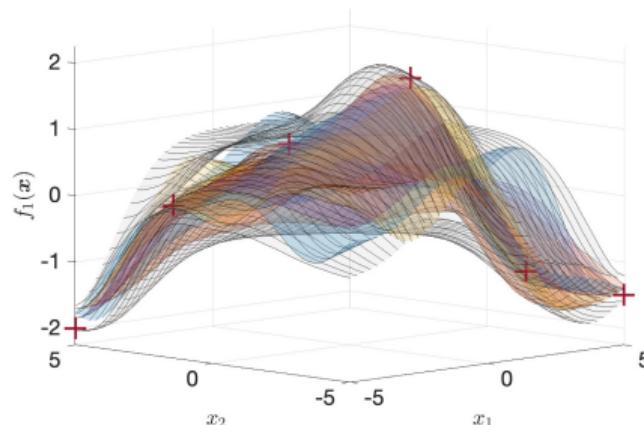
Background: Multidimensional GPs

Regression

Conditioning leads to multidimensional $\mathbf{f}(\mathbf{x}) | \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}))$ with

$$\boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) + \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\epsilon})^{-1} \text{vec}(\mathbf{Y} - \mathbf{m}(\mathbf{X}))$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{x}) = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \boldsymbol{\Sigma}_{\epsilon})^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x})$$

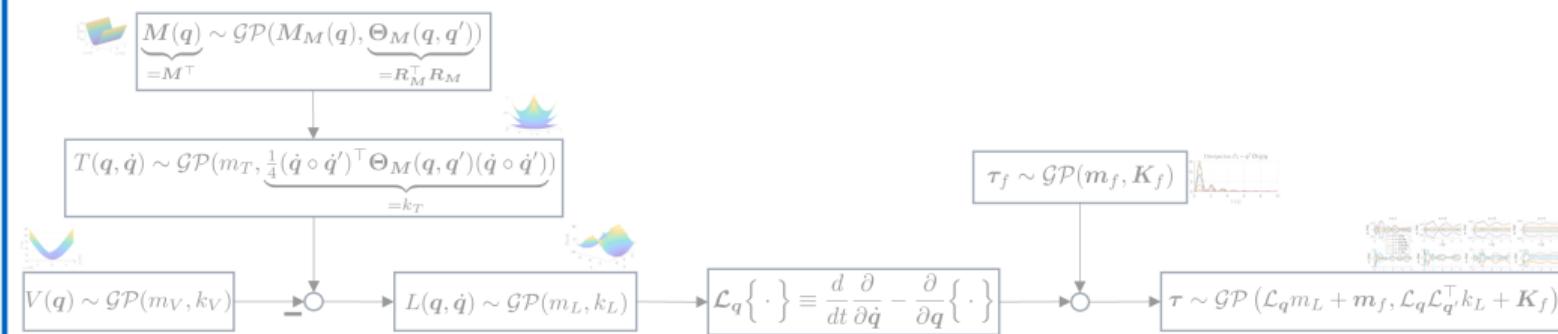


Lagrangian-Gaussian Process (L-GP) Model

Modeling Framework [Evangelisti+ 2022; Evangelisti+ 2024]

Concept: **data-driven** yet **physically consistent** model

→ enforce & leverage Lagrangian Mechanics



Giulio Evangelisti and Sandra Hirche. **Data-Driven Momentum Observers With Physically Consistent Gaussian Processes**.

In: IEEE Transactions on Robotics 40 (2024), pp. 1938–1951. DOI: 10.1109/TRO.2024.3366818.



Giulio Evangelisti and Sandra Hirche. **Physically Consistent Learning of Conservative Lagrangian Systems with Gaussian Processes**.

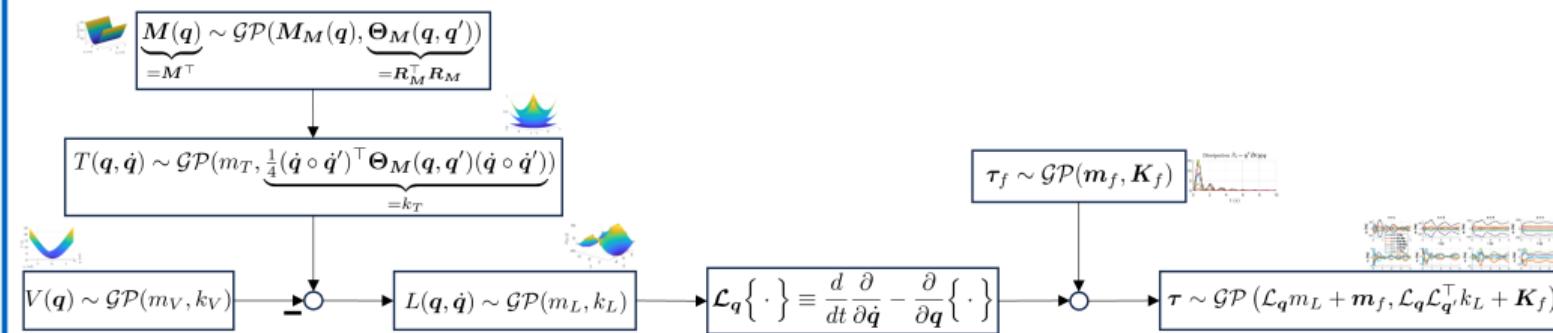
In: 2022 IEEE 61st Conference on Decision and Control (CDC). 2022, pp. 4078–4085. DOI: 10.1109/CDC51059.2022.9993123.

Lagrangian-Gaussian Process (L-GP) Model

Modeling Framework [Evangelisti+ 2022; Evangelisti+ 2024]

Concept: **data-driven** yet **physically consistent** model

→ enforce & leverage Lagrangian Mechanics

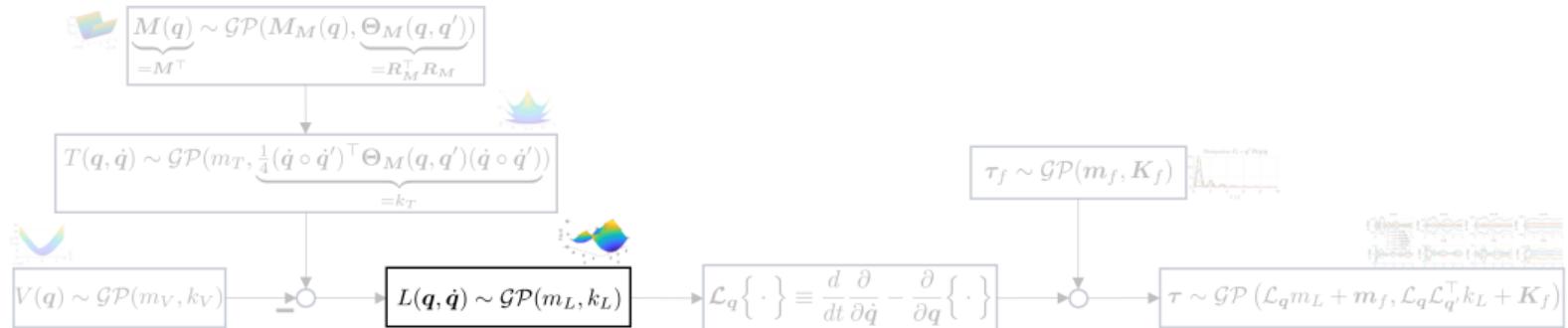


Giulio Evangelisti and Sandra Hirche. **Data-Driven Momentum Observers With Physically Consistent Gaussian Processes**.
In: IEEE Transactions on Robotics 40 (2024), pp. 1938–1951. DOI: 10.1109/TRO.2024.3366818.



Giulio Evangelisti and Sandra Hirche. **Physically Consistent Learning of Conservative Lagrangian Systems with Gaussian Processes**.
In: 2022 IEEE 61st Conference on Decision and Control (CDC). 2022, pp. 4078–4085. DOI: 10.1109/CDC51059.2022.9993123.

L-GP Modeling Framework



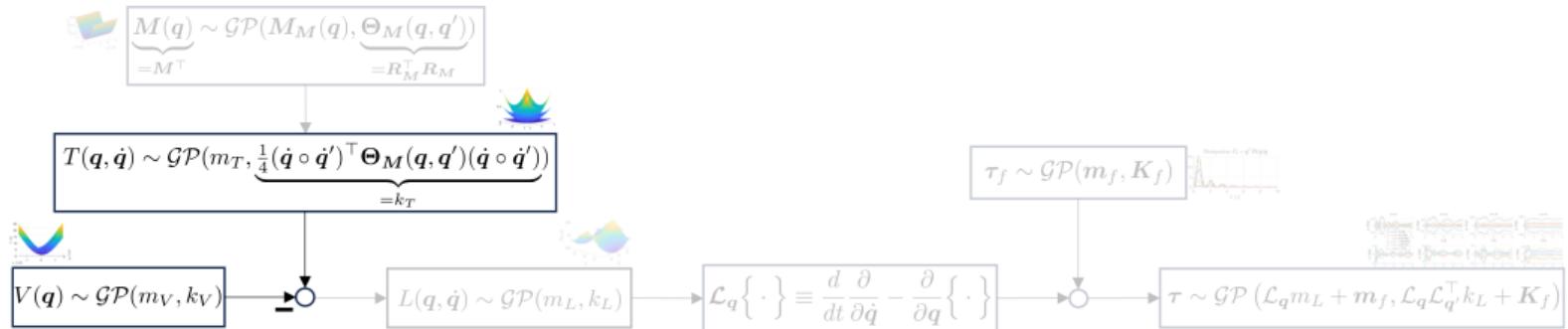
Energy structuring

Composite model $L = T - V$, where

$$L(\mathbf{q}, \dot{\mathbf{q}}) \sim GP(m_L(\mathbf{q}, \dot{\mathbf{q}}), k_L(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{q}', \dot{\mathbf{q}}'))$$

with $m_L = m_T - m_V$, $k_L = k_T + k_V$

L-GP Modeling Framework



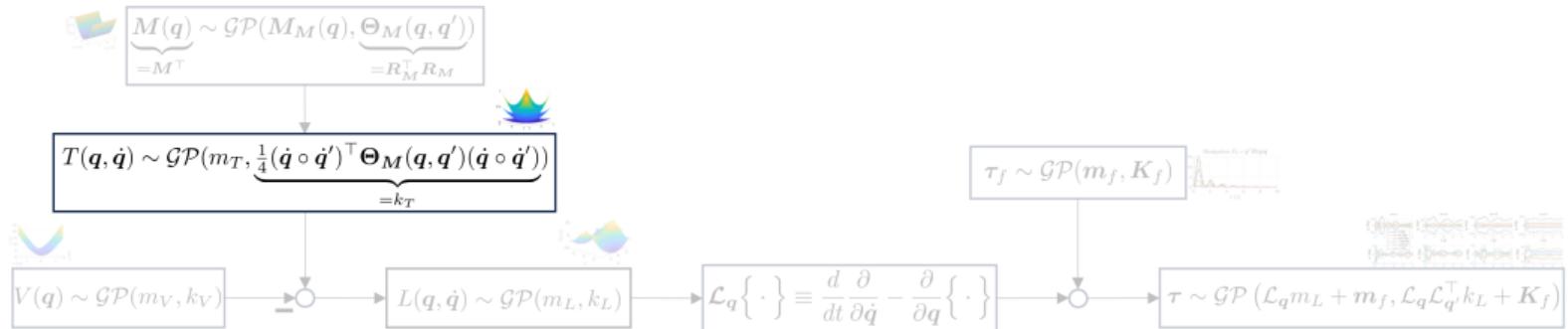
Energy structuring

Composite model $L = T - V$, where

$$L(q, \dot{q}) \sim GP(m_L(q, \dot{q}), k_L(q, \dot{q}, q', \dot{q}'))$$

with $m_L = m_T - m_V$, $k_L = k_T + k_V$

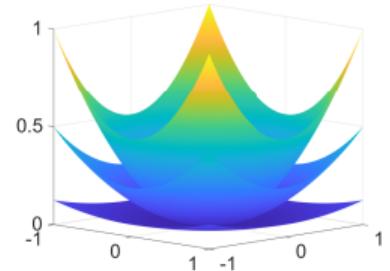
L-GP Modeling Framework



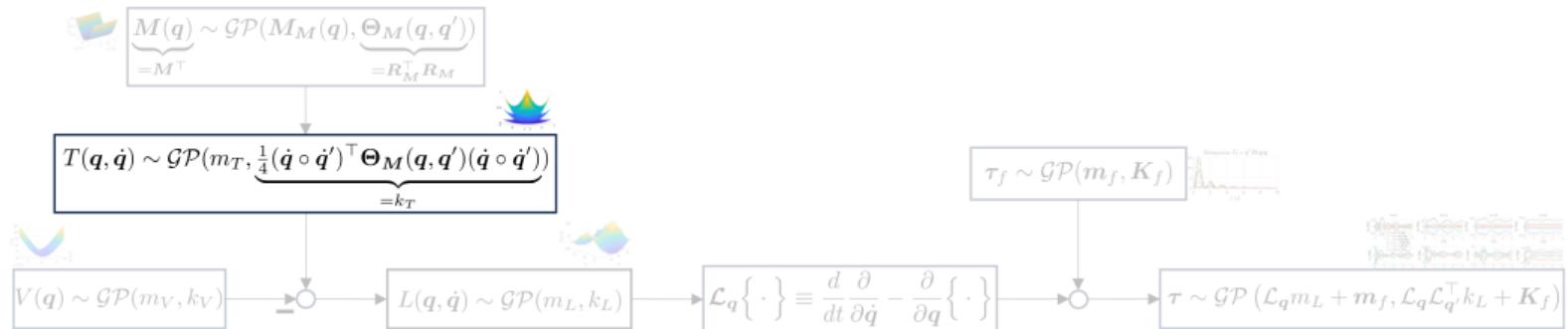
Differential-quadratic forms for $k \in \{0, 1\}$ s.t. $f_k \in \{U, T\}$:

$$f_k(\mathbf{q}) := \frac{1}{2} \mathbf{q}^{(k)\top} \mathbf{H}_k(\mathbf{q}) \mathbf{q}^{(k)}$$

with additional elastic U and gravitational energy G such that $V = G + U = G + \frac{1}{2} \mathbf{q}^\top \mathbf{S} \mathbf{q}$ and $T = \frac{1}{2} \dot{\mathbf{q}}^\top \mathbf{M} \dot{\mathbf{q}}$



L-GP Modeling Framework



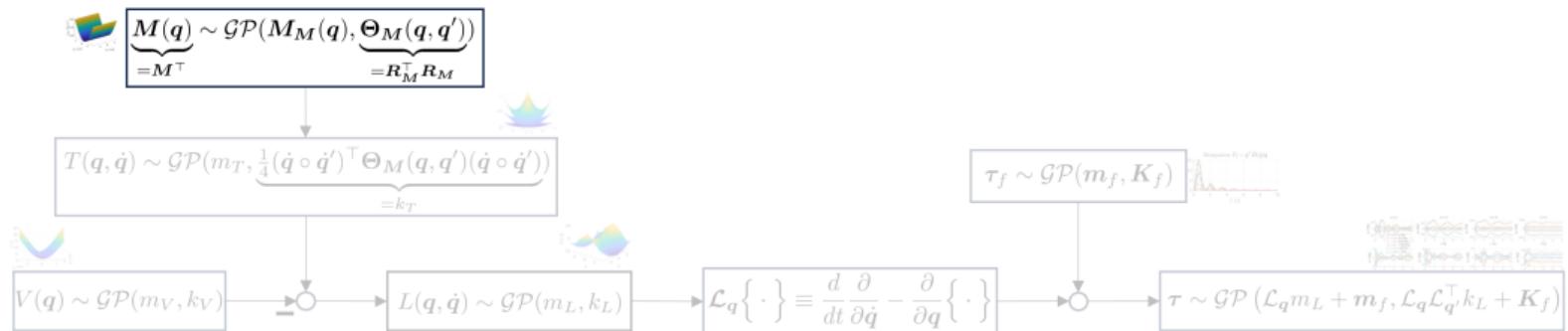
Specification of function space instead of black-box RKHS approach

Cholesky decomposed matrix kernels: $\Theta_k \equiv \mathbf{R}_k^\top(\mathbf{q}, \mathbf{q}') \mathbf{R}_k(\mathbf{q}, \mathbf{q}')$ with

$$\kappa_k(\mathbf{q}, \mathbf{q}') = \frac{1}{4} \left({}^{(k)}\mathbf{q} \circ {}^{(k)}\mathbf{q}' \right)^\top \Theta_k(\mathbf{q}, \mathbf{q}') \left({}^{(k)}\mathbf{q} \circ {}^{(k)}\mathbf{q}' \right)$$

- $\kappa_0 = k_U$ for $k = 0$ and $\kappa_1 = k_T$ for $k = 1$
- deterministic preservation of kinetic and elastic energies' quadratic form
- stochastical consistency: all energies $\sim \mathcal{GP}$, in contrast: $\mathbf{R}_M \sim \mathcal{GP} \Rightarrow \mathbf{M} \sim \chi^2$

L-GP Modeling Framework



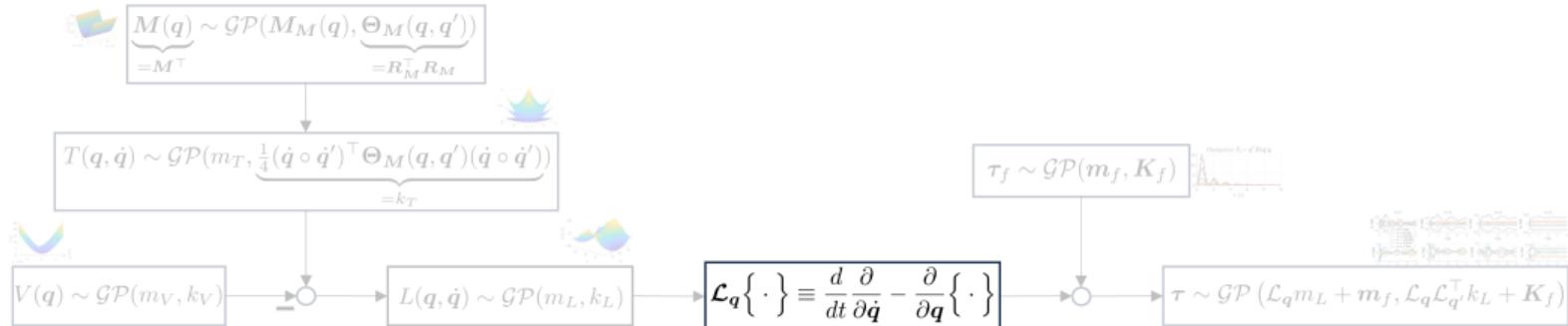
Specification of function space instead of black-box RKHS approach

Cholesky decomposed matrix kernels: $\Theta_k \equiv \mathbf{R}_k^\top(\mathbf{q}, \mathbf{q}') \mathbf{R}_k(\mathbf{q}, \mathbf{q}')$ with

$$\kappa_k(\mathbf{q}, \mathbf{q}') = \frac{1}{4} \left({}^{(k)}\mathbf{q} \circ {}^{(k)\prime}\mathbf{q}' \right)^\top \Theta_k(\mathbf{q}, \mathbf{q}') \left({}^{(k)}\mathbf{q} \circ {}^{(k)\prime}\mathbf{q}' \right)$$

Thus for stiffness and inertia matrices: $h_{knm}(\mathbf{q}) \sim \mathcal{GP}(\eta_{knm}(\mathbf{q}), \theta_{knm}(\mathbf{q}, \mathbf{q}'))$

L-GP: Differential Equation Embedding

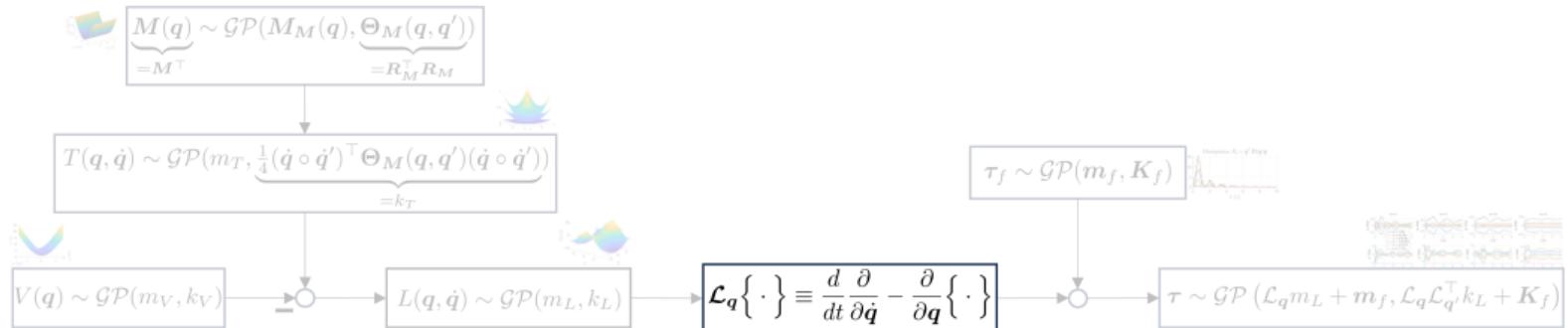


Linear Operators and GPs [Raissi+ 2017]

Applying a linear transformation operator \mathcal{T}_x leads to

$$\mathcal{T}_x f(x) \sim \mathcal{GP}\left(\mathcal{T}_x m(x), \mathcal{T}_x k(x, x') \mathcal{T}_{x'}^\top\right)$$

L-GP: Differential Equation Embedding

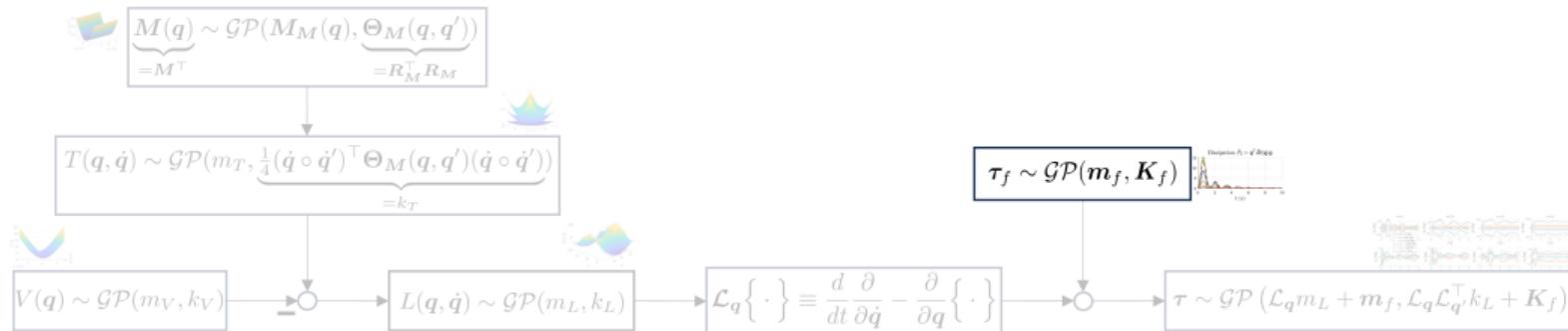


Exploiting linearity of differentiation for $\mathcal{L}_q := \left(\frac{\partial}{\partial \dot{q}^\top} \ddot{q} + \frac{\partial}{\partial q^\top} \dot{q} \right) \frac{\partial}{\partial \dot{q}} - \frac{\partial}{\partial q} \equiv \frac{d}{dt} \frac{\partial}{\partial \dot{q}} - \frac{\partial}{\partial q}$ leads to a vector-valued GP for the conservative torques:

$$\tau_c(q, \dot{q}, \ddot{q}) \sim \mathcal{GP} \left(\mathcal{L}_q m_L(q, \dot{q}), \mathcal{L}_q \mathcal{L}_{q'}^\top k_L(q, \dot{q}, q', \dot{q}') \right)$$

→ automatic differentiation tools can be applied [Geist+ 2020], as well as analytical [Lutter+ 2019; Evangelisti+ 2022], numerical [Ober-Blöbaum+ 2023] or symbolic approaches

L-GP: Dissipative Extension



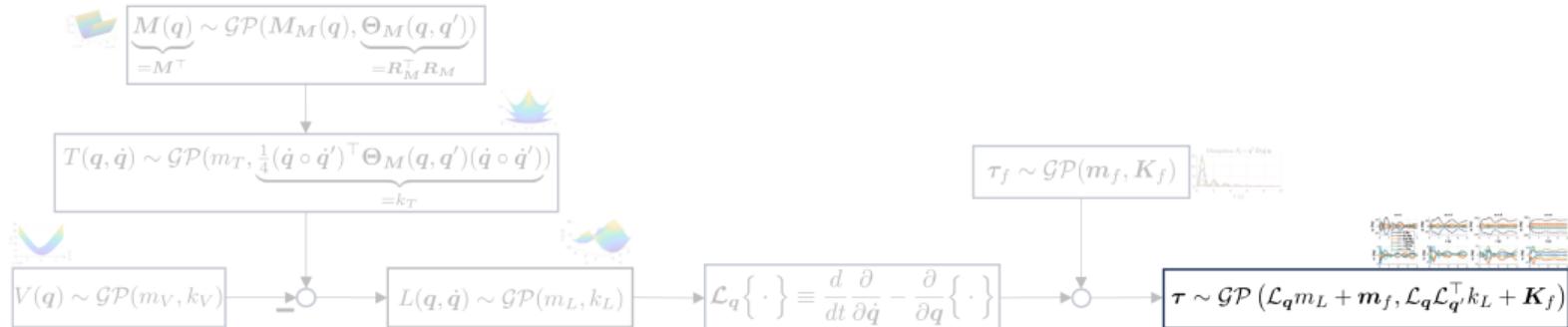
Incorporating friction

Extend composite GP structure by dissipative component $\tau_f \sim \mathcal{GP}(\mathbf{m}_f, \mathbf{K}_f)$

- for position-dependent $D(q)$: Rayleigh dissipation potential $R = \frac{1}{2}\dot{q}^\top D(q)\dot{q}$
- otherwise, e.g., $\tau_f = D_v(\dot{q})\dot{q} + D_c(\dot{q})\text{sign}(\dot{q})$ for viscous and Coulomb damping
- or black-box approach (friction as abstraction to combine various complex effects)

General friction model learning in compliance with Lagrangian mechanics is challenging!

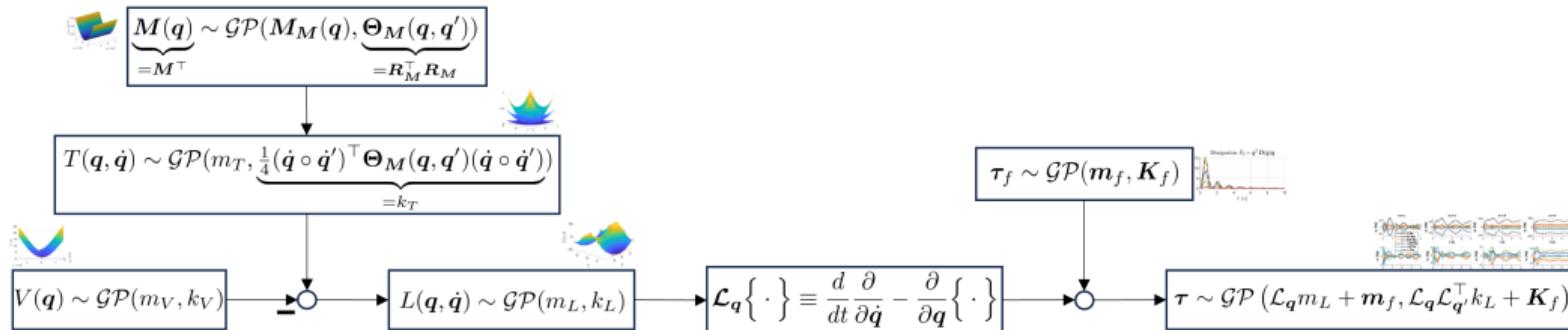
L-GP: Dissipative Extension



Multidimensional Composite GP Model

$$\tau = \tau_c + \tau_f \sim \mathcal{GP} \left(\mathcal{L}_q m_L + m_f, \mathcal{L}_q \mathcal{L}_{q'}^\top k_L + K_f \right)$$

L-GP: Dissipative Extension



Multidimensional Composite GP Model

$$\tau = \tau_c + \tau_f \sim \mathcal{GP} \left(\mathcal{L}_q m_L + m_f, \mathcal{L}_q \mathcal{L}_{q'}^\top k_L + K_f \right)$$

Estimates for $\hat{M}, \hat{C}, \hat{g}$ are obtained by leveraging $\tau_c = \hat{M}\ddot{q} + \hat{C}\dot{q} + \hat{g}$

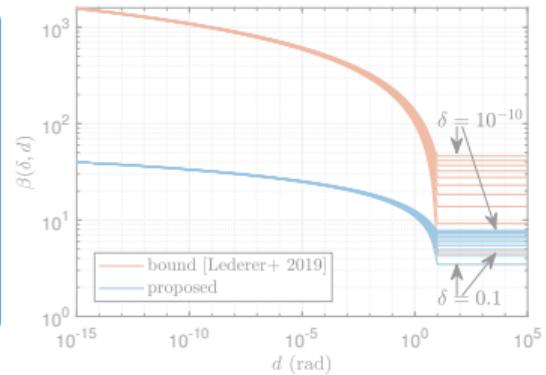
Uniform Multidimensional Error Bounds

Assumption

Unknown $\mathbf{f}: \mathcal{X} \rightarrow \mathbb{R}^N$ is drawn with hyperparameters ψ from

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}'; \psi))$$

of which $D \in \mathbb{N}$ noisy observations have been collected.



Lemma: Probabilistically Exact Error Bound

Pick $\delta \in (0, 1)$, $d \in \mathbb{R}^+$, and set $\beta(\delta, d) = \sqrt{2\Gamma_{N/2}^{-1}\left(\frac{\delta\gamma(N/2)}{M_d(\mathcal{X})}\right)}$. Then, $\forall \mathbf{x} \in \mathcal{X}$,

$$\Pr\left\{\|\mathbf{f}(\mathbf{x}) - \boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x})\| \leq \beta(\delta, d)\sqrt{\lambda(\Sigma_{\mathbf{f}}(\mathbf{x})) + s(d)} + (L_{\mathbf{f}} + L_{\boldsymbol{\mu}})d\right\} = 1 - \delta$$

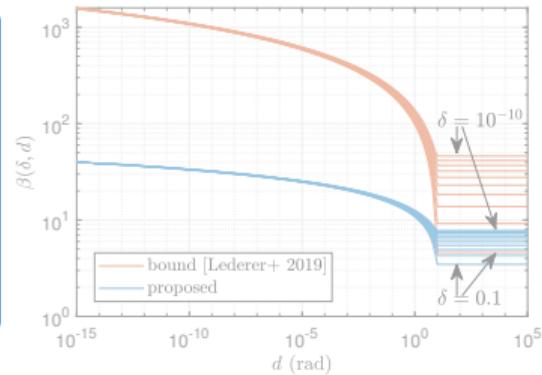
Uniform Multidimensional Error Bounds

Assumption

Unknown $\mathbf{f}: \mathcal{X} \rightarrow \mathbb{R}^N$ is drawn with hyperparameters ψ from

$$\mathbf{f}(x) \sim \mathcal{GP}(\mathbf{m}(x), \mathbf{K}(x, x'; \psi))$$

of which $D \in \mathbb{N}$ noisy observations have been collected.



Lemma: Probabilistically Exact Error Bound

Pick $\delta \in (0, 1)$, $d \in \mathbb{R}^+$, and set $\beta(\delta, d) = \sqrt{2\Gamma_{N/2}^{-1}\left(\frac{\delta\gamma(N/2)}{M_d(\mathcal{X})}\right)}$. Then, $\forall x \in \mathcal{X}$,

$$\Pr\left\{\|\mathbf{f}(x) - \mu_{\mathbf{f}}(x)\| \leq \beta(\delta, d)\sqrt{\underline{\lambda}(\Sigma_{\mathbf{f}}(x)) + s(d)} + (L_{\mathbf{f}} + L_{\mu})d\right\} = 1 - \delta$$

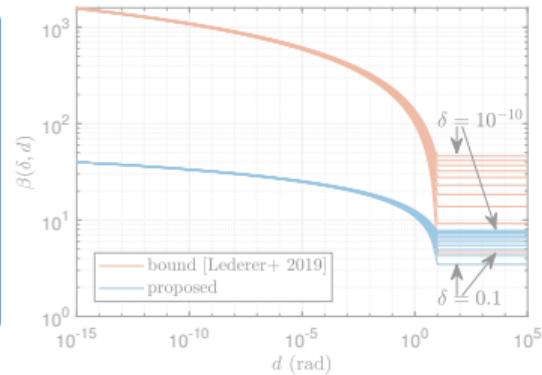
Uniform Multidimensional Error Bounds

Assumption

Unknown $\mathbf{f}: \mathcal{X} \rightarrow \mathbb{R}^N$ is drawn with hyperparameters ψ from

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}'; \psi))$$

of which $D \in \mathbb{N}$ noisy observations have been collected.



Lemma: Probabilistically Exact Error Bound

Pick $\delta \in (0, 1)$, $d \in \mathbb{R}^+$, and set $\beta(\delta, d) = \sqrt{2\Gamma_{N/2}^{-1} \left(\frac{\delta\gamma(N/2)}{M_d(\mathcal{X})} \right)}$. Then, $\forall \mathbf{x} \in \mathcal{X}$,

$$\Pr \left\{ \|\mathbf{f}(\mathbf{x}) - \boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x})\| \leq \beta(\delta, d) \sqrt{\lambda(\Sigma_{\mathbf{f}}(\mathbf{x})) + s(d)} + (L_{\mathbf{f}} + L_{\boldsymbol{\mu}})d \right\} = 1 - \delta$$

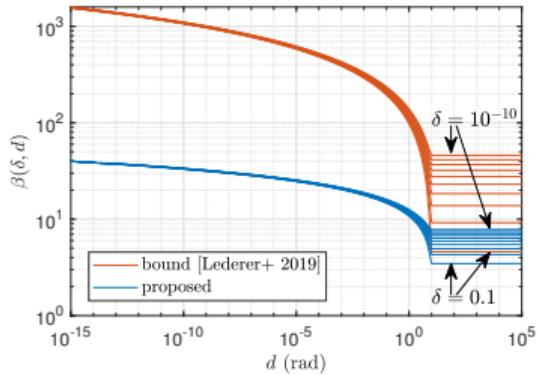
Uniform Multidimensional Error Bounds

Assumption

Unknown $\mathbf{f}: \mathcal{X} \rightarrow \mathbb{R}^N$ is drawn with hyperparameters ψ from

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}'; \psi))$$

of which $D \in \mathbb{N}$ noisy observations have been collected.



Lemma: Probabilistically Exact Error Bound

Pick $\delta \in (0, 1)$, $d \in \mathbb{R}^+$, and set $\beta(\delta, d) = \sqrt{2\Gamma_{N/2}^{-1}\left(\frac{\delta\gamma(N/2)}{M_d(\mathcal{X})}\right)}$. Then, $\forall \mathbf{x} \in \mathcal{X}$,

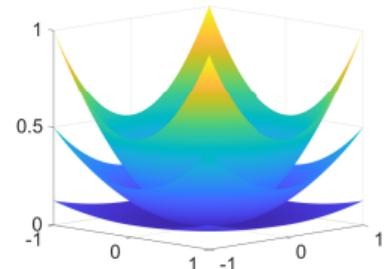
$$\Pr\left\{\|\mathbf{f}(\mathbf{x}) - \boldsymbol{\mu}_{\mathbf{f}}(\mathbf{x})\| \leq \beta(\delta, d)\sqrt{\underline{\lambda}(\Sigma_{\mathbf{f}}(\mathbf{x})) + s(d)} + (L_{\mathbf{f}} + L_{\boldsymbol{\mu}})d\right\} = 1 - \delta$$

Quadratic Form & Positive Definiteness

Lemma: Quadratic Form

Mean estimates $\hat{f}_k(\mathbf{q}) \equiv \text{E}[f_k(\mathbf{q})|\mathbf{y}, \gamma_0]$ are quadratic:

$$\hat{f}_k(\mathbf{q}) = \frac{1}{2} \mathbf{q}^{(k)\top} \hat{\mathbf{H}}_k(\mathbf{q}) \mathbf{q}^{(k)}$$



Theorem: Positive Definiteness (PDF)

PDF is guaranteed with high probability:

$$\Pr\{\hat{\mathbf{H}}_k(\mathbf{q}) \succ \mathbf{0}\} = 1 - \delta_k(\mathbf{q})$$

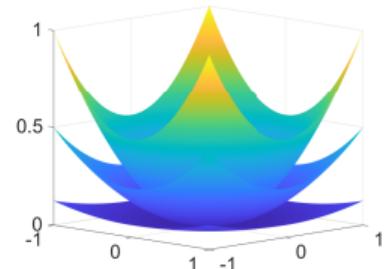
$$\text{and bounded } \delta_k(\mathbf{q}) \leq \Pr\left\{\lambda_N(\mathbf{H}_{k0}) + \exp\left[-\text{vec}(\Sigma_{d_k}^{-1})^\top \mathbf{D}(\mathbf{q})\right] \boldsymbol{\beta}_k \leq \mathbf{0}\right\}$$

Quadratic Form & Positive Definiteness

Lemma: Quadratic Form

Mean estimates $\hat{f}_k(\mathbf{q}) \equiv \text{E}[f_k(\mathbf{q})|\mathbf{y}, \gamma_0]$ are quadratic:

$$\hat{f}_k(\mathbf{q}) = \frac{1}{2} \mathbf{q}^{(k)\top} \hat{\mathbf{H}}_k(\mathbf{q}) \mathbf{q}^{(k)}$$



Theorem: Positive Definiteness (PDF)

PDF is guaranteed with high probability:

$$\Pr\{\hat{\mathbf{H}}_k(\mathbf{q}) \succ \mathbf{0}\} = 1 - \delta_k(\mathbf{q})$$

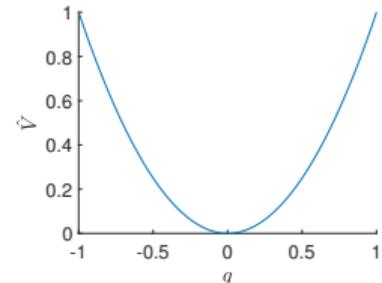
$$\text{and bounded } \delta_k(\mathbf{q}) \leq \Pr\left\{ \lambda_N(\mathbf{H}_{k0}) + \exp\left[-\text{vec}(\boldsymbol{\Sigma}_{\mathbf{d}_k}^{-1})^\top \mathbf{D}(\mathbf{q})\right] \boldsymbol{\beta}_k \leq 0 \right\}$$

Equilibria & Conservatism

Theorem: Equilibrium

L-GP based potential $\hat{V}(\mathbf{q}) \equiv E[V(\mathbf{q})|\mathbf{y}, \gamma_0]$ guarantees

$$\hat{V}(\mathbf{0}) = 0, \quad \nabla_{\mathbf{q}} \hat{V}(\mathbf{0}) = \mathbf{0}$$



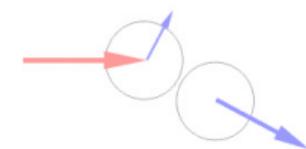
Induced by differentiation [Raissi+ 2017]: multidimensional torque GP $\hat{\tau} \equiv E[\tau|\mathbf{y}, \gamma_0]$

$$\hat{\tau}(q, \dot{q}, \ddot{q}) = \hat{M}(q)\ddot{q} + \hat{C}(q, \dot{q})\dot{q} + \hat{g}(q) \quad (1)$$

Corollary: Energy Conservation

L-GP dynamics (1) are passive/lossless w.r.t. $\hat{E} = \hat{T} + \hat{V}$:

$$\dot{\hat{E}} = \dot{\mathbf{q}}^\top \hat{\tau}$$

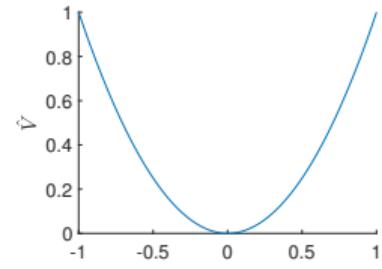


Equilibria & Conservatism

Theorem: Equilibrium

L-GP based potential $\hat{V}(\mathbf{q}) \equiv E[V(\mathbf{q})|\mathbf{y}, \gamma_0]$ guarantees

$$\hat{V}(\mathbf{0}) = 0, \quad \nabla_{\mathbf{q}} \hat{V}(\mathbf{0}) = \mathbf{0}$$



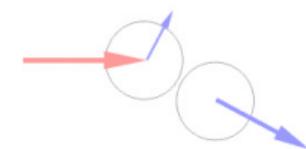
Induced by differentiation [Raissi+ 2017]: multidimensional torque GP $\hat{\tau} \equiv E[\boldsymbol{\tau}|\mathbf{y}, \gamma_0]$

$$\hat{\tau}(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}) = \hat{M}(\mathbf{q})\ddot{\mathbf{q}} + \hat{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \hat{g}(\mathbf{q}) \quad (1)$$

Corollary: Energy Conservation

L-GP dynamics (1) are passive/lossless w.r.t. $\hat{E} = \hat{T} + \hat{V}$:

$$\dot{\hat{E}} = \dot{\mathbf{q}}^\top \hat{\tau}$$

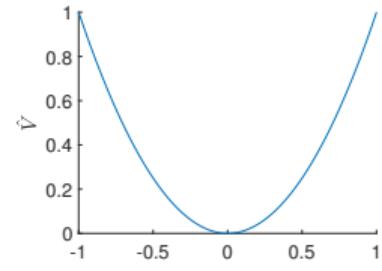


Equilibria & Conservatism

Theorem: Equilibrium

L-GP based potential $\hat{V}(\mathbf{q}) \equiv E[V(\mathbf{q})|\mathbf{y}, \gamma_0]$ guarantees

$$\hat{V}(\mathbf{0}) = 0, \quad \nabla_{\mathbf{q}} \hat{V}(\mathbf{0}) = \mathbf{0}$$



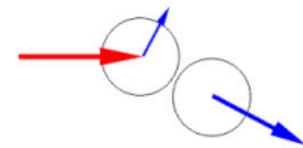
Induced by differentiation [Raissi+ 2017]: multidimensional torque GP $\hat{\tau} \equiv E[\boldsymbol{\tau}|\mathbf{y}, \gamma_0]$

$$\hat{\tau}(q, \dot{q}, \ddot{q}) = \hat{M}(q)\ddot{q} + \hat{C}(q, \dot{q})\dot{q} + \hat{g}(q) \quad (1)$$

Corollary: Energy Conservation

L-GP dynamics (1) are passive/lossless w.r.t. $\hat{E} = \hat{T} + \hat{V}$:

$$\dot{\hat{E}} = \dot{\mathbf{q}}^\top \hat{\tau}$$



Monte Carlo: Generalization Comparison

Setup:

- Conservative two-link pendulum from [Lutter+ 2023]: $N = 2$
- $\forall i$: 8 random (uniformly distributed) data sets \mathcal{D}_i with $\mathbf{q} \in [-1, 1]^N$, $\dot{\mathbf{q}} = \ddot{\mathbf{q}} = \mathbf{1}$
- DeLaN: parametrization from [Lutter+ 2023] with reference dimension 2×64

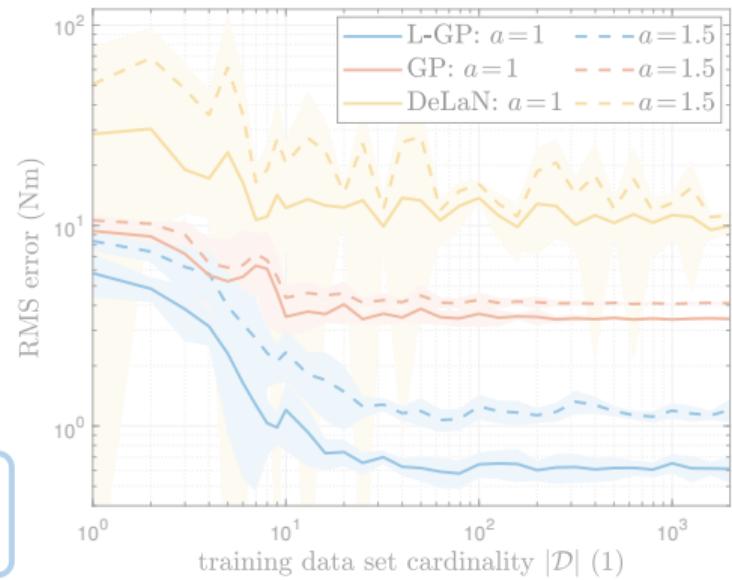
Training:

- (L)-GPs: log-likelihood
- DeLaN: sum of squared errors
- noise: $\sigma_\tau = 0.1 \text{ Nm}$, $\sigma_{\dot{\mathbf{q}}} = \pi/180 \text{ s}^{-2}$

Evaluation:

- 2 grids (10^6 samples) in domain
 $(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}) \in [-a, a]^{3N} \Rightarrow$ generalization
- metric: overall torque RMS error

physical structuring \Rightarrow generalization \uparrow
L-GP: data-efficiency \uparrow , robustness to noise \uparrow



Monte Carlo: Generalization Comparison

Setup:

- Conservative two-link pendulum from [Lutter+ 2023]: $N = 2$
- $\forall i$: 8 random (uniformly distributed) data sets \mathcal{D}_i with $\mathbf{q} \in [-1, 1]^N$, $\dot{\mathbf{q}} = \ddot{\mathbf{q}} = \mathbf{1}$
- DeLaN: parametrization from [Lutter+ 2023] with reference dimension 2×64

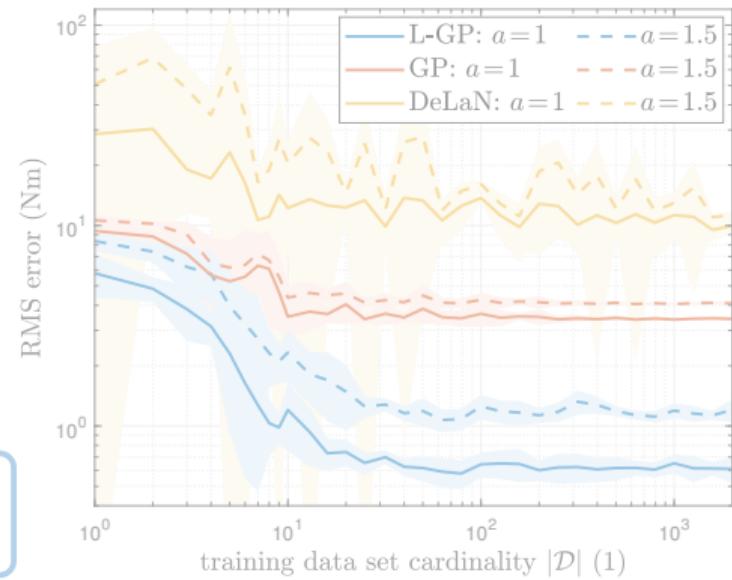
Training:

- (L)-GPs: log-likelihood
- DeLaN: sum of squared errors
- noise: $\sigma_\tau = 0.1 \text{ Nm}$, $\sigma_{\dot{\mathbf{q}}} = \pi/180 \text{ s}^{-2}$

Evaluation:

- 2 grids (10^6 samples) in domain
 $(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}) \in [-a, a]^{3N} \Rightarrow$ generalization
- metric: overall torque RMS error

physical structuring \Rightarrow generalization \uparrow
L-GP: data-efficiency \uparrow , robustness to noise \uparrow



Monte Carlo: Generalization Comparison

Setup:

- Conservative two-link pendulum from [Lutter+ 2023]: $N = 2$
- $\forall i$: 8 random (uniformly distributed) data sets \mathcal{D}_i with $\mathbf{q} \in [-1, 1]^N$, $\dot{\mathbf{q}} = \ddot{\mathbf{q}} = \mathbf{1}$
- DeLaN: parametrization from [Lutter+ 2023] with reference dimension 2×64

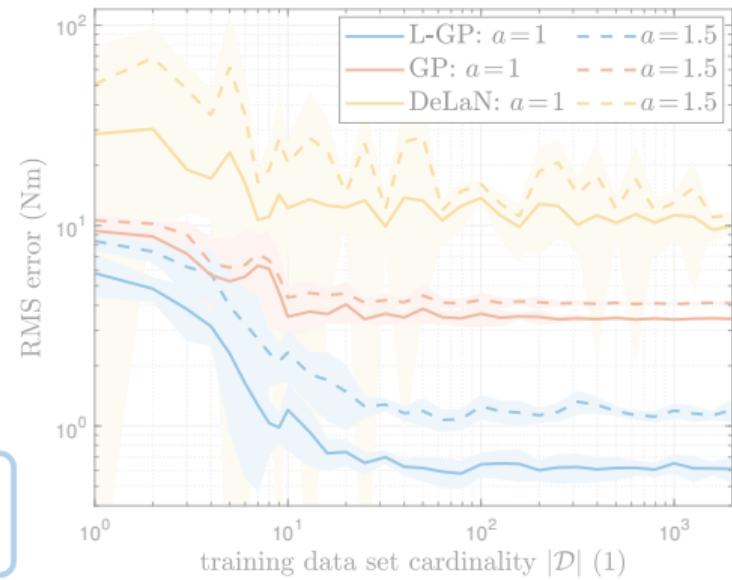
Training:

- (L)-GPs: log-likelihood
- DeLaN: sum of squared errors
- noise: $\sigma_\tau = 0.1 \text{ Nm}$, $\sigma_{\dot{\mathbf{q}}} = \pi/180 \text{ s}^{-2}$

Evaluation:

- 2 grids (10^6 samples) in domain
 $(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}) \in [-a, a]^{3N} \Rightarrow$ generalization
- metric: overall torque RMS error

physical structuring \Rightarrow generalization \uparrow
L-GP: data-efficiency \uparrow , robustness to noise \uparrow



Monte Carlo: Generalization Comparison

Setup:

- Conservative two-link pendulum from [Lutter+ 2023]: $N = 2$
- $\forall i$: 8 random (uniformly distributed) data sets \mathcal{D}_i with $\mathbf{q} \in [-1, 1]^N$, $\dot{\mathbf{q}} = \ddot{\mathbf{q}} = \mathbf{1}$
- DeLaN: parametrization from [Lutter+ 2023] with reference dimension 2×64

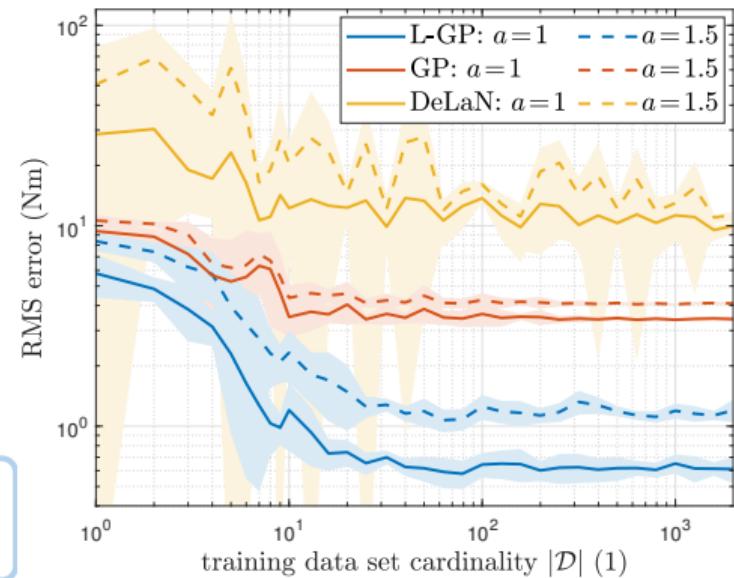
Training:

- (L)-GPs: log-likelihood
- DeLaN: sum of squared errors
- noise: $\sigma_\tau = 0.1 \text{ Nm}$, $\sigma_{\dot{\mathbf{q}}} = \pi/180 \text{ s}^{-2}$

Evaluation:

- 2 grids (10^6 samples) in domain $(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}) \in [-a, a]^{3N}$ \Rightarrow generalization
- metric: overall torque RMS error

physical structuring \Rightarrow generalization \uparrow
L-GP: data-efficiency \uparrow , robustness to noise \uparrow



Monte Carlo: Generalization Comparison

Setup:

- Conservative two-link pendulum from [Lutter+ 2023]: $N = 2$
- $\forall i$: 8 random (uniformly distributed) data sets \mathcal{D}_i with $\mathbf{q} \in [-1, 1]^N$, $\dot{\mathbf{q}} = \ddot{\mathbf{q}} = \mathbf{1}$
- DeLaN: parametrization from [Lutter+ 2023] with reference dimension 2×64

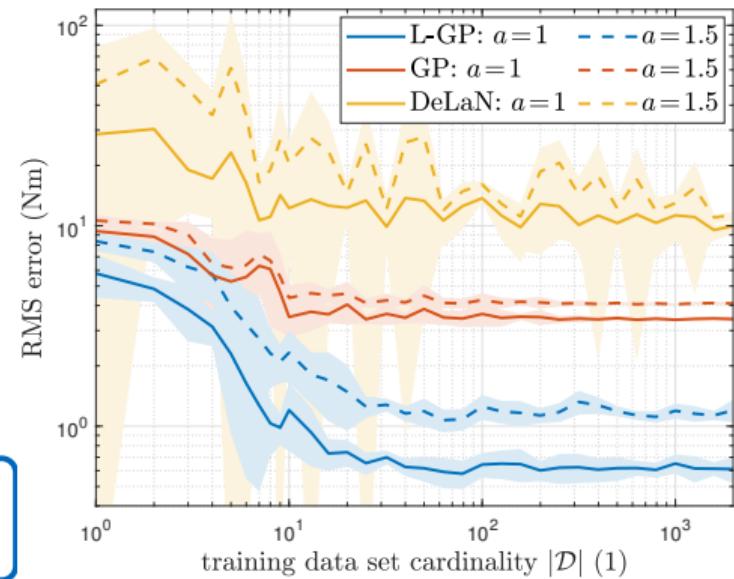
Training:

- (L)-GPs: log-likelihood
- DeLaN: sum of squared errors
- noise: $\sigma_\tau = 0.1 \text{ Nm}$, $\sigma_{\dot{\mathbf{q}}} = \pi/180 \text{ s}^{-2}$

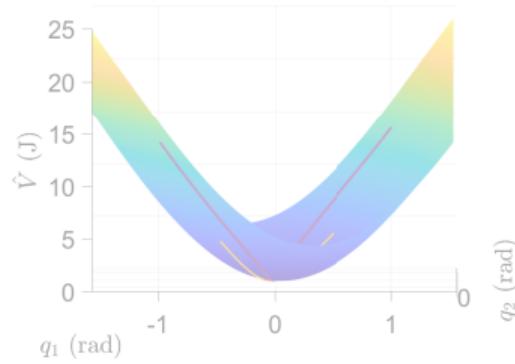
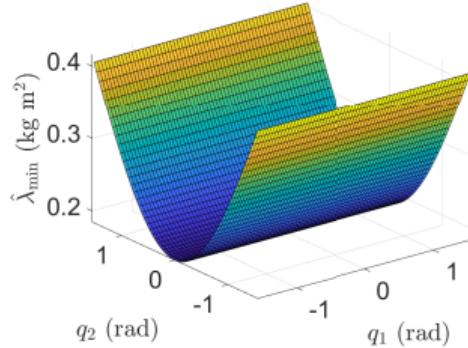
Evaluation:

- 2 grids (10^6 samples) in domain $(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}) \in [-a, a]^{3N}$ \Rightarrow generalization
- metric: overall torque RMS error

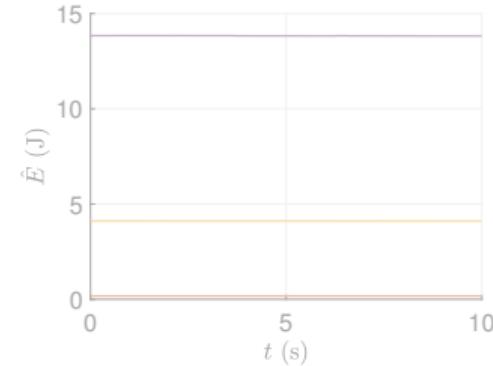
physical structuring \Rightarrow generalization \uparrow
L-GP: data-efficiency \uparrow , robustness to noise \uparrow



Conservative L-GP Validation



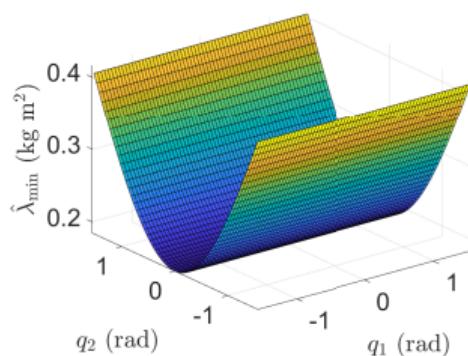
(b) Potential



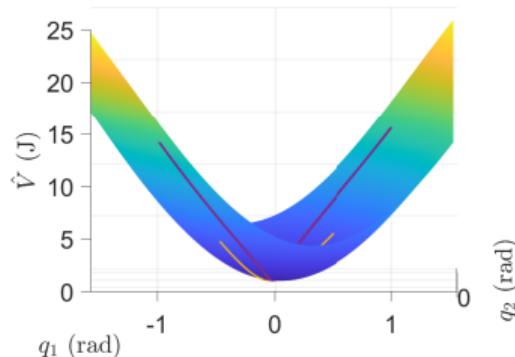
(c) Unforced dynamics

L-GP preserves PDF, conservative structure, equilibria etc. → physically consistent

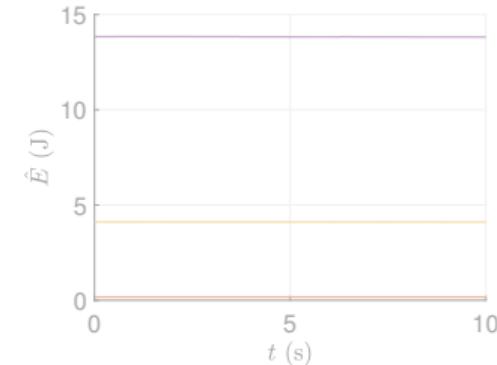
Conservative L-GP Validation



(a) Mass inertia matrix



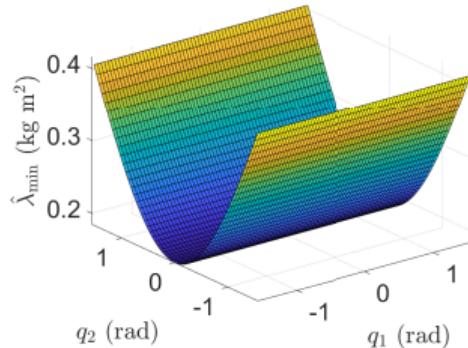
(b) Potential



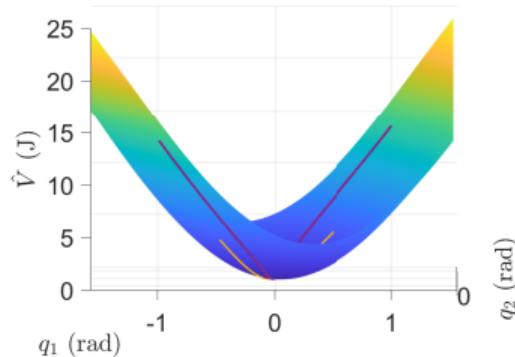
(c) Unforced dynamics

L-GP preserves PDF, conservative structure, equilibria etc. → physically consistent

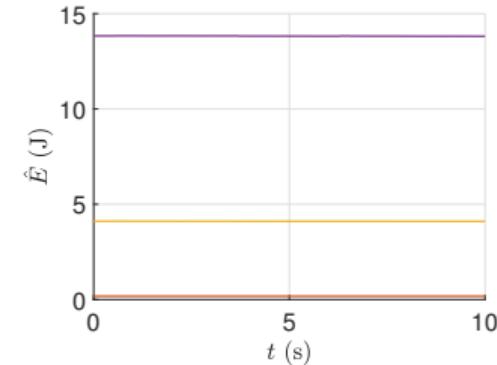
Conservative L-GP Validation



(a) Mass inertia matrix



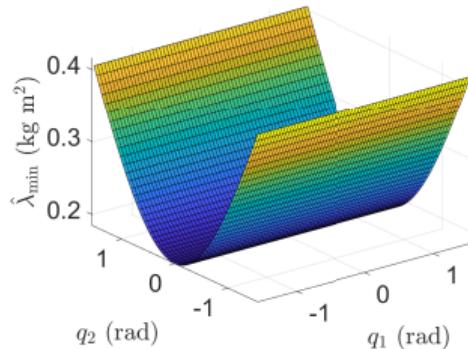
(b) Potential



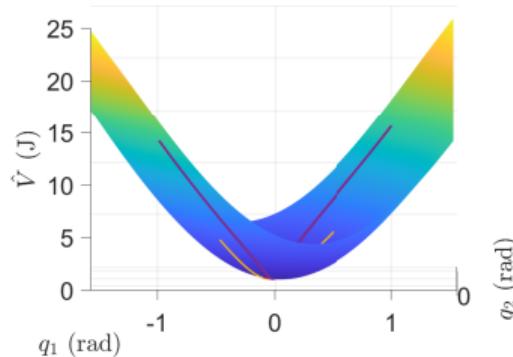
(c) Unforced dynamics

L-GP preserves PDF, conservative structure, equilibria etc. → physically consistent

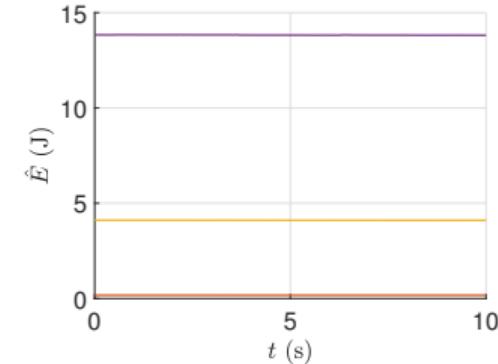
Conservative L-GP Validation



(a) Mass inertia matrix



(b) Potential



(c) Unforced dynamics

L-GP preserves PDF, conservative structure, equilibria etc. → physically consistent

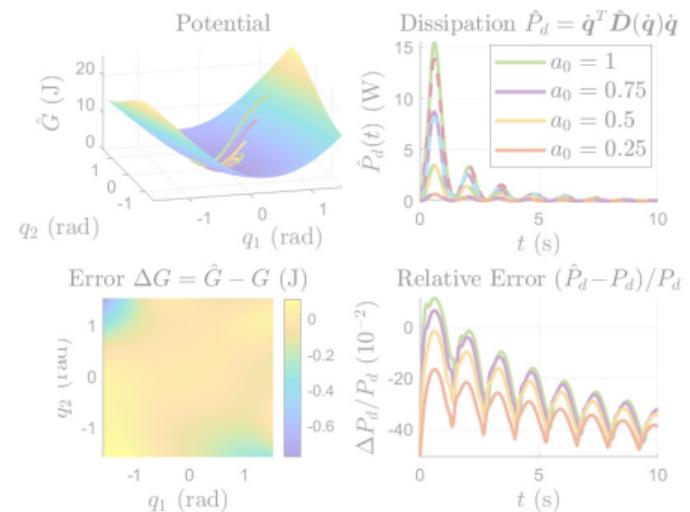
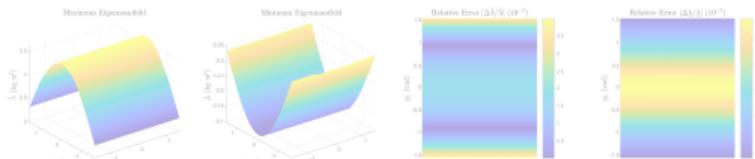
Two-link: Additional Dissipation

Setup:

- Unit two-link [Murray+ 1994] with gravity $g=10 \text{ ms}^{-2}$, dampings $d_i=1 \text{ Nm rad s}^{-1}$
- Parameter estimates: $\pm 50\%$ error

Training:

- $D = 34$ noisy measurement pairs
- noise: $\sigma_\epsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization



L-GP → accurate regression of conservative and dissipative subcomponents

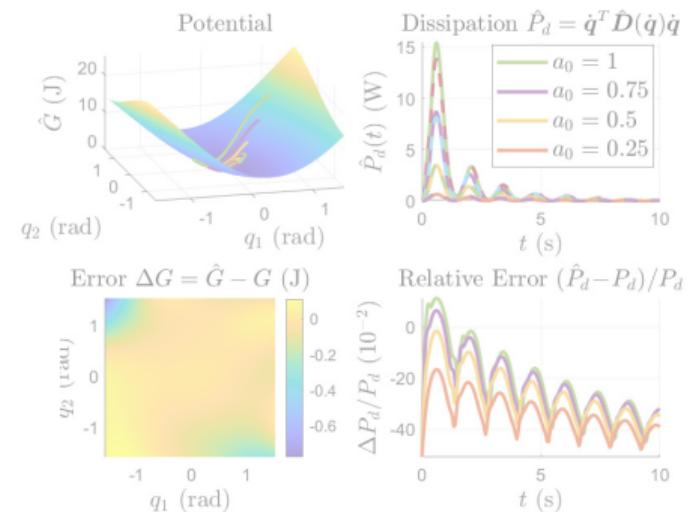
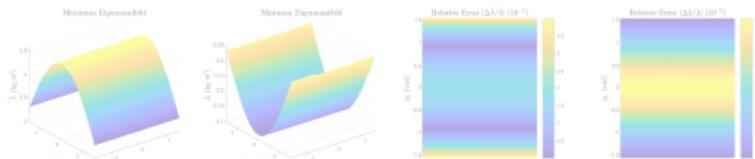
Two-link: Additional Dissipation

Setup:

- Unit two-link [Murray+ 1994] with gravity $g=10 \text{ ms}^{-2}$, dampings $d_i=1 \text{ Nm rad s}^{-1}$
- Parameter estimates: $\pm 50\%$ error

Training:

- $D = 34$ noisy measurement pairs
- noise: $\sigma_\epsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization



L-GP → accurate regression of conservative and dissipative subcomponents

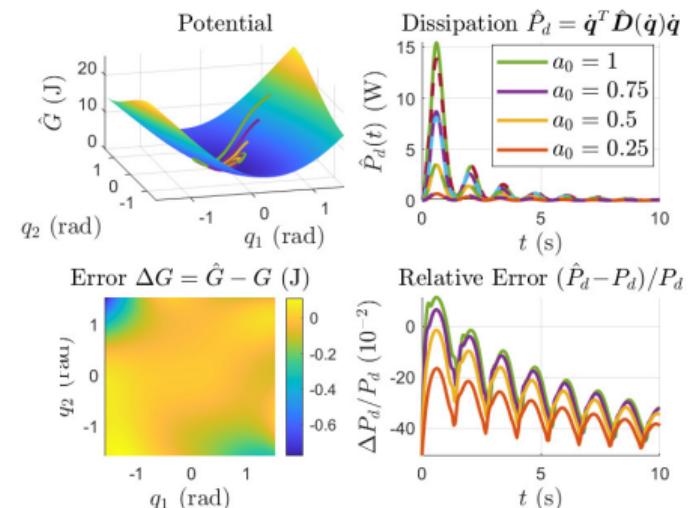
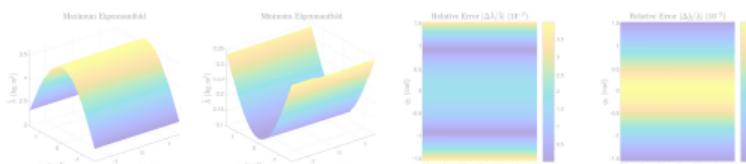
Two-link: Additional Dissipation

Setup:

- Unit two-link [Murray+ 1994] with gravity $g=10 \text{ ms}^{-2}$, dampings $d_i=1 \text{ Nm rad s}^{-1}$
- Parameter estimates: $\pm 50\%$ error

Training:

- $D = 34$ noisy measurement pairs
- noise: $\sigma_\epsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization



L-GP → accurate regression of conservative and dissipative subcomponents

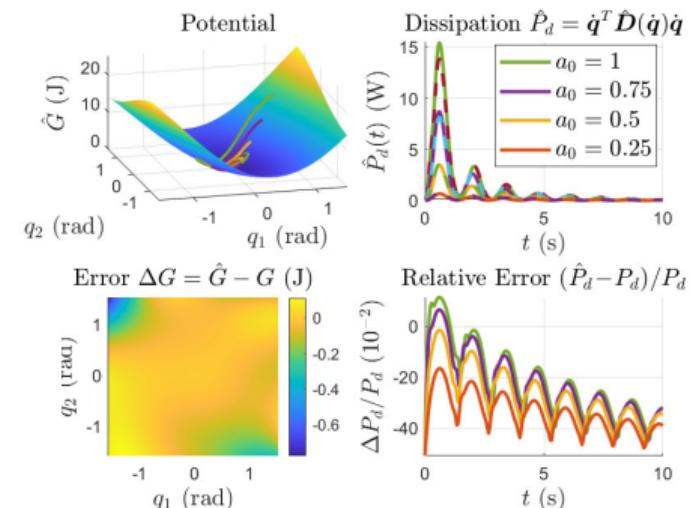
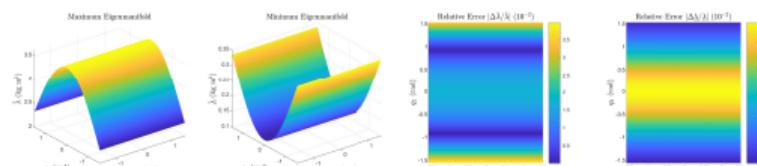
Two-link: Additional Dissipation

Setup:

- Unit two-link [Murray+ 1994] with gravity $g=10 \text{ ms}^{-2}$, dampings $d_i=1 \text{ Nm rad s}^{-1}$
- Parameter estimates: $\pm 50\%$ error

Training:

- $D = 34$ noisy measurement pairs
- noise: $\sigma_\varepsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization



L-GP → accurate regression of conservative and dissipative subcomponents

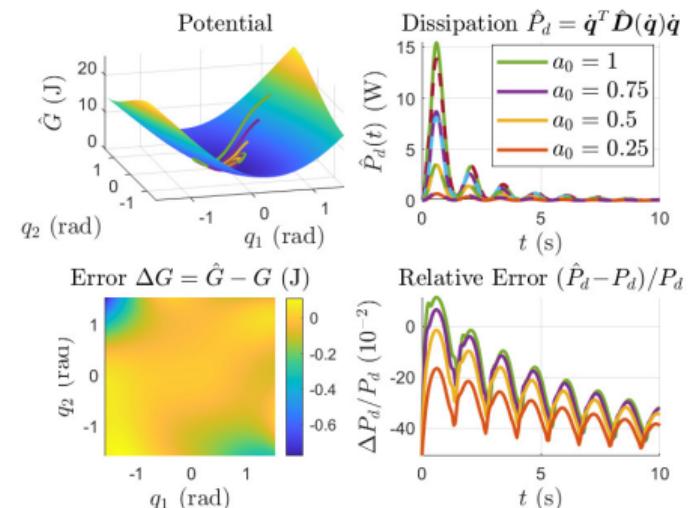
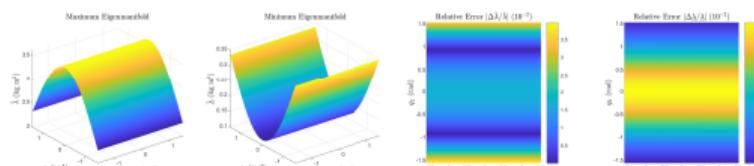
Two-link: Additional Dissipation

Setup:

- Unit two-link [Murray+ 1994] with gravity $g=10 \text{ ms}^{-2}$, dampings $d_i=1 \text{ Nm rad s}^{-1}$
- Parameter estimates: $\pm 50\%$ error

Training:

- $D = 34$ noisy measurement pairs
- noise: $\sigma_\epsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization



L-GP → accurate regression of conservative and dissipative subcomponents

L-GP-based Tracking Control

Setup:

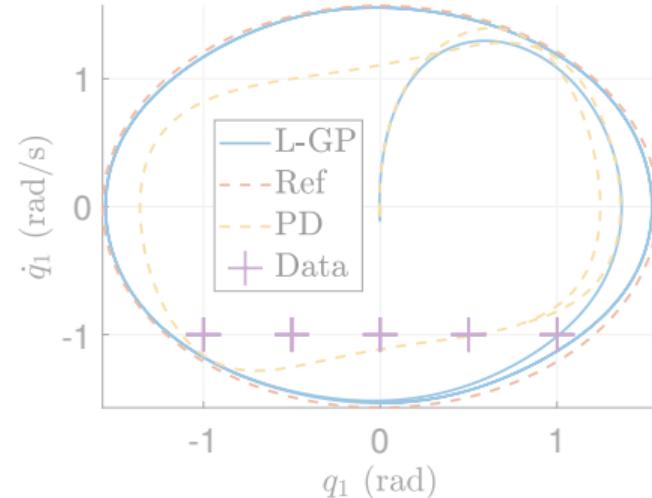
- Unit two-link from [Murray+ 1994]
- Parameter estimates: $\pm 50\%$ error

Training:

- noise: $\sigma_\varepsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization

Augmented PD law

$$\tau = \hat{M}(q)\ddot{q}_d + \hat{C}(q, \dot{q})\dot{q}_d + \hat{g}(q) - K_p e - K_d \dot{e}$$



L-GP → passivity- & energy-based control laws directly applicable

L-GP-based Tracking Control

Setup:

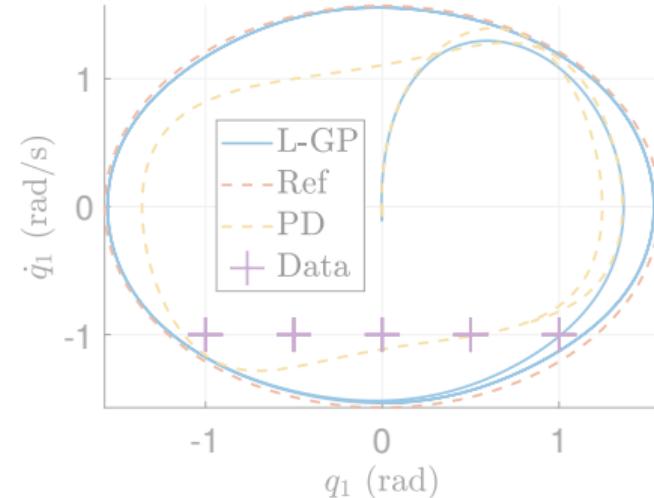
- Unit two-link from [Murray+ 1994]
- Parameter estimates: $\pm 50\%$ error

Training:

- noise: $\sigma_\varepsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization

Augmented PD law

$$\tau = \hat{M}(q)\ddot{q}_d + \hat{C}(q, \dot{q})\dot{q}_d + \hat{g}(q) - K_p e - K_d \dot{e}$$



L-GP → passivity- & energy-based control laws directly applicable

L-GP-based Tracking Control

Setup:

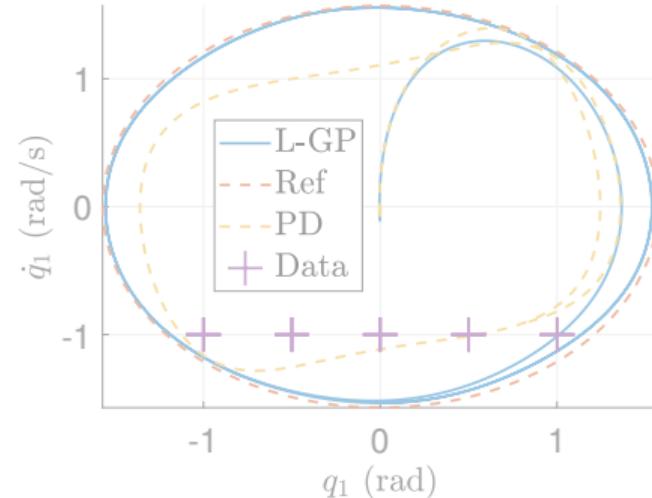
- Unit two-link from [Murray+ 1994]
- Parameter estimates: $\pm 50\%$ error

Training:

- noise: $\sigma_\varepsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization

Augmented PD law

$$\tau = \hat{\mathbf{M}}(\mathbf{q})\ddot{\mathbf{q}}_d + \hat{\mathbf{C}}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}}_d + \hat{\mathbf{g}}(\mathbf{q}) - \mathbf{K}_p \mathbf{e} - \mathbf{K}_d \dot{\mathbf{e}}$$



L-GP → passivity- & energy-based control laws directly applicable

L-GP-based Tracking Control

Setup:

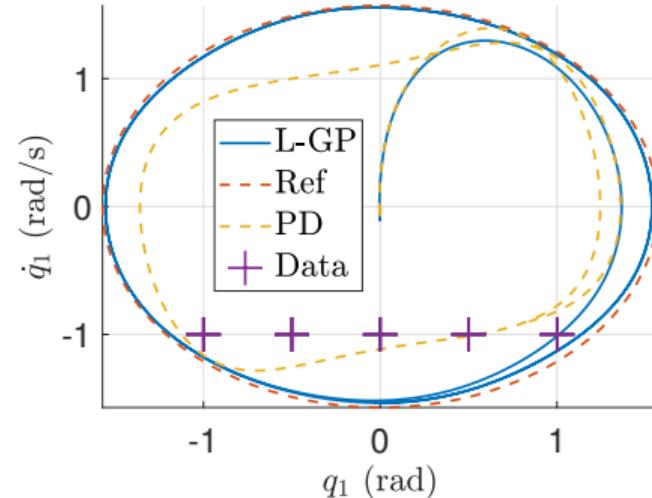
- Unit two-link from [Murray+ 1994]
- Parameter estimates: $\pm 50\%$ error

Training:

- noise: $\sigma_\varepsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization

Augmented PD law

$$\tau = \hat{M}(q)\ddot{q}_d + \hat{C}(q, \dot{q})\dot{q}_d + \hat{g}(q) - K_p e - K_d \dot{e}$$



L-GP \rightarrow passivity- & energy-based control laws directly applicable

L-GP-based Tracking Control

Setup:

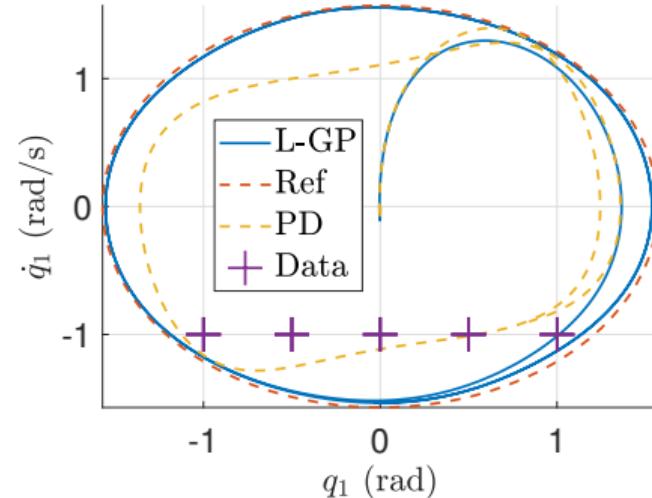
- Unit two-link from [Murray+ 1994]
- Parameter estimates: $\pm 50\%$ error

Training:

- noise: $\sigma_\varepsilon = 0.1 \text{ Nm}$, $\sigma_\alpha = \pi/180 \text{ rad/s}^2$
- log-likelihood optimization

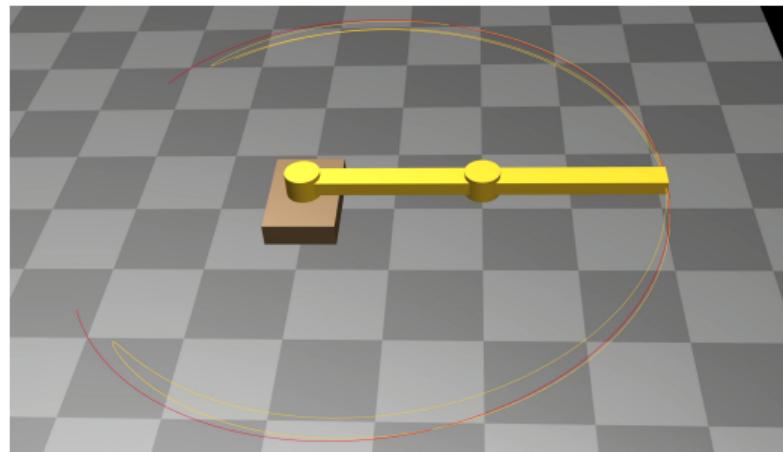
Augmented PD law

$$\tau = \hat{M}(q)\ddot{q}_d + \hat{C}(q, \dot{q})\dot{q}_d + \hat{g}(q) - K_p e - K_d \dot{e}$$

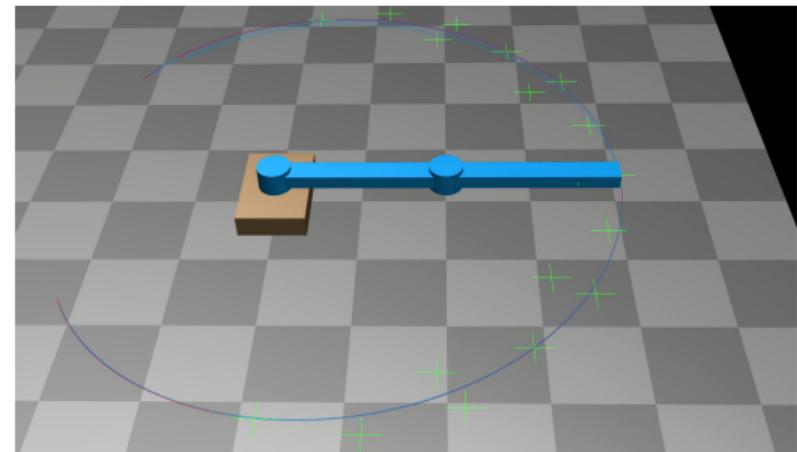


L-GP \rightarrow passivity- & energy-based control laws directly applicable

L-GP-based Tracking Control



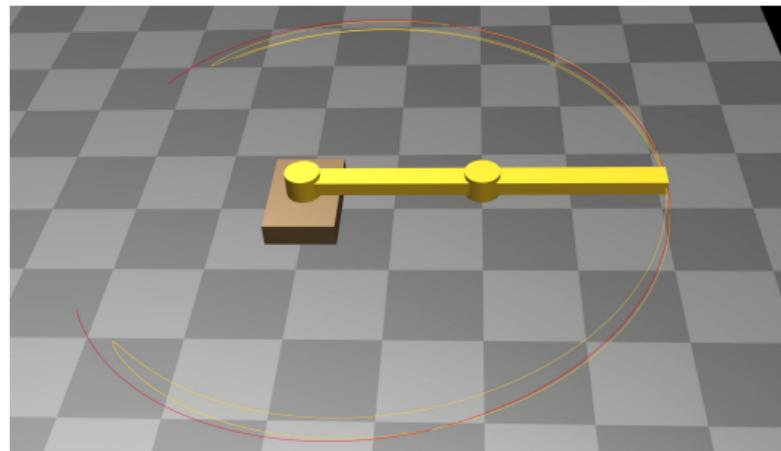
(a) Parametric model-based PD law



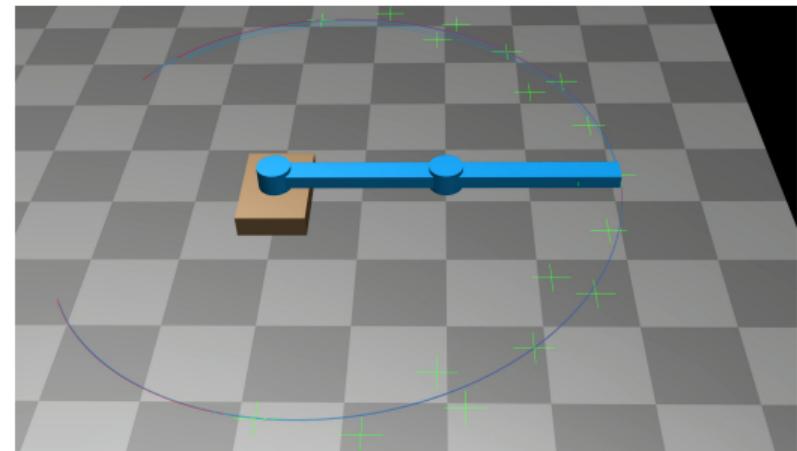
(b) L-GP-augmented PD law

L-GP → significantly higher performance

L-GP-based Tracking Control



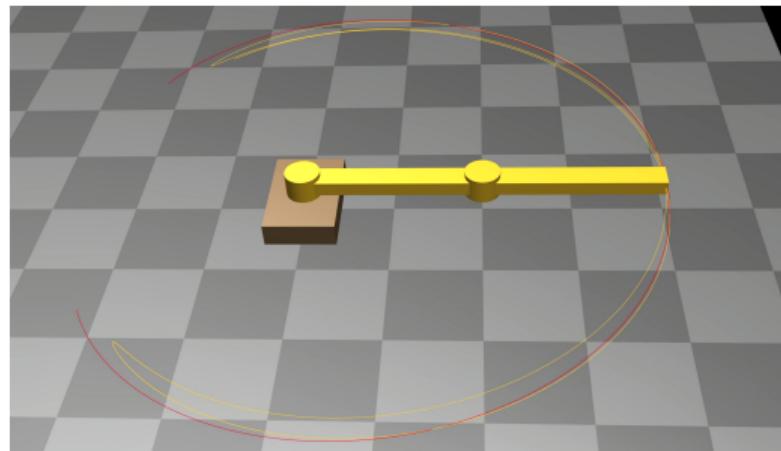
(a) Parametric model-based PD law



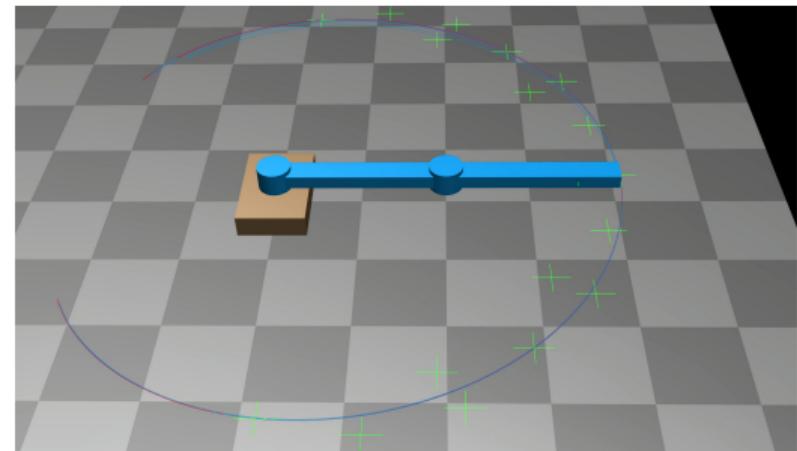
(b) L-GP-augmented PD law

L-GP → significantly higher performance

L-GP-based Tracking Control



(a) Parametric model-based PD law



(b) L-GP-augmented PD law

L-GP → significantly higher performance

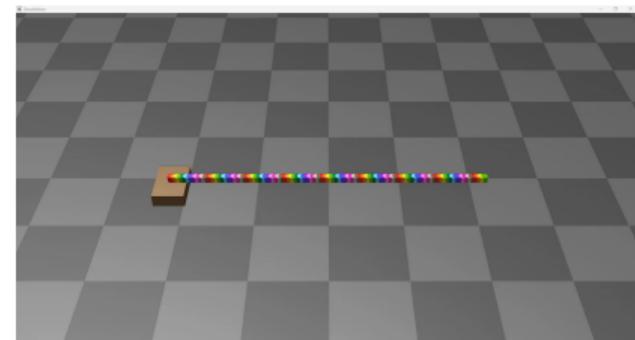
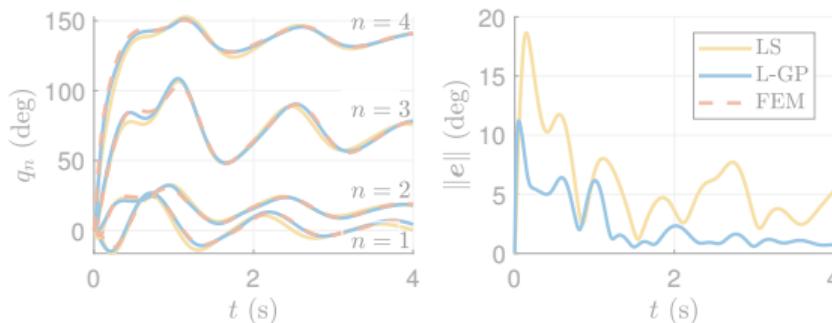
L-GP-based Control: Soft Robotics

Setup:

- FEM of soft robot from [DellaSantina+ 2020]
- Discretization into $N_{\text{FEM}} = 100$ with $k_{\text{FEM}} = 10 \text{ Nm/rad}$ and $d_{\text{FEM}} = 1 \text{ Nms/rad}$

Training:

- $D = 24$ equidistant trajectory samples
- step $a = 1 \text{ Nm}$, noise $\sigma_\epsilon = 0.01a$
- dynamical resimulation of trajectory



Accurate L-GP modeling & control with $4N = 16$ equivalent DOFs

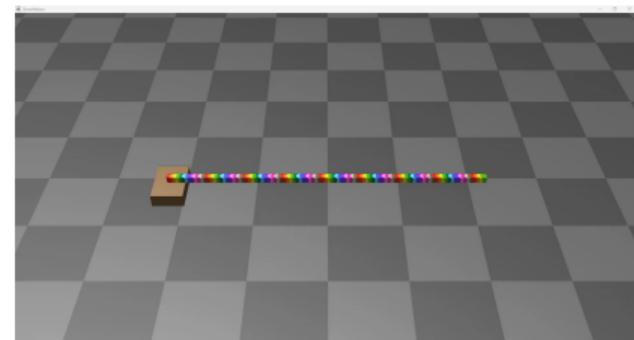
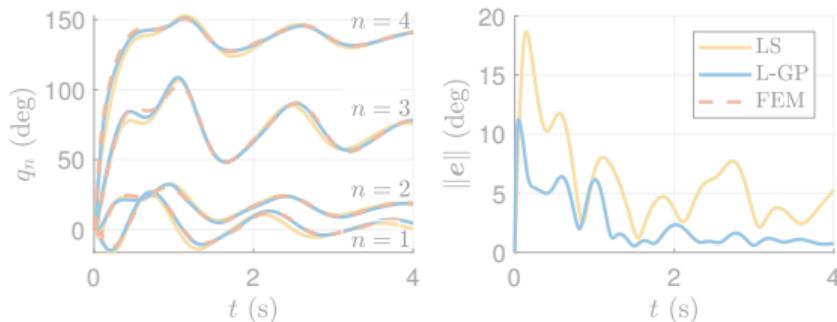
L-GP-based Control: Soft Robotics

Setup:

- FEM of soft robot from [DellaSantina+ 2020]
- Discretization into $N_{\text{FEM}} = 100$ with $k_{\text{FEM}} = 10 \text{ Nm/rad}$ and $d_{\text{FEM}} = 1 \text{ Nms/rad}$

Training:

- $D = 24$ equidistant trajectory samples
- step $a = 1 \text{ Nm}$, noise $\sigma_\epsilon = 0.01a$
- dynamical resimulation of trajectory



Accurate L-GP modeling & control with $4N = 16$ equivalent DOFs

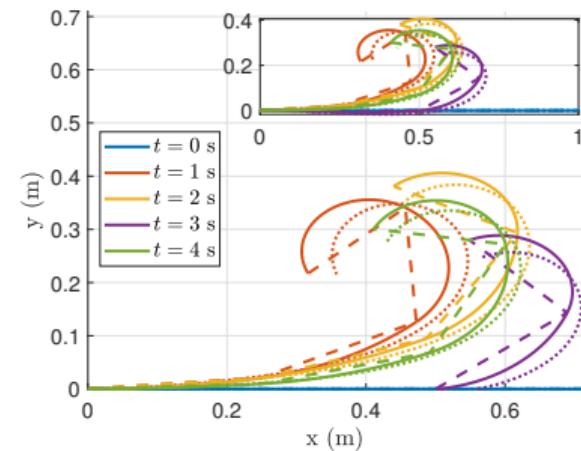
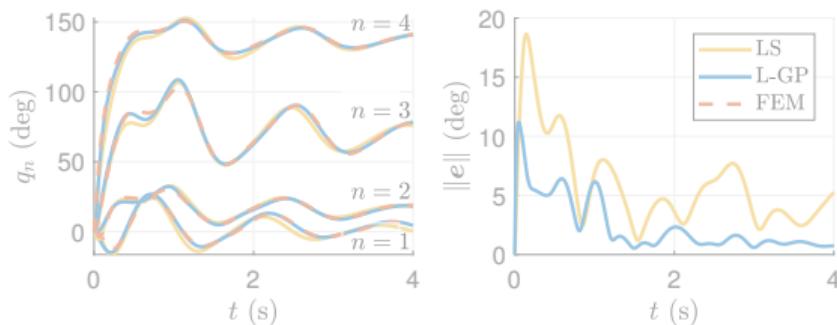
L-GP-based Control: Soft Robotics

Setup:

- FEM of soft robot from [DellaSantina+ 2020]
- Discretization into $N_{\text{FEM}} = 100$ with $k_{\text{FEM}} = 10 \text{ Nm/rad}$ and $d_{\text{FEM}} = 1 \text{ Nms/rad}$

Training:

- $D = 24$ equidistant trajectory samples
- step $a = 1 \text{ Nm}$, noise $\sigma_\epsilon = 0.01a$
- dynamical resimulation of trajectory



Accurate L-GP modeling & control with $4N = 16$ equivalent DOFs

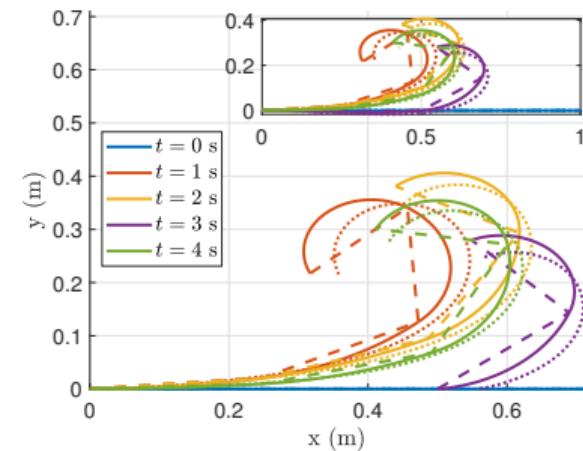
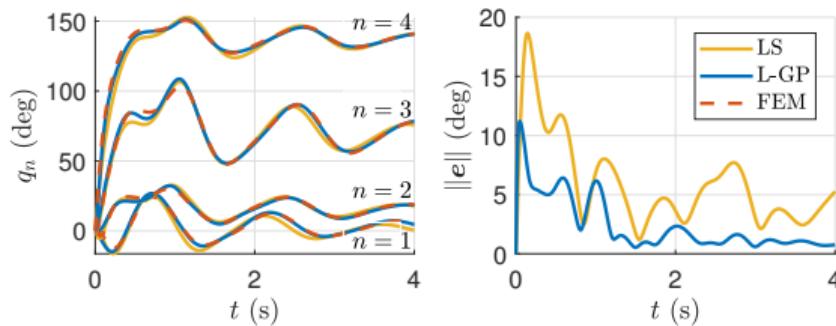
L-GP-based Control: Soft Robotics

Setup:

- FEM of soft robot from [DellaSantina+ 2020]
- Discretization into $N_{\text{FEM}} = 100$ with $k_{\text{FEM}} = 10 \text{ Nm/rad}$ and $d_{\text{FEM}} = 1 \text{ Nms/rad}$

Training:

- $D = 24$ equidistant trajectory samples
- step $a = 1 \text{ Nm}$, noise $\sigma_\epsilon = 0.01a$
- dynamical resimulation of trajectory



Accurate L-GP modeling & control with $4N = 16$ equivalent DOFs

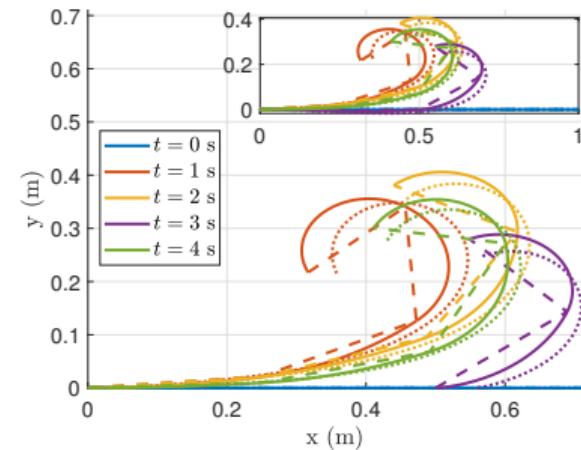
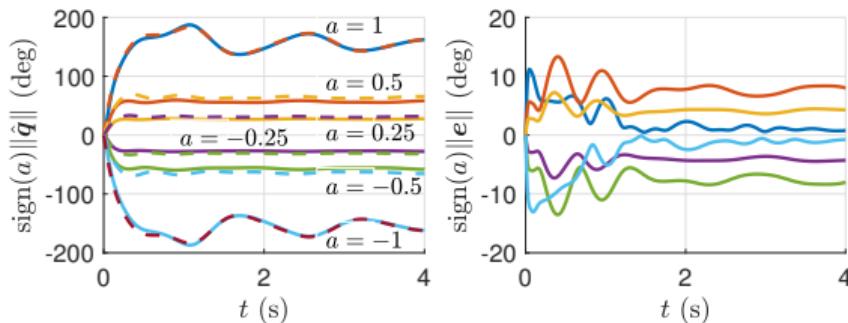
L-GP-based Control: Soft Robotics

Setup:

- FEM of soft robot from [DellaSantina+ 2020]
- Discretization into $N_{\text{FEM}} = 100$ with $k_{\text{FEM}} = 10 \text{ Nm/rad}$ and $d_{\text{FEM}} = 1 \text{ Nms/rad}$

Training:

- $D = 24$ equidistant trajectory samples
- step $a = 1 \text{ Nm}$, noise $\sigma_\epsilon = 0.01a$
- dynamical resimulation of trajectory



Accurate L-GP modeling & control with $4N = 16$ equivalent DOFs

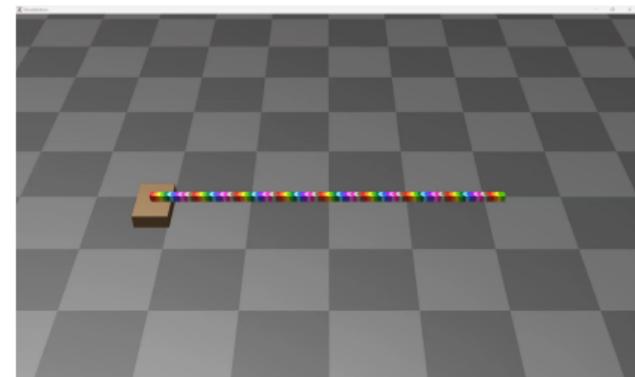
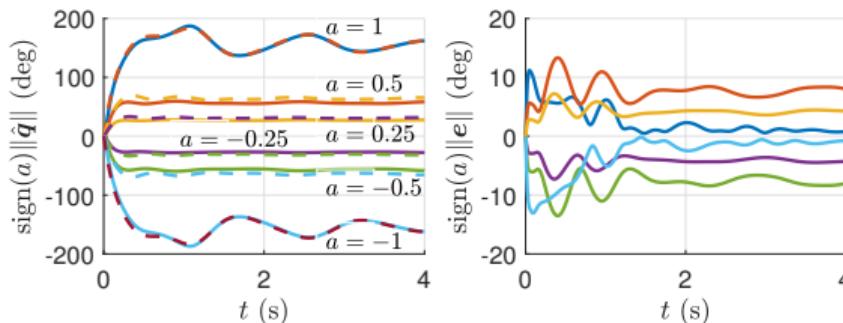
L-GP-based Control: Soft Robotics

Setup:

- FEM of soft robot from [DellaSantina+ 2020]
- Discretization into $N_{\text{FEM}} = 100$ with $k_{\text{FEM}} = 10 \text{ Nm/rad}$ and $d_{\text{FEM}} = 1 \text{ Nms/rad}$

Training:

- $D = 24$ equidistant trajectory samples
- step $a = 1 \text{ Nm}$, noise $\sigma_\epsilon = 0.01a$
- dynamical resimulation of trajectory



Accurate L-GP modeling & control with $4N = 16$ equivalent DOFs

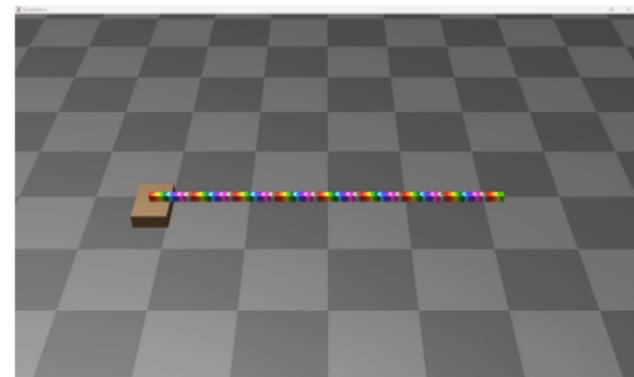
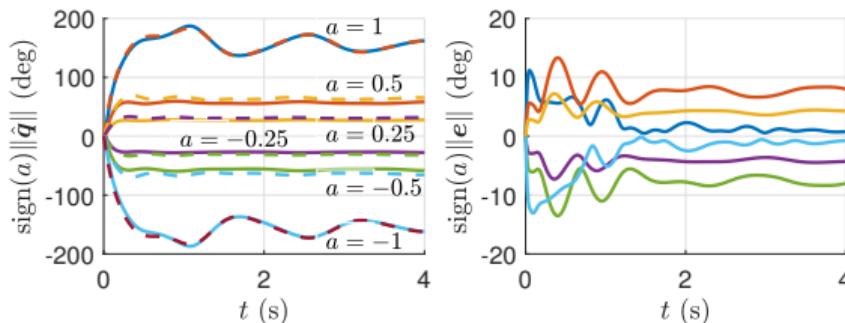
L-GP-based Control: Soft Robotics

Setup:

- FEM of soft robot from [DellaSantina+ 2020]
- Discretization into $N_{\text{FEM}} = 100$ with $k_{\text{FEM}} = 10 \text{ Nm/rad}$ and $d_{\text{FEM}} = 1 \text{ Nms/rad}$

Training:

- $D = 24$ equidistant trajectory samples
- step $a = 1 \text{ Nm}$, noise $\sigma_\epsilon = 0.01a$
- dynamical resimulation of trajectory



Accurate L-GP modeling & control with $4N = 16$ equivalent DOFs

L-GP-based Momentum Observers (L-GP-MOs)

L-GP-MO [Evangelisti+ 2024]

Consider for $\dot{\boldsymbol{p}} = \hat{\boldsymbol{M}}(\boldsymbol{q})\dot{\boldsymbol{q}}$ the L-GP-based momentum observer given by

$$\dot{\hat{\boldsymbol{p}}} = \boldsymbol{\tau}_m - \hat{\boldsymbol{g}} + \hat{\boldsymbol{C}}^T \dot{\boldsymbol{q}} - \hat{\boldsymbol{\tau}}_f + \boldsymbol{r}$$

$$\dot{\boldsymbol{r}} = \boldsymbol{K}_O(\boldsymbol{\Sigma}_{\boldsymbol{\tau}})(\dot{\boldsymbol{p}} - \dot{\hat{\boldsymbol{p}}})$$

Disturbance	$\boldsymbol{\tau}_{\text{ext}} = \mathbf{0}$	$\boldsymbol{\tau}_{\text{ext}}(t)$
Observer	STD (Nm)	STD (Nm)
KUKA	0.7785	0.8986
Parametric MO	0.6330	0.7151
Static L-GP-MO	0.5699	0.6475
Σ -adapt. L-GP-MO	0.4672	0.5697



Giulio Evangelisti and Sandra Hirche. Data-Driven Momentum Observers With Physically Consistent Gaussian Processes.
In: IEEE Transactions on Robotics 40 (2024), pp. 1938–1951. DOI: 10.1109/TR0.2024.3366818.

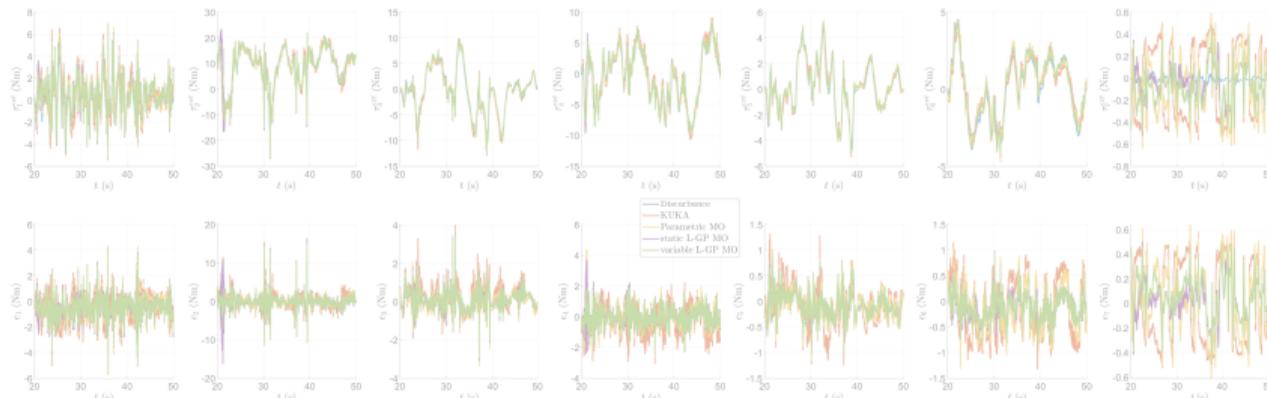
L-GP-based Momentum Observers (L-GP-MOs)

L-GP-MO [Evangelisti+ 2024]

Consider for $\dot{\boldsymbol{p}} = \hat{\boldsymbol{M}}(\boldsymbol{q})\dot{\boldsymbol{q}}$ the L-GP-based momentum observer given by

$$\begin{aligned}\dot{\hat{\boldsymbol{p}}} &= \boldsymbol{\tau}_m - \hat{\boldsymbol{g}} + \hat{\boldsymbol{C}}^T \dot{\boldsymbol{q}} - \hat{\boldsymbol{\tau}}_f + \boldsymbol{r} \\ \dot{\boldsymbol{r}} &= \boldsymbol{K}_O(\boldsymbol{\Sigma}_{\boldsymbol{\tau}})(\dot{\boldsymbol{p}} - \dot{\hat{\boldsymbol{p}}})\end{aligned}$$

Disturbance Observer	$\tau_{\text{ext}} = 0$	$\tau_{\text{ext}}(t)$
	STD (Nm)	STD (Nm)
KUKA	0.7785	0.8986
Parametric MO	0.6330	0.7151
Static L-GP-MO	0.5699	0.6475
Σ -adapt. L-GP-MO	0.4672	0.5697



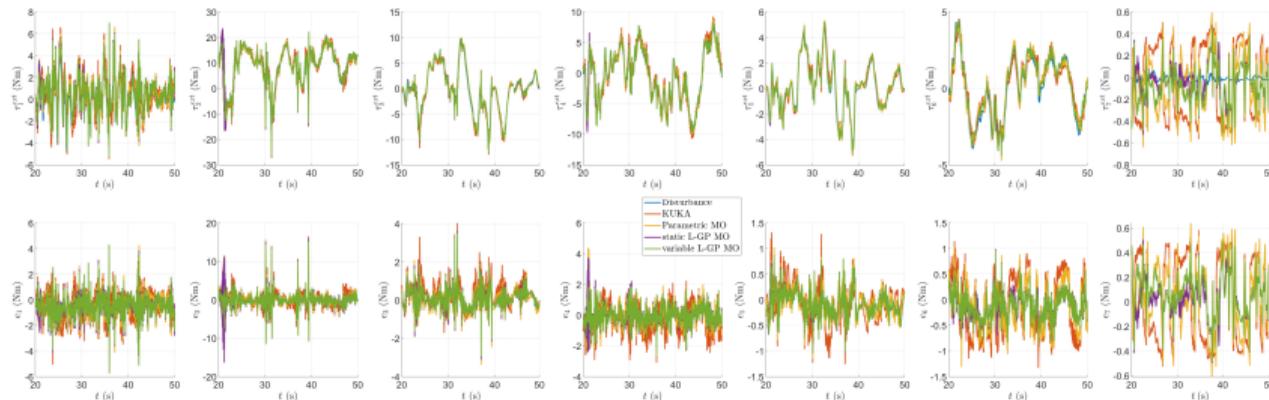
L-GP-based Momentum Observers (L-GP-MOs)

L-GP-MO [Evangelisti+ 2024]

Consider for $\dot{\boldsymbol{p}} = \hat{\boldsymbol{M}}(\boldsymbol{q})\dot{\boldsymbol{q}}$ the L-GP-based momentum observer given by

$$\begin{aligned}\dot{\hat{\boldsymbol{p}}} &= \boldsymbol{\tau}_m - \hat{\boldsymbol{g}} + \hat{\boldsymbol{C}}^T \dot{\boldsymbol{q}} - \hat{\boldsymbol{\tau}}_f + \boldsymbol{r} \\ \dot{\boldsymbol{r}} &= \boldsymbol{K}_O(\boldsymbol{\Sigma}_{\boldsymbol{\tau}})(\dot{\boldsymbol{p}} - \dot{\hat{\boldsymbol{p}}})\end{aligned}$$

Disturbance Observer	$\tau_{\text{ext}} = 0$	$\tau_{\text{ext}}(t)$
	STD (Nm)	STD (Nm)
KUKA	0.7785	0.8986
Parametric MO	0.6330	0.7151
Static L-GP-MO	0.5699	0.6475
Σ -adapt. L-GP-MO	0.4672	0.5697



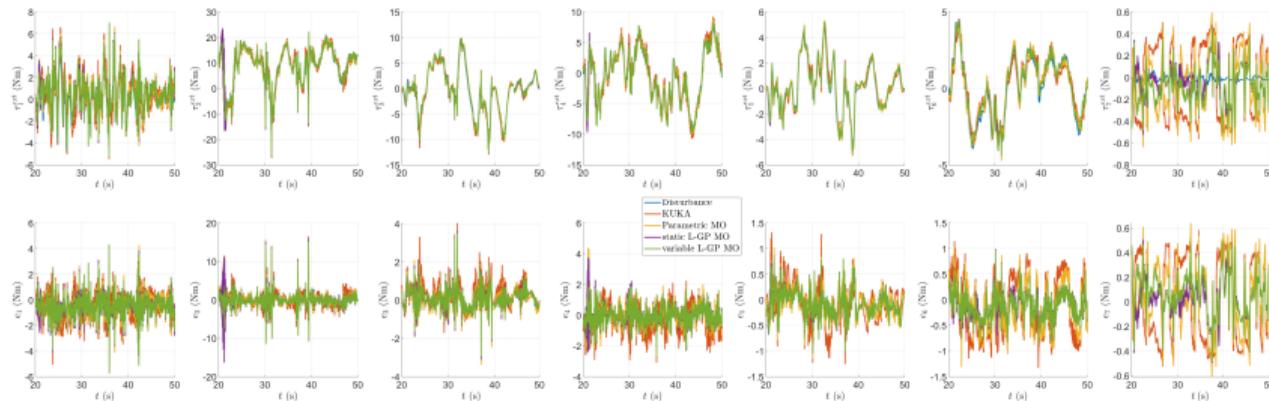
L-GP-based Momentum Observers (L-GP-MOs)

L-GP-MO [Evangelisti+ 2024]

Consider for $\dot{p} = \hat{M}(q)\dot{q}$ the L-GP-based momentum observer given by

$$\begin{aligned}\dot{\hat{p}} &= \tau_m - \hat{g} + \hat{C}^T \dot{q} - \hat{\tau}_f + r \\ \dot{r} &= K_O(\Sigma_\tau)(\dot{p} - \dot{\hat{p}})\end{aligned}$$

Disturbance Observer	$\tau_{\text{ext}} = 0$	$\tau_{\text{ext}}(t)$
STD (Nm)	STD (Nm)	
KUKA	0.7785	0.8986
Parametric MO	0.6330	0.7151
Static L-GP-MO	0.5699	0.6475
Σ -adapt. L-GP-MO	0.4672	0.5697



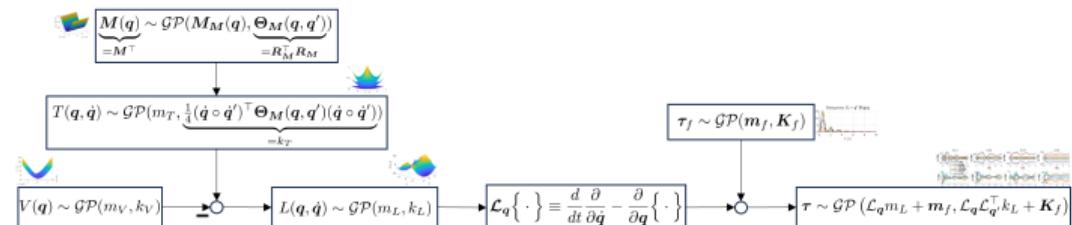
Summary

Lagrangian-Gaussian Process (L-GP):

- enables **physically consistent** identification of uncertain Lagrangian systems
- function space is tailored according to energy components of Lagrangian
- embeds differential equation structure into the model by construction
- probabilistic preservation of positive definiteness

Analytical/deterministic guarantees:

- quadratic forms
- equilibrium
- energy conservation



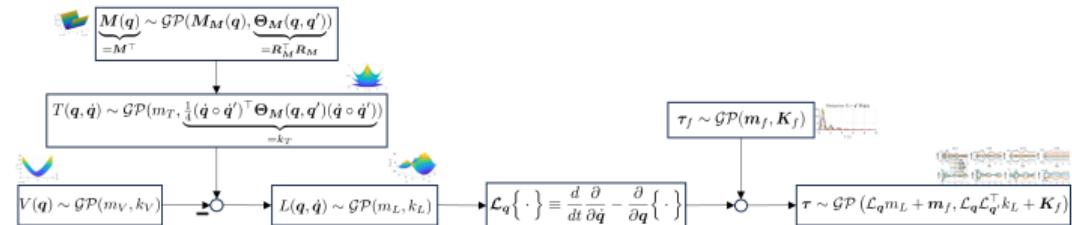
Summary

Lagrangian-Gaussian Process (L-GP):

- enables **physically consistent** identification of uncertain Lagrangian systems
- function space is tailored according to energy components of Lagrangian
- embeds differential equation structure into the model by construction
- probabilistic preservation of positive definiteness

Analytical/deterministic guarantees:

- quadratic forms
- equilibrium
- energy conservation



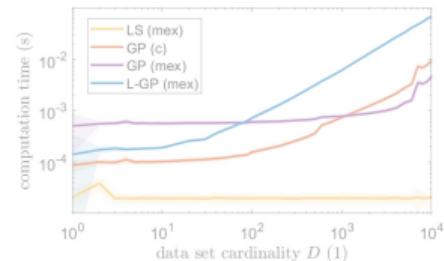
Discussion

Challenges:

- Data efficiency & reliability versus computational complexity
- Only probabilistic preservation of PDF (unavoidable due to desired stochastically consistent GP structure)

Opportunities:

- Uncertainty Quantification applied to DeLaNs
- Learning-based Optimal, Passivity- and Energy-based Control



Efficiency comparison
for $N = 7$

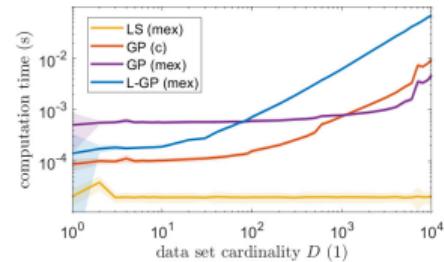
Discussion

Challenges:

- Data efficiency & reliability versus computational complexity
- Only probabilistic preservation of PDF (unavoidable due to desired stochastically consistent GP structure)

Opportunities:

- Uncertainty Quantification applied to DeLaNs
- Learning-based Optimal, Passivity- and Energy-based Control



Efficiency comparison
for $N = 7$

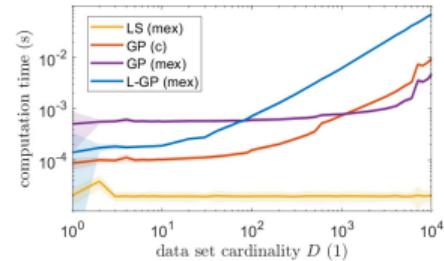
Discussion

Challenges:

- Data efficiency & reliability versus computational complexity
- Only probabilistic preservation of PDF (unavoidable due to desired stochastically consistent GP structure)

Opportunities:

- Uncertainty Quantification applied to DeLaNs
- Learning-based Optimal, Passivity- and Energy-based Control



Efficiency comparison
for $N = 7$

Acknowledgments

This work was supported by:

CO-MAN



Safe data-driven control for
human-centric systems



REHYB



Mauricio A. Álvarez, Lorenzo Rosasco and Neil D. Lawrence. **Kernels for Vector-Valued Functions: A Review.**
In: Foundations and Trends® in Machine Learning 4.3 (2012), pp. 195–266. ISSN: 1935-8237. DOI: 10.1561/2200000036.



Ching-An Cheng and Han-Pang Huang. **Learn the Lagrangian: A Vector-Valued RKHS Approach to Identifying Lagrangian Systems.**
In: IEEE Transactions on Cybernetics (2016).



Cosimo DellaSantina, Robert K Katzschmann, Antonio Bicchi and Daniela Rus.
Model-based dynamic feedback control of a planar soft robot: trajectory tracking and interaction with the environment.
In: The International Journal of Robotics Research 39.4 (2020), pp. 490–513. DOI: 10.1177/0278364919897292.



Giulio Evangelisti and Sandra Hirche. **Data-Driven Momentum Observers With Physically Consistent Gaussian Processes.**
In: IEEE Transactions on Robotics 40 (2024), pp. 1938–1951. DOI: 10.1109/TRO.2024.3366818.



Giulio Evangelisti and Sandra Hirche. **Physically Consistent Learning of Conservative Lagrangian Systems with Gaussian Processes.**
In: 2022 IEEE 61st Conference on Decision and Control (CDC). 2022, pp. 4078–4085. DOI: 10.1109/CDC51059.2022.9993123.



Roy Featherstone. **Rigid Body Dynamics Algorithms.** 1st ed. Springer New York, NY, 2007.



Andreas Rene Geist and Sebastian Trimpe. **Learning Constrained Dynamics with Gauss' Principle adhering Gaussian Processes.**
In: Proceedings of the 2nd Annual Conference on Learning for Dynamics and Control. Vol. 120. PMLR, 2020, pp. 225–234.



Samuel Greydanus, Misko Dzamba and Jason Yosinski. **Hamiltonian Neural Networks.**
In: Advances in Neural Information Processing Systems. Vol. 32. Curran Associates, Inc., 2019.



Noémie Jaquier, Viacheslav Borovitskiy, Andrei Smolensky, Alexander Terenin, Tamim Asfour and Leonel Rozo.
Geometry-aware Bayesian Optimization in Robotics using Riemannian Matérn Kernels.
In: Proceedings of the 5th Conference on Robot Learning. Ed. by Aleksandra Faust, David Hsu and Gerhard Neumann. Vol. 164. Proceedings of Machine Learning Research. PMLR, 2022, pp. 794–805.



M. Lutter, C. Ritter and J. Peters. **Deep Lagrangian Networks: Using Physics as Model Prior for Deep Learning.**
In: 7th International Conference on Learning Representations (ICLR). ICLR, May 2019.



Michael Lutter and Jan Peters. **Combining physics and deep learning to learn continuous-time dynamics models.**

In: The International Journal of Robotics Research 42.3 (2023), pp. 83–107. DOI: 10.1177/02783649231169492.

eprint: <https://doi.org/10.1177/02783649231169492>.



Richard M. Murray, Zexiang Li and S. Shankar Sastry. **A Mathematical Introduction to Robotic Manipulation.** CRC Press, 1994.



Sina Ober-Blobaum and Christian Offen. **Variational learning of Euler–Lagrange dynamics from data.**

In: Journal of Computational and Applied Mathematics 421 (2023), p. 114780. ISSN: 0377-0427.

DOI: <https://doi.org/10.1016/j.cam.2022.114780>.



Romeo Ortega, Antonio Loría, Per Johan Nicklasson and Hebertt Sira-Ramirez. **Passivity-based Control of Euler-Lagrange Systems.** 1st ed.

Communications and Control Engineering. Springer London, 1998.



Maziar Raissi, Paris Perdikaris and George Em Karniadakis. **Inferring solutions of differential equations using noisy multi-fidelity data.**

In: Journal of Computational Physics (2017).



C. E. Rasmussen and C. K. I. Williams. **Gaussian Processes for Machine Learning.** the MIT Press, 2006.



Katharina Rath, Christopher G. Albert, Bernd Bischl and Udo von Toussaint.

Symplectic Gaussian process regression of maps in Hamiltonian systems.

In: Chaos: An Interdisciplinary Journal of Nonlinear Science (2021).



Lucas Rath, Andreas René Geist and Sebastian Trimpe.

Using Physics Knowledge for Learning Rigid-body Forward Dynamics with Gaussian Process Force Priors.

In: Proceedings of the 5th Conference on Robot Learning. Vol. 164. Proceedings of Machine Learning Research. PMLR, 2022, pp. 101–111.



Mark W. Spong, Seth Hutchinson and M. Vidyasagar. **Robot Dynamics and Control.** 2nd Edition. Wiley, 2020.



Patrick M. Wensing, Sangbae Kim and Jean-Jacques E. Slotine.

Linear Matrix Inequalities for Physically Consistent Inertial Parameter Identification: A Statistical Perspective on the Mass Distribution.

In: IEEE Robotics and Automation Letters 3.1 (2018), pp. 60–67. DOI: 10.1109/LRA.2017.2729659.