An Uncertainty-Aware Minimal Intervention Control Strategy Learned from Demonstrations

João Silvério, Yanlong Huang, Leonel Rozo and Darwin G. Caldwell

***Abstract*— Motivated by the desire to have robots physically present in human environments, in recent years we have witnessed an emergence of different approaches for learning active compliance. Some of the most compelling solutions exploit a *minimal intervention control* principle, correcting deviations from a goal only when necessary, and among those who follow this concept, several probabilistic techniques have stood out from the rest. However, these approaches are prone to requiring several task demonstrations for proper gain estimation and to generating unpredictable robot motions in the face of uncertainty. Here we present a Programming by Demonstration approach for uncertainty-aware impedance regulation, aimed at making the robot compliant – and safe to interact with – when the uncertainty about its predicted actions is high. Moreover, we propose a data-efficient strategy, based on the energy observed during demonstrations, to achieve minimal intervention control, when the uncertainty is low. The approach is validated in an experimental scenario, where a human collaboratively moves an object with a 7-DoF torque-controlled robot.**

I. Introduction

Learning variable impedance controllers in the context of Programming by Demonstration (PbD) [1] has been an active topic of research during recent years. In particular, largely influenced by results from the field of motor control [2], probabilistic models have been exploited extensively to endow robots with the ability to efficiently synthesize demonstrated skills [3], [4], [5], [6]. Nonetheless, several aspects still present relevant challenges to the skill transfer problem. On one hand, by exploiting a notion of variance associated with redundancy or variability, most state-of-the-art techniques may require a significant number of demonstrations for proper control gain design, losing applicability in scenarios where few demonstrations are available or where providing them might be cumbersome. On the other hand, such solutions typically do not render the robot capable of properly handling situations where training data is missing or not available, potentially leading to dangerous and unpredictable motions.

Here we offer a novel perspective on this problem by challenging the notion that tracking precision should be a function of redundancy in training data. We instead propose that precision should be intrinsically linked to a robot's confidence about its desired actions. For this, we rely on the predictive power of Gaussian Processes (GP) to obtain estimations of the uncertainty level about a predicted desired state. Moreover, while acknowledging the relevance of the *minimal intervention control* principle proposed by Todorov [2], we provide a data-efficient strategy for control effort regulation based on the energy in the demonstrations. To this end, we exploit an optimal control formulation together with an argument based on the work-energy theorem to support our choice.

In summary, our framework aims for estimating active compliance controllers that provide:

1) uncertainty-aware robot state tracking precision, where the compliance is a function of the uncertainty of predicted actions, and
2) a data-efficient, physically-meaningful strategy that exploits the energy measured during demonstrations to prioritize the correction of the most important directions of a movement when deviations occur.

While the latter point can be seen as addressing the *how to imitate?* problem in PbD, the former pertains to the question of *when to imitate?* [7]. We here posit that the robot should imitate when it is confident about the actions it should perform and otherwise be safe to interact with. Such an intuition is particularly timely given today's high requirements for physical safety in human-robot interaction scenarios.

This paper is comprised of six sections. Section II summarizes related work, Section III describes our proposed approach in detail, while in Section IV we provide experimental results, in both simulation and a real robot, using a human-robot collaboration scenario. Finally, we provide a discussion on the proposed solution, as well as future work directions, in Section V, and close the paper in Section VI with conclusions.

II. Related Work

Motivated by the desire to have robots physically present in human environments, we have recently witnessed an emergence of different approaches for learning active compliance from demonstrations. The trend for exploiting probabilistic approaches in this context dates back a few years, to works such as [8], [9] where heuristics are proposed to set stiffness and damping gains of impedance controllers based on covariance matrices that model the variability in the data. More recently, however, optimal control formulations have started to gain traction, likely due to encouraging results from the field of motor control. According to the *minimal intervention principle*, Todorov [2] suggests that deviations from a desired behavior should be corrected in direct proportion to the amount of disturbance to the overall task performance. In the light of the work of Todorov, Medina *et al.* [3] propose a framework for endowing robots with assistive behaviors,

The authors are with the Department of Advanced Robotics, Istituto Italiano di Tecnologia, 16163 Genova, Italy (e-mail: name.surname@iit.it)

This work was supported by the Italian Ministry of Defense.

where full covariance matrices, retrieved from a Hidden Markov Model (HMM) that compactly represents demonstrations, are exploited as a proxy for the degree of correction along each movement direction. In the same spirit, Calinon *et al.* [4] exploit the variability and correlations in demonstrations, encoded in a Gaussian Mixture Model (GMM), in combination with a task-parameterized formulation, to efficiently adapt skills with variable impedance to new situations. Furthermore, Rozo *et al.* [5] extend the concept to also consider interaction forces into the control problem. One common feature among the aforementioned works is that, due to the specificities of the underlying techniques such as GMM and HMM, variance is equated with variability in demonstrations. Recently, this notion and its implications in practical scenarios have been debated to some extent as different state-of-the-art probabilistic techniques provide complementary notions of variance. In [10], Umlauft *et al.* discuss the differences between variance being interpreted as uncertainty and variability. The topic is also covered in [11], where the different notions are exploited in scenarios that require the combination of different controllers, and in [12], in the context of robot dynamics with multiple additive noise sources. The first contribution of the present work is the exploitation of the notion of variance as *uncertainty* to regulate impedance gains and render the robot compliant when uncertain about its actions. However, in the cases where the uncertainty is low, it still makes sense to follow a minimal intervention approach. Therefore, our second contribution is a strategy, that does not require variability, for selectively correcting deviations in the different degrees of freedom. The strategy exploits the work-energy theorem [15] to establish a connection between the energy that is used in the demonstrations and the control effort for each degree of freedom. Previous works set the effort manually based on the maximum desired amplitude for the commands [4], [5], while [6] set it based on the trade-off between reproduction accuracy and magnitude of the controls. In contrast, this parameter is here learned from the demonstrations and is allowed to vary throughout the task.

III. Proposed Approach

In the spirit of previous works [3], [4] who exploit optimal control techniques to design controllers using human demonstrations, we propose a formulation based on a typical Linear Quadratic Regulator (LQR) [13] to achieve the aforementioned goals (Section III-A). We then learn the parameterization of the LQR problem from demonstrations (Sections III-B and III-C).

A. Problem description

We consider linear systems $\dot{\boldsymbol{\xi}} = \boldsymbol{A}\boldsymbol{\xi} + \boldsymbol{B}\boldsymbol{u}$, where $\boldsymbol{\xi}, \dot{\boldsymbol{\xi}} \in \mathbb{R}^N$ correspond to the system state and its first-order derivative (N denotes the dimension of the state) and $\boldsymbol{u} \in \mathbb{R}^M$ is a control command, where M denotes the number of controlled degrees of freedom. Moreover, $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{B} \in \mathbb{R}^{N\times M}$ represent the state and input matrices. Here, we make the following simplifying assumptions, in line with [4], [5]:

1) We focus our approach on task space control and thus we have $\boldsymbol{\xi} = [\boldsymbol{x}^\top \ \dot{\boldsymbol{x}}^\top]^\top$, where $\boldsymbol{x}, \dot{\boldsymbol{x}} \in \mathbb{R}^3$ represent the end-effector position and linear velocity.
2) We assume a robot with perfect dynamics compensation (i.e. inertia, friction, gravity) and thus model the end-effector as a unit mass.

On the basis of these assumptions, we formulate our LQR problem using a double integrator system, i.e.:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{0}_{3\times3} \\ \boldsymbol{I}_{3\times3} \end{bmatrix}. \tag{1}$$

Moreover, we consider that the end-effector is driven to track a sequence of reference states $\hat{\boldsymbol{\xi}}_t = [\hat{\boldsymbol{x}}_t^\top \ \hat{\dot{\boldsymbol{x}}}_t^\top]^\top$, $t = 1, \dots, T$ by an impedance controller which generates a force in Cartesian space given by $\boldsymbol{F}_t = \boldsymbol{K}_t^{\mathcal{P}}(\hat{\boldsymbol{x}}_t - \boldsymbol{x}_t) + \boldsymbol{K}_t^{\mathcal{V}}(\hat{\dot{\boldsymbol{x}}}_t - \dot{\boldsymbol{x}}_t)$, where $\boldsymbol{K}_t^{\mathcal{P}}$ and $\boldsymbol{K}_t^{\mathcal{V}}$ are time-varying, positive-definite, stiffness and damping gain matrices, respectively. Under the unit mass assumption 2), we have that $\boldsymbol{u}_t = \boldsymbol{F}_t = \ddot{\boldsymbol{x}}_t$, i.e. the control command corresponds to the desired acceleration of the system. We resort to LQR to find an optimal linear state feedback control law

$$\boldsymbol{u}_t = [\boldsymbol{K}_t^{\mathcal{P}} \ \ \boldsymbol{K}_t^{\mathcal{V}}](\hat{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t) \tag{2}$$

that drives the system to track the desired states with time-varying stiffness and damping gain matrices. Such control law is computed as the solution of the quadratic cost function

$$\boldsymbol{c}(t) = \sum_{t=1}^{T} (\hat{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t)^\top \boldsymbol{Q}_t (\hat{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t) + \boldsymbol{u}_t^\top \boldsymbol{R}_t \boldsymbol{u}_t, \tag{3}$$

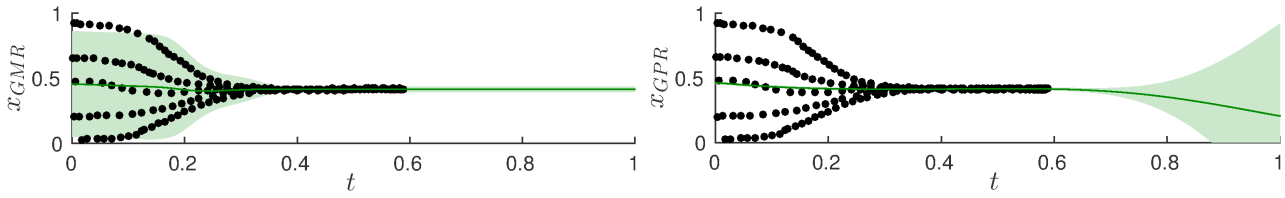
where $\boldsymbol{Q}_t$ is a $N \times N$ positive semi-definite matrix that determines how much the optimization penalizes deviations from the desired state $\hat{\boldsymbol{\xi}}_t$ and $\boldsymbol{R}_t$ is an $M \times M$ positive-definite matrix that penalizes the magnitude of the control commands or, in other words, regulates the control effort. Equation (3) corresponds to the finite horizon LQR formulation and its solution can be obtained through backward integration of the Riccati equations, when matrices $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ are known for the complete time horizon $t = 1, \dots, T$. If this is not the case, one can resort to the infinite horizon formulation

$$\boldsymbol{c}(t) = \sum_{n=t}^{\infty} (\hat{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_n)^\top \boldsymbol{Q}_t (\hat{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_n) + \boldsymbol{u}_n^\top \boldsymbol{R}_t \boldsymbol{u}_n. \tag{4}$$

which is solved iteratively using the algebraic Riccati equation.

As discussed in Section II, previous works set $\boldsymbol{Q}_t$ to the inverse of the predicted covariance matrix that models demonstration data, i.e. $\boldsymbol{Q}_t = \boldsymbol{\Sigma}_t^{-1}$, and manually set the effort regulation term $\boldsymbol{R}_t$. The main novelty in our approach is that we propose to estimate both $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ from demonstrations simultaneously. More specifically:

1) We estimate $\boldsymbol{Q}_t$ based on the uncertainty that the robot has about its actions, given some input. This is aimed at rendering the robot compliant when the uncertainty is high, by allowing for high state errors.



(a) **GMR:** The variance models the variability in the dataset. This occurs regardless of the underlying modeling technique (e.g. GMM, HMM).

(b) **GPR:** The variance models the uncertainty of the estimate (depending on the presence/absence of training datapoints in the neighborhood).

Fig. 1: For a given set of datapoints (black dots), GMR and GPR compute different and complementary notions of variance. The green line is the regressed function, while the light green contour represents the computed variance around the prediction.

2) We exploit $\boldsymbol{R}_t$ to regulate the desired control action in a way that a variation of the minimal intervention principle applied in [3], [4] is followed, when the robot in certain about its actions. In this case, the kinetic energy measured along each degree of freedom (DoF) during the demonstrations is used to regulate the amplitude of the control command of each DoF. Namely, higher/lower commands are allowed when the observed kinetic energy was higher/lower.

To do so, we here rely on the predictive power of Gaussian Processes.

B. Modeling of desired state $\hat{\boldsymbol{\xi}}_t$ and penalty term $\boldsymbol{Q}_t$

Let us consider T demonstrated datapoints, forming tuples $\{\boldsymbol{\zeta}_t^{\mathcal{I}}, \boldsymbol{\zeta}_t^{\mathcal{O}}\}_{t=1}^{T}$, where the indices $\mathcal{I}$, $\mathcal{O}$ denote input and output dimensions. Here the inputs $\boldsymbol{\zeta}_t^{\mathcal{I}}$ can represent any measurable quantities, e.g. time, interaction forces, human/robot states, while typically the outputs are a desired state for the robot, whether a pose in task space or a joint space configuration. A Gaussian Process is a distribution over functions, with a Gaussian prior on outputs $\boldsymbol{\zeta}^{\mathcal{O}}$ given by $\boldsymbol{\zeta}^{\mathcal{O}} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{\Xi}^{\mathcal{I}}), \boldsymbol{K}(\boldsymbol{\Xi}^{\mathcal{I}}, \boldsymbol{\Xi}^{\mathcal{I}}))$, where $\boldsymbol{m}(\boldsymbol{\Xi}^{\mathcal{I}})$ is a vector-valued function yielding the mean of the process, $\boldsymbol{K}(\boldsymbol{\Xi}^{\mathcal{I}}, \boldsymbol{\Xi}^{\mathcal{I}})$ denotes its covariance matrix and $\boldsymbol{\Xi}^{\mathcal{I}} = [\boldsymbol{\zeta}_1^{\mathcal{I}}, \boldsymbol{\zeta}_2^{\mathcal{I}}, \dots \boldsymbol{\zeta}_T^{\mathcal{I}}] \in \mathbb{R}^{D_{\mathcal{I}} \times T}$ is a concatenation of observed $D_{\mathcal{I}}$-dimensional inputs. The covariance matrix is computed from a kernel function $k(\cdot, \cdot)$ evaluated at the inputs, with elements $K_{ij} = k(\boldsymbol{\zeta}_i^{\mathcal{I}}, \boldsymbol{\zeta}_j^{\mathcal{I}})$. Several types of kernel functions exist; see e.g., [14]. Here, we exploit the popular squared-exponential (SE) kernel, typically used to model smooth functions and defined by

$$k(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j) = \epsilon_f^2 \exp\left(\frac{(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j)^\top (\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j)}{l^2}\right), \qquad (5)$$

where ϵ_f^2 and l are hyperparameters that represent the output variance and the input length scale.

Standard Gaussian Process Regression (GPR) allows the prediction of a scalar function $\boldsymbol{\zeta}_*^{\mathcal{O}} = f(\boldsymbol{\zeta}_*^{\mathcal{I}}) : \mathbb{R}^{D_{\mathcal{I}}} \to \mathbb{R}$, for arbitrary inputs $\boldsymbol{\zeta}_*^{\mathcal{I}} \in \mathbb{R}^{D_{\mathcal{I}}}$. In robotics, one typically requires multi-dimensional outputs, thus GPR is often employed separately for each output of a given problem. The prediction of each output dimension $d \in \{1, 2, ..., D_{\mathcal{O}}\}$ is given by

$$\mu_d = \boldsymbol{m}_* + \boldsymbol{k}_*[\boldsymbol{K} + \epsilon_n^2 \boldsymbol{I}]^{-1}(\boldsymbol{\zeta}^{\mathcal{O}_d} - \boldsymbol{m}), \qquad (6)$$

$$\sigma_d^2 = \boldsymbol{k}_{**} - \boldsymbol{k}_*[\boldsymbol{K} + \epsilon_n^2 \boldsymbol{I}]^{-1}\boldsymbol{k}_*^\top, \qquad (7)$$

where $\boldsymbol{\zeta}^{\mathcal{O}_d} \in \mathbb{R}^T$ is the vector of demonstrated outputs for dimension d, $\boldsymbol{k}_* = [k(\boldsymbol{\zeta}_*^{\mathcal{I}}, \boldsymbol{\zeta}_1^{\mathcal{I}}) \dots k(\boldsymbol{\zeta}_*^{\mathcal{I}}, \boldsymbol{\zeta}_T^{\mathcal{I}})]$, $\boldsymbol{k}_{**} = k(\boldsymbol{\zeta}_*^{\mathcal{I}}, \boldsymbol{\zeta}_*^{\mathcal{I}})$, $\boldsymbol{m} = \boldsymbol{m}(\boldsymbol{\Xi}^{\mathcal{I}})$, $\boldsymbol{m}_* = \boldsymbol{m}(\boldsymbol{\zeta}_*^{\mathcal{I}})$, $\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{\Xi}^{\mathcal{I}}, \boldsymbol{\Xi}^{\mathcal{I}})$, and ϵ_n^2 is an additional hyperparameter modeling noise in the observations (which acts as a regularization term). Furthermore, we can concatenate the predictions into one single multivariate Gaussian with mean and covariance matrix given by

$$\boldsymbol{\mu}_{\mathcal{O}} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{D_{\mathcal{O}}} \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\mathcal{O}} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{D_{\mathcal{O}}}^2 \end{bmatrix}. \qquad (8)$$

Here we use GPR to probabilistically model the demonstrated multi-dimensional desired robot states $\hat{\boldsymbol{\xi}}$ and predict them during reproduction, for new observations of the inputs, using (6). To do so, we set the output datapoints to $\boldsymbol{\zeta}_t^{\mathcal{O}} = [\boldsymbol{x}_t^\top \ \ \dot{\boldsymbol{x}}_t^\top]^\top$, a concatenation of demonstrated robot end-effector position and velocity. The definition of inputs $\boldsymbol{\zeta}_t^{\mathcal{I}}$ can however remain general (in the experimental evaluation of Section IV, for example, we make it the human hand position).

While the mean prediction (6) allows for retrieving desired states, the variance prediction (7) plays another important role in our approach. There exists one major difference between the variance predicted by GPR and the one predicted by Gaussian Mixture Regression (GMR), which is exploited in [3], [4]. We illustrate this difference in Fig. 1, where we see that the variance regressed by GMR (shown as an envelope around the mean in Fig. 1(a)) reflects the datapoint distribution in the original dataset or, in other words, the variability in the data. Figure 1(b) shows that the predicted GPR variance represents the uncertainty of the prediction or, in different terms, the absence/presence of input datapoints. On the basis of this observation, and given our goal of regulating the impedance of the robot based on the uncertainty about its desired action, we propose to exploit GPR to define the state penalty matrix $\boldsymbol{Q}_t$ in (3)−(4) as

$$\boldsymbol{Q}_t = \boldsymbol{\Sigma}_{\mathcal{O},t}^{-1} = \begin{bmatrix} \sigma_{1,t}^2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{D_{\mathcal{O}},t}^2 \end{bmatrix}^{-1}, \qquad (9)$$

where $\sigma_{d,t}^2$ denotes the variance of dimension d, predicted at t. Intuitively, based on this design, one can expect the

Fig. 2: Demonstration of the collaborative transportation task, where a bag is cooperatively carried by a human and a robot. Orange arrows indicate the main directions of the task: it starts with a movement along x_3, followed by a movement along x_2.

diagonal elements of $\boldsymbol{Q}_t$ in (9) to be small when the prediction uncertainty is high, penalizing less the deviations from the desired state and rendering the robot compliant. In the opposite way, when the uncertainty is low, the robot should track its desired state in a stiffer manner.

In some cases, one might wish to have control over how compliant the robot is when the uncertainty is high. Such a prior on the level of compliance can be set in our approach by selecting the kernel hyperparameter ϵ_f^2 accordingly. Indeed it can be easily demonstrated from (7) that, when a given input point $\boldsymbol{\zeta}_*^{\mathcal{I}}$ is far from demonstrations, i.e. $\|\boldsymbol{\zeta}_*^{\mathcal{I}} - \boldsymbol{\zeta}_i^{\mathcal{I}}\| \gg 0, \forall i \in \{1, \dots, T\}$, the predicted variance approaches the kernel variance $\sigma_d^2 \to \epsilon_f^2$, i.e., $\boldsymbol{Q}_t \to \text{diag}(\epsilon_f^2, \dots, \epsilon_f^2)^{-1}$, which can be seen as a prior on the state penalty term and, thus, on impedance.

With the proposed strategy for uncertainty-aware impedance regulation, that renders the robot compliant in the face of uncertainty, we now describe our proposed solution for regulating the effort of each degree of freedom, when performing a demonstrated task.

C. Regulating $\boldsymbol{R}_t$ using demonstrated energy profiles

The proposed effort regulation strategy originates from the hypothesis that control commands with higher amplitude must have required a high amount of energy to be generated during demonstrations. Such intuition is backed up by the work-energy theorem – whose importance in human motor control is well attested [15] – which states that the work done by a rigid body equates the variation of kinetic energy during the movement.

Let us consider an arbitrary degree of freedom m. The work done by the end-effector to move along that DoF (starting at rest) with a force F_m by a distance x_m is given by $W_m = F_m x_m$. Moreover, for a displacement that occurred with a linear velocity $\dot{x}_m$, we have a variation of kinetic energy given by $T_m = \frac{1}{2}\dot{x}_m^2$, under the unit mass assumption. According to the work-energy theorem,

$$W_m = T_m \iff F_m x_m = \frac{1}{2}\dot{x}_m^2 \tag{10}$$

must hold. Given that we here consider $u_m = F_m$, from (10) the required control effort scales with the kinetic energy, i.e. $u_m \propto \dot{x}_m^2$. Therefore we propose to regulate the control effort of each DoF as a function of the kinetic energy used in the demonstrations, computed as the square of the observed velocities. In particular we propose

$$\begin{aligned} \boldsymbol{R}_t &= \lambda_R \boldsymbol{E}_t^{-1}, \\ \text{with} \quad \boldsymbol{E}_t &= \text{diag}(\dot{x}_{1,t}^2, \dots, \dot{x}_{M,t}^2) + \lambda_E \boldsymbol{I}, \end{aligned} \tag{11}$$

where λ_R is a hyperparameter that can be adjusted to regulate the overall control effort and λ_E regularizes the energy matrix, avoiding numerical instability when the demonstrated velocities are excessively low. Through the choice of (11), degrees of freedom with high energy result in low values in the corresponding entries of $\boldsymbol{R}_t$ which, in turn, results in higher amplitude control commands. Similarly to the effect of ϵ_f^2 on $\boldsymbol{Q}_t$, λ_E can also act as a prior on $\boldsymbol{R}_t$, allowing for setting the effort penalty and, thus, impedance, when the predicted energy approaches zero.

With the GP encoding proposed in Section III-B, the complete state is predicted during task reproduction, and its velocity component is exploited in the gain estimation through (11). Note that (11) does not require variability to learn input-dependent impedance gains that may differ across degrees of freedom. This stems from our approach being built on GP, inheriting its data-efficiency when computing (6) and (7), in constrast to the approach followed in [3], [4], which requires several demonstrations to compute covariance matrices that explain the data well.

IV. Evaluation

We evaluate the proposed approach in a human-robot collaboration scenario of jointly carrying an object. For this task we used a torque-controlled 7-DoF Barrett WAM robot. The experimental results reported in this section were obtained in MATLAB, by simulating the end-effector using the dynamical system described in Section III. The reader is referred to `http://joaosilverio.weebly.com/2018iros.html` for videos using the real robot.

A. Experimental setup

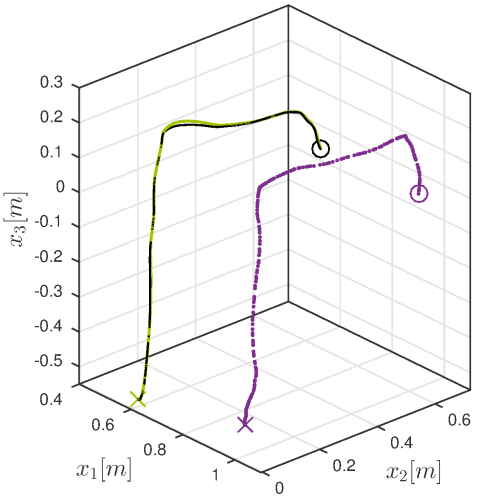
We collected one demonstration of the task of collaboratively carrying a bag from the floor onto a table, as seen in Fig. 2. In this scenario, the human right hand position $\boldsymbol{x}^H \in \mathbb{R}^3$ (with respect to the base of the robot) is tracked using an optical tracking system and the robot end-effector is kinesthetically guided by a demonstrator to fulfill the task. Notice that there are essentially two important directions of movement in this task. In the beginning, the end-effector should move vertically, along x_3, so as to lift the bag. Then,

it should move in the direction of the table which, in this case, is x_2. We chose to use one demonstration in order to highlight that we can achieve uncertainty-aware impedance regulation in a data-efficient manner, unlike typical PbD approaches which rely on covariance information and thus require several demonstrations. Notice that our focus is on the tracking of a desired reference trajectory at the end-effector and thus we intentionally overlook the dynamic aspects of this task such as contact forces or mass compensation. For this reason, we intentionally use a load with low mass.

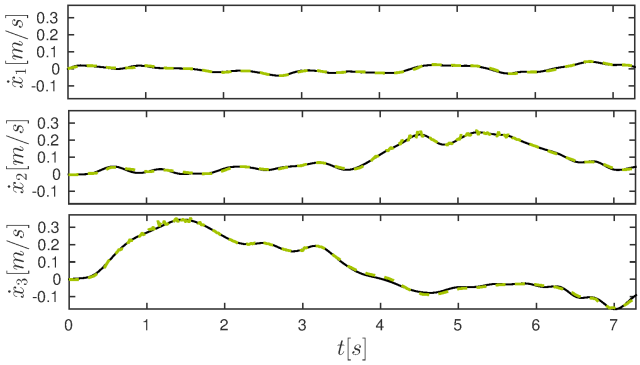
For the GP modeling, we consider the human hand as the input to GPR, i.e., $\boldsymbol{\zeta}_t^{\mathcal{I}} = \boldsymbol{x}_t^H$, $t = 1, \ldots, T$, and the demonstrated states as outputs, as described in Section III. Also, in the experiments reported here we used $\lambda_R = 1 \times 10^{-1}$, $\lambda_E = 1 \times 10^{-4}$, $l = 1 \times 10^{-1} m$, $\epsilon_f^2 = 1 \times 10^{-1}$ and $\epsilon_n^2 = 1 \times 10^{-3}$. As per Section III, control actions in task space are given by $\boldsymbol{F} = \boldsymbol{I}_{3\times3}\ddot{\boldsymbol{x}}$, with $\boldsymbol{I}_{3\times3}$ a 3×3 identity matrix that follows from the unit mass assumption. Control actions are transformed into torque commands through $\boldsymbol{\tau} = \boldsymbol{J}^\top \boldsymbol{F}$, to obtain torque references for each joint, where $\boldsymbol{J}$ is the Jacobian matrix of the end-effector [16]. In a human-robot interaction scenario such as this one, it is not possible to predict in advance the desired state of the robot over a given time horizon, as typically the robot does not know in advance the future state of the human. We therefore exploit here the infinite horizon formulation of LQR (4).

B. Learning variable stiffness and damping

In our first assessment, we used the demonstrated human hand trajectory to predict the desired robot end-effector position throughout one execution of the task. Figures 3(a) and 3(b) show that the generated trajectory closely matches the demonstrated one, both in position and velocity. This is achieved through a proper estimation of stiffness and damping gain matrices, based on the demonstrated energy along each of the three degrees of freedom of the Cartesian space. Figure 4 shows the predicted energy during the execution of the task, along with the corresponding stiffness and damping gains. For convenience of visualization, we plot the results against time, even though the trajectory was generated by taking the human hand position as input. As expected, the estimated gains are proportional to the predicted energy, increasing when the movement has higher energy, and thus validating the chosen energy-based LQR strategy. This is especially evident in the two most important moments of this task: the lifting, when the stiffness along x_3 is higher than that along the remaining directions (peaking at $t \approx 1.5s$), and moving towards the table, where the stiffness is high in the x_2 direction (peaking at $t \approx 5.5s$). In order to showcase this aspect further, we artificially applied two perturbations, simulating a force $\boldsymbol{F}_{pert} = [-50N \; -50N \; -50N]$ applied to the robot end-effector at different instants. Figure 5 shows the response of each degree of freedom after the perturbations are applied. One can observe that the obtained responses are in line with the estimated stiffness profiles from Fig. 4: during the first perturbation, the x_3 direction is practically not affected, with x_2 resulting in a similar observation for



(a) Demonstrated human hand (purple) and end-effector (black) positions. The light green curve represents the robot end-effector trajectory during one reproduction. Initial and final points of each trajectory are denoted by '×' and 'o' respectively.



(b) Linear velocities in operational space: demonstrated (black lines) and observed during one reproduction (dotted green lines).

Fig. 3: Positions and velocities of the end-effector generated using the proposed approach, given demonstrations from a collaborative transportation task where the human hand is used to predict desired robot actions.

the second case. The accompanying video clearly shows this aspect in the real robot.

C. Uncertainty-dependent gains

Subsequently, we tested the capabilities of the proposed framework in rendering the robot compliant when it is uncertain about its actions. Three points, simulating a human hand gradually moving further away from the demonstrations, were used to query the GP for an energy prediction and subsequent stiffness and damping gain matrices. The selected points are $0.1m, 0.2m$ and $0.5m$ away from an arbitrarily chosen demonstrated human hand position, along the $-x_1$ direction: $\boldsymbol{p}_1 = [0.8162 \;\; 0.1053 \;\; -0.1760]$, $\boldsymbol{p}_2 = [0.7162 \;\; 0.1053 \;\; -0.1760]$, $\boldsymbol{p}_3 = [0.4162 \;\; 0.1053 \;\; -0.1760]$. Moving along the $-x_1$ direction corresponds to approaching the base of the robot, i.e., a region where the human should be able to safely interact with the manipulator. The obtained stiffness and damping gains at these points are shown in Fig. 7 for direction x_2 (the results were equivalent in all the remaining directions). We can observe a decrease in both stiffness and damping as the distance to the training

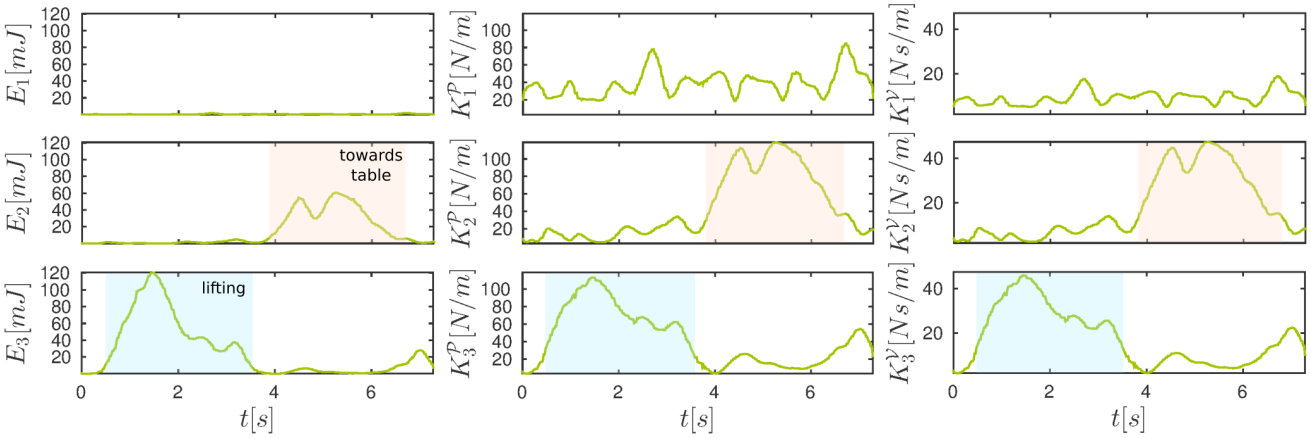


Fig. 4: Estimated energy and consequent impedance gains during one reproduction of the collaborative transportation task, using the proposed approach. **Left:** Energy, **center:** stiffness gains, **right:** damping gains. Each row corresponds to one position degree of freedom in task space.

data increases. Similar observations occur for movements of the human hand along all directions of the operational space, as one can verify in the accompanying video. These results clearly suggest that our approach permits rendering the robot compliant when the inputs to its desired actions differ from those observed during demonstrations.

D. Comparison with GMM

The ability of the proposed approach to render the robot compliant in the face of uncertainty is here tested against the approach based on GMM introduced by Calinon *et al.* [4]. In that framework, we have $\boldsymbol{Q}_t = \boldsymbol{\Sigma}_t^{-1}$, where $\boldsymbol{\Sigma}_t$ is a full covariance matrix, predicted by GMR. Moreover, the penalty term on the control commands $\boldsymbol{R}_t$ is set to a constant value. In order to model the collaborative transportation task using GMM, further demonstrations are required and thus we added 4 more to the dataset. We trained a model with 4 Gaussian components, chosen empirically, yielding the representation of the robot end-effector position shown in Fig. 6. The previously selected points $\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3$ were used again to compute the impedance gains, in regions where demonstrations are not present. Using this approach we obtained the stiffness and damping gains shown in Fig. 7 (green columns). Once more, we here show the results only for direction x_2, but the same trends were present in all directions. We observe that the estimated stiffness and damping gains are consistently high, regardless of the distance to the region where demonstrations were provided. This follows from the fact that covariance matrices model variability and correlation among state variables, unlike variance in GPR predictions which, as we discussed in Section III, models the uncertainty.

V. Discussion and Future Work

Despite that Section IV showed the merits of the proposed approach in a realistic scenario, some points deserve a more detailed discussion. Firstly, throughout the paper we have referred to the distance to the demonstrations in a rather qualitative sense. However, it should be noted that the notion of distance can be regulated through the length scale of the SE kernel l. Indeed, by increasing/decreasing l, one

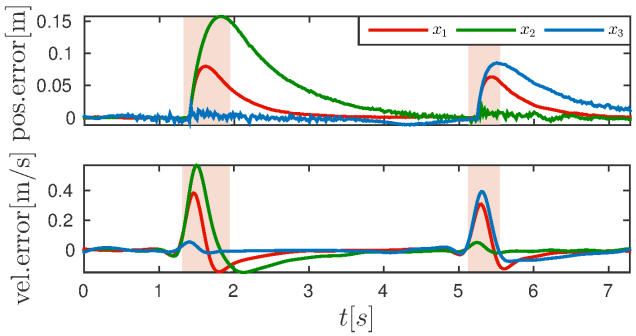


Fig. 5: Effect of perturbations on the end-effector trajectory, measured by position and velocity tracking errors. Shaded areas highlight the moments when perturbations occur. Notice the different responses along each degree of freedom, which follow from the learned impedance profiles.

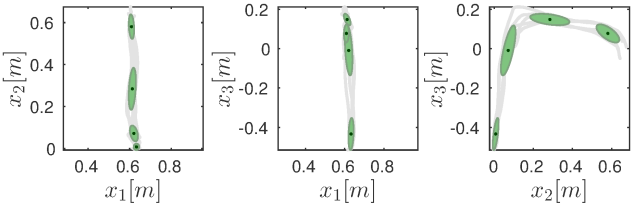


Fig. 6: Five demonstrations of the collaborative transportation task, modeled by a GMM with 4 states. The gray points correspond to the end-effector position during the demonstrations while the green ellipses depict Gaussian distributions, plotted to a width of two standard deviations.

can make the robot exhibit the default compliant behavior further away/closer to the demonstrations. Note, however, that setting this hyperparameter to a too small value, might result in good tracking only when the input coincides exactly with the demonstrations, which might be hard to achieve in physical human-robot interaction scenarios since humans may exhibit some degree of variation in their actions.

Secondly, our strategy for minimal intervention control does not consider correlated movements between DoFs. It is a well-known fact that Gaussian Processes – the backbone of our framework – are homoscedastic, i.e. they do not encode input-dependent covariance. Works such as [10], [17], which propose formulations for regressing both uncertainty and correlation, may provide the possibility to simultaneously consider both aspects into our approach, which can potentially allow us to study the complementarity of the two

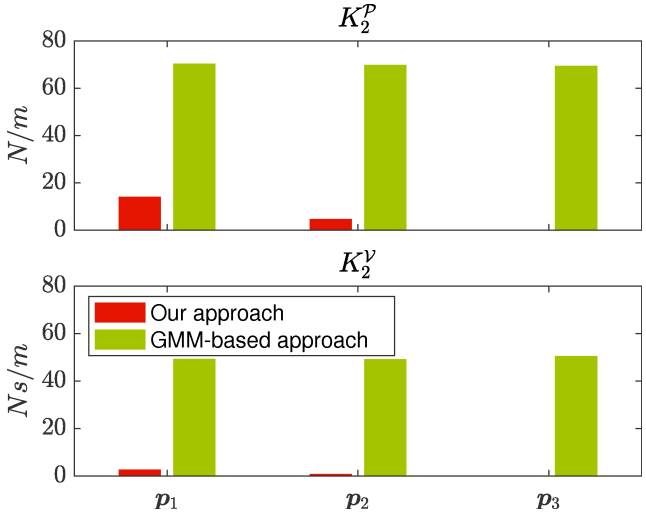


Fig. 7: Stiffness and damping gains along x_2, estimated for three points at increasing distances from the training data. The red columns show the results using our approach while the green ones correspond to the GMM-based approach from [4]. The decreasing trend using our approach is present in all other directions, despite not being plotted here.

notions of minimal intervention control.

Thirdly, it is important to highlight another aspect where the strategy presented in Section III-C intrinsically differs from that exploited in [3], [4], [5], [6]. With our approach, since impedance gains do not depend on the consistency in demonstrations, it may occur that the robot is too compliant in directions where the demonstrator intended it to exhibit high stiffness by demonstrating small variability. Circumventing this problem can be easily handled by increasing the parameter λ_E. However this comes at the cost of increasing the gains of all degrees of freedom, which, depending on the scenario, may pose stability problems. It should be noted that similar cases can also occur in the aforementioned techniques. Namely, one might inadvertently exhibit low variability in degrees of freedom where, in practice, high compliance may be desired.

Finally, it should also be noted that exploiting Gaussian Processes does not constrain the experimenter to use one single demonstration, as we did in Section IV. Providing more demonstrations can improve the generalization capabilities of the approach. However, since the computational cost of GPR increases with the number of training datapoints, one should be aware of the trade-off between the number of demonstrations and the computational load.

In future work we plan to study formulations of the problem in joint space, where energy is directly linked to physical properties of the real system such as motor power and link mass. Moreover, we understand there is potential of application in the context of task-parameterized skill learning [4], which may be exploited with the aim of increasing the extrapolation capabilities of the proposed framework.

VI. Conclusions

In this paper we proposed an approach for learning variable impedance gains from demonstrations while considering the uncertainty of the actions predicted by the robot. This was achieved through a combination of optimal control and Gaussian Process Regression. In particular, we formulated the gain estimation as a typical LQR problem, with parameterization obtained from a Gaussian Process that serves as a prior on demonstrations. The proposed framework was validated in a human-robot collaborative transportation scenario, in which we saw that the approach effectively rendered the robot compliant when the human was outside the region where demonstrations were provided. We also verified that the robot was able to adapt its stiffness and damping gains when the uncertainty was low, following a minimum intervention strategy to control its trajectory by tracking the reference state with higher gains along the most important directions of the movement.

References

[1] A. G. Billard, S. Calinon, and R. Dillmann, "Learning from humans," in *Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Secaucus, NJ, USA: Springer, 2016, ch. 74, pp. 1995–2014, 2nd Edition.

[2] E. Todorov, "Optimality principles in sensorimotor control," *Nature Neuroscience*, vol. 7, no. 9, pp. 907–915, 2004.

[3] J. R. Medina, D. Lee, and S. Hirche, "Risk-sensitive optimal feedback control for haptic assistance," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, May 2012, pp. 1025–1031.

[4] S. Calinon, D. Bruno, and D. G. Caldwell, "A task-parameterized probabilistic model with minimal intervention control," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, Hong Kong, China, May-June 2014, pp. 3339–3344.

[5] L. Rozo, D. Bruno, S. Calinon, and D. G. Caldwell, "Learning optimal controllers in human-robot cooperative transportation tasks with position and force constraints," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, Sept.-Oct. 2015, pp. 1024–1030.

[6] M. Zeestraten, S. Calinon, and D. G. Caldwell, "Variable duration movement encoding with minimal intervention control," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, Stockholm, Sweden, May 2016, pp. 497–503.

[7] C. Nehaniv and K. Dautenhahn, *Imitation in Animals and Artifacts*. Boston, USA: MIT Press, 2002.

[8] K. Kronander and A. Billard, "Learning compliant manipulation through kinesthetic and tactile human-robot interaction," *IEEE Transactions on Haptics*, vol. 7, no. 3, pp. 367–380, 2014.

[9] A. X. Lee, H. Lu, A. Gupta, S. Levine, and P. Abbeel, "Learning force-based manipulation of deformable objects from multiple demonstrations," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, Seattle, Washington, USA, May 2015, pp. 177–184.

[10] J. Umlauft, Y. Fanger, and S. Hirche, "Bayesian uncertainty modeling for Programming by Demonstration," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, Singapore, May–June 2017, pp. 6428–6434.

[11] J. Silvério, Y. Huang, L. Rozo, S. Calinon, and D. G. Caldwell, "Probabilistic learning of torque controllers from kinematic and force constraints," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, Madrid, Spain, October 2018.

[12] J. R. Medina and S. Hirche, "Uncertainty-dependent optimal control for robot control considering high-order cost statistics," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, September–October 2015, pp. 3995–4002.

[13] E. Todorov, "Optimal control theory," *Bayesian brain: probabilistic approaches to neural coding*, pp. 269–298, 2006.

[14] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Cambridge, MA, USA: MIT Press, 2006.

[15] J. Hamill, K. Knutzen, and T. Derrick, *Biomechanical Basis of Human Movement*. Wolters Kluwer Health, 2014.

[16] O. Khatib, "A unified approach for motion and force control of robot manipulators: The operational space formulation," *IEEE Journal on Robotics and Automation*, vol. 3, no. 1, pp. 43–53, 1987.

[17] A. G. Wilson and Z. Ghahramani, "Generalised Wishart Processes," in *Proc. of the 27th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 2011.