Learning Optimal Controllers in Human-robot Cooperative Transportation Tasks with Position and Force Constraints

Leonel Rozo[1], Danilo Bruno[1], Sylvain Calinon[1,2] and Darwin G. Caldwell[1]

Abstract— **Human-robot collaboration seeks to have humans and robots closely interacting in everyday situations. For some tasks, physical contact between the user and the robot may occur, originating significant challenges at safety, cognition, perception and control levels, among others. This paper focuses on robot motion adaptation to parameters of a collaborative task, extraction of the desired robot behavior, and variable impedance control for human-safe interaction. We propose to teach a robot cooperative behaviors from demonstrations, which are probabilistically encoded by a task-parametrized formulation of a Gaussian mixture model. Such encoding is later used for specifying both the desired state of the robot, and an optimal feedback control law that exploits the variability in position, velocity and force spaces observed during the demonstrations. The whole framework allows the robot to modify its movements as a function of parameters of the task, while showing different impedance behaviors. Tests were successfully carried out in a scenario where a 7 DOF backdrivable manipulator learns to cooperate with a human to transport an object.**

I. INTRODUCTION

The robots' role in our daily life is becoming more prominent as robots are getting safer, more user-friendly and versatile. This allows robots to share and populate human environments such as hospitals, houses, factories, etc. In these places, the robot is aimed at assisting or collaborating with people to facilitate, improve and/or speed up specific tasks that usually are carried out by a group of humans exclusively. In this context, some of the robot duties can involve physical contact, for example, in hand-over tasks [1], or when a robot cooperatively carries an object with a human partner [2]. This physical interaction provokes a rich exchange of haptic information, involves compliant robot movements [3], and sometimes requires the robot to follow a desired trajectory [4]. These aspects are not straightforward to program, and so is the inclusion of all the possible variations the robot might face. Additionally, the type of jobs carried out by the robot may frequently vary. Therefore, a robotic assistant is required to be easily and rapidly re-programmed several times according to specific needs.

Programming by demonstration (PbD) [5], [6] emerges as a promising alternative solution allowing the natural and intuitive transfer of human knowledge about a task to a collaborative robot. In this paper we exploit PbD in human-robot collaboration (HRC) by tackling four problems: *(i)* the encoding of human demonstrations that vary according to parameters of the task (e.g., objects to grasp, location of obstacles in the workspace, etc), *(ii)* the extraction of the desired state of the robot during the reproduction of the collaborative behavior, *(iii)* the estimation of different compliance levels over time, and *(iv)* a safe interaction with the human. Thus, our aim is to provide a PbD framework that allows the robot to adapt both its impedance and desired state according to the task constraints, while safely cooperating with users.

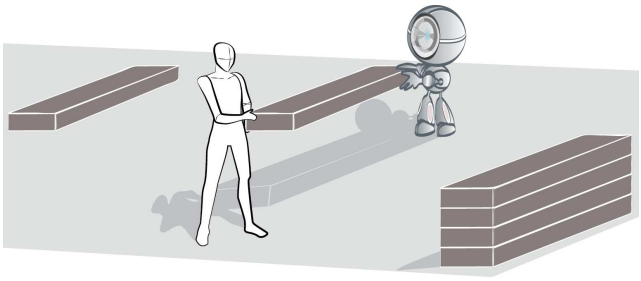


Fig. 1: Illustration of a human-robot cooperative transportation task. Several bulky planks need to be carried to a specific location. The height at which each plank is placed varies as more planks are placed at the final position, which is a parameter influencing parts of the collaborative task.

Specifically, the proposed approach encodes the human demonstrations with a task-parametrized formulation of the Gaussian mixture model [7], that allows the robot to shape its behavior or motion as function of parameters of the task. The novelties are threefold. We first propose a principled approach to handle robot states composed of position, velocity and force, in human-robot collaborative tasks. Secondly, our approach exploits the statistical representation of task constraints in multiple coordinate systems within an optimal control strategy to automatically estimate Cartesian impedance gains as full matrices of the controller governing the cooperative robot behavior. These time-varying gain matrices likewise allow the robot to automatically prioritize to specific parts of the task, or to more precisely compensate for errors of particular desired state variables. Thirdly, the proposed optimal controller permits to minimize not only the robot effort, but also the human intervention along the task.

We successfully test our approach in a real-world scenario where a 7 DOF robotic manipulator learns to perform a cooperative manipulation task requiring different constraints to be satisfied, while behaving with different levels of compliance.

[1]Department of Advanced Robotics, Istituto Italiano di Tecnologia (IIT), Via Morego 30, 16163 Genova, Italy. name.surname@iit.it

[2]Idiap Research Institute, Rue Marconi 19, PO Box 592, CH-1920 Martigny, Switzerland. sylvain.calinon@idiap.ch

This work was supported by the European projects STIFF-FLOP (FP7-ICT-287728) and DexROV (H2020-EU.3.2-635491).

The rest of the paper is organized as follows: Section II reviews works related to our problem, while Sections III, IV and V respectively present the interaction model of the robot, the learning framework, and the optimal estimation of the controller parameters. The experimental setting and results are introduced in Section VI. Finally, conclusions and future routes of research are given in Section VII.

II. RELATED WORK

Recently, research on PbD for human-robot collaborative scenarios has gained increased interest in the Robotics community. In this field, new challenges need to be solved, such as dealing with novel types of information (e.g., haptic perception, human gaze, etc.), role allocation, intention prediction, among others. For instance, in [8] a probabilistic framework based on Gaussian mixture models (GMM) and Gaussian mixture regression (GMR) was proposed to respectively encode and reproduce robot collaborative behaviors. Demonstrations of leader/follower roles during a cooperative lifting task were provided by teleoperation. GMM encapsulated the robot motion and the sensed forces, while GMR generated the reference inputs corresponding to a given sensed force during reproduction. The same task was studied by Gribovskaya *et al.* [9], who proposed a hybrid structure based on PbD and adaptive control that drives the robot using an adaptive impedance controller. First, a feed-forward model of the task was learned from demonstrations encoded by a GMM. Then, the impedance parameters were adapted as function of the kinematic and force errors generated during the execution of the task.

Medina *et al.* [10] introduced a cognitive system with segmentation, encoding and clustering capabilities for demonstrations of collaborative behavioral primitives. These were represented by a primitive graph and a primitive tree using hidden Markov models that were incrementally updated during reproduction. One of the main differences with respect to [8] is that the robot started its behavior as a follower, with a role progressively becoming more proactive as it acquired more knowledge about the task. In [11], dynamic movement primitives (DMP) were used for driving the robot motion in cooperative tasks. The DMP depended not only on a given reference to follow, but also on an obstacle avoidance force and an interaction term. The latter was learned so that the interaction forces were minimized.

Later on, Ben Amor *et al.* [12] proposed a probabilistic encoding of the DMP parameters that allowed for adaption and correlation of the robot motion based on predictions of the human intention from partial observations. Their formulation used dynamic time warping for shaping the future robot actions according to the partner actions. DMPs have also been used in human-in-the-loop robot learning [13]. Such approach considered an online learning strategy where the human tutor taught a cyclic motion and different compliance levels through teleoperation. The former was obtained with motion capture systems, while the latter was computed from electromyography signals.

PbD and risk-sensitive optimal control have also been combined for designing robotic assistants [14]. The idea is to predict the human motion and accordingly set the trajectory reference of a risk-sensitive controller. Using this framework, the robot minimizes the human contribution along the task while slightly adapting to unexpected behaviors of the user.

These works were mainly focused on learning either the robot's role or an adaptive varying impedance or movement. In contrast, we here present an approach that not only permits to extract the position, velocity and force constraints of the task from kinesthetic demonstrations, but also to shape the robot motion as a function of task parameters, therefore extending our previous work [15] where only position feedback was considered and the controller gains were manually set. The proposed model is combined with an optimal controller that exploits the variability observed in the demonstrations to continuously adapt a feedback control law. In other words, the robot compliance level is updated according to the precision that is required to track its desired state over time. Note that our approach also differs from [14] in that we consider the force perceptions as an additional task constraint, which becomes relevant when specific force profiles are required for manipulating particular objects, which consequently leads us to a new optimal control formulation. In [14] the force perceptions are instead considered as an independent noise input to the system due to uncertain predictions of the model, and moreover the robot was not able to adapt its behavior according to varying task parameters.

III. INTERACTION MODEL

To formalize the problem, the joint space dynamics model of the robot under interaction with the environment (e.g., the human partner) is defined as

$$\boldsymbol{H}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) = \boldsymbol{\tau} + \boldsymbol{J}(\boldsymbol{q})^{\top}\boldsymbol{f}, \qquad (1)$$

where $\boldsymbol{H}(\boldsymbol{q})$, $\boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})$ and $\boldsymbol{g}(\boldsymbol{q})$ are the inertia matrix, the vector of centrifugal and Coriolis forces, and the gravity components, respectively. The pose of the robot in joint space is denoted by $\boldsymbol{q}$, $\boldsymbol{\tau}$ is the actuation torques vector, $\boldsymbol{J}(\boldsymbol{q})$ is the Jacobian of the robot, and $\boldsymbol{f}$ is the vector of external forces applied to the end-effector, that can be obtained by a sensor on the robot's tool. Also, let us define the joint space controller $\boldsymbol{\tau}$ as

$$\boldsymbol{\tau} = \boldsymbol{J}(\boldsymbol{q})^{\top}\boldsymbol{\Lambda}(\boldsymbol{q})\boldsymbol{u} + \boldsymbol{\tau}_g, \qquad (2)$$

where $\boldsymbol{u}$ represents a desired control acceleration at the robot's end-effector, $\boldsymbol{\Lambda}(\boldsymbol{q}) = \left(\boldsymbol{J}(\boldsymbol{q})\boldsymbol{H}(\boldsymbol{q})^{-1}\boldsymbol{J}(\boldsymbol{q})^{\top}\right)^{-1}$ is the Cartesian inertia matrix, and $\boldsymbol{\tau}_g$ the torque commands to compensate for the effect of gravity.

During a collaborative task, constraints at position, velocity and force may arise. So, in order for the robot to fulfill such constraints, we propose a controller $\boldsymbol{u} = \boldsymbol{u}_m + \boldsymbol{u}_f$, where $\boldsymbol{u}_m$ and $\boldsymbol{u}_f$ respectively represent motion and force control commands. These controllers compensate for motion and force feedback errors, and are defined as

$$\boldsymbol{u}_m = \boldsymbol{K}^{\mathcal{P}}\left(\bar{\boldsymbol{x}} - \boldsymbol{x}\right) + \boldsymbol{K}^{\mathcal{V}}\left(\bar{\dot{\boldsymbol{x}}} - \dot{\boldsymbol{x}}\right), \qquad (3)$$

$$\boldsymbol{u}_f = \boldsymbol{K}^{\mathcal{F}}\left(\bar{\boldsymbol{f}} - \boldsymbol{f}\right), \qquad (4)$$

where the matrices $\boldsymbol{K}^{\mathcal{P}}$, $\boldsymbol{K}^{\mathcal{V}}$ and $\boldsymbol{K}^{\mathcal{F}}$ are full stiffness, damping and force gain matrices, respectively. In addition, $\bar{\boldsymbol{x}}$, $\bar{\dot{\boldsymbol{x}}}$ and $\bar{\boldsymbol{f}}$ are the reference or desired Cartesian position, velocity and sensed force, that can be obtained from human demonstrations of the desired collaborative behavior (see Section IV). From the definition of $\boldsymbol{u}$, we can reorganize the whole controller in a matrix notation as

$$\boldsymbol{u} = -\begin{bmatrix} \boldsymbol{K}^{\mathcal{P}} & \boldsymbol{K}^{\mathcal{V}} & \boldsymbol{K}^{\mathcal{F}} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\dot{\boldsymbol{x}}} \\ \tilde{\boldsymbol{f}} \end{bmatrix}, \tag{5}$$

where $\tilde{\boldsymbol{x}} = (\boldsymbol{x} - \bar{\boldsymbol{x}})$, $\tilde{\dot{\boldsymbol{x}}} = (\dot{\boldsymbol{x}} - \bar{\dot{\boldsymbol{x}}})$ and $\tilde{\boldsymbol{f}} = \big(\boldsymbol{f} - \bar{\boldsymbol{f}}\big)$. Such an expression shares similarities with the feedback term of a linear quadratic regulator (LQR), where the controller is expressed as a proportional gain multiplying the error of the system state (in other words, a state-feedback controller, see [16]).

IV. TASK LEARNING WITH TP-GMM

The robot collaborative behavior is learned from human demonstrations, which are encoded with a *task-parametrized* version of the Gaussian mixture model (TP-GMM) [7]. This model allows us to consider task constraints in different frames of reference (i.e., the parameters of the task), which is particularly advantageous when the robot behavior is conditioned by, for example, position of objects or users, changes in the environment and changes of configurations of another robot parts. Formally, the task parameters are represented as P coordinate systems, defined at time step t by $\{\boldsymbol{b}_{t,j}, \boldsymbol{A}_{t,j}\}_{j=1}^{P}$, representing respectively the origin of the observer and a set of basis vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots\}$ forming a transformation matrix $\boldsymbol{A} = [\boldsymbol{e}_1 \boldsymbol{e}_2 \cdots]$.

A demonstration $\boldsymbol{\xi} \in \mathbb{R}^{D \times T}$ is observed from these different frames, forming a third order tensor dataset $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{D \times T \times P}$, composed of P trajectory samples $\boldsymbol{X}^{(j)} \in \mathbb{R}^{D \times T}$ observed in P candidate frames, corresponding to matrices composed of D-dimensional observations at T time steps. The parameters of the proposed TP-GMM with K components are defined by $\{\pi_i, \{\boldsymbol{\mu}_i^{(j)}, \boldsymbol{\Sigma}_i^{(j)}\}_{j=1}^{P}\}_{i=1}^{K}$ (π_i are the mixing coefficients, $\boldsymbol{\mu}_i^{(j)}$ and $\boldsymbol{\Sigma}_i^{(j)}$ are the center and covariance matrix of the i-th Gaussian component in frame j).

Learning of the parameters is achieved with the constrained problem of maximizing the log-likelihood under the constraints that the data in the different frames are generated from the same source, resulting in an expectation-maximization (EM) process to iteratively update the model parameters until convergence [7]. The model parameters are initialized with a *k-means* procedure redefined using a similar process to that used for the modified EM algorithm. The learned model can further be used to reproduce movements in other situations (for new positions and orientations of candidate frames). The model first retrieves at each time step t a GMM by computing a product of linearly transformed Gaussians

$$\mathcal{N}(\boldsymbol{\mu}_{t,i}, \boldsymbol{\Sigma}_{t,i}) \propto \prod_{j=1}^{P} \mathcal{N}\Big(\boldsymbol{A}_{t,j}\boldsymbol{\mu}_i^{(j)} + \boldsymbol{b}_{t,j},\ \boldsymbol{A}_{t,j}\boldsymbol{\Sigma}_i^{(j)}\boldsymbol{A}_{t,j}^{\top}\Big), \tag{6}$$

$$\boldsymbol{\Sigma}_{t,i} = \Big(\sum_{j=1}^{P} (\boldsymbol{A}_{t,j}\boldsymbol{\Sigma}_i^{(j)}\boldsymbol{A}_{t,j}^{\top})^{-1}\Big)^{-1}, \tag{7}$$

$$\boldsymbol{\mu}_{t,i} = \boldsymbol{\Sigma}_{t,i}\sum_{j=1}^{P} (\boldsymbol{A}_{t,j}\boldsymbol{\Sigma}_i^{(j)}\boldsymbol{A}_{t,j}^{\top})^{-1}(\boldsymbol{A}_{t,j}\boldsymbol{\mu}_i^{(j)} + \boldsymbol{b}_{t,j}). \tag{8}$$

With the temporary GMM representation computed in Eq. (6), a reference movement or an average collaborative behavior can be estimated as a regression problem [17]. We define the superscripts $\mathcal{I}$ and $\mathcal{O}$ as the sets of dimensions that span for input and output variables (that will be used as exponents in vectors and matrices). At each iteration step t, the datapoint $\boldsymbol{\xi}_t$ can be decomposed as two subvectors $\boldsymbol{\xi}_t^{\mathcal{I}}$ and $\boldsymbol{\xi}_t^{\mathcal{O}}$ spanning for the input and output variables, respectively. With this notation, a block decomposition of the datapoint $\boldsymbol{\xi}_t$, vectors $\boldsymbol{\mu}_{t,i}$ and matrices $\boldsymbol{\Sigma}_{t,i}$ can be written as

$$\boldsymbol{\xi}_t = \begin{bmatrix} \boldsymbol{\xi}_t^{\mathcal{I}} \\ \boldsymbol{\xi}_t^{\mathcal{O}} \end{bmatrix}, \quad \boldsymbol{\mu}_{t,i} = \begin{bmatrix} \boldsymbol{\mu}_{t,i}^{\mathcal{I}} \\ \boldsymbol{\mu}_{t,i}^{\mathcal{O}} \end{bmatrix}, \quad \boldsymbol{\Sigma}_{t,i} = \begin{bmatrix} \boldsymbol{\Sigma}_{t,i}^{\mathcal{I}} & \boldsymbol{\Sigma}_{t,i}^{\mathcal{I}\mathcal{O}} \\ \boldsymbol{\Sigma}_{t,i}^{\mathcal{O}\mathcal{I}} & \boldsymbol{\Sigma}_{t,i}^{\mathcal{O}} \end{bmatrix}. \tag{9}$$

Given the temporary GMM that encodes the joint distribution $\mathcal{P}(\boldsymbol{\xi}_t^{\mathcal{I}}, \boldsymbol{\xi}_t^{\mathcal{O}}) \sim \sum_{i=1}^{K} \pi_i \mathcal{N}(\boldsymbol{\mu}_{t,i}, \boldsymbol{\Sigma}_{t,i})$ of the dataset $\boldsymbol{\xi}$, at each reproduction step t, $\mathcal{P}(\boldsymbol{\xi}_t^{\mathcal{O}} | \boldsymbol{\xi}_t^{\mathcal{I}})$ is computed as the conditional distribution

$$\mathcal{P}(\boldsymbol{\xi}_t^{\mathcal{O}} | \boldsymbol{\xi}_t^{\mathcal{I}}) \sim \sum_{i=1}^{K} \gamma_i(\boldsymbol{\xi}_t^{\mathcal{I}})\, \mathcal{N}\Big(\hat{\boldsymbol{\mu}}_{t,i}^{\mathcal{O}}(\boldsymbol{\xi}_t^{\mathcal{I}}), \hat{\boldsymbol{\Sigma}}_{t,i}^{\mathcal{O}}\Big), \tag{10}$$

$$\text{with} \quad \hat{\boldsymbol{\mu}}_{t,i}^{\mathcal{O}}(\boldsymbol{\xi}_t^{\mathcal{I}}) = \boldsymbol{\mu}_{t,i}^{\mathcal{O}} + \boldsymbol{\Sigma}_{t,i}^{\mathcal{O}\mathcal{I}}{\boldsymbol{\Sigma}_{t,i}^{\mathcal{I}}}^{-1}(\boldsymbol{\xi}_t^{\mathcal{I}} - \boldsymbol{\mu}_{t,i}^{\mathcal{I}}), \tag{11}$$

$$\hat{\boldsymbol{\Sigma}}_{t,i}^{\mathcal{O}} = \boldsymbol{\Sigma}_{t,i}^{\mathcal{O}} - \boldsymbol{\Sigma}_{t,i}^{\mathcal{O}\mathcal{I}}{\boldsymbol{\Sigma}_{t,i}^{\mathcal{I}}}^{-1}\boldsymbol{\Sigma}_{t,i}^{\mathcal{I}\mathcal{O}}, \tag{12}$$

$$\text{and} \quad \gamma_i(\boldsymbol{\xi}_t^{\mathcal{I}}) = \frac{\pi_i \mathcal{N}(\boldsymbol{\xi}_t^{\mathcal{I}} |\ \boldsymbol{\mu}_{t,i}^{\mathcal{I}}, \boldsymbol{\Sigma}_{t,i}^{\mathcal{I}})}{\sum_k^K \pi_k \mathcal{N}(\boldsymbol{\xi}_t^{\mathcal{I}} |\ \boldsymbol{\mu}_{t,k}^{\mathcal{I}}, \boldsymbol{\Sigma}_{t,k}^{\mathcal{I}})}. \tag{13}$$

Note that (10) represents a multimodal distribution that can be approximated by a single Gaussian distribution $\mathcal{N}(\hat{\boldsymbol{\mu}}_t^{\mathcal{O}}, \hat{\boldsymbol{\Sigma}}_t^{\mathcal{O}})$ with parameters

$$\hat{\boldsymbol{\mu}}_t^{\mathcal{O}} = \sum_{i=1}^{K} \gamma_i \hat{\boldsymbol{\mu}}_{t,i}^{\mathcal{O}}, \tag{14}$$

$$\hat{\boldsymbol{\Sigma}}_t^{\mathcal{O}} = \sum_{i=1}^{K} \gamma_i \Big[\hat{\boldsymbol{\Sigma}}_{t,i}^{\mathcal{O}} + \hat{\boldsymbol{\mu}}_{t,i}^{\mathcal{O}}(\hat{\boldsymbol{\mu}}_{t,i}^{\mathcal{O}})^{\top}\Big] - \hat{\boldsymbol{\mu}}_t^{\mathcal{O}}(\hat{\boldsymbol{\mu}}_t^{\mathcal{O}})^{\top}. \tag{15}$$

Hence, we can obtain the reference state of the robot in an online manner during the cooperative task by GMR. The desired position $\bar{\boldsymbol{x}}$, velocity $\bar{\dot{\boldsymbol{x}}}$, and forces $\bar{\boldsymbol{f}}$ will be used in (5), thus establishing the feedback control law from the cooperative behavior previously demonstrated.

V. OPTIMAL CONTROLLER GAINS ESTIMATION

Once the reference state of the robot has been learned for the collaborative task at hand, it is crucial to determine how the robot will follow this reference state during reproduction. First, let us define the whole state of the robot

as $\boldsymbol{\zeta} = [\,\boldsymbol{x}^\top\ \dot{\boldsymbol{x}}^\top\ \boldsymbol{f}^\top\,]^\top$, recalling that $\boldsymbol{x}$, $\dot{\boldsymbol{x}}$ and $\boldsymbol{f}$ are the Cartesian position, velocity and sensed force of the robot end-effector. We also define the inputs to the system as the vector $\boldsymbol{\nu} = [\,\boldsymbol{u}^\top\ \boldsymbol{v}^\top\,]^\top$, where $\boldsymbol{u}$ is the control input expressed as (5), and $\boldsymbol{v}$ represents an external input to the system. Note that, unlike [7], [14], this additional input allows us not only to establish a dynamic equation for the sensed forces $\boldsymbol{f}$, but also to include the influence of an external input $\boldsymbol{v}$ on the system dynamics. In the HRC context, such an input can represent the interaction of the human with the robot during the cooperative task. Furthermore, let us assume that the end-effector becomes equivalent to a unit mass after gravity compensation, so the state space representation of the robot in task space can be written as[1]

$$\dot{\boldsymbol{\zeta}} = \overbrace{\begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}}^{\boldsymbol{A}} \boldsymbol{\zeta} + \overbrace{\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}}^{\boldsymbol{B}} \boldsymbol{\nu}, \tag{16}$$

namely $\frac{d}{dt}\boldsymbol{x} = \dot{\boldsymbol{x}}$, $\frac{d}{dt}\dot{\boldsymbol{x}} = \boldsymbol{u} + \boldsymbol{f}$, and $\frac{d}{dt}\boldsymbol{f} = \boldsymbol{v}$. Note that the latter equation indicates that the variation of the sensed forces depends on the external input $\boldsymbol{v}$, in other words, the physical interaction between the human and the robot directly influences the variation of the robot's force perception. Lastly, we denote the column space of the input matrix $\boldsymbol{B} = [\boldsymbol{B}_1\ \boldsymbol{B}_2]$.

Once reference position, velocity and force profiles have been obtained for the current time step, the controller gains can be estimated with an optimal control strategy. Optimal feedback controllers allow the robot to plan a feedback control law tracking the desired state. Formally, the problem is stated as finding the optimal input $\boldsymbol{\nu}$ that minimizes the cost function

$$J_t = \sum_{n=t}^{\infty} (\boldsymbol{\zeta}_n - \bar{\boldsymbol{\zeta}}_t)^\top \boldsymbol{Q}_t (\boldsymbol{\zeta}_n - \bar{\boldsymbol{\zeta}}_t) \ + \ \boldsymbol{\nu}_n^\top \boldsymbol{R}_t \, \boldsymbol{\nu}_n, \tag{17}$$

where $\bar{\boldsymbol{\zeta}}_t$ represents the reference or desired state obtained by GMR, while the matrices $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ are weighting matrices that determine the proportion in which the tracking errors and control inputs affect the minimization problem. The aforementioned problem is typically known as an infinite horizon LQR [16]. We exploit (17) in two new manners. First, we take advantage of the variability observed during the demonstrations to adapt on-the-fly the error costs in (17). This can also be interpreted as shaping the gain matrices according to the precision required by the task across the robot state variables. On the other hand, the minimization of the second term of (17) implies that both the robot control command $\boldsymbol{u}$ and the external input $\boldsymbol{v}$ are minimum, and consequently minimizing the robot effort and human intervention.

We define

$$\boldsymbol{Q}_t = \left(\hat{\boldsymbol{\Sigma}}_t^{\mathcal{O}}\right)^{-1}, \quad \boldsymbol{R}_t = \begin{bmatrix} \boldsymbol{R}_t^u & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_t^v \end{bmatrix}, \tag{18}$$

[1] $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices defining the dynamical system, not to be confounded with the $\boldsymbol{A}_{t,j}$ and $\boldsymbol{b}_{t,j}$ defining the coordinate systems in (6).

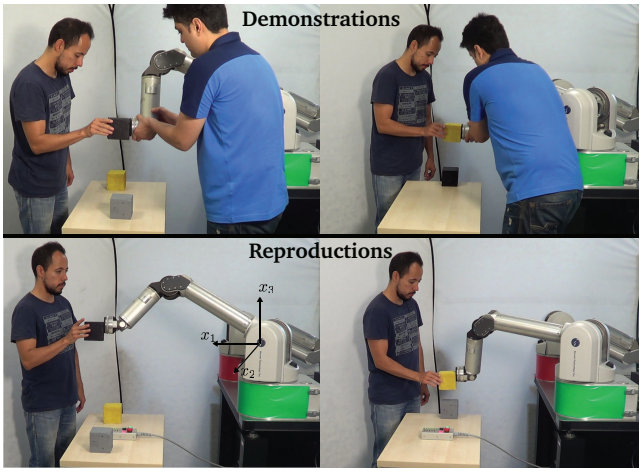


Fig. 2: Experimental setting of the human-robot transportation task: (*top*) kinesthetic demonstrations, and (*bottom*) reproduction phase.

using (15). In our experiment, $\boldsymbol{R}_t$ is defined as a diagonal matrix. Nevertheless, it is worth highlighting that the structure of the matrix $\boldsymbol{R}_t$ permits to vary the relevance given to the minimization of the robot control command and the external human input through the submatrices $\boldsymbol{R}_t^u$ and $\boldsymbol{R}_t^v$, respectively. Such features can be significantly exploited in physical human-robot interaction, where the matrix $\boldsymbol{R}_t^u$ can be shaped over time according to the safety level demanded by the task. For instance, the higher the $\boldsymbol{R}_t^u$ values, the lower the control forces applied by the robot, and therefore the safer the interaction. This issue will be thoroughly explored in future works.

Note that the cost function is updated at each time step t to compute the next control command. This formulation is better suited for HRC in weakly structured environments, where the robot actions might be updated swiftly based on the state and/or actions of the user, and the state of the environment. In contrast to infinite horizon LQR, finite-horizon requires the recursive computation of an ordinary differential equation, and is thus better suited for planning situations in which the candidate frames are not expected to move. The minimization of (17) can be solved through the algebraic Riccati equation, providing an optimal feedback controller in the form of (5) with full stiffness, damping and force gain matrices. Specifically, the LQR solution for our problem is represented by

$$\boldsymbol{\nu}_t = \boldsymbol{R}_t^{-1} \boldsymbol{B}^\top \left[-\boldsymbol{S}_t (\,\boldsymbol{\zeta}_t - \bar{\boldsymbol{\zeta}}_t\,) + \boldsymbol{d}_t \right], \tag{19}$$

where the robot controller is obtained as

$$\boldsymbol{u}_t = {\boldsymbol{R}_t^u}^{-1} \boldsymbol{B}_1^\top \left[-\boldsymbol{S}_t \left(\boldsymbol{\zeta}_t - \bar{\boldsymbol{\zeta}}_t \right) + \boldsymbol{d}_t \right], \tag{20}$$

with $\boldsymbol{S}_t$ and $\boldsymbol{d}_t$ as solutions of the equations

$$\boldsymbol{A}^\top \boldsymbol{S}_t + \boldsymbol{Q}_t + \boldsymbol{S}_t \boldsymbol{A} - \boldsymbol{S}_t \boldsymbol{B} \boldsymbol{R}_t^{-1} \boldsymbol{B}^\top \boldsymbol{S}_t = \boldsymbol{0}, \tag{21}$$

$$-\boldsymbol{A}^\top \boldsymbol{d}_t + \boldsymbol{S}_t \boldsymbol{A} \boldsymbol{\zeta}_t + \boldsymbol{S}_t \boldsymbol{B} \boldsymbol{R}_t^{-1} \boldsymbol{B}^\top \boldsymbol{d}_t - \boldsymbol{S}_t \dot{\boldsymbol{\zeta}}_t = \boldsymbol{0}, \tag{22}$$

and $\boldsymbol{B}_1$ belonging to the column space of $\boldsymbol{B}$, as specified previously. In the above, $\boldsymbol{d}_t$ is the feedforward term, which

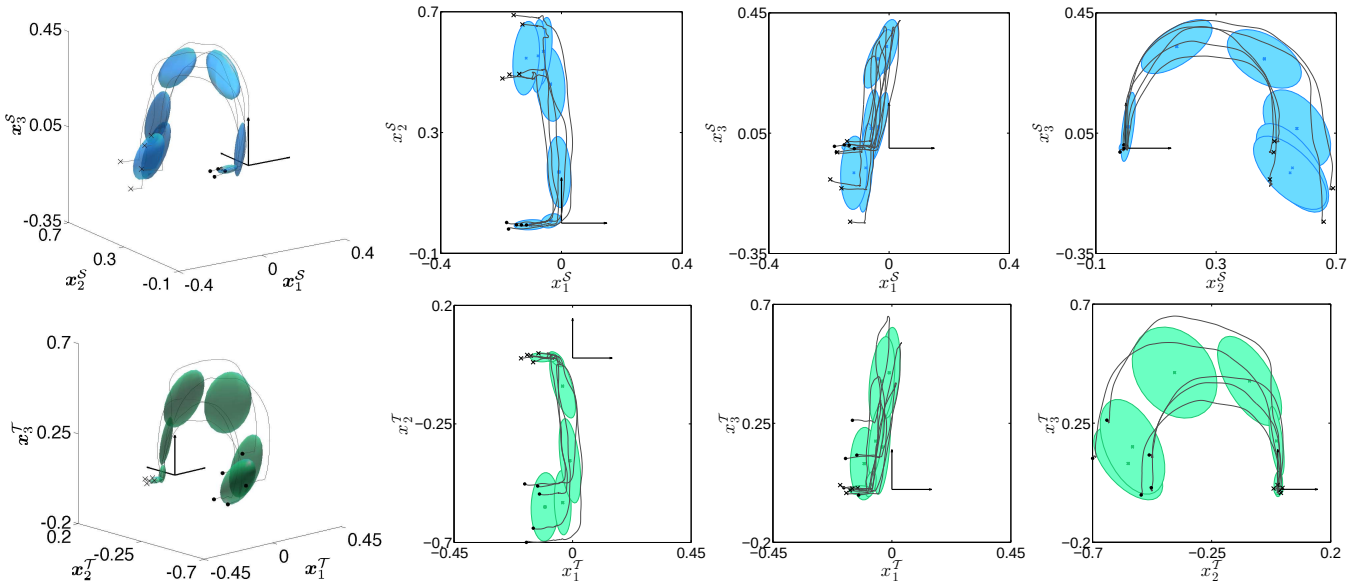


Fig. 3: Encoding of the demonstrations in the different coordinate systems. The local models in the *start* and *target* frames are respectively shown in the first and second rows (projected on the position subspace of the data). The gray lines depict the end-effector trajectory (in meters) observed from the different frames. The ellipsoids represent the Gaussian components of the model. Black dots depict the beginning of the demonstration.

can optionally be neglected for low dynamic movements. The solution for $\boldsymbol{u}_t$ provides optimal feedback gains $\boldsymbol{K}^{\mathcal{P}}$, $\boldsymbol{K}^{\mathcal{V}}$ and $\boldsymbol{K}^{\mathcal{F}}$, which allow the robot to optimally track its desired state during the cooperative task in a stable manner, while shaping its compliance level according to the invariant characteristics of the demonstrations. This permits the robot to perform the task precisely. It is worth mentioning that an equation similar to (20) is obtained for $\boldsymbol{v}_t$, which can be theoretically interpreted as the optimal interaction input generated by the human, that can be used for simulation purposes. In the HRC context, such an external input is directly given by the human counterpart.

VI. EXPERIMENT

We test the performance of our approach in an experiment where a human-robot dyad transports an object from an initial location to a desired target, similarly to [15]. However, in [15] the robot was not able to behave with different compliance levels, because both the stiffness and damping matrices were manually set by the user. The robot was also not endowed with force feedback during reproduction, and the proposed formulation did not consider the minimization of the robot effort and human intervention.

The detailed description about the setting, the demonstration and reproduction phases as well as the obtained results are given below.

A. Description

The experiment consists of teaching a robot to simultaneously handle position, velocity and force constraints arising when a human and a robot cooperatively manipulate/transport an object (see Fig. 2). At the beginning of the transportation task, two participants reach for the object. Once they make contact with the load, they start jointly transporting the object along a bell-shaped path to reach the target location. When the object gets to the final position, the two persons release it and move away from it. Note that both the starting and goal object positions vary across repetitions. The aim is to introduce a robot into such a task by replacing one of the human participants by a robot.

For this experiment, we used a torque-controlled 7 DOF WAM robot endowed with a 6-axis force/torque sensor. The robot controller is defined by (2) and (5). In the demonstration phase, the gravity-compensated robot is kinesthetically guided by the teacher while cooperatively achieving the task with the other human partner, as shown in Fig. 2. The teacher shows the robot both the path to be followed and the force pattern it should use while transporting the load.

In this task two candidate coordinate systems ($P = 2$) are considered, namely, the frames representing the initial and target locations of the object. They are respectively defined as $\{\boldsymbol{b}^{\mathcal{S}}, \boldsymbol{A}^{\mathcal{S}}\}$ and $\{\boldsymbol{b}^{\mathcal{T}}, \boldsymbol{A}^{\mathcal{T}}\}$. Here, $\boldsymbol{b}^{\mathcal{S}} = [0\, \boldsymbol{x}^{\mathcal{S}\top}\, \boldsymbol{0}_{1\times 6}]^\top$ and $\boldsymbol{b}^{\mathcal{T}} = [0\, \boldsymbol{x}^{\mathcal{T}\top}\, \boldsymbol{0}_{1\times 6}]^\top$, where $\boldsymbol{x}^{\mathcal{S}}$ and $\boldsymbol{x}^{\mathcal{T}}$ are the Cartesian positions where the object is picked up and then released. Similarly, the transformation matrices are defined as $\boldsymbol{A}^{\mathcal{S}} = \text{blockdiag}(1, \boldsymbol{R}^{\mathcal{S}}, \boldsymbol{R}^{\mathcal{S}}, \boldsymbol{R}^{\mathcal{S}})$ and $\boldsymbol{A}^{\mathcal{T}} = \text{blockdiag}(1, \boldsymbol{R}^{\mathcal{T}}, \boldsymbol{R}^{\mathcal{T}}, \boldsymbol{R}^{\mathcal{T}})$, where $\boldsymbol{R}^{\mathcal{S}}$ and $\boldsymbol{R}^{\mathcal{T}}$ respectively represent the initial and final orientation of the object with rotation matrices.[2] The motion in this experiment is time-driven, therefore each datapoint $\boldsymbol{\xi}_t$ is defined as $\boldsymbol{\xi}_t^{\mathcal{I}} = t$ and $\boldsymbol{\xi}_t^{\mathcal{O}} = \boldsymbol{\zeta}_t$, where t and $\boldsymbol{\zeta}_t$ are time and the whole state of the robot, respectively.

During reproduction of the task, the start and target frames are given to the model in order to obtain the temporary

[2]The positions and orientations of the object were predefined in this experiment, but these can alternatively be obtained using a vision system.

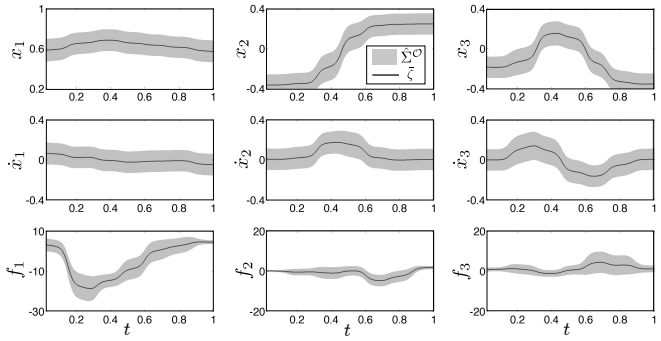


Fig. 4: Estimated desired state of the robot $\bar{\boldsymbol{\zeta}}$ and its associated variance $\hat{\boldsymbol{\Sigma}}^{\mathcal{O}}$ along a reproduction of the cooperative transportation. Position, velocity and forces are respectively given in meters, meters per second, and Newtons.

GMM parameters using (6)-(8). The orientation of the end-effector is fixed. Then, the robot and the user transport the object towards the target location. Here, for each time step t, the robot obtains an updated reference state $\bar{\boldsymbol{\zeta}}$ (see Section IV) along with optimal stiffness, damping and force gain matrices, that generate a new desired acceleration in the operational space of the robot.

B. Results

A set of five examples of the collaborative behavior are given to the robot. The demonstrations are then used for training a TP-GMM ($K=7$ empirically determined). Figure 3 shows the resulting encoding of the position trajectories observed from the two different candidate frames. Notice how the multiple demonstrations are locally consistent when the robot approaches the initial location of the object (i.e., frame $\mathcal{S}$), and when the manipulator moves away once the load has been placed at its target position (i.e., frame $\mathcal{T}$). This is reflected by the small and narrow ellipsoids in these parts of the task. Given this model and a new set of task parameters (i.e., initial and target locations), it is possible to compute, in an online manner, the desired state of the robot and associated covariance at each time step, as described in Section IV. Figure 4 displays the desired position, velocity and interaction force and their corresponding variances obtained by GMR. Note that the reference position trajectory can vary as the initial and/or target locations change, as evidenced in Fig. 5 where several reproductions with different task parameters are shown. Also, note that the interaction forces at the beginning and the end of the reproduction are expected to be zero, coinciding with the parts of the task when the robot is approaching or releasing the object (see Fig. 4).

Figure 6 shows how $\boldsymbol{K}^{\mathcal{P}}$, $\boldsymbol{K}^{\mathcal{V}}$ and $\boldsymbol{K}^{\mathcal{F}}$ vary over time along one of the reproduction attempts shown in Fig. 5, with $\boldsymbol{R}_t = r\boldsymbol{I}_{6\times 6}$ and $r=0.01$. Notice that at the beginning and at the end of the task, the robot behaves less stiffly along $\boldsymbol{x}_1$, while being stiffer along the axes $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$. The robot does not allow high variations on the plane $(\boldsymbol{x}_2, \boldsymbol{x}_3)$, guaranteeing that the object is picked up and released by passing through trajectories consistent with the demonstrations. In contrast,

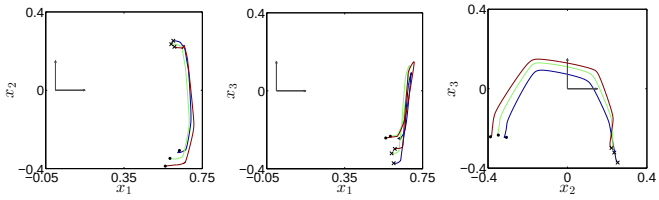


Fig. 5: Reproductions with varying initial and target locations of the object.

as expected, when the human-robot dyad is cooperatively transporting the load, the robot behaves stiffly along $\boldsymbol{x}_1$, while allowing deviations on the plane $(\boldsymbol{x}_2, \boldsymbol{x}_3)$. Lastly, Fig. 7 displays the behavior of the robot for two different cases, namely, when the interaction force is similar to the reference force profile obtained by GMR, and when some perturbation forces are applied at the end-effector. In the latter case, the user exerted different force profiles along the three Cartesian axes (the shaded area in Fig. 7 shows when the perturbations occurred). Note that by checking the difference between the robot's trajectory and the desired path (first row in Fig. 7), it is observed that the robot reacts to deviations along the axis $\boldsymbol{x}_1$, while perturbations along $\boldsymbol{x}_3$ are slightly permitted in the middle of the task. This is coherent with the demonstrated task constraints and with the feedback gain profiles shown in Fig. 6. A video accompanying this paper shows the results of the experiment, and is available at `http://programming-by-demonstration.org/IROS2015/`.

VII. CONCLUSIONS AND FUTURE WORK

We introduced a PbD framework for learning cooperative robot skills in the context of human-robot object transportation. Our approach brings together the advantages of probabilistic encoding, generalization capability of the task-parametrized GMM, and robustness of optimal control used with both position and force constraints. This framework allows the robot to automatically encode the human demonstrations and their interconnection with parameters of the task. Moreover, in contrast to [15], the robot is able to exploit the observed variability for estimating different optimal compliance levels over time, while determining the precision with which the state variable errors need to be compensated for. Lastly, our approach reduces robot effort and human intervention, thus favoring safer interactions. Experiments on a real setting showed the strengths and practical use of the approach.

The proposed model was used to learn a time-driven robot motion, where part of the desired state also depend on the parameters of the task. In this sense, we plan in future work to avoid the explicit time dependence by taking advantage of methods that also encapsulate the sequential information of the task. We also plan to study how to exploit the structure of our cost function in order to include safety constraints as a function of the interaction with the human. Moreover, further work is required to investigate how the state of the user could be included into the loop, so that the robot could react in various ways to its human partner's actions.

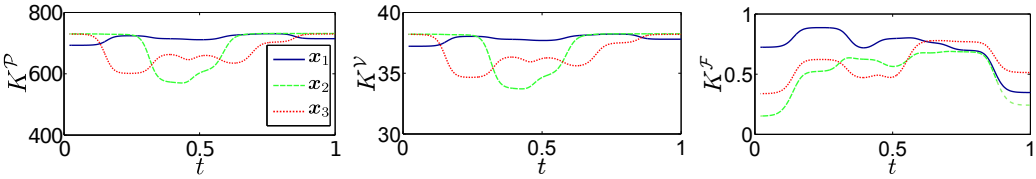
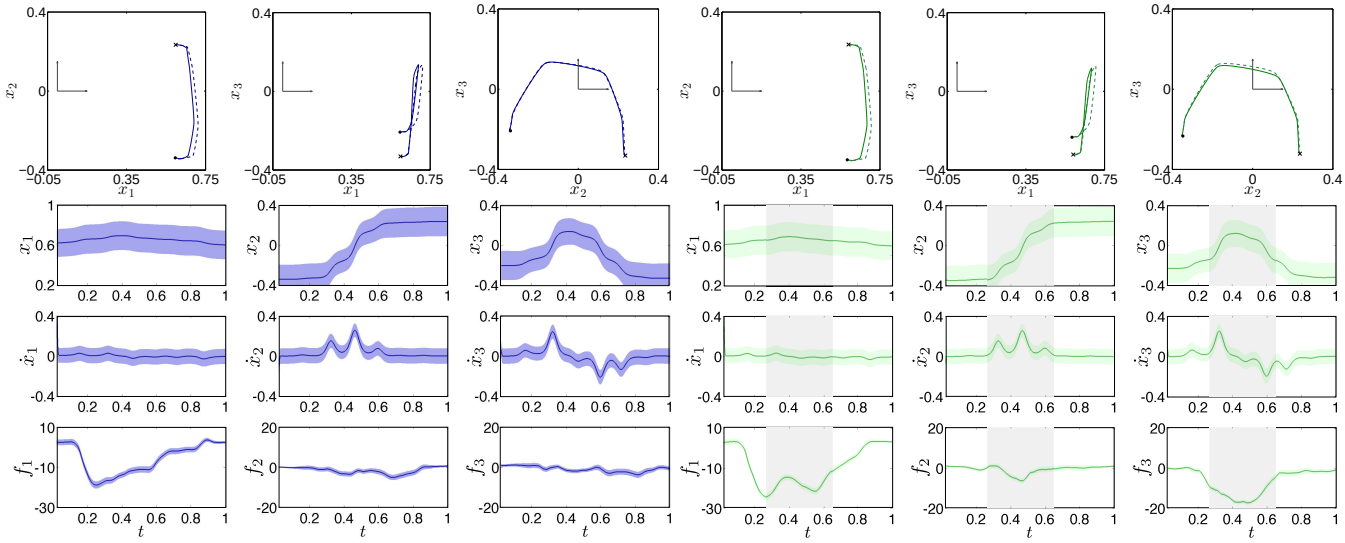


Fig. 6: Profiles of the diagonal values of the estimated stiffness, damping and force gain matrices during the reproduction of the cooperative transportation.



(a) Interaction force similar to demonstrations. (b) Interaction with perturbations.

Fig. 7: Reproductions of the cooperative transportation. The first row shows the robot trajectory (solid line) and the desired path (dashed line) in the three planes of the Cartesian space. The remaining rows show the position, velocity and sensed force of the robot over time (solid lines), along with corresponding feedback gain (represented as an envelope surrounding the reproduced trajectory).

REFERENCES

[1] H. Admoni, A. Dragan, S. Srinivasa, and B. Scassellati, "Deliberate delays during robot-to-human handovers improve compliance with gaze communication," in *ACM/IEEE Intl. Conf. on Human-Robot Interaction (HRI)*, 2014, pp. 49–56.

[2] D. Agravante, A. Cherubini, A. Bussy, and A. Kheddar, "Human-humanoid joint haptic table carrying task with height stabilization using vision," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2013, pp. 4609–4614.

[3] L. Rozo, S. Calinon, D. G. Caldwell, P. Jiménez, and C. Torras, "Learning collaborative impedance-based robot behaviors," in *AAAI Conf. on Artificial Intelligence*, 2013, pp. 1422–1428.

[4] A. D. Santis, B. Siciliano, A. D. Luca, and A. Bicchi, "An atlas of physical human-robot interaction," *Mechanism and Machine Theory*, vol. 43, no. 3, pp. 253–270, 2008.

[5] B. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.

[6] L. Rozo, P. Jiménez, and C. Torras, "A robot learning from demonstration framework to perform force-based manipulation tasks," *Journal of Intelligent Service Robotics, Special Issue on Artificial Intelligence Techniques for Robotics: Sensing, Representation and Action, Part 2*, vol. 6, no. 1, pp. 33–51, 2013.

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

[8] P. Evrard, E. Gribovskaya, S. Calinon, A. Billard, and A. Kheddar, "Teaching physical collaborative tasks: Object-lifting case study with a humanoid," in *IEEE/RAS Intl. Conf. on Humanoid Robots (Humanoids)*, 2009, pp. 399–404.

[9] E. Gribovskaya, A. Kheddar, and A. Billard, "Motion learning and adaptive impedance for robot control during physical interaction with humans," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2011, pp. 4326–4332.

[10] J. Medina, M. Lawitzky, A. Mortl, D. Lee, and S. Hirche, "An experience-driven robotic assistant acquiring human knowledge to improve haptic cooperation," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2011, pp. 2416–2422.

[11] T. Kulvicius, M. Biehlc, M. J. Aein, M. Tamosiunaite, and F. Wörgötter, "Interaction learning for dynamic movement primitives used in cooperative robotic tasks," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1450–1459, 2013.

[12] H. B. Amor, G. Neumann, S. Kamthe, O. Kroemer, and J. Peters, "Interaction primitives for human-robot cooperation tasks," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2014, pp. 2831–2837.

[13] L. Peternel, T. Petrič, E. Oztop, and J. Babič, "Teaching robots to cooperate with humans in dynamic manipulation tasks based on multi-modal human-in-the-loop approach," *Autonomous Robots*, vol. 36, no. 1–2, pp. 123–136, 2014.

[14] J. Medina, M. Lawitzky, A. Molin, and S. Hirche, "Dynamic strategy selection for physical robotic assistance in partially known tasks," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 1180–1186.

[15] L. Rozo, S. Calinon, and D. G. Caldwell, "Learning force and position constraints in human-robot cooperative transportation," in *IEEE Intl. Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2014, pp. 619–624.

[16] E. Todorov, "Optimal control theory," in *Bayesian Brain*, K. Doya, Ed. MIT Press, 2006, pp. 1–28.

[17] Z. Ghahramani and M. Jordan, "Supervised learning from incomplete data via EM approach," in *Neural Information Processing Systems (NIPS)*, 1994, pp. 120–127.