Online Verification of Impact-Force-Limiting Control for Physical Human-Robot Interaction

Stefan B. Liu, and Matthias Althoff

Abstract— **Humans must remain unharmed during their interaction with robots. We present a new method guaranteeing impact force limits when humans and robots share a workspace. Formal guarantees are realized using an online verification method, which plans and verifies fail-safe maneuvers through predicting reachable impact forces by considering all future possible scenarios. We model collisions as a coupled human-robot dynamical system with uncertainties and identify reachset-conforming models based on real-world collision experiments. The effectiveness of our approach for human-robot co-existence is demonstrated for the human hand interacting with the end effector of a six-axis robot manipulator with force sensing. By integrating a human pose detection system, the efficiency of robot movements increases.**

I. Introduction

Humans and robots are sharing their workspaces, collaborating, and interacting with each other. Trending application areas include collaborative manufacturing, assistive robotics for rehabilitation and elderly care, and robotic surgery. When designing robot controllers, safety is one of the top priorities; humans should never be harmed or injured. To mitigate pain, ISO/TS 15066 [1] defined interaction-force thresholds for each body part, which should not be exceeded. However, guaranteeing safety through limiting these forces is a challenging task:

- The human body is capable of performing a range of movements, making it difficult to predict exact collision scenarios.
- Interaction forces depend on the mechanical properties of robots and humans. These are subject to uncertainty, e.g., stiffnesses changes according to muscle activity.
- Varying tasks and diverse environments create many possibilities for collision, thus, offline assessments become infeasible. Therefore, an online approach should be preferred, considering only the current situation.
- To guarantee safety properties despite uncertainties, formal methods should be used.

We propose to tackle these challenges through an online verification approach based on human pose detection, fail-safe planning, and reachability analysis. The fail-safe planner decides whether an upcoming motion command can be executed by verifying the safety of possible fail-safe maneuvers. The online proofs are based on reachability analysis, which checks, whether all possible interaction forces are within specified limits. Reachable sets are computed using a model of the coupled human-robot interaction dynamics. Uncertainties in the system, such as human velocity, collision time, varying stiffnesses, control performance are all modeled as sets, and the interaction dynamics are identified in a way that preserves reachset conformance with real behaviors.

Most of the previous approaches only assess safety without proving thresholds. Shivakumar et al. [2] propose that impact forces with environmental objects can be predicted using a spring-damper model or an energy-based model. Yamada et al. [3] describe how to design the thickness of a viscoelastic coat for robots to avoid exceeding pain limits during collisions. Ikuta et al. [4] introduce a danger index relative to the maximum tolerable collision force at the end effector, which depends on factors such as the robot's mass and velocity and joint- and and coating elasticities. Heinzmann and Zelinski [5] propose an online safety controller that derives admissible control torques from the maximum collision forces of a rigid robot, coupled with scaling of the robot velocity. Models used in [3]–[5], however, assume that human is a rigid obstacle, which reduces uncertainty but contributes to a conservative force estimation. Post-collision force-limiting strategies in Navarro et al. [6] and Li et al. [7] focus on reactive behavior for overshoots during the interaction, however, it cannot guarantee impact-force limits. Some non-mentioned works use the model provided in ISO/TS 15066 [1] to guarantee force limits. However, Kirschner et al. [8] reported that the model is inaccurate and unsuitable for estimating collision forces. In contrast to these non-formal studies, we consider impedance models with reachset conforming uncertainties to provide formal guarantees.

In addition, alternative metrics for reducing impact injury have been proposed, involving velocity [9], [10] or energy and power [11]–[14], which are easier to evaluate, since only the robot model is required. Haddadin et al. [9] realized that injury occurrence is directly related to the impact velocity beyond a certain robot mass. A database has been implemented by Mansfeld et al. [10], which can be used for online and offline injury assessments based on collision speeds and robot modeling. Meguenani et al. [11] indirectly limit impact force by limiting the kinetic and potential energy of the robot. Raiola et al. [12] scale the stiffness and damping of impedance controllers to guarantee energy and power limits. The port-Hamiltonian formulation of coupled human-robot dynamics in [13], [14] allows one to directly control energy in physical interaction to preserve passivity. The difficulty with speed, energy, and power metrics is that suitable limits are unavailable, or are based on the non-formal derivations from [1]. In contrast, we verify established force-based pain

Authors are with Cyber-Physical Systems Group, Department of Informatics, Technical University of Munich, 85748 Garching, Germany [stefan.liu,althoff]@tum.de

limits [1] for humans.

Our study is the first one that uses formal methods to verify controllers for physical human-robot interaction. In addition, we provide an identification method for models and uncertainties based on real-world experiments. Also, other methods for formal verification, such as differential dynamic logic theorem-proving [15] and inevitable collision states [16] consider uncertainties in dynamical systems.

This study is structured as follows: we define the safety properties to be verified in Sec. II. Modeling and identification of the coupled human-robot dynamics are discussed in Sec. III. The impact-force-limiting controller is presented in Sec. IV. The experimental evaluation in Sec. V demonstrates the effectiveness of our approach on a real interaction scenario, followed by the conclusions in Sec. VI.

II. Safety objectives

This section poses the safety problem that is encountered inbetween humans and robots. We denote sets in calligraphic letters (e.g., $\mathcal{A}$), matrices with upper case letters (e.g., A), vectors by $\vec{\cdot}$, and scalar values by lower case letters (e.g., a). Considering a system with state vector $\vec{z}$, input vector $\vec{u}$, and parameters $\vec{p}$, of which the dynamical equation is $\dot{\vec{z}} = \vec{f}(\vec{z}, \vec{u}, \vec{p})$. We make use of reachable sets, which are defined as follows:

Definition 1 (Reachable Set)**.** Given the initial set $\mathcal{Z}_0$, the uncertain input set $\mathcal{U}$, and the non-deterministic parameter set $\mathcal{P}$, the reachable set of $\dot{\vec{z}} = \vec{f}(\vec{z}, \vec{u}, \vec{p})$ at time t is

$$\mathcal{R}(t) = \Big\{ \int_0^t \vec{f}(\vec{z}(\tau), \vec{u}(\tau), \vec{p}) d\tau + \vec{z}(0) \Big| \vec{z}(0) \in \mathcal{Z}_0, \\ \forall \tau \in [0, t] : \vec{u}(\tau) \in \mathcal{U}, \vec{p}(\tau) \in \mathcal{P} \Big\}.$$

To compute $\mathcal{R}(t)$ (also denoted as $\texttt{reach}(\mathcal{Z}_0, \mathcal{U}, \mathcal{P})$), we use an optimized version of the software CORA [17].

We regard systems consisting of humans sharing a workspace with a robot manipulator. From the goal that a robot should not actively cause harm to the human, we derive three safety objectives:

1) A non-moving robot cannot actively cause harm to a human. Consider a robot manipulator with n degree of freedoms, where $\vec{q}, \dot{\vec{q}} \in \mathbb{R}^n$ are the joint position and velocity of the robot, and $\vec{x} = [\vec{q}, \dot{\vec{q}}]^T$ is its state. Let us define the predicate $\mathrm{standstill}(t)$ indicating whether the system is safe:

$$\mathrm{standstill}(t) \iff \vec{x}(t) \in \mathcal{ISS} := \mathbb{R}^n \times \vec{0},$$

where $\vec{0}$ is a vector of n zeros. We refer to the set on the right hand side as an *invariably safe set*, implying that our system is safe for an infinite time horizon when it is reached.

2) We consider that a robot cannot cause harm to the human, if they are not physically interacting, i.e., the occupied space of the human does not overlap with the occupied space of the robot. We denote $\mathcal{M}(t)$ and $\mathcal{H}(t)$ as occupancy sets of the robot and human, respectively:

$$\mathrm{noInteraction}(t) \iff \mathcal{M}(t) \cap \mathcal{H}(t) = \emptyset.$$

To predict occupancy sets of humans, a tracking system is required. For additional information, we refer to our previous work in [18], [19].

3) We consider that harm is caused to the human if force thresholds are violated during an impact. For ISO/TS 15066, two limits are defined: a *transient force* limit $f_{\mathrm{tra,lim}}$, which is the peak at the beginning of a collision, and the quasi-static force $f_{\mathrm{qs,lim}}$ limit, which is the converged stationary force acting on a clamped human. We introduce the reachable set of the absolute force $\mathcal{F}_{\mathrm{coll}}(\tau) \subseteq \mathbb{R}, t < \tau < t + t_e$, where t_e is a prediction horizon. Our system is safe if

$$\mathrm{safeForce}(t) \iff \sup\left(\mathcal{F}_{\mathrm{coll}}(\tau)\right) \leq f_{\mathrm{tra,lim}} \\ \wedge \lim_{\tau \to t_e} \sup(\mathcal{F}_{\mathrm{coll}}(\tau)) \leq f_{\mathrm{qs,lim}},$$

where sup is the supremum, and t_e needs to be large to converge to the quasi-static force.

We consider a system to be verified as safe, if any of the above three conditions hold at all times:

$$\forall t : \mathrm{standstill}(t) \vee \mathrm{noInteraction}(t) \\ \vee \mathrm{safeForce}(t) \iff \text{safe}. \quad (1)$$

The remaining part of this paper focuses on the prediction of reachable forces to evaluate the predicate $\mathrm{safeForce}(t)$. For the other predicates, we refer to [18], [19].

III. Interaction Modeling

To represent physical interaction, we state the dynamical models with uncertainties in Sec. III-A, and present its reachset-conforming model identification in Sec. III-B.

A. Physical interaction modeling

The goal of the model is to predict the set of reachable forces $\mathcal{F}_{\mathrm{coll}}$, given the planned robot trajectory, and the human and robot collision velocities. We make the following assumptions:

- We model the case of a hand interacting with the robot end-effector, which is controlled by a Cartesian impedance controller.
- The collision is a blunt impact with any part of the end effector from any direction, for which the force limits apply [1]. We do not consider robots with sharp edges; for their safety analysis, pressure limits apply [1].
- The collision is unintended, i.e., the human does not push against the robot, and remains passive after impact.

In addition, we use a scalar model to represent the dynamics in all possible (three-dim.) spatial directions. A projection operator over-approximatively transforms three-dimensional inputs of our models into a one-dimensional interval:

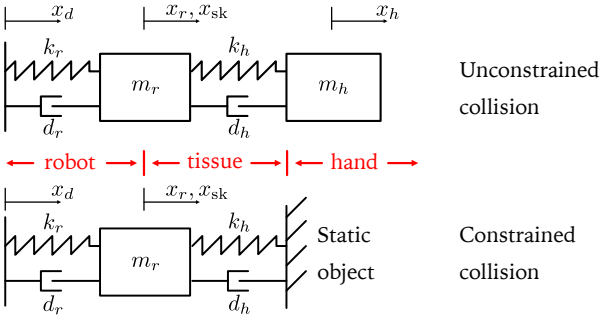

Fig. 1. Physical interaction is modelled as mass-spring-damper systems.

Definition 2 (Scalar projection of a set)**.** The scalar projection of a three-dimensional set $\mathcal{S}$ is defined as an interval

$$\texttt{proj}(\mathcal{S}) := [-\|\mathcal{S}\|_2, \|\mathcal{S}\|_2],$$

where the 2-norm of a set is $\|\mathcal{S}\|_2 := \sup\{\|s\|_2 | s \in \mathcal{S}\}$.

Our modeling approach takes the following steps: 1) derive a human model, 2) derive a robot model, and 3) couple the dynamics and introduce uncertainty into the model. We distinguish between two types of collisions [20]:

1) *Unconstrained collision*: the human hand can move away after a collision, i.e., it is not clamped.
2) *Constrained collision*: the human hand is clamped between the robot end-effector and another static object.

We choose mass-spring-damper systems to model both interactions (Fig. 1). For the unconstrained collision, the human hand is modeled by a moving mass m_h, where x_{sk} are the hand and skin position, respectively, and the impact force is $f_{\text{coll},1}$. The skin has a tissue stiffness k_h, and a damping d_h:

$$m_h \ddot{x}_h = \underbrace{k_h(x_{\text{sk}} - x_h) + d_h(\dot{x}_{\text{sk}} - \dot{x}_h)}_{f_{\text{coll},1}}, \tag{2}$$

For the constrained collision, the hand position is assumed to be fixed, thus cannot move ($x_h, \dot{x}_h, \ddot{x}_h = 0$). Therefore, we define the dynamical equation as

$$f_{\text{coll},2} = k_h x_{\text{sk}} + d_h \dot{x}_{\text{sk}}. \tag{3}$$

To model the robot, we consider the rigid-body dynamics

$$M(\vec{q})\ddot{\vec{q}} + C(\vec{q}, \dot{\vec{q}})\dot{\vec{q}} + \vec{g}(\vec{q}) = \vec{\tau} + J(\vec{q})^T \vec{f}_{\text{ext}}, \tag{4}$$

where $\vec{q}$ is the joint position, $M(\vec{q})$ the mass matrix, $C(\vec{q}, \dot{\vec{q}})$ the Coriolis and centripetal matrix, $\vec{g}(\vec{q})$ the gravity torques, $J(\vec{q})$ the Jacobian, $\vec{f}_{\text{ext}}$ the measured external force at the end effector, and $\vec{\tau}$ the input torque. To track the desired trajectory $\vec{x}_d(t)$, we use a Cartesian impedance controller [21]—a prominent method for controlling human-robot interaction [22]—given by

$$\begin{aligned}
\vec{\tau} =& \vec{g}(\vec{q}) + J(\vec{q})^T(\Lambda(\vec{q})\ddot{\vec{x}}_d + \mu(\vec{q}, \dot{\vec{q}})\dot{\vec{x}}_r) - \\
& J(\vec{q})^T \Lambda(\vec{q}) \Lambda_r^{-1}(K_r(\vec{x} - \vec{x}_d) + D_r(\dot{\vec{x}} - \dot{\vec{x}}_d)) + \\
& J(\vec{q})^T(\Lambda(\vec{q})\Lambda_r^{-1} - I)\vec{f}_{\text{ext}}, \\
\Lambda(\vec{q}) =& J(\vec{q})^{-T} M(\vec{q}) J(\vec{q})^{-1}, \\
\mu(\vec{q}, \dot{\vec{q}}) =& J(\vec{q})^{-T} \left(C(\vec{q}, \dot{\vec{q}}) - M(\vec{q}) J(\vec{q})^{-1} \dot{J}(\vec{q}) \right) J(\vec{q})^{-1},
\end{aligned}$$

where $\vec{x}_r$ is the end effector position, Λ_r is the desired mass matrix, K_r is the desired stiffness matrix, and D_r is the desired damping matrix. Thus, the end effector behaves like a mass-spring-damper system, which can be seen in the closed-loop robot dynamics [21]:

$$\Lambda_r(\ddot{\vec{x}}_r - \ddot{\vec{x}}_d) + D_r(\dot{\vec{x}}_r - \dot{\vec{x}}_d) + K_r(\vec{x}_r - \vec{x}_d) = \vec{f}_{\text{ext}}. \tag{5}$$

We consider only the translational part of the closed-loop dynamics since our interest is in translational forces, i.e., $\vec{x}_r, \vec{x}_d, \vec{f}_{\text{ext}} \in \mathbb{R}^3$ and $\Lambda_r, D_r, K_r \in \mathbb{R}^{3\times 3}$. When choosing $\Lambda_r = m_r I, D_r = d_r I$ and $K_r = k_r I$, where m_r, d_r, k_r are scalars and I is a three-dimensional identity matrix, then the following equation

$$m_r(\ddot{x}_r - \ddot{x}_d) + d_r(\dot{x}_r - \dot{x}_d) + k_r(x_r - x_d) = f_{\text{ext}} \tag{6}$$

is the orthogonal projection of (5) onto any spatial direction.

We derive the coupled dynamics of the unconstrained collision by coupling the forces $f_{\text{coll},1,2} = -f_{\text{ext}}$, connecting the end effector to the skin of the human hand $x_r = x_{\text{sk}}$, and inserting (2) into (6). Given the vector $\vec{z}_1 = [x_r, \dot{x}_r, x_h, \dot{x}_h]^T$, the state-space representation of the dynamics is:

$$\dot{\vec{z}}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{(k_r+k_h)}{m_r} & -\frac{(d_h+d_r)}{m_r} & \frac{k_r}{m_r} & \frac{d_h}{m_r} \\ 0 & 0 & 0 & 1 \\ \frac{k_h}{m_r} & \frac{d_h}{m_r} & -\frac{k_h}{m_r} & -\frac{d_h}{m_r} \end{bmatrix} \vec{z}_1 + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} u + \vec{w}_1, \tag{7}$$

$$f_{\text{coll},1} = \begin{bmatrix} k_h & d_h & -k_h & -d_h \end{bmatrix} \vec{z}_1 + v_1, \tag{8}$$

and u is an orthogonal projection of

$$\vec{u} = \ddot{\vec{x}}_d + \Lambda_r^{-1} D_r \dot{\vec{x}}_d + \Lambda_r^{-1} K_r \vec{x}_d \tag{9}$$

The coupled dynamics for the constrained collision are derived by inserting (3) into (6). Given state $\vec{z}_2 = [x_r, \dot{x}_r]^T$:

$$\dot{\vec{z}}_2 = \begin{bmatrix} 0 & 1 \\ -\frac{(k_r+k_h)}{m_r} & -\frac{(d_h+d_r)}{m_r} \end{bmatrix} \vec{z}_2 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u + \vec{w}_2 \tag{10}$$

$$f_{\text{coll},2} = \begin{bmatrix} k_h & d_h \end{bmatrix} \vec{z}_2 + v_2. \tag{11}$$

The state-space dynamics have been augmented by additive disturbances $\vec{w}_1 \in \mathcal{W}_1, \vec{w}_2 \in \mathcal{W}_2$ and $v_1 \in \mathcal{V}_1, v_2 \in \mathcal{V}_2$, which shall represent the model uncertainties. We assume that the system starts in a relaxed state $x_r(0) = x_h(0) = 0$, and without loss of generality, we set $x_r(0)$ as the origin of the variables x_r, x_h, and x_d. We can now apply Def. 1 to compute the reachable forces $f_{\text{coll},1} \in \mathcal{F}_{\text{coll},1}$ and $f_{\text{coll},2} \in \mathcal{F}_{\text{coll},2}$.

The Cartesian impedance controller is a convenient choice, since the resulting coupled dynamics are linear in the Cartesian spatial dimensions. Reachable sets of linear systems can be efficiently computed [23]. Generally, choosing other robot controllers is also possible, and the coupled dynamics can be derived similarly. Then, the systems are generally non-linear. The generalization into three-dimensional models is straightforward; the mass, spring, and damping parameters for both robots and humans are replaced by three-dimensional

matrices. The proj() operator is not needed anymore, reducing over-approximativity. The three-dimensional model is general, however, the number of parameters increases, which makes the model identification difficult. The number of states increases from 4 to 12 for the unconstrained collision model, which leads to a slower reachability analysis. A typical algorithm with zonotopic set-representation has complexity $\mathcal{O}(n^3)$ [24], where n is the number of states.

B. Reachset-conforming model identification

For our chosen interaction models in (7)–(11), only the parameters m_r, d_r, and k_r of the Cartesian impedance controller are known. The parameters m_h, d_h, and k_h, as well as the uncertainties $\mathcal{P}_1 = \{\mathcal{W}_1, \mathcal{V}_1\}, \mathcal{P}_2 = \{\mathcal{W}_2, \mathcal{V}_2\}$, are unknown.

The parameters are selected in a way that allows the reachable sets $\mathcal{F}_{\text{coll}}$ to include the behavior of the real system. We also refer to this property as *reachset conformance* [25]. We propose to ensure this property by means of testing the real system: from real collision experiments, we collect the inputs for our models, which are the initial states $\vec{z}_1(0)$, $\vec{z}_1(0)$, and u. We then make a forward prediction using a set of parameters and check if measured forces $f_m(t)$ are contained in $\mathcal{F}_{\text{coll}}(t)$ for all times. We wish to keep the reachable sets as small as possible.

Given m test cases, we formulate the identification as a constrained optimization problem minimizing the norm of the reachable sets, where $\mathcal{P}$ are the unknown parameters:

$$\min_{\mathcal{P}} \quad \sum_{1\le i\le m} \int_0^{t^*} \|\mathcal{F}^{(i)}_{\text{coll},\mathcal{P}}(t)\| dt, \tag{12a}$$

$$\text{subject to} \quad \forall i \, \forall t : f_m(t) \subseteq \mathcal{F}^{(i)}_{\text{coll},\mathcal{P}}(t). \tag{12b}$$

Because (7)–(11) are linear systems, the above optimization can be solved in a nested fashion, according to [25]: an inner loop computes the cost of optimal disturbances $\mathcal{W}$ and $\mathcal{V}$ using linear programming, given m_h, d_h, and k_h; an outer loop uses nonlinear programming to find m_h, d_h, and k_h with the smallest cost computed using the inner loop.

For safety analysis, reachset-conformant force predictions are sufficient. Requiring other variables (e.g., position trajectories) to be reachset-conformant would pose unnecessary constraints on the identification, which leads to more conservative models.

IV. Online Verification

This section describes our novel online verification procedure for our novel impact-force-limiting control, which always ensures the safety objective in (1). We first illustrate the fail-safe planning framework in Sec. IV-A, and then present our algorithm for evaluating $\text{safeForce}(t_k)$ in Sec. IV-B.

A. Fail-safe planning

The main idea of fail-safe planning [18], [26] is that during normal operation, the controller aims to generate and verify *fail-safe maneuvers*, as shown in Fig. 2. The next section

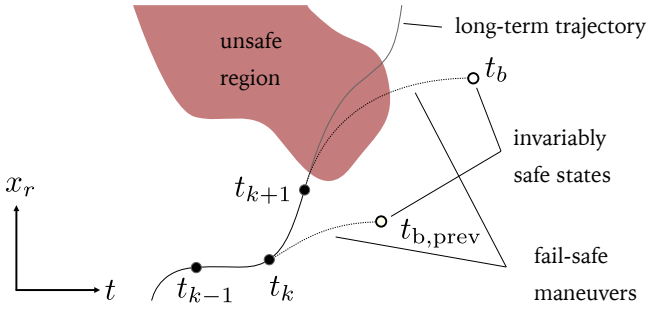

Fig. 2. *Fail-safe planning*: The robot is moving on the trajectory on $[t_{k-1}, t_k]$. The verification of the fail-safe maneuver on $[t_{k+1}, t_b]$ fails, because it passes an unsafe region. Therefore, the robot will execute the previously stored verified maneuver for $[t_k, t_{b,prev}]$.

of an intended *long-term trajectory* in the time interval $[t_k, t_{k+1}]$ can only be executed, if a consecutive and verified fail-safe maneuver to the invariably safe set $\mathcal{ISS}$ exists, while satisfying (1) upon reaching the $\mathcal{ISS}$. If a verified fail-safe maneuver cannot be found, then the previously verified one, starting at t_k, will be immediately executed, as can be seen in Fig. 2. If the robot has already been in a fail-safe maneuver during $[t_{k-1}, t_k]$, it is attempted to verify and then execute a *recovery maneuver* for $[t_k, t_{k+1}]$ to bring the system back to the long-term trajectory. Similar to [18], we limit ourselves to path-consistent fail-safe and recovery maneuvers [27] in this work to focus on the novel aspect of limiting forces.

Let us denote the time of reaching the $\mathcal{ISS}$ as t_b. A fail-safe maneuver for $[t_{k+1}, t_b]$ is verified by first checking the predicate noInteraction, by computing the reachable occupancies of the robot and the tracked human for the interval $[t_k, t_b]$. If the occupancies indicate a possible collision at $t_k \le t_c < t_b$, we also check the predicate safeForce for the interval $[t_c, t_b]$. If humans are not tracked, then we disregard noInteraction and directly evaluate safeForce by setting $t_c = t_k$. If an actual collision occurs, as measured by force sensors, then the robot brakes, until the force acting on the robot has vanished.

In practice, due to the interplay of intended trajectories, fail-safe trajectories, and recovery trajectories, the speed of the robot will always be as high as safely possible. Thus, it is not necessary to offline design a safe long-term trajectory.

B. Verifying compliance to impact-force limits

We present Alg. 1, which verifies at each time instant t_k that a fail-safe maneuver adheres to both transient and quasi-static force limits. Given is the maneuver $\vec{x}_d(t)$ for $t \in [t_k, t_b]$, which brings the robot to an $\mathcal{ISS}$. We first compute the reachable occupancies of the human $\mathcal{H}([t_k, t_b])$ and of the robot $\mathcal{M}([t_k, t_b])$, using the approach in [19], to detect possible future collisions, which trigger subsequent force evaluations. In case of a potential collision, we compute the set of reachable forces $\mathcal{F}_{\text{coll},1}$ for possible unconstrained collisions and $\mathcal{F}_{\text{coll},2}$ for possible constrained collision dynamics, as presented Sec. III-A. However, additional uncertainties have to be considered here:

Algorithm 1 Verification of safeForce(t_k)

Input: $\vec{x}_d(t), t \in [t_k, t_b]$
Output: `isSafe`

1: $\mathcal{M}(t), \mathcal{H}(t) \leftarrow$ (see [19])
2: find t_c, s.t. $\mathcal{M}([t_c, t_b]) \cap \mathcal{H}([t_c, t_b]) \neq \emptyset$
3: $\vec{x}_r(t) \leftarrow \vec{x}_d(t) + \mathcal{E}_r$ {assume tracking error bound}
4: $\dot{\vec{x}}_r(t) \leftarrow \dot{\vec{x}}_d(t) + \mathring{\mathcal{E}}_r$
5: $\mathcal{U} \leftarrow \texttt{proj}(\vec{u}([t_c, t_b]))$ {uncertain input set}
6: $\mathring{\mathcal{X}}_r \leftarrow \texttt{proj}(\dot{\vec{x}}_r([t_c, t_b]))$ {robot collision velocities}
7: $\mathring{\mathcal{X}}_h \leftarrow [-v_{\max}, v_{\max}]$ {maximum hand velocities}
8: $\mathcal{Z}_{0,1} \leftarrow 0 \times \mathring{\mathcal{X}}_r \times 0 \times \mathring{\mathcal{X}}_h$ {initial set for z_1}
9: $\mathcal{Z}_{0,2} \leftarrow 0 \times \mathring{\mathcal{X}}_r$ {initial set for z_2}
10: $\mathcal{F}_{\text{coll},1}(\tau) \leftarrow \texttt{reach}_1(\mathcal{Z}_{0,1}, \mathcal{U}, \mathcal{P}_1)$
11: $\mathcal{F}_{\text{coll},2}(\tau) \leftarrow \texttt{reach}_2(\mathcal{Z}_{0,2}, \mathcal{U}, \mathcal{P}_2)$
12: $f_{\text{tra}} = \sup(\mathcal{F}_{\text{coll},1}(\tau) \cup \mathcal{F}_{\text{coll},2}(\tau)) \quad \text{for} \quad 0 \leq \tau \leq t_e$
13: $f_{\text{qs}} = \lim_{\tau \to t_e} \sup(\mathcal{F}_{\text{coll},2}(\tau))$
14: **if** $f_{\text{tra}} \leq f_{\text{tra,lim}} \wedge f_{\text{qs}} \leq f_{\text{qs,lim}}$ **then**
15: `isSafe` $\leftarrow$ `true`
16: **else**
17: `isSafe` $\leftarrow$ `false`
18: **end if**

- Due to the acceleration capabilities of the human hand, we cannot predict its future velocity. Thus, we assume that it is bounded by an interval $\mathring{\mathcal{X}}_h := [-v_{\max}, v_{\max}]$. E.g., $v_{\max} = 2$ (m/s) complies with ISO 13855 [28].
- We assume that the robot position $\vec{x}_r$ and velocity $\dot{\vec{x}}_r$ are bounded by the errors $\mathcal{E}_r, \mathring{\mathcal{E}}_r \in \mathbb{R}^3$ around the desired trajectory before a collision.
- The collision time can be at any $t \in [t_c, t_b]$. Therefore, the collision speed of the robot is uncertain, but can be bounded by the union of all possible robot velocities $\mathring{\mathcal{X}}_r = \texttt{proj}(\dot{\vec{x}}_r([t_c, t_b]))$. Similarly, we bound the input by a $\mathcal{U} = \texttt{proj}(\vec{u}([t_c, t_b]))$.

The initial sets are defined as $\mathcal{Z}_{0,1} := 0 \times \mathring{\mathcal{X}}_r \times 0 \times \mathring{\mathcal{X}}_h$ and $\mathcal{Z}_{0,2} := 0 \times \mathring{\mathcal{X}}_r$, accounting for the above uncertainties. The operations in line 10–18 of Alg. 1 compute the reachable sets and evaluate the predicate safeForce from Sec. II.

V. Experimental results

This section experimentally evaluates our verified impact-force-limiting control for the interaction of a robot end-effector with the right hand of a human. The robot used is a Schunk LWA-4P lightweight robot using the Cartesian impedance controller from Sec. III-A, where the desired impedances are chosen as $\Lambda_r = 5I, D_r = 50I$, and $K_r = 150I$. To measure the impact force at the end effector, we designed a custom 3D-printed blunt impactor, which contains a 6-axis force-torque sensor. To measure the hand position and velocities, we use a Vicon Vero motion capture system. We show the experimental identification of the physical interaction model in Sec. V-A. We show the effectiveness of our controller in a human-robot co-existence scenario by comparing the robot performance with and without human tracking in Sec. V-B.

TABLE I
Identified parameters of unconstrained (up) and constrained (down) collision models

Dim.	$\mathcal{W}_1$	$\mathcal{V}_1$	Param.	Value
1	0.196	$[-79.27, 85.33]$	m_h	0.29
2	19.20	-	d_h	55.05
3	0	-	k_h	5434
4	0	-		
Dim.	$\mathcal{W}_2$	$\mathcal{V}_2$	**Param.**	**Value**
1	-0.498	$[-69.67, 38.37]$	d_h	719.3
2	5.459	-	k_h	29900

A. Results for model identification

We use our approach in Sec. III-B to identify reachset conforming model parameters. For that, two series of tests are conducted with the impedance-controlled robot, one for the unconstrained collision model, and the other for the constrained collision model. In the first experiment, multiple collisions of a hand with the end effector are initiated from random directions, and with random parts of the hand. In the second experiment, the robot collides with a resting hand on a table at different velocities. The hand is moved around in-between experiments, such that different parts of the hand are clamped. Due to safety reasons, we only did a reduced amount of tests, and these experiments were only conducted by the first author of this paper. Forty-three collisions have been evaluated for identifying the unconstrained collision model, whereas 41 collisions for the constrained collision model. The identified parameters are shown in Tab. I.

The results for a few randomly selected test cases are plotted in Fig. 3. For the unconstrained collision model, we only test until $t_e = 0.06$ seconds, since the impact transient has finished for all test cases at that time. For the constrained collision model, we test until $t_e = 0.5s$, because we are interested in the quasi-static force, to which our system converges. Regarding the values in Tab. I, we observe that the identified stiffnesses are smaller than in other works (e.g., [1]), and the damping values are high. The reason is that the identification algorithm decided that it is more effective (i.e., smaller reachable sets) to let the nominal parameters m_h, k_h, and d_h model the low-frequency dynamics, whereas high-frequency dynamics resulting from high stiffness and low damping are lumped inside the sets $\mathcal{V}_{1,2}$.

B. Results for the impact-force-limiting control

We demonstrate the effectiveness of the impact-force-limiting control using our online verification approach described in Sec. IV. We consider two scenarios. In the first scenario, we assume that human hand tracking is available to the controller, i.e., the collision time $t_c \geq t_k$. In the second scenario, tracking is not available, i.e., we set $t_c = t_k$. The robot moves between the joint angles $q_1 = [\frac{\pi}{2}, \frac{\pi}{6}, -\frac{\pi}{2}, -\frac{\pi}{2}, \frac{\pi}{2}, 0]^T$ and $q_2 = [-\frac{\pi}{4}, \frac{\pi}{6}, -\frac{\pi}{2}, -\frac{\pi}{2}, \frac{\pi}{2}, 0]^T$, for three times. At the first time, the human does not intervene. At the second time, the human intervenes without a collision, and at the third time with a collision. We set the transient force limit to 220 N, and quasi-static force limit

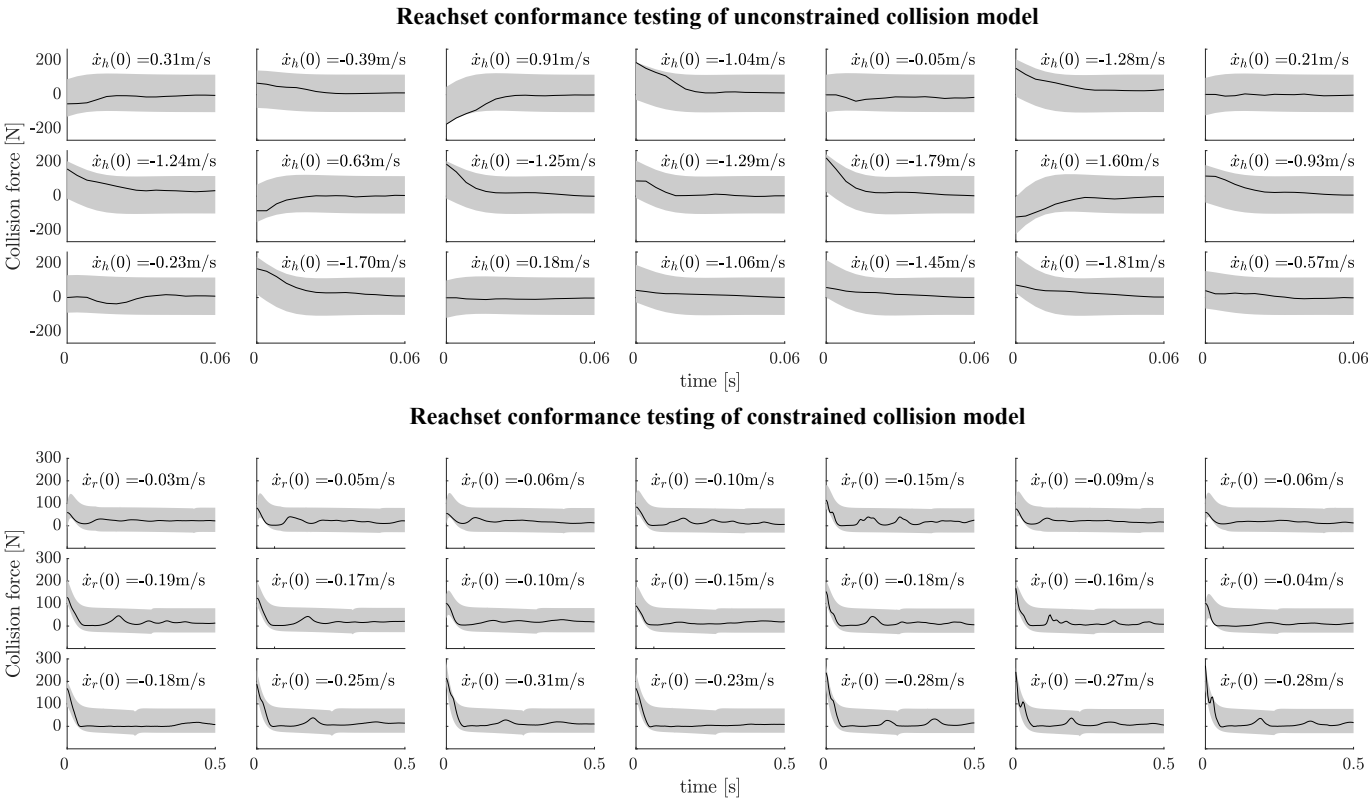

Fig. 3. *Reachset conformance testing of physical interaction models.* The model identified in Tab. I is reachset conformant. The reachable sets (gray) of of the models always over-approximate the force profiles (black) of real collision experiments. For each test case, the collision velocity is shown.

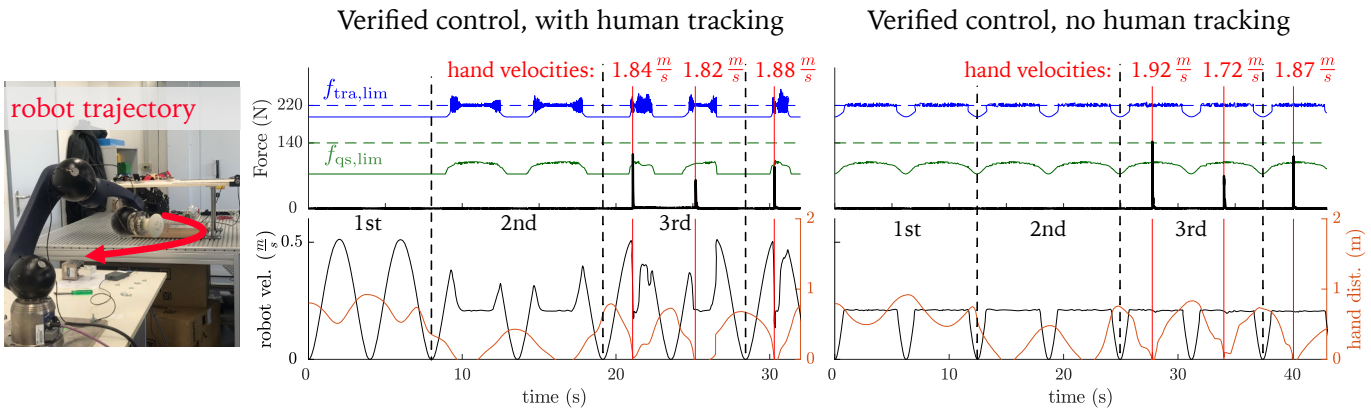

Fig. 4. *Verified impact-force-limiting control.* Upper graph: Transient force estimations are shown in green, quasi-static force estimations are shown in blue, force measurements are shown in **bold**, and collisions are shown in red. Lower graph: robot velocity is shown in black, and the relative distance of the hand to the robot is shown in orange All numbers are absolute values.

to 140 N. The resulting behavior of the robot is shown in Fig. 4. A video recording of the experiments is provided in supplementary materials.

For the scenario with tracking, we observe that in the first run, the robot can run at full speed, although the human-robot distance is closer than one meter. In the second run, we place our hand directly on the path of the robot. The robot automatically slows down because the reachable force predictions hit the transient force limit. However, as soon as the robot approaches a certain distance to the hand, it speeds up again. In the third run, we initiate collisions with the robot. The collision force never exceeds the estimated maximum transient.

For the scenario without human tracking, the results are similar, however, the robot remains at a slow speed for all times, and thus needs more time to complete its task. Human tracking benefits the efficiency of the robot.

VI. Conclusions and Future Work

This study presents the first work on guaranteeing impact-force limits during possible unintentional collisions between the human hand and robot end-effectors, despite uncertain-

ties. We believe that this concept can be extended to the entire human body and robot arm using similar coupled interaction models. The primary innovation is the prediction of the impact forces using reachability analysis, combined with a fail-safe motion planning. Our model identification method ensures that the interaction model is reachset conformant with the real interaction. In an experiment, we demonstrated that the impact force limit criterion allows robot motion, even if humans work closely to the robot. In addition, we have shown the advantages of tracking humans, which allows the robot to move faster when humans are distant to the robot. Multiple extensions to this work are possible:

- To extend this approach to the entire human body, the identification experiments need to be repeated for every body part. Additionally, high-volume testing and experiments on more diverse human tissues are needed to make sure that edge cases of the model are covered.
- We have not regarded the fact, that the robot closed-loop dynamics can be uncertain. In this study, such uncertainties were lumped inside the sets $\mathcal{W}_{1,2}$ and $\mathcal{V}_{1,2}$. Thus, our approach is only applicable to the controller used in the identification experiments. To verify variable impedance controllers, the uncertain dynamics of the robot and the human should be separately identified, and the coupling between these should be created online to analyze interaction forces.
- Online verification can also be combined with any other safety metric, i.e., by exchanging predicate $\mathrm{safeForce}(t)$ with power, energy, or safe velocity limits.
- An interesting extension is the verification of continuous physical interaction, where the dynamics of the human arm are also usually modeled as impedances.

ACKNOWLEDGMENT

The authors gratefully acknowledge partial financial support by the Central Innovation Programme of the German Federal Government under grants ZF4086004LP7, ZF4086012DB9, and the European Commission project CONCERT under grant number 101016007.

REFERENCES

[1] ISO/TS 15066:2016, "Robots and robotic devices - collaborative robots," Int. Org. for Standardization, Geneva, Switzerland, 2016.

[2] K. N. Shivakumar, W. Elber, and W. Illg, "Prediction of impact force and duration due to low-velocity impact on circular composite laminates," *J. of Applied Mechanics*, vol. 52, no. 3, pp. 674–680, 1985.

[3] Y. Yamada, Y. Hirasawa, S. Huang, Y. Umetani, and K. Suita, "Human-robot contact in the safeguarding space," *IEEE/ASME Trans. on Mechatronics*, vol. 2, no. 4, pp. 230–236, 1997.

[4] K. Ikuta, H. Ishii, and M. Nokata, "Safety evaluation method of design and control for human-care robots," *Int. J. of Robotics Research*, vol. 22, no. 5, pp. 281–297, 2003.

[5] J. Heinzmann and A. Zelinsky, "Quantitative safety guarantees for physical human-robot interaction," *Int. J. of Robotics Research*, vol. 22, no. 7-8, pp. 479–504, 2003.

[6] B. Navarro, A. Cherubini, A. Fonte, R. Passama, G. Poisson, and P. Fraisse, "An ISO10218-compliant adaptive damping controller for safe physical human-robot interaction," in *Proc. of ICRA*, 2016, pp. 3043–3048.

[7] Z. J. Li, H. B. Wu, J. M. Yang, M. H. Wang, and J. H. Ye, "A position and torque switching control method for robot collision safety," *Int. J. of Automation and Computing*, vol. 15, no. 2, pp. 156–168, 2018.

[8] R. J. Kirschner, N. Mansfeld, S. Abdolshah, and S. Haddadin, "Experimental Analysis of Impact Forces in Constrained Collisions According to ISO / TS 15066," in *Proc. of ISR*, 2021.

[9] S. Haddadin, A. Albu-Schäffer, and G. Hirzinger, "Requirements for safe robots: measurements, analysis and new insights," *Int. J. of Robotics Research*, vol. 28, no. 11-12, pp. 1507–1527, 2009.

[10] N. Mansfeld, M. Hamad, M. Becker, A. G. Marin, and S. Haddadin, "Safety map: A unified representation for biomechanics impact data and robot instantaneous dynamic properties," *IEEE Rob. Autom. Letters*, vol. 3, no. 3, pp. 1880–1887, 2018.

[11] A. Meguenani, V. Padois, J. Da Silva, A. Hoarau, and P. Bidaud, "Energy-based control for safe human-robot physical interaction," in *Int. Symp. on Exp. Robotics*. Cham: Springer, 2017, pp. 809–818.

[12] G. Raiola, C. A. Cardenas, T. S. Tadele, T. De Vries, and S. Stramigioli, "Development of a safety and energy aware impedance controller for collaborative robots," *IEEE Rob. Autom. Letters*, vol. 3, no. 2, pp. 1237–1244, 2018.

[13] M. Geravand, E. Shahriari, A. De Luca, and A. Peer, "Port-based modeling of human-robot collaboration towards safety-enhancing energy shaping control," in *Proc. of ICRA*, 2016, pp. 3075–3082.

[14] M. Angerer, S. Music, and S. Hirche, "Port-hamiltonian based control for human-robot team interaction," in *Proc. of ICRA*, 2017, pp. 2292–2299.

[15] S. Mitsch, K. Ghorbal, D. Vogelbacher, and A. Platzer, "Formal verification of obstacle avoidance and navigation of ground robots," *Int. J. of Robotics Research*, vol. 36, no. 12, pp. 1312–1340, 2017.

[16] S. Petti and T. Fraichard, "Safe motion planning in dynamic environments," in *Proc. of IROS*, 2005, pp. 2210–2215.

[17] M. Althoff, "An introduction to CORA 2015," in *Proc. of Workshop on Applied Verification for Cont. and Hybr. Systems*, 2015, pp. 120–151.

[18] D. Beckert, A. Pereira, and M. Althoff, "Online verification of multiple safety criteria for a robot trajectory," in *Proc. of IEEE Conf. on Decision and Control*, 2017, pp. 6454–6461.

[19] M. Althoff, A. Giusti, S. B. Liu, and A. Pereira, "Effortless creation of safe robots from modules through self-programming and self-verification," *Science Robotics*, vol. 4, no. 31, 2019, eaaw1924.

[20] S. Haddadin, A. Albu-Schäffer, and G. Hirzinger, "The role of the robot mass and velocity in physical human-robot interaction - Part I: Non-constrained blunt impacts," in *Proc. of ICRA*, 2008, pp. 1331–1338.

[21] C. Ott, *Cartesian Impedance Control of Redundant and Flexible-Joint Robots*, ser. Springer Tracts in Advanced Robotics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, vol. 49.

[22] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schäffer, K. Kosuge, and O. Khatib, "Progress and prospects of the humanrobot collaboration," *Autonomous Robots*, vol. 42, no. 5, pp. 957–975, 2018.

[23] M. Althoff, G. Frehse, and A. Girard, "Set Propagation Techniques for Reachability Analysis," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, 2021.

[24] A. Girard, C. Le Guernic, and O. Maler, "Efficient Computation of Reachable Sets of Linear Time-Invariant Systems with Inputs," in *Lecture Notes in Comp. Sci.*, 2006, vol. 3927 LNCS, pp. 257–271.

[25] S. B. Liu, B. Schrmann, and M. Althoff, "Reachability-based identification, analysis, and control synthesis of robot systems," 2021, arXiv:2103.01626.

[26] M. Althoff, S. Maierhofer, and C. Pek, "Provably-Correct and Comfortable Adaptive Cruise Control," *IEEE Trans. on Intelligent Vehicles*, vol. 6, no. 1, pp. 159–174, 2021.

[27] T. Kröger and F. Wahl, "Online trajectory generation: basic concepts for instantaneous reactions to unforeseen events," *IEEE Trans. on Robotics*, vol. 26, no. 1, pp. 94–111, 2010.

[28] ISO 13855:2010, "Safety of machinery positioning of safeguards with respect to the approach speeds of parts of the human body," Int. Org. for Standardization, Geneva, Switzerland, 2010.