Learning Force Control for Legged Manipulation

Tifanny Portela[12], Gabriel B. Margolis[1], Yandong Ji[13], and Pulkit Agrawal[1]

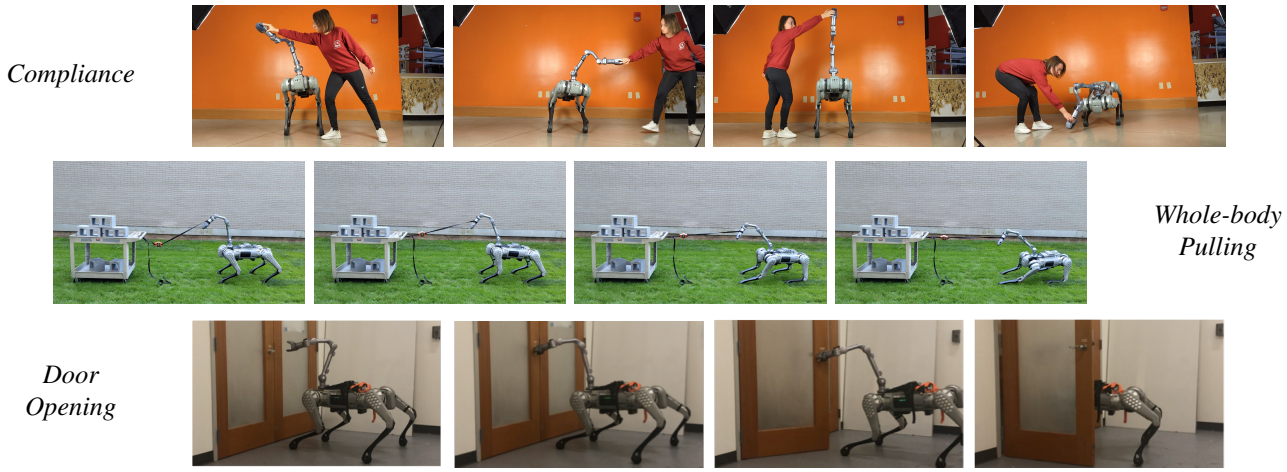

Fig. 1: We train whole-body policies to control *contact force* as well as for end effector position control in *noncontact* scenarios. The force control mode enables compliant interaction (top) and whole-body force application (middle). The position controller, shown opening a door (bottom), completes a pipeline for teleoperation of locomotion, grasping, and force-controlled manipulation.

***Abstract*— Controlling the contact force during interactions is an inherent requirement for locomotion and manipulation tasks. Current reinforcement learning approaches to locomotion and manipulation rely implicitly on forceful interaction to accomplish tasks but do not explicitly regulate it. This paper proposes a reinforcement learning task specification that focuses on matching desired contact force levels. Integrating force control with the coordination of a robot's body and arm, we present an end-to-end policy for legged manipulator control. Force control enables us to realize compliant gripper and whole-body pulling movements that have not been previously demonstrated using a learned policy. It also facilitates a characterization of the force-tracking performance of learned policies in simulation and the real world, indicating their performance potential for force-critical tasks. Video is available at the project website: `https://sites.google.com/view/learning-compliance`.**

I. Introduction

Large humanoids and quadruped manipulators will soon perform contact-rich tasks in complex environments or environments with humans. Reinforcement learning has recently accomplished state-of-the-art robotic control in domains of locomotion, manipulation, and drone flight [1]–[4]. For large and strong robots, it is important to regulate the amount of force they apply to their environment for safe and controlled interaction. However, existing learning-based formulations of loco-manipulation successful at tasks such as posture control, object manipulation, and fall recovery [5]–[7] do not explicitly regulate the amount of applied force. Compared to other forceful manipulation systems [8]–[12], typically learned policies use a coarse action space of position targets for a low-gain PD controller, a relatively low control frequency, and do not incorporate a force-torque sensor on the contact body. Regardless, reinforcement learning controllers are frequently used to generate forceful interactions such as during walking, running, parkour, fall recovery, object re-orientation, and others. In this work, we propose a task specification for reinforcement learning where the desired amount of contact force is directly commanded. Leveraging this specification, we train a policy capable of applying the desired force through a manipulator mounted on a large quadruped. We characterize the force tracking and estimation capability of such a system. We also investigate the ability of the quadruped to coordinate the body and legs to increase both its reach and the applied force magnitude compared to using the manipulator in isolation.

Force control underpins many classical robotic manipulation tasks such as pushing, pulling, grinding, drawing, wiping, and insertion and has been widely studied for both fixed-base and mobile manipulators [13]–[17]. Some recent works in legged locomotion achieved impressive performance using end-to-end reinforcement learning [6], [18]–[23]. This motivates revisiting force control obtained by coordinating the body and arm of legged manipulators using learning methods, in the hope of performing force-controlled tasks under challenging conditions where reinforcement learning excels, like in-the-wild terrains, using a single neural network policy, which allows computationally efficient accommoda-

Research was conducted in the Improbable AI Lab at MIT. Author affiliations: [1] Improbable AI Lab. [2] EPFL. [3] University of California, San Diego.

tion of the full system dynamics. Our work enables a legged manipulator to perform tasks that involve force control, from a compliant mode for kinesthetic demonstration or safe operation around humans to high-force tasks like pulling and lifting, by coordinating the entire body with an end-to-end policy.

We advance the science of legged manipulator control through the following contributions:

1) Propose a task specification for learning end-to-end policies for end effector force control (see Section IV).
2) Demonstrate the first deployment of learning-based whole-body force application control in legged manipulators (see Section V).
3) Characterize the effectiveness of force tracking and force estimation for learned policies in simulation and reality (Section V).

II. Related Work

A. Reinforcement Learning for Loco-Manipulation

Prior work has controlled a quadruped with mounted arm using sim-to-real reinforcement learning by formulating the task as simultaneous end effector position control and locomotion velocity control [5], [24]. Our work is highly influenced by [5], which was deployed on a real robot and displayed an increased workspace but was not suitable for forceful and compliant tasks due to the lack of force modeling during training, the competing constraint arising from explicitly commanding the gripper position relative to the body, and the smaller arm with a maximum payload rating of $0.2\,\mathrm{kg}$. Other works have learned a policy for a downstream task that requires forceful interaction, such as preventing and recovering from falls with a mounted arm [7] or dribbling a ball with the feet [6], [25]. These works demonstrate that forceful interaction is possible but optimize it in service of a single task rather than defining an interface that can be used to teleoperate forceful tasks. Moreover, without an explicit way of commanding the interaction force, we cannot repurpose these controllers for other force application tasks, and the precision of the force control in simulation and reality cannot be directly evaluated. Another line of work has used hierarchical architecture to push large objects using the robot's body, without an arm [26]. While the body and legs alone can perform useful environment interactions, a mounted arm can increase the workspace and allow the robot to control the direction and magnitude of force and torque application more precisely.

B. Forceful Control Primitives

For some manipulation tasks, particularly those involving contact interactions, it is hard to teleoperate by commanding position setpoints due to the unknown geometry of the contact surface or a desired degree of compliance in the motion. Instead, various parameterizations of control that regulate contact force have been proposed. Classic work on compliance and force control [13]–[16] established impedance control and hybrid force-velocity control. The purpose of these techniques is to define how the robot should reconcile force and position tracking errors given that these quantities cannot be independently controlled during contact. Force control methods may be implemented with closed-loop feedback from a wrist force sensor or with less precision by estimating the contact force from proprioceptive actuators and robot model [8].

One use case for force control is kinesthetic teaching, wherein a human physically interacts with a robot to demonstrate the movements and forces it should apply on the environment to accomplish a task [27], [28]. To receive a kinesthetic demonstration, it is necessary for the robot to comply with the operator's guidance and also estimate the amount of force exerted. This method of teaching has previously been considered challenging for legged robots and high-dimensional systems because of the cognitive challenge of reasoning about many degrees of freedom [29], which provides a motivation for developing low-level controllers that assume some control authority and ease the task.

C. Whole-body Control

Effective model-based approaches have been developed for controlling legged manipulators in a variety of tasks [9]–[12], [30]–[35]. Such works typically optimize a whole-body inverse dynamics model to stabilize a reference trajectory computed using model-predictive control or trajectory optimization. A few works have specialized in forceful interaction with a legged manipulator [9]–[12]. Early approaches [9] used trajectory optimization to generate open-loop behaviors for manipulating heavy objects and a separate online tracking controller to realize the motion. Subsequent work [10] incorporated contact point planning and control to manipulate objects using the entire body of a humanoid robot. More recently, ALMA [12] proposed a motion planning and control framework capable of coordinating dynamic and compliant locomotion and mobile manipulation based on a whole-body control task hierarchy. They demonstrated accurate contact force control through the end effector using the whole body. The ALMA system was also shown to support compliant behavior in service of a collaborative payload-carrying task.

III. Materials

Robot Quadruped: We implement our method on the Unitree B1 quadruped robot. This $55\,\mathrm{kg}$ robot stands $0.64\,\mathrm{m}$ tall. It has 12 identical electric actuators – each equipped with a joint position encoder – and an inertial measurement unit in its body to provide orientation. An onboard NVIDIA Jetson Xavier NX computer runs our control policy.

Robot Arm: We mount the Unitree Z1 robot arm above the B1's base. This 6 degree-of-freedom robot arm weighs $5.3\,\mathrm{kg}$ and has a maximum reach of $0.74\,\mathrm{m}$.

Teleoperation Joystick: To control the robot and arm, we use the Oculus Meta Quest 2 headset with its two controllers (Figure 2). The motion of the right controller is mapped to position commands for the end effector. The robot's body motion, force application, and position commands at the end effector are controlled by the thumb joysticks and buttons of both controllers.

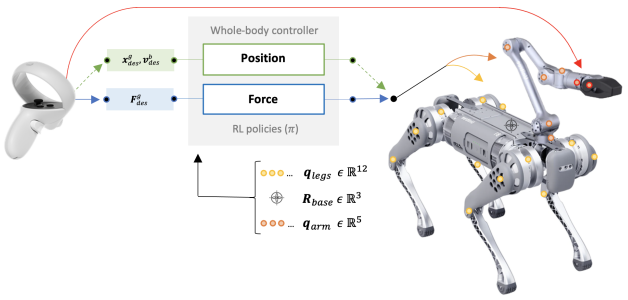

Fig. 2: System diagram depicting the command flow initiated by the Oculus controller. We implement two modes of learned control: a base velocity $\mathbf{v}^b_{des}$ and end-effector position $\mathbf{x}^g_{des}$ controller (green), and an end-effector force $\mathbf{F}^g_{des}$ controller (blue) for a robot with consistent foot-ground contact. While Oculus directly modulates the last two arm joints, the learned policy handles the 12 leg joints and the first five joints of the arm.

Simulator: We use NVIDIA Isaac Gym to collect simulated data to train all policies.

IV. Method

Our system is comprised of two types of learned policies. A force controller (Section IV-D) is the main focus of our work and enables compliant and force-controlled tasks. In order to apply forces, the robot first needs to walk to and grasp objects, so we implement an end-effector positioning and locomotion controller (Section IV-C) with a task specification inspired from prior work [5]. The controllers use a common policy architecture (IV-B) and action and observation space (IV-A) except the task-related command.

A. Action and Observation Space

Control policies act in a seventeen-dimensional space ($\mathbf{a}_t \in \mathbb{R}^{17}$), controlling position targets for a proportional-derivative controller in each of the robot's joints: thigh, calf, and hip joints of the B1 robot as well as the first five joints of the arm. The zero angles correspond to the robot's default standing pose with its arm raised. The sixth and seventh joints, controlling the roll orientation and opening angle of the gripper, are controlled independently using the right joystick of the Oculus headset.

The observation, denoted as o^t, consists of the base orientation $R^t_{base} \in \mathbb{R}^3$, the feet clock timings $c^t_{feet} \in \mathbb{R}^4$, the joint positions, $q^t \in \mathbb{R}^{17}$, the joint velocities, $\dot{q}^t \in \mathbb{R}^{17}$, the actions, $a^t \in \mathbb{R}^{17}$ and the previous actions $a^{t-1} \in \mathbb{R}^{17}$:

$$o^t = [R^t_{base}, c^t_{feet}, q^t, \dot{q}^t, a^t, a^{t-1}] \in \mathbb{R}^{75} \quad (1)$$

The observation history $o^{[t-H\ldots t]}$ is concatenated to the history of task-associated commands $t^{[t-H\ldots t]}_{cmd}$ as input to the policy. The dimension N_c of the task-related command varies among the whole-body position and force controllers and will be defined for each in the following sections.

B. Policy Architecture and Optimization

The policy consists of three modules: the actor-network, the critic network, and the state estimation module. First, the state estimation module predicts an estimate $\hat{e}_t$ approximating the privileged state e_t from the observation history. This style of concurrent state estimation can improve optimization performance [36]. Then, the actor-network inputs the observation history and the state estimate $\hat{e}_t$ and outputs the action. Separately, the critic network inputs the observation history and the true privileged state e_t. The estimator, actor, and critic network are multilayer perceptions with `elu` nonlinearities and hidden layer dimensions $[512, 256, 128]$, $[512, 256, 128]$, and $[256, 128]$ respectively. The state estimator is trained with a supervised loss while the actor and critic networks are optimized using Proximal Policy Optimization [37] with 4096 parallel environments.

C. End-Effector Positioning and Locomotion Task

The end-effector positioning and locomotion controller serves multiple purposes toward our main goal of forceful manipulation: (i) It supports walking to objects and grasping them. (ii) It also generates diverse high-quality initialization poses for the force controller. This controller should have a wide workspace and intuitive teleoperation interface for reaching any point in the environment. This controller reimplements elements from prior work [5] with some adaptation to our much larger robot.

1) Task definition: The end-effector positioning and locomotion controller takes two commands as input:

- The desired end effector position in the body frame, expressed in spherical coordinates: $(r, \theta, \phi)^{cmd} \in \mathbb{R}^3$.
- The desired linear velocity in the base frame x- and y-axes and the angular velocity command in the yaw axis: $v^{\text{cmd}}_b \in \mathbb{R}^3$.

This task-associated command $t_{cmd} = [(r, \theta, \phi)^{cmd}, v^{\text{cmd}}_b]$ is six-dimensional (i.e $N_c = 6$).

2) Reward design: The reward is expressed as $(\mathbf{r}^b_v + \mathbf{r}^g_x)e^{\mathbf{r}_l + \mathbf{r}_g}$ with separate terms for the body velocity task ($\mathbf{r}^b_v$), gripper position tracking task ($\mathbf{r}^g_x$), safety and smoothness criteria ($\mathbf{r}_l$), and gait pattern heuristic ($\mathbf{r}_g$) [21] (Table I). The exponential form from [36] ensures the reward will remain positive by lifting the negative penalty terms into the exponential.

A key feature of the proposed approach is that when the magnitude of the commanded base velocities falls below 0.1, the desired contact schedule for each foot is set to zero, ensuring all feet remain grounded. This characteristic prevents the robot from stepping in place for minimal velocities, easing the teleoperation of the arm.

3) Initialization and task sampling: The robot is initialized with joint angles randomized about a nominal standing pose. In each episode, the desired base velocity is sampled uniformly in the range $v_x \in [-1, 1\,\text{m/s}]$, $v_y \in [-1, 1\,\text{m/s}]$, $\omega_z) \in [-1, 1\,\text{rad/s}]$.

We adopt the end effector command parameterization from [5], where the end effector commands are defined in a body

Term	*Equation*	*Weight*
End-Effector Positioning and Locomotion (IV-C)		
$\mathbf{r}_x^g$: gripper position	$\exp\{-\lvert(r,\theta,\phi)-(r,\theta,\phi)^{\text{cmd}}\rvert/0.5\}$	5.0
$\mathbf{r}_v^b$: body velocity	$\exp\{-\lvert v_b - v_b^{\text{cmd}}\rvert/0.25\}$	1.0
Force Control (IV-D)		
$\mathbf{r}_f^g$: gripper force	$\exp\{-\lvert F - F^{\text{cmd}}\rvert/20\}$	5.0
Safety and Smoothness ($\mathbf{r}_l$)		
collision penalty	$1_{\text{collision}}$	-5.0
arm joint limit	$1_{q_a>0.9*q_a^{max}\lvert\lvert q_a<0.9*q_a^{min}}$	-3.0
leg joint limit	$1_{q_l>0.9*q_l^{max}\lvert\lvert q_l<0.9*q_l^{min}}$	-1.0
joint velocities	$\lvert\dot{q}\rvert^2$	$-8e{-}4$
joint acceleration	$\lvert\ddot{q}\rvert^2$	$-3e{-}7$
action smoothing	$\lvert a_{t-1} - a_t\rvert$	-0.05
–	$\lvert a_{t-2} - 2a_{t-1} + a_t\rvert^2$	-0.02
arm torque limit	$1_{\tau_a>0.8*\lvert\tau_a^{max}\rvert}$	-0.0015
Gait Pattern Heuristic ($\mathbf{r}_g$)		
swing phase	$\sum_{\text{foot}}[1 - C_i^{cmd}(t)]\exp\{-\lvert\mathbf{f}^{\text{foot}}\rvert^2/4\}$	0.9
stance phase	$\sum_{\text{foot}}[C_i^{cmd}(t)]\exp\{-\lvert\mathbf{v}_{xy}^{\text{foot}}\rvert^2/4\}$	4.0

TABLE I: Reward terms for learning position and forceful whole-body control.

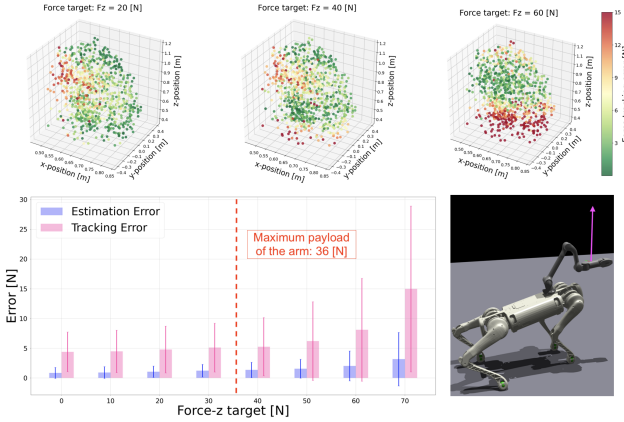

Fig. 3: Visualization of force tracking errors in simulation across all training setpoints for three target vertical forces: 20, 40, and 60 N. In the heatmap, elevated tracking errors are depicted in red, while optimal tracking is shown in green. As the target force increases, tracking accuracy diminishes, particularly for points proximate to the ground. Below is a bar plot illustrating the average force tracking and estimation errors for each target force across all training setpoints. The accompanying figure presents a robot with an external force applied at the gripper along the vertical axis, exemplifying the practical implications of these tracking errors.

height, roll, and pitch agnostic frame that is relative to the body in translation and yaw. At the start of each episode, the target gripper waypoint is initialized to the initial gripper position. Then, we repeatedly sample new waypoint positions from the range $r^{\text{cmd}} \in [0.3\,\text{m}, 0.9\,\text{m}]$, $\theta^{\text{cmd}} \in$ [-0.4π rad, 0.4π rad] and $\phi^{\text{cmd}} \in$ [-0.6π rad, 0.6π rad]. Each time a new position is sampled, we linearly interpolate the position target from the previous waypoint to the new one over a duration of four seconds.

4) Early episode termination: The maximum episode length is 20 seconds. In order to save computational resources and speed up the training process, early episode termination is introduced if the agent reaches a failure state. For this whole-body position controller, the episode terminates prematurely if the gripper mover is in collision or if the body height falls below 0.3m.

5) Concurrent state estimation: The state estimation target used for this controller consists of the robot's body velocity and the end effector position in the base frame.

D. Force Control Task

The force controller regulates the amount of force applied to the environment through the end effector. In contrast with end effector position control, the force control task offers some distinct benefits: (1) the robot can learn to be compliant and yield to external forces, even when carrying a payload, and (2) the robot's whole body and arm posture can adapt to facilitate force application without competing objectives.

1) Task definition: The force controller takes one command as input, the desired force vector at the end-effector $F^{\text{cmd}} \in \mathbb{R}^3$. This desired force is defined in Cartesian coordinates in the world frame and is sampled between -70 and 70 N for each axis. For this policy, the task-associated command, $t_{cmd} = [F^{\text{cmd}}]$, is three-dimensional (i.e $N_c = 3$).

Every 10 seconds, a force target is sampled, as well as a duration, t_F, which defines the duration of the force application and ranges between 2.0 and 4.0 seconds. The force command undergo linear interpolation from 0 to the sampled values in t_F seconds. Then, they maintain these values for a duration of 1.5 seconds, after which they are linearly interpolated back to 0 in t_F seconds. This process repeats every 10 seconds. To ensure that fully compliant behavior is experienced frequently, each time an environment is reset, it has a 20% chance of receiving a zero force target.

2) Reward design: The reward is expressed as $(\mathbf{r}_f^g)e^{\mathbf{r}_l+\mathbf{r}_g}$ with separate terms for the gripper force control task ($\mathbf{r}_f^g$), safety and smoothness criteria ($\mathbf{r}_l$), and gait pattern heuristic ($\mathbf{r}_g$) (Table I). To encourage firm stance during force application, the desired contact schedule for each foot is set to zero.

3) Initialization and task sampling: A key feature of our force control training is an emulated soft contact between the gripper and the environment, where the force application can be controlled by slight adjustments of the position. We simulate an external force $F_e \in \mathbb{R}^3$ that pulls on the gripper as a function of its displacement from a force setpoint. The force is modulated using a proportional-derivative control scheme on the position and velocity difference between the gripper position, $\mathbf{x}_g^t \in \mathbb{R}^3$ and the setpoint $\mathbf{x}_s \in \mathbb{R}^3$:

$$F_e^t = K_p(\mathbf{x}_s - \mathbf{x}_g^t) + K_d(\dot{\mathbf{x}_s} - \dot{\mathbf{x}_g}^t) = K_p(\mathbf{x}_s - \mathbf{x}_g^t) - K_d\dot{\mathbf{x}_g}^t. \quad (2)$$

The chosen gains, $K_p = 700$, $K_d = 6$, were tuned by inspecting the behavior of the simulator to ensure that applied force is large enough to bring the gripper to the setpoint but small enough to avoid extreme torques or oscillations.

If we initialize the force setpoint randomly, it may be unreachable or in collision with the robot's body. To avoid

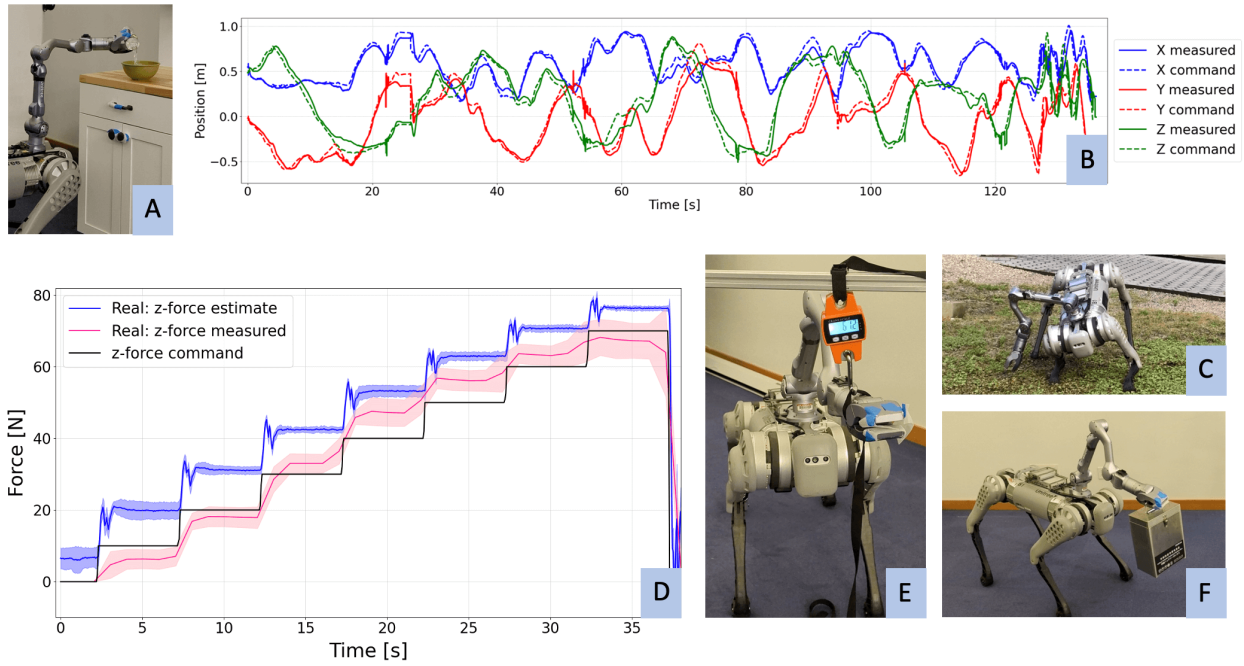

Fig. 4: Demonstrations of versatility and precision of the whole-body position controller include water pouring in a kitchenette (A), quantitative trajectory position tracking of the end-effector in Cartesian coordinates (B), and outdoor whole-body control (C). The effectiveness of the end-effector force controller is captured in its interaction with a dynamometer, with a graphical representation of the measured, commanded and estimated force over time (D) and a view of the actual setup (E). The force controller is able to hold a 4kg box, showcasing its gravity compensation capability (F)."

this, we instead initialize the force setpoint at the gripper's initial position. To obtain diverse initial positions, using the whole-body position controller (Section IV-C), we recorded the postures adopted by the robot to reach 1000 randomly sampled points in front of the robot in a restricted workspace with depth $40\,\mathrm{cm}$, width $80\,\mathrm{cm}$, and height $75\,\mathrm{cm}$. During this sampling procedure, each agent is initialized with a randomized pose. Then, the whole body position controller operates for fixed number of iterations, maintaining a constant end effector position command. If no collision is recorded during this period, the robot's current pose and joint configuration are saved. These saved postures serve as the initialization states for the training episodes of the whole-body force controller.

4) Concurrent state estimation: The state estimation target used for this controller consists of (1) the three-dimensional force applied to the gripper in the world frame and (2) the error between the targeted and actual joint position, which helped reduce the arm and leg torques.

5) Early episode termination: For the whole-body force controller, an episode terminates prematurely under several conditions:

- If the robot's calf, hip, arm, or gripper are in collision.
- If the body height drops below 0.3m.
- If the arm torques exceed 90% of their limits.
- If the distance between the end effector and the setpoint increases beyond 0.4 meters.

V. Results

We evaluate two separate learned force controllers: one to regulate force in the z-axis, another to regulate force in the xy-plane. We also characterize the performance of the end effector position and locomotion controller and show that it is sufficient to complete our system for forceful teleoperation.

A. Applying Large Forces

When lifting or pulling objects, an effective controller should realize a large applied force without undesired transients or inefficient postures that would result in the motor exceeding its safety limits. To evaluate this characteristic, the average force tracking and estimation errors for each target force across all training setpoints are reported in Figure 3. For z-axis force control, the mean absolute tracking error is below $5\,\mathrm{N}$ for low force targets and increases at high forces, while the estimation error constantly falls below $5\,\mathrm{N}$.

To evaluate the policy on the real platform, we attach our robot's end effector to a rope and gradually increase the downward force command through the entire training range (0-$70\,\mathrm{N}$). We use a dynamometer to record the applied force across five trials with the gripper in high, low, left, right, and middle positions (Figure 4-E). The applied force is within one standard deviation of the simulated error distribution (Figure 3, Figure 4-D). The estimated force tends to overshoot, suggesting a moderate sim-to-real gap.

To evaluate whether our controller can coordinate the body with the legs to increase the applied force, we initialized the arm directly in front of the robot, ramped the force command

in the x-axis, and recorded the highest applied pulling force (Figure 1). The observed force of 90 N is greater than the arm's rated payload of 36 N.

B. Applications: Compliance and Kinesthetic Teaching

1) Compliant end effector state: When the z-axis force controller IV-D is commanded to apply zero force, this corresponds to a fully compliant mode where the posture of the body and arm coordinate to drive the gripper force application to zero in all three axes. When released, the gripper remains suspended in place, with the system exhibiting gravity compensation. The supplementary video shows an operator manipulating the system to reach a variety of points.

Our result supports that fully compliant behavior is possible using the standard reinforcement learning architecture, which has not been previously shown. Realizing compliant behavior in learned motor policies is likely to improve the safety of robots around humans and robustness against unexpected disturbances. For example, a running quadruped that bumps its gripper into a wall or obstacle can reduce damage to itself or the environment. Compliance also makes it feasible to perform kinesthetic teaching, in which a human operator directly manipulates a robot's limbs in order to demonstrate a task without writing code or learning to use a teleoperation interface.

2) Compliant manipulation of heavy objects: By modulating the force application command in the z-axis, the compliant mode can be extended to the scenario where the robot is lifting an object with its arm. We tested the compliant mode with payloads of 0 kg, 2 kg, 4 kg, 6 kg as shown in the supplementary video and Figure 4-F. With a command force of zero, nonzero payloads cause the gripper to sink to the ground, which is expected since it should not apply a resistive force. With a zero force command, it takes substantial force for the human to lift the gripper with the object in grasp because the gripper will not apply any lifting force to the grasped object. Next, we increased the vertical force command to match the payload weight. For payloads up to 4 kg, this resulted in restored gravity compensation against the payload and compliance in all axes. With the highest payload of 6 kg, the gripper is less compliant and drifts to the center of the robot when released to the side.

C. End-Effector Positioning and Locomotion Performance

The policy for end-effector position control serves to establish an initial grasp or contact with the force application target. To manipulate objects that are higher or lower than the robot's body, we would like to achieve a large manipulation workspace through coordination of the body and arm. Prior work has shown this for a much smaller robot [5].

To measure the workspace of the controller, we send a sequence of commands at the limit of the training distribution and record the actual gripper positions achieved. As a baseline, we measure the workspace of the arm while the base is fixed in place by randomly sampling 3000 arm configurations and recording the resulting gripper positions. To characterize the workspace in each scenario, we fit a convex hull to the reached points and compute its volume. The whole-body controller improves the workspace by 71.5 %, demonstrating the significant benefit of whole-body coordination (Figure 5).

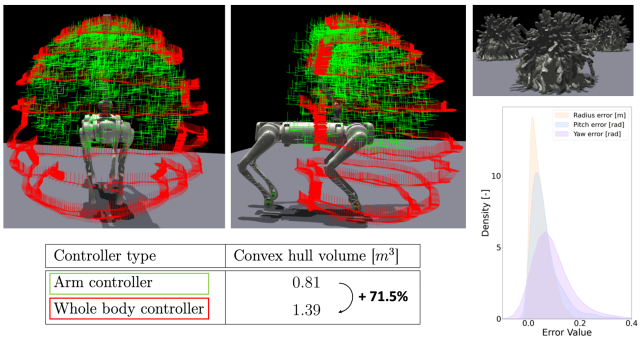

Fig. 5: Whole body position controller (red) enlarges the arm workspace (green) by 71.5%. The position error distribution in spherical coordinates shows the tracking accuracy, with its parallel evaluation method shown in the top right corner.

To grasp objects for manipulation, the policy for end-effector position control should also be accurate. We measure the distribution of error between the end effector position and the target position in spherical coordinates across a set of 2200 end-effector position commands, uniformly sampled from the training distribution. 90 % of error values fall below $(r, \phi, \theta) = (0.11\,\text{m},\ 0.19\,\text{rad},\ 0.24\,\text{rad})$.

The end-effector tracking performance is evaluated on hardware by teleoperating the arm across a variety of commands within the training distribution. We record the trajectory of commands as well as the trajectory of end-effector positions captured via a motion-tracking system. The path, expressed in Cartesian coordinates, can be seen in Figure 4-B. As shown, the commanded and actual end-effector positions align closely. For this trajectory, the average positional errors on the x, y, and z axes are 4.42 cm, 5.37 cm, and 6.86 cm, respectively. In a series of qualitative experiments, we successfully teleoperated the end-effector positioning controller for door opening (Figure 1) and water pouring (Figure 4-A).

VI. Discussion

We have demonstrated that learned whole-body manipulation policies can acquire a degree of compliance and perform force control at the end effector using only the minimal sensor configuration of joint encoders and body IMU. The force tracking performance is sufficient for some force-controlled tasks, including kinesthetic teaching with a weight and whole-body pulling. In quantitative experiments, we demonstrate good accuracy of a learned force estimator and the force application command tracking, with tracking performance degrading when the applied force reaches double the arm's rated payload. We also characterize the sim-to-real gap in our system. In combination with the end-effector positioning and locomotion controller, our work provides a teleoperation framework that is suitable for teleoperation and kinesthetic teaching of forceful loco-manipulation tasks.

In the future, it will be promising to explore hybrid control modes where the robot performs compliant trajectory tracking for improved teleoperation and imitation learning pipelines where the robot learns to accomplish compliant and forceful behaviors autonomously from demonstrations.

REFERENCES

[1] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Science Robotics*, vol. 4, no. 26, p. eaau5872, 2019.

[2] T. Chen, J. Xu, and P. Agrawal, "A system for general in-hand object re-orientation," in *Conference on Robot Learning*. PMLR, 2022, pp. 297–307.

[3] T. Chen, M. Tippur, S. Wu, V. Kumar, E. Adelson, and P. Agrawal, "Visual dexterity: In-hand reorientation of novel and complex object shapes," *Science Robotics*, vol. 8, no. 84, p. eadc9244, 2023.

[4] Y. Song, A. Romero, M. Müller, V. Koltun, and D. Scaramuzza, "Reaching the limit in autonomous racing: Optimal control versus reinforcement learning," *Science Robotics*, vol. 8, no. 82, p. eadg1462, 2023.

[5] Z. Fu, X. Cheng, and D. Pathak, "Deep whole-body control: learning a unified policy for manipulation and locomotion," in *Conference on Robot Learning*. PMLR, 2023, pp. 138–149.

[6] Y. Ji, G. B. Margolis, and P. Agrawal, "Dribblebot: Dynamic legged manipulation in the wild," *arXiv preprint arXiv:2304.01159*, 2023.

[7] Y. Ma, F. Farshidian, and M. Hutter, "Learning arm-assisted fall damage reduction and recovery for legged mobile manipulators," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 149–12 155.

[8] S. Chiaverini, B. Siciliano, and L. Villani, "A survey of robot interaction control schemes with experimental comparison," *IEEE/ASME Transactions on mechatronics*, vol. 4, no. 3, pp. 273–285, 1999.

[9] M. P. Murphy, B. Stephens, Y. Abe, and A. A. Rizzi, "High degree-of-freedom dynamic manipulation," in *Unmanned Systems Technology XIV*, vol. 8387. SPIE, 2012, pp. 339–348.

[10] M. Murooka, S. Nozawa, Y. Kakiuchi, K. Okada, and M. Inaba, "Whole-body pushing manipulation with contact posture planning of large and heavy object for humanoid robot," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 5682–5689.

[11] B. U. Rehman, M. Focchi, J. Lee, H. Dallali, D. G. Caldwell, and C. Semini, "Towards a multi-legged mobile manipulator," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 3618–3624.

[12] C. D. Bellicoso, K. Krämer, M. Stäuble, D. Sako, F. Jenelten, M. Bjelonic, and M. Hutter, "Alma-articulated locomotion and manipulation for a torque-controllable robot," in *2019 International conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 8477–8483.

[13] M. T. Mason, "Compliance and force control for computer controlled manipulators," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 11, no. 6, pp. 418–432, 1981.

[14] M. H. Raibert and J. J. Craig, "Hybrid position/force control of manipulators," 1981.

[15] T. Yoshikawa, "Dynamic hybrid position/force control of robot manipulators–description of hand constraints and calculation of joint driving force," *IEEE Journal on Robotics and Automation*, vol. 3, no. 5, pp. 386–392, 1987.

[16] N. Hogan, "Impedance control: An approach to manipulation," in *1984 American control conference*. IEEE, 1984, pp. 304–313.

[17] ——, "Impedance control: An approach to manipulation: Part ii—implementation," 1985.

[18] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, "Sim-to-real: Learning agile locomotion for quadruped robots," *arXiv preprint arXiv:1804.10332*, 2018.

[19] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science robotics*, vol. 5, no. 47, p. eabc5986, 2020.

[20] G. B. Margolis, T. Chen, K. Paigwar, X. Fu, D. Kim, S. Kim, and P. Agrawal, "Learning to jump from pixels," *Conference on Robot Learning*, 2021.

[21] G. B. Margolis and P. Agrawal, "Walk these ways: Tuning robot control for generalization with multiplicity of behavior," in *Conference on Robot Learning*. PMLR, 2023, pp. 22–31.

[22] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "RMA: Rapid motor adaptation for legged robots," *Robotics: Science and Systems*, 2021.

[23] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," *Science Robotics*, vol. 7, no. 62, p. eabk2822, 2022.

[24] S. Lee, S. Jeon, and J. Hwangbo, "Learning legged mobile manipulation using reinforcement learning," in *International Conference on Robot Intelligence Technology and Applications*. Springer, 2022, pp. 310–317.

[25] T. Haarnoja, B. Moran, G. Lever, S. H. Huang, D. Tirumala, M. Wulfmeier, J. Humplik, S. Tunyasuvunakool, N. Y. Siegel, R. Hafner *et al.*, "Learning agile soccer skills for a bipedal robot with deep reinforcement learning," *arXiv preprint arXiv:2304.13653*, 2023.

[26] S. Jeon, M. Jung, S. Choi, B. Kim, and J. Hwangbo, "Learning whole-body manipulation for quadrupedal robot," *arXiv preprint arXiv:2308.16820*, 2023.

[27] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 286–298, 2007.

[28] P. Kormushev, S. Calinon, and D. G. Caldwell, "Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input," *Advanced Robotics*, vol. 25, no. 5, pp. 581–603, 2011.

[29] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, "Recent advances in robot learning from demonstration," *Annual review of control, robotics, and autonomous systems*, vol. 3, pp. 297–330, 2020.

[30] H. Dai, A. Valenzuela, and R. Tedrake, "Whole-body motion planning with centroidal dynamics and full kinematics," in *2014 IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2014, pp. 295–302.

[31] L. Sentis and O. Khatib, "Synthesis of whole-body behaviors through hierarchical control of behavioral primitives," *International Journal of Humanoid Robotics*, vol. 2, no. 04, pp. 505–518, 2005.

[32] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 69–81, 2014.

[33] J.-P. Sleiman, F. Farshidian, M. V. Minniti, and M. Hutter, "A unified mpc framework for whole-body dynamic locomotion and manipulation," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4688–4695, 2021.

[34] M. P. Polverini, A. Laurenzi, E. M. Hoffman, F. Ruscelli, and N. G. Tsagarakis, "Multi-contact heavy object pushing with a centaur-type humanoid robot: Planning and control for a real demonstrator," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 859–866, 2020.

[35] J.-P. Sleiman, F. Farshidian, and M. Hutter, "Versatile multicontact planning and control for legged loco-manipulation," *Science Robotics*, vol. 8, no. 81, p. eadg5014, 2023.

[36] G. Ji, J. Mun, H. Kim, and J. Hwangbo, "Concurrent training of a control policy and a state estimator for dynamic and robust legged locomotion," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4630–4637, 2022.

[37] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

APPENDIX

A. Oculus: the teleoperation interface

In this work, the Oculus Meta Quest 2 Virtual Reality (VR) headset from Meta is used to teleoperate the robot with an attached arm. This VR headset comes with two controllers, one for each hand of the user. However, it's important to note that the visual aspect of the product is not employed in this project. Instead, only specific features, including the buttons and joysticks of both controllers, as well as the position tracking of the right controller are used to control the legged mobile manipulator.

In this setup, the VR headset remains static and is placed on a table. The position of the right controller in relation to the goggles defines the end-effector target position in the arm frame. This process is visually represented in Figure 7, where one can envision the arm's movement synchronized with that of the right controller.

Figure 8 illustrates the the buttons utilized to control the system. The gripper opening angle can be controlled using the Trigger A and Button A (highlighted in red in Figure 8). By pressing Trigger A, the gripper incrementally opens or closes by 10 degrees. On the other hand, Button A determines the direction of the angle increment, whether the gripper opens or closes. Similarly, the control of Joint 6 is achieved through the use of Trigger and Button B (highlighted in green in Figure 8)

The left and right joysticks (highlighted in blue in Figure 8) are used to define the base velocity of the B1 legged robot. More specifically, the left joystick defines the base linear velocities along the x (v_x) and y (v_y) axes and the right joystick controls the angular velocity around the z axis (v_{yaw}). Figure 9 illustrates this control system.

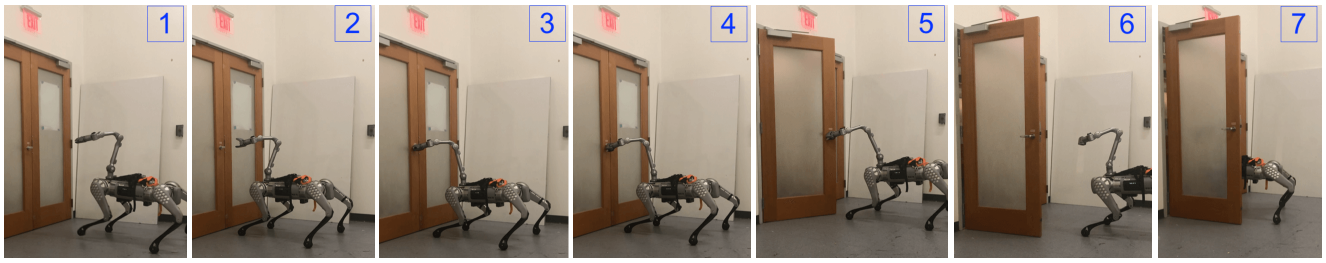

Fig. 6: Door opening

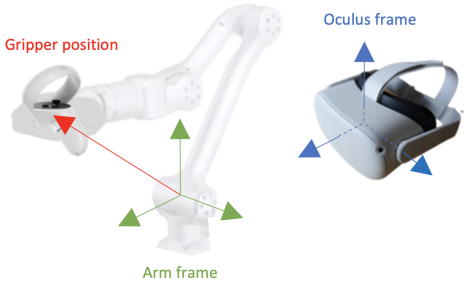

Fig. 7: Oculus controller to gripper position mapping

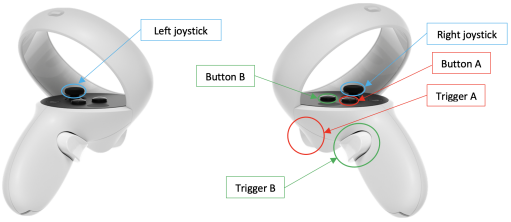

Fig. 8: Oculus controller button description

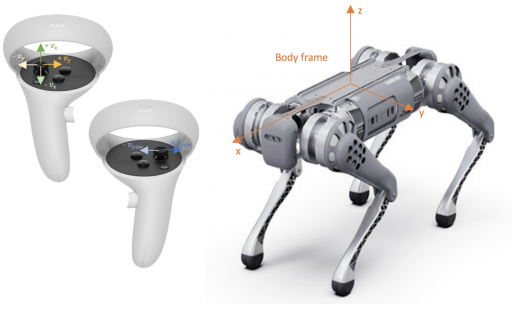

Fig. 9: Base velocity commands mapping

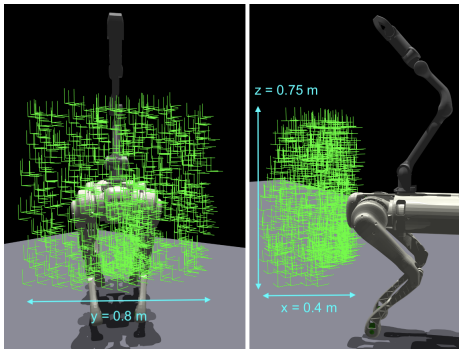

Fig. 10: Gripper setpoints employed for training the whole-body force controller

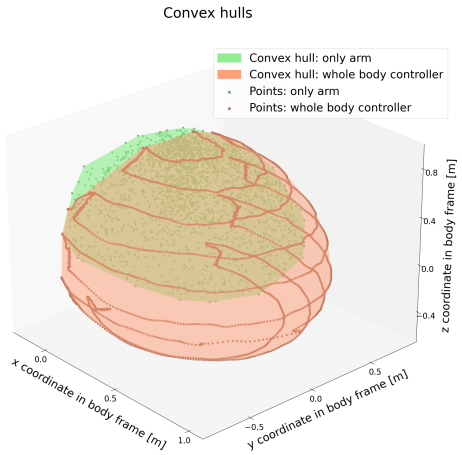

Fig. 11: Convex hulls

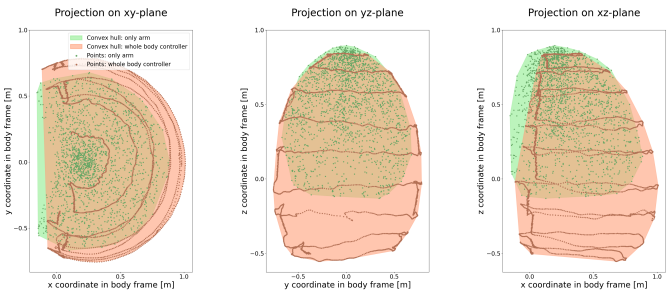

Fig. 12: 2D projection of the convex hulls