

# Cross-Entropy Optimization of Physically Grounded Task and Motion Plans

Andreu Matoses Gimenez <sup>1</sup>, Graduate Student Member, IEEE, Nils Wilde <sup>2</sup>, Member, IEEE, Chris Pek <sup>1</sup>, and Javier Alonso-Mora <sup>1</sup>, Senior Member, IEEE

**Abstract**—Autonomously performing tasks often requires robots to plan high-level discrete actions and continuous low-level motions to realize them. Previous TAMP algorithms have focused mainly on computational performance, completeness, or optimality by making the problem tractable through simplifications and abstractions. However, this comes at the cost of the resulting plans potentially failing to account for the dynamics or complex contacts necessary to reliably perform the task when object manipulation is required. Additionally, approaches that ignore effects of the low-level controllers may not obtain optimal or feasible plan realizations for the real system. We investigate the use of a GPU-parallelized physics simulator to compute realizations of plans with motion controllers, explicitly accounting for dynamics, and considering contacts with the environment. Using cross-entropy optimization, we sample the parameters of the controllers, or actions, to obtain low-cost solutions. Since our approach uses the same controllers as the real system, the robot can directly execute the computed plans. We demonstrate our approach for a set of tasks where the robot is able to exploit the environment’s geometry to move an object.

**Index Terms**—Task and motion planning, mobile manipulation, manipulation planning, parallel simulation.

## I. INTRODUCTION

**J**OINTLY planning tasks and motions for real-world robots presents significant challenges. It involves determining both what to do, i.e., the sequence of high-level actions, and how to do it, i.e., the specific motions associated with each action. This joint problem is usually complex as early actions can impact the performance of later actions or even the overall plan feasibility. Task and Motion Planning (TAMP) methods aim to simultaneously satisfy symbolic constraints (relations between discrete, high-level actions in the task plan) and geometric constraints (requirements on their continuous, physical realization) [1], [2].

Consider, for example, a mobile manipulator that needs to move an object from  $Table_1$  to  $Table_2$  and finally move to a dedicated *Exit* location, as shown in Fig. 1(a). In such a scenario,

Received 2 November 2025; accepted 23 March 2026. Date of publication 20 April 2026; date of current version 29 April 2026. This article was recommended for publication by Associate Editor H. Zhang and Editor O. Stasse upon evaluation of the reviewers’ comments. This work was supported by the European Union through the ERC project INTERACT under Grant 101041863. (Corresponding author: Andreu Matoses Gimenez.)

Andreu Matoses Gimenez, Chris Pek, and Javier Alonso-Mora are with the Department of Cognitive Robotics, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: A.MatosesGimenez@tudelft.nl; c.pek@tudelft.nl; j.alonsomora@tudelft.nl).

Nils Wilde is with the Faculty of Computer Science, Dalhousie University, Halifax, NS B3H 4R2, Canada (e-mail: nils.wilde@dal.ca).

Website and code: [andreamatoses.github.io/research/parallel-realization](https://andreamatoses.github.io/research/parallel-realization)  
This article has supplementary downloadable material available at <https://doi.org/10.1109/LRA.2026.3685463>, provided by the authors.

Digital Object Identifier 10.1109/LRA.2026.3685463

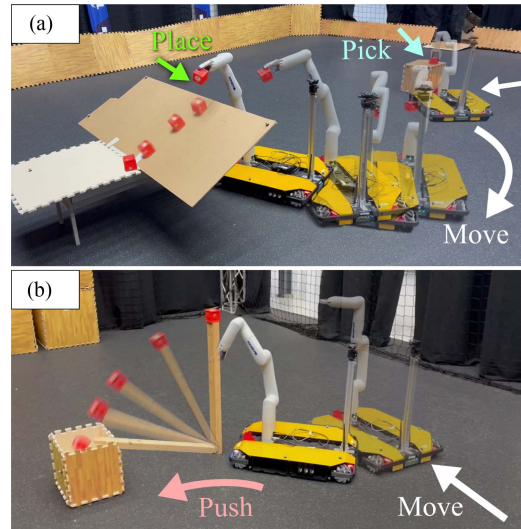


Fig. 1. (a) Mobile manipulator performing a sequential pick and place task. The robot can exploit the obstacle’s slope to find a faster solution, instead of going around. (b) Performing a move and push to put the block in the box.

the robot needs to move in sequence to  $Table_1$ , perform a grasping action, move to  $Table_2$  while holding the object, perform a placing action, and finally move to the *Exit*. To perform this autonomously, the robot would need to a) compute a high level plan, given by a *sequence of actions*, while considering its capabilities and its goal, and b) decide how to perform each action, by selecting a suitable *low-level controller*. The robot could, for example, approach  $Table_1$  from the right or from the left side, it could go around the obstacle in many different ways, or even better, it could exploit the geometry of the ramp-shaped obstacle to achieve its goal faster. However, the robot’s ability to decide how to perform actions is limited by the type of controllers available and the *action parameters* that define the motions generated by the controllers. For example, grasp actions can use full-body geometric controllers which take as input desired final end-effector poses. Move actions can use path planning algorithms such as  $A^*$  with a final position as input parameter.

Solving these TAMP problems is computationally expensive and could require extensive modeling. The robot has to come up with a sequence of actions for the plan, suitable action parameters, and also evaluate whether such a plan is feasible and efficient. Consequently, TAMP problems require the robot to consider the large combinatorial set of possible plans that accomplish the high-level task specification, as well as the set of possible action parameters. To deal with such large planning spaces in a computationally efficient way, most

TAMP algorithms use simplifications and abstractions, such as sampling-based discretization of the state space [3], simplified dynamics [4], or limited modeling of contacts [5], if any. Furthermore, sampling-based TAMP approaches typically rely on rejection-sampling “streams” (generators) to synthesize continuous parameters. While effective for geometric constraints, these streams become inefficient when valid parameters lie in narrow manifolds determined by complex dynamics.

These assumptions and simplifications come at the cost that the resulting plans may not adequately account for the dynamics or complex contacts necessary to reliably perform the task. Additionally, when ignoring the effect of the robot’s low-level controllers, such approaches may obtain suboptimal or even infeasible plans for the real system.

We investigate the use of parallelized physics simulators [6] to compute low-cost realizations of a plan with realistic controllers. The combination of such simulations with efficient and parallelizable controllers allows us to consider the dynamics, complex contacts, and limitations of the controllers themselves when evaluating plan realizations [7], [8]. We can see in Fig. 1(a) how the robot takes advantage of the ramp to let the block slide to the table, without the need to go around, and Fig. 1(b), where the robot is able to push a stick to let the block fall into a box.

To this end, we formulate the task planning problem using the widely used Planning Domain Definition Language (PDDL) [9] due to its expressivity and the availability of off-the-shelf symbolic planners. We then associate appropriate controllers and their input parameters to the operators used in the PDDL formulation. We explicitly frame the synthesis of these controller parameters as a “Stream” (or generative sampling strategy). However, instead of using standard rejection sampling, we implement this stream using Cross-Entropy optimization to actively search for feasible and low-cost parameters. This enables the discovery of plan realizations that exploit complex dynamics and contacts without significantly increasing the modeling effort at the symbolic level. Since our approach uses the same controllers as the real system, the robot can directly execute our realized plans.

### A. Related Work

There are a variety of integrated TAMP solvers [2], [10], which can be categorized into constraint-based, sampling-based, or optimization-based TAMP solvers.

Constraint-based TAMP [11] uses general-purpose Satisfiability Modulo Theory (SMT) solvers, where motion constraint information is incrementally added at the task level. These methods focus on feasibility satisfaction of the resulting plan, and require detailed knowledge about the symbolic and geometric representations of the world, which need to be sufficient to ground a given plan. Optimization-based TAMP solvers often divide the problem into hierarchical feasibility (and optimality) stages with progressively more detail to filter the space of symbolic plans. In particular, [1], [12] use tree search over sequences of symbolic actions and evaluate the feasibility and optimality of each node with increasing detail as an heuristic. Detailed motion calculation is based on carefully crafted mathematical constraints and kinematic models that can be leveraged by specialized optimizers [13], [14]. For some problems these solvers may fail to find a solution, and computational cost can increase prohibitively if more detail is considered, such as dynamics and contacts.

Sampling-based TAMP methods focus on efficiently generating action parameters which can then be used at the task planning

stage. In [3], [15], conditional samplers are used in combination with Planning Domain Definition Language (PDDL) [9] symbolic planners to incrementally generate PDDL sub-problems from discretized action parameters. [4] introduces an optimal sampling-based approach using asymmetric bidirectional sampling, which significantly improves the efficiency of finding action parameters for plans without the need for discretization. These methods require the creation of domain-specific samplers that generate constraint-satisfying configurations of the robot and objects to handle complex dynamic interactions. This requires considerable engineering effort when adapting to new problems without significant simplifications. For instance, analytically modeling the physical constraints for a single, generic “place” action that covers all possible dynamic outcomes is often intractable. Consequently, prior methods typically require discretizing these interactions into finer-grained, explicitly modeled actions (e.g., requiring both *place\_flat* and *place\_sloped*). Recent work also explores the use of sampling-based optimization to find plan realizations solving a non-linear program [16], [17]. [16] explores the use of specialized MCMC sampling strategies and how they could potentially be used in the context of TAMP. [17] investigates the use of evolutionary strategies to generate high-quality samples of these complex distributions. These non-linear sampling-based optimization approaches are particularly promising in TAMP settings where solutions lie in multi-modal, disconnected regions of the parameter space.

There is growing interest in integrating closed-loop controllers into TAMP solvers to facilitate real-world deployment. In [18] a sequence of actions is executed by switching between MPC controllers with the corresponding sets of constraints. Once the set of goal states are deemed unreachable, a backtracking mechanism brings the execution back to the last action from which the plan is feasible. Most recently, [19] utilized pre-defined behaviors to handle informational uncertainty (e.g., occlusions). In contrast, we address dynamic complexity. By optimizing generic actions via physics simulation, our framework automatically discovers complex interactions (e.g., sliding) without requiring them to be explicitly encoded as separate symbolic behaviors. In [20], a framework for uncertainty-aware TAMP is presented. The uncertainty in the execution of the controllers is estimated by simulating the actions and recording successes. These approaches can be implemented in real-world scenarios, as closed-loop controllers are expected to compensate for disturbances arising from unmodeled complex dynamics and contact interactions. However, while these bring TAMP solvers closer to practical deployments, planning while considering these effects remains a challenge.

### B. Contributions

This paper contributes a task and motion planning method that accounts for the robot’s dynamics and physical constraints, environment, contacts, and controllers. We

- 1) combine PDDL [9] with information on low-level controllers and required parameters to formulate symbolic tasks plans, treating the parameter synthesis as a closed-box stream;
- 2) propose a cross-entropy optimization algorithm to optimize the parameters of the controllers that realize the candidate plan and minimize the cost function. Cross-entropy optimization is performed in a simulator to consider dynamics, contacts, and the controllers’ limitations when finding plan realizations;
- 3) demonstrate the method on both a pick-and-place and pushing task in a simulator and on a real robot. Our results

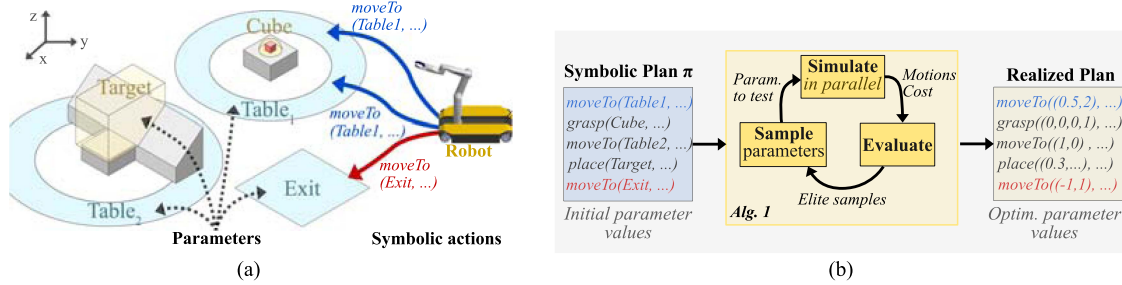


Fig. 2. Overview of our TAMP framework. (a) The scenario specification involves defining symbolic actions the robot can perform, e.g., moving to something, and parameters for the actions, e.g., a region corresponding to Table1 or the Exit. Note that parameterized actions such as *moveTo(Table1, ...)* may result in different motions, shown in blue. (b) We optimize the parameters of the symbolic plan using parallel simulations and cross entropy optimization.

show that we can find solutions that take advantage of contact with the environment, which would require explicit modeling in traditional methods.

## II. PROBLEM FORMULATION

We consider TAMP problems, where a mobile manipulator needs to execute a sequence of  $k$  actions, performed by low-level controllers, to achieve a high-level goal while minimizing a cost function. Each controller requires the input values of several parameters to fully define the resulting robot motion. Additionally, we focus on TAMP problems where the robot may be able to exploit the environment’s geometry and dynamics, e.g., by making contacts, to find low-cost plans.

The configuration space of the system is  $\mathcal{X} = \mathcal{Q} \times SE(3)^m$ , where  $\mathcal{Q}$  denotes the configuration space of the robot, and there are  $m$  objects in the environment. We use  $\mathbf{x}_t \in \mathcal{X}$  to denote the state of the whole system at time  $t$ . We assume that the problem can be solved within a time horizon  $T$ , if a solution exists. The task specification  $\psi_g$  encodes the goal, which is achieved if  $\psi_g(\mathbf{x}_T) = \text{True}$  for the final time  $T$ . Formally, the problem is stated as follows.

**Problem 1 (Sequence of Controllers):** Given a set of controllers  $\Gamma$ , cost functions (1a), goal specification  $\psi_g(\mathbf{x}_T)$ , and system dynamics (1b), we aim to find a sequence of  $k$  controllers  $c_{1:k}$ , each selected from  $\Gamma$ , and their parameters  $z_{1:k}$  that solve the following optimization problem:

$$\min_{c_{1:k}, z_{1:k}} \sum_{i=0}^k \sum_{t=t_{0,i}}^{T_i} J(x_t, \dot{x}_t, c_i, z_i) + J_{\text{end}}(x_T) \quad (1a)$$

$$\text{s.t. } x_{t+1} = f(x_t, c_i, z_i), \quad (1b)$$

$$\psi_g(\mathbf{x}_T) = \text{True} \quad (1c)$$

where  $t_{0,i}$  and  $T_i$  are the initial and final time of controller  $c_i$  execution. The next controller’s execution will start thereafter, i.e.,  $t_{0,i+1} = T_i + 1$ . The number of controllers  $k$  is not known a priori, and assumed to be bounded.

## III. METHOD

Solving Problem 1 requires simultaneously finding the discrete sequence of controllers  $c_{1:k}$  and their continuous parameters  $z_{1:k}$  in a high-dimensional planning space. Further, we must consider non-linear dynamics, as well as discontinuities introduced by contact events and controller switches. These

additional considerations make it intractable to use direct search methods to find the sequence of controllers that achieve the goal specification.

To address this issue, we propose to divide the problem into two levels, illustrated in Fig. 2. First, an abstracted symbolic task planning step, where we define the robot’s actions and associated parameters for the task, and find a candidate sequence of actions and their associated controllers. This step effectively provides a good candidate for the assignment of  $c_{1:k}$  in Problem 1, and narrows down the parameters. We refer to this candidate plan as symbolic plan  $\pi$ . Secondly, we perform cross-entropy optimization to determine values for the parameters that render the plan feasible with low cost. We call this step finding a plan realization, which provides values to the parameters  $z_{1:k}$  consistent with the task plan. In this work, we assume that the candidate symbolic plan is physically feasible. We define high-level actions broadly (e.g., a generic `place` action), leaving the complex geometric details to be solved by the optimizer. This generality maximizes the chance of finding a working solution. While our optimization method could theoretically search over multiple candidate plans simultaneously, we restrict the current scope to a single plan to maximize the computational efficiency of the parallel physics simulation.

### A. Symbolic Task Planning

The objective of the symbolic task planning step is to provide a candidate symbolic plan  $\pi$  that is 1) general enough to allow many possible realizations, and 2) informative to the subsequent plan realization step. To this end, we deliberately model the set of possible actions as general as possible. For example, in Fig. 1, we consider actions such as “move”, “pick”, or “place”, without specific references to the environment, e.g., to objects or regions. Thus, without requiring expert knowledge of the environment.

We use PDDL [9] to model the symbolic planning problem. This choice gives us an action-focused notation with flexibility in the level of detail considered in the abstractions. However, we need to add additional information to distinguish between different parameters for the actions. More formally, we define the task planning problem with:

*a) Parameters:* We define the set of symbolic parameters  $\mathcal{Z} = \{z_1, z_2, \dots\}$ . Each parameter  $z_j \in \mathcal{Z}$  is defined as a tuple  $z_j = \langle \text{sym}_j, \mathcal{M}_j, \mathcal{C}_j \rangle$ , where:  $\text{sym}_j$  is the unique identifier (e.g., `Table1`, `Cube`, `Start`) used in the PDDL domain.  $\mathcal{M}_j$  is the underlying manifold (e.g.,  $\mathbb{R}^2$ ,  $SE(3)$ ) defining the continuous space of the parameter. We use  $\perp$  to denote a parameter that requires no realization, indicating it is used solely for symbolic

logic and not as an optimization variable.  $\mathcal{C}_j$  is a set of constraint functions  $h(z) \leq 0$  defining the valid range or region (e.g., circular region around a table). During symbolic planning,  $z_j$  is treated as a discrete token. During plan realization, we optimize over values in  $\mathcal{M}_j$  only for parameters where  $\mathcal{M}_j \neq \perp$ .

b) *Action Schema*: The robot can execute different controllers  $c_i$ . The parameters of a controller result in different actions, e.g., the different *moveTo* actions in Fig. 2(a). We define action schemas as abstractions of high-level skills (e.g., grasping or placing) and the associated controllers and parameters. Action schemas  $\alpha_i$  are defined by the tuple

$$\alpha_i = (\mathcal{Z}_{\alpha_i}, c_{\alpha_i}, J_{\alpha_i}(\mathbf{x}_{t_{0,i}:T_i}), \text{success}_{\alpha_i}, \psi_{\text{pre}}, \psi_{\text{eff}}), \quad (2)$$

where  $\mathcal{Z}_{\alpha_i} \subseteq \mathcal{Z}$  is the subset of parameters consistent with the action's type of input. For example, if the action schema uses a point in  $\mathbb{R}^2$  for its tracking controller, the type of parameter must also be in  $\mathbb{R}^2$ .

The  $c_{\alpha_i}$  uses the parameters to calculate the action motion and control the robot. The action schema has a cost  $J_{\alpha_i}(\mathbf{x}_t)$ , which can be continuously evaluated or only once when the action is finished. We can determine an action's success by evaluating the Boolean function  $\text{success}_{\alpha_i} \in \{\text{True}, \text{False}\}$ , depending on the executed motion.

Following the PDDL convention,  $\psi_{\text{pre}}$  is the set of precondition predicates, which are evaluated on the symbolic parameters. The action can be executed if the preconditions are met. Similarly  $\psi_{\text{eff}}$  is the set of effect predicates that abstract the effects of a controller in the environment by describing which predicates change value after execution. By chaining actions of compatible effects and preconditions, we can limit the possible sequences of actions to the ones that are at least feasible at the abstract level.

As the plan's feasibility is only verified in the realization step, see Section III-B, we do not require the  $\psi_{\text{pre}}$  and  $\psi_{\text{eff}}$  predicates to accurately model complex physical interactions in the environment. For example, when performing a grasping action, we do not enforce that a grasp is physically stable. Instead we limit the predicates to model logic at the symbolic level, such as not being able to perform the grasp action if the robot gripper is not empty.

When a compatible assignment of parameters  $\mathcal{Z}_{\alpha_i} \subseteq \mathcal{Z}$  is made to instantiate an action schema, we obtain a symbolic action  $a_i$ . As an example, consider the symbolic actions resulting from a different choice of parameters for the schema: *moveTo*( $\cdot$ , *Exit*), *moveTo*( $\cdot$ , *Table*<sub>1</sub>). The action is *realized* when a value is chosen for its parameters  $\mathcal{Z}_{\alpha_i}$ .

A symbolic plan is a sequence of symbolic actions  $\pi = (a_1, \dots, a_k)$  with its associated parameters,  $\mathcal{Z}_\pi = (\mathcal{Z}_{\alpha_1}, \dots, \mathcal{Z}_{\alpha_k})$ . We use a standard symbolic planner [21] to generate the symbolic plan that accomplishes goal predicates  $\psi_g$ , which we assume models  $\psi_g(\mathbf{x}_T)$  at the symbolic level.

## B. Plan Realization

Once the solver generates a candidate symbolic plan, we need to find a feasible and low-cost realization of the actions to execute the plan on the real robot. A plan is realized when all its actions are realized, and it is feasible if all actions can be executed and the goal specification  $\psi_g(\mathbf{x}_T)$  is achieved. The total cost of a realized plan is the sum of all realized action costs plus a final cost.

*Problem 2 (Symbolic Plan Realization)*: Given a symbolic plan  $\pi = a_{1:k}$ , goal specification  $\psi_g(\mathbf{x}_T)$  and system dynamics

$f$ , find a realization of  $\mathcal{Z}_\pi = \mathcal{Z}_{\alpha_{1:k}}$  that solves

$$\begin{aligned} \min_{\mathcal{Z}_{\alpha_{1:k}}} \quad & J_{\text{end}}(\mathbf{x}_T, \mathcal{Z}_{\alpha_{1:k}}) + \sum_{i=1}^k \sum_{t=t_{0,i}}^{T_i} J_{\alpha_i}(\mathbf{x}_t) \\ \text{s.t.} \quad & \text{success}_{\alpha_i} = \text{True}, \forall \alpha_i \in \pi, \\ & \psi_g(\mathbf{x}_T) = \text{True}, \\ & \mathbf{x}_{t+1} = f(\mathbf{x}_t, a_i), \end{aligned}$$

where  $t_{0,i}$  and  $T_i$  represent the start and end time of  $\alpha_i$ .

In order to optimize the values of the parameters  $\mathcal{Z}_\pi$  of a given plan  $\pi$  in Problem 2, we use Cross-Entropy optimization (CE) [22]. The method iteratively refines the estimation of the optimal parameter values by using importance sampling.

In CE optimization, we assume that there exists a target distribution over the parameter space from which samples have a high chance of being low cost. Our aim is to move an approximated distribution closer to this target distribution after each iteration, to generate low cost parameter samples. To this end, we assign a probability distribution to each parameter of an action  $\mathbf{z}_k \sim \phi(\theta_k) \forall \mathbf{z}_k \in \mathcal{Z}_{\alpha_i}$  parametrized by  $\theta_k$ . Let  $\phi(\theta_\pi)$  be the resulting stacked vector of probability distributions of all the parameters of a plan  $\pi$ . For convenience, we refer to sampling all the parameters of the plan as sampling the vector  $\mathcal{Z}_\pi \sim \phi(\theta_\pi)$ .

After sampling an  $n_{\text{env}}$ -sized batch of parameter values  $\mathcal{Z}_\pi$ , we roll out the plans in parallel using a physics simulator, and subsequently evaluate the samples feasibility and performance. The top  $N_e$  elite samples, that is the feasible ones with lowest cost, are used to find a new  $\theta_\pi$  such that the KL divergence with respect to the target distribution is minimized [22]. With the updated approximated distribution, we can sample again and repeat until convergence. This approach allows for gradient free optimization and to consider the discontinuities and nonlinearity of the dynamics.

Algorithm 1 outlines our CE optimization algorithm. The initial sampling stage is critical for the algorithm's performance, as ideally it provides enough feasible samples representing the potential multi-modal realizations. Consider, for example avoiding an obstacle from the left, or from the right; these are two realizations that may exist in disconnected regions of the parameter spaces. For this reason, the initial distribution  $\phi(\theta_\pi^0)$  used in Algorithm 1 is uniformly distributed inside the region of the parameter  $h(z_k) \leq 0$ , which acts as an initial guess of areas where solutions may exist. Fig. 2(a) illustrates examples of such parameter regions, e.g., a circular approach region around the table. In subsequent iterations of the algorithm, the constraints on parameter values may not be enforced, as we assume that the next samples will not deviate excessively from the previous feasible samples. In line 1, if the number of feasible solutions  $n_{\text{feasible}}$  is lower than  $N_e$ , only the  $n_{\text{feasible}}$  samples are used to update the distribution. If only one solution is found, the successful sample is used with added artificial noise to enforce exploration around the sample in the subsequent iterations.

In cases where an action's parameter space is high-dimensional and its region is poorly chosen, such as grasp poses, we may not find feasible plan realizations during the initial sampling step. Even if a valid grasping pose is sampled, it may not satisfy  $\psi_g(\mathbf{x}_T)$  in combination with the other realized plan parameters. In these cases, we use the plans where at least all actions are feasible, i.e.,  $\text{success}_{\alpha_i} = \text{True}, \forall \alpha_i \in \pi$ , for the update of the sampling distribution. We hypothesize

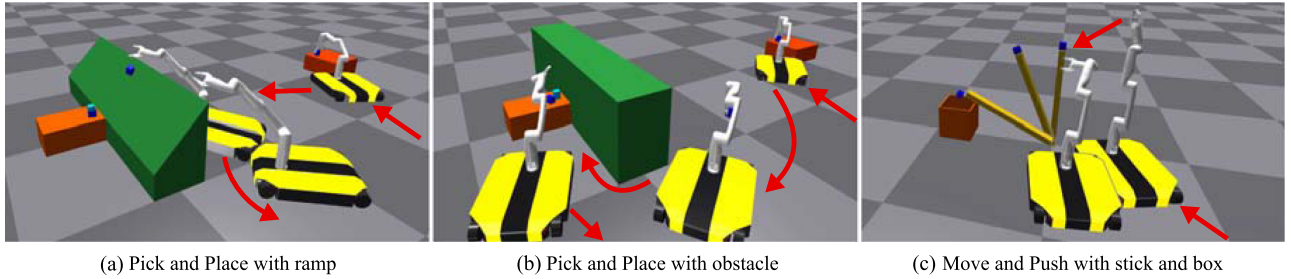


Fig. 3. Simulation results of: a) plans for ramp-shaped obstacle, and b) without ramp. When the ramp is present, the robot uses it to let the cube slide, avoiding going around the obstacle. When the obstacle is a box, the robot only finds solutions that go around it. In c), the move and push scenario.

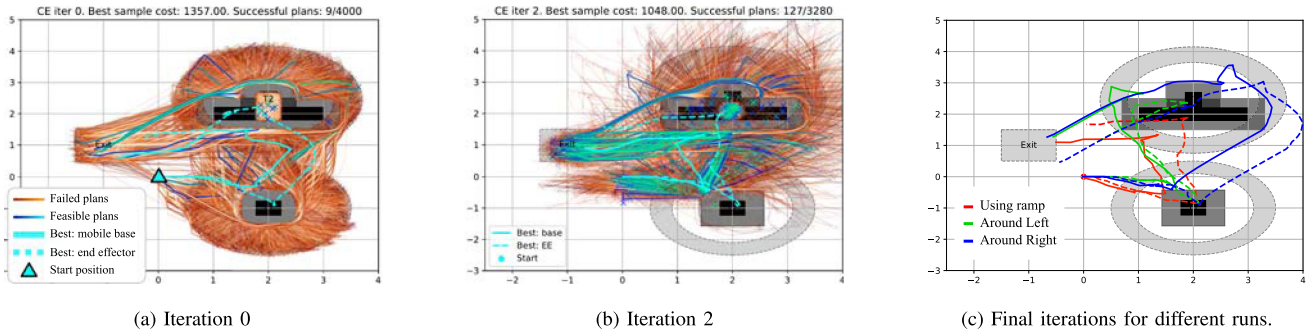


Fig. 4. Plans simulated from the sampled parameters for different iterations of the CE optimization. The trajectories displayed in orange tones show the trajectory of the mobile base of failed plans. The trajectories in cold tones show the successful plans. The best-sampled plan is shown in cyan, both base and end effector trajectory (EE), in continuous and dashed lines, respectively. The cross markers show the value of the parameter. The parameter areas are in clear grey. The tables and obstacles are in black, with 20 cm of additional padding, as used by the A\* algorithm.

### Algorithm 1: Plan Parameter Cross-Entropy Optimization.

#### Inputs:

- 1:  $\pi$ : Symbolic plan
- 2:  $\mathcal{Z}_\pi$ : Symbolic parameters used by the actions
- 3:  $\phi(\theta_\pi^0)$ : Parameterized initial distribution
- 4:  $n_{env}(j)$ : Number of parallel plans for each iteration  $j$
- 5:  $N_e$ : Number of important samples

#### Main loop:

- 6:  $j = 0$   $\triangleright$  CE iteration counter
- 7: **while**  $\theta_\pi^j$  not converged **do**
- 8:  $S \leftarrow$  Sample  $n_{env}(j)$  times  $\mathcal{Z}_\pi \sim \phi(\theta_\pi^j)$
- 9:  $C \leftarrow$  Evaluate successful plan costs of samples in  $S$
- 10:  $\theta_\pi^{j+1} \leftarrow$  Update distribution from top  $N_e$  samples
- 11:  $j = j + 1$
- 12: **end while**
- 13: **return** Best sample  $Z_\pi \in S$ , and final  $\theta_\pi^{j+1}$

that the parameters of a plan where all actions are individually successful are strongly correlated to parameters that render the whole plan feasible, i.e.,  $\psi_g(\mathbf{x}_T) = \text{True}$ . As a result, subsequent sampling iterations have a higher chance of generating feasible plan realizations.

## IV. RESULTS

We evaluate our method for a pick and place task of a cube with a mobile manipulator, and for a move and push task. These scenarios were selected to demonstrate the method's capability

to resolve complex hybrid dynamics where defining explicit analytical constraints is non-trivial. We show simulation and real-world results, Figs. 3 and 6.

### A. Experimental Setup

a) *Robot and dynamics model*: The robot in our experiments is a Dingo-O with Kinova 6DoF arm and gripper. To model the dynamics and complex contacts of the robot interacting with objects in the environment, we use the GPU-based physics simulator IsaacGym [6]. The simulator allows for computationally efficient parallelized execution of plans.

To keep computation efficient, we require the use of robot controllers that can be evaluated efficiently and in parallel, together with the simulator. For our experiment, we consider:

b) *Whole-body control*: We employ Geometric fabrics for fast computation of motion for the base and the robot arm when end-effector pose control with collision avoidance is required. Geometric fabrics [23], [24] are based on Riemannian Motion Policies (RMPs) [25] where the desired behaviors of the robot, such as collision avoidance and convergence to a goal pose, are composed into a differential equation:

$$\tilde{M}(\mathbf{q}, \dot{\mathbf{q}})\ddot{\mathbf{q}} + \tilde{\mathbf{f}}(\mathbf{q}, \dot{\mathbf{q}}) + \gamma\partial_{\mathbf{q}}\psi + \mathbf{B}\dot{\mathbf{q}} = 0, \quad (3)$$

where  $\tilde{M}$  and  $\tilde{\mathbf{f}}$  encode the system that is forced by a potential  $\psi$  towards a goal with attractor weight  $\gamma$  and damping  $\mathbf{B}$ . Solving (3) yields the trajectory generation policy  $\ddot{\mathbf{q}} = \tilde{\pi}(\mathbf{q}, \dot{\mathbf{q}})$  for whole-body control of the robot, which allows for parallelized evaluation.

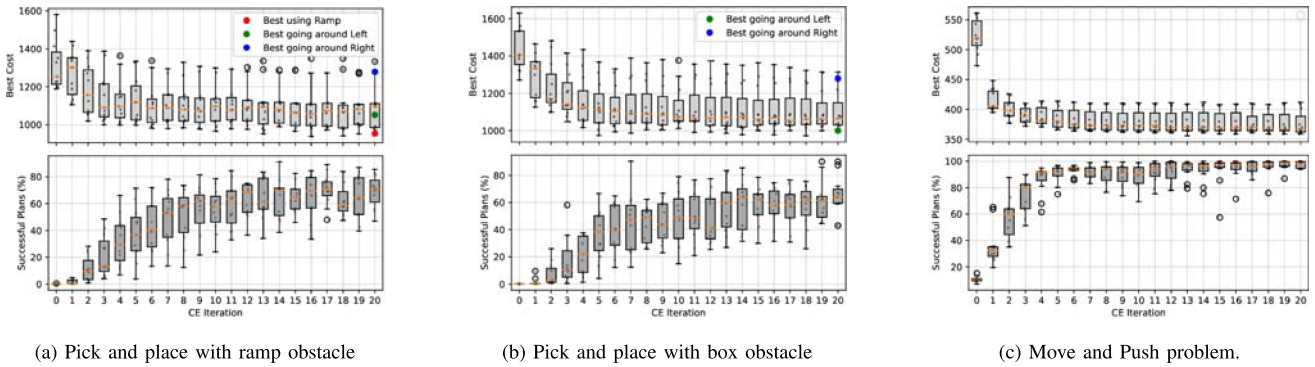


Fig. 5. Top: Cost evolution of the best-sampled plan realization for each scenario. Bottom: Percentage change of sampled, feasible plans.

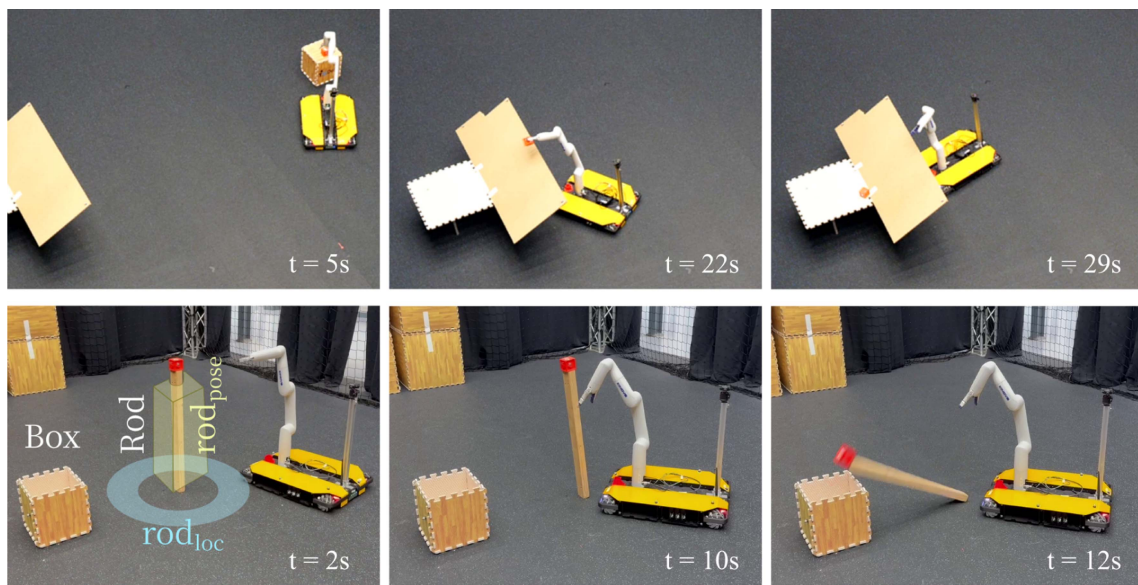


Fig. 6. Plan executed in the real system, using the same low level controllers. (Top row): Pick and place scenario with the ramp-shaped obstacle. (Bottom row): Move and Push problem.

*c) Path planning for the base:* We employ  $A^*$  [26] to compute collision-free paths for the mobile base when moving between locations. The generated waypoints are tracked with a PID controller and evaluated in parallel.

*d) TAMP problem:* Our pick and place task with one object and two tables is illustrated in Fig. 2(a). The goal specification is for the cube to be on Table 2, and the robot needs to be in the exit zone by the end of the plan. Between the tables, there is an obstacle, the geometry of which could be exploited to find better solutions. We consider two setups: Setup 1) the obstacle has a slanted side, effectively creating a ramp, and setup 2), the obstacle has no slanted side, making it a rectangular prism, which we call box. Fig. 2(a) shows a diagram of the environment setup with the ramp obstacle.

We define the following symbolic parameters that represent regions and grasps of interest

- $Table_1 \in \mathbb{R}^2$ : Represents the region of points around Table 1. The initial uniform sampling is taken within a ring ( $r = 1.1$ ,  $R = 1.5$ )m centered around the table.

- $Table_2 \in \mathbb{R}^2$ : Region of points around Table 2. The initial uniform sampling is taken from within a ring ( $r = 1.2$ ,  $R = 1.7$ )m centered around the table.
- $Cube \in SO(3)$ : Represents the orientation of a grasp pose, located in the center of the object to be picked. The orientation is encoded as a quaternion initially sampled uniformly within a 0.3 size hyper-rectangle centered around a top-down grasp pose.
- $Target \in SE(3)$ : Represents the position and orientation of the end effector when the grasped object is released. It is encoded by a location  $p \in \mathbb{R}^3$ , initially sampled from a hyper-rectangle of size (0.2,0.4,0.6)m, located on top of the table and partially on the obstacle, and a quaternion initially sampled as in  $Cube$ .
- $Exit \in \mathbb{R}^2$ : Represents a rectangular area of size (1,1)m where the robot needs to be at the end of the plan. The values are initially sampled uniformly in the rectangle.

We define the following action schemata to abstract moving, picking and placing and generate the possible symbolic actions

together with the previously defined parameters.

$$\begin{aligned} & \text{move\_to}(\{p_1 \in \perp, p_2 \in \mathbb{R}^2\}, \mathbf{A}^*, J_m(x_t), \text{succ}_m(p), \\ & \quad \psi_{\text{pre}} = \{\text{robot\_at}(p_1)\}, \\ & \quad \psi_{\text{eff}} = \{\text{robot\_at}(p_2), \neg \text{robot\_at}(p_1)\}) \end{aligned} \quad (4)$$

$$\begin{aligned} & \text{grasp}(\{q \in SO(3), p \in \perp\}, \text{Fabrics}, J_g(x_t), \text{succ}_g(q), \\ & \quad \psi_{\text{pre}} = \{\text{robot\_at}(p), \text{object\_at}(q, p), \text{ee\_empty}\}, \\ & \quad \psi_{\text{eff}} = \{\neg \text{object\_at}(q, p), \neg \text{ee\_empty}\}) \end{aligned} \quad (5)$$

$$\begin{aligned} & \text{place}(\{q \in SE(3), p \in \perp\}, \text{Fabrics}, J_p(x_t), \text{succ}_p(p), \\ & \quad \psi_{\text{pre}} = \{\text{robot\_at}(p), \neg \text{ee\_empty}\}, \\ & \quad \psi_{\text{eff}} = \{\text{object\_at}(q, p), \text{ee\_empty}\}) \end{aligned} \quad (6)$$

The robot arm remains static during *moveTo*. The cost  $J_m$  is set to 1 for every time step the action is not successful. The success condition is true if the mobile base is at the goal location within a tolerance  $\text{succ}_m(p_2) : \|x_{\text{base}} - p_2\| \leq \epsilon$

For (5) and (6), we use as input parameters an orientation and a grasp pose respectively. Both use the full body (mobile base + arm) parallelized geometric fabrics controller. As before, the cost is simply 1 every time step the action is not finished. The success condition is true if the end effector has reached the grasp/place pose within a tolerance  $\text{succ}_g(q) : \|q - p_{ee}\| \leq \epsilon$ . For grasping, we additionally check for the gripper to enclose the object.

To generate candidate symbolic plans, we consider the symbolic goal *robot\_at(Exit) AND object\_at(Table<sub>2</sub>)*. The resulting realized plan is considered successful if  $\psi_g(\mathbf{x}_T) : x_{\text{robot}}(T) \in \text{Exit}$  and  $x_{\text{cube}} \in \mathcal{X}_{\text{Table\_Surface}}$ , i.e. the cube is on the upper surface of Table 2, and the robot is at the exit region. Notice how the requirement of a successful plan does not enforce the robot to be in the annular regions *Table<sub>1</sub>*, *Table<sub>2</sub>* after executing the move to actions. The regions serve as an initial guess for sampling.

We create a candidate symbolic plan using the off-the-shelf PDDL solver [27]. The resulting plan accomplishes at the symbolic level the goal of moving the cube to the second table, and ends with the robot in the designated exit area.

The plan sequence is defined as follows, where parameters in **bold** denote continuous variables optimized by our framework, while non-bold parameters represent fixed symbolic context ( $\mathcal{M} = \perp$ ):

$$\begin{aligned} \pi_{p\&camp;p} = & \text{moveTo}(\text{Start}, \mathbf{Table}_1), \text{grasp}(\mathbf{Cube}, \text{Table}_1), \\ & \text{moveTo}(\text{Table}_1, \mathbf{Table}_2), \text{place}(\mathbf{Target}, \text{Table}_2), \\ & \text{moveTo}(\text{Table}_2, \mathbf{Exit}) \end{aligned} \quad (7)$$

Similarly, for the move and push problem, we have an open box, a wooden rod, and a red block on its top. The goal is to place the red block inside the box as fast as possible. The candidate symbolic plan is

$$\pi_{m\&camp;p} = \text{moveTo}(\text{Start}, \mathbf{rod}_{\text{loc}}), \text{push}(\mathbf{rod}_{\text{loc}}, \mathbf{rod}_p) \quad (8)$$

We initialize the symbolic location  $\text{rod}_{\text{loc}}$  as an annular region around the rod. The *push* action is parameterized by a target end-effector pose  $\text{rod}_p \in SE(3)$ . Its controller reaches the target and opens the gripper. The symbolic preconditions require the robot to be located at  $\text{rod}_{\text{loc}}$  and the block to be on the rod. This pose,  $\text{rod}_p$ , is initially sampled inside a cuboid centered around the rod, as seen in Fig. 6

e) *Algorithm setting*: We run the CE Algorithm 1, by initially sampling values uniformly within the bounds of the parameters. For subsequent iterations, we model the distribution of the parameters as an independent normal distribution. The resulting  $\mu$  and  $\sigma$  of the elite parameter samples are used for the next iteration. We initially sample  $n_{\text{envs}} = 3000$  and decrease linearly to 300 in iteration 10, constant afterwards.  $N_e = 50$  elite samples are used. We repeat each experiment for 10 different random seeds.

## B. Simulation Results

Fig. 4 shows sampled plan realizations at different CE stages for the pick and place problem (ramp setup). The sampled parameters and trajectories are projected onto the ground plane. The best sample's end-effector trajectory is overlaid.

In the pick and place problem with the ramp obstacle, the best solution exploits the geometry by sliding the cube down the slope to the table, avoiding going around it (Fig. 3(a)). When the obstacle is a box instead, the robot goes around it to directly place the cube (Fig. 3(b)).

Fig. 5(a)–(b) shows the success rate and cost of the best sampled plan over CE iterations for both setups. We observe that initially, only  $\approx 0.1\%$  of the sampled plans are feasible. Feasible values for parameters like grasp and place poses occupy highly constrained, disconnected regions of the parameter space, causing low initial success rates. On the other hand, parameters such as locations are more often successful, as the controller used can accomplish the move action reliably for most parameter's values. Over iterations, successful plan realization samples significantly increase and costs decrease. For these reported metrics, we utilized a fixed budget of 20 iterations to ensure asymptotic stability, though we observe that the solution cost typically converges ( $< 1\%$  variance) within the first 5-10 iterations.

We also observe the multi-modality of the realizations found in the pick and place problem. The solution modes can be divided in three, ordered by increasing average cost: 1) using the ramp (only possible with the ramp), 2) going around the obstacle from the left, and 3) going around the obstacle from the right. Fig. 4(c) shows a converged example for each mode. As we use normal distributions for sampling the parameters, they eventually collapse to single local minima corresponding to a mode. If the initial samples cover the full range of modes equally, thus providing information about the different costs, Algorithm 1 converges to the mode that provides the best costs. However, as few feasible plans are found in the initial iterations, some modes may be overrepresented, and the updated distribution will move towards them even if they create higher cost realizations. In Fig. 5(a), with the ramp, we observe that out of the 10 final solutions, 6 use the ramp, 3 go around the left, and 1 goes around the right. For the box setup, Fig. 5(b), we observe 7 solutions going around the left, 3 going around the right.

Evaluating the computational cost on an NVIDIA 4060 (Laptop), we find the bottleneck is parallel controller execution ( $\approx 80\%$  of total time) compared to physics simulation ( $\approx 19\%$ ). However, the time per iteration decreases drastically as the optimization converges. For the pick and place scenario, the compute time drops from  $\approx 6$  min (Iteration 0) to 50 s (Iteration 10+) as the number of active environments decreases and trajectories become more efficient. Similarly, the Push scenario decreases from 1.5 min to 15 s per iteration. The cost of the plan realizations follows a steep decrease during the initial iterations Fig. 5, which

allows the possibility to obtain a low cost solution with relatively few iterations.

### C. Real World Experiments

We report results for 1) the pick and place (ramp setup), and 2) move and push problem on a real-world system. To bridge the sim-to-real gap, we rely on three design choices. First, identical closed-loop controllers in simulation and reality compensate for minor dynamic mismatches. Second, object-centric reference frames to ensure actions remain valid relative to the object even if global localization varies. Third, we utilize Domain Randomization (DR) to robustify plans against bounded uncertainty in friction or mass [28], though deterministic optimization proved sufficient given accurate motion-capture estimation. Simulator inputs include geometries of the objects in the environment and their locations. Fig. 6 shows key execution moments.

The real-world execution consistently matched the optimal simulation strategy (Fig. 3). While the pick-and-place task took longer (35 s vs. 21 s) due to ground friction and less capable actuators than in simulation, the move-and-push task showed only a 3 s difference. The solution showcases that complex contacts and dynamics such as sliding a cube can be considered at the plan realization stage in a real system, without injecting expert knowledge at the symbolic stage.

## V. CONCLUSION

Our work computes plan realizations of task plans while implicitly accounting for robot dynamics and contacts with the environment. We employ a parallelized physics simulator and cross-entropy optimization to sample the parameters used by the controllers, or actions, and obtain feasible and low-cost solutions. The results of our simulations and real-world experiments showcase that our method is capable of finding low-cost plan realizations within a few CE iterations and with little prior information. The generated solutions consider geometrical constraints, such as the inability of the robot to reach objects from certain positions. Instead of explicitly modeling these constraints in the problem, our method implicitly models them as a consequence of the physics simulations. This allows us to consider the effects of the utilized controllers, e.g., as seen when our mobile base is repositioned while pushing against the table during grasping and placing. This behavior resulting from environment contacts and robot controllers could not be modeled in previous sampling-based methods [3], [4].

*Limitations:* Despite GPU parallelization, evaluating full-horizon physics rollouts remains computationally expensive, currently limiting real-time planning, or evaluating different symbolic plans efficiently. Furthermore, the approach relies on informative initial regions; without these, the optimizer may fail to converge in sparse solution spaces.

## ACKNOWLEDGMENT

Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

## REFERENCES

- [1] M. Toussaint, "Logic-geometric Programming: An optimization-based approach to combined task and motion planning," in *Proc. Int. Conf. AI*, Buenos Aires, Argentina, 2015, pp. 1930–1936.
- [2] C. R. Garrett et al., "Integrated task and motion planning," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 4, pp. 265–293, 2020.
- [3] C. R. Garrett, T. Lozano-Pérez, and L. P. Kaelbling, "PDDLStream: Integrating symbolic planners and blackbox samplers via optimistic adaptive planning," in *Proc. Int. Conf. Automated Plan. Scheduling*, 2020, pp. 440–448.
- [4] W. Thomason, M. P. Strub, and J. D. Gammell, "Task and motion informed trees (TMIT\*): Almost-surely asymptotically optimal integrated task and motion planning," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 11370–11377, Oct. 2022.
- [5] M. Toussaint, K. Allen, K. Smith, and J. Tenenbaum, "Differentiable physics and stable modes for tool-use and manipulation planning," in *Proc. 14th Conf. Robot. Sci. Syst.*, 2018, pp. 6231–6235.
- [6] V. Makoviychuk et al., "Isaac Gym: High performance GPU-based physics simulation for robot learning," in *Proc. Neural Inf. Process. Syst. Track Datasets Benchmarks*, 2021.
- [7] M. Noseworthy et al., "FORGE: Force-guided exploration for robust contact-rich manipulation under uncertainty," *IEEE Robot. Automat. Lett.*, vol. 10, no. 5, pp. 4436–4443, May 2025.
- [8] J. Xu et al., "Accelerated policy learning with parallel differentiable simulation," in *Proc. Int. Conf. Learn. Representations*, 2022.
- [9] D. McDermott et al., "PDDL-The planning domain definition language," Yale Center Comput. Vis. Control, New Haven, CT, USA, Tech. Rep. CVC TR-98-003/DCS TR-1165, 1998.
- [10] Z. Zhao et al., "A survey of optimization-based task and motion planning: From classical to learning approaches," *IEEE/ASME Trans. Mechatron.*, vol. 30, no. 4, pp. 2799–2825, Aug. 2025.
- [11] N. T. Dantam, Z. K. Kingston, S. Chaudhuri, and L. E. Kavraki, "An incremental constraint-based framework for task and motion planning," *Int. J. Robot. Res.*, vol. 37, no. 10, pp. 1134–1151, Sep. 2018.
- [12] M. Toussaint and M. Lopes, "Multi-bound tree search for logic-geometric programming in cooperative manipulation domains," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2017, pp. 4044–4051.
- [13] M. Toussaint, "Newton methods for K-order Markov constrained motion problems," Jul. 2014, *arXiv:1407.0414*.
- [14] R. Holladay, T. Lozano-Perez, and A. Rodriguez, "Force-and-motion constrained planning for tool use," in *Proc. IEEE Int. Conf. Intel. Robots Syst.*, 2019, pp. 7409–7416.
- [15] C. R. Garrett, T. Lozano-Pérez, and L. P. Kaelbling, "Sampling-based methods for factored task and motion planning," *Int. J. Robot. Res.*, vol. 37, no. 13/14, pp. 1796–1825, 2018.
- [16] M. Toussaint, C. V. Braun, and J. Ortiz-Haro, "NLP Sampling: Combining MCMC and NLP methods for diverse constrained sampling," Jul. 2024, *arXiv:2407.03035*.
- [17] C. V. Braun, R. T. Lange, and M. Toussaint, "Stein variational evolution strategies," in *Proc. 41st Conf. Uncertainty Artif. Intell.*, PMLR, vol. 286, 2025, pp. 398–420.
- [18] M. Toussaint, J. Harris, J.-S. Ha, D. Driess, and W. Hönig, "Sequence-of-constraints MPC: Reactive timing-optimal control of sequential manipulation," in *Proc. IEEE Int. Conf. Intel. Robots Syst.*, 2022, pp. 13753–13760.
- [19] T. Pan, R. Shome, and L. E. Kavraki, "Task and motion planning for execution in the real," *IEEE Trans. Robot.*, vol. 40, pp. 3356–3371, 2024.
- [20] A. Curtis et al., "Partially observable task and motion planning with uncertainty and risk awareness," in *Proc. 20th Conf. Robot. Sci. Syst.*, Delft, The Netherlands, 2024.
- [21] M. Helmert, "The fast downward planning system," *J. AI Res.*, vol. 26, pp. 191–246, 2006.
- [22] Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, and P. L'Ecuyer, "The cross-entropy method for optimization," in *Handbook of Statistics*, vol. 31. Amsterdam, The Netherlands: Elsevier, 2013, pp. 35–59.
- [23] K. Van Wyk et al., "Geometric fabrics: Generalizing classical mechanics to capture the physics of behavior," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 3202–3209, Apr. 2022.
- [24] M. Spahn, M. Wisse, and J. Alonso-Mora, "Dynamic optimization fabrics for motion generation," *IEEE Trans. Robot.*, vol. 39, no. 4, pp. 2684–2699, Aug. 2023.
- [25] N. D. Ratliff, J. Issac, D. Kappler, S. Birchfield, and D. Fox, "Riemannian motion policies, 2018," *arXiv:1801.02854*.
- [26] P. Hart, N. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Trans. Syst. Sci. Cybern.*, vol. 4, no. 2, pp. 100–107, Jul. 1968.
- [27] M. Helmert, "The fast downward planning system," *J. Artif. Intell. Res.*, vol. 26, pp. 191–246, Jul. 2006.
- [28] C. Pezzato, C. Salmi, E. Trevisan, M. Spahn, J. Alonso-Mora, and C. H. Corbato, "Sampling-based model predictive control leveraging parallelizable physics simulations," *IEEE Robot. Autom. Lett.*, vol. 10, no. 3, pp. 2750–2757, Mar. 2025.