

# Scalable Differentiable Physics for Learning and Control

Anonymous Authors<sup>1</sup>

## Abstract

Differentiable physics is a powerful approach to learning and control problems that involve physical objects and environments. While notable progress has been made, the capabilities of differentiable physics solvers remain limited. We develop a scalable framework for differentiable physics that can support a large number of objects and their interactions. To accommodate objects with arbitrary geometry and topology, we adopt meshes as our representation and leverage the sparsity of contacts for scalable differentiable collision handling. Collisions are resolved in localized regions to minimize the number of optimization variables even when the number of simulated objects is high. We further accelerate implicit differentiation of optimization with nonlinear constraints. Experiments demonstrate that the presented framework requires up to two orders of magnitude less memory and computation in comparison to recent particle-based methods. We further validate the approach on inverse problems and control scenarios, where it outperforms derivative-free and model-free baselines by at least an order of magnitude.

## 1. Introduction

Differentiable physics enables gradient-based learning systems to strictly adhere to physical dynamics. By making physics simulation differentiable, we can backpropagate through the physical consequences of actions. This enables agents to quickly learn to achieve desired effects in the physical world. It is also a principled and effective approach to inverse problems that involve physical systems (Degrave et al., 2019; de Avila Belbute-Peres et al., 2018; Toussaint et al., 2018; Schenck & Fox, 2018).

Recent efforts have significantly advanced the understanding of differentiable physics in machine learning and robotics. Automatic differentiation and analytical methods have been applied to derive a variety of differentiable simulation engines (Ingraham et al., 2019; Liang et al., 2019; Holl et al., 2020; Hu et al., 2020). Yet existing differentiable frameworks are still limited in their ability to simulate complex

scenes composed of many detailed interacting objects. For example, modeling assumptions have limited some existing frameworks to restricted object classes, such as balls or two-dimensional polygons (de Avila Belbute-Peres et al., 2018; Degrave et al., 2019). Other approaches adopt expressive grid- or particle-based formulations, which are general in principle but have not scaled to detailed simulation of large scenes due to the cubic growth rates of volumetric grids and particles (Mrowca et al., 2018; Hu et al., 2019; Li et al., 2019a). For example, a grid of size  $1024^3$  has on the order of a billion elements, and this resolution is still quite coarse in comparison to the scale and detail of physical environments we encounter in daily life, such as city streets. Particle-based representations likewise suffer from explosive growth rates: Macklin et al. (2014) suggest a maximum size ratio of 1:10 between the smallest and largest feature, or placing particles only around boundaries, which runs the risk of object tunneling.

In this paper, we develop a differentiable physics framework that addresses these limitations. We adopt *meshes* as a general representation of objects. Meshes are the most widely used specification for object geometry in computer graphics and scientific computing. Meshes are inherently sparse, can model objects of any shape, and can compactly specify environments with both large spatial extent and highly detailed features (Botsch et al., 2010).

The use of meshes brings up the challenge of collision handling, since collisions can occur anywhere on the surface of the mesh. Prior frameworks adopted global LCP-based formulations, which severely limited scalability (Degrave et al., 2019; de Avila Belbute-Peres et al., 2018). In contrast, we leverage the structure of contacts by grouping them into localized impact zones, which can be efficiently represented and processed (Bridson et al., 2002; Harmon et al., 2008). This substantially reduces the number of variables and constraints involved in specifying dynamical scenes and dramatically speeds up backpropagation.

Following prior work, we use implicit differentiation to compute gradients induced by an optimization problem embedded in the simulation (Amos & Kolter, 2017; Liang et al., 2019). One of our contributions is an acceleration scheme that can handle the nonlinear constraints encountered in our system.

We evaluate the scalability and generality of the presented approach experimentally. Controlled experiments that vary the number of objects in the scene and their relative scales demonstrate that our approach is dramatically more efficient than state-of-the-art differentiable physics frameworks. Since meshes can represent both rigid and deformable objects, our work is also the first differentiable physics framework that can simulate two-way coupling of rigid bodies and cloth. We further demonstrate example applications to learning and control scenarios, where the presented framework outperforms derivative-free and model-free baselines by at least an order of magnitude. Code and video are provided in the supplement.

## 2. Related Work

Backpropagating through physical dynamics is an enticing prospect for embedding physical reasoning into learning processes. Degrave et al. (2019) articulated the potential of this approach and applied automatic differentiation to rigid body systems where contacts are assumed to only happen between balls and planes. Although this assumption simplifies collision detection and response, it greatly limits the scope of application. de Avila Belbute-Peres et al. (2018) introduced implicit differentiation for gradient computation, obviating the need to explicitly differentiate through all steps of the forward optimization solver. By designing specific rules for collision response, their framework supports simple two-dimensional shapes such as circles and polygons, but was not extended to general 3D objects. Toussaint et al. (2018) utilized differentiable primitives to develop a robot reasoning system that can achieve user-defined tasks. It is designed for path planning with sphere-swept shapes and frictionless interactions, but not for general control tasks.

Building on the implicit differentiation methodology, Liang et al. (2019) developed a differentiable simulator for cloth. Their formulation assumes that each node can move independently and is only tied to others by soft constraints. This is true for cloth but not for rigid bodies. Carpenter & Mansard (2018) developed algorithms for efficient computation of analytical derivatives of rigid-body dynamics, but did not handle collision. Millard et al. (2020) implemented an articulated body simulator in a reverse-mode automatic differentiation framework, but likewise did not handle dynamic contact and collision response, which are at the heart of our work.

Hu et al. (2019) developed a differentiable simulator for deformable objects based on the material point method (MPM). A follow-up work applies this framework to controller design for soft robots (Spielberg et al., 2019). This approach is effective in handling soft bodies, but has limited ability to enforce rigid structure. Since the method is based on particles and grids, it has limited scalability and has not been

applied to large scenes with interacting objects. In contrast, our method uses meshes to represent objects, naturally enforces precise detailed geometry, and scales to scenes with both large spatial extent and fine local features.

Hu et al. (2020) introduce a domain-specific language for building differentiable physical simulators. This work develops a programming language and compiler that can be broadly useful throughout the differentiable physics community, including to projects such as ours. However, at present this framework does not deal with general mesh-based rigid and deformable objects. Their rigid-body scenario uses rectangles constrained by springs, with contact responses only possible between rectangles and the ground plane. Their three-dimensional scenario uses soft bodies modeled by particles and grids and inherits the aforementioned limitations of MPM-based modeling.

A line of work developed particle-based frameworks for differentiable simulation of fluids (Schenck & Fox, 2018; Ummenhofer et al., 2020) and unified treatment of fluids and solids (Mrowca et al., 2018; Li et al., 2019a). These frameworks resort to approximations in handling rigid bodies and detailed geometry, and suffer from high growth rates in the scale of objects and scenes. The generality of particle-based approaches is appealing, but their scalability limitations are well-known and are explicitly discussed in the literature (Macklin et al., 2014). We take a different tack and model objects as meshes, thus harnessing the scalability and generality of this representation.

Some of the aforementioned works, as well as others (Battaglia et al., 2016; Chang et al., 2017; Sanchez-Gonzalez et al., 2018; Li et al., 2019b), fit function approximators to demonstrations of physical dynamics. Such approximators are valuable, but may not fully capture the true underlying physics and may degenerate outside the training distribution. Our work follows the program exemplified by Degrave et al. (2019), de Avila Belbute-Peres et al. (2018), and others in that we aim to backpropagate through the true physical dynamics. In this approach, physical correctness is enforced by construction.

## 3. Overview

Our differentiable physics engine can function as a layer embedded within a broader differentiable pipeline, with other neural network layers in the loop. This scenario is illustrated in Figure 1. Given observations from simulation, the network can compute control signals and send them to the simulator. The physics engine performs time integration, evaluates collision response, and updates the mesh states. The simulation results can be used to evaluate the loss and provide new observations to the network, thus closing the loop. All operations in our physics engine are differen-

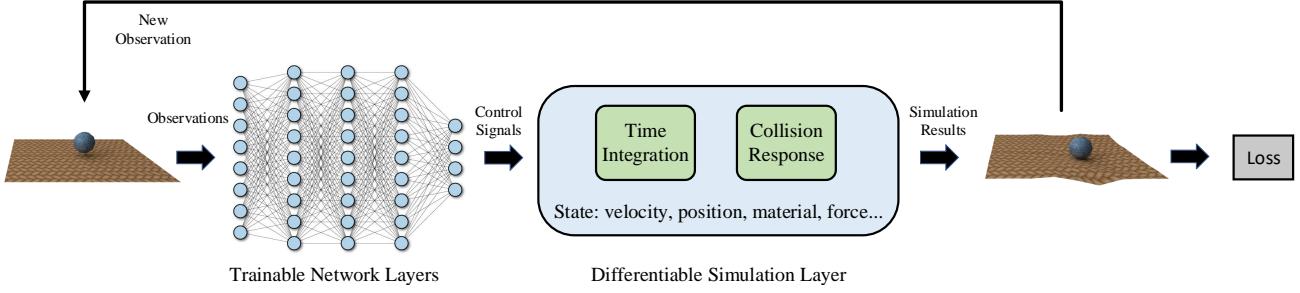


Figure 1. Differentiable simulation embedded in a neural network. The loss can be backpropagated through the physics simulator to the control signal, thus enabling end-to-end gradient-based training of the controller network. This approach enables faster convergence with higher accuracy in learning and control of physical systems.

tiable, supporting end-to-end gradient-based optimization of the complete pipeline. As mentioned earlier, state-of-the-art differentiable simulation algorithms often suffer from scalability issues. They either use particle representations, which have cubic growth rates with respect to object sizes, or compute custom dynamic responses for a limited number of object primitives, which have limited ability to represent complex shapes. Our method resolves these issues to advance scalability and generality.

Section 4 introduces our notation and the basic forward integration scheme. To facilitate scalability and generality, we adopt meshes as our basic object representation. Simulation of complex scenes brings up the challenge of collision detection and response. We describe an efficient collision handling scheme in Section 5. Observing that collisions are often sparse, we utilize the impact zone method and handle collisions and contact forces locally in independent collision areas. This collision handling scheme has linear computational complexity with respect to the number of constraints rather than the total number of nodes. Our unified collision handling formulation also enables differentiable two-way coupling of rigid bodies and cloth. In Section 6, we describe the challenges involved in differentiation and introduce an acceleration scheme for implicit differentiation with nonlinear constraints.

## 4. Preliminaries

During the presentation of our method, the configuration of the whole system is collected into one vector  $\mathbf{q} = [\mathbf{q}_1^\top, \mathbf{q}_2^\top, \dots, \mathbf{q}_n^\top]^\top$ , where  $\mathbf{q}_{1 \leq i \leq n}$  are the generalized coordinates of all objects in the system. Similarly, we use  $\dot{\mathbf{q}} = [\dot{\mathbf{q}}_1^\top, \dot{\mathbf{q}}_2^\top, \dots, \dot{\mathbf{q}}_n^\top]^\top$  to denote the velocity of the corresponding coordinates. For a rigid body, the generalized coordinates  $\mathbf{q}_i$  have 6 degrees of freedom (DOF), corresponding to position and orientation. Accordingly,  $\dot{\mathbf{q}}_i$  can be interpreted as a vector of linear and angular velocities. For a cloth node,  $\mathbf{q}_j$  has 3 DOF, corresponding to its Cartesian coordinates, and  $\dot{\mathbf{q}}_j$  is the velocity vector.

After construction of the system, the dynamics can be written as

$$\frac{d}{dt} \begin{pmatrix} \mathbf{q} \\ \dot{\mathbf{q}} \end{pmatrix} = \begin{pmatrix} \dot{\mathbf{q}} \\ \ddot{\mathbf{q}} \end{pmatrix} = \begin{pmatrix} \dot{\mathbf{q}} \\ \mathbf{M}^{-1} \mathbf{f}(\mathbf{q}, \dot{\mathbf{q}}) \end{pmatrix}, \quad (1)$$

where  $\mathbf{M}$  is the mass matrix and  $\mathbf{f}$  is the generalized force vector. We incorporate contact forces (Harmon et al., 2008), internal forces including bending and stretching (Narain et al., 2012), and external forces including gravity and control input. For numerical simulation, the dynamical system can be discretized with time step  $h$ . At time  $t_0$ , with  $\mathbf{q}_0 = \mathbf{q}(t_0)$  and  $\dot{\mathbf{q}}_0 = \dot{\mathbf{q}}(t_0)$ , we can compute the increment of the coordinates  $\Delta\mathbf{q} = \mathbf{q}(t_0 + h) - \mathbf{q}(t_0)$  and velocities  $\Delta\dot{\mathbf{q}} = \dot{\mathbf{q}}(t_0 + h) - \dot{\mathbf{q}}(t_0)$  using the implicit Euler method (Witkin & Baraff, 1997):

$$\begin{pmatrix} \Delta\mathbf{q} \\ \Delta\dot{\mathbf{q}} \end{pmatrix} = h \begin{pmatrix} \dot{\mathbf{q}}_0 + \Delta\dot{\mathbf{q}} \\ \mathbf{M}^{-1} \mathbf{f}(\mathbf{q} + \Delta\mathbf{q}, \dot{\mathbf{q}} + \Delta\dot{\mathbf{q}}) \end{pmatrix}. \quad (2)$$

With a linear approximation, we can solve for  $\Delta\dot{\mathbf{q}}$ ,

$$\left( h^{-1} \mathbf{M} - \frac{\partial \mathbf{f}}{\partial \dot{\mathbf{q}}} - h \frac{\partial \mathbf{f}}{\partial \mathbf{q}} \right) \Delta\dot{\mathbf{q}} = \mathbf{f}_0 + h \frac{\partial \mathbf{f}}{\partial \mathbf{q}} \dot{\mathbf{q}}_0, \quad (3)$$

which can then be used to compute  $\mathbf{q}(t_0 + h)$  and  $\dot{\mathbf{q}}(t_0 + h)$  for the next time step.

## 5. Scalable Collision Handling

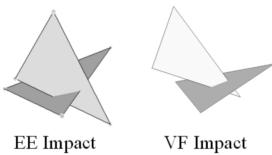
One of the most time-consuming and complex parts of physics simulation is collision handling. The difficulty, especially for differentiable physics, lies in how to compute the *correct* derivatives *efficiently* even in a large scene with many interacting objects. As observed by Hu et al. (2020), naive discrete-time impulse-based collision response can lead to completely incorrect gradients. We apply continuous collision detection to circumvent this problem. We also employ a bounding volume hierarchy to localize and accelerate dynamic collision detection.

Traditional collision response algorithms for rigid bodies commonly employ global LCP-based solvers to resolve

165 collisions (de Avila Belbute-Peres et al., 2018). This works  
 166 well for small-to-medium-size systems, but does not scale  
 167 well when the number of objects increases. The problem  
 168 is exacerbated when the derivatives of all DOFs in this  
 169 large system need to be computed simultaneously, which is  
 170 exactly what backpropagation does.

171 Our key observation is that collisions are sparse in the  
 172 majority of simulation steps, which means that the number of  
 173 collisions is much smaller than the total DOF. As shown  
 174 in Figure 2, collisions are commonly localized around the  
 175 scene. With this observation, we resolve collision locally  
 176 among the objects involved, rather than setting up a global  
 177 optimization, updating coordinates, and computing the  
 178 gradients for all objects in the environment.

180 One way to represent local  
 181 collisions is to use impact  
 182 zones (Bridson et al., 2002;  
 183 Harmon et al., 2008). An  
 184 impact is a pair of primitives  
 185 colliding with each other. It  
 186 can be an edge-edge (EE) or a vertex-face (VF) pair. For  
 187 these two cases, the non-penetration constraints can be written  
 188 as



$$190 \quad 0 \leq C_{ee} = \mathbf{n} \cdot [(\alpha_3 \mathbf{x}_3 + \alpha_4 \mathbf{x}_4) - (\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2)] \\ 191 \quad 0 \leq C_{vf} = \mathbf{n} \cdot [\mathbf{x}_4 - (\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \alpha_3 \mathbf{x}_3)], \quad (4)$$

193 respectively, where  $\mathbf{x}_i$  are the positions of the mesh vertices,  
 194  $\alpha_i$  are the barycentric coordinates, and  $\mathbf{n}$  is the normal of  
 195 the colliding plane.

196 Impacts may share vertices. All the impacts in one  
 197 connected component are said to form an impact zone. Each  
 198 impact zone is a local area that can be treated independently.  
 199 The red areas in Figure 2 are impact zones. Collision reso-  
 200 lution in one impact zone can be written as an optimization  
 201 problem:

$$203 \quad \underset{\mathbf{x}'}{\text{minimize}} \quad \frac{1}{2} (\mathbf{x} - \mathbf{x}')^\top \mathbf{M} (\mathbf{x} - \mathbf{x}') \\ 204 \quad \text{subject to} \quad \mathbf{Gx}' + \mathbf{h} \leq \mathbf{0}. \quad (5)$$

207 Here  $\mathbf{x}$  is the concatenation of vertex positions in the impact  
 208 zone before collision handling;  $\mathbf{x}'$  are the resolved positions  
 209 that satisfy constraints;  $\mathbf{M}$  is the mass matrix;  $\mathbf{G}$  and  $\mathbf{h}$   
 210 are a matrix and a vector derived from the constraints in  
 211 Equation 4.

212 The impact zone method is fail-safe (Bridson et al., 2002)  
 213 and is commonly used in cloth simulation, where each ver-  
 214 tex in the cloth is free and can be explicitly optimized when  
 215 solving Equation 5. However, we cannot directly use this  
 216 formulation for rigid bodies, because all vertices in a rigid  
 217 body are tied and cannot be optimized separately. A possi-  
 218 ble treatment is to use additional constraints on the relative  
 219

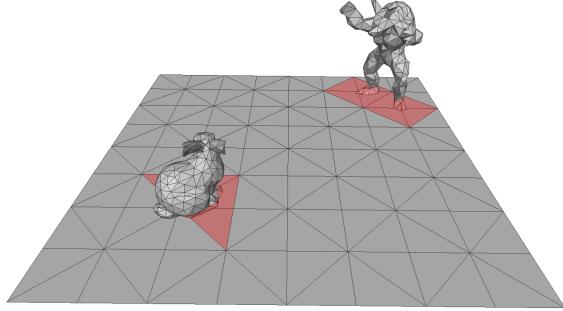


Figure 2. Collision visualization. Collided regions are shown in red. In most cases, the number of collisions is sparse compared to the number of degrees of freedom in the scene.

positions of vertices to enforce rigidity. However, the number of vertices in a rigid body is usually far larger than its DOF. We therefore take a different approach that minimizes the number of optimization variables in the system. This is critical for scalable backpropagation in such systems.

We make a distinction between contact vertices  $\mathbf{x}$  and the objects' generalized coordinates  $\mathbf{q}$ . The actual optimization variables are replaced by  $\mathbf{q}$ . For a rigid body  $i$  with rotation  $\mathbf{r} = (\phi, \theta, \psi)^\top$  and translation  $\mathbf{t} = (t_x, t_y, t_z)^\top$ , the generalized coordinates are  $\mathbf{q} = [\mathbf{r}^\top, \mathbf{t}^\top]^\top \in \mathbb{R}^6$ . The block-diagonal mass matrix  $\hat{\mathbf{M}} \in \mathbb{R}^{6 \times 6}$  is accordingly changed into angular and linear inertia. The constraints are rewritten using the new variables. A function  $\mathbf{f}(\cdot)$  is defined that maps generalized coordinates  $\mathbf{q}$  to a contact point  $\mathbf{x}$ , including both rotation and translation. Detailed formulations of  $\hat{\mathbf{M}}$  and  $\mathbf{f}(\cdot)$  are given in Appendices A and B. The optimization problem for collision resolution becomes

$$\begin{aligned} & \underset{\mathbf{q}'}{\text{minimize}} \quad \frac{1}{2} (\mathbf{q} - \mathbf{q}')^\top \hat{\mathbf{M}} (\mathbf{q} - \mathbf{q}') \\ & \text{subject to} \quad \mathbf{Gf}(\mathbf{q}') + \mathbf{h} \leq \mathbf{0}. \end{aligned} \quad (6)$$

## 6. Fast Differentiation

Gradients for the sparse linear system in Equation 3 can be computed via implicit differentiation (Amos & Kolter, 2017; Liang et al., 2019). Other basic operations can be handled by automatic differentiation (Paszke et al., 2019). However, the heaviest computation is induced by collision handling, which can take up to 90% of the runtime for a single simulation step. Collision resolution can involve multiple iterations of the optimization specified in Equation 6 for each impact zone. Liang et al. (2019) proposed a method that reduces computation, but it only deals with linear constraints and does not apply in our case due to the the nonlinearity induced by  $\mathbf{f}(\cdot)$ . We now develop an appropriate acceleration scheme that works in our setting.

Consider the differentiation in our case. At a local minimum  $\mathbf{z}^*$  with Lagrange multiplier  $\lambda^*$ , the following KKT

220 conditions hold:

$$\begin{aligned} \hat{\mathbf{M}}\mathbf{z}^* - \hat{\mathbf{M}}\mathbf{q} + \nabla f^\top \mathbf{G}^\top \lambda^* &= 0 \\ D(\lambda^*)(\mathbf{G}f(\mathbf{z}^*) + \mathbf{h}) &= 0, \end{aligned} \quad (7)$$

221 where  $D(\cdot)$  packages a vector into a diagonal matrix. The  
222 key idea here is to linearize  $f(\cdot)$  around a small neighbor-  
223 hood of  $\mathbf{z}^*$ . The implicit differentiation is formulated as  
224

$$\begin{bmatrix} \hat{\mathbf{M}} & \nabla f^\top \mathbf{G}^\top \\ D(\lambda^*)\mathbf{G}\nabla f & D(\mathbf{G}f(\mathbf{z}^*) + \mathbf{h}) \end{bmatrix} \begin{bmatrix} \frac{\partial \mathbf{z}}{\partial \lambda} \\ \frac{\partial \mathbf{L}}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{M}}\partial \mathbf{q} - \nabla f^\top \partial \mathbf{G}^\top \lambda^* \\ -D(\lambda^*)(\partial \mathbf{G} \cdot f(\mathbf{z}^*) + \partial \mathbf{h}) \end{bmatrix}. \quad (8)$$

225 As derived by Amos & Kolter (2017), by solving for  $\mathbf{d}_z$ ,  $\mathbf{d}_\lambda$   
226 in the following equation  
227

$$\begin{bmatrix} \hat{\mathbf{M}} & \nabla f^\top \mathbf{G}^\top D(\lambda^*) \\ \mathbf{G}\nabla f & D(\mathbf{G}f(\mathbf{z}^*) + \mathbf{h}) \end{bmatrix} \begin{bmatrix} \mathbf{d}_z \\ \mathbf{d}_\lambda \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \mathbf{z}}^\top \\ \mathbf{0} \end{bmatrix}, \quad (9)$$

228 we can get the backward gradients:  
229

$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}} = \mathbf{d}_z^\top \hat{\mathbf{M}} \quad (10)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{G}} = -D(\lambda^*)\mathbf{d}_\lambda f(\mathbf{z}^*)^\top - \lambda^* \mathbf{d}_z^\top \nabla f^\top \quad (11)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{h}} = -\mathbf{d}_\lambda^\top D(\lambda^*). \quad (12)$$

230 Assume that in one impact zone, there are  $n$  DOFs and  
231  $m$  constraints. In Equation 9, the size of the system is  
232  $n + m$ . The solution of this system would typically take  
233  $O((n + m)^3)$ : prohibitively expensive in large scenes. We  
234 therefore need to accelerate the computation of  $\mathbf{d}_z$  and  $\mathbf{d}_\lambda$ .  
235 Liang et al. (2019) proposed to use a QR decomposition.  
236 However, it only copes with quadratic programming with  
237 linear constraints. We can extend this method to nonlinear  
238 constraints by incorporating the Jacobian  $\nabla f$ .

239 Recall that a rigid body has coordinates  $\mathbf{q} = [\mathbf{r}, \mathbf{t}] \in \mathbb{R}^6$ .  
240 The rotation is represented by Euler angles  $\mathbf{r} = (\phi, \theta, \psi)$ .  
241 The Jacobian at vertex  $\mathbf{x} = (x, y, z)^\top$  of the rigid body  
242

$$\nabla f_k = \begin{bmatrix} \frac{\partial x}{\partial \phi} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \psi} & 1 & 0 & 0 \\ \frac{\partial y}{\partial \phi} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \psi} & 0 & 1 & 0 \\ \frac{\partial z}{\partial \phi} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \psi} & 0 & 0 & 1 \end{bmatrix}. \quad (13)$$

243 The exact formulation of  $\nabla f_k$  can be found in Appendix C.

244 After computing  $\nabla f$ , we apply a QR decomposition  
245  $\sqrt{\hat{\mathbf{M}}}^{-1} \nabla f^\top \mathbf{G}^\top = \mathbf{Q}\mathbf{R}$  and derive  $\mathbf{d}_z$ ,  $\mathbf{d}_\lambda$  from Equation 9 as  
246

$$\mathbf{d}_z = \sqrt{\hat{\mathbf{M}}}^{-1} (\mathbf{I} - \mathbf{Q}\mathbf{Q}^\top) \sqrt{\hat{\mathbf{M}}}^{-1} \frac{\partial \mathbf{L}}{\partial \mathbf{z}}^\top \quad (14)$$

$$\mathbf{d}_\lambda = D(\lambda^*)^{-1} \mathbf{R}^{-1} \mathbf{Q}^\top \sqrt{\hat{\mathbf{M}}}^{-1} \frac{\partial \mathbf{L}}{\partial \mathbf{z}}^\top. \quad (15)$$

The cost of the QR decomposition is  $O(nm^2)$ , reducing the computation from  $O((n + m)^3)$ .

## 7. Experiments

We conduct a variety of experiments to evaluate the presented approach. We begin by comparing to other differentiable physics frameworks, with particular attention to scalability and generality. We then conduct ablation studies to evaluate the impact of the techniques presented in Sections 5 and 6. Lastly, we provide case studies that illustrate the application of the presented approach to inverse problems and learning control.

The automatic differentiation is implemented in PyTorch 1.3 (Paszke et al., 2019). All experiments are run on an Intel(R) Xeon(R) W-2123 CPU @ 3.60GHz.

### 7.1. Scalability

We construct two benchmark scenes and vary them in controlled fashion to evaluate scalability. The scenes are illustrated in Figure 3.

In the first, objects fall from the air, hit the ground, and finally settle. To test scalability, we increase the number of objects while maintaining the stride between objects. As the number of objects increases, the spatial extent of the scene expands accordingly.

In the second scene, a rigid bunny strikes a deformable cloth. We vary the relative scale of the bunny and the cloth to test the ability of simulation frameworks to handle objects with geometric features at different resolutions.

Few differentiable simulation frameworks are capable of modeling these scenes. de Avila Belbute-Peres et al. (2018) only simulate 2D scenes with a restricted repertoire of primitives. Liang et al. (2019) simulate cloth but cannot handle rigid body dynamics. We therefore choose the simulation framework of Hu et al. (2019) as our main baseline. This framework uses the material point method (MPM), which leverages particles and grids to model solids. We use the state-of-the-art implementation from the high-performance DiffTaichi library (Hu et al., 2020).

The results are shown in Figure 3. The first row reports the runtime and memory consumption of the two frameworks as the number of objects in the scene increases from 20 to 1000. Our runtime and memory consumption increase linearly in the complexity of the scene, while the MPM-based method scales cubically until it runs out of memory at 200 objects and a  $640^3$  grid.

The second row reports runtime and memory consumption as the relative sizes of the cloth and the bunny are varied

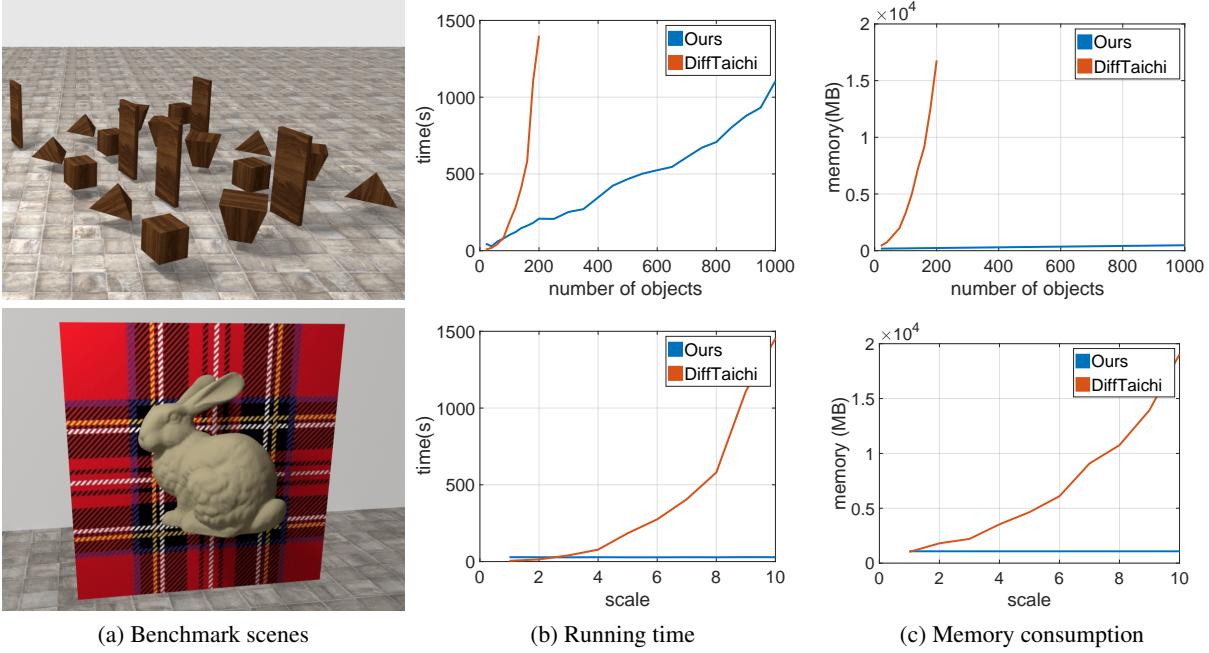


Figure 3. Scalability. (a) Benchmark scenes. Few differentiable simulation frameworks are capable of modeling these scenes. We compare to the expressive MPM-based framework of Hu et al. (2019), implemented in the high-performance DiffTaichi library (Hu et al., 2020). (b,c) Runtime and memory consumption as the scenes are varied in controlled fashion. Top: the number of objects in the scene increases from 20 to 1000. Bottom: the relative scale of the cloth and the bunny increases from 1:1 to 10:1. The MPM-based framework consumes up to two orders of magnitude more memory and computation before it runs out of memory.

from 1:1 to 10:1. The runtime and memory consumption of our method stay constant. In contrast, as the size of the cloth grows, the MPM-based framework is forced to allocate more and more memory and expand greater and greater computation.

These experiments indicate that scalability is a significant advantage of our method. Since we do not need to quantize space, the extent of the scene or the relative sizes of objects do not dominate our runtime and memory consumption. Since our method dynamically detects and handles collisions locally, as needed, rather than setting up a global optimization problem, the runtime scales linearly (rather than quadratically) with scene complexity.

## 7.2. Ablation Studies

We proposed localized collision handling in Section 5 and fast differentiation in Section 6 to make our method scalable to large scenes with many degrees of freedom. We conduct dedicated ablation studies to assess the contribution of these techniques.

**Localized collision handling.** We compare with LCP-based rigid body differentiable simulation developed by de Avila Belbute-Peres et al. (2018). This baseline also uses implicit differentiation to compute derivatives of optimization solvers (Amos & Kolter, 2017). The environment

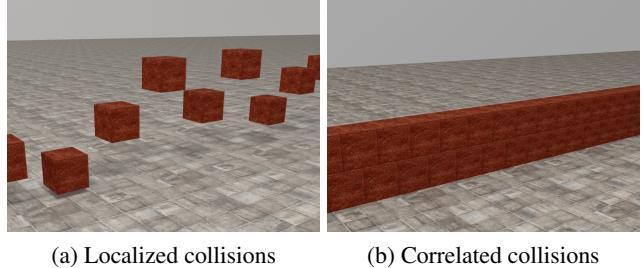


Figure 4. Ablation studies. (a) Comparing global and local collision handling. Cubes are released from the air and fall to the ground. (b) Evaluating the contribution of the fast differentiation scheme. Cubes are densely stacked, forming a big connected component. All collisions need to be solved simultaneously because a move of one cube can affect all others.

is shown in Figure 4(a).  $N (= 100, 200, 300)$  cubes are released above the ground plane. They fall down and hit the ground. We use the same environment for both methods, although the LCP-based framework only simulates in 2D, thus having only half the degrees of freedom compared to our 3D simulation (3 versus 6 per object, accounting for translation and rotation). We disabled our fast differentiation method in this experiment in order to neutralize its effect and conduct a controlled comparison between global and local collision handling.

330 The implementation of de Avila Belbute-Peres et al. (2018)  
 331 uses four threads while our method only uses one. Re-  
 332 sults are reported in Table 1. Our sparse collision handling  
 333 method runs up to 5 times faster than the LCP-based ap-  
 334 proach, and the performance gap widens as the complexity  
 335 of the scene increases.

| # of cubes | 100                                  | 200                                  | 300                                  |
|------------|--------------------------------------|--------------------------------------|--------------------------------------|
| LCP        | $0.73s \pm 0.017s$                   | $2.87s \pm 0.103s$                   | $8.42s \pm 0.190s$                   |
| Ours       | <b><math>0.56s \pm 0.009s</math></b> | <b><math>1.11s \pm 0.012s</math></b> | <b><math>1.65s \pm 0.025s</math></b> |

340 *Table 1.* Runtime of backpropagation (in seconds per simulation  
 341 step) with LCP-based collision handling (de Avila Belbute-Peres  
 342 et al., 2018) and our approach. Simulation of the scene shown in  
 343 Figure 4(a). The LCP-based framework simulates in 2D and uses  
 344 4 threads, while our implementation simulates in 3D and uses 1  
 345 thread. Our approach is faster, and the performance gap widens  
 346 with the complexity of the scene.

347 **Fast differentiation.** We now evaluate the contribution  
 348 of the acceleration scheme described in Section 6. The  
 349 environment is shown in Figure 4(b). N (= 100, 200, 300)  
 350 cubes are stacked in two layers. During collision handling,  
 351 all contacts between cubes form one connected component.  
 352 All constraints need to be solved in one big optimization  
 353 problem. Therefore, the linear system in Equation 9 will be  
 354 large.

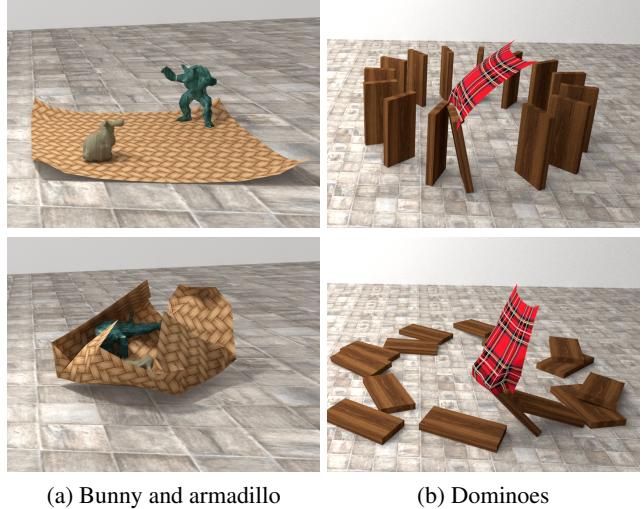
| # of cubes | 100                                  | 200                                  | 300                                  |
|------------|--------------------------------------|--------------------------------------|--------------------------------------|
| W/o FD     | $1.43s \pm 0.015s$                   | $7.76s \pm 0.302s$                   | $21.88s \pm 0.125s$                  |
| Ours       | <b><math>0.41s \pm 0.004s</math></b> | <b><math>0.86s \pm 0.008s</math></b> | <b><math>1.30s \pm 0.008s</math></b> |
| Speedup    | 3.49x                                | 9.02x                                | 16.83x                               |

361 *Table 2.* Runtime of backpropagation (in seconds per simulation  
 362 step) with and without our fast differentiation scheme. Simulation  
 363 of the scene shown in Figure 4(b). N (= 100, 200, 300) cubes  
 364 are stacked in two layers. The impact of the acceleration scheme  
 365 increases with the complexity of the scene.

366 We evaluate the runtime of backpropagation in this scene  
 367 with and without the presented acceleration scheme. The  
 368 ablation condition is referred to as ‘W/o FD’, where FD  
 369 refers to fast differentiation. In this condition, the derivative  
 370 of Equation 6 is calculated by solving Equation 9 directly.  
 371 The results are reported in Table 2. The fast differentiation  
 372 accelerates backpropagation by up to 16x in this scene, and  
 373 the impact of the technique increases with the complexity  
 374 of the scene.

### 375 7.3. Two-way coupling

376 One of the advantages of this presented framework is that it  
 377 can handle both rigid bodies and deformable objects such as  
 378 cloth. This is illustrated in the two scenes shown in Figure 5.  
 379 In Figure 5(a), a Stanford bunny and an armadillo stand on  
 380 a piece of cloth. The cloth is lifted by its corners, envelopes  
 381 the figurines, and lifts them up. In Figure 5(b), a piece  
 382 of cloth strikes a domino, which begins a chain reaction  
 383 that propagates until the last domino strikes the cloth from  
 384 behind. The scenes are also shown in the supplementary  
 385 video.



(a) Bunny and armadillo

(b) Dominoes

386 *Figure 5.* Two-way coupling of rigid bodies and cloth. Top: initial  
 387 state. Bottom: towards the end of the simulation. (a) Two figurines  
 388 are lifted by a piece of cloth. (b) Cloth strikes domino and is  
 389 later struck back. No prior differentiable simulation framework is  
 390 capable of modeling these scenes properly. See the supplementary  
 391 video.

392 of cloth strikes a domino, which begins a chain reaction  
 393 that propagates until the last domino strikes the cloth from  
 394 behind. The scenes are also shown in the supplementary  
 395 video.

396 No prior differentiable simulation framework is capable of  
 397 modeling these scenes properly. Liang et al. (2019) can sim-  
 398 ulate the cloth, but its motion does not affect the figurines.  
 399 The framework of Hu et al. (2019; 2020) does not enforce  
 400 rigidity and suffers from interpenetration. In contrast, our  
 401 approach simulates the correct dynamics and the two-way  
 402 coupling between them.

## 7.4. Applications

403 **Inverse problem.** Differentiable simulation naturally lends  
 404 itself to gradient-based optimization for inverse problems.  
 405 A case study is shown in Figure 6, in which a marble is  
 406 supported by a soft sheet. The sheet is fixed at the four  
 407 corners. In each time step, an external force is applied to  
 408 the marble. The goal is to find a sequence of forces that  
 409 drives the marble to a target position in 2 seconds, while  
 410 minimizing the total amount of applied force. The vertical  
 411 component of the external force is enforced to be 0 so that  
 412 the marble has to interact with the cloth before reaching  
 413 the target. We compare with a derivative-free optimiza-  
 414 tion algorithm, CMA-ES (Hansen, 2016). Our framework  
 415 enables gradient-based optimization in this setting, which  
 416 converges in 4 iterations, reaching a lower objective value  
 417 than achieved by CMA-ES in two orders of magnitude more

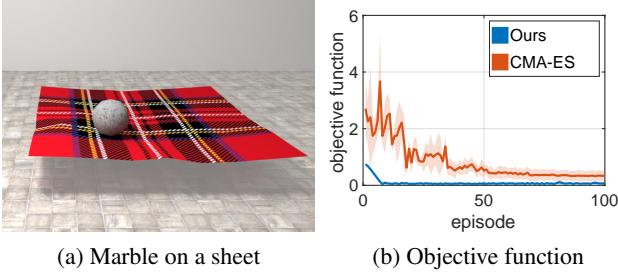


Figure 6. Inverse problem. (a) The goal is to bring the marble to a desired position with minimal force. (b) Our differentiable simulation framework enables gradient-based optimization for this problem, which quickly converges to a lower objective value than derivative-free optimization (CMA-ES). We perform 5 runs with different random seeds; shaded areas represent one standard deviation.

iterations.

**Learning control.** Our second set of case studies uses the differentiable simulator to backpropagate gradients in a neural network that must learn to manipulate objects to achieve desired outcomes. The scenarios are shown in Figure 7. The neural network controls a pair of sticks and a piece of cloth, respectively, which must be used to manipulate an object. The goal is to bring the object to a desired position within 1 second. The optimization objective is the L2 distance to the target. The neural network is an MLP with 50 nodes in the first layer and 200 nodes in the second, with ReLU activations.

For reference, we report the performance of DDPG, a model-free reinforcement learning algorithm (Lillicrap et al., 2016). In each episode, we fix the initial position of the manipulator and the object while the target position is randomized. For both methods, the input to the network is a concatenated vector of relative distance to the target, speed, and remaining time. The DDPG reward is the negative L2 distance to the target.

Our method updates the network once at the end of each episode, while DDPG receives a reward signal and updates the network weights in each time step. As shown in the loss curves in Figure 7, our method quickly converges to a good control policy, while DDPG fails to learn the task on a comparable time scale. This example illustrates the power of gradient-based optimization with our differentiable physics framework.

## 8. Conclusion

We developed a scalable differentiable simulation method that has linear complexity with respect to the number of objects and collisions, and constant complexity with respect to spatial resolution. We use meshes as our core representa-

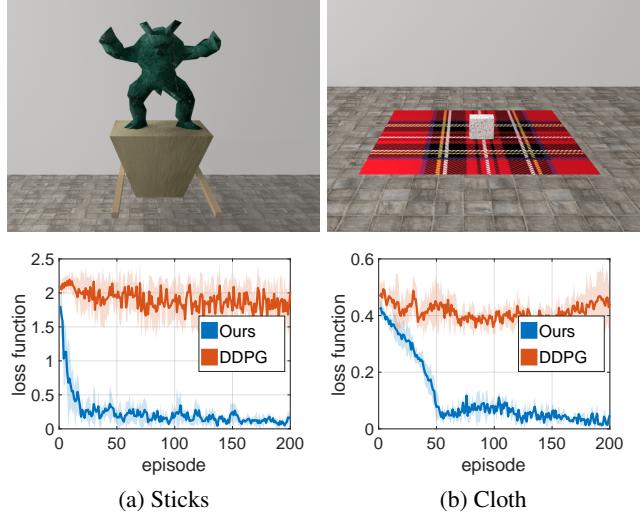


Figure 7. Learning control. A neural network must learn to control (a) a pair of sticks or (b) a piece of cloth. The trained controller must be able to bring the manipulated object to any desired location specified at test time. Our method enables effective gradient-based training of the controller. A model-free reinforcement learning baseline (DDPG) fails to learn the task on a comparable time scale. We perform 5 runs with different random seeds; shaded areas represent one standard deviation.

tion, enabling support for objects of arbitrary shape at any scale. A unified local collision handling scheme enables fast coupling between objects with different materials. An acceleration scheme speeds up backpropagation by implicit differentiation. Experiments show that the presented techniques speed up backpropagation by large multiplicative factors, and their impact increases with the complexity of simulated scenes.

We demonstrated the application of the presented method in a number of case studies, including end-to-end gradient-based training of neural network controllers. Training with differentiable simulation was shown to be significantly more effective than gradient-free baselines.

One avenue for future work is to extend the presented framework to other types of objects, materials, and contact models. The framework is sufficiently general to support deformable solids and articulated bodies, as long as they have mesh-based representations with repulsive contact forces. We plan to extend our implementation to incorporate such object classes.

## References

- Amos, B. and Kolter, J. Z. OptNet: Differentiable optimization as a layer in neural networks. In *ICML*, 2017.
- Battaglia, P. W., Pascanu, R., Lai, M., Rezende, D., and Kavukcuoglu, K. Interaction networks for learning about ob-

- 440 projects, relations and physics. In *Neural Information Processing Systems*, 2016.  
 441  
 442 Botsch, M., Kobbel, L., Pauly, M., Alliez, P., and Lévy, B. *Polygon mesh processing*. AK Peters/CRC Press, 2010.  
 443  
 444 Bridson, R., Fedkiw, R., and Anderson, J. Robust treatment of collisions, contact and friction for cloth animation. *ACM Trans. Graph.*, 2002.  
 445  
 446 Carpenterier, J. and Mansard, N. Analytical derivatives of rigid body dynamics algorithms. In *Robotics: Science and Systems (RSS)*, 2018.  
 447  
 448 Chang, M. B., Ullman, T., Torralba, A., and Tenenbaum, J. B.  
 449 A compositional object-based approach to learning physical dynamics. In *ICLR*, 2017.  
 450  
 451 de Avila Belbute-Peres, F., Smith, K. A., Allen, K. R., Tenenbaum,  
 452 J., and Kolter, J. Z. End-to-end differentiable physics for learning  
 453 and control. In *Neural Information Processing Systems*, 2018.  
 454  
 455 Degrave, J., Hermans, M., Dambre, J., and wyffels, F. A differentiable  
 456 physics engine for deep learning in robotics. *Frontiers in Neurorobotics*, 13, 2019.  
 457  
 458 Hansen, N. The CMA evolution strategy: A tutorial. *arXiv:1604.00772*, 2016.  
 459  
 460 Harmon, D., Vouga, E., Tamstorf, R., and Grinspun, E. Robust  
 461 treatment of simultaneous collisions. *ACM Trans. Graph.*, 2008.  
 462  
 463 Holl, P., Koltun, V., and Thuerey, N. Learning to control PDEs  
 464 with differentiable physics. In *ICLR*, 2020.  
 465  
 466 Hu, Y., Liu, J., Spielberg, A., Tenenbaum, J. B., Freeman, W. T.,  
 467 Wu, J., Rus, D., and Matusik, W. ChainQueen: A real-time  
 468 differentiable physical simulator for soft robotics. In *ICRA*, 2019.  
 469  
 470 Hu, Y., Anderson, L., Li, T., Sun, Q., Carr, N., Ragan-Kelley, J.,  
 471 and Durand, F. DiffTaichi: Differentiable programming for  
 472 physical simulation. In *ICLR*, 2020.  
 473  
 474 Ingraham, J., Riesselman, A., Sander, C., and Marks, D. Learning  
 475 protein structure with a differentiable simulator. In *ICLR*, 2019.  
 476  
 477 Li, Y., Wu, J., Tedrake, R., Tenenbaum, J. B., and Torralba, A.  
 478 Learning particle dynamics for manipulating rigid bodies, de-  
 479 formable objects, and fluids. In *ICLR*, 2019a.  
 480  
 481 Li, Y., Wu, J., Zhu, J.-Y., Tenenbaum, J. B., Torralba, A., and  
 482 Tedrake, R. Propagation networks for model-based control  
 483 under partial observation. In *ICRA*, 2019b.  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494
- Liang, J., Lin, M. C., and Koltun, V. Differentiable cloth simulation for inverse problems. In *Neural Information Processing Systems*, 2019.  
 495  
 496 Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa,  
 497 Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. In *ICLR*, 2016.  
 498  
 499 Macklin, M., Müller, M., Chentanez, N., and Kim, T. Unified  
 500 particle physics for real-time applications. *ACM Trans. Graph.*, 2014.  
 501  
 502 Millard, D., Heiden, E., Agrawal, S., and Sukhatme, G. S. Auto-  
 503 matic differentiation and continuous sensitivity analysis of rigid  
 504 body dynamics. *arXiv:2001.08539*, 2020.  
 505  
 506 Mrowca, D., Zhuang, C., Wang, E., Haber, N., Li, F., Tenenbaum,  
 507 J., and Yamins, D. L. Flexible neural representation for physics  
 508 prediction. In *Neural Information Processing Systems*, 2018.  
 509  
 510 Narain, R., Samii, A., and O'Brien, J. F. Adaptive anisotropic  
 511 remeshing for cloth simulation. *ACM Trans. Graph.*, 2012.  
 512  
 513 Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan,  
 514 G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. Py-  
 515 Torch: An imperative style, high-performance deep learning  
 516 library. In *Neural Information Processing Systems*, 2019.  
 517  
 518 Sanchez-Gonzalez, A., Heess, N., Springenberg, J. T., Merel, J.,  
 519 Riedmiller, M., Hadsell, R., and Battaglia, P. Graph networks  
 520 as learnable physics engines for inference and control. In *ICML*,  
 521 2018.  
 522  
 523 Schenck, C. and Fox, D. SPNets: Differentiable fluid dynamics  
 524 for deep neural networks. In *Conference on Robot Learning (CoRL)*, 2018.  
 525  
 526 Spielberg, A., Zhao, A., Hu, Y., Du, T., Matusik, W., and Rus, D.  
 527 Learning-in-the-loop optimization: End-to-end control and co-  
 528 design of soft robots through learned deep latent representations.  
 529 In *Neural Information Processing Systems*, 2019.  
 530  
 531 Toussaint, M., Allen, K., Smith, K., and Tenenbaum, J. Differentiable  
 532 physics and stable modes for tool-use and manipulation  
 533 planning. In *Robotics: Science and Systems (RSS)*, 2018.  
 534  
 535 Ummenhofer, B., Prantl, L., Thürey, N., and Koltun, V. Lagrangian  
 536 fluid simulation with continuous convolutions. In *ICLR*, 2020.  
 537  
 538 Witkin, A. and Baraff, D. Physically based modeling: Principles  
 539 and practice. SIGGRAPH Course Notes, 1997.

## A. Mass Matrix

For a rigid body  $i$  with rotation  $\mathbf{r} = (\phi, \theta, \psi)^\top$  and translation  $\mathbf{t} = (t_x, t_y, t_z)^\top$ , the generalized coordinates are  $\mathbf{q} = [\mathbf{r}^\top, \mathbf{t}^\top]^\top \in \mathbb{R}^6$ . Its kinetic energy can be computed by  $\mathbf{E} = \frac{1}{2}\dot{\mathbf{q}}\hat{\mathbf{M}}\dot{\mathbf{q}}$ , where  $\hat{\mathbf{M}}$  is a diagonal blocked mass matrix composed of angular and linear inertia,

$$\hat{\mathbf{M}} = \begin{bmatrix} \mathcal{I}_a & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_l \end{bmatrix}. \quad (16)$$

The linear inertia  $\mathcal{I}_l$  is simply  $m\mathbf{I}_{3\times 3}$ , where  $m$  is the total mass of the rigid body.

When the rigid body's distribution is approximated by a set of particles, an angular inertial is given by,

$$\mathcal{I}' = \sum_i m_i (\mathbf{p}_i^\top \mathbf{p}_i \mathbf{I}_{3\times 3} - \mathbf{p}_i \mathbf{p}_i^\top), \quad (17)$$

where  $m_i$  is the mass of the particle  $i$ ;  $\mathbf{p}_i$  is the vector from the center of mass to this particle. Note that this angular inertia corresponds to the axis-angle velocity  $\omega$  in the *world frame*. The angular momentum is  $\mathcal{I}'\omega$ . However, We cannot directly use this formula since we have to represent the angular velocity in terms of velocities of Euler angles.

We used the RPY convention for Euler angle representation: given the Euler angle  $\mathbf{r} = (\phi, \theta, \psi)^\top$ , the rigid body will first rotate about the  $Z$  axis by  $\psi$ , then rotate about the new  $Y'$  axis by  $\theta$ , and finally rotate about the new  $X''$  axis by  $\phi$ .

By the above definition, the angular velocity vector  $\Omega$  can be represented by Euler angles,

$$\Omega = \dot{\psi}\mathbf{e}_Z + \dot{\theta}\mathbf{e}_{Y'} + \dot{\phi}\mathbf{e}_{X''} \quad (18)$$

We convert the angular velocity in local frame back to the world frame so that it matches with the angular inertia in the world frame:

$$\begin{aligned} \Omega = & (\cos \theta \cos \psi \dot{\phi} - \sin \psi \dot{\theta})\mathbf{e}_X \\ & + (\cos \theta \cos \psi \dot{\phi} + \cos \psi \dot{\theta})\mathbf{e}_{Y'} \\ & + (-\sin \theta \dot{\phi} + \dot{\psi})\mathbf{e}_Z \end{aligned} \quad (19)$$

The matrix form is given as:

$$\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} = \begin{bmatrix} \cos \theta \cos \psi & -\sin \psi & 0 \\ \cos \theta \sin \psi & \cos \psi & 0 \\ -\sin \theta & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix}. \quad (20)$$

We denote this transformation as  $\omega = \mathbf{T}\dot{\mathbf{r}}$ , so the angular momentum is reformed as  $\mathcal{I}'\omega = \mathcal{I}'\mathbf{T}\dot{\mathbf{r}}$ . Therefore, in the Euler angle representation, the angular inertia becomes

$$\mathcal{I}_a = \mathbf{T}^\top \mathcal{I}' \mathbf{T}. \quad (21)$$

The new mass matrix  $\hat{\mathbf{M}}$  for the generalized coordinates is,

$$\hat{\mathbf{M}} = \begin{bmatrix} \mathbf{T}^\top \mathcal{I}' \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{3\times 3} \end{bmatrix}. \quad (22)$$

## B. Represent a Vertex using Generalized Coordinates

For a vertex  $p$  attached to the rigid body, its coordinate in the body frame is  $\mathbf{p}_0$ . The origin of the body frame is set to be the center of mass (COM).  $\mathbf{p}_0 = (p_x, p_y, p_z)^\top$  is the relative displacement of  $p$  w.r.t. the COM in the first frame of simulation. The coordinates of the vertex in the world frame are then given by,

$$\mathbf{p} = \mathbf{f}(\mathbf{q}) = [\mathbf{r}]\mathbf{p}_0 + \mathbf{t}, \quad (23)$$

where  $[\mathbf{r}]$  is a rotation matrix represented by the Euler angle  $\mathbf{r} = (\phi, \theta, \psi)^\top$ . The corresponding rotation matrix  $[\mathbf{r}] = \mathbf{R}_{3\times 3}$  is computed by

$$\begin{aligned} \mathbf{R}_{11} &= \cos \theta \cos \psi \\ \mathbf{R}_{12} &= -\cos \phi \sin \psi + \sin \phi \sin \theta \cos \psi \\ \mathbf{R}_{13} &= \sin \phi \sin \psi + \cos \phi \sin \theta \cos \psi \\ \mathbf{R}_{21} &= \cos \theta \sin \psi \\ \mathbf{R}_{22} &= \cos \phi \cos \psi + \sin \phi \sin \theta \sin \psi \\ \mathbf{R}_{23} &= -\sin \phi \cos \psi + \cos \phi \sin \theta \sin \psi \\ \mathbf{R}_{31} &= -\sin \theta \\ \mathbf{R}_{32} &= \sin \phi \cos \theta \\ \mathbf{R}_{33} &= \cos \phi \cos \theta \end{aligned}$$

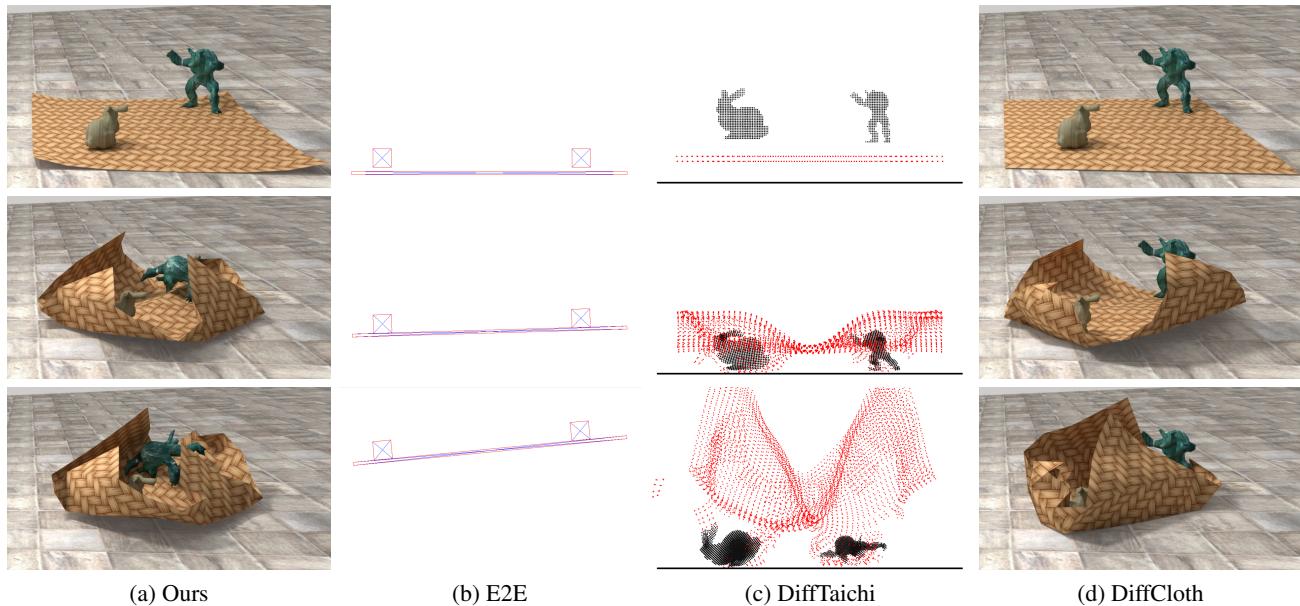
## C. Computation of Derivatives

To backpropagate the gradients at vertex  $\mathbf{p}$  to the generalized coordinates  $\mathbf{q}$ , we need to compute the derivatives  $\partial \mathbf{f}(\mathbf{q}) / \partial \mathbf{q}$ .

The coordinates of a vertex  $(x, y, z)^\top = \mathbf{f}([\phi, \theta, \psi, t_x, t_y, t_z]^\top)$  are given by Equation 23. Therefore, the Jacobian matrix for the partial derivatives is:

$$\nabla \mathbf{f} = \begin{bmatrix} \frac{\partial x}{\partial \phi} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \psi} & 1 & 0 & 0 \\ \frac{\partial y}{\partial \phi} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \psi} & 0 & 1 & 0 \\ \frac{\partial z}{\partial \phi} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \psi} & 0 & 0 & 1 \end{bmatrix}. \quad (24)$$

Assuming that the relative displacement of  $p$  w.r.t. the COM in the first frame is  $\mathbf{p}_0 = (p_x, p_y, p_z)$ , the corresponding elements in the Jacobian matrix are:



**Figure 8.** Qualitative comparison on complex models and interactions. In this scene, we simulate two complex rigid bodies, bunny and armadillo on a cloth. (a) is the simulation result of our method, and the comparisons, (b) is E2E ([de Avila Belbute-Peres et al., 2018](#)); (c) is DiffTaiChi ([Hu et al., 2020](#)); (d) is DiffCloth ([Liang et al., 2019](#)). Our method gets the most reasonable results because we can deal with both rigid body and cloth dynamics. Meanwhile, the mesh representation enables us to model different objects precisely.

$\frac{\partial x}{\partial \psi} = -p_x \cos \theta \sin \psi + p_y (-\cos \phi \cos \psi - \sin \phi \sin \theta \sin \psi) + p_z (\sin \phi \cos \psi - \cos \phi \sin \theta \sin \psi)$

$\frac{\partial x}{\partial \theta} = -p_x \sin \theta \cos \psi + p_y \sin \phi \cos \theta \cos \psi + p_z \cos \phi \cos \theta \cos \psi$

$\frac{\partial x}{\partial \phi} = p_y \sin \phi \sin \psi + \cos \phi \sin \theta \cos \psi + p_z (\cos \phi \sin \psi - \sin \phi \sin \theta \cos \psi)$

$\frac{\partial y}{\partial \psi} = p_x \cos \theta \cos \psi + p_y (-\cos \phi \sin \psi + \sin \phi \sin \theta \cos \psi) + p_z (\sin \phi \sin \psi + \cos \phi \sin \theta \cos \psi)$

$\frac{\partial y}{\partial \theta} = -p_x \sin \theta \sin \psi + p_y \sin \phi \cos \theta \sin \psi + p_z \cos \phi \cos \theta \sin \psi$

$\frac{\partial y}{\partial \phi} = p_y (-\sin \phi \cos \psi + \cos \phi \sin \theta \sin \psi) + p_z (-\cos \phi \cos \psi - \sin \phi \sin \theta \sin \psi)$

$\frac{\partial z}{\partial \psi} = 0$

$\frac{\partial z}{\partial \theta} = -p_x \cos \theta - p_y \sin \phi \sin \theta - p_z \cos \phi \sin \theta$

## D. Qualitative Comparisons

Figure 8 shows qualitative comparisons between our differentiable simulation and rigid bodies interacting with a soft cloth. In this comparison, our differentiable simulation produces realistic results, while E2E cannot simulate the bunny and armadillo from the table. E2E cannot simulate the bunny and armadillo because it only supports rigid bodies. DiffTaichi cannot preserve the *rigid* constraint of the objects because it only supports rigid bodies. Moreover, the cloth would tear during the simulation and obey the dynamics. Moreover, the cloth would tear during the simulation and obey the dynamics. The MPM method cannot keep connectivity fully during the simulation. Diffcloth preserves the motion of cloth, but the cloth cannot interact with the bunny and armadillo, therefore they appear as separate entities.

#### D. Qualitative Comparisons

Figure 8 shows qualitative comparisons of two complex rigid bodies interacting with a soft cloth.

In this comparison, our differentiable physics gets the most realistic results, that both the bunny and armadillo are lifted from the table. E2E cannot simulate the detailed motion of the objects because it only supports basic primitives. DiffTaichi cannot preserve the *rigid* property of bunny and armadillo because these objects are composed of small particles during the simulation and obey deformable body dynamics. Moreover, the cloth would break up because the MPM method cannot keep connectivity information faithfully during the simulation. Diffcloth is able to model the motion of cloth, but the cloth cannot have an impact on the bunny and armadillo, therefore they can not be lifted.