

Synthesizing Diverse and Physically Stable Grasps with Arbitrary Hand Structures using Differentiable Force Closure Estimator

Tengyu Liu¹, Zeyu Liu¹, Ziyuan Jiao¹, Yixin Zhu^{2,3,†}, and Song-Chun Zhu^{2,3,4}

Abstract—Existing grasp synthesis methods are either analytical or data-driven. The former one is oftentimes limited to specific application scope. The latter one depends heavily on demonstrations [1], thus suffers from generalization issues; e.g., models trained with human grasp data would be difficult to transfer to 3-finger grippers. To tackle these deficiencies, we formulate a fast and differentiable force closure estimator, capable of producing diverse and physically stable grasps with arbitrary hand structures, *without any training data*. Although force closure has commonly served as a measure of grasp quality, it has not been widely adopted as an optimization objective for grasp synthesis primarily due to its high computational complexity; in comparison, the proposed differentiable method can test a force closure within milliseconds. In experiments, we validate the proposed method’s efficacy in six different settings.

Index Terms—Grasp synthesis, Dexterous manipulation, Energy-based model, Optimization, Force closure

I. INTRODUCTION

GRASP synthesis has been a challenging task due to the complexity of hand kinematics. Although force closure has been commonly accepted to evaluate the quality of the generated grasps, researchers usually avoid using it as an optimization objective: Computing force closure requires solving for contact forces, which is an optimization problem itself. As a result, using force closure as the optimization objective in grasp synthesis would produce a notoriously slow and nested optimization problem. Instead, researchers have primarily turned to analytical or data-driven methods [1].

Analytical methods use manually derived algorithms. Due to the intrinsic complexity of the grasp synthesis, these methods [2–4] typically perform only in limited settings (usually on power grasps [5]) and are only applicable to specific robotic hand structures. Modern approaches focus more on data-driven methods [6, 7], which rely on large datasets of human demonstrations. Although these methods can reproduce (and even interpolate) similar but different grasps compared to human demonstrations, they are inherently difficult to generalize (especially to extrapolate) to arbitrary hand kinematics and

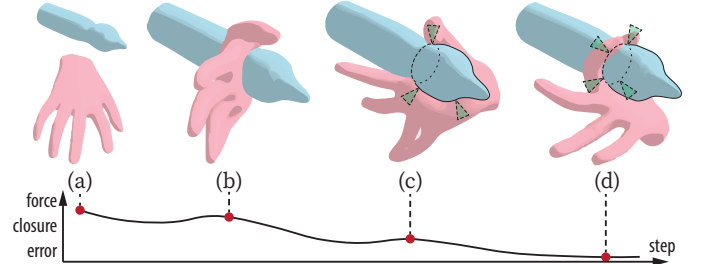


Fig. 1: Grasp synthesis process by minimizing the force closure error. The green trianglelets in (c)(d) denote the friction cones at contact points used to calculate force closure.

unseen grasp types. Furthermore, these data-driven methods usually do not consider the physical stability in producing grasps, making them difficult to deploy on physical robots.

In this paper, we rethink the grasp synthesis problem and derive a differentiable force closure estimator, computed within **milliseconds** on modern desktops. Such fast computation of force closure opens a new venue for grasp synthesis. Since it does not rely on training data or restrict to specific robotic hand structures, our method can be applied to arbitrary hand structures to synthesize physically stable and diverse grasps.

Specifically, our method is on the basis of two simple yet reasonable and effective assumptions: **zero friction** and **equal magnitude of contact forces**, which avoid solving the contact forces as an optimization problem. Intuitively, such assumptions indicate that the contact force on each contact point becomes simply the object’s surface normal on that point. Consequently, the overall nested optimization problem is converted to minimizing the errors that violate the above assumptions; see an example in Fig. 1. In a series of experiments, we demonstrate that our estimated error reflects the difference between surface normal vectors and force closure contact force vectors. We further devise a grasp energy function based on the estimated force closure and validate the force-closure grasp synthesis by minimizing the energy function in six settings.

This paper makes three primary contributions: (i) We formulate a fast and differentiable estimator of force closure, computed within milliseconds. (ii) We synthesize diverse grasps with arbitrary hand structures without any training data. (iii) Since our method is independent of specific hand structures, grasping and manipulation algorithms built upon our method would be easily transferable among competitions and benchmarks that require different end-effectors.

Manuscript received: April, 1st, 2021; Revised July, 16th, 2021; Accepted August, 19th, 2021.

This paper was recommended for publication by Editor Tamim Asfour upon evaluation of the Associate Editor and Reviewers’ comments.

† Corresponding email: y@bigai.ai. See additional material on <https://sites.google.com/view/ral2021-grasp/>.

¹ UCLA Center for Vision, Cognition, Learning, and Autonomy (VCLA).

² Beijing Institute for General Artificial Intelligence (BIGAI).

³ Peking University. ⁴ Tsinghua University.

Digital Object Identifier (DOI): see top of this page.

II. RELATED WORK

Grasp synthesis literature can be roughly categorized into two schools of thought: analytic and data-driven approach.

The analytic approach generates grasps by considering kinematics and physics constraints [8]. Although force closure has been commonly adopted as the physics constraint [9–12], primary efforts focus on simplifying the search space (*e.g.*, [2–4]) as testing force closure is expensive. However, these methods are only effective in specific settings or applications.

The data-driven approach leverages recent advancements in machine learning to estimate grasp points. Despite promising progress [13–15], this approach relies heavily on large datasets to learn successful grasps, with a particular focus on grippers with limited DoF. Although recent literature [16–18] extends this approach to more complex hand models, it still relies on the expensive and tedious collection of human demonstration data. Fundamentally, it is non-trivial for a data-driven approach to generalize the learned model to other hand kinematics.

An example that does not fall into either of the above categories is the popular toolkit of GraspIt! [19]. It generates grasps by initializing hand pose randomly, squeezing the fingers as much as possible, and ranking them by a user-defined grasp metric (*e.g.*, a force closure metric). Although this method can generate valid grasps, it is highly inefficient and incapable of generating diverse grasps [20].

A **force-closure** grasp is a grasp with contact points $\{x_i \in \mathbb{R}^3, i = 1, \dots, n\}$ such that $\{x_i\}$ can resist arbitrary external wrenches with contact forces f_i , where f_i lies within the friction cones rooted from x_i . The angles of the friction cones are determined by the surface friction coefficient: The stronger the friction, the wider the cone. The force-closure metric is, therefore, irrelevant to the actual hand pose, but only relevant to the contact points and friction cones. To test whether a set of contact points form a force-closure grasp, the first step is solving an optimization problem regarding contact forces rooted from the points [21, 22]. Although various methods have been devised, they all require iterations to jointly solve an auxiliary function, *e.g.*, a support function [23], a bilinear matrix inequality [24], or a ray shooting problem [25]. As such, solving force-closure grasps under the constraint of hand kinematics and force closure is a nested optimization problem.

Various methods were proposed to fast approximate this optimization problem, including friction cone approximation with ellipsoids [26] and data-driven force closure estimation [14]. The former method is conceptually similar to this paper but would slow down exponentially as the number of contact points increases, whereas our method is more computationally robust (see Section V-B). The latter method depends heavily on training data and thus suffers from the same problems as data-driven grasp synthesis algorithms do.

Human grasps, organized into a **grasp taxonomy** [5], provide different levels of power and precision. Most existing grasp synthesis methods focus on synthesizing power grasp, either analytical [9–12] or data-driven [7]. At a high cost of annotating object-centric grasp contact information, some data-driven approaches [6, 17] demonstrate a certain level of capability to generate a broader range of grasp types.

III. DIFFERENTIABLE FORCE CLOSURE

Formally, given a set of n contact points $\{x_i \in \mathbb{R}^3, i = 1, \dots, n\}$ and their corresponding friction cones $\{(c_i, \mu)\}$, where c_i is the friction cone axis and μ is the friction coefficient, a grasp is in *force closure* if there exists contact forces $\{f_i\}$ at $\{x_i\}$ within $\{(c_i, \mu)\}$ such that $\{x_i\}$ can resist arbitrary external wrenches. We follow the notations in Dai *et al.* [24] to define a set of contact forces to be force closure if it satisfies the following constraints:

$$GG' \geq \epsilon I_{6 \times 6}, \quad (1a)$$

$$Gf = 0, \quad (1b)$$

$$f_i^T c_i > \frac{1}{\sqrt{\mu^2 + 1}} |f_i|, \quad (1c)$$

$$x_i \in S, \quad (1d)$$

where S is the object surface, and

$$G = \begin{bmatrix} I_{3 \times 3} & I_{3 \times 3} & \dots & I_{3 \times 3} \\ [x_1]_{\times} & [x_2]_{\times} & \dots & [x_n]_{\times} \end{bmatrix}, \quad (2)$$

$$[x_i]_{\times} = \begin{bmatrix} 0 & -x_i^{(3)} & x_i^{(2)} \\ x_i^{(3)} & 0 & -x_i^{(1)} \\ -x_i^{(2)} & x_i^{(1)} & 0 \end{bmatrix}. \quad (3)$$

The form of $[x_i]_{\times}$ ensures the cross product $[x_i]_{\times} f_i = x_i \times f_i$, where $f = [f_1^T f_2^T \dots f_n^T]^T \in \mathbb{R}^{3n}$ is the unknown variable of contact forces. In Eq. (1a), ϵ is a small constant. $A \geq B$ means $A - B$ is positive semi-definite, *i.e.*, it is symmetric, and all its eigenvalues are non-negative. Eq. (1a) states that G is full rank. Eq. (1b) states that the contact forces cancel out each other so that the net wrench is zero. Eq. (1c) prevents f_i from deviating from the friction cone $\{(c_i, \mu)\}$. Eq. (1d) constrains contact points to be on the object surface.

Relaxation: Of note, Eq. (1b) is bilinear on x_i and f_i . Given a set of contact points $\{x_i\}$, verification of force closure requires finding a solution of $\{f_i\}$. The time complexity for computing such a solution is linear w.r.t. the number of contact points [24]. Here, we rewrite Eq. (1b) to

$$Gf = G(f_n + f_t) = 0, \quad (4a)$$

$$G \frac{f_n}{\|f_n\|_2} = -\frac{Gf_t}{\|f_n\|_2}, \quad (4b)$$

$$Gc = -\frac{Gf_t}{\|f_n\|_2}, \quad (4c)$$

where f_n and f_t are the normal and tangential components of contact force f in the force closure model, and $c = [c_1^T c_2^T \dots c_n^T]^T$ is the set of friction cone axes. We obtain c_i as the surface normal of the object on x_i , which is easily accessible in many shape representations. We use Gc to approximate Gf , and therefore relax Eq. (1) to

$$GG' \geq \epsilon I_{6 \times 6}, \quad (5a)$$

$$\|Gc\|_2 < \delta, \quad (5b)$$

$$x_i \in S, \quad (5c)$$

where δ is the maximum allowed error introduced from our relaxation. By adopting Eq. (5), we no longer need to solve the unknown variable f . The constraints of x_i becomes quadratic.

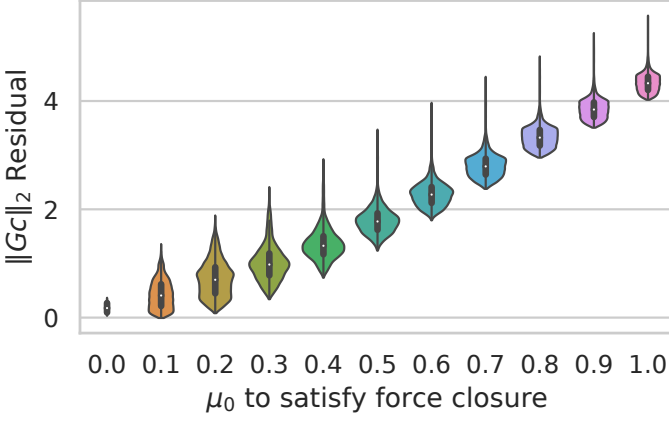


Fig. 2: $\|Gc\|_2$ residual from Eq. (5b) (y-axis) against minimum friction coefficient μ_0 (x-axis). The violinplot is the distribution of $\|Gc\|_2$ residuals of all examples that require a minimum friction coefficient μ_0 to pass the classic force closure test. Overall, these two values are linearly correlated; $\|Gc\|_2 \approx 4.035\mu_0$.

Hence, the verification of force closure can now be computed extremely fast. The residual in $\|Gc\|_2$ reflects the difference between contact forces and friction cone axes.

To allow gradient-based optimization, we further cast Eq. (5) as a soft constraint in the form

$$FC(x, O) = \lambda_0(GG' - \epsilon I_{6 \times 6}) + \|Gc\|_2 + w \sum_{x_i \in x} d(x_i, O), \quad (6)$$

where $\lambda_0(\cdot)$ gives the smallest eigenvalue, and $d(x, O)$ returns the distance from point x to the surface of object O . The scalar w controls the weight of the distance term. By minimizing the three terms, we are looking for $\{x_i\}$ that satisfies the three constraints in Eq. (5), respectively.

Implications of Assumptions: Using surface normal vectors to approximate contact forces implies zero friction and equal magnitude contact forces. Such an assumption may seem to eliminate a large pool of force-closure contact-point compositions. In practice, however, this is not the case: A residual in $\|Gc\|_2$ indicates that the existence of friction f_t and difference in force magnitude f_n on contact forces. By allowing the residual to be smaller than a reasonable threshold δ , we allow the tangential and normal components of the contact forces to deviate within a reasonable range.

To further verify our interpretation, we randomly sample 500,000 grasps, each containing three contact points on the surface of a unit sphere. For each grasp, we compute the *minimum* friction coefficient μ_0 required for the grasp to satisfy the classic force closure constraints described in Eq. (1). Fig. 2 plots the residual $\|Gc\|_2$ from Eq. (5b) against μ_0 ; it shows an almost linear relation between μ_0 and $\|Gc\|_2$.

Force-closure Contact-point Generation: By directly minimizing the soft force closure constraint, we can synthesize force closure contact points with *arbitrary* shapes. Specifically, we run gradient descent on contact point positions x to minimize Eq. (6). Fig. 3 shows the computed contact points on a unit sphere and some daily objects. Despite our assumptions, minimizing our force closure estimation can indeed properly generate force-closure contact points.

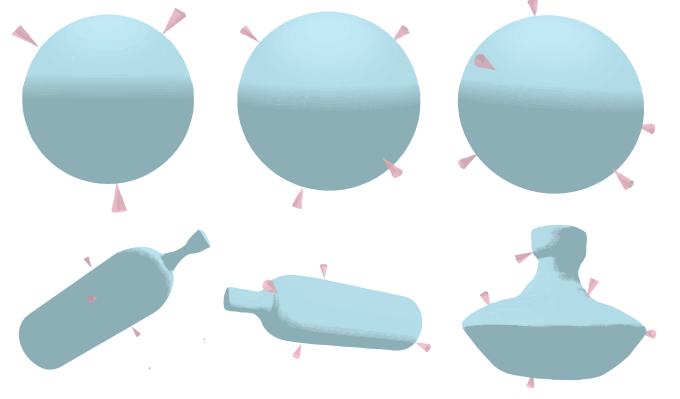


Fig. 3: Force-closure contact-point generations on unit spheres (top) and daily objects (bottom) by minimizing Eq. (6). Objects in each columns have 3, 4, and 5 contact points, respectively.

IV. GRASP SYNTHESIS

In this section, we first describe how to leverage our differentiable force closure estimator to formulate a probability distribution of grasping. Next, we devise an optimization algorithm to sample diverse grasps from the distribution.

Formulation: We formulate the grasp synthesis problem as sampling from a conditional Gibbs distribution:

$$P(H|O) = \frac{P(H, O)}{P(O)} \propto P(H, O) = \frac{1}{Z} \exp^{-E(H, O)}, \quad (7)$$

where Z denotes the intractable normalizing constant, H the hand, O the object, and $E(H, O)$ the energy function. We rewrite $E(H, O)$ as the minimum value of the energy function $E_{\text{grasp}}(H, x, O)$ w.r.t. contact point choices x :

$$\begin{aligned} E(H, O) &= \min_{x \subset S(H)} E_{\text{grasp}}(H, x, O) \\ &= \min_{x \subset S(H)} FC(x, O) + E_{\text{prior}}(H) + E_{\text{pen}}(H, O), \end{aligned} \quad (8)$$

where $S(H)$ is a set of points sampled uniformly from the surface of a hand with pose H . We denote the selected contact points from hand surface as $x \subset S(H)$. $FC(x, O)$ is the soft constraint from Eq. (6). $E_{\text{prior}}(H)$ is the *energy prior* of the hand pose. Its exact form depends on the hand definition. The *penetration energy* is defined as $E_{\text{pen}}(H, O) = \sum_{v \in S(H)} \sigma(v, O)$, where $\sigma(v, O)$ is a modified distance function between a point v and an object O :

$$\sigma(v, O) = \begin{cases} 0 & \text{if } v \text{ outside } O \\ |d| & \text{otherwise} \end{cases}, \quad (9)$$

where d is the distance from v to surface of O .

Algorithm: Due to the complexity of human hand kinematics, our grasp energy suffers from a complex energy landscape. A naïve gradient-based optimization algorithm is likely to stop at sub-optimal local minima. We use a modified Metropolis-adjusted Langevin algorithm (MALA) to overcome this issue; see the algorithm details in Algorithm 1. The random walk aspect of Langevin dynamics provides the chance of escaping bad local minima. Our algorithm starts with random initialization of hand pose H and contact points $x \subset S(H)$. Next, we run our algorithm L iterations to update H, x and maximize

$P(H, O)$. In each iteration, our algorithm randomly decides to update either the hand pose by Langevin dynamics or one of the contact points to a point uniformly sampled from the hand surface. The updates are accepted or rejected according to the Metropolis-Hastings algorithm, in which a lower-energy update is more likely to be accepted than a higher-energy one.

Algorithm 1: Modified MALA Algorithm

Input: Energy function E_{grasp} , object shape O , step size η , Langevin steps L , switch probability ρ

Output: grasp parameters H, x

```

1 Initialize  $H, x$ 
2 for step = 1 :  $L$  do
3   if rand() <  $\rho$  then
4     Propose  $H^*$  according to Langevin dynamics
       
$$H^* = H - \frac{\eta^2}{2} \frac{\partial}{\partial H} E_{\text{grasp}}(H, x, O) + \eta \epsilon,$$

       where  $\epsilon \sim N(0, 1)$  is a Gaussian noise
5   else
6     Propose  $x^*$  by sampling from  $S(H)$ 
7   end
8   Accept  $H \leftarrow H^*, x \leftarrow x^*$  by Metropolis-Hastings
   algorithm using energy function  $E_{\text{grasp}}$ 
9 end

```

Of note, different compositions of contact points correspond to different grasp types as they contribute to some of the classification basis of the grasp taxonomy, including virtual finger assignment and opposition type. Hence, sampling contact points on Line 6 in Algorithm 1 is crucial for exploring different types of grasps. In practice, we also empirically find that this step is essential for escaping bad local minima.

V. SIMULATION

We detail experimental setup with analysis in simulation.

A. Simulation Setup

Hand Model: We use MANO [27] to model the humanoid hand. It is a parameterized 3D hand shape model that maps low-dimensional hand poses to 3D human hand shapes. We use the norm of the PCA weights of the hand pose as $E_{\text{prior}}(H)$. Since MANO vertices are distributed uniformly across the hand surface, we sample points from the hand surface by directly sampling from MANO vertices.

Object Model: We use the DeepSDF model [28] to model the objects to be grasped. DeepSDF is a densely connected neural network that implicitly represents the shape surface and estimates the signed distance from a position to an object surface; the signed distance is negative if the point is inside the object, and vice versa. The 0-level set composes the surface of the object. We obtain the object surface normal by taking the derivative of the signed distance w.r.t. the input position.

We test our grasp synthesis algorithm on various bottles retrieved from ShapeNet dataset [29]. Given the pre-trained DeepSDF model of an object, we randomly initialize a MANO hand and use Algorithm 1 to sample the hand pose and contact points from $P(H|O)$. We set the step size $\eta = 0.1$, switch probability $\rho = 0.85$, distance weight $w = 1$, and Langevin

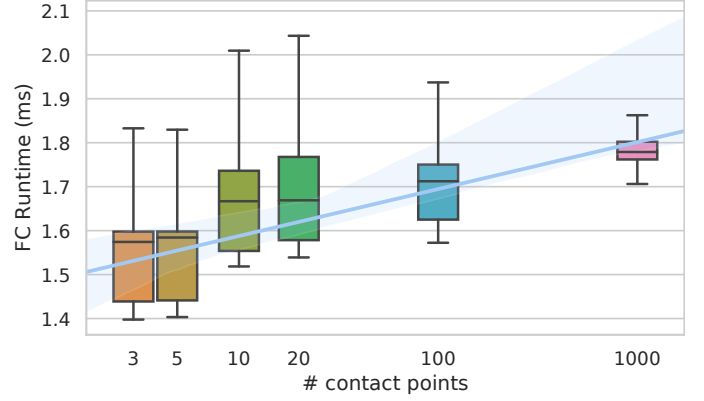


Fig. 4: **Boxplot and log-linearly fitted curve of the runtime of Eq. (6) w.r.t. to the number of contact points.** We run a simulated test of force closure with 3, 5, 10, 20, 100, and 1000 contact points for 1,000 iterations. X-axis is the number of contact points in log scale. Y-axis is the runtime of our force closure error estimate. The shaded area denotes the 95% confidence interval. The light blue line denotes that the estimated relation between the FC runtime t and the number of contact points n is $t \approx 0.107 \times \log_{10} n + 1.502$.

steps $L = 10^6$. We filter out samples in bad local minima by keeping samples that satisfy following empirical constraints:

$$\|Gc\|_2 < 0.5 \quad (10a)$$

$$\sum_{x_i \in x} d(x_i, O)^2 < 0.02 \quad (10b)$$

$$E_{\text{pen}}(H, O) < 0.02 \quad (10c)$$

where x is the set of contact points on the hand surface, and c the friction cone axes at contact points.

B. Runtime Analysis

Fig. 4 shows the time complexity of testing force closure using Eq. (6) and our simulation setup, wherein we further fit a log-linear curve of the running time w.r.t. the number of contact points. Each test takes 1-2ms to run on an NVIDIA 3090 GPU, significantly faster than the exact solution [24]. We also observe that roughly 80% of the total runtime is spent at the computation of surface normal; this operation is particularly slow because it takes a derivative of the DeepSDF model. Taken together, these empirical results in simulation indicate that a further improvement in runtime efficiency would be achievable with a more computationally tractable object shape representation, an interesting future research direction.

We further examine the efficiency of Algorithm 1 under different settings in Fig. 5. We synthesize 512 examples for 10^4 steps under each setting and count the number of successful synthesis results that satisfy Eq. (10) at each step. We observe that the algorithm is more likely to succeed with a smaller search space of contact point selection (less contact points and less candidates) and with simpler object shapes. For complex shapes such as bottles, our method only produced 5 successful syntheses after 10^4 steps, and the first acceptable synthesis emerged at 1047th step. On average, each step takes 224.4ms for bottles, of which over 200ms is spent on computing the gradient of the force closure estimate—it involves computing the second-order derivative of the DeepSDF function.

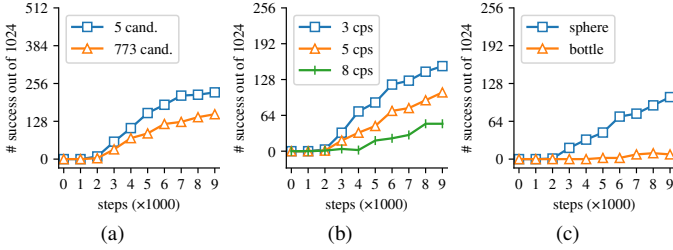


Fig. 5: **Runtime analysis of grasp synthesis.** (a) Grasp spheres with 3 contact point candidates (one on each fingertip) and 773 candidates (uniformly distributed over entire hand surface). (b) Grasp spheres with 3, 5, and 8 contact points. (c) Grasp spheres and ShapeNet bottles with 3 contact points.

C. Refinement

While our modified MALA algorithm can produce realistic results, we still observe physical inconsistencies in the synthesized examples such as penetrations and gaps between contact points and object surface. To resolve these issues, we further refine the synthesized results by minimizing E_{grasp} using gradient descent on H . We do not update the contact point selection x in this step, since we hope to focus on optimizing the physical consistency in this step rather than exploring the grasp landscape for diverse grasp types.

VI. RESULTS

We showcase our method’s capabilities in simulation.

A. Grasp Synthesis

Fig. 6 shows synthesis results with and without the refinement step: Higher values of our force closure estimation corresponds to non-grasps, whereas force closure estimations close to zero are as good as the ones with force closure estimations equal to zero. **This observation confirms our previous analysis.** We also notice cases when the synthesis is trapped in bad local minima; we show two examples in the last column of Fig. 6. These examples exhibit large values in our force closure estimator, which happened due to the non-convexity of the optimization problem; one cannot avoid every bad minimum with gradient-based methods. Fortunately, we can identify these examples by their high force closure scores.

B. Physical Stability

We verify the physical stability of our synthesized examples by simulating the samples in PyBullet. Specifically, we set gravity to be $[0, 0, -10]m/s^{-2}$ and use the default values of friction coefficients in PyBullet. We assumed both the hand and the object to be rigid bodies. An example is deemed to be a successful grasp if the object’s vertical drop is less than $0.3m$ after 1000 steps of simulation, or $16.67s$.

A grasp’s physical stability depends on the force closure score of the contact points and whether the contact points are close enough to the object surface. We set two thresholds on the contact point distance; Table I tabulates detailed comparisons of the success rate between our method against state-of-the-art algorithms: To our best knowledge, Przybylski *et*

TABLE I: Grasp success rates: ours vs. state-of-the-art methods

method	success rate
Unions of Balls [30]	72.53%
Visuo-Haptic [31]	85.00%
Ours ($\sigma < 0.0015$)	76.98%
Ours ($\sigma < 0.0005$)	85.00%

al. [30] is the state-of-the-art analytic approach, whereas Ottenhaus *et al.* [31] is the state-of-the-art data-driven approach.

Of note, although Ottenhaus *et al.* [31] reported 95% success rate in the original paper, many of the objects being tested have simple shapes, such as a sphere or a box; the success rate would drop to 85% when we remove these simple objects. Additionally, neither of the two state-of-the-art methods has demonstrated the ability to synthesize diverse types of grasps. Although some other data-driven methods have demonstrated a certain level of diverse grasp synthesis, they fail to report their physical stability as it is not their primary focus.

C. Diversity of the Grasp Types

To evaluate the diversity of the grasps generated by the proposed method, we examine the energy landscape of our grasp energy function; we use ADELM algorithm [32] to build the energy landscape mapping of our grasps energy function $E(H, O)$. Below, we show that grasps defined by our energy function loosely aligns with the carefully defined grasps taxonomy [5] when applied to humanoid hands.

Specifically, we collected 371 synthesized grasp examples and adopted the ADELM algorithm [32] to find minimum energy pathways (MEPs) between them. We project the MEPs between examples to a disconnectivity graph in Fig. 7. In the disconnectivity graph, each circle at the bottom represents a local minima group. The size of the circle indicates how many synthesized examples fall into this group. The height of the horizontal bar between two groups represent the maximum energy (or energy barrier) along the MEPs between two groups. The MEPs with lowest barriers connect smaller groups into larger groups, and this process is repeated until all examples are connected. The produced disconnectivity graph is an estimation of the true landscape of the energy function. Energy landscape mapping in Fig. 7 shows that the local minima with low energy barriers between them have similar grasps, and those with high energy barriers between them tend to have different grasps. We also observe that the energy landscape contains all three categories in the power and the precision dimension as described in Feix *et al.* [5].

To provide a more comprehensive understanding of the alignment between our energy landscape and the existing taxonomy, we further plot the local minima groups as a 2D graph in Fig. 8, which supplements the 1D energy landscape shown in Fig. 7. In Fig. 8, each node represents a local minima group. We arrange the nodes so that groups with low energy barriers between them are placed closer to each other. The edges between nodes also indicate the energy barriers: Thicker edges indicate lower barriers, and no edge between two nodes means no pathway between the two groups is formed.

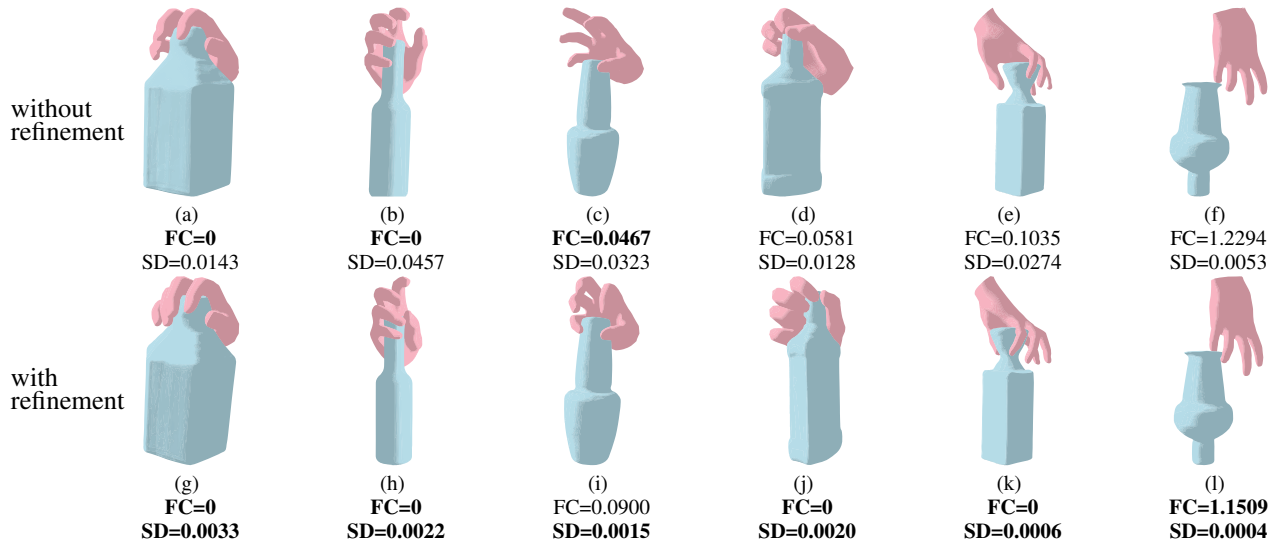


Fig. 6: **Examples of synthesized grasps.** Top: synthesized grasps before refinement. Bottom: the same set of synthesized grasps after refinement. FC: estimated force closure error. SD: mean distance from each contact point to the object surface. Left to right: examples with zero FC error, small FC error, and high FC error qualitatively illustrate how our estimation of force closure correlates to grasp quality.

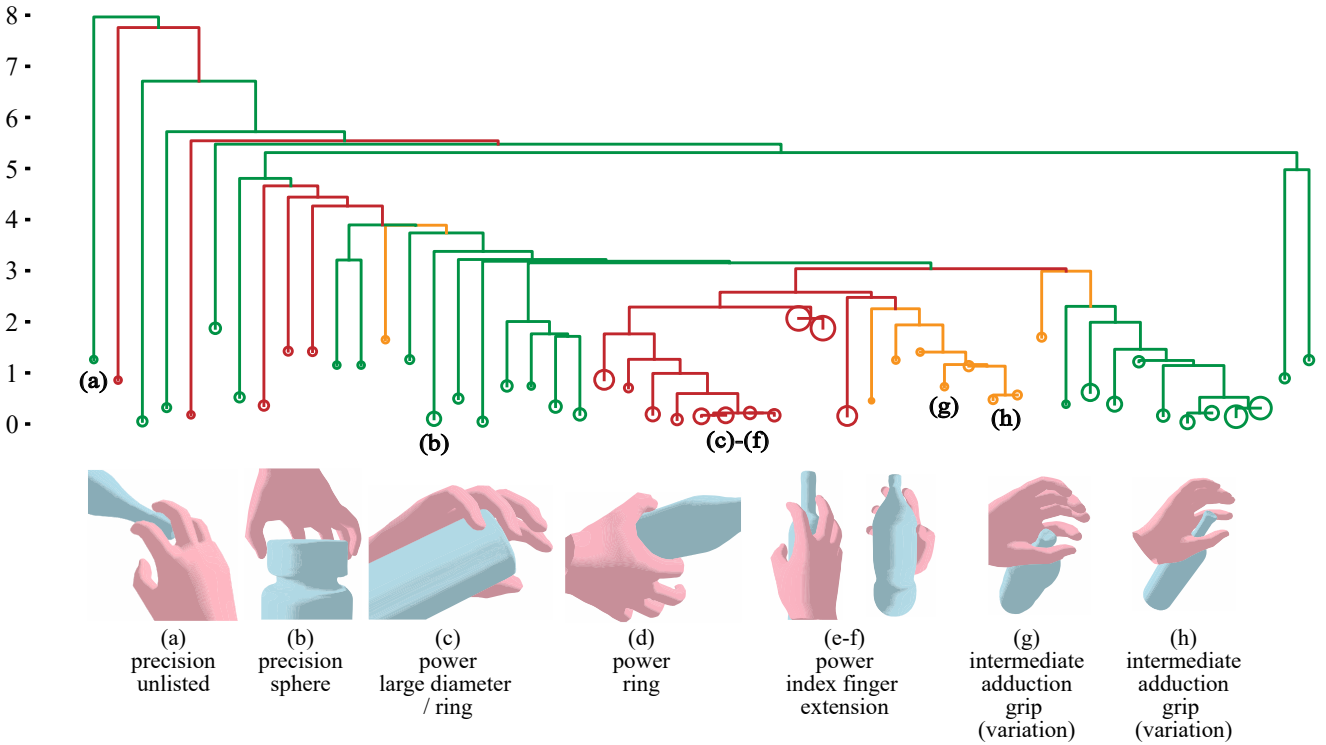


Fig. 7: **Energy landscape mapping generated by the ADELM algorithm [32]; best viewed in color.** Top: disconnectivity diagram of the energy landscape of our energy function $E(H, O)$. Green minima denote precision grasps, red power grasps, and yellow intermediate grasps. Bottom: examples from selected local minima; minima with lower energy barriers in between have similar grasps. We also label the grasp taxonomy of each example according to Feix *et al.* [5]. Examples marked as *unlisted* do not belong to any manually classified type.

Specifically, Fig. 8a shows that the power grasps and precision grasps are mostly separate from each other, indicating a high energy barrier between the two. One interpretation is that there is no smooth transition between a power grasp and a precision grasp without a non-force-closure grasp along the transition. Intermediate grasps are scattered around. Nodes that are not colored are grasp types not listed in any existing grasp taxonomy, indicating the manually-defined grasp taxonomy,

though carefully collected and designed, may still fall short when facing a large variety of grasps in various applications.

In Fig. 8b, we draw various types of power grasps in different colors. Only the power grasps close to the precision grasps belong to the power sphere type. This observation matches our intuition as a power sphere grasp is similar to a precision sphere grasp, with a slight difference in the distance between the object and the palm. In other words, there exists

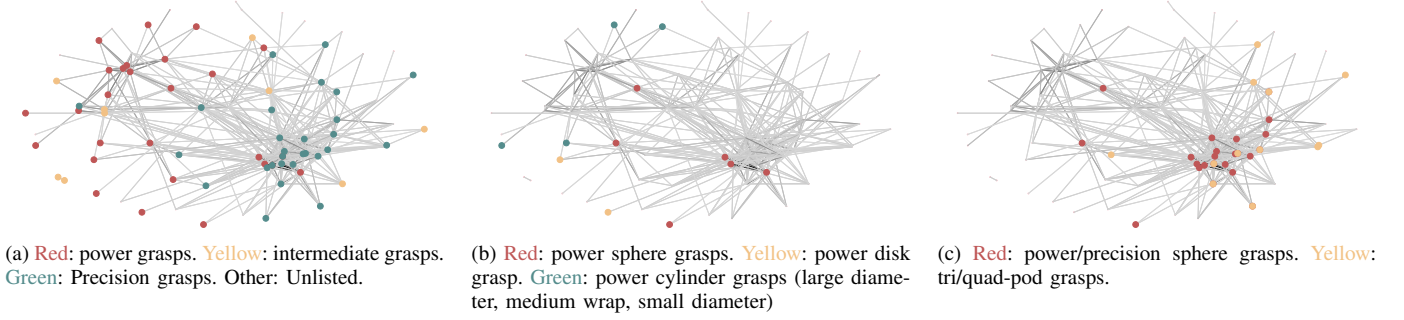


Fig. 8: Alignment between our energy landscape and existing grasp taxonomy [5]; best viewed in color.

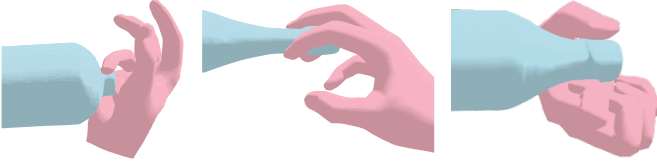


Fig. 9: **Examples of novel grasp poses.** To the best of our knowledge, these newly discovered grasp poses do not correspond to any grasp types in existing human-designed grasp taxonomy (*e.g.*, [6, 20]).

a smooth transition between a precision sphere grasp and a power sphere grasp, such that all snapshots along the transition are force-closure grasps. Please refer to Feix *et al.* [5] for details about the power and precision sphere grasps.

In Fig. 8c, we observe that sphere grasps and tri- or quadrupod grasps are close to each other. This observation is also expected since many sphere grasps can be converted to tri- or quadrupod grasps by merely lifting one or two fingers.

We further demonstrate that our algorithm can find natural but novel grasps in Fig. 9. These grasps are rarely collected in any of the modern 3D grasp datasets (*e.g.*, [6, 20]), since they do not belong to any type as defined in the grasp taxonomy. However, these grasps are valid grasps and could well exist during physical manipulations. For example, the left example in Fig. 9 is commonly used to twist-open a bottle when some of the fingers are occupied or injured. The second example would occur if one is already holding something (*e.g.*, a ball) in the palm while picking up another bottle-like object.

These grasps occur because the human hand is excellent in doing multiple tasks simultaneously, which have not been recognized or explored in grasp literature as we always assumed otherwise. Such limitation would hinder a robotic hand’s capacity from developing to its full potential. Our method paves the way to explore grasp types beyond the grasp taxonomy, which is a crucial step toward exploiting the total capacity of a complex hand structure such as human hands.

D. Grasp Synthesis for Arbitrary Hand Structures

Although above experiments primarily rely on MANO for hand modeling and grasp taxonomy, our method in fact makes no assumption on the hand kinematics except for having a differentiable mapping between pose and shape. As a result, we can synthesize grasps for arbitrary hand so long as there exists such a mapping. In Fig. 10, our method, without

modifications, can directly synthesize grasps of a MANO hand with its thumb removed and a Robotiq 3-finger gripper. Specifically, for the 3-finger gripper, we used a differentiable forward kinematics [33] as the mapping from joint states to the hand shape. These examples demonstrate that our method can explore a wide range of grasps for arbitrary hand structure, which could provide valuable insights for understanding the task affordance of prosthetic or robotic hands, and hands with injuries or disabilities. Our method is also applicable to animations for grasps of non-standard hands or claws.

E. Limitations

We show two representative failure cases in Fig. 11, wherein an unstable or unrealistic grasp receives a low force closure score. In our experiment, most failure cases are caused by concavities in object shapes. For concave shapes, the force closure requirement is sometimes satisfied with a single finger in the concavity, providing contact forces in opposing directions. The issue may be eliminated with manually defined heuristics, such as enforcing contact points on different fingers or encouraging contact points to have larger distances between each other. Another common failure comes from model intersections. We test penetration by computing the signed distance between hand surface vertices and the object shape. When the vertices are sparse, or the object has a pointy part, it is possible for the object to penetrate the hand without being detected. This issue can be addressed with a dense sample of hand surface vertices or by adopting a differentiable mesh intersection algorithm.

Another primary limitation of our approach lies in the gap between the simulation and the reality. Our algorithm assumes perfect knowledge of the object shape and its signed distance field. Inferring such properties from perception is non-trivial. We plan to address this issue in future studies.

VII. CONCLUSION

We formulated a fast and differentiable approximation of the force closure test computed within milliseconds, which enables a new grasp synthesis algorithm. In a series of experiments, we verified that our force closure estimator correctly reflects the quality of a grasp, and demonstrated the proposed grasp synthesis algorithm could generate diverse and physically stable grasps with arbitrary hand structures. The diversity of the generated grasps is validated by its alignment with widely accepted grasp taxonomy with newly discovered grasp types.

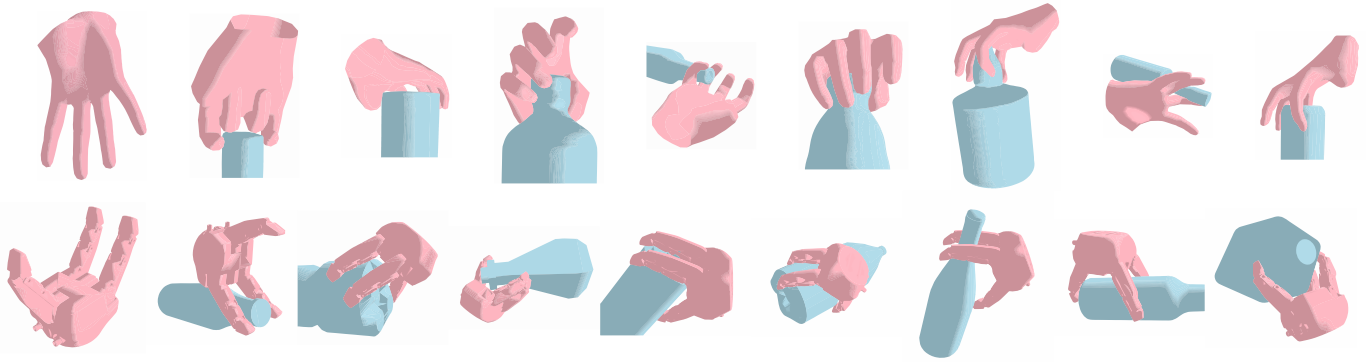


Fig. 10: **Synthesized grasps of different hands using our formulation.** Top: A MANO hand with its thumb removed. Bottom: A Robotic 3-finger gripper. The left-most figure shows the hand used in each row.

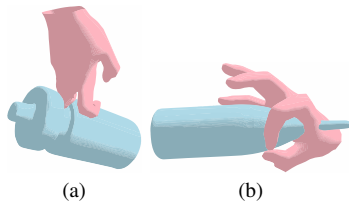


Fig. 11: **Examples of failure cases.** (a) Concave shape results in force closure configuration that is not a grasp. (b) Sparse penetration detection leads to intersection.

We believe that exploring different grasp types is crucial for future work of understanding the hand's total functional capacity, whether it is a prosthetic hand, a robotic hand, or an animated character's hand.

REFERENCES

- [1] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis—a survey," *T-RO*, vol. 30, no. 2, pp. 289–309, 2013.
- [2] J. Ponce, S. Sullivan, J.-D. Boissonnat, and J.-P. Merlet, "On characterizing and computing three- and four-finger force-closure grasps of polyhedral objects," in *ICRA*, 1993.
- [3] J. Ponce, S. Sullivan, A. Sudsang, J.-D. Boissonnat, and J.-P. Merlet, "On computing four-finger equilibrium and force-closure grasps of polyhedral objects," *IJRR*, vol. 16, no. 1, pp. 11–35, 1997.
- [4] J.-W. Li, H. Liu, and H.-G. Cai, "On computing three-finger force-closure grasps of 2-d and 3-d objects," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 1, pp. 155–161, 2003.
- [5] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The grasp taxonomy of human grasp types," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 66–77, 2015.
- [6] O. Taheri, N. Ghorbani, M. J. Black, and D. Tzionas, "GRAB: A dataset of whole-body human grasping of objects," in *ECCV*, 2020.
- [7] K. Karunratanakul, J. Yang, Y. Zhang, M. Black, S. Tang, and K. Muandet, "Grasping field: Learning implicit representations for human grasps," in *3DV*, 2020.
- [8] A. Sahbani, S. El-Khoury, and P. Bidaud, "An overview of 3d object grasp synthesis algorithms," *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 326–336, 2012.
- [9] A. Rodriguez, M. T. Mason, and S. Ferry, "From caging to grasping," *IJRR*, vol. 31, no. 7, pp. 886–900, 2012.
- [10] D. Prattichizzo, M. Malvezzi, M. Gabiccini, and A. Bicchi, "On the manipulability ellipsoids of underactuated robotic hands with compliance," *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 337–346, 2012.
- [11] C. Rosales, R. Suárez, M. Gabiccini, and A. Bicchi, "On the synthesis of feasible and prehensile robotic grasps," in *ICRA*, 2012.
- [12] R. M. Murray, *A mathematical introduction to robotic manipulation*. CRC press, 2017.
- [13] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *IJRR*, vol. 27, no. 2, pp. 157–173, 2008.
- [14] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," in *RSS*, 2017.
- [15] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *IJRR*, vol. 37, no. 4-5, pp. 421–436, 2018.
- [16] Y. Lin and Y. Sun, "Robot grasp planning based on demonstrated grasp strategies," *IJRR*, vol. 34, no. 1, pp. 26–42, 2015.
- [17] S. Brahmabhatt, A. Handa, J. Hays, and D. Fox, "Contactgrasp: Functional multi-finger grasp synthesis from contact," in *ECCV*, 2020.
- [18] P. Grady, C. Tang, C. D. Twigg, M. Vo, S. Brahmabhatt, and C. C. Kemp, "Contactopt: Optimizing contact to improve grasps," in *CVPR*, 2021.
- [19] A. T. Miller and P. K. Allen, "Graspit! a versatile simulator for robotic grasping," *IEEE Robotics & Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.
- [20] E. Corona, A. Pumarola, G. Alenya, F. Moreno-Noguer, and G. Rogez, "Ganhand: Predicting human grasp affordances in multi-object scenes," in *CVPR*, 2020.
- [21] S. P. Boyd and B. Wengert, "Fast computation of optimal contact forces," *T-RO*, vol. 23, no. 6, pp. 1117–1132, 2007.
- [22] L. Han, J. C. Trinkle, and Z. X. Li, "Grasp analysis as linear matrix inequality problems," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 6, pp. 663–674, 2000.
- [23] Y. Zheng and C.-M. Chew, "Distance between a point and a convex cone in n -dimensional space: Computation and applications," *T-RO*, vol. 25, no. 6, pp. 1397–1412, 2009.
- [24] H. Dai, A. Majumdar, and R. Tedrake, "Synthesis and optimization of force closure grasps via sequential semidefinite programming," in *Robotics Research*. Springer, 2018, pp. 285–305.
- [25] Y.-H. Liu, "Qualitative test and force optimization of 3-d frictional form-closure grasps using linear programming," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 1, pp. 163–173, 1999.
- [26] T. Tsuji, K. Harada, and K. Kaneko, "Easy and fast evaluation of grasp stability by using ellipsoidal approximation of friction cone," in *IROS*, 2009.
- [27] J. Romero, D. Tzionas, and M. J. Black, "Embodied hands: Modeling and capturing hands and bodies together," *TOG*, vol. 36, no. 6, pp. 1–17, 2017.
- [28] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *CVPR*, 2019.
- [29] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su *et al.*, "Shapenet: An information-rich 3d model repository," *arXiv preprint arXiv:1512.03012*, 2015.
- [30] M. Przybylski, T. Asfour, and R. Dillmann, "Unions of balls for shape approximation in robot grasping," in *IROS*, 2010.
- [31] S. Ottenhaus, D. Renninghoff, R. Grimm, F. Ferreira, and T. Asfour, "Visuo-haptic grasping of unknown objects based on gaussian process implicit surfaces and deep learning," in *International Conference on Humanoid Robots (Humanoids)*, 2019.
- [32] M. Hill, E. Nijkamp, and S.-C. Zhu, "Building a telescope to look into high-dimensional image spaces," *Quarterly of Applied Mathematics*, vol. 77, no. 2, pp. 269–321, 2019.
- [33] G. Sultanto, A. Wang, Y. Lin, M. Mukadam, G. Sukhatme, A. Rai, and F. Meier, "Encoding physical constraints in differentiable newton-euler algorithm," in *Learning for Dynamics and Control*, 2020.