Robot Model Identification and Learning: A Modern Perspective

Taeyoon Lee,[1] Jaewoon Kwon,[2] Patrick M. Wensing,[3] and Frank C. Park[4]

[1]Naver Labs, Seongnam, South Korea, 13561; email: ty-lee@naverlabs.com
[2]Naver Labs, Seongnam, South Korea, 13561; email: jaewoon.kwon@naverlabs.com
[3]Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, Indiana, USA, 46556; email: pwensing@nd.edu
[4]Mechanical Engineering, Seoul National University, Seoul, South Korea, 08826; email: fcp@snu.ac.kr

Annu. Rev. Control Robot. Auton. Syst. 2024. 7:1–26

https://doi.org/10.1146/((please add article doi))

Keywords

system identification, model learning, geometry, inductive bias

Abstract

In recent years, the increasing complexity and safety-critical nature of robotic tasks have highlighted the importance of accurate and reliable robot models. This trend has led to a growing belief that, given enough data, traditional physics-based robot models can be replaced by appropriately trained deep networks or their variants. Simultaneously, there has been a renewed interest in physics-based simulation, fueled by the widespread use of simulators to train reinforcement learning algorithms in the sim-to-real paradigm. The primary objective of this review is to present a unified perspective on the process of determining robot models from data, commonly known as system identification or model learning in different subfields. The review aims to illuminate the key challenges encountered and highlight recent advancements in system identification for robotics. Specifically, we focus on recent breakthroughs that leverage the geometry of the identification problem and incorporate physics-based knowledge beyond mere first-principles model parameterizations. Through these efforts, we strive to provide a contemporary outlook on this problem, bridging classical findings with the latest progress in the field.

1. INTRODUCTION

In recent years, the field of robotics has witnessed a significant increase in the complexity and importance of safety in robotic tasks. Notable examples include legged robots capable of running and performing highly athletic motions, as well as manipulators and multi-fingered hands executing dexterous, non-prehensile manipulation tasks. The growing complexity of these tasks has highlighted the necessity for accurate and dependable robot models.

Traditionally, physics-based dynamic models have been the go-to approach for approximating the input-output behavior of robots. These models rely on principles derived from classical mechanics to describe the robot's motion and interaction with the environment. Recently, pure data-driven approaches, e.g., from machine learning, have gained momentum as they offer the possibility of learning robot models directly from data without explicitly modeling the underlying physics. These approaches aim to reconstruct input-output behavior solely based on observed data, leveraging the power of statistical learning techniques. This shift towards pure data-driven models has sparked a discussion about their potential to augment or even replace traditional physics-based models in certain contexts.

In fact, the problem of approximating the input-output behavior of a system from data, referred to as the system identification or model learning problem, has a long history in robotics. Earlier approaches to robot model identification aimed to improve the accuracy of the physical parameters of physics-based models by collecting joint kinematics and torque data. After all, the availability, coverage, and quality of the data samples play a central role in accurately approximating reality in both physics-based and physics-free approaches. Unfortunately, acquiring extensive and sufficiently rich data from real robotic systems can be challenging, particularly for complex systems and when safety concerns arise. Even for mechanics-based models, identifying a minimal set of parameters that represents the rigid-body dynamics of a standard humanoid structure (often requiring hundreds of parameters) becomes highly challenging, particularly in terms of generalizing the model predictions across unforeseen initial states and system inputs.

One of the primary objectives of this review is to provide a unified perspective on the system identification problem across the robotics and machine learning literature. It aims to shed light on the challenges encountered in identifying accurate and reliable models for different types of robots, as well as the recent advancements in system identification within the robotics domain. The intended audience comprises individuals with a background in machine learning who seek to comprehend the challenges and recent advances in system identification for robotics, as well as those with a background in robotics who are interested in the recent progress of methods that go beyond mechanics-based parameterizations for modeling physical systems.

Of course, this review does not aim to provide an exhaustive review of all system identification and model learning methods. Rather, it focuses on recent advances that appeal more to incorporating domain-specific (physics-based) knowledge to develop sample-efficient and robust methods that go beyond classical parameterizations, yet without treating system identification as a complete black-box function approximation problem. By doing so, it aims to provide a modern perspective on the problem.

The remainder of the paper is structured as follows. Section 2 gives a background on the problem of model learning for nonlinear systems, emphasizing a general viewpoint on the problem often adopted in the machine learning literature. We use this section to motivate the main challenges facing dynamic modeling and identification, and then, in Section 3 detail the more specific structure of the problem for robot model identification. We concentrate

on the classical problem of identifying inertial parameters of a mechanism, and in Section 4 we describe recent advances that leverage the geometric structure of rigid-body inertial parameters to improve the efficiency and learning of dynamic models. Section 5 then moves beyond this classical setting and discusses more recent paradigms that are physics informed, while offering more model-form flexibility. Section 6 concludes the article.

2. BACKGROUND ON DYNAMIC MODEL IDENTIFICATION AND LEARNING

In this Section, we briefly overview the essential procedures and challenges for general nonlinear system identification problems with the primary aim of bridging approaches to system identification in robotics and model learning in machine learning.

Consider the true state-space nonlinear system dynamics of the form,

$$\begin{aligned} x_{t+1} &= f(x_t, u_t, \omega_t) \\ y_t &= h(x_t, u_t, \epsilon_t) \end{aligned} \qquad 1.$$

where $x_t \in \mathcal{X} \subseteq \mathbb{R}^n$ is the state, $u_t \in \mathcal{U}$ is the control input, $\omega_t \in \mathcal{W}$ is the process noise, and $y_t \in \mathcal{Y}$ is the output measurement with $\epsilon_t \in \mathcal{E}$ being the measurement noise. The system identification problem typically begins by *selecting the set of model candidates*, which is assumed to include within it the true system dynamics. Let us consider parametric model classes $\mathcal{M}_f = \{f_\theta : \mathcal{X} \times \mathcal{U} \times \mathcal{W} \to \mathcal{X} \mid \theta \in \Theta\}$ and $\mathcal{M}_h = \{h_\theta : \mathcal{X} \times \mathcal{U} \times \mathcal{E} \to \mathcal{Y} \mid \theta \in \Theta\}$ where a particular instantiation of the system dynamics with parameter θ is given as follows[1][2],

$$\begin{aligned} x_{t+1} &= f_\theta(x_t, u_t, \omega_t) \\ y_t &= h_\theta(x_t, u_t, \epsilon_t) \end{aligned} \qquad 2.$$

By executing a series of control inputs $u_1, \cdots, u_N$, one *collects a finite number of data samples*, $\mathcal{D} = \{u_1, y_1, \cdots, u_N, y_N\}$, from the true system in the real world. Then, the goal is to find the model parameter vector θ that best matches the input-output behavior of the true system dictated by the collected data and the selected model class $\mathcal{M}$. For this, it is essential to *define an error criterion* to be optimized between the true system and the model based on the finite number of data samples. Such a criterion should provide a statistically meaningful objective and/or be informed by the specific usage of the identified model in practice. It is worth noting that, in reality, this process is often not a one-shot ordeal. During validation of the identified model, deficiencies can be exposed due to multiple factors such as the choice of model class $\mathcal{M}$, quality and sufficiency of data samples, the choice of error criterion, or even the parameter estimation algorithms employed.

[1]The present description of the model is often called a grey-box model in nonlinear control systems literature (1). For a more general, comprehensive treatment of nonlinear system models and the associated identification problem, readers are encouraged to refer to (2).

[2]Also, we note that the collection of state variables x_t or their representation within the dynamics model may not generally be identical to the one for the true system. For example, one may choose to neglect the effect of thermodynamics or aerodynamics for modeling the rigid body dynamics of a robot manipulator. The state representation itself can also be auxiliary and learned from data, which is common in learning-based approaches with camera images as outputs (3)

2.1. Parameter Identifiability (...That Matters, or Does it?)

The concept of structural identifiability asks the question of whether different model instances within the model class "uniquely" represent different input-output behaviors. More formally, a particular parameterized model class is said to be *structurally identifiable*, if the parameters can be uniquely determined from the input-output behavior, i.e., if $y(\theta_1) \equiv y(\theta_2)$ for all inputs, then $\theta_1 = \theta_2$. This property naturally leads to the uniqueness and existence of the solution θ to the identification problem, yet with the conditions that the true system is within the specified model class and an infinite amount of data that covers all cases is available. Given that the amount of data collectible from robots is always finite, a more viable description of identifiability to practitioners may be described by accounting for the statistical precision of the parameter estimates. *Practically identifiable* parameters for a given dataset can be considered as ones with confidence intervals smaller than some threshold.

Meanwhile, it is noted that the concept of parameter identifiability pertains to a parameter-centric view of the model identification process; how accurate do the estimated parameters $\hat{\theta}$ revealed from data match the true values? This aspect is less relevant with data-driven learning using black-box models such as deep neural network models since parameters therein are not interpretable, and do not often provide a unique construction of input-output behavior. On the other hand, parameters within mechanics-based models, such as mass, inertia, and stiffness, have tangible and physical meanings that analytically relate to the *global* description of physical behaviors of a system.

Above all, system identification is all about the statistical approximation of the input-output behavior of a system. Then, how can we systematically integrate our understanding of physical parameters into statistical parameter identification of mechanics-based models for improved robustness? Further, how can we construct an extended class of parameterized models that globally relate to our understanding of physical behaviors of robotic systems? Here, we hold our discussion (which is to be resumed in Sections 4 and 5), and from below, we present generic formulations to the data-driven model identification problem and source of errors in the identification process.

2.2. On the Choice of Error Criterion

Employing statistical motivations to construct the error criterion is common in system identification. To explain, one can define a joint probability density function for the output data samples, $p_\theta(y_{1:N}|u_{1:N})$, induced from the stochastic nonlinear dynamics model in Equations 2 and 3.[3] Then, the Maximum Likelihood Estimation (MLE) formulation for approximating θ can be written as minimizing the negative log-likelihood function $L(\theta)$, i.e.,

$$L(\theta) = -\log p_\theta(y_{1:N}|u_{1:N}) \tag{3.}$$

While the joint probability density function is, in general, computationally intractable, the incorporation of practically viable structure in the model allows computationally tractable algorithms for estimating θ in practice. For instance, let us consider the case of a noise-free full-state measurement model, i.e., $x_{t+1} = f_\theta(x_t, u_t, \omega_t),\ y_t = x_t$. Then, the MLE objective

[3]Here, we omit the dependence of the joint probability density with respect to the initial state x_0, and assume it is given.

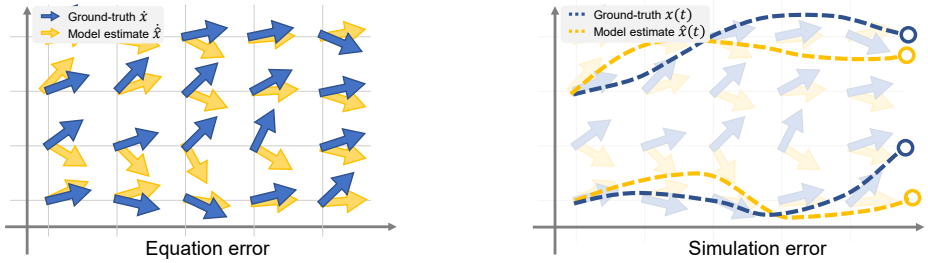

Figure 1

Equation error criteria consider the error in a one-step prediction, while simulation error criteria consider the accumulation of model mismatch over trajectories or multi-step predictions. From a continuous-time view, the equation error criteria correspond to the difference between the vector fields of the dynamic models, while the simulation error is calculated based on the difference between their integral curves.

can be decomposed as follows:

$$L(\theta) = -\sum_{t=1}^{N} \log p_\theta(x_t | x_{t-1}, u_{t-1}), \tag{4.}$$

where the probability density function $p_\theta(x_t|x_{t-1}, u_{t-1})$ can now be solely defined by the dynamic model, Equation 2. A different likelihood function (given that x_{t+1} and ω_t is one-to-one for every fixed x_t, u_t) can be considered as[4],

$$L(\theta) = -\sum_{t=1}^{N} \log p(\omega_t), \quad \text{s.t.} \quad x_{t+1} = f_\theta(x_t, u_t, \omega_t) \quad \text{for} \quad t = 0, \cdots, N-1 \tag{5.}$$

These types of formulations are commonly referred to as employing an *equation error* or *one-step-ahead prediction error* criteria. Loosely speaking, the parameters are optimized to fit the average error in the local state transition dynamics.

On the other hand, another class of error criteria that deserves attention is the so-called *simulation error* criterion, which aims at minimizing the deviation between the measured outputs and the ones simulated or integrated using the model (4, 5, 6). To illustrate, the simulation error criterion can be typically given in the form,

$$L(\theta) = \sum_{i=1}^{N} \|\hat{x}_i - x_s(i, \theta)\|^2, \tag{6.}$$

where $\hat{x}_i$ is the state measurement and $x_s(i, \theta)$ is the i-step ahead prediction of the state by simulating with the model, i.e., $x_s(i, \theta) = f_\theta(x_s(i-1, \theta), u_{i-1})$ for $i = 1, \cdots, N$ with $x_s(0, \theta) = x_0$. It should be noted that the simulation error also admits the maximum likelihood interpretation under full-state measurement with white Gaussian noise and noise-free state dynamics, i.e., $x_{t+1} = f(x_t, u_t)$, $y_t = \hat{x}_t = x_t + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, \sigma^2 I)$. More

[4]The objective in Equation 5 is, in general, different from that in Equation 4, since $\log p_\theta(x_{t+1}|x_t, u_t) = \log p(\omega_t) - \log\det(\partial x_{t+1}/\partial \omega_t)$. However, if $\partial x_{t+1}/\partial \omega_t$ is constant, e.g., $x_{t+1} = f_\theta(x_t, u_t) + \omega_t$, they are identical.

generally, an *output error* criterion can be defined as minimizing the simulated error on the output, i.e., $L(\theta) = \sum_{i=1}^{N} \|y_i - h_\theta(x_s(i,\theta))\|^2$.

Without statistical grounding, the simulation error criterion is inherently different from the equation error criterion in that it reflects the parametric estimation error more globally, rather than locally, by directly integrating the system dynamics model over time. In that regard, the simulation error criterion better exposes the nature of the system identification problem in that it is not simply a "static" estimation problem but involves dynamics that relate to the sequential input-output behavior of the system. Some errors in the parameters that contribute to minor effects in the equation error can lead to severe deviation from the true system dynamics over time. For instance, consider a simple 1-D mass-damper system, i.e., $m\ddot{q} + d\dot{q} = 0$, where m is the mass and d is the damping coefficient. While any possible bounded estimation error on the mass and damping coefficients only leads to bounded equation error, errors that lead to negative values in any of these parameters lead to an unstable forward simulation of the dynamics that can produce exponentially diverging simulation error over time (i.e., $\dot{q} = \dot{q}_0 \exp(-td/m)$). For this reason, minimizing simulation errors are generally more appealing to increase the simulation fidelity or for use in receding-horizon predictive control, or model-based RL. However, compared to equation error approaches, they are usually highly computationally demanding, requiring iterative simulation of the model in the computation loop, and are subject to a nonconvex optimization landscape, which may not be suitable for, e.g., online identification or adaptive control.

2.3. Source of Errors

Suppose the true system dynamics is within the model candidate set, i.e., there exists $\bar{\theta} \in \Theta$ such that $f = f_{\bar{\theta}} \in \mathcal{M}_f$. Then, as the sample size tends to infinity and remains complete, i.e., $L(\theta)$ has a unique global minimum, the maximum likelihood estimate of θ is known to converge to $\bar{\theta}$. In such a case, the estimator is said to be statistically consistent. However, it should be noted that these assumptions rarely hold true in practice; 1) Modeling can almost always be inaccurate to some degree (i.e., introducing structural error), 2) the data can be noisy (i.e., introducing random error) and 3) incomplete. Below we discuss how these problems manifest more specifically.

2.3.1. Structural Error. In scenarios where the problem of model mismatch, i.e., $f \notin \mathcal{M}_f$, is prevalent, the conventional approach of conducting statistical analysis on parameter estimation performance tends to be less viable. Even when abundant and comprehensive data samples are available, uncorrectable biases in the model structure can corrupt parameter estimates in unpredictable ways.

There are many components in real robotic systems that are difficult to model accurately using standard mechanics-based models. For instance, complex frictions involving stick, slip, and state-dependency, nonlinear hysteresis, and slackening effects from tendon-driven mechanisms (7), as well as cascaded actuator dynamics (8), are among the components that standard mechanics-based models struggle to describe with sufficient accuracy.

Also, it should be emphasized that the problem of model mismatch may arise not only from factors related to difficult-to-model effects. If the fixed parameters in a physical model actually contain significant errors, failure to identify them along with other parameters may lead to structural bias issues. For instance, as noted in (9), kinematic parameter error in the dynamics model can cause significant structural bias in estimating dynamic parameters.

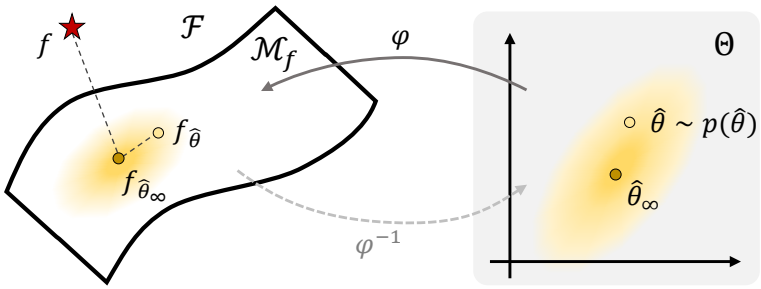

Figure 2

A conceptual picture of the source of errors. Through a mapping $\varphi : \Theta \to \mathcal{M}_f$, the parameter θ is related to the dynamic model, i.e., $f_{\hat{\theta}} = \varphi(\hat{\theta}) \in \mathcal{M}_f$. Its inverse mapping φ^{-1} exists only if the model is structurally identifiable. $\mathcal{F}$ represents the ambient space wherein the model set $\mathcal{M}_f$ is embedded. The ground-truth description of the system $f \in \mathcal{F}$ (denoted by the red star) is not necessarily within the model set. The yellow gradation represents the probability distribution $p(\hat{\theta})$ of the estimated parameters that is induced by the stochasticity of the finite data samples and the specific identification method employed, i.e., $\hat{\theta} = \mathrm{ID}(\mathcal{D})$, $\mathcal{D} \sim p(y_{1:N}|u_{1:N})$. Structural error persists even when the random error is resolved with an infinite number of samples, i.e., $f_{\hat{\theta}_\infty} \in \mathcal{M}_f$.

2.3.2. Data: Randomness and Incompleteness. Random error refers to the estimation error or variance that is incurred by the stochasticity in the system dynamics or noise in the output measurements. Apart from structural error, random error can, in principle, be resolved by having a sufficient amount of and complete measurement data.

Then, given a diverse and complete sampling, one should be able to answer how many data samples are needed to ensure a practically reliable estimation of the model parameters. The Fisher information matrix is given as,

$$\mathcal{I}(\bar{\theta}, u_{1:N}) = \mathbb{E}_{p_{\bar{\theta}}} \left[\nabla_\theta \log p_\theta(y_{1:N}|u_{1:N}) \cdot \nabla_\theta \log p_\theta(y_{1:N}|u_{1:N})^T \mid \bar{\theta} \right], \qquad 7.$$

which measures the amount of information that data samples carry about the unknown true parameter vector $\bar{\theta}$. This definition importantly relates to the Cramer-Rao bound that sets the theoretically achievable estimation covariance of the parameters, i.e., $\mathrm{var}[\hat{\theta}] \succeq \mathcal{I}(\bar{\theta})^{-1}$ when $\hat{\theta}$ denotes an arbitrary unbiased estimator for θ. Accordingly, this measure is commonly used to discern a practically identifiable parameter set from data, or (as noted in the sidebar titled Data Collection Problem) to guide the design of an optimal data collection strategy before identification. Herein, we refer to data being incomplete in the case where the information matrix $\mathcal{I}(\theta)$ (conditioned on the specific data collection experiment $u_{1:N}$) being degenerate along certain subspaces within the structurally identifiable parameter space, i.e., $\delta^T \mathcal{I}(\theta)\delta \equiv 0$ for $\delta \neq 0$, $\delta \in \mathbb{R}^{\dim(\Theta)}$. For example, when data is collected from static postures of a robot, it inherently lacks complete information concerning the robot's inertial properties, irrespective of their structural identifiability.

3. SYSTEMS AND METHODS IN ROBOT MODEL IDENTIFICATION

In this Section, we review methods in robot dynamic identification that are based on traditional mechanics-based models, with a particular focus on rigid-link robots; interested

THE DATA COLLECTION PROBLEM

The fact that the information matrix offers a measure of how the data samples impact the parameter estimation performance makes it possible to optimize the data collection strategy, which is known as optimal experimental design or the optimal excitation problem (10, 5); that is,

$$\max_{u_{1:N}} \quad \sigma\left(\mathcal{I}(\bar{\theta}, u_{1:N})\right). \tag{8.}$$

An intricate issue regarding constructing a physically meaningful scalar measure $\sigma : \mathbb{R}^{\dim(\Theta)\times\dim(\Theta)\to\mathbb{R}}$ out of the matrix-valued information measure $\mathcal{I}(\bar{\theta})$ will be discussed in Section 4.4. The information matrix is in general dependant on the unknown parameter $\bar{\theta}$, which, in practice, typically requires nominal values of θ to proceed with optimization. We note that there are also experimental design criteria that are based on the output prediction error variance (11), as opposed to parameter error variance, essentially by projecting the influence of parametric error to some distribution of output predictions.

Meanwhile, it is worth noting that solving data collection problems for identifying unknown parameters is somewhat akin to a chicken-and-egg problem in general robotics applications. Specifically, while we require trajectory data samples that excite the full spectrum of the dynamics to obtain a good model estimate, designing and executing dynamic motions safely without an accurate dynamics model can be challenging. Consequently, optimal excitation of robot motions has mainly been studied on fixed-base, fully-actuated robots, such as robot manipulators (5), which are less safety-critical than floating-base, underactuated robots like humanoids (12). Considering safety in high-performance control under model uncertainty is itself an important and active area of research, which readers can refer to in the recent review by (13).

readers may also consult the review on soft robot modeling in (14). The main focus of this review herein is to highlight how generic identification formulations give rise to specific challenges for the identification of robotic systems, taking into account their unique characteristics, and different applications. It is important to note that the presentation of recent insights and advancements in robot dynamic identification methods, which follow this section, are not mutually exclusive to the ones reviewed here but rather complement them.

3.1. Systems without Contact or with Contact Force Measurements

3.1.1. Fixed-Base Systems. The seminal work (15) led to a popular linear least squares objective for robot dynamic model identification, which follows the equation error formulation (Eq. 5). Their key contribution recognized that the mass-inertial parameters ϕ (Eq. 14) of robot links and loads appear linearly in the "inverse" dynamics equation of rigid-body systems. To illustrate, for a N_L-link fixed-based robot manipulator with joint configuration variables $q \in \mathbb{R}^n$, joint input torque vector $u \in \mathbb{R}^n$ with additive Gaussian noise ω, i.e., $w \sim \mathcal{N}(0, \Sigma)$, the second-order inverse dynamic equation can be described as,

$$u + \omega = M(q, \psi)\ddot{q} + b(q, \dot{q}, \psi) \equiv Y(q, \dot{q}, \ddot{q})\psi, \tag{9.}$$

where $M(q, \psi) \in \mathbb{R}^{n\times n}$ is the mass matrix, $b(q, \dot{q}, \psi) \in \mathbb{R}^n$ denotes the vector of Coriolis and gravitational forces, and $\psi = [\phi_1, \cdots, \phi_{N_L}]$ denotes the complete set of mass and

STRUCTURALLY IDENTIFIABLE PARAMETERS OF RIGID BODY DYNAMIC MODELS

Multi-body dynamics models of the form $u + w = Y(q, \dot{q}, \ddot{q})\psi$ are not structurally identifiable when ψ contains the ten standard inertial parameters of each body (17). This property is due to the fact that each connecting joint creates ambiguity in how mass/inertia can be assigned to links on either side of the joint without affecting the dynamics (16). Formally, the ambiguity can be characterized via the set:

$$\mathcal{N} = \{\delta \mid Y(q, \dot{q}, \ddot{q})\delta = 0, \forall q, \dot{q}, \ddot{q}\} \quad 11.$$

If $\overline{\psi}$ denotes the true parameters, then any ψ in the affine subspace $\overline{\psi}+\mathcal{N}$ gives the same dynamic model and measurements. To recover structural identifiability, a re-parameterization is often pursued using so-called *base parameters*, denoted as θ. This task can be accomplished via choosing a basis for $\mathcal{N}^\perp$ (the orthogonal complement of $\mathcal{N}$) as follows. Suppose a fixed full-rank matrix B such that $\text{Range}(B^T) = \mathcal{N}^\perp$. Then, for this selection, $\theta_i = \sum_j B_{ij}\psi_j$ gives the i-th base parameter as an identifiable linear combination of the standard parameters ψ_j. While $\mathcal{N}^\perp$ is unique, the choice of B, and thus the choice of base parameters, is not. B can be chosen in row-reduced echelon form (or similar) so that each base parameter represents a regrouping (18) into a standard parameter, which can be desirable for efficient simulation (19).

Many methods exist to re-parameterize a model with base parameters. Base parameter sets can be constructed using symbolic methods via the analysis of the dynamics equations (17, 18, 20), numerically via QR or SVD decompositions applied to regressors of assumed maximally exciting data (21), or geometrically via recursively characterizing how each link is able to be excited (16). Many symbolic methods do not guarantee that the resulting re-parameterization is structurally identifiable except in special cases, while numerical methods will underestimate the number of base parameters if the data used is not fully exciting.

inertial parameters of each of the links (and also typically includes joint Coulomb and viscous friction parameters). A structurally identifiable set of parameters (see the sidebar titled Structurally Identifiable Parameters of Rigid Body Dynamic Models) linearly lumps together parameters in ψ , i.e., via $\theta = B\psi$, and the terms in Equation 9 can be collected and represented in the most compact form as (16),

$$u + \omega = M(q, \theta)\ddot{q} + b(q, \dot{q}, \theta) \equiv \Gamma(q, \dot{q}, \ddot{q})\theta, \quad 10.$$

where $\Gamma(q, \dot{q}, \ddot{q}) \equiv Y(q, \dot{q}, \ddot{q})B$. Then, given the measurements of states $y = x = (q, \dot{q})$ and $\ddot{q}$ (through appropriate numerical differentiation and filtering) at multiple points $t = 1, \cdots, N$ along some reference trajectory, the negative log-likelihood objective can be given as $L(\theta) = \sum_{i=1}^{N} -\log p(\omega_i) \sim \sum_{i=1}^{N} \|\Gamma(q_i, \dot{q}_i, \ddot{q}_i)\theta - u\|^2_{\Sigma^{-1}}$ (c.f., Equation 5).

Such inherent linear-in-parameters property for the robot dynamics equation has prompted the adoption of a rich class of robust identification techniques that rely on linear models. Many of these approaches relax the noise-free assumptions on the regressor matrix (22), the statistical independence assumption between the regressor matrix $\Gamma(q, \dot{q}, \ddot{q})$ and the input u, (5, 23), impose a non-Gaussian distribution on ω to derive robust linear regression methods (24), or employ set membership uncertainty to estimate the parameter error bound amenable within, e.g., robust control (25). To any extent, all these methods concern with fitting the equation error that represents the local error of the dynamics.

There have also been methods that explore the use of simulation/output error criterion (4) or composite ones (6) (e.g., approximating the simulation error solution in a compu-

tationally efficient way using an equation error formulation) in the identification of robot manipulators, and proven to result in more robust identification results compared to the ones that are based purely on equation error. These methods are usually performed using a closed-loop dynamics simulation, i.e., PD-controlled manipulator dynamics with a reference trajectory signal as input, to enhance the stability of the forward simulation computation within identification (26). For a more comprehensive review on parameter identification of robot manipulators, readers are encouraged to refer to the review (27).

3.1.2. Floating-Base Systems. Floating-base systems like humanoid or quadruped robots are inherently underactuated for which the dimension of configuration space $q = (q_j, q_r) \in \mathbb{R}^{n+6}$ is larger than the number of input torque actuation $u \in \mathbb{R}^n$, where $q_j \in \mathbb{R}^n$ denotes the joint configuration and $q_r \in \mathbb{R}^6 \simeq SE(3)$ represents the pose and orientation of a floating base (root) link. The dynamic equation subject to (known) contact forces λ_i for $i = 1, \cdots, n_c$ can be described as,

$$\begin{bmatrix} u \\ 0 \end{bmatrix} + \sum_{i=1}^{n_c} \begin{bmatrix} J^c_{\mathrm{j},i}(q)^T \\ J^c_{\mathrm{r},i}(q)^T \end{bmatrix} \lambda_i \;=\; M(q,\theta)\ddot{q} + b(q,\dot{q},\theta) \;\equiv\; \begin{bmatrix} \Gamma_j(q,\dot{q},\ddot{q}) \\ \Gamma_r(q,\dot{q},\ddot{q}) \end{bmatrix} \theta, \qquad 12.$$

where $J^c_i = [J^c_{j,i}, J^c_{r,i}]$ denotes the Jacobian of the contact points. As shown in the second line of the equation above, a remarkable finding from (28) was that, in the case of open-chain rigid-body systems, the structurally identifiable set of parameters θ of the full dynamic equation can be equally identified solely from the base-link dynamics, i.e., $\sum_{i=1}^{n_c} J^c_{r,i}(q)^T \lambda_i = \Gamma_r(q,\dot{q},\ddot{q})\theta$. Identifying parameters without employing joint-space dynamics is particularly appealing for systems in which joint torque input is unobservable like human subjects (29) or when other structural errors such as from joint friction models can severely bias the parameter identification. However, the reduced nature of identification without joint-torque measurements generally implies a challenge in that there is less information to be employed for accurately identifying the parameters in a way that generalizes to prediction with the full dynamic model. Not surprisingly, a study by (30) demonstrates that the identification using only contact force measurements generalized less accurately for predicting the joint space dynamics than the identification employing contact force and joint torque data.

A fundamental challenge within generic system identification of floating-base systems, as also noted in the sidebar titled The Data Collection Problem, is that collecting sufficiently rich, dynamic data is often unattainable compared to the case for, e.g., fixed-based robot manipulators. This can often lead to highly biased parameter estimates that are not physically consistent and less generalizable for use within high-performance simulation and control (31). Constrained optimization approaches to guarantee physically consistent estimates (32, 33, 34) and regularized formulations to exploit the nominal parameter information, e.g., from CAD data, have been studied (35, 36). These methods were later reimagined and improved following the geometric perspective presented in Section 4.

3.2. Systems with Contact and without Contact Force Measurements

System identification under contact without contact force measurements remains challenging due, in large part, to the non-smooth and hybrid nature of the dynamics,

$$Bu + \sum_{i=1}^{n_c} J^c_i(q)^T \lambda_i = M(q,\theta)\ddot{q} + b(q,\dot{q},\theta). \qquad 13.$$

which is further confounded by redundancy for the contact forces in many contact scenarios. Such a model encompasses the description of legged robots walking and running in contact with the ground, non-prehensile manipulation of objects, etc. The fact that only the motion of the system is observed , e.g., $y = (q, \dot{q}) + \epsilon$, essentially results in added complexity to the problem in which robust determination of the contact states, i.e., contact points, modes and forces, has to be addressed within the parameter identification process.

In (37), the authors explored the use of a time-stepping linear complementarity problem (LCP) formulation to determine the contact states within the equation error formulation. Also, parameter identifiability was revealed under several cases with contact modes known a priori. However, due to the nature of the equation error formulation, in which perfect, noise-free, state measurements are implicitly assumed, the contact timing, mode, and points are forced to be determined directly via the raw state measurements. Incorrect specification of the contact states can bias the resulting parameter estimates, which further has a significant impact on the accuracy of any forward simulations using the identified model.

To address this issue, it would be more appropriate to expose these contact uncertainties incurred by the possible state measurement errors under the simulation error formulation, in which the state estimation problem is implicitly involved (38) (c.f., Equation 6). Modern differentiable physics simulators provide gradients of the state evolution with respect to the model parameters, with which system identification of various (rigid and soft) robotic systems undergoing contact has been studied. These methods are most often based on the simulation error criterion and have primarily relied on shooting methods (39, 40, 41, 42, 43, 44, 45). However, reliable gradient-based optimization of state trajectories undergoing contacts is, in general, an open problem due to the inherent discontinuous nature of the contact dynamics and the associated combinatorial nature of optimization over contacts. This area remains one of active research (45, 46, 47, 48).

4. GEOMETRIC TAKE ON ROBOT MODEL PARAMETERS

4.1. Revisiting Parameter-Centric View on Identification Error

Continuing the discussion initiated in Section 2.1, in the context of the general model parameter identification problem, it is difficult to foresee the impact of parameter errors on model performance beyond the available limited and noisy data samples.

That being said, the physical nature of the parameters to be identified in physics-based robot models offers a higher level of interpretability and predictability in the identification process. To elaborate, mass-inertial parameters provide a quadratic representation of kinetic energy in (interconnected) rigid-body systems. Stiffness parameters define components that store elastic potential energy, while friction parameters describe the model of dissipative forces. The inherent connection of these parameters to underlying physical phenomena can guide the identification process, preventing identified models from overfitting to aberrant aspects in the data. For instance, mass-inertial parameters resulting in a negative definite mass matrix or negative values for friction or stiffness parameters would no longer accurately capture inherent physical characteristics. Similarly, in the case of a humanoid robot, center of mass parameters that deviate significantly from its physical body would not provide a sensible estimate of the risk of falling. Given these considerations, a natural question arises: ***Is there a systematic approach to measure the amount of information the data samples carry about the parameters, yet in a way that accounts for their inherent physical meanings?***

In this Section, we show how a geometric characterization of the space of physical parameters within robot dynamic models has facilitated the recent development of robust and physically consistent identification methods. These methods have shown notable improvements in generalization performance, particularly when dealing with limited and noisy data. We begin by characterizing the space of physical parameters constituting robot dynamic models (Section 4.2). This characterization provides a bridge to show how the mass-inertial parameters of a robot reside in a curved Riemannian space (Section 4.3), which allows perturbations in the mass-inertial properties to be measured in a coordinate-invariant manner. Then, in Section 4.4, we answer the question raised above by demonstrating a systematic way to construct a set of geometric information measures, which can be used to derive a physically meaningful, invariant confidence interval for assessing practical identifiability, and also to proceed to generate optimal excitation trajectories for identification. Finally, Section 4.5 discusses geometric regularization methods within the context of both offline identification and online adaptive control.

4.2. Characterization of Physical Parameter Space

A direct approach to guaranteeing that the model produces physically consistent long-term predictions is to restrict the parameters to those realizable in the physical world. That is, we seek to define the space Θ so that every element is realizable in the physical world. For instance, mass or joint stiffness and friction coefficients should always be positive, while a spatial stiffness or damping matrix should be positive definite.

Things get a bit more involved in defining the correct necessary and sufficient condition for mass-inertial parameters of a rigid body to be physically realizable. To illustrate, the mass-inertial parameters ϕ of a single rigid body constitute ten parameters represented in vectorized form as

$$\phi = [m, h^T, I^{xx}, I^{yy}, I^{zz}, I^{xy}, I^{yz}, I^{zx}]^T \in \mathbb{R}^{10}, \tag{14.}$$

where $m \in \mathbb{R}$ is the mass, $h = m \cdot p \in \mathbb{R}^3$ is the first mass moment with p being the position of center of mass, $I \in \mathbb{R}^{3\times 3}$ denotes the 3×3 symmetric tensor representation of rotational inertia. A necessary condition for physical consistency was first employed in the context of robot mass-inertial parameter identification by (49, 50, 51); that is, the mass should be positive, $m > 0$, and the rotational inertia at the center of mass should be positive definite, $I - [h][h]^T/m \succ 0$, where $[\cdot]$ is a skew-symmetric representation of a 3-D vector (52). This condition was later formalized by (32) as a linear matrix inequality (LMI) constraint, which allowed the nonlinear physical consistency condition to be incorporated within a convex semidefinite programming formulation for identification. Then, (33) pointed out the complete necessary and sufficient condition for physical consistency condition on inertial parameters, which can also be found in the rigid-body dynamics literature (53). The remaining condition for sufficiency was the triangle inequality condition on the eigenvalues of the rotational inertia tensor. More recently, (34) formalized the full physical consistency condition in an LMI as

$$P(\phi) \triangleq \begin{bmatrix} \Sigma & h \\ h^T & m \end{bmatrix} \succ 0\,. \tag{15.}$$

where there exists a 1-1 linear correspondence between the above *pseudo-inertia matrix* $P(\phi)$ and ϕ via $\Sigma = \frac{1}{2}\mathrm{tr}(I)\mathbb{I}_3 - I$ and $I = \mathrm{tr}(\Sigma)\mathbb{I}_3 - \Sigma$. Today's modern robot simulators,

such as MuJoCo (54), support built-in validation for inertial parameters to satisfy this condition for improved simulation fidelity and better practice of model-based engineering.

4.3. Geometry of Physical Parameter Space

Importantly, the condition in Equation 15 precisely represents the fact that mass-inertial parameters should be strictly realizable from a nonnegative mass density function of a rigid body (34). Based on this enlightenment, (55) proposed a coordinate-invariant distance metric on the mass-inertial parameters that measures perturbations to the underlying mass distribution in a physically meaningful way. Specifically, the affine-invariant Riemannian metric (56) between two mass-inertial parameter vectors ϕ_1 and ϕ_2 can be given as,

$$d(\phi_1, \phi_2)^2 = \frac{1}{2}\text{tr}\left(\text{Log}\left((P(\phi_1)^{-1}P(\phi_2)\right)^2\right), \tag{16.}$$

where Log denotes the matrix Logarithm.

As noted in (57), while there can be many other possible choices of distance metric that encode different useful physical meanings, coordinate invariance is firmly required since an arbitrary choice of body-fixed coordinates or units can represent the mass-inertial parameters. To explain, given a generic pseudo-inertia $P \in \mathcal{P}(4)$ (i.e., the set of 4×4 positive definite matrices), any change of coordinate frame, or change in physical units/scale, transforms the pseudo-inertia P to GPG^T for some $G \in GL(4)$, where $GL(n)$ denotes the set of $n \times n$ invertible matrices (see Equations 10-13 in (55)). This operation represents a $GL(4)$ group action * on $\mathcal{P}(4)$ defined by $G * P \triangleq GPG^T$. A key property of the affine-invariant metric is that it is invariant under this group action (56)[5]. Since coordinate/scale changes represent a subset of all such transformations GPG^T, this property guarantees coordinate and scale/unit invariance of the proposed metric in Eq. 16. While numeric values transform according to the transformation of coordinates, the intrinsic measure of how two mass-inertial parameters differ should not depend on this choice.

As will be discussed in the following Sections, some non-trivial convex approximations to the Riemannian distance metric have proven to be useful for developing computationally tractable and efficient algorithms. The Bregman divergence (57) associated with the negative log determinant $F(\phi) = -\log(|P(\phi)|)$ provides a second-order approximation to the squared Riemannian distance given as,

$$d_F(\phi_1 \| \phi_2)^2 = \log\frac{|P(\phi_2)|}{|P(\phi_1)|} + \text{tr}\left(P(\phi_2)^{-1}P(\phi_1)\right) - 4\,, \tag{17.}$$

which is convex in its first argument. Also, a quadratic approximation can be given by considering a constant (differential) Riemannian metric evaluated at some given nominal value ϕ_0 as,

$$d_0(\phi_1, \phi_2)^2 = \frac{1}{2}\text{tr}\left(\left[(P(\phi_0)^{-1}P(\phi_1 - \phi_2)\right]^2\right) \equiv (\phi_1 - \phi_2)^T g(\phi_0)(\phi_1 - \phi_2), \tag{18.}$$

where $g(\cdot)$ is the pullback of the affine-invariant Riemannian metric on $\mathcal{P}(4)$ to $\mathbb{R}^{10}$ under the mapping $P(\cdot)$. This squared metric is also a case of a Bregman divergence, but associated

[5]That is, given matrices $P_1, P_2 \in \mathcal{P}(n)$ and any $G \in GL(n)$, the geodesic distance $d_{\mathcal{P}(n)} : \mathcal{P}(n) \times \mathcal{P}(n) \to \mathbb{R}$ corresponding to the affine-invariant Riemannian metric satisfies $d_{\mathcal{P}(n)}(P_1, P_2) = d_{\mathcal{P}(n)}(G * P_1, G * P_2)$.

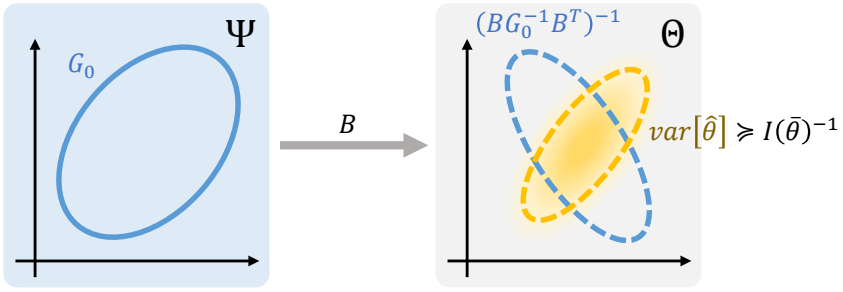

Figure 3

Illustration of how the parameter estimation variance or information (indicated as yellow dashed ellipsoid) can be quantified relative to the metric $H_0 = (BG_0^{-1}B^T)^{-1}$ (indicated as blue dashed ellipsoid) inherited from the metric G_0 originally defined on the standard parameter space Ψ.

with the function $F(\phi) = \phi^T g(\phi_0)\phi$. Importantly, both of these approximate distance measures admit the same coordinate invariance property as the Riemannian distance.

As previously mentioned, other physical parameters such as stiffness, friction, and damping can be readily identified as positive tensors. It is possible to demonstrate that the same affine-invariant Riemannian metric and its approximations provide well-defined coordinate-invariant distance measures for these parameters. Further, as noted in section 5.3 in (58) the affine-invariant Riemannian manifold structure can be generalized to be imposed on an arbitrary convex set of mass-inertial parameters. This generalization can be accomplished by inducing a Riemannian metric as the Hessian of a strictly convex, twice differentiable barrier function on the set, which then gives rise to a Hessian manifold structure. This approach can be used, for example, to impose variable bounds or other linear constraints on the parameters beyond those related to physical consistency (58).

4.4. Geometric Information Measure

Referring to Equation 8, selecting a meaningful scalar measure $\sigma(\cdot)$ as a function of the information matrix amounts to selecting a meaningful distance metric to quantify the variability of the estimation error in the parameter space. To explain, the so-called A-optimality criterion (10) is constructed by choosing the inverse-trace operator for σ, i.e., $\mathrm{tr}(\mathcal{I}(\bar{\theta})^{-1})$. Noting that the inverse of the information matrix serves as the covariance matrix $\mathrm{var}[\hat{\theta}]$ for the efficient unbiased estimator $\hat{\theta}$, one can rewrite it as,

$$\mathrm{tr}(\mathrm{var}[\hat{\theta}]) = \mathbb{E}_{\hat{\theta}}\left[\|\bar{\theta} - \hat{\theta}\|^2\right]. \tag{19.}$$

Clearly, the standard A-optimality criterion exhibits the standard Euclidean metric as a distance metric on parameters. This choice is a sensible one if the representation of the parameters θ exhibits a canonical coordinate choice under which each entries of the parameters are in a similar scale. However, as mentioned earlier, physical parameters like mass-inertial parameters, or more accurately the base parameters, exhibit arbitrary choice of coordinates and linear reparametrizations by the user. The standard Euclidean metric is not only coordinate-dependant but also does not capture the multi-scale nature of physical parameters that come with different units and scales.

Instead, an A-optimality criterion that is akin to the form, $\mathbb{E}_{\hat{\theta}}\left[d(\bar{\theta}, \hat{\theta})\right]$, where $d(\cdot,\cdot)$ be-

ing some coordinate-invariant physically meaningful distance metric, is a more practically appealing choice. In (59), various choices of existing coordinate-dependant optimality criteria, including the alphabet-optimality criteria (10) and the condition number, have been reformulated via definition as (c.f., Equation 8)

$$\sigma \left(\mathcal{I}(\bar{\theta}, u_{1:N}) \cdot H_0^{-1} \right), \tag{20.}$$

where $\sigma(X)$ denotes a symmetric function of the eigenvalues of matrix X, and the constant normalization matrix $H_0 = (BG_0^{-1}B^T)^{-1}$ can be understood as the projection of the constant pullback metric G_0 (analogous to $g(\phi_0)$ in Equation 18) defined over the standard parameters $\psi = [\phi_1, \cdots, \phi_n] \in \Psi$ to the reduced structurally identifiable parameter space under the particular base parameter representation B, i.e., $\theta = B\psi \in \Theta$. As shown in Figure 3, this can also be viewed as the coordinate-invariant measure of how much the distribution of the parameter estimates is distorted in relative to the natural choice of metric predefined over the parameter space Θ. In effect, this geometric framework allows for a formal, systematic way in which to normalize the information matrix so that the ensuing optimal excitation trajectory generation problem leads to the consideration of lighter links needing to be more "excited" than the heavier ones, as mass-inertial parameter values are more likely to be negative definite under the same scale of estimation variance.

As noted in (59), the inverse of the eigenvalues[6] of the normalized information matrix can also be used to assess the practical identifiability of the parameters based on a single scale-free threshold value. Meanwhile, numerical studies provided in (59) have also shown that the fraction of practically identifiable parameters, among the structurally identifiable ones, is considerably restricted in practice, especially for high-dimensional systems such as humanoid robots. This finding highlights the need for a regularized formulation of the parameter identification problem to effectively mitigate such a practical identifiability issue within purely "data-driven" parameter estimation by incorporating prior information.

4.5. Geometric Regularization Techniques

One unique aspect of physics-based models, in comparison to generic function approximation models, is that obtaining reasonably accurate and physically plausible nominal parameters is viable prior to collecting data. This can be achieved through various means such as utilizing CAD data or making a rough guess. In light of the practical challenges due to factors such as data sufficiency, noise, and structural errors, it is crucial for practitioners to recognize that full precise identification of robot models is often impractical and that the nominal model of a robot, in addition to input-output data samples, is a valuable and viable source of information that can be leveraged in identification. More concretely, many of these difficulties can be mitigated by appealing to appropriate regularization of the parameter identification objective using the nominal parameters.

To explain, assuming that some prior distribution on the standard parameters $\psi = [\phi_1, \cdots, \phi_{N_L}]$ is given as $p(\psi)$, the Maximum a Posteriori (MAP) formulation (60) aims to optimize the posterior distribution over ψ, as, $\min_\psi -\log p(\psi|\mathcal{D}) \propto L(B\psi) - \log p(\psi)$, where the previous MLE objective $L(\theta) = L(B\psi)$ (Equation 3) defined over the base parameters is now regularized with the negative logarithm of the prior model over ψ. Consequently, the

[6]They provide scale-free measure of parameter estimation variance projected along the corresponding eigenvectors under the metric $(BG_0^{-1}B^T)^{-1}$.

GEOMETRIC REGULARIZATION IN ADAPTIVE CONTROL

Bregman divergence measures (e.g., Eqs. 17 or 18) for geometric regularization in identification have also played a key role in the design of recent geometric adaptive control laws (62, 61). In that context, one can show stability of the adaptive laws using a Lyapunov function that includes terms $D(\overline{\psi} \,\|\, \hat{\psi}) := \sum_i d_F(\overline{\phi}_i \,\|\, \hat{\phi}_i)^2$, where $\hat{\psi}$ represents the parameter estimate. Remarkably, the asymptotic solution of the adaptation laws developed therein *implicitly regularize* the solution, in that persistently exciting (PE) trajectory references lead to (61): $\lim_{t\to\infty} \hat{\psi}(t) = \operatorname{argmin}_{\psi \in \overline{\psi}+\mathcal{N}} D(\psi \,\|\, \hat{\psi}(0))$. More generally, in the absence of PE references, the parameters converge to a minimizer within a (higher-dimensional) affine subspace that contains $\overline{\psi} + \mathcal{N}$ (see (61) for detail).

regularization term is chosen to be proportional to some distance measure to the nominal parameters ψ_0, i.e., the MAP optimization takes the form

$$\min_{\psi} L(B\psi) + \gamma d(\psi, \psi_0)^2. \tag{21.}$$

The base parameter estimate can be simply recovered by $\hat{\theta} = B\hat{\psi}$, where $\hat{\psi}$ is given as the solution of the regularized identification objective above.

Here, the choice of distance metric d greatly impacts the generalization performance, which essentially dictates a scalar measure of how the multivariate parameter estimates from data samples would deviate from the nominal parameters. The use of geometric distance metric in Equation 16 or its convex approximations in Equation 17 or 18 was shown to be significantly more generalizable than, e.g., the standard Euclidean metric, as they capture the perturbations in the mass-inertial parameters in a coordinate-invariant and physically meaningful way (see Figure 2 of (57)). Recent theoretical breakthroughs regarding implicit regularization (61) have likewise translated these benefits to adaptive control settings (see Sidebar on Geometric Regularization in Adaptive Control). Overall, geometric regularization methods have shown practical significance for robustness and generalization performance in various systems and scenarios, which included online parameter identification of robot manipulators with proprioceptive actuation, generalization of fixed-base dynamics identification results for quadruped robot legs to the full floating-base dynamics model, and reduced base-link dynamics identification for a humanoid structured human model under significant levels of measurement noise.

5. REDUCING STRUCTURAL BIAS IN ROBOT MODELS

While previous section focused on system identification methods that exploit physical meanings of parameters constituting a classical parametrization of robot dynamics models, this section explores various extended classes of parameterized models that aim to alleviate the structural errors (introduced in Section 2.3.1) in a physically meaningful way.

5.1. Kinodynamic Model Identification

Existing methods for robot identification generally separate kinematic identification from dynamic identification; the kinematic parameters are either identified first or assumed to be

provided by manufacturers (e.g., via computer-aided design), followed by the identification of mass-inertial parameters. This separation is largely for reasons of convenience; given that the dynamic model is nonlinear with respect to the kinematic parameters, the linearity-in-parameter property fundamental to many existing robot model identification methods only holds if the kinematic parameters are fixed.

These sequential approaches, however, could potentially introduce bias due to errors in kinematic parameters. Specifically, as reported in (9), poorly identified kinematic parameters can lead to an uncorrectable bias in the dynamic model, leading to errors in the mass-inertial parameters that are many times that of kinematic parameter errors. To address this, the authors proposed a unified identification framework, which jointly identifies the kinematic and dynamic model parameters by minimizing the sum of both errors as

$$\min_{\theta_{\text{dyn}},\theta_{\text{kin}}} \quad L_{\text{dyn}}(\theta_{\text{dyn}},\theta_{\text{kin}}) + \alpha \cdot L_{\text{kin}}(\theta_{\text{kin}}), \tag{22.}$$

where θ_{dyn}, θ_{kin}, L_{dyn} and L_{kin} denote the dynamic and kinematic parameters(e.g., joint screws) and identification objective functions, respectively. The kinematic identification objective, for example, is an error between the end-effector poses measured by cameras and the kinematic model estimates.

Formally, this approach can be derived by expanding the observation y in Equation 2 to include the kinematic measurement (e.g., $y = (x, T) = (x, \text{FK}_{\theta_{\text{kin}}}(x) + \epsilon)$ where T is the end-effector pose measurement and $\text{FK}_{\theta_{\text{kin}}}$ denotes the forward kinematics mapping) and applying the maximum likelihood estimation argument (i.e., Equation 3). By allowing for the update of kinematic parameters to reduce the dynamic error as well, the structural error induced by the kinematic parameter error can be mitigated. Importantly, under the maximum likelihood estimation argument, the weight α can be determined to represent the relative accuracy of kinematic and dynamic sensors (e.g., camera and joint torque sensor) as $\alpha = \frac{\sigma_{\text{dyn}}}{\sigma_{\text{kin}}}$ where σ_{dyn} and σ_{kin} denote the sensor noise scales. This relationship implies that when the kinematic sensor is highly inaccurate, α becomes zero (i.e., $\sigma_{\text{kin}} \gg 1$), rendering the kinematic error less critical in identifying the kinematic parameter. Conversely, if the kinematic sensor is noise-free (for example, with a high-performance motion capture system), α approaches infinity, making the process similar to the traditional decoupled identification. This unified method is especially advantageous when precise kinematic parameters or sensors are unavailable, and substantial model bias is otherwise introduced.

Alternatively, by substituting the error criterion with the simulation error (as in Equation 6), kinodynamic model identification for contact manipulation has been explored (38). In this study, the dynamic parameters include contact model parameters (such as stiffness and damping), with both state trajectories and model parameters estimated concurrently.

Note that while the kinodynamic model identification has extended the candidate model set to deal with the structural error, it still consists of only physically-interpretable parameters and strictly adheres to first principles (i.e., the dynamics of articulated rigid-body systems). In the subsequent sections, we will explore more data-driven methods designed to compensate for residual errors that the existing laws of physics cannot fully address.

5.2. Discrepancy Modeling Approaches

Data-driven modeling approaches that partly augment existing physics-based models have proven more generalizable and data-efficient than relying on neural networks or other generic function approximators to model the complete robot kinematics and dynamics.

Some of these methods directly append deterministic data-driven models to the forward or inverse dynamics equations. These methods are designed to model residual forces acting on the bodies or joints that are difficult to model analytically, such as nonlinear state-dependent frictions or aerodynamic drag forces, or to capture non-Markovian phenomena such as hysteresis, backlash, or unmodeled elasticity with memory-based models. Examples of such methods include (7, 63, 64, 65). In addition, (8) proposed a technique to learn a nonlinear recurrent neural network mapping from low-level actuator commands to more conventional inputs (e.g., joint torques) accepted by existing physics-based simulators. This approach has been shown effective for modeling complex input-output dynamics of actuators subject to nonlinear elasticity, such as series-elastic actuators. Stochastic models based on deep generative models (66) or nonparametric models such as Gaussian Processes (67) have also been used to model the uncertainty in the residual error.

Although discrepancy modeling approaches significantly reduce the extent to which data-driven models need to learn from data compared to pure data-driven models, they do not necessarily ensure the preservation of many system characteristics inherent to the original physics-based model. Therefore, it is essential to carefully validate the model on a wide distribution of test samples to ensure its reliability and accuracy. In addition, the deliberate incorporation of inductive biases in the augmented data-driven model can also be a practical and viable solution. In the subsequent section, we highlight some notable approaches in this direction for modeling contact dynamics, which build upon the classic contact mechanics in a way that explicitly ensures the non-penetration of colliding bodies.

5.3. Putting Physics in Data-Driven Model Learning

Up to this point, we have identified physics-based models as those that characterize the input-output behavior of physical systems using fundamental principles or laws of physics. These models explicitly represent physical parameters, such as mass-inertia, joint screws, and stiffness, which are all tangible and measurable. On the other hand, data-driven modeling approaches (67) typically refer to bottom-up methods of deducing new laws from experimental observations, with the caveat that the parameters within these models do not necessarily possess an explicit physical meaning. However, this observation doesn't mean that data-driven models must be arbitrarily complex in order to learn everything from data. Recent works have shown that a new class of models can be constructed to deliberately exhibit some essential characteristics, via so-called inductive biases, of physical systems.

Below we classify a range of beneficial inductive biases pertinent to the modeling of robotic systems. While the majority of these attributes are currently satisfied and incorporated within established physics-based models, the overarching technical objective is to reconstruct these inductive biases with enhanced flexibility, thereby facilitating the modeling of an expanded repertoire of complex physical systems.

5.3.1. Energy Conservation. Embracing the concept of energy has played a pivotal role in the implementation of model-based engineering approaches within traditional mechanics-based robot models, such as those employing Lagrangian mechanics to characterize coupled rigid bodies. Key elements of these methodologies include developing stable integrators (68, 69, 70) and energy-based controllers (71, 72). Recent advancements in data-driven modeling techniques have shown ways to construct a new class of deep neural networks

that explicitly inherit essential properties, such as energy conservation and passivity, of Hamiltonian and Lagrangian mechanics by constructing a valid kinetic energy function of the system (73), i.e., $E = \frac{1}{2}\dot{q}^T M_\theta(q)\dot{q} > 0$. These models straightforwardly demonstrate that the aforementioned energy-based methodologies can be applied in a manner comparable to their traditional counterparts (74). More recently, these models have been extended to be augmented with classical contact dynamics to model legged robots and robotic manipulators involving contacts and collisions, which introduce discontinuities in the states (75).

5.3.2. Energy Dissipation. Robotic systems consist of several components that contribute to the dissipation of energy in complex ways. One of the primary factors that leads to energy dissipation is friction, which can be highly nonlinear and state-dependent. Despite its complexity, the presence of friction is often found to be useful for control applications, mainly due to its dissipative nature. Specifically, the energy $\mathcal{E}$ of an uncontrolled system with generalized coordinates q should dissipate over time, i.e., $\dot{E} = \dot{q}^T f(q, \dot{q}) \leq 0$, where $f(q, \dot{q})$ represents the generalized forces of friction. A data-driven configuration-dependent friction model has proven effective for modeling complex tendon-driven robots, in which the coupling of joints forming closed kinematic chains can cause non-linear variations of friction coefficients (7). For dynamic friction effects subject to, for instance, hysteresis or stick-slip motions, a more general condition can be derived from the strict passivity condition, i.e., $\dot{q}^T f(z, q, \dot{q}) \geq \dot{W}(z(t))$ for all t, where W is a positive storage function of the internal state z. While energy-dissipative models for dynamic friction have primarily been studied in one-dimensional systems (76), it is potentially viable to explore a more general class of multi-dimensional dynamic friction models that satisfy strict passivity.

5.3.3. Contact. Several studies have shown that standard rigid contact solvers used in state-of-the-art rigid-body dynamics simulators fail to precisely capture the contact behavior in the real-world (48). While there have been many data-driven modeling approaches to directly learn the contact dynamics from data, these approaches often fail to guarantee the most salient properties of contact interactions which are nonpenetration of bodies and energy dissipation subject to contact friction.

One of the major issues in simulating contact behaviors lies in the robust and accurate determination of contact points/normals and modes. Also, some of the ad-hoc heuristics adopted in off-the-shelf rigid body simulators are partly driven by real-time computational constraints rather than fully prioritizing simulation accuracy (77). Recently, data-driven models have augmented the classical contact solvers in a non-trivial way to preserve the non-penetrating and dissipative behavior of contact while being flexible enough to match reality. These methods include data-driven learning of robust contact mode switch detection(78) as well as contact clustering (77). Also, data-driven learning for the smooth representation of shapes has been studied within the identification of contact dynamics to alleviate shape uncertainty and contact-induced discontinuities. These data-driven shape representations include inter-body signed distance functions (79) and neural density fields (80), which can be straightforwardly adopted in standard differentiable physics engines.

5.3.4. Topology and Graph. It is evident that robots with articulated bodies possess an inherent graphical structure due to their physical connectivity. Additionally, the contact interactions between robots and their surrounding environments also exhibit a graphical structure dictated by the intermittent kinematic structure of contact (81). Recently, Graph

Neural Network (GNN) models have gained popularity as they explicitly aim to capture and leverage this relational inductive bias. These models impose this structure by appropriately assigning state variables to each node in the graph and constraining computations to propagate through pre-specified (or potentially learnable) edge connections. However, these methods are currently limited to simulation experiments (82), and scenarios involving simple particle/object interactions (83, 84). We believe one of the significant advantages that comes with the way graphical structure is embedded in classical mechanics-based models lies in their compositionality. For example, if one has a completely separate pair of models for a robot and an object, it becomes possible to evaluate their interactions in a zero-shot manner. In that regard, whether these graphically structured data-driven models can effectively demonstrate generalization performance within arbitrary compositions of graphs remains an open problem.

5.3.5. Invariance and Symmetry. It is important to note that while the choice of coordinate frames to describe the physical output variables, such as end-effector poses, or parameters, such as mass-inertia and shapes, is entirely up to the user, the results are completely invariant, or more precisely equivariant[7], to these choices. This property essentially makes the universal adoption of standard robot description formats, such as URDF, MJCF, and SDF, with arbitrary coordinate choices viable. Indeed, the result of any physical phenomenon must not be affected by the particular choices of coordinate one adopts to describe them numerically. While this may sound somewhat obvious, such a property is not easily attainable in many data-driven models that do not exhibit invariance or equivariance properties subject to certain group transformations on the input-output variables and/or parameters. Recent group-equivariant networks are beginning to be adapted to various dynamics learning problems in robotics, and have shown superior generalization performance and sample efficiency with respect to the variations in shapes and poses. The current applications range from grasp quality prediction (85), table-top object pushing and the learning of interaction dynamics (86, 87), and dynamic modeling of multi-legged robots (88).

6. CONCLUSION

Constructing accurate and reliable descriptions of robotic systems interacting in the real world is naturally posed as a data-driven learning problem from input-output data samples. Starting from the generic black-box system identification view of the dynamic model identification and learning problem, this review has pointed out various practical concerns across stages of the system identification process. As applied to robotics, we have primarily focused on addressing the availability of a sufficient amount of data and the generalizability of the modeling and identification methods.

It remains firmly established that robotics continues to deal with problems related to interactions that occur in the physical world. Undoubtedly, physics represents the most powerful domain knowledge and useful inductive bias inherent in this setting. In light of this view, we have shown how explicit considerations of physics-based knowledge in statistical

[7]A mapping $f : \mathcal{X} \to \mathcal{Y}$ is *equivariant* under the transformation group $\mathcal{G}$, if it satisfies the relation, $\mathcal{S}_g \circ y = f(\mathcal{T}_g \circ x)$, for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $g \in \mathcal{G}$ where $\mathcal{T}_g : \mathcal{X} \to \mathcal{X}$ and $\mathcal{S}_g : \mathcal{Y} \to \mathcal{Y}$ are the group actions of g on $\mathcal{X}$ and $\mathcal{Y}$ respectively. Normally, these group actions are given as the coordinate transformation rules on the respective spaces. If $\mathcal{S}_g$ is an identity mapping, f is an invariant mapping.

data-driven approaches to system identification have unpinned many recent advances for enhanced robustness and generalizability in the face of limited and noisy data. Below are the main summary points of this review.

SUMMARY POINTS

1. A recent geometric perspective on the mass-inertial parameters of rigid robot dynamics model has facilitated the development of robust and physically consistent identification methods that led to notable improvements in model generalization, particularly when dealing with limited and noisy data.
2. We identify and correct some longstanding issues with the established practice of first performing kinematic identification, followed by mass-inertial parameter identification. Specifically, poorly identified kinematic parameters can lead to an uncorrectable bias in the dynamic model, leading to errors in the dynamic parameters that are many times that of kinematic parameter errors. A unified kinodynamic identification method was described that leads to more accurate identification of both the kinematic and dynamic parameters.
3. We described ways in which robot models can be augmented with data-driven models or entirely reconstructed in such a way to respect some of the important physics-based inductive biases with enhanced flexibility; thereby facilitating the modeling of an expanded repertoire of complex physical systems.

In conclusion, it is important to recognize that system identification should not be considered an ultimate goal in and of itself, but rather a valuable tool for a wide range of targeted robotics applications. In line with this understanding, the controls and reinforcement learning community is increasingly committed to developing system identification methods that are directly aligned with control and task objectives (26, 89, 90). While this review article did not extensively explore these ideas, they represent exciting prospects for the field's future. Through these and other advancements, we hold hope that ongoing developments will facilitate the promotion of robust, sample-efficient, and generalizable robot models that are well-equipped to support the complex tasks robots are expected to perform. As these systems interact with and navigate the physical world, we have a growing opportunity to leverage the vast structural richness it provides.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Frank C. Park is supported in part by NRF-MSIT grant RS-2023-00208052, IITP-MSIT grant 2021-0-02068 (SNU AI Innovation Hub), KIAT-MOTIE grant P0020536 (HRD Program for Industrial Innovation), IITP-MSIT grant 2022-0-00480 (Training and Inference Methods for Goal-Oriented AI Agents), SNU-AIIS, SNU-IAMD, SNU BK21+ Program in Mechanical Engineering, and SNU Institute for Engineering Research. Patrick M. Wensing

was supported in part by NSF award CMMI-2220924 with a sub-award to the University of Notre Dame.

LITERATURE CITED

1. Kristensen NR, Madsen H, Jørgensen SB. 2004. Parameter estimation in stochastic grey-box models. *Automatica* 40(2):225–237
2. Schoukens J, Ljung L. 2019. Nonlinear system identification: A user-oriented road map. *IEEE Control Systems Magazine* 39(6):28–99
3. Wu P, Escontrela A, Hafner D, Abbeel P, Goldberg K. 2023. *DayDreamer: World Models for Physical Robot Learning*. In *Proceedings of The 6th Conference on Robot Learning*, ed. K Liu, D Kulic, J Ichnowski, pp. 2226–2240, vol. 205 of *Proceedings of Machine Learning Research*, pp. 2226–2240. N.p.: PMLR
4. Brunot M, Janot A, Carrillo F, Cheong J, Noël JP. 2020. Output error methods for robot identification. *Journal of Dynamic Systems, Measurement, and Control* 142(3):031002
5. Swevers J, Ganseman C, Tukel DB, De Schutter J, Van Brussel H. 1997. Optimal robot excitation and identification. *IEEE transactions on robotics and automation* 13(5):730–740
6. Gautier M, Janot A, Vandanjon PO. 2012. A new closed-loop output error method for parameter identification of robot dynamics. *IEEE Transactions on Control Systems Technology* 21(2):428–444
7. Choi K, Kwon J, Lee T, Park C, Pyo J, et al. 2020. A hybrid dynamic model for the ambidex tendon-driven manipulator. *Mechatronics* 69:102398
8. Hwangbo J, Lee J, Dosovitskiy A, Bellicoso D, Tsounis V, et al. 2019. Learning agile and dynamic motor skills for legged robots. *Science Robotics* 4(26):eaau5872
9. Kwon J, Choi K, Park FC. 2021. Kinodynamic model identification: A unified geometric approach. *IEEE Transactions on Robotics* 37(4):1100–1114
10. Pukelsheim F. 2006. *Optimal design of experiments*. SIAM
11. Pronzato L. 2008. Optimal experimental design and some related control problems. *Automatica* 44(2):303–325
12. Bonnet V, Fraisse P, Crosnier A, Gautier M, González A, Venture G. 2016. Optimal exciting dance for identifying inertial parameters of an anthropomorphic structure. *IEEE Transactions on Robotics* 32(4):823–836
13. Brunke L, Greeff M, Hall AW, Yuan Z, Zhou S, et al. 2022. Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems* 5:411–444
14. Yasa O, Toshimitsu Y, Michelis MY, Jones LS, Filippi M, et al. 2023. An overview of soft robotics. *Annual Review of Control, Robotics, and Autonomous Systems* 6(1):1–29
15. Atkeson CG, An CH, Hollerbach JM. 1986. Estimation of inertial parameters of manipulator loads and links. *The International Journal of Robotics Research* 5(3):101–119
16. Wensing PM, Niemeyer G, Slotine JJE. 2017. Observability in inertial parameter identification. *arXiv preprint arXiv:1711.03896*
17. Mayeda H, Yoshida K, Osuka K. 1988. *Base parameters of manipulator dynamic models*. In *Proceedings. 1988 IEEE International Conference on Robotics and Automation*, vol. 3, pp. 1367–1372
18. Gautier M, Khalil W. 1988. *A direct determination of minimum inertial parameters of robots*. In *Proceedings. 1988 IEEE International Conference on Robotics and Automation*, pp. 1682–1687. IEEE
19. Khalil W, Kleinfinger JF. 1987. Minimum operations and minimum parameters of the dynamic models of tree structure robots. *IEEE Journal on Robotics and Automation* 3(6):517–526
20. Ros J, Iriarte X, Mata V. 2012. 3d inertia transfer concept and symbolic determination of the base inertial parameters. *Mechanism and machine theory* 49:284–297

21. Gautier M. 1991. Numerical calculation of the base inertial parameters of robots. *Journal of robotic systems* 8(4):485–506
22. Van Huffel S, Vandewalle J. 1991. *The total least squares problem: computational aspects and analysis*. SIAM
23. Janot A, Vandanjon PO, Gautier M. 2013. A generic instrumental variable approach for industrial robot identification. *IEEE Transactions on Control Systems Technology* 22(1):132–145
24. Janot A, Vandanjon PO, Gautier M. 2009. *Using robust regressions and residual analysis to verify the reliability of LS estimation: Application in robotics*. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1962–1967. IEEE
25. Ramdani N, Poignet P. 2005. Robust dynamic experimental identification of robots with set membership uncertainty. *IEEE/ASME Transactions on Mechatronics* 10(2):253–256
26. Gevers M. 2005. Identification for control: From the early achievements to the revival of experiment design. *European journal of control* 11(4-5):335–352
27. Leboutet Q, Roux J, Janot A, Guadarrama-Olvera JR, Cheng G. 2021. Inertial parameter identification in robotics: A survey. *Applied Sciences* 11(9):4303
28. Ayusawa K, Venture G, Nakamura Y. 2014. Identifiability and identification of inertial parameters using the underactuated base-link dynamics for legged multibody systems. *The International Journal of Robotics Research* 33(3):446–468
29. Venture G, Ayusawa K, Nakamura Y. 2008. *Motion capture based identification of the human body inertial parameters*. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4575–4578. IEEE
30. Ogawa Y, Venture G, Ott C. 2014. *Dynamic parameters identification of a humanoid robot using joint torque sensors and/or contact forces*. In *2014 IEEE-RAS International Conference on Humanoid Robots*, pp. 457–462. IEEE
31. Del Prete A, Mansard N. 2016. Robustness to joint-torque-tracking errors in task-space inverse dynamics. *IEEE transactions on Robotics* 32(5):1091–1105
32. Sousa CD, Cortesao R. 2014. Physical feasibility of robot base inertial parameter identification: A linear matrix inequality approach. *The International Journal of Robotics Research* 33(6):931–944
33. Traversaro S, Brossette S, Escande A, Nori F. 2016. *Identification of fully physical consistent inertial parameters using optimization on manifolds*. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5446–5451. IEEE
34. Wensing PM, Kim S, Slotine JJE. 2017. Linear matrix inequalities for physically consistent inertial parameter identification: A statistical perspective on the mass distribution. *IEEE Robotics and Automation Letters* 3(1):60–67
35. Jovic J, Escande A, Ayusawa K, Yoshida E, Kheddar A, Venture G. 2016. Humanoid and human inertia parameter identification using hierarchical optimization. *IEEE Transactions on Robotics* 32(3):726–735
36. Ayusawa K, Venture G, Nakamura Y. 2011. *Real-time implementation of physically consistent identification of human body segments*. In *2011 IEEE International Conference on Robotics and Automation*, pp. 6282–6287. IEEE
37. Fazeli N, Kolbert R, Tedrake R, Rodriguez A. 2017. Parameter and contact force estimation of planar rigid-bodies undergoing frictional contact. *The International Journal of Robotics Research* 36(13-14):1437–1454
38. Kolev S, Todorov E. 2015. *Physically consistent state estimation and system identification for contacts*. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pp. 1036–1043. IEEE
39. Jatavallabhula KM, Macklin M, Golemo F, Voleti V, Petrini L, et al. 2021. gradsim: Differentiable simulation for system identification and visuomotor control. *arXiv preprint arXiv:2104.02646*
40. Hu Y, Anderson L, Li TM, Sun Q, Carr N, et al. 2019. Difftaichi: Differentiable programming

for physical simulation. *arXiv preprint arXiv:1910.00935*
41. Xu J, Chen T, Zlokapa L, Foshey M, Matusik W, et al. 2021. An end-to-end differentiable framework for contact-aware robot design. *arXiv preprint arXiv:2107.07501*
42. Degrave J, Hermans M, Dambre J, et al. 2019. A differentiable physics engine for deep learning in robotics. *Frontiers in neurorobotics* :6
43. Hahn D, Banzet P, Bern JM, Coros S. 2019. Real2sim: Visco-elastic parameter estimation from dynamic motion. *ACM Transactions on Graphics (TOG)* 38(6):pap. 236
44. Geilinger M, Hahn D, Zehnder J, Bächer M, Thomaszewski B, Coros S. 2020. Add: Analytically differentiable dynamics for multi-body systems with frictional contact. *ACM Transactions on Graphics (TOG)* 39(190):pap. 190
45. Werling K, Omens D, Lee J, Exarchos I, Liu CK. 2021. *Fast and feature-complete differentiable physics engine for articulated rigid bodies with contact constraints*. In *Proceedings of Robotics: Science and Systems XVII, ed. DA Shell, M Toussaint, MA Hsieh*, pp. pap. 34. N.p.: Robot. Sci. Syst. Found
46. Suh HJ, Simchowitz M, Zhang K, Tedrake R. 2022. *Do differentiable simulators give better policy gradients?* In *Proceedings of the 39th International Conference on Machine Learning*, pp. 20668–20696. N.p.: PMLR
47. Antonova R, Yang J, Jatavallabhula KM, Bohg J. 2023. *Rethinking optimization with differentiable simulation from a global perspective*. In *Proceedings of The 6th Conference on Robot Learning*, pp. 276–286. N.p.: PMLR
48. Lidec QL, Jallet W, Montaut L, Laptev I, Schmid C, Carpentier J. 2023. Contact models in robotics: a comparative analysis. *arXiv preprint arXiv:2304.06372*
49. Li W, Slotine JJE. 1989. An indirect adaptive robot controller. *Systems & Control Letters* 12(3):259–266
50. Yoshida K, Osuka K, Mayeda H, Ono T. 1996. When is the set of base-parameter values physically impossible? *Journal of the Robotics Society of Japan* 14(1):122–130
51. Yoshida K, Khalil W. 2000. Verification of the positive definiteness of the inertial matrix of manipulators using base inertial parameters. *The International Journal of Robotics Research* 19(5):498–510
52. Lynch KM, Park FC. 2017. *Modern robotics*. Cambridge University Press
53. Wittenburg J. 1980. *Dynamics of multibody systems*. In *Proceedings, XVth IUTAM/ICTAM Congress*. Springer
54. Todorov E, Erez T, Tassa Y. 2012. *Mujoco: A physics engine for model-based control*. In *2012 IEEE/RSJ international conference on intelligent robots and systems*, pp. 5026–5033. IEEE
55. Lee T, Park FC. 2018. A geometric algorithm for robust multibody inertial parameter identification. *IEEE Robotics and Automation Letters* 3(3):2455–2462
56. Moakher M. 2005. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications* 26(3):735–747
57. Lee T, Wensing PM, Park FC. 2019. Geometric robot dynamic identification: A convex programming approach. *IEEE Transactions on Robotics* 36(2):348–365
58. Lee T. 2019. Geometric methods for dynamic model-based identification and control of multibody systems. *Seoul National University*
59. Lee T, Lee BD, Park FC. 2021. Optimal excitation trajectories for mechanical systems identification. *Automatica* 131:109773
60. Presse C, Gautier M. 1992. *Bayesian estimation of inertial parameters of robots*. In *Proceedings 1992 IEEE International Conference on Robotics and Automation*, pp. 364–365. IEEE Computer Society
61. Boffi NM, Slotine JJE. 2021. Implicit regularization and momentum algorithms in nonlinearly parameterized adaptive control and prediction. *Neural Computation* 33(3):590–673
62. Lee T, Kwon J, Park FC. 2018. *A Natural Adaptive Control Law for Robot Manipulators*. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE.

http://doi.org/10.1109/IROS.2018.8593727
63. Zeng A, Song S, Lee J, Rodriguez A, Funkhouser T. 2020. Tossingbot: Learning to throw arbitrary objects with residual physics. *IEEE Transactions on Robotics* 36(4):1307–1319
64. Golemo F, Taiga AA, Courville A, Oudeyer PY. 2018. *Sim-to-real transfer with neural-augmented robot simulation.* In *Proceedings of The 2nd Conference on Robot Learning*, pp. 817–828. N.p.: PMLR
65. Heiden E, Millard D, Coumans E, Sheng Y, Sukhatme GS. 2021. *NeuralSim: Augmenting differentiable simulators with neural networks.* In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9474–9481. IEEE
66. Ajay A, Wu J, Fazeli N, Bauza M, Kaelbling LP, et al. 2018. *Augmenting physical simulators with stochastic neural networks: Case study of planar pushing and bouncing.* In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3066–3073. IEEE
67. Nguyen-Tuong D, Peters J. 2011. Model learning for robot control: a survey. *Cognitive processing* 12:319–340
68. Marsden JE, West M. 2001. Discrete mechanics and variational integrators. *Acta numerica* 10:357–514
69. Lee J, Liu CK, Park FC, Srinivasa SS. 2020. *A linear-time variational integrator for multibody systems.* In *Proceedings of the Twelfth Workshop on the Algorithmic Foundations of Robotics*, pp. 352–367. Springer
70. Kim M, Lee Y, Lee Y, Lee D. 2017. Haptic rendering and interactive simulation using passive midpoint integration. *The International Journal of Robotics Research* 36(12):1341–1362
71. Spong MW. 1996. Energy based control of a class of underactuated mechanical systems. *IFAC Proceedings Volumes* 29(1):2828–2832
72. Spong MW. 2022. An historical perspective on the control of robotic manipulators. *Annual Review of Control, Robotics, and Autonomous Systems* 5(1):1–31
73. Lutter M, Ritter C, Peters J. 2019. Deep lagrangian networks: Using physics as model prior for deep learning. *arXiv preprint arXiv:1907.04490*
74. Lutter M, Listmann K, Peters J. 2019. *Deep Lagrangian Networks for end-to-end learning of energy-based control for under-actuated systems.* In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7718–7725. IEEE
75. Zhong YD, Dey B, Chakraborty A. 2021. Extending lagrangian and hamiltonian neural networks with differentiable contact models. *Advances in Neural Information Processing Systems* 34:21910–21922
76. Johanastrom K, Canudas-De-Wit C. 2008. Revisiting the lugre friction model. *IEEE Control Systems Magazine* 28(6):101–114
77. Yoon J, Lee M, Son D, Lee D. 2022. Fast and accurate data-driven simulation framework for contact-intensive tight-tolerance robotic assembly tasks. *arXiv preprint arXiv:2202.13098*
78. Jiang Y, Sun J, Liu CK. 2022. *Data-augmented contact model for rigid body simulation.* In *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, pp. 378–390. N.p.: PMLR
79. Pfrommer S, Halm M, Posa M. 2021. *Contactnets: Learning discontinuous contact dynamics with smooth, implicit representations.* In *Proceedings of the 2020 Conference on Robot Learning*, pp. 2279–2291. N.p.: PMLR
80. Le Cleac'h S, Yu HX, Guo M, Howell T, Gao R, et al. 2023. Differentiable physics simulation of dynamics-augmented neural objects. *IEEE Robotics and Automation Letters* 8(5):2780–2787
81. Mason MT. 2018. Toward robotic manipulation. *Annual Review of Control, Robotics, and Autonomous Systems* 1(1):1–28
82. Sanchez-Gonzalez A, Heess N, Springenberg JT, Merel J, Riedmiller M, et al. 2018. *Graph networks as learnable physics engines for inference and control.* In *Proceedings of the 35th International Conference on Machine Learning*, pp. 4470–4479. N.p.: PMLR

83. Battaglia P, Pascanu R, Lai M, Jimenez Rezende D, et al. 2016. Interaction networks for learning about objects, relations and physics. *Advances in neural information processing systems* 29
84. Allen KR, Guevara TL, Rubanova Y, Stachenfeld K, Sanchez-Gonzalez A, et al. 2023. *Graph network simulators can learn discontinuous, rigid contact dynamics*. In *Proceedings of The 6th Conference on Robot Learning*, pp. 1157–1167. N.p.: PMLR
85. Zhu X, Wang D, Biza O, Su G, Walters R, Platt R. 2022. *Sample Efficient Grasp Learning Using Equivariant Models*. In *Proceedings of Robotics: Science and Systems XVIII, ed. K Hauser, D Shell, S Huang*, pp. pap. 71. N.p.: Robot. Sci. Syst. Found
86. Kim S, Lim B, Lee Y, Park FC. 2023. *SE (2)-Equivariant Pushing Dynamics Models for Tabletop Object Manipulations*. In *Proceedings of The 6th Conference on Robot Learning*, pp. 427–436. N.p.: PMLR
87. Han J, Huang W, Ma H, Li J, Tenenbaum J, Gan C. 2022. Learning physical dynamics with subequivariant graph neural networks. *Advances in Neural Information Processing Systems* 35:26256–26268
88. Lee Je, Lee J, Bandyopadhyay T, Sentis L. 2023. Sample efficient dynamics learning for symmetrical legged robots: Leveraging physics invariance and geometric symmetries. *IEEE International Conference on Robotics and Automation*
89. Richards SM, Azizan N, Slotine JJE, Pavone M. 2021. *Adaptive-control-oriented meta-learning for nonlinear systems*. In *Robotics: Science and Systems*
90. Singh S, Richards SM, Sindhwani V, Slotine JJE, Pavone M. 2021. Learning stabilizable nonlinear dynamics with contraction-based regularization. *The International Journal of Robotics Research* 40(10-11):1123–1150