

# Learning State-Dependent Losses for Inverse Dynamics Learning

Kristen Morse\*, Neha Das\*, Yixin Lin, Austin Wang, Akshara Rai, Franziska Meier

**Abstract**—Being able to quickly adapt to changes in dynamics is paramount in model-based control for object manipulation tasks. In order to influence fast adaptation of the inverse dynamics model’s parameters, data efficiency is crucial. Given observed data, a key element to how an optimizer updates model parameters is the loss function. In this work, we propose to apply meta-learning to learn structured, state-dependent loss functions during a meta-training phase. We then replace standard losses with our learned losses during online adaptation tasks. We evaluate our proposed approach on inverse dynamics learning tasks, both in simulation and on real hardware data. In both settings, the structured and state-dependent learned losses improve online adaptation speed, when compared to standard, state-independent loss functions.

## I. INTRODUCTION

Truly autonomous robots need to be able to adapt their internal models when interacting with the environment. For instance, a robot that has learned its own dynamics needs to be able to adapt to the changes in dynamics caused by picking and placing a heavy object (Figure 1). Humans are remarkably good at adapting to such changes very quickly. And studies on how humans adapt to external forces suggest that we *learn how to adapt* to such changes in dynamics, and that we remember the adaptation strategy when encountering previously experienced perturbations [1].

The larger goal of our work is to learn representations that endow robots with similar capabilities. In this work, we specifically look at fast learning or adaptation of inverse dynamics for model-based control. While many robots have analytical inverse dynamics models, they are often crude approximations to the actual dynamics of the robot. Furthermore, even if the models would be accurate enough for the robot’s dynamics, they do not model the changes in dynamics caused by interaction with the environment. Thus, a lot of research has gone into data-driven methods for model-based control [2], with the hope that learning-based methods can either learn better models altogether [3], [4], or learn error-models on top of existing analytical ones [5].

In that context, online learning or adaptation of inverse dynamics models has been studied [6], [7], [8]. When such online learning methods are deployed on real robots, one has to be concerned with computational and data efficiency of the optimizer. An online learner needs to make the most out of the recently observed data with as few parameter updates as possible. A key insight of our work is that standard losses, such as the Mean-Squared-Error (MSE), are not state-dependent, but observed errors and how much

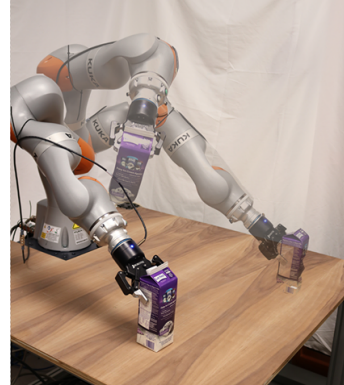


Fig. 1: Picking and placing a milk carton with KUKA.

we should correct for them can be state-dependent. For instance, an optimizer could update dynamics parameters aggressively throughout most of the state-space, but needs to be more conservative around singular configurations, or after overcoming the effects of static friction. Learned, state-dependent losses could potentially incorporate such issues and thus lead to faster inverse dynamics model learning. Thus, in this work, we aim at improving data efficiency, by replacing standard loss functions for inverse dynamics learning, with learned, state-dependent losses.

Towards this goal, we develop a meta-learning framework, that trains loss functions for fast adaptation of inverse dynamics models. Our approach builds on [9], a gradient-based meta-learning framework for learning loss functions. This work proposed learning losses that are parameterized as neural networks for various learning tasks. Here, we contribute the following: 1) we utilize [9] to formulate an offline meta-training algorithm for training loss functions for inverse dynamics learning, and 2) we propose two structured loss representations specifically suitable for learning inverse dynamics models, one of which is state-dependent.

We evaluate our proposed loss representations on several inverse dynamics learning tasks, both in simulation as well as on real hardware data of a 7 degree of freedom robot arm. Our results show that meta-learning structured losses leads to more robust meta-training of the losses, as compared to using typical neural network representations for the loss function. Furthermore, we find that our structured losses lead to faster learning and online adaptation of inverse dynamics models. Finally, especially on real robot learning experiments, the proposed state-dependent loss outperforms all other loss representations in terms of adaptation speed, providing experimental validation for the thought that learning state-dependent loss functions can improve data efficiency in

\*authors contributed equally

All authors are with Facebook AI Research. {kristenmorse, dasneha, yixinlin, wangaustin, akshararai, fmeier}@fb.com

inverse dynamics learning.

## II. BACKGROUND AND RELATED WORK

We start out by describing the control architecture used in this paper, and then summarize related works with respect to learning inverse dynamics and meta-learning.

### A. Background: Inverse Dynamics Control

A typical model-based control loop for inverse dynamics control as used in our work is shown in Figure 2. The controller accepts a target position  $q_{target}$  and outputs a plan - the desired joints, joint velocities and joint accelerations  $Q_d = \{q_d, \dot{q}_d, \ddot{q}_d\}$  for each time step  $t$  of the trajectory. At each time step, the inverse dynamics model takes the current joint position  $q$  and velocity  $\dot{q}$ , as well as the desired acceleration  $\ddot{q}_d$  as input and outputs a feed forward torque  $u_{ff}$ . The inverse dynamics controller then combines  $u_{ff}$  with a feedback  $u_{fb}$  component to produce the total desired torque  $\tau$  applied on the robot. The feedback component is computed through a PD-control law that corrects for the errors caused due to an inaccurate dynamics model:  $u_{fb} = K_p(q_d - q) + K_d(\dot{q}_d - \dot{q})$ .  $u_{fb}$  corrects for inaccuracies in the dynamics model, and pushes the robot towards the desired trajectory. Ideally, for compliant and fast motion, we want small contributions from the feedback, and most of the torque to come from  $u_{ff}$ , motivating the need for learning accurate inverse dynamics models.

### B. Learning Inverse dynamics

In this work we apply meta-learning to the problem of improving inverse dynamics models, in an attempt to train as quickly as possible. The problem of inverse dynamics learning is widely studied in robotics literature, ranging from learning analytical rigid-body models, as in [10] to over-parametric neural-network models, as in [11]. Given the computational and data constraints of learning dynamics models on real robots, there has been significant work on computationally efficient methods for learning and adapting inverse dynamics models [11], [12], [13], [3], [8], including learning *incrementally* in an online manner [6], [14], [15], [16]. Another line of work assumes a prior model (e.g. analytical), which will be inaccurate in the real world, and attempts to learn a model over this error instead of the entire inverse dynamics model [5], [7], [17], [18]. In all of these approaches, the dynamics are learned to optimize the mean-squared error, and do not benefit from a state-dependent loss function that can lead to faster learning online. We present a meta-learning framework that can learn structured, state-dependent loss functions for inverse dynamics learning, significantly improving online learning speed.

### C. Meta-Learning

Learning-to-adapt can be considered a form of meta-learning, where an agent attempts to learn *how to learn* from observations. Meta-learning has become very popular within the machine-learning community [19], [20], [9] and we can differentiate roughly between three types of approaches for

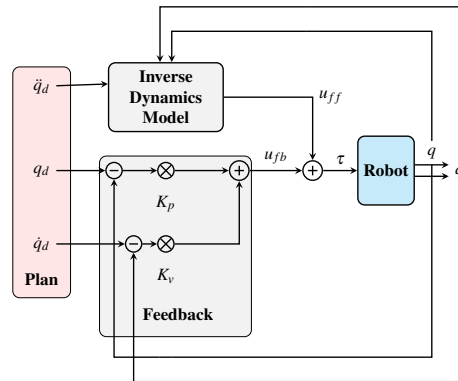


Fig. 2: In this work we learn the model used within an inverse dynamics controller. The controller accepts a target position  $q_{target}$  and outputs a plan - the desired joints, joint velocities and joint accelerations  $Q_d = \{q_d, \dot{q}_d, \ddot{q}_d\}$  for each time step  $t$  of the trajectory. At each time step the inverse dynamics model accepts the current joint position  $q$  and velocity  $\dot{q}$ , as well as the desired acceleration  $\ddot{q}_d$  and outputs a feed forward torque. The feedback controller computes a feedback torque from the current joints and joint velocities  $q$  and  $\dot{q}$ . The combination of  $u_{ff}$  and  $u_{fb}$  gives us the torque control parameters  $\tau$  required to drive the robot arm to  $q_{target}$ .

learning-to-adapt: Methods that meta-learn initialization of models from which an agent can quickly adapt to new scenarios [19], methods that learn an optimizer representation [21], [22], and finally methods that learn loss functions [9], [23], [24]. This work focuses on the latter, learning loss functions. For a more detailed overview of meta-learning approaches see [25].

In this work, we focus on learning loss functions, and build on the work by [9], which meta-learns a loss function in order to generalize across tasks. We modify their framework for training loss functions for inverse dynamics learning, and use it to efficiently update dynamics models at test time. We additionally introduce state-dependent losses, that allow for changes in the loss landscape based on the current robot state.

### D. Meta-Learning Dynamics Models

There has also been research in meta-learning with a focus on meta-learning dynamics models, either with fast fine-tuning [26] or with online learning of a memory learning rates [20]. [26] meta-learn a dynamics model initialization in simulation that can quickly be fine-tuned during test time to adapt to different settings. However, adaptation to completely novel dynamics, not captured in the meta-train task-distributions, remains a challenge for such approaches. On the other hand, [20] present an approach to learn the learning rate used in model updates online, which can increase online adaptation efficiency. But because the model that predicts learning rates is learned online, it is affected by forgetting, thus information about prior tasks is lost.

In this work, we learn a *state-dependent* loss function during an offline meta-training phase; and show that it transfers to previously unseen tasks, with novel dynamics (such as additional load at the end-effector) at meta-test

time.

### III. LEARNING LOSS FUNCTIONS FOR INVERSE DYNAMICS MODELS

In this section we present our approach to learning structured loss functions for inverse dynamics learning. We built on prior work on meta-learning [9] which has shown that learned loss functions can lead to faster learning for various supervised learning problems. Here, we propose to learn state-dependent loss functions with structure.

Our loss-learning framework comprises of two phases: *Phase 1 (meta train)*: Meta-Training is the process by which we learn our loss function. More specifically, we use data collected from sine motions on the joints to optimize the parameters  $\phi$  of our loss function  $\mathcal{L}_{\text{learn}\phi}$ . *Phase 2 (meta test)*: Meta-testing (which is similar to a typical training phase) is the process by which we train our inverse dynamics model. We use the  $\mathcal{L}_{\text{learn}\phi}$  with optimal parameters  $\phi^*$  to train  $f_\theta$  on new tasks. We start out by recapping how to learn inverse dynamics models given any loss function.

#### A. Learning Models for Inverse Dynamics Control

We aim to learn a model of the inverse dynamics - i.e given the current joint positions  $q$  and velocities  $\dot{q}$ , and the next desired joint acceleration  $\ddot{q}_d$ , we want to learn a function  $f$ , parameterized by  $\theta$  such that:

$$u_{ff} = f_\theta(q, \dot{q}, \ddot{q}_d)$$

where  $u_{ff}$  is the feed-forward torque that should be applied to achieve the desired acceleration  $\ddot{q}_d$  as per the inverse dynamics model. In this work, we model  $f$  via a neural network parameterized by  $\theta$ , and we aim to learn it from a stream of data as quickly as possible. This model is trained on a data set  $\mathcal{D} = \{Q_t = (q_t, \dot{q}_t, \ddot{q}_{t+1}), \tau_t\}_{t=1}^T$  by minimizing a loss  $\mathcal{L}_{\text{MSE}}$  between the predicted control outputs ( $u_{ff}$ ) and ground truth torque values ( $\tau$ ). The inverse dynamics model can be learned via stochastic gradient descent (SGD) by iteratively updating the parameters  $\theta$  as follows:

$$\theta_{\text{new}} = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\text{MSE}}$$

where  $\alpha$  is the chosen learning rate. The MSE loss, just like any other typical loss, is a domain-independent loss that can be used for any function approximation problem. We believe that learning loss functions for inverse dynamics learning can lead to losses that update these models more effectively.

#### B. Loss Function Representations

Instead of using unstructured neural-network architectures as in [9] to represent the loss, we propose to learn *structured* loss functions. Next, we describe in detail the different loss architectures explored in our work.

1) *Multi-Layer Perceptron (mlp)*: In [9] the learned loss is represented as a neural network. Similar to the MSE loss, the learned loss  $l = \mathcal{L}_{\text{learn}\phi}(\hat{\tau}, \tau)$  takes as input the predicted output  $\hat{\tau} = f_\theta$  and its corresponding ground truth value  $\tau$ , and outputs a loss value that computes an effective distance between  $f_\theta$  and  $\tau$ . The loss architecture is a fully-connected

---

#### Algorithm 1: Learning Loss functions at (*meta-train*)

---

```

1: Randomly initialize  $\phi$ 
2: for each epoch do
3:   for each batch in epoch do
4:     Randomly initialize  $\theta$ 
5:     Sample a batch of continuous data  $q_{t:t+K}, \dot{q}_{t:t+K}, \ddot{q}_{t:t+K}, \tau_{t:t+K}$ 
6:     Split batch into 2 batches
7:      $q_A, \dot{q}_A, \ddot{q}_A, \tau_A = q_{t:t+\frac{K}{2}}, \dot{q}_{t:t+\frac{K}{2}}, \ddot{q}_{t:t+\frac{K}{2}}, \tau_{t:t+\frac{K}{2}}$ 
8:      $q_B, \dot{q}_B, \ddot{q}_B, \tau_B = q_{\frac{K}{2}:K}, \dot{q}_{\frac{K}{2}:K}, \ddot{q}_{\frac{K}{2}:K}, \tau_{\frac{K}{2}:K}$ 
9:     for each  $i < \text{iters}_{\text{max}}$  do
10:      Calculate  $\theta_{\text{new}}$  from  $\mathcal{L}_{\text{learn}\phi}$  with the old  $\phi$ 
11:       $\theta_{\text{new}} \leftarrow \theta - \alpha \cdot \nabla_{\theta} \mathcal{L}_{\text{learn}\phi}(f_\theta(q_A, \dot{q}_A, \ddot{q}_A), \tau_A)$ 
12:      Update  $\phi$  based on  $f_{\theta_{\text{new}}}$ 's performance
13:       $\phi \leftarrow \phi - \eta \cdot \nabla_{\phi} \mathcal{L}_{\text{MSE}}(f_{\theta_{\text{new}}}(q_B, \dot{q}_B, \ddot{q}_B), \tau_B)$ 
14:    end for
15:  end for
16: end for

```

---

MLP network with 3 layers of 40 neurons each, with a ReLU activation in the hidden layer. A Softplus activation is applied to the output. This loss representation is our baseline.

2) *Structured loss (structured)*: The *mlp* architecture doesn't take into account the structure of the inverse dynamics learning problem. Here, we propose to use a parameterized loss of the form:

$$\mathcal{L}_{\text{learn}\phi} = \sum_{j=1}^J \phi_j (f_{\theta_j} - \tau_j)^2, \quad (1)$$

where  $J$  denotes the number of controllable joints and  $\phi_j \in \mathbb{R}^1$  are learn-able parameters, one per joint  $j$ . This structured learned loss takes into account to the shape of the inverse-dynamics model's outputs (i.e a predicted torque  $f_\theta$  per joint), and learns to assign different weights to the different joints. Intuitively, dynamics errors in some joints can cause poor tracking of desired joint acceleration and this loss learns to weigh torque error on crucial joints more heavily than others. Updating more important joints faster likely leads to improvement in online learning speed.

3) *State Dependent loss (state-dependent)*: Our proposed structured loss assumes constant learned weights for the whole robot state space. Here, we introduce a state-dependent loss with weights  $\phi_j$  that are a function of the current joint state  $\{q, \dot{q}\}$ :

$$\mathcal{L}_{\text{learn}\phi} = \sum_{j=1}^J \phi_j(q, \dot{q}) (f_{\theta_j} - \tau_j)^2. \quad (2)$$

where the functions  $\phi_j$  are represented learnable as neural networks with hidden layer of size 32 with ReLUs as activation functions. Our intuition is that a state-dependent loss can improve upon structured loss by adapting to (ie increasing weight near) areas of the state space where the robot experiences static friction.

#### C. Learning loss functions

These loss functions can be learned offline, during a meta-training phase. We employ stochastic gradient descent (SGD) to update the parameters  $\phi$  of the learned loss. Note that at each step of the optimization process,  $\phi$  must be updated in

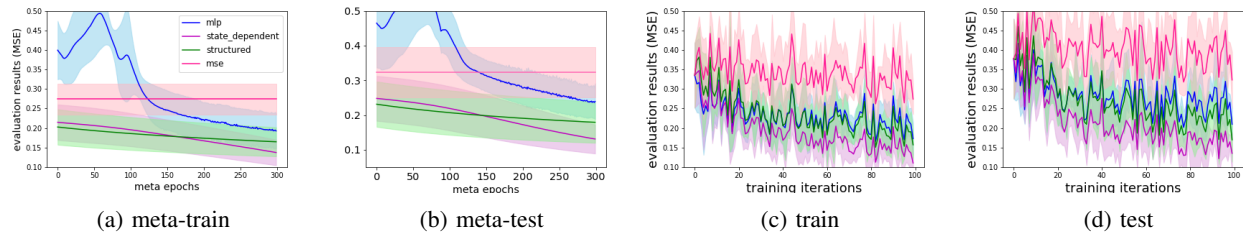


Fig. 3: **Simulation meta-training phase:** (a) and (b) plot the performance of the learned losses (on meta train data(a) and meta test data(b) collected on hardware) as a function of meta-training epochs. The losses improve over a course of 300 epochs. (c) and (d) show training curves for learning  $f_\theta$  when using the final meta (learned) losses and the fixed MSE loss. The state-dependent loss is visibly more stable to train and faster to optimize the task model with compared to other learned losses. Results averaged across 5 seeds.

a way such that training the model  $f_\theta$  with the corresponding  $\mathcal{L}_{\text{learn}_\phi}$  brings us a step closer to obtaining an accurate model of the robot’s inverse-dynamics. This can be checked by evaluating the resultant inverse-dynamics model with the  $\mathcal{L}_{\text{MSE}}$ .

More concretely, we first update the parameters  $\theta$  via the update rule  $\theta_{\text{new}} \leftarrow \theta - \alpha \nabla_\theta \mathcal{L}_{\text{learn}_\phi}$  using batches of continuous (sequential) inverse dynamics data  $q_{t:t+K}, \dot{q}_{t:t+K}, \ddot{q}_{t:t+K}, \tau_{t:t+K}$ , where  $K$  is the batch size, from the meta-train data set. This batch is halved, and the first half is used to update the parameters of the inverse dynamics model using the current state of the learned loss  $\mathcal{L}_{\text{learn}_\phi}$ . We denote the new updated parameters as  $\theta_{\text{new}}$ . We can now evaluate this updated model  $f_{\theta_{\text{new}}}$  on the 2nd half of the batch. This mimics the incremental learning at test time, where the  $\mathcal{L}_{\text{learn}_\phi}$  updates  $f_\theta$  on sequentially incoming data.

The update of the loss parameters  $\phi$ , is done implicitly through the evaluation of  $f_{\theta_{\text{new}}}$ :

$$\phi \leftarrow \phi - \eta \cdot \nabla_\phi \mathcal{L}_{\text{MSE}}(f_{\theta_{\text{new}}}(Q), \tau) \quad (3)$$

While it is possible to analytically derive this gradient (See [9]), one can use modern libraries for such meta-gradient computations. In particular, we use *higher* [27]. Our approach for meta-training a learned loss for learning inverse-dynamics models is summarized in Algorithm 1

#### IV. SIMULATION EXPERIMENTS

In this section we evaluate our proposed loss function representations for inverse dynamics learning in simulation. In particular, we aim to evaluate how well the different loss function representations train and how online adaptation of inverse dynamics with learned loss compares to using a standard MSE loss.

##### A. Meta-training: Learning loss functions

We start by training our loss function representations in simulation. In our meta-training phase, we collect data, meta-learn the loss functions, and evaluate the different types of learned losses (*mlp*, *structured*, and *state-dependent*).

1) *Data Collection:* For our simulation experiments, we record the motion of the KUKA iiwa moving in the PyBullet simulator [28]. We collect multiple sine-motion trajectories that were generated around the rest posture of the arm. More specifically, let  $q_{\text{rest}}$  be the rest posture, then we

generate the desired joint position at time step  $t$  to be  $q_t = A \sin(2\pi f \Delta t) + q_{\text{rest}}$ , where  $\Delta t$  is the length of a time step,  $f$  is the frequency of the sine motion, and  $A$  is the amplitude of the sine motion, with a value that is chosen to avoid self-collision. To get a rich data set we collect sine-motion data at various frequencies. Specifically, we collect sine motion data at frequencies  $f = [0.01, 0.02, \dots, 0.09]$ . The simulator runs at 240 Hz, and for each frequency we collect 10s of data. We split the collected data into meta-train and meta-test data sets, where meta-train data consists of data collected with frequencies  $[0.01, 0.03, 0.05, 0.06, 0.07, 0.08]$  and meta-test data of data collected with frequencies  $[0.02, 0.04, 0.09]$ .

2) *Evaluation of Meta-Training:* We first meta-train our learned loss functions *mlp*, *structured* and *state-dependent* described in Section III-B with the meta-train data set according to Algorithm 1 with hyper-parameters:

Batch Size	$\alpha$	$\eta$	iters <sub>max</sub>
256	0.001	0.01	10

After meta-training has finished, we evaluate our learned losses against Pytorch’s built in Mean Squared Error Loss [29] for comparison by training an inverse-dynamics model with each of the losses. We perform two kinds of evaluations:

*During meta-training* After each epoch, we use the current losses and use them to train inverse dynamics models, on 1) the meta-train data itself, and on 2) on the held-out meta-test data (see Figure 3 (a) and (b) respectively). We compare to using the MSE loss. Note while we use the various losses to perform the learning we evaluate the MSE achieved for all of them, to make results comparable.

*After Meta-Training:* We use all learned losses and the MSE loss to perform a final training of the inverse dynamics on both meta-train and meta-test data, for 100 gradient steps. Results of these training curves, again visualizing MSE achieved by the various losses, are shown in Figure 3 (c) and (d), respectively.

3) *Analysis:* We note from Figure 3 (a) and (b), that by 150 epochs, all three of the learned losses have achieved better results than MSE. While the *mlp* loss improves upon MSE loss, it is also less stable to train than the *structured* and *state-dependent* losses (observe how the evaluation metric for the model  $f_\theta$  varies smoothly with the number of

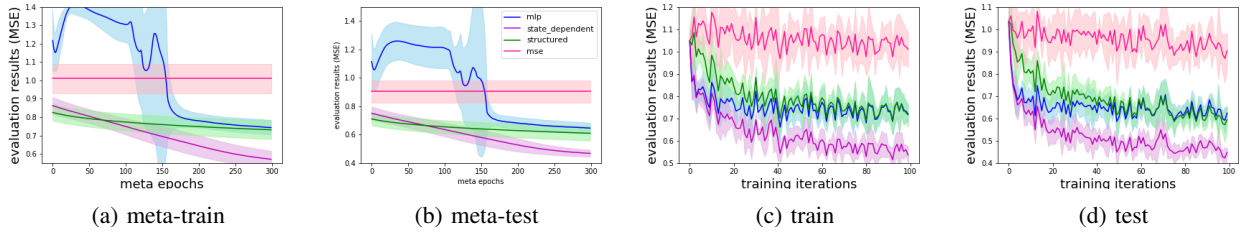


Fig. 4: **Hardware meta-training phase:**(a) and (b) plot the performance of the learned losses (on meta train data(a) and meta test data(b) collected on hardware) as a function of meta-training epochs. The losses improve over a course of 300 epochs, with 100 gradient steps per epoch. (c) and (d) show training curves for learning  $f_\theta$  when using the final meta (learned) and fixed (mse) losses. The state-dependent loss is visibly more stable to train and faster to optimize the inverse-dynamics model with, compared to other learned losses. Results are averaged across 5 seeds.

meta-epochs for these particular losses). This shows that incorporating information about the inverse dynamics model and the nature of its outputs in the loss function make the loss function easier to train. The best performer amongst all the learned losses is the *state-dependent* loss, suggesting that the loss landscape for training inverse-dynamics model is dependent on the current state.

Plots from Figure 3 (c) and (d) show that training the inverse dynamics model via the learned losses progresses much faster as compared to training via MSE loss. Therefore, these results suggest that learned losses indeed lead to faster training of inverse dynamics models. A key takeaway is that the *structured* loss performs better than the *mlp* loss, and *state-dependent* is the best performer.

### B. Online Adaptation

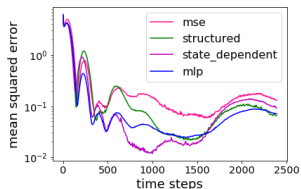


Fig. 5: Online Learning with learned losses: MSE of predicted torque of the inverse dynamics model as a function of time. At each time step the model is updated through one gradient step.

In the previous experiments, we evaluated the trained loss functions on learning inverse dynamics models from randomly sampled batches. Here, we evaluate how fast training progresses when using the learned losses to train on streaming data. For this experiment, we use the loss functions trained in the previous experiment to train a randomly initialized inverse dynamics model online. We use the meta-test trajectories obtained from the data collection process described in section IV-A.1, but instead of randomly sampling from that data, training passes through the data sequentially, in batches of size 5. In detail, at every update step  $k$ , we create a batch from the next 5 time steps, record the prediction error (MSE) made on that batch, and then perform one gradient update step on model parameters  $\theta$  based on the learned losses. We visualize this prediction error as a function of time in Figure 5. The slope and stability

of the MSE curves in the figure indicate how quickly the model is able to adapt to an incoming stream of data. We note from Figure 5 that the MSE curves for the inverse model trained via learned losses output lower values throughout the training process than the ones trained using an MSE loss. This indicates that training with a learned loss is much more suitable when the model needs to adapt in an online setting.

## V. HARDWARE EXPERIMENTS

We now turn to evaluating our proposed loss architectures on real hardware data. While simulation results are already promising, we expect the benefits of using *state-dependent* losses to be even larger on real data, as a lot of effects like stiction do not appear in simulation.

### A. Meta-training: Learning loss functions

Similar to simulation in Section IV-A.2, we first evaluate the meta-training phase - in particular, how well our learned losses can optimize inverse dynamics models on unseen data.

1) *Data Collection:* For our real hardware experiments, we collect two types of data sets from our Kuka iiwa, which is gravity compensated. Similar to our simulation experiments, we collect sine motion data at frequencies  $f = [0.02, 0.03, \dots, 0.08]$ . Our controller runs at 250 Hz, and for each frequency we collect 30s of data. We split the collected data into meta-train and meta-test data sets, where meta-train data consists of data collected with frequencies  $[0.02, 0.03, 0.04, 0.06, 0.08]$  and meta-test data of data collected with frequencies  $[0.05, 0.07]$ .

For meta-training we use the same hyper-parameters as we did in Section IV-A

2) *Evaluation and Analysis:* The evaluation statistics plotted in Figure 4 and our analysis for evaluating meta-training for the real data from them mirror those for the simulation data (Section IV-A.2). Figures 4 (a) and (b) demonstrate that meta-training learned losses with a structure, (i.e, both *structured* and *state-dependent* losses) is much more stable than meta-training the *mlp* learned loss. From Figures 4(c) and (d), we again note that inverse-dynamics models train much faster when learned-losses are used for optimization as opposed to MSE loss in the order: *state-dependent* > *structured* > *mlp*.

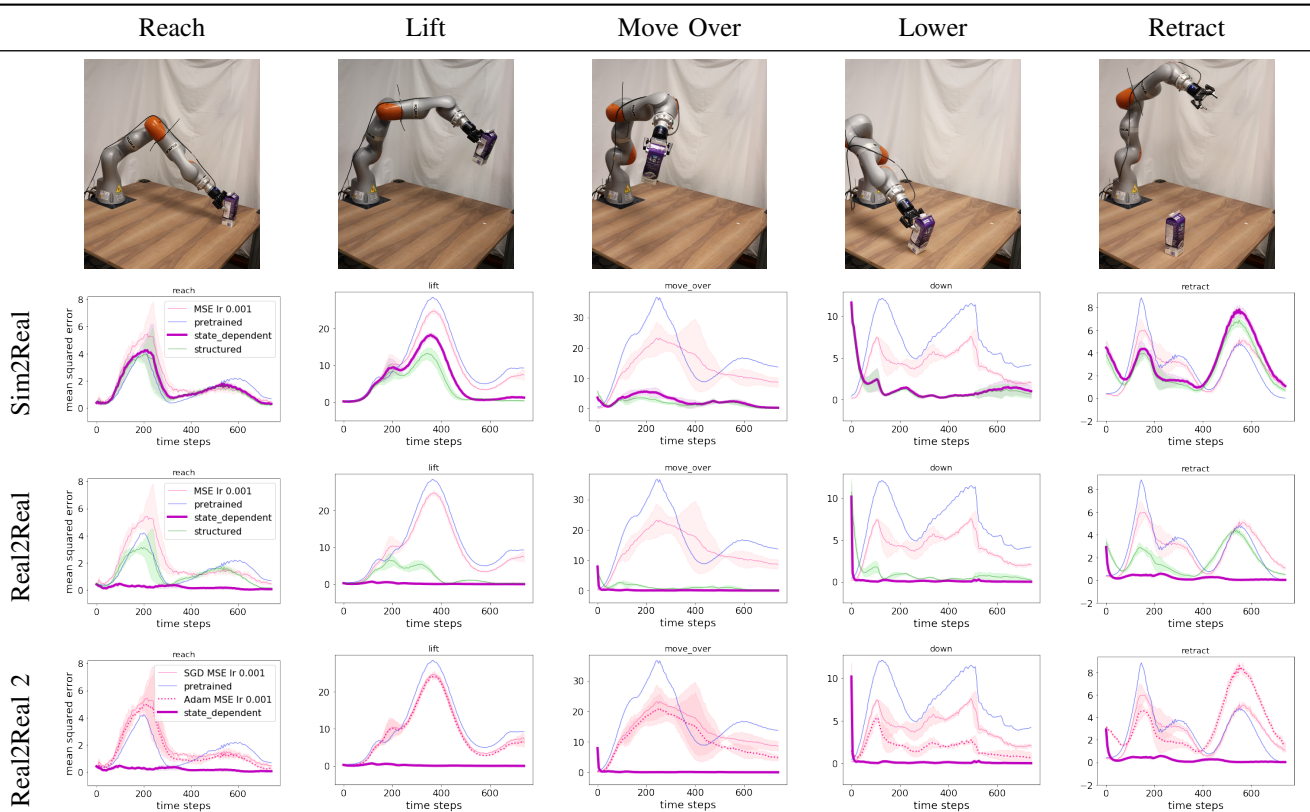


Fig. 6: The 1st row depicts a snapshot while executing each of the pick and place motions. 2nd to 4th row: Performance (MSE) of  $f_\theta$  on the 5 motions as it is being trained using  $\mathcal{L}_{\text{learn}_\theta}$  or  $\mathcal{L}_{\text{MSE}}$ . 2nd row (Sim2Real) uses loss functions trained in simulation, while the 3rd and 4th row (Real2Real) uses loss functions trained on hardware data. *Sim2Real, row 2*: The learned losses adapt the models faster/better than the MSE loss; however the *structured* and *state-dependent* loss perform similarly well (on average). *Real2Real, 3rd row*: The *state-dependent* loss significantly outperforms all other losses. *Real2Real, 4th row*: The *state-dependent* loss also significantly outperforms an adaptive optimizer. Also note the blue line: this line represents a dynamics model pretrained on all meta train data with a normal MSE loss and used as is (not adapted). Notice that the pretrained model does not perform well on this test task, highlighting the need for adaptation to this novel data distribution.

Note that while in the corresponding experiment on the simulation (Section IV-A), the *state-dependent* loss is the best performer, its improved performance on the real data is much more visible.

This significant improvement of the state-dependent loss, when compared with the experiments on simulation, can be explained by considering the fact that the hardware has to deal with real world issues, such as static friction and noise. We believe the state-dependent loss has learned how to penalize errors depending on where in the state-space the robot is.

### B. Online Adaptation on Pick and Place Task

We now use the learned losses to online learn the inverse dynamics model of a pick and place task.

1) *Data Collection and Setup*: The second data set consists of data collected for a pick-and-place task (see top row of Figure 6). We collect 3 trials of picking-and-placing a milk carton, which weighs 857 grams, and separate the task into 5 motions: reaching for the carton (*reach*), lifting the carton (*lift*), moving the carton across the table (*move-over*),

putting the carton down (*move-down*), and retracting the arm to its rest posture (*retract*).

For each trial, the inverse-dynamics model  $f_\theta$  is initialized randomly and then trained by online updating its parameters  $\theta$  using the various learned losses on the stream of sequential data points  $\{q_t, \dot{q}_t, \dot{q}_{t+1}, \tau_t\}$  we receive while undergoing each of the above 5 motions. We train our model  $f_\theta$  on each of the 5 motion phases sequentially, always warm-starting the model for the next motion. As a baseline, we also pretrain a dynamics model  $f$  on the meta-train data. In the following evaluation this pretrained model is not adapted online, and thus is an indicator on how much adaptation is needed.

2) *Evaluation*: We record the MSE between predicted and actual torques while training on each task within pick and place with the learned losses *mlp*, *structured* and *state-dependent* each with learning rates 0.001. For comparison, we plot these in Figure 6 alongside an MSE loss (with learning rate .001) and with our baseline pretrained model. The second row of Figure 6 shows the results for learned losses trained on the simulator data, indicating the ability of a sim-to-real transfer from the losses meta-trained on sim data to the model-training on real data. For the third row of

Loss Functions	Reaching		Lifting		Moving Over		Putting Down		Retracting	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
mse, lr 0.001	1.881	1.588	9.253	7.191	13.535	7.399	4.242	1.934	2.705	1.685
mse, lr 0.01	1.602	1.374	6.53	6.243	3.624	3.129	1.376	1.642	3.559	2.271
structured	1.24	0.958	2.144	2.251	1.083	0.985	0.885	1.345	1.811	1.199
state dependent	<b>0.196</b>	<b>0.142</b>	<b>0.152</b>	<b>0.174</b>	<b>0.136</b>	<b>0.661</b>	<b>0.156</b>	<b>0.849</b>	<b>0.237</b>	<b>0.311</b>
mlp	1.203	1.254	6.124	4.844	9.168	5.324	3.647	1.649	2.328	1.812

TABLE I: Online adaptation performance (MSE) of inverse-dynamics model on pick and place task. For each motion segment, results are averaged across 3 trials, and 5 random seeds per trial. The state-dependent loss is the performer across each motion trajectory both in terms of converging to the lowest error on an average as well as having the least amount of variance in terms of prediction performance.

the same Figure, we train the learned losses on data collected from a sine-motion task on real hardware. In the last row of Figure 6, we compare against an adaptive optimizer (Adam). We see that a SGD optimizer optimizing our state dependent loss outperforms an adaptive optimizer that optimizes an MSE loss, as well. Overall, we make 2 observations: 1) all learned losses outperform the standard MSE loss; 2) we see a clear improvement over the losses trained in simulation; most strikingly the state-dependent loss significantly outperforms all other losses. It is able to achieve low error predictions throughout all motions.

3) *Analysis*: Three separate conclusions emerge from the analysis of the results in Figure 6:

i) *Adaptation is needed*: The pretrained model (pre-trained on meta-train data) does not perform well on this test task. This shows that the data distribution is significantly different from the meta-train data distribution and adaptation is needed.

ii) *Sim-to-Real Transfer*: For most of the executed motions, models trained by optimizing the losses learned on the simulator data seem to perform better than models trained with MSE losses, especially when the same  $\alpha$  from meta-training (Algorithm 1) is used to train them. This indicates that the learned losses can be transferred from simulations to real data to some extent.

iii) *State-dependent losses adapt better*: Models optimized with learned losses perform consistently better than those optimized using the MSE loss at the same learning rate - 0.001. Specifically, the state-dependent loss outperforms all other losses (including the model trained with MSE loss at a higher learning rate). This can be seen in Table I, as state dependent loss has the lowest error on average and lowest variance in terms of prediction performance for each motion. Consequently, these results suggest that keeping track of the state allows our learned loss to better adapt the dynamics model (for example, in cases of stiction in our robot joints). Additionally, we observe that the evaluation curves for the state-dependent loss are much smoother and show less variance across multiple trajectories (which follow the same motion as we progress through training  $f_\theta$ ). This indicates that training a model with state-dependent loss on an online stream of data is significantly more stable than training with other learned-losses or MSE.

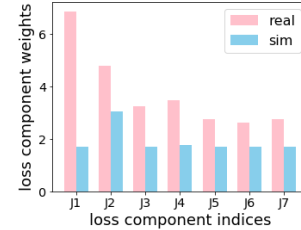


Fig. 7: This plot depicts the  $\phi$  values for *structured* loss trained on robot (real) and simulator (sim) data (loss updated in equation 1). The difference between  $\phi$ s for these data sets is likely because real data contains examples of non-ideal motion patterns with static friction and noise.

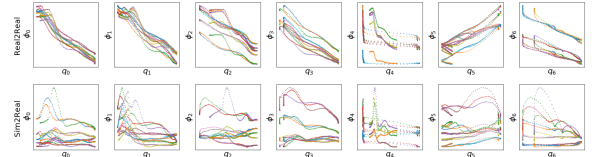


Fig. 8: We show scatter plots of the  $\phi$  against joint angles  $q$  for the *state-dependent* loss (equation 2) trained on robot (1st row) and simulator (2nd row) sine-motion data for multiple trajectories (differently colored) of the *reaching* motion on the robot. The correlation between  $\phi$  and  $q$  demonstrates how losses change as a function of the state.

### C. Analysis of Structured Losses

Our structured losses (*structured* and *state-dependent*) incorporate the notion that torques at some of the joints are more informative than others when measuring a distance between torque values and therefore learn to weigh each component of the loss function that corresponds to a torque dimension separately. We visualize the weights of the *structured* loss in Figure 7, both for the loss trained in simulation as well as the loss trained on hardware. Note that the distribution of weights differs between the simulation and hardware data. We postulate that the structured loss learned on real robot data captures some of the reality gap between hardware and simulation in its weight distribution, therefore adapting to what we might only see in a hardware experiment. It is possible that all joints have a higher weight for the real data because there exists some stiction on the joints, causing them to require higher torque values to move the same distance as in simulation. We also visualize the

weight of the first joint of the state-dependent loss (Figure 8) as a function of the joint state. We can clearly see that the weights on the joints vary as a function of the state. We believe that it is highly likely that keeping track of state captures more of these real world issues than structured loss does, allowing us to have even better performance.

## VI. CONCLUSION

In this work we contribute a framework for applying meta-learning to adaptation of inverse dynamics models and improve upon previous meta-learning loss [9] methods by adding structure and state dependency to the learned loss, resulting in better adaptation to online data. Our evaluation shows that all learned losses perform better than our baseline (MSE) on training an inverse dynamics model. However, specifically adding structure and state-dependency to the learned losses produces faster results with better (and more stable) adaptation to hardware data. This result provides experimental evidence for our intuition that a state-dependent loss can capture non-linear and unmodelled effects such as static friction.

We see several promising directions for further research involving the methods we have introduced in this work. Currently, due to safety concerns, the online adaptation experiments on the hardware were performed using data collected offline from the robot, but fed to the model sequentially, mimicking an actual online setting. In future, we plan on investigating methods to safely apply meta-learning directly on robots.

Another intriguing direction would be to investigate the *state-dependent* loss further in order to verify some of our hypotheses relating to its exceptional performance. For instance, we would like to demonstrate and visualize the inner workings of the *state-dependent* loss. In addition, we hope to explore how structured and state-dependent meta losses improve learning of other types of models for inverse dynamics - such as analytical models [30], [10]. Finally, to dive deeper into the analysis of our meta-learned losses, we intend to compare our loss-learning approach to other meta-learning methods such as [31], [32] to understand when a learned loss leads to better adaptation than a meta-trained model.

## REFERENCES

- [1] D. J. Herzfeld, P. A. Vaswani, M. K. Marko, and R. Shadmehr, "A memory of errors in sensorimotor learning." *Science*, 2014.
- [2] D. Nguyen-Tuong and J. Peters, "Model learning for robot control: a survey," *Cognitive processing*, vol. 12, no. 4, pp. 319–340, 2011.
- [3] D. Nguyen-Tuong, J. R. Peters, and M. Seeger, "Local gaussian process regression for real time online model learning," in *Advances in neural information processing systems*, 2009, pp. 1193–1200.
- [4] S. Schaal and C. G. Atkeson, "Constructive incremental learning from only local information," *Neural computation*, vol. 10, no. 8, 1998.
- [5] N. Ratliff, F. Meier, D. Kappler, and S. Schaal, "Doomed: Direct online optimization of modeling errors in dynamics," *Big data*, 2016.
- [6] O. Sigaud, C. Salaün, and V. Padois, "On-line regression algorithms for learning mechanical models of robots: a survey," *Robotics and Autonomous Systems*, vol. 59, no. 12, pp. 1115–1129, 2011.
- [7] F. Meier, D. Kappler, N. Ratliff, and S. Schaal, "Towards robust online inverse dynamics learning," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4034–4039.
- [8] A. Gijsberts and G. Metta, "Real-time model learning using incremental sparse spectrum gaussian process regression," *Neural networks*, vol. 41, pp. 59–69, 2013.
- [9] S. Bechtle, A. Molchanov, Y. Chebotar, E. Grefenstette, L. Righetti, G. Sukhatme, and F. Meier, "Meta-learning via learned loss," *arXiv preprint arXiv:1906.05374*, 2019.
- [10] C. G. Atkeson, C. H. An, and J. M. Hollerbach, "Estimation of inertial parameters of manipulator loads and links," *The International Journal of Robotics Research*, vol. 5, no. 3, pp. 101–119, 1986.
- [11] K. Hitzler, F. Meier, S. Schaal, and T. Asfour, "Learning and adaptation of inverse dynamics models: A comparison," in *Humanoids 2019*.
- [12] S. Vijayakumar and S. Schaal, "Local dimensionality reduction for locally weighted learning," in *Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97*. IEEE, 1997, pp. 220–225.
- [13] S. Schaal, C. G. Atkeson, and S. Vijayakumar, "Scalable techniques from nonparametric statistics for real-time robot learning," *Applied Intelligence*, vol. 17, no. 1, pp. 49–60, 2002.
- [14] R. Camoriano, S. Traversaro, L. Rosasco, G. Metta, and F. Nori, "Incremental semiparametric inverse dynamics learning," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 544–550.
- [15] F. Meier, P. Hennig, and S. Schaal, "Incremental local gaussian regression," in *Advances in Neural Information Processing Systems*, 2014, pp. 972–980.
- [16] L. Jamone, B. Damas, and J. Santos-Victor, "Incremental learning of context-dependent dynamic internal models for robot control," in *2014 IEEE international symposium on intelligent control (ISIC)*, 2014.
- [17] F. Meier and S. Schaal, "Drifting gaussian processes with varying neighborhood sizes for online model learning," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016.
- [18] D. Nguyen-Tuong and J. Peters, "Using model knowledge for learning inverse dynamics," in *2010 IEEE international conference on robotics and automation*. IEEE, 2010, pp. 2677–2682.
- [19] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," *arXiv e-prints:1703.03400*, 2017.
- [20] F. Meier, D. Kappler, and S. Schaal, "Online learning of a memory for learning rates," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2425–2432.
- [21] M. Andrychowicz, M. Denil, S. Gomez, M. W. Hoffman, D. Pfau, T. Schaul, and N. de Freitas, "Learning to learn by gradient descent by gradient descent," in *Advances In Neural Information Processing Systems*, 2016, pp. 3981–3989.
- [22] K. Li and J. Malik, "Learning to optimize," *arXiv preprint arXiv:1606.01885*, 2016.
- [23] R. Houthoofd, Y. Chen, P. Isola, B. Stadie, F. Wolski, O. J. Ho, and P. Abbeel, "Evolved policy gradients," in *Advances in Neural Information Processing Systems*, 2018, pp. 5400–5409.
- [24] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," *Artificial Intelligence Review*, vol. 18, no. 2, 2002.
- [25] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," *arXiv preprint arXiv:2004.05439*, 2020.
- [26] A. Nagabandi, C. Finn, and S. Levine, "Deep online learning via meta-learning: Continual adaptation for model-based rl," *arXiv preprint arXiv:1812.07671*, 2018.
- [27] E. Grefenstette, B. Amos, D. Yarats, P. M. Htut, A. Molchanov, F. Meier, D. Kiela, K. Cho, and S. Chintala, "Generalized inner loop meta-learning," *arXiv preprint arXiv:1910.01727*, 2019.
- [28] E. Coumans and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning," *GitHub*, 2016.
- [29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, pp. 8024–8035.
- [30] G. Sutanto, A. Wang, Y. Lin, M. Mukadam, G. Sukhatme, A. Rai, and F. Meier, "Encoding physical constraints in differentiable newton-euler algorithm," *arXiv preprint arXiv:2001.08861*, 2020.
- [31] A. Nagabandi, I. Clavera, S. Liu, R. S. Fearing, P. Abbeel, S. Levine, and C. Finn, "Learning to adapt in dynamic, real-world environments through meta-reinforcement learning," in *ICLR*, 2018.
- [32] R. Kaushik, T. Anne, and J.-B. Mouret, "Fast online adaptation in robotics through meta-learning embeddings of simulated priors," *arXiv preprint arXiv:2003.04663*, 2020.