

# Deep Probabilistic Movement Primitives with a Bayesian Aggregator

Michael Przystupa<sup>1</sup>, Faezeh Haghverdi<sup>1</sup>, Martin Jagersand<sup>1</sup>, Samuele Tosatto<sup>2</sup>

**Abstract**—Movement primitives are trainable parametric models that reproduce robotic movements starting from a limited set of demonstrations. Previous works proposed simple linear models that exhibited high sample efficiency and generalization power by allowing *temporal modulation* of movements (reproducing movements faster or slower), *blending* (merging two movements into one), *via-point conditioning* (constraining a movement to meet some particular via-points) and *context conditioning* (generation of movements based on an observed variable, e.g., position of an object). Previous works have proposed neural network-based motor primitive models, having demonstrated their capacity to perform tasks with some forms of input conditioning or time-modulation representations. However, there has not been a single unified deep movement primitive’s model proposed that is capable of all previous operations, limiting neural movement primitive’s potential applications. This paper proposes a deep movement primitive architecture that encodes all the operations above and uses a Bayesian context aggregator that allows a more sound context conditioning and blending. Our results demonstrate our approach can scale to reproduce complex motions on a larger variety of input choices compared to baselines while maintaining operations of linear movement primitives provide.

## I. INTRODUCTION

Learning from demonstration (LfD) is a technique that enables robots to acquire complex and adaptive skills by observing and reproducing human demonstrations [1], [20], [19], [28]. A key challenge in LfD is how to represent and generalize demonstrated motions in a robust and flexible way. Movement primitives (MPs) are a popular approach to address this challenge. MPs are parametric models that capture the essential features of a motion while allowing for variations and modifications according to different situations. The possibility of modifying the motion according to different needs is crucial to obtain an efficient mechanism to extrapolate new movements from scarce data.

This generalization has previously been achieved by using probabilistic models to represent motor skills. Treating motor skills as probabilistic models captures the uncertainty in demonstration data that deterministic models would otherwise assume are fixed. Probabilistic models further enable *variable conditioning* in which predictions can be refined through new observations. These allow movement primitives to adapt to new information during a motion.

Probabilistic movement primitives (ProMPs) were one of the first frameworks to use probabilistic modeling to learn motor skills from human demonstrations. ProMPs are

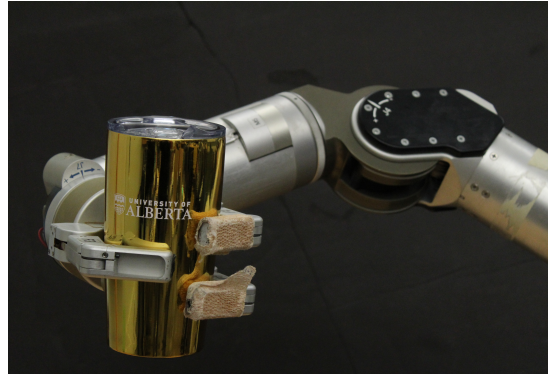


Fig. 1. Barret WAM preparing to perform rhythmic motions encoded with our deep probabilistic movement primitives to shake a Mojito.

composed of a linear-Gaussian model, providing high mathematical tractability and allowing various operations, such as *via-point* and *context conditioning*, *blending*, *temporal modulation*, and *rhythmic movements* [17], [18]. Via-point conditioning allows sampling motions that pass through some desired robot configurations, whereas context conditioning generates motions based on some external variables (e.g., the position of an object in the scene). Blending generates an unseen movement by interpolating two learned ones. Temporal modulation allows the modification of the velocity profile of a given movement. Rhythmic movements enable a continuous repetition of the learned movement.

However, Gaussian-linear models are limited to unimodal distributions and cannot represent rich, multimodal datasets or deal with high-dimensional variables like images [6], [24]. Recent approaches overcome these limitations with deep learning techniques. Some researchers have proposed autoregressive neural-networks-based movement primitives [15], [23], [22], [9], [16]. These models are *time-discrete*, which leads to performance degradation with high frequencies [3]. Other research proposes continuous-time approaches that use a deep model to generate the parameters of movement primitive models [3], [2], but the fundamental movements are still implemented with a linear model.

The pivotal work of Seker et al. [21] overcomes the limitations of previous work by proposing a continuous-time, non-linear representation of motor skills. These motor primitives utilizes Conditional Neural Processes (CNPs) [7], [8], an encoder-decoder model that allows aggregating multiple input-output pairs to form a latent representation of a function. In their framework, called Conditional Neural Movement Primitives (CNMPs), Seker et al. utilize time as

<sup>1</sup> University of Alberta, Canada. {przystupa, haghverdi, mj7}@ualberta.ca

<sup>2</sup> University of Innsbruck, Austria. samuele.tosatto@uibk.ac.at

the input variable and joint configurations as the output of CNPs. The aggregation of multiple time-joint-configuration pairs forms a latent representation of the desired motion. The latent representation is then used by a non-linear model to produce the continuous-time association between the two variables.

However, CNMPs have several limitations: 1) CNMPs use a deterministic latent variable  $\mathbf{z}$ . The stochasticity is only provided by a Gaussian output, which is well suited to capture a unimodal *aleatoric* uncertainty; but cannot represent multimodal distributions. 2) The authors do not discuss how to provide motion blending, temporal modulation, and rhythmic movements. 3) CNMPs can still produce trajectories that are far from the desired via-points.

In this paper, we aim to mitigate these weaknesses and propose a deep probabilistic movement framework that exposes all the operations of classic ProMPs while enabling working with high-dimensional variables and maintaining high precision. Our deep probabilistic movement primitives (DeepProMPs) replace the CNMP deterministic mean aggregator with a Bayesian aggregator (BA) which has been shown to have superior predictive performance [26]. We demonstrate this formulation provides a natural means to perform motion blending directly in the latent space. We assume our latent variable is independent when conditioned on either via-points or context variables. This enables flexibility when predicting motor skills and can be learned via auxiliary task learning. We further show how to implement both temporal modulation and rhythmic movement while maintaining high accuracy through state-of-the-art machine learning techniques, such as iterative optimization [27] to improve via-point conditioning accuracy, and utilization input representations similar to positional encoding [25].

In summary, our paper proposes a complete probabilistic deep-learning extension to ProMPs. Our model allows via-point and context conditioning, motion blending, temporal modulation, and rhythmic movements. Unlike ProMPs, our model can work with multimodal datasets and high-dimensional inputs. We exhaustively compare our model with ProMPs and CNMPs in virtual environments and on real robots. Our results suggest that compared to CNMP alternatives, our model is better at reconstructing demonstrations across variations of specified input data, including high-dimensional data like images. Even when CNMP variants are trained to mimic these aspects, we find our model is superior in motor skill prediction.

## II. PROBLEM STATEMENT

In this paper, we consider the problem of learning robotic movements from human demonstrations. A human demonstration can be summarized as a set of via-points  $\mathcal{A} = \{\mathbf{a}_i = (\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$ , where via-points  $\mathbf{a}_i$  indicate the configuration  $\mathbf{y}_i \in \mathbb{R}^d$  of the robot at the normalized time  $\mathbf{x}_i$ <sup>1</sup>. Furthermore, each demonstration is accompanied by a variable num-

<sup>1</sup>We denote the time variable  $\mathbf{x}$  with bold characters since rhythmic movements (Section III) have a time variable that is formed by two components.

ber  $m$  of *context variables*  $\mathcal{C} = \{\mathbf{c}_i\}_{i=1}^m$  that describe some features of the environment (e.g., the pose of an object in the scene) or some desired goals (e.g., the amount of liquid to be poured in a glass). MPs aim to learn the relation between the time variable  $\mathbf{x}$  and the joint configuration  $\mathbf{y}$  by imitating the user’s behavior in relation to the context variables  $\mathbf{c}$ . ProMPs provide a full probabilistic view on the matter where  $\mathbf{z}$ ,  $\mathbf{y}$  and  $\mathbf{c}$  are treated as stochastic variables [17], [18]. The probabilistic model allows both to sample movements similar to the one provided by the demonstration and to condition movements based on via-points and context variables, i.e.,  $p(\mathbf{y}|\mathbf{x}, \mathcal{A}, \mathcal{C}) = \int p(\mathbf{y}|\mathbf{z}, \mathbf{x})p(\mathbf{z}|\mathcal{A}, \mathcal{C})d\mathbf{z}$ . Usually, the number of conditioning variables  $n$  and  $m$  during deployment is much lower than the one in the dataset. ProMPs implement a linear-Gaussian model that allows mathematical tractability to solve the integrals in the closed form at the cost of lower expressivity. Due to their simple structure, ProMPs cannot represent complex data distributions. In the state-of-the-art literature, this limitation is overcome by using deep neural networks.

### A. Conditional Neural Movement Primitives

CNMPs aim to provide a neural-network architecture to represent movement primitives and allow via-point and context conditioning. In CNMPs, each trajectory is associated with a latent variable  $\mathbf{z}$  that represents the motion. CNMPs are, therefore, composed of two main modules: a deterministic mean aggregation encoder  $\mathbf{z} = \frac{1}{n}(\sum_{i=1}^n \phi(\mathbf{a}_i, \mathbf{c}_i))$ , where  $\phi$  is a neural network, and a movement generator (decoder)  $p(\mathbf{y}|\mathbf{z}, \mathbf{x})$ . The encoder associates a set of via-points  $\mathbf{a}_1, \dots, \mathbf{a}_n$  and context variables  $\mathbf{c}_1, \dots, \mathbf{c}_m$  with a deterministic latent representation of the movement  $\mathbf{z}$ . The encoder takes as input the time variable  $\mathbf{x}$  and the latent motion representation  $\mathbf{z}$  and outputs a Gaussian distribution over robot configurations  $\mathbf{y}$ . CNMPs can be trained with classic gradient descent techniques and allow conditioning with high-dimensional variables (e.g.,  $\{\mathbf{c}_i\}$  can be images). However, they can only learn unimodal uncertainty since  $\mathbf{y}$  is Gaussian.

### B. Bayesian Context Aggregator

Bayesian aggregation is a probabilistic aggregation technique that addresses the mean aggregator’s inability to represent uncertainty. For example, the latent variable  $\mathbf{z}$  should have a higher variance when only a few samples are provided, while it should have less variance when more samples are provided. With the mean aggregator, CNMPs postpone the problem of quantifying the uncertainty to the last layer with the neural network, which is unaware of how many conditioning variables have been presented. Therefore it cannot make a good prediction of the epistemic uncertainty. The Bayesian aggregator overcomes this issue by applying Bayesian inference directly to the latent variable.

We assume a prior distribution  $p_0(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mu_0, \sigma_0^2)$ , and

<sup>2</sup>The mean  $\mu_0$  has the same dimension of the vector  $\mathbf{z}$  and  $\sigma_0^2$  as a vector of variances (i.e., all dimensions are independent). We keep this convention through the rest of the paper.

a model  $p(\mathbf{a}_i|\mathbf{z}) = \mathcal{N}(\phi(\mathbf{a}_i)|\mathbf{z}, \sigma^2(\mathbf{a}))$  where  $\phi$  is a deterministic mapping that project  $\mathbf{a}$  in the same dimension of  $\mathbf{z}$ . BA uses the Bayesian rule to infer  $p_{i+1}(\mathbf{z}) = p(\mathbf{z}|\mathbf{a}_{i+1}) \propto p(\mathbf{a}_{i+1}|\mathbf{z})p_i(\mathbf{z})$ . This process results in  $p(\mathbf{z}|\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = p_0(\mathbf{z}) \prod_{i=1}^n p(\mathbf{a}_i|\mathbf{z})$ , which is computable in closed-form for normal distributions. Notice that the latent distribution can be equivalently rewritten as  $(\mathbf{z}|\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = p_0(\mathbf{z}) \prod_{i=1}^n p(\mathbf{z}|\mathbf{a}_i)$ , since, due to the symmetry of the Gaussian distribution,  $\mathcal{N}(\phi(\mathbf{a}_i)|\mathbf{z}, \sigma^2(\mathbf{a})) = \mathcal{N}(\mathbf{z}|\phi(\mathbf{a}_i), \sigma^2(\mathbf{a}))$ . Unlike the classic mean estimator, BA builds a probabilistic model of the latent variable based on the conditioning variable. In particular, conditioning variables that carry less information will have a lower impact on predicting the latent variable without lowering its uncertainty. In comparison, context variables that carry more information will tend to decrease model uncertainty and have a higher impact on the prediction. The latent space induced by BA is processed by subsequent nonlinear mappings and is trainable via variational inference.

### III. MODEL ARCHITECTURE AND TRAINING

An essential function of ProMPs is to learn a model of the distribution of motions (trajectories) shown by a human demonstrator. Once trained, the model can be used to sample comparable trajectories to the demonstrations. However, in most applications, one does not want only to replicate the distribution of observed movements but condition them on certain variables such as via-points and context variables.

The primary idea of this work is to use a Bayesian aggregator to model the latent space and to use variational inference to train the model similarly to variational autoencoders (VAEs) [13], [14]. This enables DeepProMPs to exploit the probabilistic model of the latent space to produce a multimodal profile of the epistemic uncertainty and learn the relation between a variable number of conditioning variables and movements.

#### A. Model Architecture

Our model is mainly composed of two elements: the *encoder*, which consists of a Bayesian aggregator that processes a variable number of via-points and context variables and builds the latent distribution of  $\mathbf{z}$ , and the *decoder*, a neural network that implements  $p(\mathbf{y}|\mathbf{z}, \mathbf{x})$ , where  $\mathbf{x}$  is a normalized time (or phase),  $\mathbf{z}$  is the latent representation of the motion, and  $\mathbf{y}$  is the desired robot configuration. Therefore, the encoder processes two different typologies of inputs: the set of via-points, denoted with  $\mathcal{A} = \{\mathbf{a}_i\}_{i=1}^n$  and the set of context variables, denoted with  $\mathcal{C} = \{\mathbf{c}_i\}_{i=1}^m$ . As described in Section II-B, the latent distribution is

$$q(\mathbf{z}|\mathcal{A}, \mathcal{C}) \propto p_0(\mathbf{z}) \prod_{i=1}^n q(\mathbf{z}|\mathbf{a}_i) \prod_{i=1}^m q(\mathbf{z}|\mathbf{c}_i), \quad (1)$$

where  $p_0(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$ ,  $q(\mathbf{z}|\mathbf{a}_i) = \mathcal{N}(\mathbf{z}|\mu_a(\mathbf{a}_i), \sigma_a^2(\mathbf{a}_i))$ , and  $q(\mathbf{z}|\mathbf{c}_i) = \mathcal{N}(\mathbf{z}|\mu_c(\mathbf{c}_i), \sigma_c^2(\mathbf{c}_i))$  where  $\mu_a, \mu_c, \sigma_a, \sigma_c$  are nonlinear projections encoded by neural networks. In

practice, the mean and the variance of the latent distribution are computed by:

$$\begin{aligned} \mu_z(\mathcal{A}, \mathcal{C}) &= \sigma_z^2(\mathcal{A}, \mathcal{C}) \left( \sum_{i=1}^n \frac{\mu_a(\mathbf{a}_i)}{\sigma_a^2(\mathbf{a}_i)} + \sum_{i=1}^m \frac{\mu_c(\mathbf{c}_i)}{\sigma_c^2(\mathbf{c}_i)} \right), \\ \sigma_z^2(\mathcal{A}, \mathcal{C}) &= \left( 1 + \sum_{i=1}^n \sigma_a^2(\mathbf{a}_i)^{-1} + \sum_{i=1}^m \sigma_c^2(\mathbf{c}_i)^{-1} \right)^{-1}. \end{aligned}$$

Notice that we assume a unique pair of functions (i.e., neural networks) to process the mean  $\mu_c$  and the variance  $\sigma_c^2$  associated with the context variables. However, in some settings, we want to have different kinds of context variables (e.g., raw images, and low-dimensional poses extracted from a motion capture), and we, therefore, employ different neural networks for the different types of input data. The decoder  $p(\mathbf{y}|\mathbf{z}, \mathbf{x})$  is implemented using a Gaussian distribution, i.e.,

$$p(\mathbf{y}|\mathbf{z}, \mathbf{x}) = \mathcal{N}(\mathbf{y}|\mu_y(\mathbf{z}, \mathbf{x}), \sigma_y^2),$$

where  $\mu_y$  is a nonlinear projection encoded with a neural network and  $\sigma_y$  is an hyper-parameter. At this point, the probability of a robotic configuration  $\mathbf{y}$  at time  $\mathbf{x}$  given a set of via-points and conditioning variables is

$$p(\mathbf{y}|\mathbf{x}, \mathcal{A}, \mathcal{C}) = \int p(\mathbf{y}|\mathbf{x}, \mathbf{z})p(\mathbf{z}|\mathcal{A}, \mathcal{C})d\mathbf{z}. \quad (2)$$

Unfortunately, (2) is not solvable in closed form, and we need to resort to *variational inference* to train the model.

#### B. Training the Model via Variational Inference

Similar to classical variational autoencoders (VAEs) [13], [14], our model is composed of three entities: model likelihood  $p(\mathbf{y}|\mathbf{z}, \mathbf{x})$ , variational posterior  $q(\mathbf{z}|\mathcal{A}, \mathcal{C})$ , and prior distribution  $p_0(\mathbf{z})$ . The model likelihood can be rewritten using the reparameterization trick,

$$\mathbf{y} = \mu_y(\mathbf{z}, \mathbf{x}) + \epsilon\sigma_y \quad \text{with } \epsilon \sim \mathcal{N}(0, \mathbf{I}), \quad (3)$$

to allow gradient training. The evidence lowerbound (ELBO) is

$$\mathcal{L}(\mathcal{A}, \mathcal{C}) = \mathbb{E}_{\mathbf{z} \sim q(\cdot|\mathcal{A}, \mathcal{C})} \left[ \sum_{i=1}^n \log p(\mathbf{y}_i|\mathbf{z}, x_i) - \log \frac{q(\mathbf{z}|\mathcal{A}, \mathcal{C})}{p_0(\mathbf{z})} \right].$$

Plugging in the definition in 1, we rewrite the ELBO as

$$\begin{aligned} \mathcal{L}(\mathcal{A}, \mathcal{C}) &= \mathbb{E}_{\mathbf{z} \sim q(\cdot|\mathcal{A}, \mathcal{C})} \left[ \sum_{i=1}^n \log p(\mathbf{y}_i|\mathbf{z}, x_i) \right] \\ &\quad - \mathbb{E}_{\mathbf{z} \sim q(\cdot|\mathcal{A}, \mathcal{C})} \left[ \sum_{i=1}^n \log \frac{q(\mathbf{z}|\mathbf{a}_i)}{p_0(\mathbf{z})} + \sum_{i=1}^m \log \frac{q(\mathbf{z}|\mathbf{c}_i)}{p_0(\mathbf{z})} \right]. \end{aligned}$$

Notice that the additional KL divergence between the variational posteriors  $q(\mathbf{z}|\mathbf{a}_i)$ ,  $q(\mathbf{z}|\mathbf{c}_i)$  and the prior  $p_0(\mathbf{z})$  are meant to satisfy the prior assumption, i.e.,  $\mathbb{E}_{\mathbf{a}}[p(\mathbf{z}|\mathbf{a}_i)] = \mathbb{E}_{\mathbf{c}}[p(\mathbf{z}|\mathbf{c}_i)] = p_0(\mathbf{z})$ . To train the neural network we maximize the ELBO with respect to the parameters of  $\mu_a, \sigma_a, \mu_c, \sigma_c$ , and  $\mu_y$ . Variational inference suffers from problems like posterior collapse [5] and aliasing from a mismatch of the posterior and prior distributions [4]. The former negatively

impacts our movement primitives’ ability to replicate the data accurately due to the over-regularizing effects of the KL divergence. The latter can lead to the model generating potentially undesirable demonstrations outside the training distribution. We address the former problem with KL-annealing and adjusting the weighting of the KL divergence. The latter can be addressed with learnable prior distributions, but we do not consider it in this work.

### C. Operations

**Generation of Movements.** DeepProMPs are a deep-learning variant of ProMPs that can depict complex movement distributions. To build a trajectory it is sufficient to sample  $\mathbf{z}$  from the prior distribution  $\mathbf{z} \sim p_0(\mathbf{z})$  and then use the model in Equation 3 to predict  $\mathbf{y}$  at the appropriate normalized time  $\mathbf{x}$ ,  $\mathbf{y} = \mu_y(\mathbf{z}, \mathbf{x})$  without including the noise  $\epsilon$ . Note that to get a smooth motion, we do not resample  $\mathbf{z}$  during trajectory generation. The variable  $\mathbf{z}$  remains constant over the life of a trajectory.

**Via-Points and Context Conditioning.** Once our model is learned, we want to use it to predict new trajectories that pass through a set of via-points  $\tilde{\mathcal{A}} \equiv \{\tilde{\mathbf{a}}_i\}$  and is conditioned on a set of external variables  $\tilde{\mathcal{C}}$ . Notice that one of the two sets can be empty (if they are both empty, we are using our model purely as generative). We use the variational posterior as in Equation 1 to sample  $\mathbf{z}$ , and then we use Equation 3 to generate the whole trajectory. However, occasionally, it might be useful to weight some via-points or some external variables differently. Taking inspiration from ProMPs, we rewrite the variational posterior, including importance weights  $\omega_i^A, \omega_i^C$ ,

$$q(\mathbf{z}|\tilde{\mathcal{A}}, \tilde{\mathcal{C}}, \omega^A, \omega^C) = \prod_{i=1}^n q_A(\mathbf{z}|\tilde{\mathbf{a}}_i)^{\omega_i^A} \prod_{i=1}^m q_C(\mathbf{z}|\tilde{\mathbf{c}}_i)^{\omega_i^C} \quad (4)$$

obtaining a Gaussian distribution again. This reweighting can be useful as the size of via-points conditioning and context conditioning are usually unbalanced, and the reweighting can help correct this bias. Notice that this importance reweighting is not possible with the mean aggregator.

The generated motions can arbitrarily violate the desired via-points. To mitigate this issue, we propose to perform a further optimization stage to adjust the distribution of generated motions to minimize the error for specified via-points. This optimization stage, carried out at deployment time, minimizes the distance between the generated trajectories and the target via-points while avoiding variance collapsing, i.e.,

$$\min_{\tilde{\mu}_z, \tilde{\sigma}_z} \sum_{i=1}^n \left\| \mathbf{y}_i - \frac{1}{k} \sum_{j=1}^k \mu_y(\mathbf{z}_j, \mathbf{x}_i) \right\|_2^2 + \|\sigma^* - \tilde{\sigma}_z\|_2^2, \quad (5)$$

where  $\mathbf{z}_{j=1, \dots, k} \sim \mathcal{N}(\tilde{\mu}_z, \tilde{\sigma}_z)$ ,  $k$  is the number of Monte-Carlo samples and  $\sigma^*$  is the target variance. We solve (5) by improving the initial solution  $\tilde{\mu}_z = \mu_z(\mathcal{A})$  and  $\tilde{\sigma}_z = \sigma_z(\mathcal{A})$  via gradient descent (Fig. 2). Further details are in Appendix.

**Blending.** As proposed in ProMPs, the blending of motions

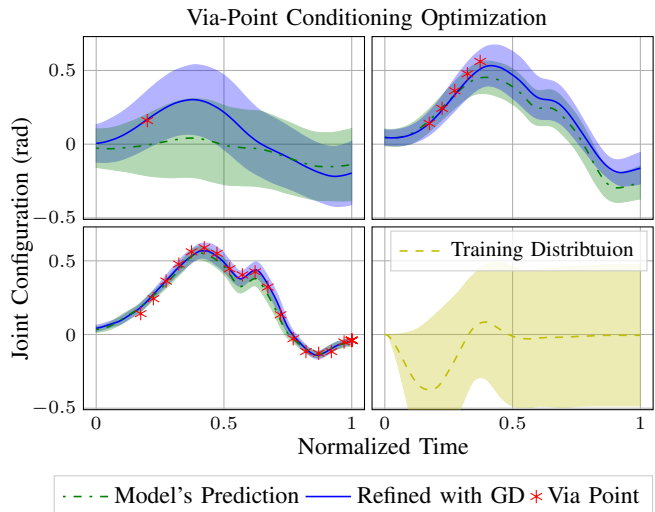


Fig. 2. Results of the deployment-time via-point error minimization on one of the joints from our real-robot data. In green is the predicted distribution from our model; in blue is the distribution refined with gradient descent. Benefits are more pronounced with fewer via-points, while more via-points improve the overall quality of the prediction. The bottom-right plot shows the distribution of the training data.

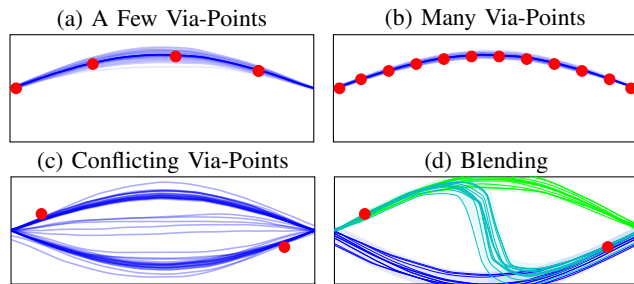


Fig. 3. DeepProMPs used in different situations. (a) Distribution of trajectories using 3 via-points conditioning. (b) With more via-points, the variance of the distribution decreases. (c) With inconsistent via-points, the model chooses to violate one of them staying close to the given dataset and avoiding unseen behavior. Note the ability of our model to generate a bi-modal distribution. (d) Generalization can be enhanced using blending.

consists of smoothly transitioning from one movement to another. Consider a time-varying weight  $\omega(\mathbf{x}) \in [0, 1]$  and two movement distributions  $q(\mathbf{z}|\mathcal{A}_1, \mathcal{C}_1)$  and  $q(\mathbf{z}|\mathcal{A}_2, \mathcal{C}_2)$ ; we derive the blending of the two movements

$$q_b(\mathbf{z}|\mathcal{A}_1, \mathcal{A}_2, \mathcal{C}_1, \mathcal{C}_2, \mathbf{x}) = q(\mathbf{z}|\mathcal{A}_1, \mathcal{C}_1)^{\omega(\mathbf{x})} q(\mathbf{z}|\mathcal{A}_2, \mathcal{C}_2)^{1-\omega(\mathbf{x})}. \quad (6)$$

using the exponential weighting introduced by ProMPs. Observe that the distribution is now dependent on time. This appears to be a problem, as sampling  $\mathbf{z}$  at different periods  $\mathbf{x}$  will result in a jerky motion. Using the reparametrization trick, however, we can resolve this issue by considering the mean  $\mu_b(\mathcal{A}_1, \mathcal{A}_2, \mathcal{C}_1, \mathcal{C}_2, \mathbf{x})$  and standard deviation  $\sigma_b(\mathcal{A}_1, \mathcal{A}_2, \mathcal{C}_1, \mathcal{C}_2, \mathbf{x})$  of the blended latent representation of the motion (6). By first sampling a standard noise  $\epsilon \sim \mathcal{N}(0, I)$  and then computing

$$\mathbf{z}(\mathbf{x}) = \mu_b(\mathcal{A}_1, \mathcal{A}_2, \mathcal{C}_1, \mathcal{C}_2, \mathbf{x}) + \epsilon \cdot \sigma_b(\mathcal{A}_1, \mathcal{A}_2, \mathcal{C}_1, \mathcal{C}_2, \mathbf{x}),$$

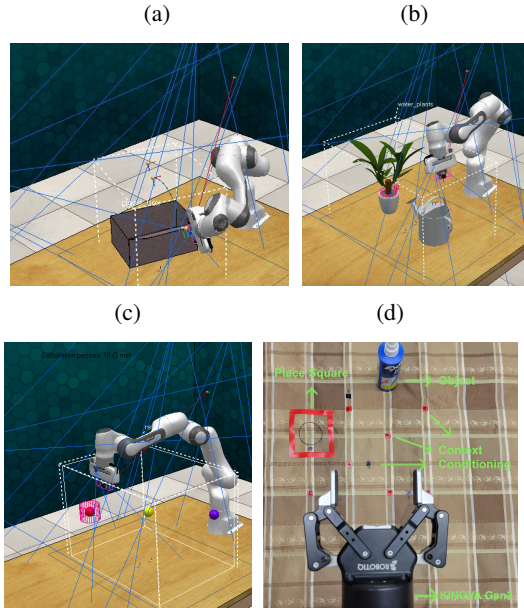


Fig. 4. (a), (b), (c): Close box, pour water, and reach from RL Bench. (d) A top view of our testing setup: The robot should grab the object and place in the designated square. Positions are encoded with 2D context variables.

where  $\cdot$  is the element-wise product, we obtain a smooth transition of  $\mathbf{z}$ .

**Time Modulation.** Time modulation is a crucial component for movement primitives. During training, time is always normalised between 0 and 1, therefore, a time modulation function  $\tau(t) : [0, T] \rightarrow [0, 1]$  can be used to translate the real-time  $t \in [0, T]$  to the *phase space*  $\mathbf{x} \in [0, 1]$ . In our implementations, we employ the linear time modulation  $\tau = t/T$ , where  $T$  is the whole duration of the movement. This linear modulation provides a straightforward method for achieving proportional velocity profiles with faster or slower movement. In general, a monotonic  $\tau$  can arbitrarily accelerate or decelerate a movement during its execution. If  $\tau$  is non-monotonic, it can *revert* the desired motions.

**Rhythmic Movements.** Rhythmic movements can be seen as a segment of movement that can be repeated indefinitely many times. Rhythmic movements require these segments to have the same initial and endpoint so that their repetition does not cause jumps. To obtain this property, the time-modulation function must have the same position and velocity at both  $t = 0$  and  $t = T$ . This effect can be obtained by using the phase  $\mathbf{x} = [\sin(2\pi t/T), \cos(2\pi t/T)]^\top$  both during training and deployment. We note this representation can be viewed as positional encoding where we use the behavior of the sinusoidal functions [25].

#### IV. EMPIRICAL ANALYSIS

In this section, we conduct several experiments to compare DeepProMPs against alternative motor primitive models. We compare each model’s reconstruction capabilities on demonstrations provided from several simulation tasks and a real robotic problem. We choose the Mean Square Error as our metric of choice to highlight how well we can

recover demonstrations depending on the conditioning. We also include experiments demonstrating the time-modulation formulation for cyclical tasks on a real robot.

We choose ProMPs, CNMPs, and VAE-CNMPs as our baselines. We include ProMPs because it is the foundational framework that motivates our work. CNMP can be viewed as a version of our model with fixed variance in the latent distribution, but that predicts output variance  $\sigma_y(\mathbf{z}, \mathbf{x})$  as done in previous work [21]. VAE-CNMP is a variation of our model without Bayesian aggregation, the via-point, and context variable independence assumption, but still uses the Isotropic Gaussian as the prior in the KL divergence regularization. For both CNMPs and VAE-CNMPs, we use a single encoder that concatenates all inputs (via points and context variables) into a single input to the encoder. We also train a version of CNMPs and VAE-CNMPs with zero-padding to compare whether or not these models are as capable of the optionality of inputs as DeeProMPs. For example, to simulate inputting only via-points with our baselines, the variational encoder’s input would be  $q(\mathbf{z}|\mathbf{a}, \mathbf{c}^0)$ , where  $\mathbf{c}_0 = \vec{0}$  or context only as  $q(\mathbf{z}|\mathbf{a}^0, \mathbf{c})$  where  $\mathbf{a}^0 = \vec{0}$  for context only inputs. We refer to these versions as CNMPs (Indep) and VAE-CNMPs (Indep) in our results.

We conduct experiments with demonstrations collected in three simulated tasks from RL Bench [11], and two sets of demonstrations performed on physical robotic manipulators. The three simulation tasks we use are: (1) *Reach*, a task where the manipulator goes to a designated target, (2) *Close Box*, a task where the manipulator closes a container, and (3) *Pour Water*, a task where the robot must pick up and pour the water from a container. Our real robot experiments include demonstrations of a pick-and-place task where a Kinova Gen-3 lite moves a designated object to a specified location. We refer to the Kinova pick-and-place task as the *Kinova* task in this section. We use the previous four tasks (three simulation and one real robot) to compare reconstruction performance. The second robot task uses a Barrett WAM<sup>®</sup> Arm to shake a container which we use to verify training and execution of cyclical tasks. The set-up for the latter simulation tasks and both real robot tasks are featured in Figure 4.

In our reconstruction experiments, we vary the number of trajectories used in each experiment. Our choices in the simulation task were motivated by preliminary experiment results and the complexity of the task. In the real robotics experiments, we used kinesthetic demonstration to collect the data. Respectively, we use 500 training examples for the Reach task, 1000 demonstrations for the Close Box and Pour Water task, and 100 demonstrations for the Kinova task. For all simulation tasks, we select the best models using a set of 100 validation examples and report results on 200 test examples. In the Kinova experiments, we report the best validation results using 10 demonstrations for all models.

For each reconstruction experiment, we train all models to take via-points, low-dimensional context variables, and images as inputs. The dimension of the low-dimensional context varies for each task. In the Reach task, this is the

target location (3 dimensions); in Close Box, this includes the angle of the box’s lid and box location (54 dimensions); for Water Plants, this is the location of the watering can and object to be watered (84 dimensions). In the Kinova experiments, this is the normalized pixel location of objects (2 dimensions) and one-hot encoding for the sub-task (pick or place). In the simulation tasks, each trajectory is associated with five images of the scene. We always use these images together at deployment. For our Kinova data, we take a picture of the object to be moved and augment this with three additional pictures that contain markers on the desired object. We found that this mitigated issues with the limited training data used. Further investigation in addressing this issue for deep motor primitives is potential future work, but out-of-scope of this paper which focuses on proposing the DeepProMPs framework generally.

**Task Reconstruction.** We evaluate ProMPs, DeepProMPs, CNMPs, VAE-CNMPs, CNMP (Indep), and VAE-CNMP (Indep). We used the Adam optimizer [12] with all models using a learning rate of  $1e-4$ . We train each model for 8000 epochs in our simulation experiments and 10000 epochs for our Kinova experiments, while always using mini-batches of size 16. All models use two hidden layer multi-layer perceptrons with ReLU activations for vector inputs (via-points and low-dimensional context variables). Each layer has 128 neurons. All models have a single decoder with two hidden layers each with 128 neurons that take in the latent variable and time-modulation variable as input. For images, we use a Resnet-50 architecture [10] to embed images in the latent space. We aggregate the representation for all images associated with a trajectory using a mean operation. Both CNMP and VAE-CNMP concatenate all context variable types to each via point as inputs. When using DeepProMP, all encoders share a single affine head to produce latent distribution parameters. We use latent dimensions of 16 for Reach and Kinova and 32 for Close Box and Pour water for all models. During training, we sub-sample combinations of input data for each model. For DeepProMP, CNMP (Indep), and VAE-CNMP (Indep) this includes an added step of choosing combinations of via-point, low-dimensional, and image context variables as inputs. For example, training with via-points and images on one update and via-points only on another. Both CNMP and VAE-CNMP models always receive a via-point and context variable combination as input because of the concatenation operation. We report results using models with the best validation error for each metric reported and train 5 of each model which results are averaged over. The one exception is ProMP, which has a deterministic solution and we instead bootstrap results to average results.

We report the average performances with radar plots in Figure 5. Smaller circles imply better performance across all metrics. We show the log-scale Mean Square Error of reconstructing demonstrations given different combinations of inputs. For CNMP and VAE-CNMP variants we achieve this with zero padding as previously described to mask out the excluded inputs. *Via Point* refers to reconstructing the trajectory without any context variables, *Image* uses only

images, *Low Dim* uses only the low dimensional context variable, and *Image + Low* uses both context variables. We also include *Aggregate* which is the average performance across the previous four combinations of inputs. We note that for each context variable input, we include the initial robot position as well.

In via-point conditioning, we see that ProMP is superior across tasks. This result is expected because ProMP’s analytic conditioning preserves all the information in the trajectory. Due to compressing the representation in a latent space, all deep models lose some information reconstructing the model. We note that results could be improved with our deployment optimization technique which was not used in these experiments.

Interestingly, we see that the zero padding training is crucial for improving baselines’ ability to do via-point conditioning across tasks but is still worse than DeepProMPs. This suggests that modularizing encoder representations to data types is better than using a single module. One advantage DeepProMPs has is great capacity with independent encoders for each data type. We trained several models with double the neurons reported in the experiments for both independent CNMP variants (256 neurons vs 128 neurons per layer), but these yielded inconsistent performance gains and losses across metrics despite having a comparable number of parameters. This is a sensible result because all models are bottle-necked by the dimensions of the latent space despite the size of the encoders.

Our results suggest that despite the independence assumptions in DeepProMP, the model provides competitive results with different input combinations or else beats other methods. We surprisingly find ProMP produces good results on low dimension reconstruction across tasks but is not as easy to extend to image data. We find that despite the independence assumption, our model still performs comparably to CNMP and VAE-CNMP, which are trained to always see all context information during training. In certain cases, the zero-pad training seems to worsen performance in scenarios we would not expect for baseline methods (see Reach results using only images). Our conclusion is that even when baselines are trained with techniques similar to DeepProMPs they come with trade-offs across conditioning choices whereas DeepProMPs do not.

**Real-Robot Make Mojito.** This experiment aims to highlight the ability of DeepProMPs to learn and replay cyclical behaviors. To this end, we collect a small data set of cyclical shaking motions and train DeepProMPs using the sinusoidal phase modulation proposed in Section III. In Figure 6, we report the most salient dimension of the learned motion, corresponding to the robotic manipulator’s second joint. Our model generalized the provided demonstration to perform it in a smooth, cyclic pattern. We find that our model realizes a smooth transition even if the gap between the demonstration’s starting point and the endpoint is large.

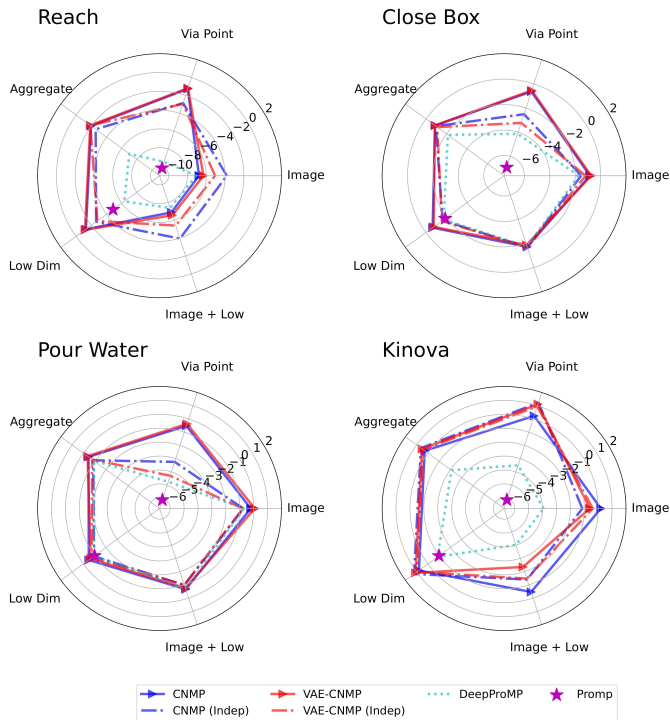


Fig. 5. Radar plots comparing reconstruction performance across different motor primitive models. Smaller circles are indicative of better performance across potential data types. We consider conditioning on images, low-dimensional context variables, the combination of both, the full trajectory as via points and the average across the four former types of inputs. Measurements are in log scale of the mean square error.

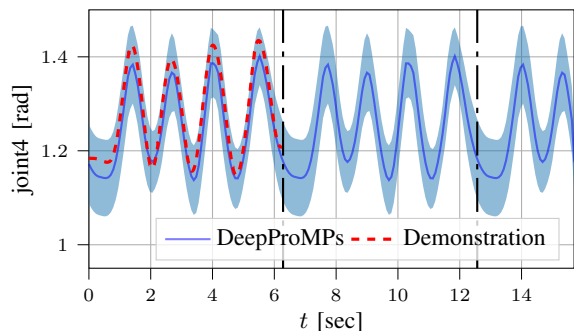


Fig. 6. Demonstrated motion and learned cyclical behavior.

## V. LIMITATIONS

Although DeepProMPs is a promising framework, they do not come without limitations and trade-offs. Deep neural networks and variational inferences add sources of approximation, which were not present in original ProMPs. Many quantities (like via-point and context conditioning) that cannot be found in closed form require sub-sampling training schemes before deployment to achieve these properties. Using an approximated posterior introduces a bias in the parameter estimation. However, this is often the price for obtaining a more flexible tool. As we chose to use an isotropic Gaussian for our latent distributions, covariant relations between latent dimensions may not be captured by our model. This could limit the expressiveness of our

learned representations. Our model also requires Monte-Carlo sampling in order to estimate the trajectory distribution statistics, which can slow inference time. This Monte Carlo sampling is used in our proposed iterative optimization to improve via-point conditioning. There is potential room to study alternative optimization approaches to perform this trajectory post-processing step.

## VI. CONCLUSION

This paper proposes Deep Probabilistic Motor Primitives as a deep learning variant of ProMPs. Our model is capable of both via-point and context conditioning independent of each. DeepProMPs are more robust to this feature compared to the baseline method trained with zero padding to achieve the same behavior. Our model can also blend motor skills together because of the Bayesian aggregation we incorporated in our model. We demonstrated DeepProMPs' capability of rhythmic motion modulation, which has otherwise been ignored in previous works. Our work is a step towards improving deep motor primitive models for applications in robotics. Future work could include incorporating other complex data types in deep motor primitives like text, consider alternative latent distribution choices to the Gaussian distribution in our Bayesian aggregator, and deploying DeepProMPs in downstream robotic applications.

## REFERENCES

- [1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A Survey of Robot Learning from Demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.
- [2] S. Bahl, A. Gupta, and D. Pathak, "Hierarchical Neural Dynamic Policies," *arXiv preprint arXiv:2107.05627*, 2021.
- [3] S. Bahl, M. Mukadam, A. Gupta, and D. Pathak, "Neural Dynamic Policies for End-to-End Sensorimotor Learning," *arXiv preprint arXiv:2012.02788*, 2020.
- [4] M. Bauer and A. Mnih, "Resampled priors for variational autoencoders," in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and M. Sugiyama, Eds., vol. 89. PMLR, 16–18 Apr 2019, pp. 66–75. [Online]. Available: <https://proceedings.mlr.press/v89/bauer19a.html>
- [5] S. R. Bowman, L. Vilnis, O. Vinyals, A. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," in *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 10–21. [Online]. Available: <https://aclanthology.org/K16-1002>
- [6] A. Colomé and C. Torras, "Dimensionality Reduction in Learning Gaussian Mixture Models of Movement Primitives for Contextualized Action Selection and Adaptation," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3922–3929, 2018, publisher: IEEE.
- [7] M. Garnelo, D. Rosenbaum, C. Maddison, T. Ramalho, D. Saxton, M. Shanahan, Y. W. Teh, D. Rezende, and S. A. Eslami, "Conditional Neural Processes," in *International Conference on Machine Learning*. PMLR, 2018, pp. 1704–1713.
- [8] M. Garnelo, J. Schwarz, D. Rosenbaum, F. Viola, D. J. Rezende, S. Eslami, and Y. W. Teh, "Neural Processes," *arXiv preprint arXiv:1807.01622*, 2018.
- [9] A. Gupta, V. Kumar, C. Lynch, S. Levine, and K. Hausman, "Relay Policy Learning: Solving Long-Horizon Tasks via Imitation and Reinforcement Learning," in *Conference on Robot Learning*. PMLR, 2020, pp. 1025–1037.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: <http://arxiv.org/abs/1512.03385>

- [11] S. James, Z. Ma, D. R. Arrojo, and A. J. Davison, “RLBench: The Robot Learning Benchmark & Learning Environment,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3019–3026, 2020, publisher: IEEE.
- [12] D. P. Kingma and J. Ba, “ADAM: A Method for Stochastic Optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [13] D. P. Kingma and M. Welling, “Stochastic Gradient VB and the Variational Auto-Encoder,” in *Second International Conference on Learning Representations, ICLR*, vol. 19, 2014, p. 121.
- [14] —, “An Introduction to Variational Autoencoders,” *arXiv preprint arXiv:1906.02691*, 2019.
- [15] T. Kipf, Y. Li, H. Dai, V. Zambaldi, A. Sanchez-Gonzalez, E. Grefenstette, P. Kohli, and P. Battaglia, “CompILE: Compositional Imitation Learning and Execution,” in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, June 2019, pp. 3418–3428. [Online]. Available: <https://proceedings.mlr.press/v97/kipf19a.html>
- [16] M. Noseworthy, R. Paul, S. Roy, D. Park, and N. Roy, “Task-Conditioned Variational Autoencoders for Learning Movement Primitives,” in *Conference on Robot Learning*. PMLR, 2020, pp. 933–944.
- [17] A. Paraschos, C. Daniel, J. Peters, and G. Neumann, “Probabilistic Movement Primitives,” in *Advances in Neural Information Processing Systems (NIPS)*. mit press, 2013.
- [18] —, “Using Probabilistic Movement Primitives in Robotics,” *Autonomous Robots (AURO)*, no. 3, pp. 529–551, 2018.
- [19] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, “Recent Advances in Robot Learning from Demonstration,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 297–330, 2020, publisher: Annual Reviews.
- [20] S. Schaal, “Is Imitation Learning the Route to Humanoid Robots?” *Trends in cognitive sciences*, vol. 3, no. 6, pp. 233–242, 1999.
- [21] M. Y. Seker, M. Imre, J. H. Piater, and E. Ugur, “Conditional Neural Movement Primitives,” in *Robotics: Science and Systems*, vol. 10, 2019.
- [22] T. Shankar and A. Gupta, “Learning Robot Skills with Temporal Variational Inference,” in *International Conference on Machine Learning*. PMLR, 2020, pp. 8624–8633.
- [23] M. Sharma, A. Sharma, N. Rhinehart, and K. M. Kitani, “Directed-Info GAIL: Learning Hierarchical Policies from Unsegmented Demonstrations using Directed Information,” in *International Conference on Learning Representations*, 2018.
- [24] S. Tosatto, G. Chalvatzaki, and J. Peters, “Contextual Latent-Movements Off-Policy Optimization for Robotic Manipulation Skills,” *arXiv preprint arXiv:2010.13766*, 2020.
- [25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, \. Kaiser, and I. Polosukhin, “Attention is All You Need,” in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.
- [26] M. Volpp, F. Flürenbrock, L. Grossberger, C. Daniel, and G. Neumann, “Bayesian Context Aggregation for Neural Processes,” in *International Conference on Learning Representations*, 2020.
- [27] W. Xia, Y. Zhang, Y. Yang, J. Xue, B. Zhou, and M. Yang, “Gan inversion: A survey,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 45, no. 03, pp. 3121–3138, mar 2023.
- [28] Y. Zhou, J. Gao, and T. Asfour, “Learning Via-Point Movement Primitives with Inter-and Extrapolation Capabilities,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 4301–4308.

## APPENDIX

In Section III-C, we explained that the trajectories predicted by the model are not necessarily compliant with the desired via-point. For this reason, we provide a further optimization step at deployment time, by applying gradient descent on the parameters of the latent distribution (mean and variance) to find a distribution of movements that best matches the desired via points.

At this point one can question the role of model prediction, since the parameters of the distribution can be found at deployment time. The model prediction serves as a good

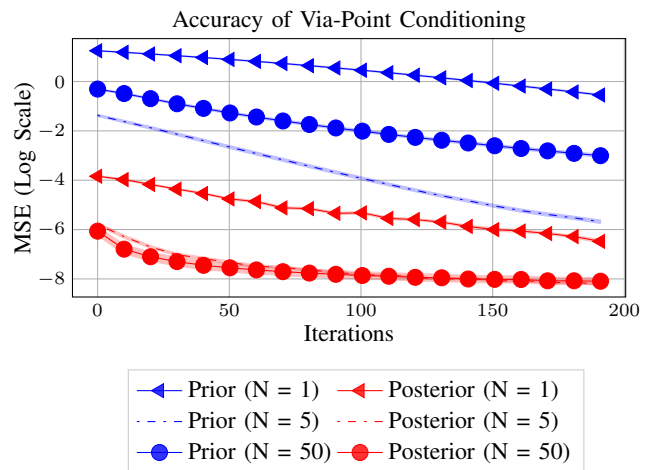


Fig. 7. Post-processing optimization results for generating trajectories that pass through via-points. We compare  $N = 1, 5, 50$  context points. For each setting we use fifty different examples and average performance over 200 iterations. We plot the standard-error bars to give confidence in the mean performance. Posterior is the predicted parameters with the variational posterior and Prior uses Isotropic Gaussian parameters and initialization.

initialization for the gradient descent optimization. In the experiment in Fig 2 and 7, we trained DeepProMPs on the Close Box task. We then queried our model to satisfy  $N = \{1, 5, 50\}$  via points. For a small number of via points (e.g., 1 and 5) our model does not predict good trajectories, as it is also possible to observe in Fig. 2. Gradient descent allows refining such prediction. If we initialize gradient descent with the prior distribution  $\mathcal{N}(0, I)$ , the optimization procedure will need many training steps to find trajectories compliant with the via points. But when initialized with the distribution predicted by the model, the gradient optimization can find trajectories that meet the via-points with accuracy ranging  $10^{-6} - 10^{-8}$  radians in only 200 steps, as shown in Fig. 7.