Temporally-Continuous Probabilistic Prediction using Polynomial Trajectory Parameterization

Zhaoen Su
Uber ATG
suzhaoen@uber.com

Chao Wang
Uber ATG
chao.wang@uber.com

Henggang Cui
Uber ATG
hcui2@uber.com

Nemanja Djuric
Uber ATG
ndjuric@uber.com

Carlos Vallespi-Gonzalez
Uber ATG
cvallespi@uber.com

David Bradley
Uber ATG
dbradley@uber.com

Abstract

A commonly-used representation for motion prediction of actors is a sequence of waypoints (comprising positions and orientations) for each actor at discrete future time-points. While this approach is simple and flexible, it can exhibit unrealistic higher-order derivatives (such as acceleration) and approximation errors at intermediate time steps. To address this issue we propose a simple and general representation for temporally continuous probabilistic trajectory prediction that is based on polynomial trajectory parameterization. We evaluate the proposed representation on supervised trajectory prediction tasks using two large self-driving data sets. The results show realistic higher-order derivatives and better accuracy at interpolated time-points, as well as the benefits of the inferred noise distributions over the trajectories. Extensive experimental studies based on existing state-of-the-art models demonstrate the effectiveness of the proposed approach relative to other representations in predicting the future motions of vehicle, bicyclist, and pedestrian traffic actors.

1 Introduction

In robotics in general and self-driving vehicle (SDV) applications in particular, anticipating the motion of other actors around the robot plays a critical role in planning safe paths to navigate the environment (1). Recently, significant improvements have come from exploring the input representation of the sensor data (2; 3; 4; 5; 6) and the neural network structures (7; 8; 9). Likewise, the output representation for trajectories has seen extensions to account for multimodality (10; 11; 12) and for modeling probability distributions over their future locations (9; 13). However, these output representations generally provide predictions of locations only at discrete and prefixed time-points which may also lack the physics constraints that govern object motion in the real world.

Prediction representations should offer enough flexibility to approximate the motion of various actors types, while still providing regularization that encourages physical realism in the predicted actor motion. Physical realism, such as realistic velocities and accelerations in the trajectories, is particularly important for safety-critical applications. Some representation choices provide such regularization (14), but also make learning more difficult by creating a more complex optimization surface. Additionally, prediction representations should be able to express multi-modal probability distributions over trajectories that reflect the uncertainty of the prediction. In robotics systems, trajectory predictions are often used to compare the probable future positions of other actors against possible future trajectories for the robot in order to find an efficient and low-risk trajectory for the

Machine Learning for Autonomous Driving Workshop at the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

robot (15; 16). Many algorithms for this collision checking can be made more computationally efficient if actor locations can be accessed at arbitrary time-points in parallel within the prediction horizon.

In this paper we propose an approach for trajectory prediction that exhibits these desirable properties. It expresses the time-varying distributions over actor future motion in terms of rigid transformations parameterized with polynomials. We show that low-order polynomials are effective for accurately representing labeled motion of various actor types. When they are applied in extensive supervised learning tasks on two large-scale SDV data sets, the comparison of prediction performance shows that low-order polynomials are as effective as other representations at the fixed time-points, while providing better prediction accuracy on interpolated time-points, low-count actors, and learning tasks with large supervision time intervals. Moreover, we demonstrate that this representation implicitly provides effective regularization that improves physical realism in the predicted actor motion.

2 Related work

Commonly-used representations for trajectory predictions include waypoints and occupancy maps. The waypoint approach describes the probable future locations of an actor at some fixed, usually periodic, time-points (1). In order to take the multimodality into account, multiple trajectories can be predicted for an actor (10; 11; 12). When the uncertainty of the prediction is considered, a spatial probability distribution is provided at each of the given time-points independently (9; 13). The mathematical details can also be found in the following section. The occupancy map representation expresses the multimodality and uncertainty of future actor motion by creating a spatially discretized grid around the actor. Each of the grid cells estimates the probability of the actor occupying this cell at a particular time-point (17; 18; 19). Both representations, continuous in the spatial dimensions or not, are discrete temporally, which may lead to suboptimal performance in many real world applications where the actors usually behave smoothly. In this paper, we explore a prediction representation option that is continuous in both the spatial and temporal domains.

Physical realism of motion prediction is another important research topic. Studies in (20; 21) improve the feasibility of human motion prediction by constructing the graph of skeleton joints and applying constraints on the graph edges. Physical realism such as collision avoidance and kinematic feasibility is studied for vehicle trajectory prediction. In the interactive vehicle following scenario, physical models are embedded in a network to avoid collision (22). The authors in (14) build a vehicle kinematic model for the waypoint representation with constraints and regularization to enforce kinematic feasibility. We show that the proposed low-order polynomial representation leads to good physical realism in predicted trajectories, without enforcing additional physical models, constraints, or regularization.

Lastly, it is well-known that low-dimensional parametric approximation is widely used in many scientific and engineering fields, such as the spectrum approximation based on Gaussian, Lorentzian, or Voigt functions in spectroscopy (23; 24; 25), or the value function approximation in reinforcement learning (26; 27; 28). Its benefits such as providing implicit regularization, compact expression, and avoiding the curse of dimensionality, are also well-understood. In this work, we successfully apply this methodology to the task of motion prediction.

3 Approach

3.1 Trajectory representation

The waypoint trajectory representation $\mathcal{P}$ can be generally expressed as a sequence of rigid transformations, SE3 in general or SE2 for 2D applications,

$$\mathcal{P} = \{(\mathbf{T}_t, \mathbf{R}_t)\}, t \in \{0, t_1, t_2, \ldots, T\}, \tag{1}$$

each of which denotes the translation and rotation of the actor at time t. The probabilistic representation describes the full probability distribution in terms of a sequence of $(p(\mathbf{T}_t), p(\mathbf{R}_t))$ each of which denotes the spatial probability densities of the position and orientation at time t. A probability distribution can be described with its sufficient statistics in terms of moments or distribution parameters if it has an analytical expression, denoted as $\mathbf{M}$. Note that the non-probabilistic representation in

(1) can be viewed as a special case where only the zeroth moments (i.e., the means) are considered. While the waypoint representation is very flexible to express an arbitrary trajectory at the predicted time-points, intermediate time-points need to be interpolated, with a common choice being linear interpolation (i.e., the trajectory is interpreted as a linear spline). As linear interpolation introduces approximation errors on accelerating objects, this might be mitigated by predicting many time-points which increases the computational complexity.

We propose another general prediction representation which parameterizes the prediction distributions over time based on polynomial approximation. More specifically, we represent each scalar element m of $\mathbf{M}$ independently with a low-order polynomial function of time as follows,

$$m(t) = f_m\left(\sum_{n=0}^{N_m} a_{m,n}\left(\frac{t}{T}\right)^n\right), \tag{2}$$

where $a_{m,n}$ are the coefficients of the polynomial of degree N_m, and $f_m(\cdot)$ can be an identity function or a function to ensure the validity of $\mathbf{M}$, if necessary. The normalization over the maximum prediction horizon of interest T, is often desired in practice, particularly for large t's. In this paper, we explore the trajectory prediction of traffic actors in SDV applications where each actor can be approximated with a fixed polygon and a time-varying SE2 transformation from the frame of the polygon to a shared *world* frame. The SE2 transformation consists of translation in the x-y plane and yaw rotation around the vertical axis, denoted as $(c_{xt}, c_{y_t}, \sin\theta_t, \cos\theta_t)$. We model each of the components independently with a univariate distribution. Using the Laplace distribution as an example, the probabilistic prediction for a specific component v can be expressed as

$$p_v(t) = \mathcal{L}(v|\mu_v(t), b_v(t)), \tag{3}$$

whose means over time t are parameterized by a polynomial

$$\mu_v(t) = \sum_{n=0}^{N_{\mu_v}} a_{\mu_v,n}\left(\frac{t}{T}\right)^n. \tag{4}$$

One physics insight into the coefficients can be found by noticing that the important moments of the trajectory can be expressed analytically and computed from the coefficients at arbitrary time-points (without finite differencing), including position, velocity, acceleration, lateral acceleration, curvature, etc., for $N_{\mu_v} > 1$. The waypoint representation with linear interpolation, in contrast, is equivalent to using polynomials with $N_{\mu_v} = 1$ over each time interval, and it only allows for direct computation of position. Another view is that predicting the polynomial coefficients is equivalent to predicting $N_{\mu_v} + 1$ control points that determine a trajectory through polynomial interpolation. Besides μ_v, the diversity parameter b_v over time can be parameterized as

$$b_v(t) = \exp\sum_{n=0}^{N_{b_v}} a_{b_v,n}\left(\frac{t}{T}\right)^n. \tag{5}$$

The additional exponential function ensures the positiveness. Note that the generality and simplicity of the representation allow its application in common non-probabilistic and probabilistic prediction models of various spatial distributions, and straightforward replacement of the representation in the models that output waypoints with few changes.

3.2 Label trajectory approximation

The Weierstrass Approximation Theorem (29) states that any continuous function can be approximated to arbitrary accuracy on a closed and bounded interval with a polynomial of sufficiently high degree. On the in-house data set (see its details in the next section), we investigate the maximum approximation error using (4) to fit label trajectories with various low-order polynomials and prediction horizons. The label trajectory of each actor in the data set can be represented by a bounding box of fixed size with a sequence of SE2 transformations at 10Hz. We parameterize the four components (c_x, c_y, $\sin\theta$, $\cos\theta$) of the time-varying SE2 transform as follows, using two polynomials for the centroid translation (i.e., c_x and c_y) and two polynomials which are normalized to produce the sine and cosine components of the rotation. We find the best-fit polynomials by minimizing the total L2

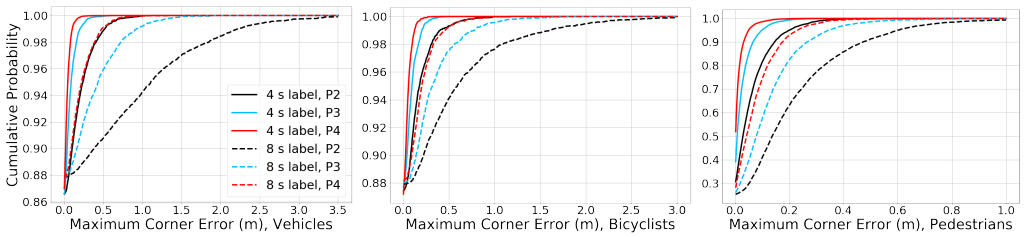

Figure 1: Approximation errors of polynomial representation to fit label trajectories. Cumulative fraction as a function of max corner errors for 4s and 8s trajectories of vehicles, bicyclists, and pedestrians, with polynomials of degrees 2-4 (P2-4)

error between the labeled bounding box corner and the associated fitted corner over all corners and all 4-second time-points at 10Hz. While for metrics, we compute the maximum corner error for each trajectory defined as the maximum L2 distance over all corners and all time-points, which examines the worst approximation. Note this error metric includes both translation and orientation errors, and bounds the maximum error that a collision-checking algorithm could encounter.

Fig. 1 shows the maximum corner error using polynomials of degrees 2 through 4 to fit label trajectories over 4s and 8s time horizons, for vehicles, bicyclists, and pedestrians. As expected, polynomials of higher orders yield lower approximation errors. However, what is surprising is how quickly the maximum corner error drops below the average predicted centroid displacement error as the order of the polynomial is increased, suggesting that even very low-order polynomials are able to represent traffic actor motion with small approximation error relative to the expected prediction error of state-of-the-art prediction models. For instance, the average displacement error for vehicles at 8s is approximately 1.3m (see Table 2), and 96.0% of quadratic trajectories and 99.5% of cubic trajectories have maximum corner error less than that value. Further analysis shows that the high approximation errors occur mainly on maneuvers that have high jerk, or switch between being stopped and moving. For models that cannot achieve good predictions for such hard maneuvers, the low-order polynomials would not hurt; higher-order polynomials or splines of polynomials can be used to provide more representational capacity when necessary. This paper focuses on low-order polynomials as the results in Table 2 show no significant gains in prediction accuracy between 3rd and 4th order polynomials for prediction horizons of up to 8s.

3.3 Applying the representation in supervised learning

Next, we study the proposed representation in supervised trajectory prediction tasks by replacing the waypoint representation with the polynomial representation using (3), and compare prediction performances using the different representations. We adapt MultiXNet(9), which is a deep model with competitive performance designed to detect traffic actors around a SDV and predict their future trajectories. Like most works in trajectory prediction (30; 10; 31; 32; 33; 34; 35), MultiXNet outputs trajectory prediction at periodic time-points $t \in \{0, \tau, 2\tau, \ldots, n\tau\}$. The model assumes a univariate Laplace distribution for each of the $(c_x, c_y, \sin\theta, \cos\theta)$ components at each of the $n+1$ time-points, i.e, $(n+1) \times 4 \times 2$ values are regressed for the means (μ's) and diversity parameters (b's). By contrast, in models with the polynomial representation, the regression values are the $N_\mu + 1$ coefficients in (4) and the $N_b + 1$ coefficients in (5), for the four components individually. If multimodal prediction is modeled, such as for the vehicles in MultiXNet, independent polynomial representation is applied to each separate mode.

Note that the data sets used in this paper provide prediction label as bounding boxes at periodic time-points. To train the polynomial models, we sample waypoints from the polynomial prediction at the same time-points as the waypoint models, and the regression loss is applied to these sampled waypoints. While it would be possible to compute the target polynomial coefficients by fitting the label trajectories and define regression loss on the polynomial coefficients directly, our goal in this study is to compare representations with minimal modeling differences, therefore we keep the regression targets and loss functions unchanged.

4 Evaluation

4.1 Experimental setups

Implementation details. We evaluate the proposed representation by adapting the MultiXNet as the waypoint representation (WP) baseline. In each comparison experiment group, the supervision is provided at the same periodic time-points. The baseline WP model outputs waypoint probabilistic prediction in terms of four univariate Laplace distributions for $(c_{xt}, c_{y_t}, \sin\theta_t, \cos\theta_t)$ at each of the time-points (note however that we do not model heading for pedestrians). To simplify the discussion, we use polynomial representation of a same degree d for all of the four means, denoted as Pd, and polynomial of a same degree k for their diversity parameters, denoted as Pk(b). Using the 8-second prediction as an example, a waypoint model would regress 648 values for one trajectory, while the polynomial representation for ($d = 2$, $k = 1$) would provide 20 values as the model output instead. We use this setting as a default for the polynomial models, unless specified differently. We also implement the vehicle kinematic model (KM) proposed in (14; 36). The KM model outputs the longitudinal acceleration and curvature for each waypoint, clipped within $[-8, 8]$ m/s^2 and $[-0.2, 0.2]$ m^{-1}, respectively, with L2 regularization with weight 0.1 applied on the outputs to encourage smoothness.

Data. The experiments are carried out on the open-sourced nuScenes data (37) (7000 scenes of 20s in the training split with 2Hz annotations) with the results shown in Table 1. We also used the larger Uber ATG in-house data set (14000 scenes of 25s in the training split with 10Hz annotations) throughout the studies, as it yields lower metric variances and the finer label interval facilitates finite difference computation. We explore short-term (4s) and mid-term (8s) prediction for (a) vehicles that are the most common traffic actors, (b) bicyclists that are as rare as about 2% of vehicles in the data sets, and (c) pedestrians whose motion can change abruptly.

Metrics. Detection is evaluated using average-precision (AP) with the intersection-over-union (IoU) matching threshold of 0.5, 0.3, 0.1 for vehicles, bicyclists, and pedestrians, respectively. We report prediction performance in terms of displacement errors (DE's) for centroids, and angle error ($\Delta\theta$'s) for headings. The models within each comparison group have close AP's and are thus not reported. To compare prediction fairly, the prediction metrics are computed with the detection probability threshold set to yield a recall of 0.8 as the operational point for the models trained on the in-house data set, and 0.6, 0.3, and 0.6 for vehicles, bicyclists, and pedestrians, respectively on the nuScenes data set.

4.2 Prediction performance

The first block in Table 1 shows the experiments on nuScenes data with prediction supervision at $t \in \{0, 0.5, 1.0, \ldots, 4.0\}$ seconds. Within the metric variance, P2 and P3 achieve performance similar to WP. P1 has worse performance for vehicles and pedestrians, which we attribute to its lack of representational power. P1 outperforms other models for predicting bicycle centroids, which might be explained by the low population of bicyclists in the data set and the strong regularization provided by polynomials of degree 1. Similarly 4-second prediction is studied on the in-house data set with interval 0.1s, and presented in the second block. With lower metric variance, we again see that the performance of P2 and P3 is close to that of WP for vehicles and pedestrians. The under-performance of P1 for vehicles and pedestrians is further confirmed on this data set. Furthermore, the in-house data set uses a finer supervision interval which might explain why P1 does not outperform the other models on the bicyclists class. Then, we further extend the prediction supervision horizon to 8s (See Table 2). Except for P1, which is too simple to express 8s trajectories, the polynomial models still perform as well as the model using waypoints. To show that the representation works effectively with other model designs and loss functions, we study the representations in single-stage single-modal MultiXNet variants where the second stage network is removed, each actor is modeled with a single trajectory, and the Kullback–Leibler divergence for trajectory regression is replaced by the displacement error without uncertainty learning using the smooth-L1 loss (1). For the models with these common but less optimal techniques, the prediction accuracy is also close using the polynomials and waypoints (see Table 3).

We measure the calibration of the probabilistic predictions using waypoint representation (WP(b)), and polynomials of degrees 0-2 (P0-2(b)) for the diversity parameters. Fig. 2 provides one example by their reliability diagrams (1) for the cross-track dimension at 0s, 2s, and 4s. Except for P0(b), the

Table 1: Four-second models on nuScenes and the in-house data sets with different representations for the means. First block: model comparison on nuScenes data set with a supervision interval of 0.5s. The models using waypoints and polynomials of degrees 1-3 for the means are denoted as WP and P1-3. Second block: model comparison on the in-house data set with a supervision interval of 0.1s. On top of WP models, KM is only applied to vehicle prediction. DE is in meters, and $\Delta\theta$ is in degrees. Lowest errors within metric variance are in bold.

	Vehicles				**Bicyclists**				**Pedestrians**	
Method	**2s DE**	**4s DE**	**2s $\Delta\theta$**	**4s $\Delta\theta$**	**2s DE**	**4s DE**	**2s $\Delta\theta$**	**4s $\Delta\theta$**	**2s DE**	**4s DE**
WP	**0.73**	1.67	**2.39**	**3.33**	**1.6**	3.6	7.7	**10.5**	**0.51**	1.06
P1	0.75	1.74	2.50	3.49	**1.5**	**3.1**	8.0	10.7	**0.51**	**1.05**
P2	**0.73**	1.68	**2.37**	**3.39**	**1.6**	**3.4**	**7.4**	**10.3**	**0.51**	1.06
P3	**0.74**	**1.65**	**2.41**	**3.41**	1.9	3.6	**7.6**	**10.6**	**0.50**	**1.04**
WP	**0.307**	**0.565**	**1.47**	**1.76**	**0.27**	0.52	**5.8**	**6.0**	**0.378**	**0.807**
P1	0.328	0.672	1.51	1.83	**0.27**	**0.51**	6.2	6.4	**0.380**	0.815
P2	0.311	**0.563**	**1.45**	**1.79**	**0.28**	**0.50**	**5.9**	**6.1**	**0.379**	**0.811**
P3	0.311	0.568	1.49	**1.79**	**0.28**	**0.51**	6.1	6.4	**0.380**	0.812
KM	0.313	0.575	1.64	1.80	0.30	0.53	**5.9**	**6.1**	**0.381**	0.812

Table 2: Eight-second models on the in-house data with a supervision interval of 0.1s. The models using waypoints and polynomials of degrees 1-4 for the means are denoted as WP and P1-4.

	Vehicles				**Bicyclists**				**Pedestrians**	
Method	**4s DE**	**8s DE**	**4s $\Delta\theta$**	**8s $\Delta\theta$**	**4s DE**	**8s DE**	**4s $\Delta\theta$**	**8s $\Delta\theta$**	**4s DE**	**8s DE**
WP	**0.580**	1.362	**1.78**	**2.21**	0.70	1.41	**6.5**	6.8	**0.828**	**1.903**
P1	0.684	1.618	1.85	2.36	0.67	1.28	6.7	7.0	0.832	1.926
P2	0.593	1.295	1.83	2.31	**0.58**	**1.13**	6.7	6.9	**0.826**	**1.899**
P3	0.590	**1.291**	1.82	2.28	**0.59**	1.21	**6.5**	**6.7**	**0.827**	**1.899**
P4	0.595	**1.287**	1.82	2.28	**0.60**	1.26	**6.4**	**6.6**	**0.829**	1.913

Table 3: Comparison of using waypoints (WP) and polynomials of degree 2 (P2) on simplified variants that are single-stage and single-modal, and uses displacement errors as the regression losses. The experiments are performed on the in-house data set and have four-second predictions with interval 0.1s.

	Vehicles				**Bicyclists**				**Pedestrians**	
Method	**2s DE**	**4s DE**	**2s $\Delta\theta$**	**4s $\Delta\theta$**	**2s DE**	**4s DE**	**2s $\Delta\theta$**	**4s $\Delta\theta$**	**2s DE**	**4s DE**
WP	**0.337**	**0.653**	**1.68**	**2.05**	**0.29**	**0.52**	**6.5**	**6.8**	**0.414**	0.874
P2	**0.339**	**0.654**	**1.68**	**2.04**	**0.28**	**0.51**	**6.6**	**6.9**	**0.411**	**0.870**

probabilistic predictions are well calibrated in all models, as their curves are close to the reference lines. P0(b) has stationary diversity parameters, which yields under confident predictions at the start of the prediction horizon, and over-confident prediction at the end of the horizon. Notice that polynomials of degree 1 suffice for representing b in these 4-second models.

4.3 Continuous prediction representation

We hypothesize that the polynomial representation improves accuracy over linear interpolation at time-points where supervision is not provided during training. To demonstrate this, we study models with waypoint representations (WP) and polynomials of degree 2 (P2) for the means with supervision at 0s, 2s, and 4s. We compare their performance at 1s, 2s, 3s, and 4s. The predictions of WP at 1s and 3s are computed by linear interpolation. Table 4 focuses on the actors faster than 0.2m/s. For the prediction of vehicles, P2 has slightly better performance at 2s and 4s where regression supervision is available, while it outperforms marginally the interpolated prediction of WP at 1s and 3s. WP has significantly worse performance in centroid prediction of bicyclists at all time-points, which can be explained by the large supervision interval and low population in the training, while the regularization provided by low-order polynomials mitigates those problems. Notice that for predictions of vehicles and bicyclists, the polynomial representation exhibits strength over the waypoint representation even at the fixed time-points, when the supervision time intervals are large, in additional to the better

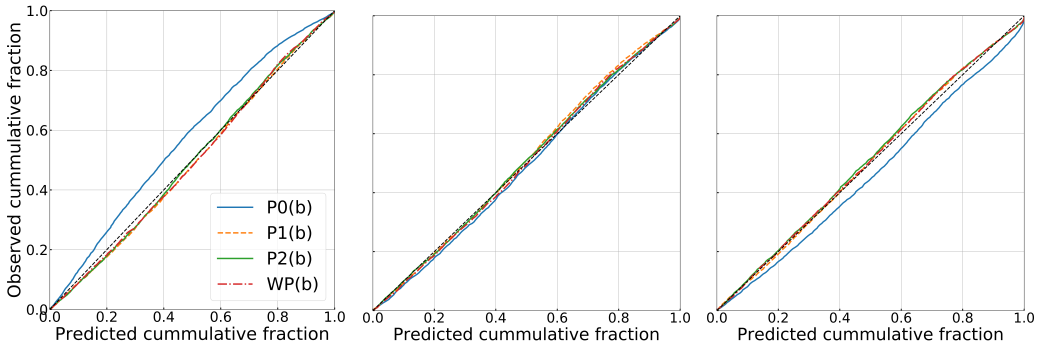

Figure 2: Reliability diagrams for the cross-track dimension at 0s (left), 2s (middle) and 4s (right) of models using waypoints (WP(b)) and polynomials of degrees 0-2 (P0-2(b)) for the diversity parameters.

Table 4: Comparison of the continuous prediction using polynomial of degree 2 (P2) and the discrete prediction using waypoints (WP) for the means on the in-house data set. The regression supervision is provided at 0, 2, and 4s. DE in meters is provided in the first block. $\Delta\theta$ in degrees is in the second block. The predictions at 1 and 3s of WP are computed by linear interpolation. Metrics are computed only for actors that are faster than 0.2m/s.

	Vehicles				**Bicyclists**				**Pedestrians**			
Method	**1s DE**	**2s DE**	**3s DE**	**4s DE**	**1s DE**	**2s DE**	**3s DE**	**4s DE**	**1s DE**	**2s DE**	**3s DE**	**4s DE**
WP	0.99	**1.92**	3.43	4.97	3.9	7.7	11.1	14.2	**0.34**	**0.63**	**0.96**	**1.30**
P2	**0.92**	**1.90**	**3.21**	**4.85**	**2.9**	**5.1**	**6.8**	**8.3**	**0.34**	**0.64**	**0.95**	**1.29**
	1s $\Delta\theta$	**2s** $\Delta\theta$	**3s** $\Delta\theta$	**4s** $\Delta\theta$	**1s** $\Delta\theta$	**2s** $\Delta\theta$	**3s** $\Delta\theta$	**4s** $\Delta\theta$	**1s** $\Delta\theta$	**2s** $\Delta\theta$	**3s** $\Delta\theta$	**4s** $\Delta\theta$
WP	1.97	2.92	4.13	**5.59**	**4.6**	6.6	8.2	9.0	-	-	-	-
P2	**1.87**	**2.81**	**4.06**	**5.56**	**4.6**	**6.4**	**7.9**	**8.7**	-	-	-	-

accuracy at intermediate time-points. The two models perform similarly for pedestrians, suggesting that the second order states for pedestrians are either not captured by the models or less important in pedestrian prediction.

4.4 Physical feasibility

To measure the physical realism of a trajectory, we analyze the maximum and the minimum longitudinal acceleration ($\mathbf{a}_t \cdot \mathbf{v}_t/|\mathbf{v}_t|$), maximum lateral acceleration ($\mathbf{a}_t \times \mathbf{v}_t/|\mathbf{v}_t|$), and maximum lateral speed ($|\mathbf{v}_t \times \mathbf{h}_t|$) over all prediction time-points for each trajectory, where $\mathbf{a}_t$ is the centroid acceleration vector, $\mathbf{v}_t$ is the centroid velocity vector, and $\mathbf{h}_t = (\cos\theta_t, \sin\theta_t)$ is the unit heading vector, measured at time-point t. For motion of common traffic actors these quantities are usually tightly constrained, as shown in Fig. 3 by their distribution for the label trajectories on the in-house data set. We compare this label distribution to the trajectory distributions predicted by models using different representations.

Fig. 3 shows their distributions in normalized histograms that focus on trajectories of non-static vehicles. One can see that the majority of the trajectories of WP have infeasible maximum and minimum accelerations, while the trajectories of label and P2-3 have distributions concentrating near zero. Interestingly, KM trajectories peak at about $\pm 1\text{m/s}^2$ instead of close to zero. In the maximum lateral acceleration plot, a large portion of the WP trajectories have less feasible lateral acceleration, while trajectories of the polynomial models demonstrate closer distributions to that of labels. Because there are no constraints on lateral acceleration in KM, it has considerable amount of trajectories showing unfeasible lateral acceleration. Lastly, the trajectories of WP show the greatest divergence from the label distribution in lateral speed. All trajectories produced by KM have zero lateral speed as it is enforced by the explicit vehicle model. Note that non-zero lateral speed is expected in the label trajectories and those of P2-3 because the speed is computed for the centroids instead of the rear axle centers that are not annotated. Feasible longitudinal acceleration and lateral speed in KM are achieved by hand-crafted constraints and regularization on the controls, which may contribute to unnatural behaviors such as the distribution peaks at around 1m/s^2 and 5m/s^2 in the maximum

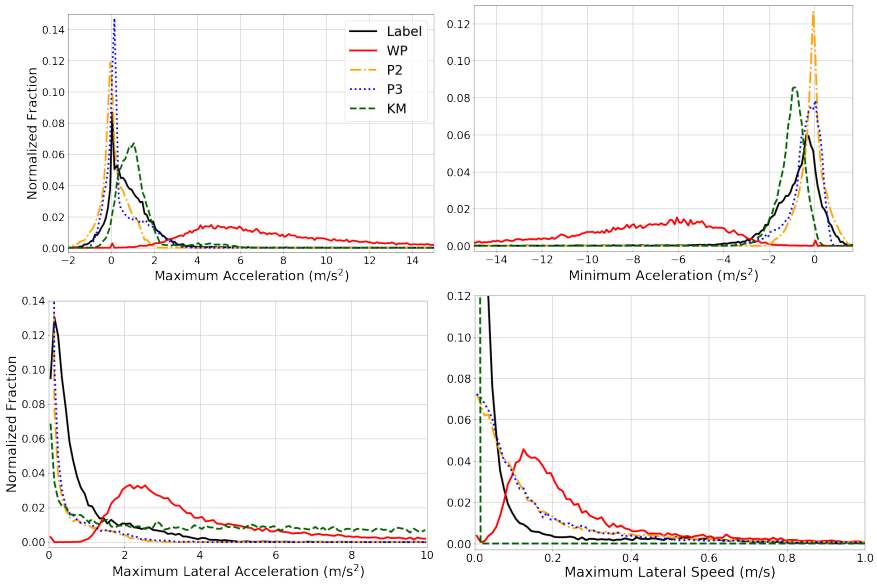

Figure 3: Normalized histograms of maximum acceleration, minimum acceleration, maximum lateral acceleration, and maximum lateral speed of label trajectories (Label) and prediction trajectories by the models (WP, P2, P3 and KM). Y-axis values of the plots are fractions of non-static vehicle trajectories per 0.1, 0.1, 0.1 m/s^2, and 0.01 m/s interval, respectively. The label and all models have 4-second trajectories. The prediction trajectories of P2-3 are sampled with the same 0.1s interval for the metric computation that uses finite difference.

longitudinal distribution, and the prediction performance regression shown in the second block in Table 1. By contrast, the polynomial representation does not require extra regularization to produce physically realistic trajectories.

5 Conclusion

We proposed a simple and general trajectory representation that expresses continuous time-varying probabilistic predictions with polynomial parameterization. Detailed studies show that the polynomial representation is broadly effective, and can outperform the waypoint representation for low-count actors and large temporal supervision intervals. Moreover, we discussed the strength of the parametric representation in providing continuous and precise predictions, which is highly desired in applied robotics systems. Lastly, studying the physical feasibility of the predicted trajectories shows that the polynomial representation exhibits physical realism intrinsically without additional constraints or explicit regularization.

References

[1] N. Djuric, V. Radosavljevic, H. Cui, T. Nguyen, F.-C. Chou, T.-H. Lin, and J. Schneider, "Uncertainty-aware short-term motion prediction of traffic actors for autonomous driving," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020.

[2] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1907–1915.

[3] S. Casas, W. Luo, and R. Urtasun, "Intentnet: Learning to predict intention from raw sensor data," in *Conference on Robot Learning*, 2018, pp. 947–956.

[4] Y. Zhou, P. Sun, Y. Zhang, D. Anguelov, J. Gao, T. Ouyang, J. Guo *et al.*, "End-to-end multi-view fusion for 3d object detection in lidar point clouds," *arXiv preprint arXiv:1910.06528*, 2019.

[5] G. P. Meyer, J. Charland, S. Pandey, A. Laddha, C. Vallespi-Gonzalez, and C. K. Wellington, "Laserflow: Efficient and probabilistic object detection and motion forecasting," *arXiv preprint arXiv:2003.05982*, 2020.

[6] J. Gao, C. Sun, H. Zhao, Y. Shen, D. Anguelov, C. Li, and C. Schmid, "Vectornet: Encoding hd maps and agent dynamics from vectorized representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 525–11 533.

[7] W. Luo, B. Yang, and R. Urtasun, "Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net," in *Proc. of the IEEE CVPR*, 2018, pp. 3569–3577.

[8] Z. Zhang, J. Gao, J. Mao, Y. Liu, D. Anguelov, and C. Li, "Stinet: Spatio-temporal-interactive network for pedestrian detection and trajectory prediction," 2020.

[9] N. Djuric, H. Cui, Z. Su, S. Wu, H. Wang, F.-C. Chou, L. S. Martin, S. Feng, R. Hu, Y. Xu, A. Dayan, S. Zhang, B. C. Becker, G. P. Meyer, C. Vallespi-Gonzalez, and C. K. Wellington, "Multixnet: Multiclass multistage multimodal motion prediction," 2020.

[10] B. Ivanovic and M. Pavone, "The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 2375–2384.

[11] H. Cui, V. Radosavljevic, F.-C. Chou, T.-H. Lin, T. Nguyen, T.-K. Huang, J. Schneider, and N. Djuric, "Multimodal trajectory predictions for autonomous driving using deep convolutional networks," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2090–2096.

[12] H. Zhao, J. Gao, T. Lan, C. Sun, B. Sapp, B. Varadarajan, Y. Shen, Y. Shen, Y. Chai, C. Schmid, C. Li, and D. Anguelov, "Tnt: Target-driven trajectory prediction," 2020.

[13] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 961–971.

[14] H. Cui, T. Nguyen, F.-C. Chou, T.-H. Lin, J. Schneider, D. Bradley, and N. Djuric, "Deep kinematic models for kinematically feasible vehicle trajectory predictions," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10 563–10 569.

[15] W. Xu, J. Pan, J. Wei, and J. M. Dolan, "Motion planning under uncertainty for on-road autonomous driving," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 2507–2512.

[16] P. Raksincharoensak, T. Hasegawa, and M. Nagai, "Motion planning and control of autonomous driving intelligence system based on risk potential optimization framework," *International Journal of Automotive Engineering*, vol. 7, no. AVEC14, pp. 53–60, 2016.

[17] B. Kim, C. M. Kang, S. H. Lee, H. Chae, J. Kim, C. C. Chung, and J. W. Choi, "Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network," *arXiv preprint arXiv:1704.07049*, 2017.

[18] H. Xue, D. Q. Huynh, and M. Reynolds, "Ss-lstm: A hierarchical lstm model for pedestrian trajectory prediction," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 1186–1194.

[19] G. Oh and J.-S. Valois, "Hcnaf: Hyper-conditioned neural autoregressive flow and its application for probabilistic occupancy map forecasting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[20] Q. Cui, H. Sun, and F. Yang, "Learning dynamic relationships for 3d human motion prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6519–6527.

[21] M. Li, S. Chen, Y. Zhao, Y. Zhang, Y. Wang, and Q. Tian, "Dynamic multiscale graph neural networks for 3d skeleton based human motion prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 214–223.

[22] C. Tang, J. Chen, and M. Tomizuka, "Adaptive probabilistic vehicle trajectory prediction through physically feasible bayesian recurrent neural network," *2019 International Conference on Robotics and Automation (ICRA)*, May 2019. [Online]. Available: http://dx.doi.org/10.1109/ICRA.2019.8794130

[23] E. E. Whiting, "An empirical approximation to the voigt profile," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 8, no. 6, pp. 1379–1384, 1968.

[24] Z. Su, A. Zarassi, J.-F. Hsu, P. San-Jose, E. Prada, R. Aguado, E. J. H. Lee, S. Gazibegovic, R. L. M. Op het Veld, D. Car, S. R. Plissard, M. Hocevar, M. Pendharkar, J. S. Lee, J. A. Logan, C. J. Palmstrøm, E. P. A. M. Bakkers, and S. M. Frolov, "Mirage andreev spectra generated by mesoscopic leads in nanowire quantum dots," *Phys. Rev. Lett.*, vol. 121, p. 127705, Sep 2018.

[25] Z. Su, R. Žitko, P. Zhang, H. Wu, D. Car, S. R. Plissard, S. Gazibegovic, G. Badawy, M. Hocevar, J. Chen, E. P. A. M. Bakkers, and S. M. Frolov, "Erasing odd-parity states in semiconductor quantum dots coupled to superconductors," *Phys. Rev. B*, vol. 101, p. 235315, Jun 2020.

[26] L. Baird, "Residual algorithms: Reinforcement learning with function approximation," in *Machine Learning Proceedings 1995*. Elsevier, 1995, pp. 30–37.

[27] G. J. Gordon, "Stable function approximation in dynamic programming," in *Machine Learning Proceedings 1995*. Elsevier, 1995, pp. 261–268.

[28] R. S. Sutton, D. A. McAllester, S. P. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," in *Advances in neural information processing systems*, 2000, pp. 1057–1063.

[29] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*. Cambridge University Press, 1988, pp. 446–448.

[30] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. Torr, and M. Chandraker, "Desire: Distant future prediction in dynamic scenes with interacting agents," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 336–345.

[31] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Multi-agent generative trajectory forecasting with heterogeneous data for control," *arXiv preprint arXiv:2001.03093*, 2020.

[32] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2255–2264.

[33] T. Zhao, Y. Xu, M. Monfort, W. Choi, C. Baker, Y. Zhao, Y. Wang, and Y. N. Wu, "Multi-agent tensor fusion for contextual trajectory prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 126–12 134.

[34] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, "Sophie: An attentive gan for predicting paths compliant to social and physical constraints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1349–1358.

[35] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, H. Rezatofighi, and S. Savarese, "Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks," in *Advances in Neural Information Processing Systems*, 2019, pp. 137–146.

[36] J. Kong, M. Pfeiffer, G. Schildbach, and F. Borrelli, "Kinematic and dynamic vehicle models for autonomous driving control design," in *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2015, pp. 1094–1099.

[37] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 621–11 631.