Trajectories as Markov-States for Long Term Traffic Scene Prediction

Jörg Reichardt

Abstract: We propose the use of traffic participant trajectories with constant time horizon as Markov states for tracking and trajectory prediction. We show how constant time length trajectories can be tracked with minimal computational overhead over kinematic state tracking. The same representation can be used to model future trajectories. In conjunction with multi-object multi-hypotheses tracking architectures it allows for an efficient representation of multi-modal distributions over the future of traffic scenes with heterogeneous participants.

Keywords: Bayes filter, tracking, trajectory prediction, situation interpretation

1 Introduction

Safe and comfortable navigation of an autonomous vehicle necessitates anticipatory planning, i.e. the ability to form expectations and make predictions about the future behavior of dynamic agents in the environment.

The basis for such predictions is an accurate estimate of the present state of dynamic agents based on past observations. Naturally, such state estimates are probabilistic due to uncertainties in the measurement process or unobservable quantities such as agent intent.

State space models are uniquely suited for this task as they provide a solid probabilistic framework to sequentially absorb observations into estimates of the agents' current state and track their motion over time.

The standard technique for this are Bayesian Filter architectures of which the Kalman Filter is the prime example [1]. Under the assumption of Gaussian densities for states and observations together with linear motion and observation models, it yields closed form prediction and update equations that are both numerically efficient and stable. Commonly, kinematic states of agents are tracked. Figure 1 illustrates that these are generally not sufficient to predict the future evolution of a traffic scenario. Many contributions exist that augment the kinematic state space with latent variables modeling agent intent that need to be inferred from observations [2, 3]. Intent is not directly measurable and often not signaled unequivocally and thus has to be inferred from past observations. However, that is a challenging problem as it is not clear how to describe the space of driver intentions. Should one use maneuver intentions that are mutually exclusive and collectively exhaustive? Does that mean one cannot make a lane change while taking a turn? Also, does this apply to all traffic participants? What are the maneuvers of a cyclist or a pedestrian?

*Continental AG, Siemensstr. 12, 93055 Regensburg (e-mail: joerg.reichardt@continental-corporation.com)

Since all the obtainable information about a driver's intent must be in its past behavior, i.e. its past trajectory, it is tempting to simply maintain a list of past kinematic states and infer driver intent on the fly based on this "historic" data. Note that a list of past states is not the same as a list of past observations - state tracking algorithms disentangle the uncertain data associations that arise in multi-object tracking giving unique physical objects unique tracking ids without which single object trajectories cannot be formed.

While promising, using such lists has the obvious drawback of being both memory and compute intensive. This is particularly the case when tracking multiple objects with multi-hypotheses tracking algorithms that potentially maintain hundreds of data association hypotheses in order to be able to track effectively through clutter and occlusion [4, 5, 6]. Since each data association hypothesis effectively represents a different history, one would have to maintain hundreds of lists of past kinematic states.

We seize upon the idea of taking the trajectory of an object as the basis for predicting its future motion. However, instead of maintaining a list of kinematic states, we promote an object's past trajectory over a constant time horizon Δt to its *Markov state*. We assume $\Delta t \gg \delta t$, i.e. much larger than the cycle time δt at which new measurements are obtained and the filter updates are executed. The compatibility of tracked past and hypothesised future trajectory in the light of current evidence then plays the decisive role in estimating the probability of future trajectories.

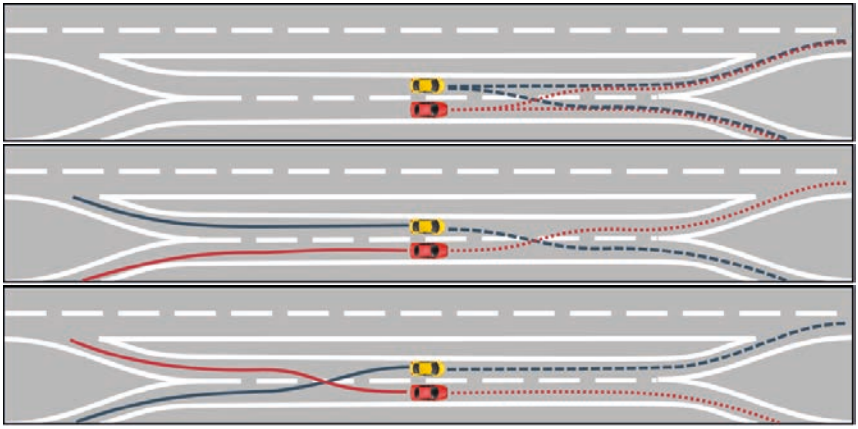

Figure 1: **Top:** Given only kinematic states, each vehicle has 2 plausible path options, resulting in a total of 4 equally likely future scenarios for the traffic scene. **Middle and Bottom:** With past trajectories given, the uncertainty about future scenarios may be largely resolved.

Our contribution is outlined as follows: we introduce a parametric trajectory representation and corresponding linear motion and observation models for its parameters, which, together with Gaussian densities over parameters, yield closed form Kalman update equations. We show that the parameters of this representation have a spatial interpretation. We then employ this representation in predicting future trajectories and finally demonstrate how this representation can be integrated into a multi-hypotheses tracking frame-

work that permits the estimation of multi-modal mixture distributions over the futures of entire traffic scenarios.

It is understood that all results presented are elementary and as such are not new. However, we feel that the exposition unifies a number of existing ideas and possibly provides a fresh perspective and novel synthesis that may still be of interest.

2 Trajectory Tracking

In contrast to paths, which are curves in space, trajectories are curves in space *and* time. The proposed parameterization rests on separating these into a set of *fixed* basis functions of time only, and a parameterization that has spatial semantics.

We choose a set of $n+1$ fixed basis functions $\phi_k(\tau) : \mathbb{R} \to \mathbb{R}$ to describe a trajectory in each of d dimensions of space. We differentiate in $\tau \in \mathbb{R}$ a re-scaled time variable that denotes time along a tracked past trajectory with $\tau = 0$ corresponding to $t - \Delta t$ and $\tau = 1$ corresponding to t. It is convenient to introduce $\mathbf{\Phi}(\tau) \in \mathbb{R}^{n+1}$ as the vector of basis functions $\mathbf{\Phi}(\tau) = [\phi_0(\tau), \phi_1(\tau), \phi_2(\tau), \cdots, \phi_n(\tau)]$ evaluated at τ.

Next, we introduce our state vector $\mathbf{x}_t \in \mathbb{R}^{(n+1)d}$. We can generate any point $\mathbf{c}_t(\tau) \in \mathbb{R}^d$ on a trajectory from linear combinations of the basis functions $\mathbf{c}_t(\tau) = (\mathbf{\Phi}^{\mathrm{T}}(\tau) \otimes \mathbf{I}_d)\mathbf{x}_t$ and thus identify our state variables as the coefficients of this linear combination. Here $\mathbf{I}_d$ is a $d \times d$ identity matrix and $\otimes$ denotes the Kronecker product that simply distributes our basis functions across the d spatial dimensions. An agent's current position at time t will be $\mathbf{c}_t(1)$ and an agent's position at time $t - \Delta t$ will be $\mathbf{c}_t(0)$.

It is instructive to interpret the entries in our state vector $\mathbf{x}_t$ as *control points*, i.e. as having spatial semantics. Separated into individual coordinates and basis vectors $\mathbf{e}_i$ of the d dimensional space we have (dropping the index t for brevity):

$$\mathbf{c}(\tau) = \sum_{i=1}^{d} \sum_{k=0}^{n} \phi_k(\tau) x_{kd+i} \mathbf{e}_i = \sum_{k=0}^{n} \phi_k(\tau) \mathbf{p}_k$$

Thus, a trajectory can also be interpreted as a weighted combination of $(n+1)$ control points $\mathbf{p}_k$ with weights that vary with time via the basis functions. The first d entries in $\mathbf{x}_t$ correspond to $\mathbf{p}_0$, the second d entries correspond to $\mathbf{p}_1$, etc. If we arrange the control points as the rows of a matrix $\mathbf{P} \in \mathbb{R}^{(n+1)\times d}$ we can write most succinctly:

$$\mathbf{c}(\tau) = \mathbf{\Phi}^{\mathrm{T}}(\tau)\mathbf{P}$$

We thus have a compact representation of a trajectory in terms of its control points. We will make use of all notations introduced in our further discussion. We stress the importance of having a parameterization that has spatial semantics: Trajectory prediction and scenario prediction in particular are about the interaction of curves extending through time and space with purely spatial features. Hence, a common footing can provide the basis for effectively modelling the interactions of trajectories and the static environment.

We are free in our choice of basis functions and can even learn them from data. A classical choice are monomials $\phi_k = \tau^k$. Then, we are effectively using a Taylor expansion of the trajectory in each spatial dimension. This provides provable limits on the approximation error if we are able to establish limits on the higher derivatives. This is possible for physical agents due to physical constraints. We can further express preference for

comfortable trajectories by placing zero mean priors on the value of the coefficients corresponding to jerk and higher derivatives [7, 8, 9]. We denote the vector of monomial basis functions by $\mathbf{\Phi}_{\mathrm{M}}(\tau)$. Another classical choice is to use Bernstein Polynomials, effectively using Bézier curves as trajectory/state representation [10, 11, 12, 13]. The corresponding vector of Bernstein Polynomials is $\mathbf{\Phi}_{\mathrm{B}}^{\mathrm{T}}(\tau) = \mathbf{\Phi}_{\mathrm{M}}^{\mathrm{T}}(\tau)\mathbf{M}$ with $\mathbf{M} \in \mathbb{Z}^{(n+1)\times(n+1)}$ defined as [14]:

$$M_{ij} = \begin{cases} \begin{pmatrix} n \\ j-1 \end{pmatrix} \begin{pmatrix} n-j+1 \\ i-j \end{pmatrix} (-1)^{((i+j) \mod 2)} & \text{if } j \leq i \\ 0 \quad \text{otherwise} \end{cases} \tag{1}$$

This choice renders the location of the control points especially intuitive with $\mathbf{p}_0$ and $\mathbf{p}_n$ always at the end points, $\mathbf{p}_0 - \mathbf{p}_1$ and $\mathbf{p}_n - \mathbf{p}_{n-1}$ tangent to the trajectory at the endpoints and the entire trajectory confined to the convex hull of the control points.

Figure 2 shows an example of such a trajectory representation with Bernstein Polynomials as basis functions.

Due to fixed basis functions, we have a linear observation model for $\mathbf{c}_t(\tau)$ and its time derivatives by construction. Our formalism allows to make observations of the trajectory at any point in time. However, for tracking applications, we are most interested in the observation model at $\tau = 1$, the end of the trajectory at current time t. The natural observable is of course $\mathbf{c}_t(\tau)$ and its derivatives with respect to time, i.e. velocities and accelerations. Due to our re-scaling of time, we have $\mathrm{d}t = \Delta t \mathrm{d}\tau$:

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n}\mathbf{c}_t(\tau) = \underbrace{\frac{1}{(\Delta t)^n}\left(\frac{\mathrm{d}^n}{\mathrm{d}\tau^n}\mathbf{\Phi}^{\mathrm{T}}(\tau) \otimes \mathbf{I}_d\right)}_{\mathbf{H}_n(\tau)} \mathbf{x}_t \tag{2}$$

This amounts to a constant observation matrix for every value of τ that can be easily combined. For example, let us assume we are observing an object's position x, y and velocity v_x, v_y organized in an observation vector $\mathbf{o}_t = [x, y, v_x, v_y]$, the corresponding linear observation model $\mathbf{H}$ is then given by stacking the rows of $\mathbf{H}_0$ and $\mathbf{H}_1$ from above. The corresponding observation noise $\mathbf{R}$ can then be used to reflect measurement noise.

Let us stress the benefits of linearity at this point. To determine the $n+1$ linear coefficients necessary for our state representation, only $n+1$ linear measurements at arbitrary times τ are necessary. For example, using $n+1 = 6$ polynomial basis functions, obtaining position, velocity and acceleration at start and end point completely determines the entire trajectory in between. This effectively corresponds to a two point Taylor expansion of the trajectory to degree 2 which results in the same accuracy as a Taylor expansion to degree 4 at a single point [15]. Together with the physically imposed limits on the higher derivatives of the motion of massive objects, our formalism should allow a highly accurate representation of physical trajectories over time horizons Δt of a few seconds. We will come back to this point when discussing trajectory prediction.

Ego motion compensation is equally trivial as the control points transform as points fixed in space under sensor movement. Augmenting the state vector to homogeneous coordinates renders ego motion compensation another linear transform. This enables tracking of the dynamic environment directly from the perspective of the autonomous vehicle, i.e. in ego-coordinates. In moving sensor applications, the ability of being able to transform our state representation to any new coordinate systems is a key advantage

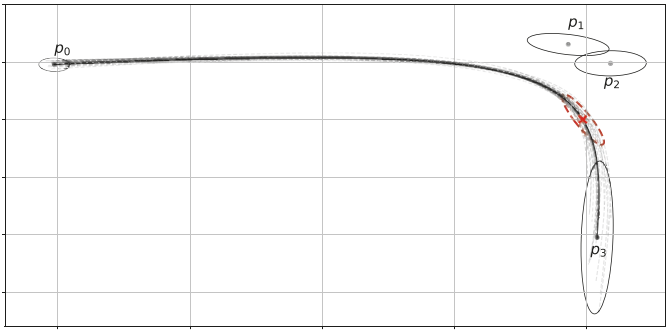

Figure 2: Example of a trajectory representation with $n = 3, d = 2$. The solid line corresponds to the mean trajectory and the dashed lines are samples from the density. Parameters, i.e. control points, are indicated at the mean values and with 95% confidence intervals. The red dashed ellipse encloses a 95% confidence region for the position of the object at $\tau = 0.7$.

over alternative state representations such as LSTM recurrent neural networks that are often employed in trajectory prediction tasks.

The next component we need in a filtering architecture is the motion model. We will consider two ways to derive a linear motion model. First, let us consider the case in which we can find a linear derivative operator $\mathbf{D}$ for our vector of basis functions:

$$\dot{\mathbf{\Phi}}^{\mathrm{T}}(\tau) = \mathbf{\Phi}^{\mathrm{T}}(\tau)\mathbf{D} \tag{3}$$

For the monomial basis functions, this derivative operator is

$$\mathbf{D}_{\mathrm{M}} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 2 & 0 & \cdots \\ 0 & 0 & 0 & 3 & \cdots \\ & & \cdots & & \ddots \end{bmatrix}$$

We now observe the associativity in:

$$\dot{\mathbf{c}}(\tau) = \dot{\mathbf{\Phi}}^{\mathrm{T}}(\tau)\mathbf{P} = (\mathbf{\Phi}^{\mathrm{T}}(\tau)\mathbf{D})\mathbf{P} = \mathbf{\Phi}^{\mathrm{T}}(\tau)(\mathbf{D}\mathbf{P}) = \mathbf{\Phi}^{\mathrm{T}}(\tau)\dot{\mathbf{P}} \tag{4}$$

which gives us a linear motion model for the control points $\dot{\mathbf{P}} = \mathbf{D}\mathbf{P}$ and consequently for our state vector:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = \frac{1}{\Delta t}(\mathbf{D} \otimes \mathbf{I}_d)\mathbf{x} \tag{5}$$

for a given $\delta t = \Delta t \delta \tau$ this can be integrated analytically as a matrix exponential to yield:

$$\begin{aligned} \mathbf{F}_{\delta t} &= \exp\left(\frac{\delta t}{\Delta t}(\mathbf{D} \otimes \mathbf{I}_d)\right) \\ \mathbf{x}_{t+\delta t} &= \mathbf{F}_{\delta t}\mathbf{x}_t \end{aligned} \tag{6}$$

We denote the derivative for Bernstein Polynomials can be obtained from $\mathbf{D}_\mathrm{M}$ via a similarity transform: $\mathbf{D}_\mathrm{B} = \mathbf{M}^{-1}\mathbf{D}_\mathrm{M}\mathbf{M}$.

The second way of generating a motion model is derived from a fit procedure: Assume at time t, we had $m \leq n+1$ samples of a trajectory $\mathbf{c}_t(\tau_i)$ from different times $\tau_i,\ i \in \{1, .., m\}$ between $\tau_1 \geq 0$ and $\tau_m \leq 1$. We arrange these samples as the rows of an $m \times d$ matrix $\mathbf{C}_t$. We now form an $m \times (n+1)$ matrix $\mathbf{B}$ so that row i of $\mathbf{B}$ corresponds to $\mathbf{\Phi}(\tau_i)$. Then, we could estimate the control points, i.e. our state vector, as a Bayesian least square fit to the samples of the trajectory:

$$\mathbf{P}_t = (\mathbf{B}^\mathrm{T}\mathbf{R}^{-1}\mathbf{B} + \mathbf{\Sigma}_\mathrm{P}^{-1})^{-1}\mathbf{B}^\mathrm{T}\mathbf{R}^{-1}\mathbf{C}_t \tag{7}$$

Here $\mathbf{\Sigma}_\mathrm{P}$ is the covariance matrix of a possible zero mean Gaussian prior and $\mathbf{R}$ the covariance of Gaussian observation noise. Note that a zero mean prior is a natural assumption if one of our basis functions is constant as is the case for both $\mathbf{\Phi}_\mathrm{M}$ and correspondingly $\mathbf{\Phi}_\mathrm{B}$. Further, this prior is the natural expression for any inductive biases one may wish to express for the motion model.

With this in mind, the way to propagate a trajectory forward in time by δt now is the following: From the current state estimate, i.e. the control points $\mathbf{P}_t$, we generate samples along the mean of the trajectory at $n+1$ equal spaced $\tau_i' = \delta t/\Delta t + i(1 - \delta t/\Delta t)/n,\ 0 \leq i \leq n$ and then re-estimate the control points from these samples pretending they were obtained at $\tau_i = \tau_i' - \delta t/\Delta t$. To do this, as before, we construct the $(n+1) \times (n+1)$ matrices $\mathbf{B}$ and $\mathbf{B}'$ from $\mathbf{\Phi}(\tau_i)$ and $\mathbf{\Phi}(\tau_i')$, respectively. We then get transformed control points from

$$\mathbf{P}_{t+\delta t} = (\mathbf{B}^\mathrm{T}\mathbf{B} + \mathbf{\Sigma}_\mathrm{P}^{-1})^{-1}\mathbf{B}^\mathrm{T}\mathbf{B}'\mathbf{P}_t \tag{8}$$

Since we are estimating the new control points from samples of the mean, we don't need to consider the observation covariance matrix $\mathbf{R}$ in this expression. Without a prior, this formula reduces to

$$\mathbf{P}_{t+\delta t} = \mathbf{B}^{-1}\mathbf{B}'\mathbf{P} \tag{9}$$

and it should be noted that (9) is consistent with our earlier approach (6) if a linear derivative operator exists:

$$\mathbf{B}^{-1}\mathbf{B}' = \exp\left(\frac{\delta t}{\Delta t}\mathbf{D}\right) \tag{10}$$

We thus have a linear motion model $\mathbf{F}_{\delta t}$ for the control points that only depends on our choice of basis functions and correspondingly for the state vector $\mathbf{x}_t$ we find

$$\begin{aligned} \mathbf{F}_{\delta t} &= ((\mathbf{B}^\mathrm{T}\mathbf{B} + \mathbf{\Sigma}_\mathrm{P}^{-1})^{-1}\mathbf{B}^\mathrm{T}\mathbf{B}') \otimes \mathbf{I}_d \\ \mathbf{x}_{t+\delta t} &= \mathbf{F}_{\delta t}\mathbf{x}_t \end{aligned} \tag{11}$$

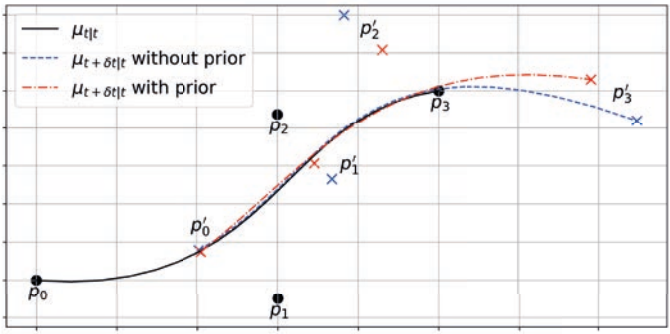

Figure 3: Example of the action of the motion model with $\delta t = 0.3\Delta t$. Note how the motion model without prior follows the trajectory exactly for all points in the past, while the model with prior deviates slightly from the past estimation.

Figure 3 illustrates this on a cubic Bézier curve and exemplifies the differences of the motion model with and without the use of a prior.

With this, we have a compact representation of object trajectories together with linear observation and motion models that enables closed form updates under a Gaussian density assumption. Thus, it can serve as a drop in replacement in any object tracker, both single object or multi-object. Observations at the most recent time point $\tau = 1$ are naturally absorbed into the tracked trajectory via the observation matrix $\mathbf{H}(\tau)$ and the linear motion model $\mathbf{F}_{\delta t}$ propagates the trajectory forward in time.

3 Trajectory Prediction

So far, we only absorb actual observations and thus are limited to track past trajectories. However, if the proposed representation is accurate for past trajectories, then it is equally adequate for future trajectories. For future trajectories, we use the semantics of $\tau = 0$ corresponding to t and $\tau = 1$ corresponding to $t + \Delta t$.

As discussed, for $n+1$ basis functions, we only need $n+1$ linear measurements to fully determine the trajectory over the entire duration of Δt. With $n + 1 = 6$ and position, velocity and acceleration given at the start $\tau = 0$, we only need 3 more observations and the natural choice is position, velocity and acceleration at the end $\tau = 1$. We denote these hypothetical measurements "pseudo-observations" $\mathbf{o}^f$ at $\tau = 1$ to differentiate them from actual observations $\mathbf{o}$ used for tracking. This concept of pseudo-observations at a constant time horizon is motivated by the rich literature on driver modeling [7] and preview control of vehicles [16, 17].

For vehicles, the road topology gives us a finite number of path options in the form of center-lines and we choose to restrict the position of our endpoints to lie on these

path options [18, 19, 20]. We further choose to restrict velocities and accelerations at the endpoints to be tangent to the path option. We thus only have to find the position along the path option, the longitudinal velocity and acceleration. This choice deliberately rules out the possibility for pseudo-observations on lane boundaries for example. For the time horizons considered here, there will generally not be enough evidence to distinguish such a hypothesis from either a completed lane change or a keep lane hypothesis. Note that this does not rule out predicting an object's position away from center-lines at intermediate times between t and $t + \Delta t$ - it just rules this out for pseudo-observations at $t + \Delta t$. If an object indeed crosses a lane boundary at $t + \Delta t$, our measurements will certainly lend support to the hypothesis of a lane change at some later point in time $t' \in (t, t + \Delta t)$ when we consider pseudo-observations at $t' + \Delta t$.

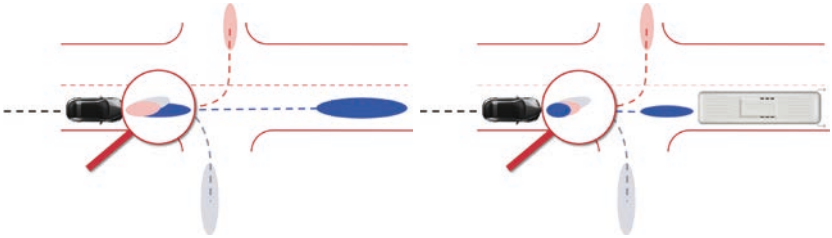

Figure 4: The static and dynamic environment shape the long term path options for a traffic participant. The distribution of vehicle states at $t+\Delta t$ is generally multi-modal and influenced not only by the static environment but also by other traffic participants. This multi-modal structure of the long term options shapes the expected state distribution in short time horizons $t + \delta t$.

Figure 4 illustrates this concept for a single vehicle in two different environmental contexts. In each environment, there are $K = 3$ path options for the vehicle. The pseudo-observations along these path options are uncertain in position, velocity and acceleration. However, since we can assume continuity of trajectories in position, velocity and acceleration, the hypothesized object state at $\tau = 1$ and the current vehicle state at $\tau = 0$ fully define a trajectory and we can propagate back the uncertainty of the pseudo-observation from $\tau = 1$ to $\tau = \delta t/\Delta t$ when we will receive the next actual observation of the vehicle dynamics. Note that due to the multitude of path options, we obtain a *mixture* distribution for expected vehicle dynamics in a short time horizon. This means, conditioned on the K pseudo-observation, we could consider K different motion models:

$$P^k_{t+\delta t|t} = \int \mathrm{d}\mathbf{x}_t P(\mathbf{x}_{t+\delta t}|\mathbf{x}_t, \mathbf{o}^f_k) P_{t|t} \tag{12}$$

The probability, which of the path options is likely taken, is then evaluated in light of new evidence together with a prior probability $P(\mathbf{o}^f_k)$ of the path option (which may be uniform):

$$P(\mathbf{o}^f_k|\mathbf{o}_{t+\delta t}) = \frac{P(\mathbf{o}^f_k) \int \mathrm{d}\mathbf{x}_{t+\delta t} P(\mathbf{o}_{t+\delta t}|\mathbf{x}_{t+\delta t}) P^k_{t+\delta t|t}}{\sum_{k'} P(\mathbf{o}^f_{k'}) \int \mathrm{d}\mathbf{x}_{t+\delta t} P(\mathbf{o}_{t+\delta t}|\mathbf{x}_{t+\delta t}) P^{k'}_{t+\delta t|t}} \tag{13}$$

A sensible choice for the pseudo-observations is to select the furthest point in the path together with velocity and acceleration that is reachable without violating comfort levels [8, 21, 22]. We are free to choose the uncertainties at $\tau = 1$. This approach is similar to [18, 23] but instead of forward generating the hypotheses, we interpolate back from a single step prediction. Another excellent choice is to employ a Gaussian Process conditioned on the past trajectory and the path ahead to provide an estimate of both mean and covariance of our pseudo-observations [24]. The covariances of the pseudo-observations must be large enough as to cover the option space plausibly, but small enough as to be discriminative in the near future.

Figure 5 illustrates this concept numerically in a toy example. A single vehicle approaches the crossing. The road topology provides three path options. We construct plausible end states as pseudo-observations at $t + \Delta t$ with $\Delta t = 5s$ for each of these options together with associated uncertainties. From a common initial condition, we simulate three actual vehicle trajectories along these path options. Note how these actual trajectories deviate from the mean hypothesized trajectories due the holonomic constraints of the vehicle and the cost function of the vehicle controller.

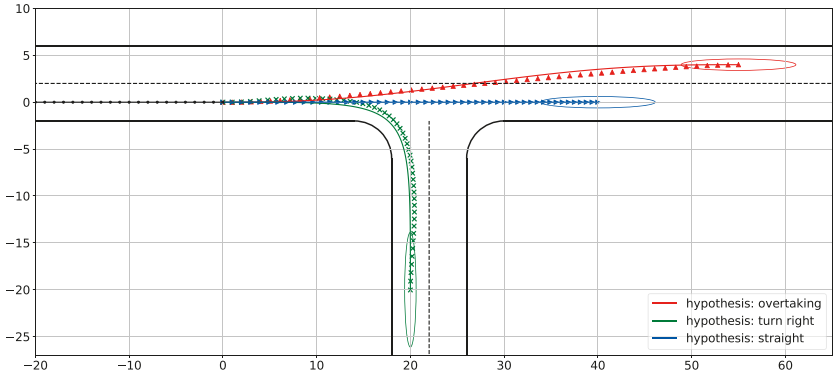

Figure 5: A vehicle approaches a crossing at time $t = 0$ with $v_0 = 10m/s$, $a_0 = 0$ with three hypotheses of future motion: a right turn, going straight across and decelerating possibly due a leading vehicle, and changing lanes and accelerating to overtake the leading vehicle. We show 95% confidence ellipses for pseudo-observations at $t+5s$ and the resulting mean trajectory for each hypothesis. With individual markers, we denote three actual trajectories obtained from simulating a vehicle along these path options for $5s$.

We are now interested at what point along the vehicle trajectory, it is possible to differentiate which of the options is actually chosen. Figure 6 illustrates this. If we only compare the vehicle's position to the expectation from each hypothesis, we reach high confidence at around $1s$ into the scene, (Figure 6, top). If we additionally consider the velocity, high confidence is reached at about $0.5s$ (Figure 6, middle). Also note how the turn right maneuver is competing with the decelerating straight maneuver early into the

scene as both maneuvers are decelerating and the simulated vehicle veers slightly to the left to increase the turn radius before the right turn. The bottom of Figure 6 shows the marginal covariances of the hypothesized trajectories at $t = 1s$ into the scene together with the actual measurement of the three simulated trajectories.

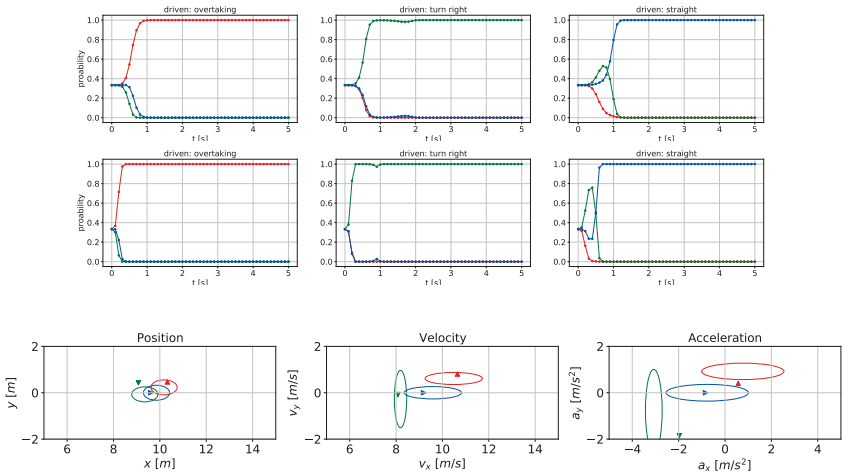

Figure 6: **Top:** The posterior probability of each hypothesis from Figure 5 for each of the actually driven trajectories when considering only position information. **Middle:** When considering both position and velocity. **Bottom:** Distribution of the hypotheses dynamical variables at $t = 1s$ into the situation together with measured quantities from driven trajectories. Colors and markers correspond to Figure 5.

In a live system running at sub second cycle times δt the difference between future options may be too small to differentiate between them, in particular as current state estimates and pseudo-observations are updated at this cycle time, too, and by construction, the end of a tracked trajectory is always 100% compatible with all future trajectories resulting form pseudo-observations. Hence, the current vehicle state does not provide any evidence about which of the future options an object is going to take. We can, however, check the compatibility of the past trajectory with a future option at the current point in time. For this, we form a transition trajectory from observations made at $t - \alpha\Delta t$ on the past tracked trajectory and at $t+(1-\alpha)\Delta t$ on the hypothesised trajectory. Typical values for $\alpha\Delta t$ are in the range of $0.5s$ identified in our preceding analysis. On this transition trajectory, we find the expectation for the current kinematic object state at $\tau = \alpha$. The higher the compatibility between past and future trajectory, the closer this expectation will match the current kinematic vehicle state. This effectively corresponds to an "observation likelihood" for future trajectories. This approach is similar in spirit to multiple model filters [25], but here, the "models" are constructed from pseudo-observations for given path options.

We can leverage this observation likelihood directly in the framework of a multi-hypotheses tracking algorithm. Before the data association step, we augment each local hypothesis (corresponding to a single tracked object) with its possible path-options and pseudo-observations. The entries in the cost matrix during the data association step are then calculated from the observation likelihood of the transition trajectories as described above [5]. That is, we effectively employ (12) in the calculation of the observation likelihood. The standard data association algorithm (Murty's Algorithm [26, 27]) now produces a ranking of the most likely assignments of object detections to local hypotheses under the constraint that all object detections are assigned and each local hypothesis is assigned at most one observation. Each of these assignments forms a new global hypothesis weighted by its total likelihood and, thus, provides a multi-modal multi-object representation of the scene. We alter this algorithm slightly by adding the additional group constraint that in each global hypothesis, only one of the possible futures for a local hypothesis can be present. This constraint can be fulfilled trivially for the most likely global hypothesis by considering only the maximum entry in each group of future trajectories. As Murty's algorithm generates alternative assignments from this globally optimal solution, we only need to adapt the exchange rules in order to reflect the additional constraints. The resulting ranking of data associations represents a probability weighted multi-modal multi-object representation of the current traffic scene and its future development. The ranking can then be further pruned by applying additional constraints such as traffic rules. The pruning of highly probable hypotheses that violate traffic rules can be taken as a consideration of an imminent violation of traffic rules.

4 Conclusion

We have introduced a parametric representation for object trajectories as Markov states for object tracking together with corresponding linear motion and observation models that allow closed form Kalman filtering under a Gaussian density approximation. Albeit we have illustrated and motivated these concepts primarily with vehicles, they are by no means limited to represent vehicle trajectories. Rather, they apply to all physical traffic participants that cannot change their state of motion arbitrarily fast. We have further shown how this representation can be used in trajectory prediction tasks and how it can be integrated into multi-object multi-hypothesis trackers to represent consistent probabilistic multi-modal distributions over the future of entire traffic scenarios.

5 Acknowledgements

The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Climate Action within the project "KI Wissen – Entwicklung von Methoden für die Einbindung von Wissen in maschinelles Lernen". The author would like to thank the consortium for the successful cooperation and Jörg Dietrich for a careful review of the manuscript.

References

[1] S. S. Haykin, ed., *Kalman filtering and neural networks*. Adaptive and learning systems for signal processing, communications, and control, New York: Wiley, 2001.

[2] T. Gindele, S. Brechtel, and R. Dillmann, "A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments," in *13th International IEEE Conference on Intelligent Transportation Systems*, (Funchal, Madeira Island, Portugal), pp. 1625–1631, IEEE, Sept. 2010.

[3] M. Schreier, V. Willert, and J. Adamy, "An Integrated Approach to Maneuver-Based Trajectory Prediction and Criticality Assessment in Arbitrary Road Environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, pp. 2751–2766, Oct. 2016.

[4] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple Hypothesis Tracking Revisited," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015.

[5] A. F. Garcia-Fernandez, J. L. Williams, K. Granström, and L. Svensson, "Poisson Multi-Bernoulli Mixture Filter: Direct Derivation and Implementation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, pp. 1883–1901, Aug. 2018.

[6] K. Granström, L. Svensson, Y. Xia, J. Williams, and F. García-Fernández, "Poisson Multi-Bernoulli Mixtures for Sets of Trajectories," *arXiv:1912.08718 [cs, eess, stat]*, Dec. 2019. arXiv: 1912.08718.

[7] C. C. Macadam, "Understanding and Modeling the Human Driver," *Vehicle System Dynamics*, vol. 40, pp. 101–134, Jan. 2003.

[8] I. Bae, J. Moon, J. Jhung, H. Suk, T. Kim, H. Park, J. Cha, J. Kim, D. Kim, and S. Kim, "Self-Driving like a Human driver instead of a Robocar: Personalized comfortable driving experience for autonomous vehicles," *arXiv:2001.03908 [cs, eess]*, Jan. 2020. arXiv: 2001.03908.

[9] H. Hayati, D. Eager, A.-M. Pendrill, and H. Alberg, "Jerk within the Context of Science and Engineering—A Systematic Review," *Vibration*, vol. 3, pp. 371–409, Oct. 2020.

[10] F. de Dilectis, D. Mortari, and R. Zanetti, "Bézier Description of Space Trajectories," *Journal of Guidance, Control, and Dynamics*, vol. 39, pp. 2535–2539, Nov. 2016.

[11] J. J. Faraway, M. P. Reed, and J. Wang, "Modelling three-dimensional trajectories by using Bézier curves with application to hand motion: *Modelling Three-dimensional Trajectories*," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 56, pp. 571–585, Nov. 2007.

[12] R. Hug, W. Hübner, and M. Arens, "Modeling continuous-time stochastic processes using N-Curve mixtures.," *CoRR*, 2019.

[13] M. Werling, J. Ziegler, S. Kammel, and S. Thrun, "Optimal trajectory generation for dynamic street scenarios in a Frenét Frame," in *2010 IEEE International Conference on Robotics and Automation*, (Anchorage, AK), pp. 987–993, IEEE, May 2010.

[14] S. Ray and P. Nataraj, "A Matrix Method for Efficient Computation of Bernstein Coefficients.," *Reliab. Comput.*, vol. 17, no. 1, pp. 40–71, 2012.

[15] R. Estes and E. Lancaster, "Two-point taylor series expansions," tech. rep., Goddard Space Flight Center, Greenbelt, Maryland, 1966.

[16] M. F. Land and D. N. Lee, "Where we look when we steer," *Nature*, vol. 369, pp. 742–744, June 1994.

[17] O. Lappi, "Future path and tangent point models in the visual control of locomotion in curve driving," *Journal of Vision*, vol. 14, pp. 21–21, Oct. 2014.

[18] D. Petrich, T. Dang, D. Kasper, G. Breuel, and C. Stiller, "Map-based long term motion prediction for vehicles in traffic environments," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, (The Hague, Netherlands), pp. 2166–2172, IEEE, Oct. 2013.

[19] C. Yang, M. Bakich, and E. Blasch, "Nonlinear constrained tracking of targets on roads," in *2005 7th International Conference on Information Fusion*, vol. 1, pp. 8–pp, IEEE, 2005.

[20] C. Yang and E. Blasch, "Fusion of tracks with road constraints," tech. rep., AIR FORCE RESEARCH LAB WRIGHT-PATTERSON AFB OH SENSORS DIRECTORATE, 2008.

[21] G. Arechavaleta, J.-P. Laumond, H. Hicheur, and A. Berthoz, "An Optimality Principle Governing Human Walking," *IEEE Transactions on Robotics*, vol. 24, pp. 5–14, Feb. 2008.

[22] A. Takahashi, T. Hongo, Y. Ninomiya, and G. Sugimoto, "Local Path Planning And Motion Control For Agv In Positioning," in *Proceedings. IEEE/RSJ International Workshop on Intelligent Robots and Systems '. (IROS '89) 'The Autonomous Mobile Robots and Its Applications*, (Tsukuba, Japan), pp. 392–397, IEEE, 1989.

[23] Q. Tran and J. Firl, "Modelling of traffic situations at urban intersections with probabilistic non-parametric regression," in *2013 IEEE Intelligent Vehicles Symposium (IV)*, (Gold Coast City, Australia), pp. 334–339, IEEE, June 2013.

[24] J. Wiest, M. Hoffken, U. Kresel, and K. Dietmayer, "Probabilistic trajectory prediction with Gaussian mixture models," in *2012 IEEE Intelligent Vehicles Symposium*, (Alcal de Henares , Madrid, Spain), pp. 141–146, IEEE, June 2012.

[25] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: a survey," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, pp. 103–123, Jan. 1998.

[26] K. G. Murty, “Letter to the Editor—An Algorithm for Ranking all the Assignments in Order of Increasing Cost,” *Operations Research*, vol. 16, pp. 682–687, June 1968.

[27] M. Motro and J. Ghosh, “Scaling data association for hypothesis-oriented mht,” in *2019 22th International Conference on Information Fusion (FUSION)*, pp. 1–8, IEEE, 2019.