Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Parameter identification for nonlinear models from a state-space approach ⋆

Jules Matz [*,**] **Abderazik Birouche** [*] **Benjamin Mourllion** [*]
Fethi Bouziani [**] **Michel Basset** [*]

[*] *Laboratoire IRIMAS EA 7499, Université de Haute-Alsace, 68093 Mulhouse, France, (e-mail: name.surname@uha.fr)*
[**] *PSA Groupe, Centre technique de Belchamp, 25420 Voujeaucourt, France, (e-mail: name.surname@mpsa.com)*

Abstract: A new approach to parameter estimation of dynamical models is proposed. Its objective is to approximate at best the different dynamics of the system, instead of approximating at best the system output in time. This leads to a weighting of the error depending on the samples location in the state-space and input space. A possible implementation is proposed and applied for estimating the parameters of a two degrees of freedom vehicle dynamics model. The proposed approach is shown to better approximate the fast transient dynamics, at the cost of a degraded performance on steady-states.

Keywords: parameter identification, state-space models, nonlinear system, vehicle dynamics.

1. INTRODUCTION

System identification aims at modelling dynamical systems using measurement data. It can be divided into different steps: designing the experiment, collecting the data, choosing the model structure, choosing a fitting criterion to assess the model quality, optimising the model parameters and possibly the model structure according to this criterion (Ljung (1999), Walter and Pronzato (1997)). This criterion is often constructed as a sum of errors between all samples and model outputs (Åström and Eykhoff (1971)), hence the criterion evaluates how closely the model output fits the measurements in time (since measurements are usually obtained periodically in time). However, in most cases all samples do not have equal importance, especially if the measurements are obtained from experiments with input signals not specifically designed for system identification. To address this issue, one can compute the criterion as a weighted sum, for example Walter and Pronzato (1997) consider the case of measurements obtained using a step input: weighting the samples by $t(k)$ favours the fitting of the steady state, while weighting by $1/t(k)$ favours the fitting of the transient. In essence, this simple example is close to the approach presented in this paper.

As mentionned above, classical methods use a quality criterion to evaluate how closely the model output fits the measurements in time. A different approach is proposed, where the model quality is assessed by the fit between the model behavior and the system behavior observed on the measurements. The concept of model behavior is from Polderman and Willems (1997). The proposed approach leads to a similar objective function as usual approaches but with an additional weighting of measurement samples depending on their location in the state and input space.

In the automotive field, simulation of vehicle dynamics is increasingly used to assess the performance of a vehicle. To have meaningful simulations, the model needs to behave as closely as possible to the real vehicle. The model parameters for a particular vehicle are obtained from various sources (test bench, suppliers, other departments, etc.). It is observed that simply inserting these parameter values into the model does not yield simulations close enough to test-track measurements, especially for aggressive sollicitations (e.g. for electronic stability control (ESC) performance evaluation). This error is mostly attributed to the parameter values being imprecise, and a specific attention is devoted to the tyres. Tyre model parameters are usually obtained from bench tests, which are not sufficiently representative of the tyre behavior onto a vehicle on the road (Montrouge et al. (2018)). Also other parameters (e.g. inertia, mass, position of center of gravity, etc.) depends on the operating conditions of the vehicle (number of passenger, payloads, etc.). For all these reasons, a parameter tuning step is done to improve the model accuracy, based on test-track measurements (Porcel and Macchi (2012)). However, signals obtained from test-track measurements only have few samples at the operating points of interest. This motivates for the study of parameter estimation algorithms that are robust to insufficient model structure and that are efficient even with measurement signals having very few samples at the operating conditions of interest.

The paper is structured as follows: first, the parameter estimation framework is introduced and some relevant methods are reviewed. Then the proposed approach is detailed, leading to a parameter estimation method. The difference between the proposed method and classical ones is discussed on an example. Finally, a conclusion and perspectives are drawn.

⋆ Research funded by PSA Groupe and the French National Association of Research and Technology (ANRT).

2. PARAMETER ESTIMATION

2.1 Model

The model class considered here is nonlinear deterministic time-invariant models. Such models can be described in the continuous-time state-space form

$$\dot{x}(t) = f(x(t), u(t), \theta) \tag{1}$$
$$y(t) = h(x(t), \theta) \tag{2}$$

Also the discrete-time form can be used

$$x(k+1) = f_d(x(k), u(k), \theta) \tag{3}$$
$$y(k) = h(x(k), \theta). \tag{4}$$

where $x(k) \in \mathbb{R}^n$ is the state vector, $u(k) \in \mathbb{R}^m$ is the input vector, $y(k) \in \mathbb{R}^p$ is the output vector, and θ is a constant parameter vector.

In this paper the following assumption is made: all states are measured, the output function is

$$y(t) = x(t). \tag{5}$$

This assumption permits to easily develop the proposed approach without additional complexity related to an observer synthesis. The general case where the output function has the form (2) will be dealt with in future work. It is noted that using this assumption is the same as considering the model to be a nonlinear autoregressive exogenous model (NARX) since replacing $y(k) = x(k)$ in (3) yields

$$y(k+1) = f_d(y(k), u(k), \theta). \tag{6}$$

2.2 Parameter estimation framework

The problem of parameter estimation can be seen as an optimisation problem. An objective function $J(\theta)$ is constructed to evaluate how closely the model approximates the system, and the optimal parameter estimate $\hat{\theta}$ is obtained by minimizing $J(\theta)$. Depending on how the error between the model output and the system measurements is defined, and on the assumptions made by the model, different approaches can be employed to tackle the problem of parameter estimation. However, usual approaches have a common feature: the error is considered as a function of time, and the objective function is constructed as the integral over time of some appropriate measure of this error

$$J(\theta) = \int_0^T \ell(e(t, \theta))dt \tag{7}$$

where $e(t, \theta)$ is the error between the model outputs and the measurements, ℓ is a suitable operator expressing how the error penalises the cost function $J(\theta)$, and T is the time length of the measurements. For discrete samples uniformly taken over time, the objective function becomes

$$J(\theta) = \sum_{k=1}^{N} \ell(e(k, \theta)) \tag{8}$$

where $e(k, \theta)$ is the error between the model output and measurements for the k-th sample, N is the total number of samples.

2.3 Prediction error

An often used error definition for e is the prediction error, especially for transfer function and NARX models (Ljung (2002)). It compares the one-step ahead prediction of the model with the measurements, using the error

$$e(k, \theta) = y(k) - f_d(y(k-1), u(k-1), \theta) \tag{9}$$
$$= y(k) - y^{\text{pred}}(k, \theta) \tag{10}$$

where $y^{\text{pred}}(k)$ is the model output prediction at time k. The prediction error method is depicted in Fig. 1, the blocks z^{-1} delay the signal by one sample period. Once all errors $e(1), \ldots, e(N)$ are obtained, the objective function is evaluated and the optimisation algorithm updates the parameter vector θ.

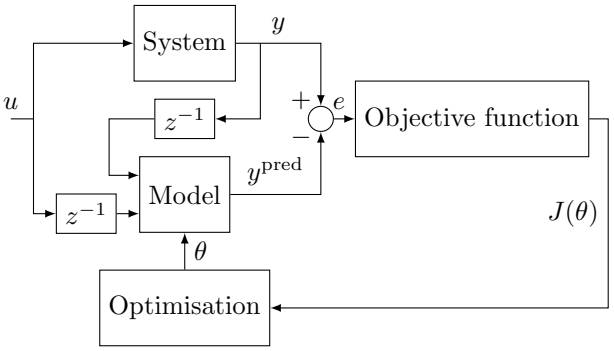


Fig. 1. Scheme of the prediction error method

In the special case where the system is the model plus a gaussian noise $\eta \sim \mathcal{N}(0, \sigma)$ on its output, the maximum likelihood approach can be used to derive an appropriate objective function

$$J(\theta) = \sum_{k=2}^{N} \frac{\left(y(k) - y^{\text{pred}}(k, \theta)\right)^2}{\sigma^2}. \tag{11}$$

In practice the difference between the model and the system is usually different from a gaussian noise, however the objective function (11) is still often used. In the particular case where the output is linear in the parameters to estimate, the parameter vector θ is directly obtained using the least-square method.

The prediction error mehod can be interpreted as follow : the objective is to find θ such that the model (6) agrees at best with the measurements. One way to achieve this is to input the measurements in (6) and optimize θ so the equality in (6) is as verified as possible. To see how well the equality is verified, the error (9) is computed as the difference between the left and right term of (6).

If the continous model (1) is used, for example if the state derivative is measured or approximated, as in Swartz and Bremermann (1975), the equation error method amounts to minimizing the difference between $\dot{x}(k)$ and $f(x(k), u(k), \theta)$ for all samples $k = 1, \ldots, N$. The relation between the continuous and discrete cases can be observed using the Euler discretization, where (1) is approximated by

$$\frac{x(k+1) - x(k)}{t(k+1) - t(k)} = f(x(k), u(k), \theta) \tag{12}$$
$$x(k+1) = \underbrace{x(k) + (t(k+1) - t(k))\, f(x(k), u(k), \theta)}_{f_d(x(k), u(k), \theta)}, \tag{13}$$

hence assessing how well the equality (1) or (3) is verified amounts approximately to the same. Note that here we

used the notation x instead of y as the two quantites are considered to be the same in this paper.

3. PROPOSED METHOD

In this approach, the interpretation given to the parameter estimation problem is "the model should approximate at best the system behavior", and not "the model should approximate at best some specific measurements in time". While classical objective functions equally weights all samples (for measurements uniformly sampled in time), the proposed method attempts at equally weighting all the different dynamics observed on the measurement data. To do this, weights are added, that are function of the samples location in the state-space and input space. The objective function becomes

$$J(\theta) = \sum_{k=1}^{N} w(k)\ \ell(e(k)). \tag{14}$$

The scheme of the method is described in Fig. 2.

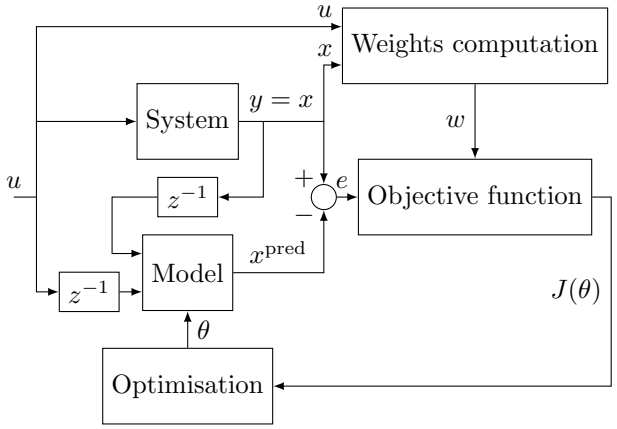


Fig. 2. Scheme of the proposed method

3.1 System behavior approximation

According to Polderman and Willems (1997), the behavior of a dynamical system is defined as the set of signals compatible with the laws governing the system. Similarly, the behavior of a model is defined as the set of signals compatible with the laws governing the model. In order for the model behavior to be as close as possible to the system behavior, the set signals compatible with the laws governing the model should be as close as possible to the set of signals compatible with the laws governing the system. Hence, the laws govering the model should be adapted to be as compatible as possible with the set of signals measured on the system. In the context of parameter estimation, the adapation of the laws governing the model is done by optimising the parameter vector θ.

In the case of a deterministic state-space model of the form (1) with $y = x$, the signals corresponds to (x, u), while the laws governing the system corresponds to $f(\cdot)$ parameterized by θ. The set of signals compatible with the model, that is the behavior of the model, is

$$\mathfrak{B}_{\mathrm{m}}(\theta) = \{(x, u) \mid \dot{x} = f(x, u, \theta), x \in \mathbb{R}^n, u \in \mathbb{R}^m\}. \tag{15}$$

An error function which can be evaluated at different operating points (x, u) can be constructed to assess how well the laws governing the system are compatible with the set of measured signals. The parameter estimation problem becomes

$$\hat{\theta} = \arg\min_{\theta} \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} \ell(e(x, u, \theta))\ dxdu \tag{16}$$

where ℓ is a suitable measure expressing how the error penalises the cost function. The error $e(x, u, \theta)$ gauges the error between the model evolution and the system evolution at the operating point x and for the input u.

3.2 Weights computation

In practice, the integral in (16) cannot be evaluated, as the measurements are a set of finite samples and are not uniformly distributed over the state-space and input space. Measurement samples are usually collected uniformly over time, for a particular initial state of the system and a specific signal input. Consequently, some regions of the state-space and input space are rich in samples, while other only have few, and many do not contain any sample. In practice, to construct an objective function that addresses the problem of approximating the system behavior, the samples are weighted depending on the sample density in the state-space and input space. The samples in densely (resp. sparsely) sampled regions are less (resp. more) weighted. Different methods are tried out to quantify the measurements density around a given sample, they are presented here.

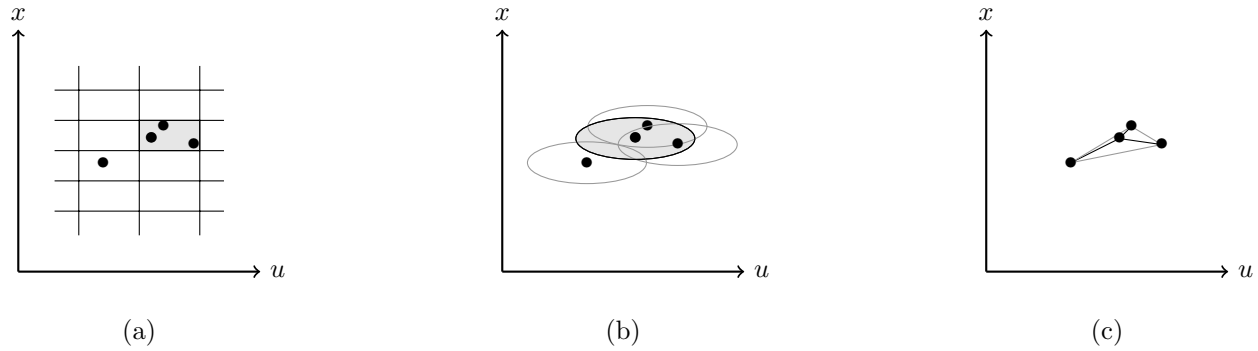
Slicing into a grid. The state and input space is sliced into a grid (Fig. 3 (a)). Each cell of the grid is considered to represent different dynamics of the system. The weights on each sample are computed such that the different grid cells crossed by the system trajectory are equally represented in the objective function. This leads to the weights

$$w(k) = \frac{1}{\#\text{ samples in the same cell}} \tag{17}$$

where the denominator is the number of samples in the same grid cell as the k-th sample. For example in Fig. 3 (a), the weight of the three samples that are in the same cell is $\frac{1}{3}$, while the sample alone in its cell have a weight of 1.

The user has to choose the grid scale. Note that refining at the extreme the grid scale or on the opposite having a way too large grid scale yields a similar objective function. In the case there is only one region, each sample is weighted by $\frac{1}{N}$. In the case of an excessively refined scale, each sample ends up in a different region (it is assumed not two are exactly equal), hence each sample is weighted by 1. The two objective functions are proportionnal, so the obtained $\hat{\theta}$ are the same. It is necessary to find a balance between these two extreme cases.

A slightly less complete variant is to use only one sample per region. Some information is lost since not all measurements are used, but it has the advantage of being faster to compute since the model is only evaluated once for each grid cell. This is especially interesting if samples close to one another do not add much information (low noise and disturbances, good enough model structure). It is a compromise between computational cost and accuracy.



(a) (b) (c)

Fig. 3. Illustration of the different weighting techniques: (a) grid in the state and input space, (b) ellipses around the samples, (c) distances between samples. Samples are represented as •.

One drawback to this method is the absolute positionning of the grid in the state and input space: if the grid is only slightly shifted, it changes the distribution of the samples in the cells. For example, slightly shifting the grid in Fig. 3 (a) to the left would yield the rightmost sample to belong to a different grid cell, changing the weights.

Ellipses around the samples. A different method which do not have the problem of being sensitive to an absolute positioning is detailed here. Similarly to the previous approach, the samples are weighted by

$$w(k) = \frac{1}{\#\text{ samples in the ellipse}} \tag{18}$$

where the denominator is the number of samples in the ellipse drawn around the k-th sample (Fig. 3 (b)). The axes of the ellipse are along the state and input dimensions, and the length of the axes are chosen as a fraction (user-defined ratio) of the total variation of the measured data along each dimension. The two extreme cases of having too large ellipses or too small ones yields the classical objective function, as all sample have same weight (similarly to the grid method).

Distances to all samples. This method does not require parametrisation by the user, contrary to the other ones. In order to measure how different two samples are, the 2-norm is used in the joint state-space and input space (Fig. 3 (c)). Since numerical values of the different states and inputs are expressed on different scales as they represent different physical quantities, some scale factors are introduced. The distance between two samples $(x(k), u(k))$ and $(x(j), u(j))$ is defined as

$$d(k,j) = \sqrt{\sum_{i=1}^{n} \frac{(x_i(k) - x_i(j))^2}{p_i^2} + \sum_{i=1}^{m} \frac{(u_i(k) - u_i(j))^2}{q_i^2}} \tag{19}$$

where $x_i(k)$ is the component of the state of sample $k \in \{1, \dots, N\}$ along the i-th dimension of the state-space. Similarly, $u_i(k)$ is the component of the input of sample k along the i-th dimension of the input space, and p_i and q_i are scale factors. The scale factors for the states are defined as the range of variation of the states along each component

$$p_i = \max_k x_i(k) - \min_k x_i(k)\ , \quad \forall i = 1, \dots, n \tag{20}$$

and similarly for the inputs,

$$q_i = \max_k u_i(k) - \min_k u_i(k)\ , \quad \forall i = 1, \dots, m. \tag{21}$$

The weight for the k-th sample is computed as the sum of the distances from this sample to all the other samples

$$w(k) = \sum_{j=1}^{N} d(k,j) \tag{22}$$

and then the weights are normalized to verify

$$\sum_{k=1}^{N} w(k) = N. \tag{23}$$

In this way, samples that are close to many others (many small distances d) have smaller weights than samples that are isolated (many large distances d).

3.3 Objective function

The objective function (14) is further modified to take into account additional features. First of all, since the error for a single sample is actually a vector of length n, and since the different state derivatives represent different physical quantites in different scales, a normalisation is required. This is done in the same way as (20) and (21). In the case of a prediction error, the scale factors are

$$r_i = \max_k(x_i(k+1) - x_i(k)) - \min_k(x_i(k+1) - x_i(k)) \quad \forall i = 1, \dots, n. \tag{24}$$

The user can also adjust the weights to better approximate some specific state components.
Also another weight can be added to take into account a possible knowledge on the accuracy of the samples, this could prove useful if the accuracy of the samples changes over time or is dependent on the state value (e.g. a sensor's accuracy deteriorates outside its nominal range). Furthermore, some user-designed weights can be added so the model better approximates the system in specific regions of the state-space and input space, for example if the user only wants to run the model around some specific operating points. Finally, using the prediction error (9) for the k-th sample the objective function becomes

$$J(\theta) = \sum_{i=1}^{n} \frac{1}{r_i} \sum_{k=2}^{N} w(k) \left(x_i(k) - f_{d,i}(x(k-1), u(k-1), \theta)\right)^2 \tag{25}$$

where the usual squared error is used, $\ell(\cdot) = (\cdot)^2$.

4. ILLUSTRATIVE EXAMPLE

An example is proposed which emphasizes the difference between the proposed and the classic approaches. In this example, the model structure does not contain the system, and the model can only approximate the system in a very limited way. The considered model is a linear vehicle bicycle model with a nonlinear tyre model. The system measurements are produced using the vehicle simulation software IPG CarMaker (release version 7.0) with the default vehicle model *DemoCar*.

4.1 Vehicle model

A vehicle bicycle model with a nonlinear tyre model is used in this study (Fig. 4). As it has only two degrees of freedom, it is possible to visualise the path of the system in the state-space. This two-wheel model makes strong assumptions such as constant non-zero longitudinal velocity, no weight transfer, vehicle motion in a plane, small angles, etc. The two states of the model are the vehicle slip angle α at its center of mass, and the vehicle yaw rate $\dot{\psi}$.

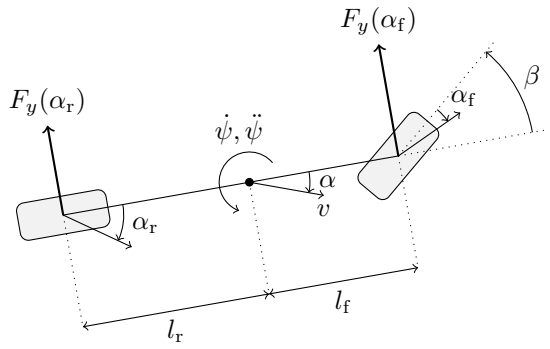


Fig. 4. Vehicle bicycle model

The structure of the model is

$$\dot{\alpha} = \frac{F_y(\alpha_\mathrm{f}) + F_y(\alpha_\mathrm{r})}{M v_x} - \dot{\psi} \tag{26}$$

$$\ddot{\psi} = \frac{l_\mathrm{f} F_y(\alpha_\mathrm{f}) - l_\mathrm{r} F_y(\alpha_\mathrm{r})}{I_{zz}} \tag{27}$$

where α_f and α_r are the slip angle at the front and rear tyres, obtained as

$$\alpha_\mathrm{f} = \alpha + \frac{l_\mathrm{f}\dot{\psi}}{v_x} - \beta \tag{28}$$

$$\alpha_\mathrm{r} = \alpha - \frac{l_\mathrm{r}\dot{\psi}}{v_x} \tag{29}$$

A nonlinear function F_y is used to model the tyres (simplified Magic Formula tyre model, Bakker et al. (1987)):

$$F_y(\alpha) = D \sin\left[C \arctan\left\{(1-E)\, B\alpha + E \arctan\left(B\alpha\right)\right\}\right]. \tag{30}$$

A single set of coefficients B, C, D, E is used for both the front and rear tyres. Due to the tyres operating in different conditions (e.g. different vertical load, longitudinal slip etc.), using two sets of coefficients would yield better results, but the distinction between the different methods is better emphasized this way. The front wheel steering angle β is obtained by multiplying the steering wheel angle u by a constant parameter:

$$\beta = r_\mathrm{steer} u. \tag{31}$$

4.2 Parameter estimation

The simulation contains two manoeuvers: a sine with dwell and a steady state cornering (Fig. 5). The sine with dwell manoeuver lasts about 2.5s, the tyres exhibit a strongly nonlinear behavior. The steady state manoeuver lasts about 100 seconds. The objective function (25) is used with and without the weights $w(k)$, corresponding respectively to the proposed method and the prediction error method. The parameters to be estimated are

$$\theta = [r_\mathrm{steer}, B, C, D, E]. \tag{32}$$

The optimisation is done using Nelder-Mead downhill simplex algorithm implemented in the Matlab function *fminsearch*.

4.3 Results

Different parameter initialisation were tried and yields the same estimated parameters for both methods, although there is no guarantee that it corresponds to a global minimum, for both methods. The bicycle model is then simulated with the obtained parameters, for the same input (Fig. 6). Note that values between time $t = 10$s and $t = 100$s are not shown since input and outputs are constant during this time interval. It can be seen that the prediction error method minimises the average error over the duration of the simulation (Fig. 6), whereas the proposed method minimises the average error over the different dynamics in the state-space and input space (Fig. 7, although input space is not shown). Hence the classical method better approximates the steady state cornering manoeuver as it lasts a long time, and the proposed approach better approximates the sine with dwell manoeuver as it exhibits diverse dynamics. If a better excitation signal had been used, or if the model structure had been more capable of approximating the system, the difference between the two methods would have been reduced. The different methods presented for the computation of sample weights yield similar results, only the method *distance to all samples* is shown on figures.

5. CONCLUSION AND PERSPECTIVES

A different interpretation of the problem of system identification has been proposed, following the concept of *model behavior*. It consists in approximating the different dynamics of the system, instead of approximating all measurement data equally. This approach was implemented using the prediction error method by weighting differently the samples according to their coordinates in the state-space and input space. Different implementations could be used. The proposed approach is relevant if the estimated model is required to approximate the system for various dynamics that may be poorly represented in the measurement data. Future work will address the case of having outputs that are nonlinear functions of the states. Also the weight computation could be improved, as the three methods presented here are purely empirical.

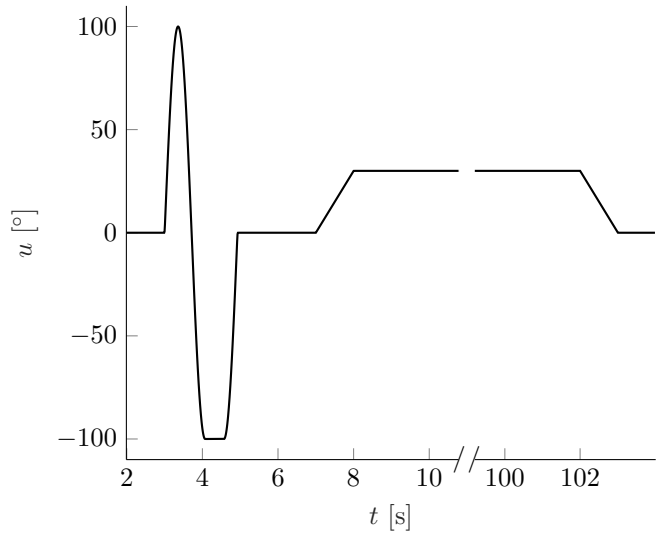


Fig. 5. Simulation input: steering wheel angle u

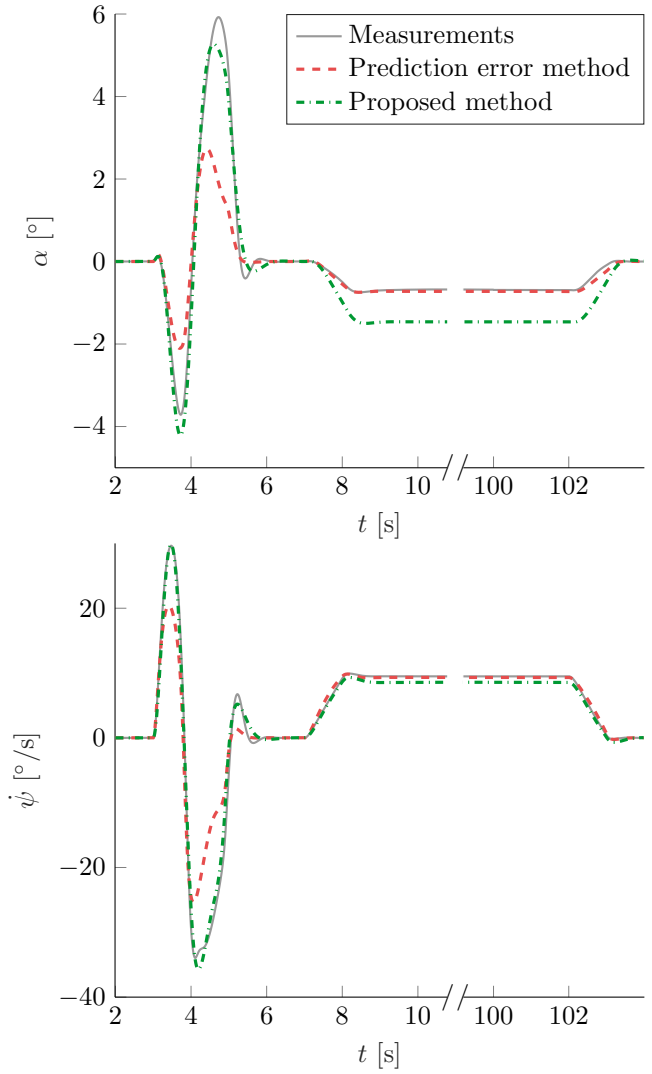


Fig. 6. Simulations outputs in time: yaw rate $\dot{\psi}$ and slip angle α

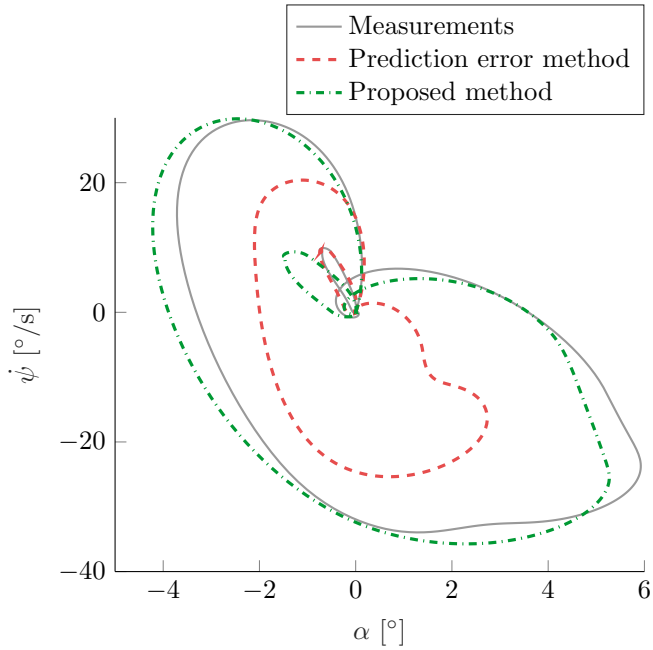


Fig. 7. Trajectories in the state-space

REFERENCES

Bakker, E., Nyborg, L., and Pacejka, H.B. (1987). Tyre modelling for use in vehicle dynamics studies. *SAE Transactions*, 190–204.

Ljung, L. (1999). *System identification: theory for the user*. Prentice Hall information and system sciences series. Prentice Hall PTR.

Ljung, L. (2002). Prediction error estimation methods. *Circuits, Systems and Signal Processing*.

Montrouge, A., Birouche, A., and Lamy, C. (2018). Development of a tire temperature model from test bench measurements. In *International Symposium on Advanced Vehicle Control*.

Polderman, J.W. and Willems, J.C. (1997). *Introduction to mathematical systems theory: a behavioral approach*. Springer.

Porcel, A. and Macchi, O. (2012). First type approval of electronic stability control in passenger cars by means of vehicle dynamics simulation in accordance with ECE 13-H – challenges, innovation and benefits. In *Apply & Innovate 2012*.

Åström, K.J. and Eykhoff, P. (1971). System identification - a survey. *Automatica*, 7(2), 123–162.

Swartz, J. and Bremermann, H. (1975). Discussion of parameter estimation in biological modelling: algorithms for estimation and evaluation of the estimates. *Journal of Mathematical Biology*, 1(3), 241–257.

Walter, E. and Pronzato, L. (1997). *Identification of parametric models from experimental data*. Springer Verlag.